

FUNDAMENTALS OF ASTRODYNAMICS

Karel F. Wakker

**Faculty of Aerospace Engineering
Delft University of Technology**

The document was created in WordPerfect X5 and converted to a PDF document. The compressed PDF document file can be downloaded from <http://repository.tudelft.nl>. As a result of the conversion process, formulas appear in bold and text in normal font when the document is viewed on a computer screen. However, when the file is printed this difference in font will largely disappear.

Published by
Institutional Repository
Library
Delft University of Technology
Delft - The Netherlands

ISBN: 978-94-6186-419-2
UUID: 3fc91471-8e47-4215-af43-718740e6694e

January 2015

Copyright © 2015 by K.F. Wakker

All rights reserved. The Open Access Policy of Delft University of Technology allows individuals to copy parts of this book for personal non-commercial use. If the text is reproduced in any form or is utilized by any information storage and retrieval system, full credit should be given to the author, his affiliation and the title of the book.

PREFACE

This book deals with the motion of the center of mass of a spacecraft, which is an application of the theory of celestial mechanics to spaceflight. This discipline is generally called astrodynamics. Celestial mechanics has always attracted many mathematicians and physicists. A large number of mathematical techniques, which are presently well known and widely used, have been developed specially to solve problems in celestial mechanics. Both in classical celestial mechanics and in astrodynamics it is supposed that when the initial conditions of bodies and the forces acting on them are known with sufficient accuracy, then the motion of the bodies can be computed accurately. Although we know that this picture of deterministic mechanics is theoretically not correct, it is still applicable to solve most problems in celestial mechanics and astrodynamics, and produces results that are in agreement with our observations of the motion of celestial bodies and spacecraft. Over the years, astrodynamics has achieved fantastic and very visible results. The orbit of satellites about the Earth can be computed with centimeter accuracy; spacecraft have explored the solar system, have navigated through the natural satellite systems of Jupiter and Saturn, have landed on the Moon, Mars, Venus and Titan, have performed flybys of and landings on asteroids and comets, and even have left the solar system and have entered interstellar space.

This book focuses on an analytical treatment of the motion of spacecraft and provides insight into the fundamentals of spacecraft orbit dynamics. A large number of topics are treated in a uniform and consistent way. The text is intended for senior undergraduate or graduate engineering students. It is a typical student study book: the knowledge of mathematics and mechanics required from the reader corresponds to that of students having a B.Sc. degree, and full derivations of the formulas are given. In this respect, this book differs from most other books on astrodynamics, in which often useful equations are given but the reader is referred to other books for the derivation of these equations. However, the book is also useful for astrodynamicists and is a valuable resource for anyone interested in astrodynamics.

The text starts with a treatment of the foundation of dynamics. It continues with the classical topics of the many-body problem and the three-body problem, and modern applications of the three-body problem for spaceflight are presented. Then, it is proved that the motion of planets, satellites and interplanetary spacecraft can generally be approximated by a two-body problem. This problem is analyzed in full detail and many useful relations for circular, elliptical, parabolic and hyperbolic motion are derived. Next, the motion of a satellite relative to another satellite is discussed and analyzed. After this, the more modern topic of regularization is treated. Then, the basic astronomical concepts of reference frames, coordinate systems, orbital elements and time are presented, and various topics which are crucial in modern astrodynamics are addressed. Then, the application of rocket engines to change the orbit of a spacecraft is treated, both for coplanar and for three-dimensional maneuvers, and various characteristic transfers from an initial orbit to a final orbit are analyzed. Subsequently, the theory of phasing orbits, which are required to reach a specified position in a final orbit, is presented and various cases are analyzed. Next, rendez-vous flights between two satellites are analyzed. Then, the launching of satellites is discussed and analyzed, as well as the execution of lunar and interplanetary flights, and the flight of spacecraft along low-thrust trajectories. The last four chapters deal with various aspects of orbit perturbations. First, the perturbing forces acting on a satellite are discussed and special and general perturbations methods are described. Then, an elementary analysis of the characteristic effects of the perturbing forces on a satellite orbit is given. The general method of variation of orbital elements is described in detail and an application to orbit maneuvers is presented. Finally, a detailed analysis of orbit perturbations due to the Earth's gravity field is given. The book concludes with three appendices, containing additional information.

The topics of using observations to determine the spacecraft's position and velocity at a particular moment (orbit determination) and to improve the dynamical model applied for the orbit computation are not covered. The main reason being that, although these topics are very important for practical orbit computations, they are primarily applications of statistical estimation theory and less of the theory of dynamics, which forms the backbone of this book. The orbit dynamics methods presented in this book, however, constitute an essential ingredient of any orbit determination procedure.

The text is based on course notes that I have used in various versions since 1976 for the course *Motion of Spacecraft* (in Dutch) and since 1997 for the course *Astrodynamics* (until 2002 in Dutch) for M.Sc. students at the Faculty of Aerospace Engineering of Delft University of Technology. I retired from the university in February 2009, but continued teaching this course until April 2015. During all these years, I have studied many classical and modern books on celestial mechanics and astrodynamics. It is therefore inevitable that some material from these books is duplicated here. Because it is impossible to check where that has been done, I have listed in Appendix A the books which I have studied often. I advise anyone interested in astrodynamics to study these books. They contain much additional information and topics that could not be included in this book.

I like to thank all secretaries who have typed parts of the many versions of the text since 1976. The final editing has been done by me, so I am to be blamed for spelling and grammar errors. I also like to thank Mr. J.A. Jongenelen and Mr. W. Spee of the Faculty of Aerospace Engineering for hand-drawing the original graphs and diagrams, and Mr. A. Pfeifer of SRON Netherlands Institute for Space Research for transferring these hand-drawn graphs and diagrams into digital format.

Karel F. Wakker
emeritus professor of astrodynamics
Faculty of Aerospace Engineering
Delft University of Technology

January 2015

Comments, remarks and questions on this text may be sent to K.F.Wakker@tudelft.nl.

CONTENTS

| | |
|--|-----|
| 1. Some basic concepts | 1 |
| 1.1. Newton's laws of motion | 1 |
| 1.2. Inertial reference frames | 3 |
| 1.3. Deterministic and chaotic motion | 6 |
| 1.4. Newton's law of gravitation | 8 |
| 1.5. Gravity field of a thin spherical shell and a sphere | 13 |
| 1.6. External gravity field of a body with arbitrary mass distribution | 15 |
| 1.7. Maneuvers with rocket thrust | 20 |
| 1.8. Astronomy and the solar system | 24 |
| 2. Many-body problem | 25 |
| 2.1. Integrals of motion | 26 |
| 2.2. Motion relative to the barycenter | 31 |
| 2.3. Polar moment of inertia, angular momentum and energy | 33 |
| 2.4. Evolution of n -body systems | 36 |
| 2.5. Total collision | 41 |
| 2.6. Pseudo-inertial reference frames | 42 |
| 2.7. Angular momentum in the two-body problem | 43 |
| 3. Three-body problem | 45 |
| 3.1. Equations of motion | 45 |
| 3.2. Central configuration solutions | 49 |
| 3.3. Circular restricted three-body problem | 55 |
| 3.4. Jacobi's integral | 60 |
| 3.5. Copenhagen problem | 61 |
| 3.6. Surfaces of Hill | 62 |
| 3.7. Lagrange libration points | 64 |
| 3.8. Motion after leaving a surface of Hill | 68 |
| 3.9. Stability in the libration points | 70 |
| 3.10. Motion about the libration points | 74 |
| 3.11. Application of Jacobi's integral for lunar trajectories | 88 |
| 3.12. Ballistic capture, weak stability boundary and invariant manifold | 91 |
| 3.13. Phenomena at and use of libration points | 97 |
| 4. Relative motion in the many-body problem | 103 |
| 4.1. Equations of motion | 103 |
| 4.2. Relative perturbing acceleration of the Earth and of an Earth satellite | 107 |
| 4.3. Sphere of influence | 112 |
| 5. Two-body problem | 117 |
| 5.1. Conservation laws | 118 |
| 5.2. Shape of the orbit | 122 |
| 5.3. Conic sections | 124 |
| 5.4. Kepler's laws | 126 |
| 5.5. From geocentrism to heliocentrism | 128 |
| 5.6. Velocity components | 135 |
| 5.7. Eccentricity vector | 137 |
| 5.8. Stability of Keplerian orbits | 139 |
| 5.9. Roche limit | 143 |
| 5.10. Relativistic effects | 145 |

| | |
|---|-----|
| 5.11. Solar radiation pressure and the Poynting-Robertson effect | 149 |
| 6. Elliptical and circular orbits | 157 |
| 6.1. Geometry, energy and angular momentum | 157 |
| 6.2. Circular orbit | 160 |
| 6.3. Velocity and orbital period | 162 |
| 6.4. Kepler's third law | 165 |
| 6.5. Kepler's equation | 167 |
| 6.6. Graphical and analytical solution of Kepler's equation | 171 |
| 6.7. Lambert's theorem | 174 |
| 7. Parabolic orbits | 181 |
| 7.1. Escape velocity | 181 |
| 7.2. Flight path angle, total energy and velocity | 183 |
| 7.3. Barker's equation | 184 |
| 7.4. Euler's equation | 187 |
| 8. Hyperbolic orbits | 191 |
| 8.1. Geometry, energy and angular momentum | 191 |
| 8.2. Velocity | 193 |
| 8.3. Relation between position and time | 194 |
| 8.4. Numerical and graphical solution of the transcendental equation | 198 |
| 8.5. Comparison of the expressions for elliptical and hyperbolic orbits | 201 |
| 9. Relative motion of two satellites | 203 |
| 9.1. Clohessy-Wiltshire equations | 203 |
| 9.2. Analytical solution of the Clohessy-Wiltshire equations | 206 |
| 9.3. Characteristics of unperturbed relative motion | 209 |
| 9.4. Relative motion after an impulsive shot | 214 |
| 10. Regularization | 219 |
| 10.1. Singularity and numerical instability | 219 |
| 10.2. Method of Burdet for a Keplerian orbit | 223 |
| 10.3. Generalized time transformation | 226 |
| 10.4. Method of Burdet for a perturbed Keplerian orbit | 227 |
| 10.5. Method of Stiefel for perturbed two-dimensional motion | 230 |
| 10.6. Method of Stiefel for perturbed three-dimensional motion | 234 |
| 11. Reference frames, coordinates, time and orbital elements | 243 |
| 11.1. Position on the Earth's surface | 243 |
| 11.2. Astronomical concepts | 245 |
| 11.3. Topocentric, geocentric and heliocentric reference frames | 250 |
| 11.4. Definition and measurement of time | 252 |
| 11.5. Orbital elements | 260 |
| 11.6. Relation between geocentric and geodetic latitude | 263 |
| 11.7. Transformation from rectangular coordinates to geodetic coordinates | 267 |
| 11.8. Transformation from orbital elements to rectangular coordinates | 268 |
| 11.9. Transformation from spherical coordinates to orbital elements | 274 |
| 11.10. Transformation from rectangular coordinates to orbital elements | 276 |
| 11.11. f - and g -series | 278 |
| 12. Transfer between two coplanar orbits | 285 |
| 12.1. Optimum transfer between two circular orbits | 285 |
| 12.2. Faster transfer between two circular orbits | 291 |
| 12.3. Multiple Hohmann transfer between two circular orbits | 294 |

| | |
|--|-----|
| 12.4. Hohmann transfer from a circular orbit to an elliptical orbit | 299 |
| 12.5. Hohmann transfer from an elliptical orbit to a circular orbit | 301 |
| 12.6. Transfer from a circular orbit to an intersecting elliptical orbit | 304 |
| 12.7. Gravity losses | 306 |
| 13. Transfer between two orbits in different orbital planes | 313 |
| 13.1. Geometry of orbital plane changes | 313 |
| 13.2. Changing Ω without changing i | 315 |
| 13.3. Changing i without changing Ω | 317 |
| 13.4. Transfer between two circular orbits with different inclinations | 318 |
| 13.5. Inclination change at high altitude | 322 |
| 13.6. Orbital plane changes in practice | 326 |
| 14. Phasing orbits | 331 |
| 14.1. Geometric aspects of direct transfer orbits | 332 |
| 14.2. Geometric aspects of low phasing orbits | 336 |
| 14.3. Selection of a low phasing orbit | 339 |
| 14.4. Geometric aspects of high phasing orbits | 344 |
| 14.5. Selection of a high phasing orbit | 347 |
| 15. Rendez-vous flights | 353 |
| 15.1. General rendez-vous aspects | 353 |
| 15.2. Flight from a large distance to the last hold point | 357 |
| 15.3. Flight from the last hold point | 362 |
| 15.4. Last phase of the rendez-vous flight | 366 |
| 15.5. Typical Space Shuttle rendez-vous flight | 369 |
| 15.6. Space Shuttle rendez-vous with ISS | 374 |
| 15.7. ATV rendez-vous with ISS | 376 |
| 16. Launching of a satellite | 379 |
| 16.1. Launch vehicle ascent trajectories | 379 |
| 16.2. Orbit injection and injection constraints | 381 |
| 16.3. Launching of the Space Shuttle | 385 |
| 16.4. Performance of launch vehicles | 387 |
| 16.5. Effects of variations in the in-plane injection parameters | 393 |
| 16.6. Effects of injection errors | 396 |
| 17. Lunar flights | 405 |
| 17.1. Earth-Moon system | 410 |
| 17.2. Two-body lunar trajectories | 413 |
| 17.3. Patched-conic two-dimensional lunar trajectories | 417 |
| 17.4. Three-dimensional lunar trajectories | 423 |
| 17.5. Non-Hohmann low-energy transfer trajectories | 430 |
| 18. Interplanetary flights | 437 |
| 18.1. General aspects of an interplanetary trajectory | 439 |
| 18.2. Launching of interplanetary spacecraft | 442 |
| 18.3. Assumptions and approximations | 444 |
| 18.4. Optimum two-dimensional direct-transfer trajectories | 447 |
| 18.5. Insertion into an orbit about the target planet | 451 |
| 18.6. Faster two-dimensional direct-transfer trajectories | 456 |
| 18.7. Launch opportunities | 459 |
| 18.8. Three-dimensional direct-transfer trajectories | 463 |
| 18.9. Gravity losses | 469 |

| | |
|---|-----|
| 18.10. Approach and flyby of a planet | 471 |
| 18.11. Swingby flights | 476 |
| 18.12. Non-Hohmann low-energy transfer trajectories | 492 |
| 19. Low-thrust trajectories | 497 |
| 19.1. Electric propulsion systems and missions | 498 |
| 19.2. Equations of motion | 508 |
| 19.3. Trajectories with radial thrust | 511 |
| 19.4. Trajectories with tangential thrust | 514 |
| 19.5. Initial part of a trajectory with tangential thrust | 518 |
| 19.6. Approximate solution of an escape trajectory with tangential thrust | 520 |
| 20. Perturbing forces and perturbed satellite orbits | 527 |
| 20.1. Earth's gravitational force | 527 |
| 20.2. Atmospheric drag | 534 |
| 20.3. Gravitational attraction by other celestial bodies | 540 |
| 20.4. Radiation force | 541 |
| 20.5. Electromagnetic force | 544 |
| 20.6. Special and general perturbations methods | 549 |
| 20.7. Historical development of the theory of orbit perturbations | 553 |
| 21. Elementary analysis of orbit perturbations | 555 |
| 21.1. Basic equations | 555 |
| 21.2. Perturbations due to the J_2 -term of the gravity field | 558 |
| 21.3. Perturbations due to lunar and solar attraction | 561 |
| 21.4. Perturbations due to atmospheric drag | 567 |
| 21.5. Perturbations due to the solar radiation force | 571 |
| 21.6. Perturbations due to the electromagnetic force | 575 |
| 21.7. Perturbations due to the $J_{2,2}$ -term of the gravity field | 578 |
| 22. Method of variation of orbital elements | 585 |
| 22.1. Lagrange's planetary equations | 586 |
| 22.2. Canonic form of the planetary equations | 594 |
| 22.3. Use of the mean anomaly | 595 |
| 22.4. Singularities of the planetary equations | 597 |
| 22.5. Gauss' form of the planetary equations | 598 |
| 22.6. Approximate analytical solution of the planetary equations | 603 |
| 22.7. Application of Gauss' equations to orbit maneuvers | 604 |
| 23. Orbit perturbations due to the Earth's gravity field | 613 |
| 23.1. Application of conservation laws | 613 |
| 23.2. Characteristics of the variation of the orbital elements | 615 |
| 23.3. First-order secular variation of the orbital elements | 619 |
| 23.4. Periodic variations of the orbital elements | 628 |
| 23.5. Full spectrum of orbit perturbations | 634 |
| 23.6. Application of Kaula's theory | 645 |
| 23.7. Specialized orbits | 650 |
| Appendix A: Consulted and recommended books | 669 |
| Appendix B: Astronomical and geophysical constants and parameters | 671 |
| Appendix C: Compilation of some expressions for elliptical, parabolic and hyperbolic orbits | 681 |

1. SOME BASIC CONCEPTS

In this book the translational motion of the center of mass of a spacecraft is treated. This topic is an application of a classical branch of astronomy: *celestial mechanics*, to spaceflight; this discipline is generally indicated by the term *astrodynamics*. Celestial mechanics, and thus astrodynamics, is based upon four laws: Newton's three *laws of motions* and Newton's *law of gravitation*. In this Chapter, the laws of motion and the law of gravitation will be discussed in some detail and their application to the computation of the trajectories of spacecraft and the approximative modeling of the gravity field of celestial bodies will be presented. In addition, some basic aspects of the application of rocket engines to changing the trajectory of a spacecraft will be presented.

1.1. Newton's laws of motion

The three laws of motion, which were formulated by I. Newton (1643-1727) in his book *Philosophiae Naturalis Principia Mathematica*, usually abbreviated to *Principia*, in 1687, read in modern terminology:

First law: Every particle continues in its state of rest or uniform motion in a straight line relative to an inertial reference frame, unless it is compelled to change that state by forces acting upon it.

Second law: The time rate of change of linear momentum of a particle relative to an inertial reference frame is proportional to the resultant of all forces acting upon that particle and is collinear with and in the direction of the resultant force.

Third law: If two particles exert forces on each other, these forces are equal in magnitude and opposite in direction (action = reaction).

It is remarkable that Abu Ali al-Hasan (also Al-Haytham; latinized: Alhacen; 965-1039) already enunciated the concept of inertia (Newton's first law of motion) and developed the concept of momentum. Newton's first two laws were, in fact, already known to Galileo Galilei (1564-1642) and C. Huygens (1629-1695), but in Newton's *Principia* they were published for the first time together in a complete and consistent way.

Newton's first law introduces some fundamental concepts: *force*, *particle* (or *point mass*), *time*, *uniform motion* and *inertial reference frame*. Although the concepts of force and time, and to some extent also the concept of point mass, are difficult to fully understand and have some metaphysical aspects, they are widely used and we will therefore assume that they are clear to us. Various aspects of the concept of time will be discussed in Section 11.4, and we define a point mass, or particle, as a body with negligible dimensions but a finite mass and mass density.

Newton calls the laws of motion *axioms* and, after giving each in his *Principia*, makes a few remarks concerning its import. Later writers regard them as inferences from experience, but accept Newton's formulation of them as practically final. A number of writers, among whom is E. Mach (1838-1916), have given profound thought to the fundamental principles of mechanics, and have concluded that they are not only inductions or simply conventions, but that Newton's statement of them is somewhat redundant, and lacks scientific directness and simplicity. Other fundamental laws may be, and indeed have been, employed; but they involve more-difficult mathematical principles at the very start. There is no suggestion, however, that Newton's laws of motion are not in harmony with ordinary astronomical experience, or that they cannot serve as the basis for celestial mechanics. But in some branches of physics certain phenomena are not fully consistent with the Newtonian principles, and they have led A. Einstein (1879-1955) and

others to the development of the so-called *principle of relativity*. The astronomical consequences of this modification of the principles of mechanics are very slight unless the time under consideration is very long.

In the first law, the statement that a particle subject to no forces moves with uniform motion, may be regarded as a definition of time. The second part of the law, which affirms that the motion is in a straight line when the particle is not subjected to forces, may be taken as a definition of a straight line, if it is assumed that it is possible to determine when a particle is subject to no forces. This part may also be taken as showing, together with the first part, whether or not forces are acting, if it is assumed that it is possible to give an independent definition of a straight line. Either alternative leads to troublesome difficulties when an attempt is made to employ strict and consistent definitions.

In the second law, the statement that the rate of change of linear momentum is proportional to the force impressed, may be regarded as a definition of the relation between force and matter by means of which the magnitude of a force, or the amount of matter in a particle, can be measured, according as one or the other is supposed to be independently known. In the statement of the second law it is implied that the effect of a force is exactly the same in whatever condition of rest or of motion the particle may be, and to whatever other forces it may be subject. Hence, the implication in the second law is, if any number of forces act simultaneously on a particle, whether it is at rest or in motion, each force produces the same change of linear momentum that it would produce if it alone acted on the particle at rest. It is apparent that this principle leads to great simplifications of mechanical problems, for in accordance with it the effects of the various forces can be considered separately.

Newton derived in his *Principia* also the parallelogram of forces from the second law of motion. He reasoned that as forces are measured by the accelerations which they produce, the resultant of, say, two forces should be measured by the resultant of their accelerations. One of the most frequent applications of the parallelogram of forces is in the subject of *statics*, which, in itself, does not involve the ideas of motion and time. In it the idea of mass can also be entirely eliminated.

The first two of Newton's laws are sufficient for the determination of the motion of one particle subject to any number of known forces; but another principle is needed when the investigation concerns the motion of a system of two or more particles subject to their mutual interactions. The third law of motion expresses precisely this principle. It is that if one particle presses against another, the second resists the action of the first with the same force. And also, though it is not so easy to conceive it, if one particle acts upon another through any distance, the second reacts upon the first with an equal and oppositely directed force.

In the *Scholium* appended to his discussion on the laws of motion, Newton made some remarks concerning an important feature of the third law. This was first stated in a manner in which it could actually be expressed in mathematical symbols by J.B. le Rond d'Alembert (1717-1783) in 1742. In essence, the statement reads "When a particle is subject to an acceleration, it may be regarded as exerting a force which is equal and opposite to the force by which the acceleration is produced." This may be considered as being true whether the force arises from another particle forming a system with the one under consideration, or has its source exterior to the system. In general, in a system of any number of particles, the resultants of all the applied forces are equal and opposite to the reactions of the respective particles. In other words, the *impressed* forces and the *reactions*, or the *expressed* forces, form systems which are in equilibrium for each particle and for the whole system. This makes the whole science of

dynamics, in form, one of *statics*, and formulates the conditions so that they are expressible in mathematical terms. This phrasing of the third law of motion has been made the starting point for the elegant and very general investigations of J.L. Lagrange (1736-1813) in the subject of dynamics.

1.2. Inertial reference frames

The concept of an inertial reference frame deserves some special attention. It is clear that it is not possible to have a fruitful discussion on motion if one does not define a reference frame with respect to which this motion is described. A very special kind of reference frame is an *inertial reference frame*, also called a *Newtonian reference frame*. The formal definition of an inertial reference frame can be derived from Newton's first law: "An inertial reference frame is a reference frame with respect to which a particle remains at rest or in uniform rectilinear motion if no resultant force acts upon that particle." Thus, one may state that Newton's first law actually defines a reference frame with respect to which Newton's second law is valid. There is, however, a circular reasoning in this formulation. The first law defines the concept of uniform rectilinear motion with the help of the concept of an inertial reference frame, while this inertial reference frame is defined with the help of the concept of uniform rectilinear motion. In this book, this philosophical aspect will not be dealt with and it is assumed that Newton's laws are completely clear to us.

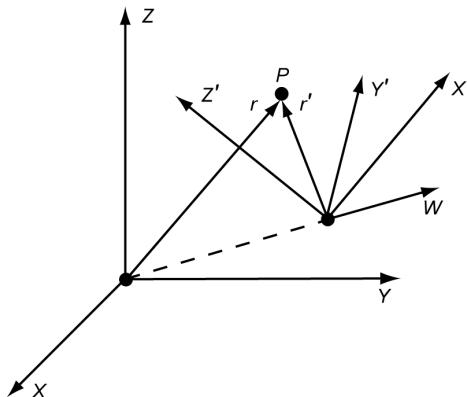


Figure 1.1: Inertial reference frame XYZ and a reference frame $X'Y'Z'$ that moves with constant velocity relative to the XYZ frame.

From the definition of an inertial reference frame it follows that if one inertial reference frame is known, immediately an entire class of inertial reference frames is known; namely all reference frames that perform a uniform rectilinear translational (no rotational) motion with respect to the original inertial reference frame and in which the time differs only by a constant from the time in the original inertial reference frame. This can be proved as follows. Suppose that in Figure 1.1 XYZ is an inertial reference frame and that $X'Y'Z'$ is a non-rotating reference frame that is moving with a constant velocity \bar{W} with respect to XYZ and in which the time differs by a constant T from the time in XYZ . For simplicity, it is assumed that the origins of both reference frames coincide on t_0 . Then, the following transformation relations hold:

$$\bar{r}' = \bar{r} - \bar{W}(t - t_0) ; \quad t' = t + T \quad (1.1)$$

These relations are known as the *Galilei transformations*. For the velocity of particle P in both reference frames we may write

$$\bar{V} = \frac{d\bar{r}}{dt} ; \quad \bar{V}' = \frac{d\bar{r}'}{dt'} \quad (1.2)$$

Combination of (1.1) and (1.2) yields

$$\bar{V}' = \frac{d\bar{r}'}{dt} \frac{dt}{dt'} = \frac{d\bar{r}}{dt} - \bar{W} = \bar{V} - \bar{W} \quad (1.3)$$

Since XYZ is an inertial reference frame, \bar{V} is constant when no (resulting) force is acting on P . Because we have assumed that also \bar{W} is constant, \bar{V}' must be constant too. Using the definition of an inertial reference frame, we thus may conclude that $X'Y'Z'$ is also an inertial reference frame.

Newton's second law expresses the relation between the force, \bar{F} , acting upon a particle and the particle's motion under the influence of this force. In mathematical terms, the law is expressed by

$$\bar{F} = \frac{d}{dt}(m \bar{V}) = \frac{d}{dt}\left(m \frac{d\bar{r}}{dt}\right) \quad (1.4)$$

Naturally, we must require that the second law of motion is invariant when applied in different inertial reference frames. Therefore, in the reference frames XYZ and $X'Y'Z'$ the following relations must hold:

$$\bar{F} = \frac{d}{dt}(m \bar{V}) ; \quad \bar{F} = \frac{d}{dt'}(m \bar{V}') \quad (1.5)$$

where it has been assumed that force and mass are invariant in different inertial reference frames. Substitution of (1.1) and (1.3) into the second equation of (1.5) yields

$$\bar{F} = \frac{d}{dt}(m \bar{V}) - \bar{W} \frac{dm}{dt} \quad (1.6)$$

which shows that Newton's second law is only invariant in the different inertial reference frames if $dm/dt = 0$, i.e. for a particle of *constant mass*. In that case, relation (1.6) can be rewritten as

$$\bar{F} = m \frac{d\bar{V}}{dt} \quad (1.7)$$

This proves that the well-known relation (1.7) is only valid for particles with constant mass and when their motion is considered with respect to an inertial reference frame.

It is noted that in Einstein's *special relativity theory*, the Galilean/Newtonian idea of absolute time running at an equal rate in all inertial reference frames is replaced by the concept that time runs differently in different inertial reference frames, in such a way that the speed of light has the same measured value in all of them. In both Newtonian physics and special relativity theory, inertial reference frames are preferred because physical laws are most simple when written in terms of inertial coordinates. In Einstein's *general relativity theory* time (and even space-time) is influenced not only by velocity but also by gravity fields, and there are no preferred reference frames. However, for an infinitely small space-time region around an observer (considered to be a massless point), one can introduce so-called 'locally inertial reference frames' where, according to Einstein's *equivalence principle*, all physical laws have the same form as in an inertial reference frame in special relativity theory. Such locally inertial reference frames are used to

describe observations taken by the point-like observer.

In reality, spacecraft are not point masses but bodies with finite dimensions. In addition, we often consider the motion of spacecraft relative to a non-inertial reference frame, and the mass of the spacecraft will vary with time when a rocket engine is thrusting. The reformulation of Newton's second law of motion such that it can also be applied in these cases is a major topic of classical theoretical mechanics. As such, it is outside the scope of this book. Therefore, only a brief survey will be given of those aspects that are relevant to astrodynamics.

A spacecraft of finite dimensions can be thought of as a continuous mass system consisting of discrete point masses. When Newton's second law of motion is applied to the motion of this system of point masses relative to an inertial reference frame, we find for a *rigid body*:

$$\bar{F} = M \frac{d\bar{V}_{cm}}{dt} \quad (1.8)$$

where \bar{F} is the net external force acting on the body, M is the total mass of the body and the index cm refers to the center of mass of the body. In deriving this expression, the velocity and acceleration of an element of the body relative to its center of mass (flexibility effects) have been neglected. To good approximation, spacecraft may be considered as rigid bodies, except in cases where the spacecraft is subjected to shocks and e.g. large solar panels may perform oscillatory motion relative to the spacecraft. Planets, moons and other celestial bodies may, to first-order approximation, also be considered as rigid bodies. This means that for all practical cases in celestial mechanics and astrodynamics, where we analyze the translational dynamics (no rotations) of celestial bodies or spacecraft, we may consider the body as a point mass located at the center of mass of the body and with a mass equal to the mass of the body. Therefore, in this book we will use the words 'point mass', 'particle' and 'body' interchangeably.

When the motion of a spacecraft is described relative to a non-inertial reference frame, we can still use Newton's second law of motion, provided that we add suitably selected *apparent forces* to the net *natural force*, \bar{F} . These apparent forces are also called fictitious forces, pseudo forces, d'Alembert forces or inertial forces. In this book we will use the term apparent force exclusively. Four apparent forces are well-known: one caused by a rectilinear acceleration of the origin of the reference frame, two caused by a rotation of the reference frame (centrifugal force and Coriolis force), and a forth caused by a variable rate of rotation of the frame (Euler force). All apparent forces are proportional to the mass of the body upon which they act, which is also true for gravity. This led Einstein to wonder whether gravity was an apparent force as well. He was able to formulate a theory with gravity as an apparent force; the apparent acceleration due to gravity is then attributed to the curvature of space-time. This idea underlies Einstein's *general theory of relativity*. In our analyses, we will always start from the equations of motion relative to an inertial reference frame. The relevant apparent forces will then automatically show up after we have applied the coordinates transformations needed to obtain the equations of motion relative to a non-inertial reference frame.

When a rocket engine on the spacecraft is thrusting, mass is expelled and the mass of the spacecraft is not constant but will decrease with time. However, the mass of the spacecraft plus the rocket engine combustion products, which have been expelled by the rocket engine, is still constant. When we apply Newton's second law of motion to the motion of all point masses constituting this entire mass system relative to an inertial reference frame, we eventually find

$$\bar{F} - \dot{m} \bar{V}_j = M \frac{d\bar{V}_{cm}}{dt} \quad (1.9)$$

where \bar{F} is the net external (natural) force acting on the body, M is the *instantaneous mass* of the spacecraft, \dot{m} is the mass flow leaving the rocket engine nozzle per unit of time, and \bar{V}_j is the effective exhaust velocity (relative to the spacecraft), which consists of an impulsive term and a pressure term. The effective exhaust velocity may, generally, be considered constant, in particular for a rocket engine thrusting outside an atmosphere. Note that the velocity vector \bar{V}_j points away from the spacecraft. The second term on the left-hand side of (1.9) including the minus-sign formally is an apparent force, which has to be included to allow the application of Newton's second law of motion to a spacecraft with a time-varying mass. This force is called the *thrust* of the rocket engine. It has the magnitude

$$F_{thr} = \dot{m} V_j \quad (1.10)$$

and acts in the direction opposite to the flow of linear momentum through the rocket engine nozzle exit area. We conclude that we may apply (1.8) for the motion of a spacecraft with a thrusting rocket engine, if M is considered to represent the instantaneous mass of the spacecraft and if the thrust is considered as a real external force that is added to the other external forces acting on the spacecraft. It is emphasized that in arriving at (1.9) a number of assumptions and approximations had to be introduced: 1) the spacecraft itself is considered as a rigid body; 2) the Coriolis force acting on the combustion product particles due to a rotation of the rocket engine is negligible with respect to the thrust; 3) the flow of combustion products leaving the rocket engine nozzle is stationary; 4) the velocity of the center of mass of the spacecraft relative to the spacecraft body is negligible with respect to the exhaust velocity of the combustion products.

1.3. Deterministic and chaotic motion

A fundamental aspect of Newton's theory, and the basis of classical mechanics, is the idea that the computation of the motion of an object is a deterministic problem. This means that we assume that if we know the initial position and velocity of a body accurately enough and if we can determine the forces acting on the body accurately enough, then we can compute the position and velocity of that body at any given time with high accuracy. It is true that in the past one has realized that an uncertainty in the initial conditions may yield a divergence of the computed position and velocity at a certain time (Section 5.8 and Section 10.1), but this error was still viewed as a deterministic phenomenon. In the last decennia, however, one started to realize that *chaotic behavior* also plays a certain role in mechanical problems. Although, for the classical problems in celestial mechanics and for the time intervals associated with these problems, these chaotic aspects usually are of no importance, a short examination of chaos will be given below in order to provide some understanding of the limitations of deterministic mechanics.

Chaos arises in deterministic systems because of their specific instability. For example, imagine a billiard game. The player sends the ball into the usual array of other balls. The slightest variation in the direction of the original push will send the ball down quite a different path and the difference will not attenuate but will grow with time. Each collision of the balls with each other will further amplify this divergence. To prolong the motion, let us assume that the loss of energy is small. Newton's laws do determine the trajectory of each ball and the sequence of collisions. But the prediction will be completely wrong after a certain number of collisions, even if the initial push is defined with an error as small as the gravitational effect of a single electron

on the margin of the galaxy. The deviation grows exponentially in time, so that prediction is impossible at any level of precision of the initial conditions. If the boards are convex (*Sinay billiard*) even a single ball reduced to a material point will display the same instability. With a multitude of such turning points, a dynamic system may display erratic, complicated behavior, which looks and is called *chaotic*. Though deterministic, it will be unpredictable, because prediction would require paradoxical precision of the initial conditions. This is not an abstract extravaganza. On the contrary, chaotic systems can be surprisingly simple, like the nonlinear pendulum, for example.

In his research on the three-body problem (Chapter 3), J.H. Poincaré (1854-1912) became the first person to discover a chaotic deterministic system. Given the law of gravitational attraction and the initial positions and velocities of the three bodies, the subsequent positions and velocities are fixed; so the three-body system is deterministic. However, Poincaré found that the evolution of such a system is often chaotic in the sense that a small perturbation in the initial state, such as a slight change in one body's initial position, might lead to a radically different later state than would be produced by the unperturbed system. If the slight change is not detectable by our measuring instruments, then we would not be able to predict which final state will occur. So, Poincaré's research proved that the problem of determinism and the problem of predictability are distinct problems. The scientific line of research that Poincaré opened was neglected until 1963, when meteorologist E.N. Lorenz (1917-2008) rediscovered a chaotic deterministic system: the Earth's atmosphere. Earlier, Poincaré had already suggested that the difficulties of reliable weather predicting are due to the intrinsic chaotic behavior of the atmosphere. Amazingly, this kind of chaos does contain inherent regularities. These regularities can be understood, and some integral traits of chaotic behavior can even be predicted. One fundamental regularity in chaotic behavior was discovered by Lorenz in his study on thermal convection in the atmosphere. Figure

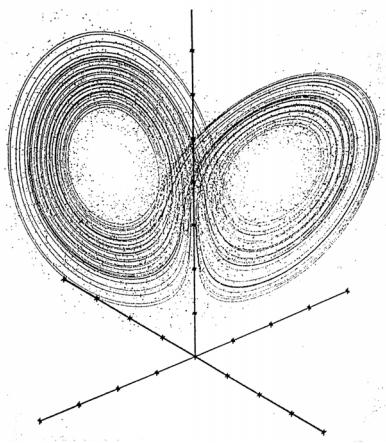


Figure 1.2: Lorenz strange attractor.

1.2 shows a phase space for this system; its three coordinates characterize the intensity of the convection stream, and horizontal and vertical temperature gradients. This means that a point completely defines the state of the system at some moment of time; the evolution in time is defined by a trajectory. The lines (curves) show that trajectory. The (hardly visible) dots in Figure 1.2 show 10,000 different states at some moment of time. They evolved from 10,000 initial states, which were so close that they are all merged into one dot in Figure 1.2 (somewhere in the top right corner). In other words, microscopic initial perturbation leads to macroscopic divergence, and prediction is impossible. However, there exists an inherent regularity: all states eventually congregate around the configuration represented by the lines. These lines are the asymptotic

trajectories; the evolution of the system will gradually be attracted to them. They occupy the subspace called the *chaotic* or *strange attractor*. Nowadays, we know that the shape of such a distribution is fractal-like. However, the scientific study of fractals did not begin until B.B. Mandelbrot's (1924-2010) work in 1975, a century after Poincaré's first insight.

Chaos in the solar system is associated with gravitational resonances. The simplest case of such resonance occurs when the orbital periods of two bodies are about in the ratio of two small integers. The solar system is full of this type of resonance. For instance, the orbital periods of Venus and Earth are in the ratio of about 13:8, of Venus and Mars about 3:1, of Jupiter and Saturn about 5:2, of Uranus and Neptune about 2:1, and of Neptune and Pluto about 3:2. Jupiter's moons Io, Europa and Ganymede have orbital periods in the ratio of 4:2:1 and Saturn's moons Enceladus and Dione in the ratio of about 2:1. There are other more subtle gravitational resonances associated with the precession of the orbits (Section 23.3) of the bodies in addition to their orbital period. Resonances thread the entire solar system in a complex web and it is therefore remarkable that our solar system proved to be rather stable over extended periods of time (Section 5.8).

1.4. Newton's law of gravitation

Partially based on the observed motions of the planets around the Sun, Newton formulated his law of gravitation and published it also in his *Principia*:

- Two particles attract each other with a force directly proportional to their masses and inversely proportional to the square of the distance between them.

Mathematically, this law can be expressed as follows:

$$F = G \frac{m_1 m_2}{r^2} \quad (1.11)$$

where r is the distance between the two particles. Note that when the distance between the particles approaches zero, the gravitational force as expressed by (1.11) would approach infinity. This implies that there must be some small distance at which the gravitation equation breaks down, perhaps at quantum distances. Newton himself did not use the law of gravitation in the form of (1.11); rather he worked with ratios so that the constant G cancels out. Later, the law of gravitation took its modern form. The proportionality constant G is called the *universal gravitational constant*; it appears both in Newton's gravity law and in Einstein's general relativity theory. In celestial mechanics, the effects of the gravitational forces between celestial bodies are studied. These bodies move at relatively large distances from each other, and most of them have an almost spherical shape and a nearly radially-symmetric mass density distribution. As will be shown in Section 1.5, such bodies may be considered as point masses located at the centers of these bodies, as far as the mutual gravitational attraction force is concerned, and (1.11) may be applied. When (1.11) is applied to compute the gravitational force between a celestial body and a spacecraft, the spacecraft certainly may be considered as a point mass, since its dimensions are much smaller than those of (most) celestial bodies. To give an example of the magnitude of the gravitational force and the resulting acceleration of a body, consider the case of a satellite with a mass of 10,000 kg at an altitude of 1000 km above the Earth. Substituting the values of the universal gravitational constant, the mass of the Earth, and the radius of the Earth, which are given in Appendix B, we find from (1.11) for the attraction force: 73.3 kN. For the acceleration of the satellite and of the Earth due to this gravitational force we then find: 7.33 m/s² and 1.23*10⁻²⁰ m/s², respectively. Just as could be expected, the acceleration of the Earth is

extremely small, which is a direct result of its large mass. As a second example, we consider two spacecraft, each with a mass of 10,000 kg and at a distance of 1 km from each other. If we assume that both spacecraft may be considered as point masses as far as their gravitational attraction is concerned, we find for the attraction force on each spacecraft: $6.67 \cdot 10^{-9}$ N, and for the acceleration of each spacecraft: $6.67 \cdot 10^{-13}$ m/s². This very small value shows that the gravitational attraction between spacecraft can be neglected when computing their motion.

The determination of the value of G requires a very delicate experiment measuring the gravitational force between two masses. H. Cavendish (1731-1810) used in 1798 a special torsion balance to determine the mean mass density of the Earth, which was an important scientific problem at the time. He found that the Earth's mean mass density is about $\rho \approx 5.4$ gr/cm³, which is much larger than the density of rocks at the Earth's surface. This observation was one of the first strong indications that density must increase substantially towards the center of the Earth. In 1894, C.V. Boys (1855-1944) published the first calculation of the value of G . Rather than performing a new experiment, he used Cavendish's torsion balance measurements and found $G = 6.74 \cdot 10^{-11}$ m³/kg s². Nowadays, the adopted value is $G = 6.67428 \cdot 10^{-11}$ m³/kg s². It is usual to assume that G is independent of scale and that Newton's inverse-square law of gravitational attraction holds on both laboratory and planetary scales. So, G is considered as a fundamental constant of physics; it is, however, the least well determined fundamental physical constant owing to the intrinsic weakness of the gravitational force. Indeed, the limited accuracy available for G limits the accuracy of the determination of the mass of the Sun and the planets. In Chapter 5 the *gravitational parameter* of a celestial body, $\mu = GM$, where M is the mass of the body, is introduced. That parameter is known with much higher precision than the values of G and M individually. The gravitational parameter of the Earth is $\mu = 398600.4418$ km³/s².

It is interesting to note that the idea of planetary motion about the Sun, where the planets are attracted toward the Sun by a force proportional to the inverse square of the distance between planet and Sun, was already advocated by Hipparchos (~190-120 B.C.). His supposition was based on ideas from various cultures long before him. The inverse square dependence on the distance came from the assumption that the attraction is propagated along rays emanating from the surfaces of the bodies. Brahmagupta (598-668) and Abu Ja'far Muhammad ibn Musa (~803-873) proposed that there is a force of attraction between the Sun and the heavenly bodies. A vague idea of a gravitational force that diminishes with distance was proposed by Johannes Scotus Eriugena (815-877). Alhacen discussed the theory of attraction between masses, and it seems that he was aware of the magnitude of acceleration due to gravity. Ismael Bullialdus (born as Ismael Boulliau, 1605-1694) supported the hypothesis published by J. Kepler (1571-1630) in 1609 that the planets move in elliptical orbits around the Sun (Section 5.4), but argued against Kepler's proposal that the strength of the force exerted on the planets by the Sun would decrease in inverse proportion to their distance from it. He argued in 1640 that if such a force existed it would instead have to follow an inverse-square law. However, Bullialdus did not believe that any such force did exist! R. Hooke (1635-1703) wrote in 1680 that all planets are pulled towards the Sun with a force proportional to their mass and inversely proportional to the square of their distance to the Sun. By that time, this assumption was rather common and had been advanced by a number of scientists for different reasons. In 1687, Newton published his *Principia*, in which he hypothesizes the inverse-square law of universal gravitation between any two bodies.

Many scientists have philosophized about the nature of the gravitational force. R. Descartes (1596-1650) and Huygens used vortices to explain gravitation. Hooke and J. Challis (1803-1882) assumed that every body emits waves which lead to an attraction of other bodies. N. Fatio de Duillier (1664-1753) and G.L. le Sage (1724-1803) proposed a corpuscular model, using some

sort of screening or shadowing mechanism. Later, a similar model was developed by H.A. Lorentz (1853-1928), who used electromagnetic radiation instead of corpuscular radiation. Newton and G.F.B. Riemann (1826-1866) argued that aether streams carry all bodies to each other. Newton and L. Euler (1707-1783) proposed a model, in which the aether loses density near the masses, leading to a net force directing to the bodies. W. Thomson (Lord Kelvin; 1824-1907) proposed that every body pulsates, which then could explain gravitation and electric charges. In 2010, E.P. Verlinde (1962-) argued that gravity is linked to the amount of information associated with matter and its location, measured in terms of entropy. Changes in this entropy when matter is displaced then leads to a reaction force that we know as gravity.

In Einstein's general relativity theory, gravitation is not a force but a phenomenon resulting from the curvature of space-time. This curvature is caused by the presence of matter (objects). Einstein proposed that free-falling objects are moving along locally straight paths in curved space-time (this type of path is called a *geodesic*). The more massive the object is, the greater the curvature it produces and hence the more intense the gravitation. As celestial objects move around in space-time, the curvature changes to reflect the changed locations of those objects. In certain circumstances, the time-varying accelerations of compact massive bodies (e.g. neutron stars, black holes) in binary star systems, or produced by neutron star mergers or black hole formations, may create fluctuations in the curvature of space-time. These fluctuations generate *gravitational waves*, which propagate outwards at the speed of light and transport energy as *gravitational radiation*. When a gravitational wave passes an observer, that observer will find space-time distorted and will measure distances between free objects to increase and decrease rhythmically as the wave passes, at a frequency corresponding to that of the wave. The amount of gravitational radiation emitted by the solar system is far too small to measure. In theory, the loss of energy through gravitational radiation makes the Earth orbit to slowly spiral in at a rate of about 10^{-15} m/day. At this rate, it would take the Earth approximately 10^{13} times more than the current age of the universe to spiral onto the Sun, while the Earth is predicted to be swallowed by the Sun in the red giant stage of its life in a few billion years time. Gravitational radiation has been indirectly observed as an energy loss over time in binary pulsar systems. In the past decades, gravitational radiation observatories have been built to measure this type of radiation, but no confirmed detections have yet been made. Space-based interferometers aiming at measuring gravitational waves, such as the NGO system that is proposed by ESA for launch after 2022, are being studied. The NGO mission aims at measuring gravitational waves over a broad band at low frequencies, from about 100 μ Hz to 1 Hz. The mission will employ three spacecraft forming a rotating nearly equilateral triangle with an arm length of 10^6 km, positioned in heliocentric, Earth-trailing orbits with a radius of about 1 AU and with the plane of the constellation inclined at 60° to the ecliptic. Lasers in each of the spacecraft will measure changes in path length between free falling test masses housed in the three spacecraft to picometer accuracy.

In this book, we will apply Newton's law of gravitation, which assumes that G is a constant and that the gravitational force acts instantaneously; i.e. irrespective of the position or velocity of two bodies, the force will always act along the instantaneous straight line connecting the two bodies. So, we may say that this law assumes that the speed of gravity to be infinite. Consequently, orbit computations must use true, instantaneous positions of all celestial bodies when computing the gravitational attraction by the bodies. For example, even though we know that the Earth is at a distance of about 500 light-seconds from the Sun, Newtonian gravitation theory describes the force on Earth directed towards the Sun's position 'now', not its position 500 s ago. This aspect has already worried P.S. Laplace (1749-1827) and he made in 1805 an attempt to combine a finite gravitational speed with Newton's law of gravitation. He found that, for a stable solar system, the speed of gravitational interactions should be at least 7×10^6 times the speed of

light. However, his analysis was fundamentally incorrect.

The infinite speed of gravity in Newtonian theory seems to contradict Einstein's relativity theory, which forbids any effect to propagate faster than the speed of light. However, we should realize that general relativity is conceptually very different from Newtonian gravitation theory. Loosely stated, it tells us that for any mass that moves uniformly relative to an inertial frame its gravity field appears static relative to the mass itself—i.e., it moves as if attached to the mass. For weak fields, which occur in our solar system and in ‘normal’ stellar systems, one finds that the ‘force’ in general relativity is not quite central—it does not point directly towards the source of the gravity field—and that it depends on velocity as well as position. The net result is that the effect of propagation delay is almost exactly cancelled, and general relativity theory very nearly reproduces the Newtonian gravity result. It is noted that when the orbit of a celestial body has been computed, the position where we ‘see’ that body can be computed by allowing for the delay of light traveling from that body to Earth. This *aberration* effect is discussed in Section 5.11.

It is remarkable that the parameter ‘mass’ is present both in the equation of motion (1.7) and in the equation for the attracting force (1.11). Intuitively, we always assume that the meaning of ‘mass’ is identical in both equations. However, this is not trivial! According to (1.7), we have to exert a force to change an object’s velocity. The necessary force is proportional to the *inertial mass* of the object; the more massive the object, the larger the necessary force. The gravitational force exerted by two objects on each other is proportional to their *gravitational masses*. Newton assumed that the gravitational mass of an object is identical to its inertial mass. As a result, the acceleration of a body acted upon by gravitational forces is independent of the mass of that body, as we will see later on. However, Newton realized that the assumption that inertial mass and gravitational mass are identical is not self-evident. The effort it takes us to move an object, does not necessarily have to be dependent on the gravitational mass that determines the force that the object exerts itself. Consider, for example, an electrical force; this force is proportional to the electric charge and not to its inertial mass. L. von Eötvös (1848-1919) and R.H. Dicke (1916-1997) have verified that materials of different composition and mass experience exactly the same acceleration in a gravity field, which indicates that inertial mass and gravitational mass are very much equal; their results had an accuracy of 10^{-8} and 10^{-11} , respectively. Presently, experiments even achieve an accuracy of 10^{-12} . Einstein has proved that in the supposition of a constant speed of light for all observers (which is the substance of the *special theory of relativity*), gravitational mass and inertial mass are identical indeed.

If we now consider the force acting on particle m_2 due to the mutual gravitational attraction between m_1 and m_2 , we may write (1.11) as

$$\bar{F}_2 = -G \frac{m_1 m_2}{r_2^3} \bar{r}_2 \quad (1.12)$$

where \bar{r}_2 is the position vector from m_1 to m_2 . We can imagine the force acting on m_2 to be caused by a *gravity field* generated by m_1 . The force per unit of mass of m_2 at the location of m_2 is called the *field strength*, \bar{g}_2 , of the gravity field generated by m_1 :

$$\bar{g}_2 = -G \frac{m_1}{r_2^3} \bar{r}_2 \quad (1.13)$$

We now introduce a scalar quantity

$$U_2 = -G \frac{m_1}{r_2} + U_{2,0} \quad (1.14)$$

where $U_{2,0}$ is an arbitrary constant. Note that U_2 is a function of the position of body m_2 relative to body m_1 only. From (1.13) and (1.14) follows:

$$\bar{g}_2 = -\bar{\nabla}_2 U_2 \quad (1.15)$$

where $\bar{\nabla}_2$ is the nabla operator (del operator, gradient), i.e. the derivative of U_2 in three-dimensional space with respect to the coordinates of body m_2 . From theoretical mechanics we know that if the local field strength can be found by partial differentiation of a scalar function of position coordinates to these position coordinates, then this function is called a *potential*. Therefore, U_2 is the potential of the force field generated by body m_1 at the location of body m_2 . The potential energy of body m_2 is $m_2 U_2$. In celestial mechanics, it is customary to choose the potential at infinity equal to zero, which means that $U_{2,0}=0$. Thus, at any other distance the gravitational potential is negative, and the gravitational potential of a particle m_1 at an arbitrary distance, r , can be expressed as

$$U = -G \frac{m_1}{r} \quad (1.16)$$

From theoretical mechanics we know that if a potential is not *explicitly* depending on time, then the force field is *conservative* and the sum of potential and kinetic energy of a body moving in that force field is constant. So, the Newtonian gravity field described by (1.16) is conservative. Notice that when computing the Earth's gravitational force acting on an object on the surface of the Earth, of course, the centrifugal force due to the rotation of the Earth should be added to the gravitational force. At the equator, this centrifugal force leads to an outward acceleration of 3.39 cm/s². In that case, the centrifugal potential $-\frac{1}{2}\dot{\theta}^2 d^2$, where $\dot{\theta}$ is the rotational velocity of the Earth and d is the distance of the object from the rotation axis of the Earth, should be added to the gravitational potential in (1.15).

An arbitrary body L with finite dimensions can be viewed as a collection of particles. Since potential functions may be added, the gravity field of such a body may be written as

$$U = \sum_L dU \quad (1.17-1)$$

where

$$dU = -G \frac{dm}{r} \quad (1.17-2)$$

and dm is a particle from the collection constituting L . In reality, one will use an integration over the entire body instead of the summation as given in (1.17-1).

For a body of arbitrary shape and mass distribution it is not possible to find a closed-form analytical solution for the gravitational potential of that body and one is forced to use series expansions. Exceptions are spherical shells of constant mass density and spheres with a radially symmetric mass density distribution. We know that the shape of the stars, the Sun, the Moon and the planets closely resembles that of a sphere, and that, to first approximation, the mass density distribution of these celestial bodies can be assumed to be radially symmetric. Therefore, we will derive in the next Section expressions for the gravity field of a thin spherical shell and of a

sphere.

1.5. Gravity field of a thin spherical shell and a sphere

Let us first consider the gravitational potential of a homogeneous thin spherical shell in a point P within that shell. The radius of the shell is R and its thickness is t ; for the analysis given below we assume $t \ll R$. In Figure 1.3 (left) a thin ring perpendicular to the line CP is shown, where C is the center of the spherical shell and l is the distance between C and P . All points on the ring are at a distance r from P . The circumference of the ring is $2\pi R \sin\theta$, and the mass of the ring is given by

$$dm = (2\pi R \sin\theta)(R d\theta) t \rho$$

where ρ is the mass density of the shell. The gravitational potential of the ring at P is

$$dU_P = -\frac{G 2\pi R^2 t \rho \sin\theta d\theta}{r} \quad (1.18)$$

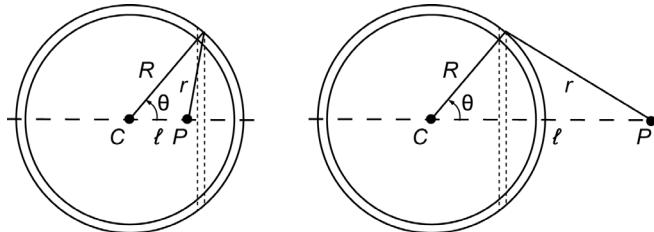


Figure 1.3: Geometry for the computation of the internal (left) and external (right) gravity field of a thin spherical shell.

For r we write

$$r^2 = R^2 + l^2 - 2Rl \cos\theta \quad (1.19)$$

and for the mass of the thin spherical shell:

$$M = 4\pi R^2 t \rho$$

Substitution of these relations into (1.18) yields for the gravitational potential of the shell at P :

$$U_P = -\frac{1}{2} G M \int_{\theta=0}^{\pi} \frac{\sin\theta d\theta}{\sqrt{R^2 + l^2 - 2Rl \cos\theta}} \quad (1.20)$$

Differentiation of (1.19) yields for $R > l$:

$$dr = \frac{R l \sin\theta d\theta}{\sqrt{R^2 + l^2 - 2Rl \cos\theta}} \quad (1.21)$$

Combination of (1.20) and (1.21) gives

$$U_P = -\frac{1}{2} \frac{GM}{Rl} \int_{r=R-l}^{R+l} dr$$

or, after integration,

$$U_P = -\frac{GM}{R} \quad (1.22)$$

This equation shows that within a spherical shell the gravitational potential is constant, i.e. independent of the position of P . Since the total force on particle P with mass m_p in an arbitrary direction x can be expressed as

$$F_{P_x} = -m_p \frac{\partial U_p}{\partial x} \quad (1.23)$$

it can be concluded that the resulting attracting force on P is equal to zero.

Next, the case of P outside the spherical shell is considered (Figure 1.3 (right)). In a similar way as for the case of P within the shell, the following expression can be derived:

$$U_P = -\frac{1}{2} \frac{GM}{Rl} \int_{r=l-R}^{l+R} dr \quad (1.24)$$

or

$$U_P = -\frac{GM}{l} \quad (1.24)$$

Note that in this case the gravitational potential is dependent on the position of P . The resulting attracting force on P is directed along l and is according to (1.23) and (1.24) equal to

$$F_P = -\frac{GMm_p}{l^2} \quad (1.25)$$

So, a spherical shell attracts a mass m_p outside the shell as if the mass of the shell were concentrated in the center of the spherical shell.

A spherical body with a radially symmetric mass density distribution can be considered as a series of thin spherical shells with the same center and each with its own constant mass density. Using (1.24), the external gravity field of such a body can be written as

$$U_P = -\frac{G}{l} \sum_i M_i = -\frac{GM_T}{l} \quad (1.26)$$

where M_i is the mass of a spherical shell and M_T is the total mass of the sphere. According to (1.23), the total attracting force on P can be expressed as

$$F_P = -\frac{GM_T m_p}{l^2} \quad (1.27)$$

where the force is again directed along l . Note that the gravitational potential and the gravitational force are independent of the mass density distribution within the sphere, as long as this distribution is radially symmetric. Also note that the sphere attracts point mass m_p as if the entire mass of the sphere were concentrated at the center of the sphere. This is a very important result, because it demonstrates that, to first-order approximation, we may consider celestial bodies as point masses as far as their gravitational attraction is concerned.

The slight deviations in the internal mass distribution of the Sun and the planets will hardly

have any effect on the motion of the planets around the Sun, because of the very large distances between them. Totally different is the case of the motion of a satellite about the Earth. Many satellites orbit the Earth at distances from the Earth's surface that are small compared to the radius of the Earth. Therefore, the slight deviations from a radially-symmetric mass distribution of the Earth will have a clearly observable effect on the satellite orbit. A second difference between satellite orbits and planetary orbits is the fact that satellites often move so close to the Earth's surface that their orbits are significantly influenced by atmospheric forces. It are these two effects, i.e. the non-spherical mass distribution of the Earth and the occurrence of atmospheric forces, that make precise orbit computations of satellites about the Earth more difficult than precise orbit computations of planets about the Sun. Fortunately, both effects can be considered as perturbations of the orbit. For satellite orbit computations, we may, to first-order approximation, consider the Earth as purely spherical and its mass density distribution as radially symmetric, and we may neglect the atmospheric forces. The perturbations of satellite orbits and the computation of perturbed satellite orbits will be dealt with in Chapters 20 to 23.

1.6. External gravity field of a body with arbitrary mass distribution

As mentioned before, it is not possible to derive a closed-form analytical solution for the external gravitational potential of a body with an arbitrary shape and mass distribution. Therefore, the gravitational potential is usually expressed through series expansion. To gain some insight in the character of these series expansions, we will discuss a few special cases.

Consider body L that has an arbitrary shape and internal mass density distribution. Connected to this body is a reference frame XYZ , of which the origin, O , coincides with the center of mass of L (Figure 1.4). An element Q of L has a mass dm and coordinates x, y, z . At some distance ℓ from O there is a particle P , which has a mass m_1 and coordinates x_1, y_1, z_1 . It is assumed that ℓ is larger than the largest dimension of L ; i.e. P is positioned outside a sphere around O that fully contains body L .

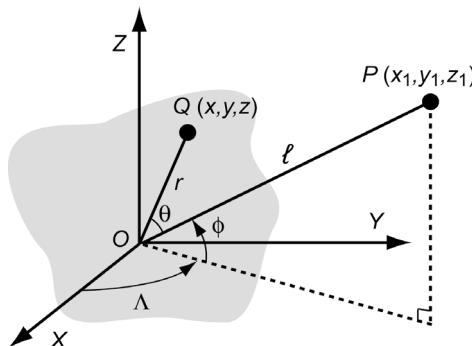


Figure 1.4: Arbitrary body L and an external point P . The origin of reference frame XYZ coincides with the center of mass of L .

According to Newton's law of gravitation, the force acting on P due to the mutual attraction between P and Q can be expressed as

$$\bar{F} = G \frac{m_1 dm}{(PQ)^3} \bar{PQ}$$

where G is the universal gravitational constant and \bar{PQ} is a vector along the line connecting P

and Q , which is directed from P to Q and has a magnitude equal to the distance between P and Q . Therefore, the gravitational potential of body L at the position of P is

$$U = -G \int \frac{dm}{PQ} \quad (1.28)$$

where the integral has to be taken over the entire body L . For the distance PQ and the distances from P and Q to the origin O , the following expressions hold:

$$(PQ)^2 = (x_1 - x)^2 + (y_1 - y)^2 + (z_1 - z)^2$$

$$(OP)^2 = l^2 = x_1^2 + y_1^2 + z_1^2 \quad ; \quad (OQ)^2 = r^2 = x^2 + y^2 + z^2$$

Combination of these expressions gives

$$(PQ)^2 = l^2 \left[1 - 2 \left(\frac{x_1 x + y_1 y + z_1 z}{l r} \right) \frac{r}{l} + \left(\frac{r}{l} \right)^2 \right] \quad (1.29)$$

When the notations

$$\alpha = \frac{r}{l} \quad ; \quad q = \cos \theta = \frac{\bar{l} \cdot \bar{r}}{l r} = \frac{x_1 x + y_1 y + z_1 z}{l r} \quad (1.30)$$

are introduced, where θ is the angle POQ , (1.29) can be rewritten as

$$(PQ)^2 = l^2 (1 - 2q\alpha + \alpha^2)$$

Substitution of this relation into (1.28) yields

$$U = -G \int \frac{dm}{l(1 - 2q\alpha + \alpha^2)^{1/2}} \quad (1.31)$$

Since every point Q must satisfy $\alpha < 1$ and since also: $q \leq 1$, it is possible to use the following series expansion for (1.31):

$$U = -\frac{G}{l} \left(\int P_0(q) dm + \int \alpha P_1(q) dm + \int \alpha^2 P_2(q) dm + \dots \right) \quad (1.32)$$

where

$$P_0(q) = 1 \quad ; \quad P_1(q) = q \quad ; \quad P_2(q) = \frac{1}{2}(3q^2 - 1) \quad ; \quad \text{etc.} \quad (1.33)$$

The structure of these expressions reveals that P_i are *Legendre polynomials* in q . Evaluating the first integral in (1.32) yields

$$U_0 = -\frac{G}{l} \int dm = -\frac{GM}{l} \quad (1.34-1)$$

where M is the total mass of L . With (1.33) the second integral in (1.32) yields

$$U_1 = -\frac{G}{l} \int q \alpha dm = -\frac{G}{l^3} \left(x_1 \int x dm + y_1 \int y dm + z_1 \int z dm \right)$$

Since O is the center of mass of L , the result is

$$U_1 = 0 \quad (1.34-2)$$

Evaluation of the third integral in (1.32) gives

$$U_2 = -\frac{1}{2} \frac{G}{l} \int (3q^2 - 1) \alpha^2 dm = -\frac{1}{2} \frac{G}{l^3} \left(2 \int r^2 dm - 3 \int r^2 \sin^2 \theta dm \right) \quad (1.35)$$

The moments of inertia A , B , C and D of body L about the X -, Y - and Z -axis and the line OP , respectively, are defined as

$$\begin{aligned} A &= \int (y^2 + z^2) dm ; \quad B = \int (x^2 + z^2) dm \\ C &= \int (x^2 + y^2) dm ; \quad D = \int r^2 \sin^2 \theta dm \end{aligned} \quad (1.36)$$

Using these expressions, (1.35) may be rewritten as

$$U_2 = -\frac{1}{2} \frac{G}{l^3} (A + B + C - 3D) \quad (1.34-3)$$

When the series expansion in (1.32) is truncated after the third term, substitution of relations (1.34) into (1.32) results in:

$$U = -\frac{GM}{l} - \frac{1}{2} \frac{G}{l^3} (A + B + C - 3D) \quad (1.37)$$

which leads to the following expression for the (central) attracting force on P per unit of mass:

$$g = \frac{F}{m_1} = -\frac{GM}{l^2} - \frac{3}{2} \frac{G}{l^4} (A + B + C - 3D) \quad (1.38)$$

where g is the acceleration due to gravity. Equation (1.37) has first been published by J. MacCullagh (1809–1847) in 1849 and is generally known as *MacCullagh's formula*; it gives a first-order approximation of the external gravity field of a body with arbitrary shape and mass density distribution. Equation (1.38) shows that the gravitational attraction of an irregular body has two contributions; the first is the attraction of a point mass with mass M located at the center of mass of the body, the second term depends on the moments of inertia around the principal axes, which in turn depend completely on the mass distribution of the body. The first term decays as $1/l^2$ with increasing distance, l , while the second term decays as $1/l^4$. So, at large distances the gravity field approaches that of a point mass and becomes less and less sensitive to aspherical variations in the mass distribution of the body. This is the reason why the CHAMP, GRACE and GOCE (Section 19.1) satellites, which were launched in 2000, 2002, 2009, respectively, and were dedicated to measuring the gravity field of the Earth accurately, have flown at altitudes as low as 300 km, 450 km and 250 km, respectively.

For a sphere or a spherical shell

$$A = B = C = D$$

which leads to

$$U = -\frac{GM}{l} ; \quad F = -\frac{GMm_1}{l^2} \quad (1.39)$$

These results are, of course, identical to the ones found in Section 1.5. For the general case, we

can, using (1.30) and (1.36), derive the following expression for D :

$$D = \int \left(r^2 - \frac{1}{l^2} (x x_1 + y y_1 + z z_1)^2 \right) dm$$

According to Figure 1.4 :

$$x_1 = l \cos \phi \cos \Lambda ; \quad y_1 = l \cos \phi \sin \Lambda ; \quad z_1 = l \sin \phi$$

Then, D can be written as

$$\begin{aligned} D = & \int \{ x^2 + y^2 + z^2 - (x^2 \cos^2 \phi \cos^2 \Lambda + y^2 \cos^2 \phi \sin^2 \Lambda + z^2 \sin^2 \phi \\ & + 2xy \cos^2 \phi \sin \Lambda \cos \Lambda + 2xz \cos \phi \sin \phi \cos \Lambda + 2yz \cos \phi \sin \phi \sin \Lambda) \} dm \end{aligned} \quad (1.40)$$

We now assume that the reference frame is oriented such that XYZ are the *principal axes of inertia* of body L. In that case, the products of inertia are zero and all terms containing xy , xz and yz can be set equal to zero. This yields

$$D = \int (x^2 + y^2 + z^2 - x^2 \cos^2 \phi \cos^2 \Lambda - y^2 \cos^2 \phi \sin^2 \Lambda - z^2 \sin^2 \phi) dm \quad (1.41)$$

After some trigonometric manipulation this relation can be rewritten as

$$D = \int \{ (y^2 + z^2) \cos^2 \phi \cos^2 \Lambda + (x^2 + z^2) \cos^2 \phi \sin^2 \Lambda + (x^2 + y^2) \sin^2 \phi \} dm$$

or, with (1.36),

$$D = A \cos^2 \phi \cos^2 \Lambda + B \cos^2 \phi \sin^2 \Lambda + C \sin^2 \phi \quad (1.42)$$

With this relation it is possible to calculate D of an arbitrary body of which the principal moments of inertia A , B and C are known, and consequently determine an approximation of the external gravity field of L at point P .

As a simplification, we now assume that the principle moments of inertia A and B are equal. This is the case for a body of which both the shape and the internal mass density distribution is rotationally symmetric about the Z-axis. Using relation (1.42) we then find

$$D = A \cos^2 \phi + C \sin^2 \phi \quad (1.43)$$

Substitution of $A = B$ and (1.43) into (1.37) and (1.38) yields

$$\begin{aligned} U &= -\frac{GM}{l} \left[1 - \frac{1}{2} \frac{C-A}{Ml^2} (3 \sin^2 \phi - 1) \right] \\ g &= -\frac{GM}{l^2} \left[1 - \frac{3}{2} \frac{C-A}{Ml^2} (3 \sin^2 \phi - 1) \right] \end{aligned} \quad (1.44)$$

The term $\frac{1}{2}(3\sin^2 \phi - 1)$ is the second-degree *Legendre polynomial* in $\sin \phi$. If the series expansion in (1.32) would have been continued after the third term, we would have found a gravitational potential consisting of a sum of Legendre polynomials (Section 20.1).

The expressions (1.44) are often used as a first-order approximation of the Earth's external gravity field. This is allowed, because for the Earth the higher-order terms are indeed small of the second order (Section 20.1). Equations (1.44) can also be written as

$$\begin{aligned} U &= -\frac{GM}{l} \left[1 + \frac{1}{2} \frac{C-A}{MR^2} \left(\frac{R}{l} \right)^2 (1 - 3 \sin^2 \phi) \right] \\ g &= -\frac{GM}{l^2} \left[1 + \frac{3}{2} \frac{C-A}{MR^2} \left(\frac{R}{l} \right)^2 (1 - 3 \sin^2 \phi) \right] \end{aligned} \quad (1.45)$$

where R is the radius of the body in the XY -plane. We may write

$$\frac{C-A}{MR^2} = \frac{C-A}{C} \frac{C}{MR^2} \quad (1.46)$$

The ratio $(C - A)/C$ may be found from observations on the luni-solar precession of the Earth (Section 11.2), while the ratio C/MR^2 may be estimated from the theory of hydrostatic equilibrium of the Earth. This method was used by G.H. Darwin (1845-1912) and further developed by W. de Sitter (1872-1934); its limitation depends on the degree of applicability of the theory of hydrostatic equilibrium to the Earth. When we substitute in (1.46) modern values of the relevant parameters listed in Appendix B, we find $(C - A)/(MR^2) \approx 1.082 \times 10^{-3}$, which demonstrates that the effect of the non-radially-symmetric mass distribution is very small indeed.

Equation (1.45-2) shows that the acceleration of a body due to gravity is independent of the mass of that body. For small altitude variations at the surface of the Earth ($l \approx R$) we find from (1.45-2) that, at a certain location on Earth, the acceleration of the body is constant. Ancient scholars were not aware of this fact. As an example, Aristotle (~384-322 B.C.), who probably was the first writer who gave a quantitative description of falling motion, wrote that an object falls at a constant speed, attained shortly after being released, and that this speed is proportional to the weight of the body. After him, many scientists discussed the motion of free falling bodies. Galileo Galilei (1564-1642) was the first who realized in 1604 that, if air resistance and buoyancy can be neglected, all falling bodies experience a constant acceleration, irrespective of their mass. He also tried to measure the acceleration of free falling bodies, and of objects rolling down an inclined plane. However, he never got accurate results due to the lack of an accurate clock. The invention by Huygens of an accurate pendulum clock in 1656 afforded the first practical means of measuring the acceleration due to gravity. To honor Galilei for his studies on the acceleration due to gravity, the unit often used in gravimetry is the gal (symbol Gal): $1 \text{ Gal} = 1 \text{ cm/s}^2 \approx 10^{-3} \text{ g}$. Until the second half of the nineteenth century it was virtually impossible to measure gravity with an accuracy of better than 1 Gal. Around 1900 a measurement precision of 1 mGal was achieved; nowadays, routine relative gravity field measurements (difference in gravity between two locations) have an accuracy of about 5 μGal ; in a laboratory an accuracy of down to about 0.1 μGal is achieved.

When we substitute in (1.45-2) modern values of the relevant parameters listed in Appendix B, we find $g = 9.814 \text{ m/s}^2$ at the equator and $g = 9.832 \text{ m/s}^2$ at the poles. The first observations that gravity varies at different points on Earth was made in 1672 by J. Richer (1630-1696), who took a pendulum clock to Cayenne, French Guiana, and found that it lost 2.5 min per day. In 1687, Newton showed in his *Principia* that this was because the Earth has a slightly oblate shape (flattened at the poles; Section 11.1). In 1737, P. Bouguer (1698-1758) made a series of pendulum observations in the Andes mountains, Peru, at three different altitudes, from sea level to the top of the high Peruvian plateau. His measurements showed that gravity fell off slower than with the inverse square of the distance from the center of the Earth. He correctly attributed the ‘extra’ gravity to the gravitational field of the huge Peruvian plateau. Nowadays, we know that the shape of the Earth can be approximated by an ellipsoid of revolution (Section 11.1), with its

minor axis, having a length of 6356.752 km, oriented along the Earth's rotation axis and its major axis, having a length of 6378.137 km, oriented in the equatorial plane. Due to the oblateness of the Earth and its rotational velocity, the total acceleration experienced by a body at this reference surface varies from 9.780 m/s² at the equator to 9.832 m/s² at the poles.

1.7. Maneuvers with rocket thrust

On several places in this book the application of rocket engines for launch vehicles and spacecraft maneuvers will be discussed. To prevent that elements of this subject have to be treated more than once, this Section deals with those elements. The analysis concerns high-thrust propulsion systems, such as chemical and future thermal-nuclear rocket engines. In Chapter 19, an analysis will be presented for low-thrust propulsion systems, such as ion rocket engines. In the discussion, it is assumed that the reader already has some general knowledge of trajectories of rockets and spacecraft.

The thrust of a rocket engine is, according to (1.10), given by $F = \dot{m} V_j$, where \dot{m} is the mass flow rate, i.e. the mass of exhaust gases that flow per second through the nozzle, and V_j is the (effective) exhaust velocity. With $f = F/M$ and $\dot{m} = -dM/dt$, where f is the *thrust acceleration* and M is the (instantaneous) mass of the space vehicle, we find from (1.10):

$$\frac{dM}{M} = -\frac{f}{V_j} dt$$

Integration from $t = 0$, i.e. the time that the engine starts thrusting, to the time that the engine stops thrusting, t_e , gives

$$\ln \frac{M_0}{M_e} = \int_0^{t_e} \frac{f}{V_j} dt \quad (1.47)$$

The initial mass of the spacecraft, M_0 , may be assumed to consist of three components: the mass of the construction (including the rocket engines and propellant tanks), M_c , the mass of the payload, M_l , and the mass of the propellants, M_p . So, $M_0 = M_c + M_l + M_p$. For the mass at t_e , when all propellants have been consumed, we may write $M_e = M_c + M_l$. From these relations we find

$$\frac{M_0}{M_e} = \left(\frac{M_e}{M_0} \right)^{-1} = \left(\frac{M_l}{M_0} + \frac{M_c}{M_0} \right)^{-1} > 1$$

This expression shows that, for a given value of M_c/M_0 , a maximum payload ratio, M_l/M_0 , corresponds to a minimum value of M_0/M_e . This means a minimum value of the integral on the right-hand side of (1.47). Note that a higher exhaust velocity will always lead to a larger payload ratio. For these thrusting systems the exhaust velocity may, generally, be considered constant during the thrusting phase, which means that, for a given spacecraft with specified initial mass and engine exhaust velocity, the payload ratio is maximum if

$$\int_0^{t_e} f dt \text{ is minimum} \quad (1.48)$$

It is emphasized that the integral given above, generally, is not equal to the velocity increment of the vehicle at the end of the thrusting period. The reason is that other forces are acting on the

spacecraft.

To find the optimum thrust program (magnitude and direction of the thrust) and the optimum trajectory of the spacecraft, we have to integrate the equations of motion for specified initial and final conditions (including flight time), and using (1.48) as a minimization criterion. In this book, we will not address the computation of optimum trajectories. When, as a zeroth-order approximation, we assume that no other forces are acting on the spacecraft (this condition is generally referred to as *gravity-free space*) and that the (effective) exhaust velocity is constant, then integration of (1.47) leads to

$$\Delta V = V_j \ln \frac{M_0}{M_e} \quad (1.49)$$

where ΔV is the *ideal velocity increment* produced by the propulsion system during the period t_0 to t_e . This equation is known as *Tsiolkovski's law*. It was published by K.E. Tsiolkovski (1857-1935) in 1903, and represents for gravity-free space the relation between the amount of propellants used, the (effective) exhaust velocity and the velocity increment achieved. The ideal velocity change is independent of the thrust program, as long as the exhaust velocity is constant.

In reality, the gravitational attraction by the Earth is acting on the spacecraft. In Figure 1.5 a section of a trajectory is sketched, where it is assumed that during the time interval t_0 to t_e a thrust \bar{F} acts upon the spacecraft continuously. The gravitational attraction force is indicated by G . The angle between the thrust vector and the normal to the position vector (*thrust angle*), and the angle between the velocity vector and the normal to the position vector (*flight path angle*) are indicated by δ and γ , respectively.

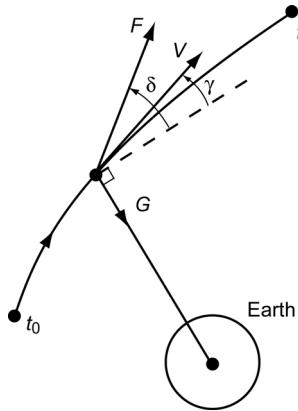


Figure 1.5: Geometry of powered flight and a definition of the flight path angle, γ , and the thrust angle, δ .

Starting from (1.9), we can write for the motion along the trajectory

$$\frac{dV}{dt} = \frac{\dot{m} V_j}{M} \cos(\delta - \gamma) - g \sin \gamma \quad (1.50)$$

where $g = G/M$. For satellite launchers and high-thrust spacecraft orbit change maneuvers, the optimum flight profile usually requires that the thrust is directed (approximately) tangentially to the trajectory ($\delta \approx \gamma$). If the exhaust velocity is assumed to be constant, integration of (1.50) then gives

$$\Delta V = V_j \ln \frac{M_0}{M_e} - \int_{t_0}^{t_e} g \sin \gamma \, dt \quad (1.51)$$

This equation shows that the true velocity increment is smaller than the ideal velocity increment. The second term on the right-hand side of (1.51) represents the *gravity loss*:

$$\Delta V_G = \int_{t_0}^{t_e} g \sin \gamma \, dt \quad (1.52)$$

This gravity loss is determined by the thrust program, which determines g and γ as a function of time, and by the total thrusting time $t_e - t_0$. The gravity loss is equal to zero when the thrust vector is continuously directed perpendicularly to the position vector ($\gamma = 0^\circ$). In reality, trajectory optimization and mission requirements will not permit the thrust vector to be directed perpendicularly to the position vector during the entire propelled flight, and thus gravity losses will occur. Quantitative information about the magnitude of the gravity loss occurring during orbit maneuvers will be provided in Section 12.7 and Section 18.9.

Satellite launch vehicles will also experience a *drag loss*, ΔV_D , due to their motion through the atmosphere, and for these vehicles we may write

$$\Delta V = \Delta V_{id} - \Delta V_G - \Delta V_D$$

Both the drag and the gravity loss are sensitive to the initial thrust-to-mass ratio, $(F/M)_0$. Low thrust-to-mass ratios cause the gravity loss to be high because the vehicle spends more time in ascent, while high thrust-to-mass ratios cause the drag loss to be high because of the higher velocities achieved in the lower atmosphere. For medium-to-large launch vehicles flying optimum trajectories the gravity loss amounts to 0.7 - 1.5 km/s and the drag loss to 20 - 50 m/s.

In analyses of spacecraft maneuvers, the concept of an *impulsive shot* is often used. This concept is based on the fact that for these maneuvers a rocket engine usually operates only for a (very) short time interval and that we therefore may assume that the maneuver generates a discontinuous change in velocity, while the position of the spacecraft remains unaltered. In other words, by using the impulsive shot approximation, we assume that the kinetic energy per unit of mass of the spacecraft is changed, while the potential energy per unit of mass remains the same. It is clear that for an impulsive shot both the gravity loss and the drag loss are zero. In that case, (1.49) shows that, for a given initial mass and construction mass, a maximum payload mass is realized if the ideal velocity increment is minimum. So, the optimization criterium becomes a minimization of ΔV ; this criterion will be used many times in this book. The concept of an impulsive shot can, of course, only be applied when using chemical and thermal-nuclear propulsion, of which the thrusting time is short in comparison to the travel time for a certain mission. Using these types of propulsion systems, the larger part of the flight is performed *gravitating* without thrust. If electric propulsion is used, the propulsion system operates during a large part of the flight (Chapter 19) and the concept of an impulsive shot loses its meaning.

Figure 1.6 refers to a satellite that orbits the Earth. At time t_0 , the position vector and the velocity vector are indicated by \bar{r}_0 and \bar{V}_0 , respectively. We assume that at that time an impulsive shot $\Delta \bar{V}$ is applied. The position vector and velocity vector just after the impulsive shot are \bar{r}_1 and \bar{V}_1 , respectively, with

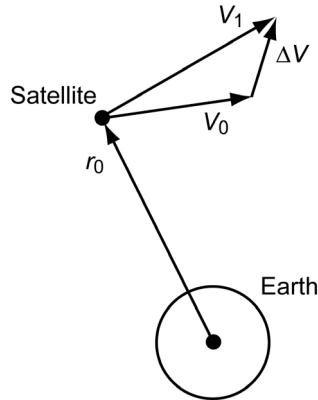


Figure 1.6: Geometry of an impulsive shot.

$$\bar{V}_1 = \bar{V}_0 + \Delta\bar{V} \quad (1.53)$$

As stated above, an optimum maneuver requires, for a given initial mass and a specified orbital change, that ΔV is minimum. Here, an elementary analysis will be presented for two basic maneuvers: a maneuver to change the spacecraft's orbital angular momentum and a maneuver to change the spacecraft's orbital energy. We assume that the initial mass of the spacecraft and its initial orbit about the Earth are known. We want to determine both the position where the maneuver has to be executed and the direction of the thrust vector, such that the change of the angular momentum or of the orbital energy, both per unit of mass, is maximum for a prescribed value of ΔV . The change in orbital angular momentum per unit of mass is given by

$$\Delta\bar{H} = \bar{r}_0 \times \bar{V}_1 - \bar{r}_0 \times \bar{V}_0 = \bar{r}_0 \times \Delta\bar{V} \quad (1.54)$$

The change in the orbital energy per unit of mass is equal to the change in kinetic energy per unit of mass:

$$\Delta\mathcal{E} = \frac{1}{2}(V_1^2 - V_0^2) = \frac{1}{2}(\bar{V}_0 + \Delta\bar{V}) \cdot (\bar{V}_0 + \Delta\bar{V}) - V_0^2 = \frac{1}{2}(\Delta\bar{V})^2 + \bar{V}_0 \cdot \Delta\bar{V} \quad (1.55)$$

From expressions (1.54) and (1.55) some interesting conclusions can be drawn:

- For a given magnitude of $\Delta\bar{V}$, the maximum change in orbital angular momentum is achieved if the impulsive shot is executed when the spacecraft is farthest away from the Earth and if $\Delta\bar{V}$ is perpendicular to \bar{r}_0 .
- If the direction of the orbital angular momentum vector should not be changed, $\Delta\bar{V}$ should be directed in the initial orbital plane. If the direction of the angular momentum vector should be changed, a component of $\Delta\bar{V}$ should be directed perpendicular to the initial orbital plane.
- For a given magnitude of $\Delta\bar{V}$, the maximum change in (total) orbital energy is achieved if the impulsive shot is executed at the point in the orbit where the velocity reaches a maximum value, and if $\Delta\bar{V}$ is directed along the velocity vector \bar{V}_0 , i.e. tangentially to the (initial) orbit. These conclusions strictly only hold for an impulsive shot maneuver, but they are to first-order approximation also valid for most realistic spacecraft maneuvers.

The last conclusion listed above can be applied directly for the analysis of interplanetary trajectories (Chapter 18). For such missions, an important topic is the optimum impulsive shot maneuver to leave a circular parking orbit about the Earth and to enter an interplanetary trajectory, or to enter a parking orbit about another planet upon arrival at that planet. In both cases, the goal of the impulsive shot is to increase or decrease the total orbital energy of the

spacecraft, and that maneuver, of course, has to be executed in such a way that it requires a minimum amount of propellants. For the case of leaving a circular parking orbit about the Earth, we know that for each position in the parking orbit the initial velocity vector of the spacecraft, \bar{V}_0 , is directed perpendicular to the local satellite position vector, \bar{r}_0 , and that its magnitude is the same for each position in the parking orbit. Consequently, the impulsive shot can be applied at any point in the parking orbit and, according to (1.55), the impulsive shot should be directed tangentially to the circular parking orbit and in the direction of the initial velocity, \bar{V}_0 . For the case of entering an (elliptical) parking orbit about the target planet, the deceleration impulse should, according to (1.55), be applied at the point in the flyby trajectory where the velocity in that trajectory is a maximum, and in the direction opposite to the velocity vector at that point in the trajectory. We will use these results in Chapter 18.

1.8. Astronomy and the solar system

To understand the problems in astrodynamics, it is necessary to have some knowledge of astronomy and a global overview of the structure and physical characteristics of our solar system. However, these subjects falls outside the scope of this book and the reader is referred to the many introductory books on this subject. Some physical, astronomical and geophysical data that are useful in astrodynamics are summarized in Appendix B.

2. MANY-BODY PROBLEM

Let us consider a system composed of n bodies, which may be considered as point masses (Figure 2.1). Assume that body i with mass m_i has the coordinates x_i, y_i, z_i with respect to an inertial reference frame. For any other body j the corresponding parameters are m_j, x_j, y_j, z_j . The position of body j relative to body i can be expressed as

$$\bar{r}_{ij} = \bar{r}_j - \bar{r}_i \quad (2.1)$$

where the magnitude of vector \bar{r}_{ij} is

$$r_{ij} = [(x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2]^{1/2} \quad (2.2)$$

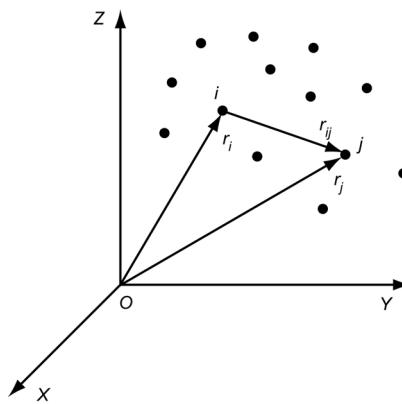


Figure 2.1: Position of n point masses relative to an inertial reference frame XYZ.

When we assume that no external forces act on the system, and that within the system of n bodies only gravitational forces occur, then, applying Newton's second law of motion and Newton's law of gravitation, the motion of body i with respect to the inertial reference frame can be written as

$$m_i \frac{d^2 \bar{r}_i}{dt^2} = \sum_{j \neq i} G \frac{m_i m_j}{r_{ij}^3} \bar{r}_{ij} \quad (2.3)$$

where the summation is taken from $j = 1$ to $j = n$, excluding $j = i$. This equation shows that the motion of body i is governed largely by those bodies j for which the ratio m_j/r_{ij}^2 is large, i.e. bodies that have a large mass and that are close to body i . The equation of motion of body i may be written as three scalar second-order differential equations. Similarly, for the motion of n bodies $3n$ second-order differential equations can be written. The n -body problem can then be formulated as follows: "Consider n point masses in three-dimensional physical space. Suppose that the force of attraction experienced between each pair of point masses is Newtonian. Then, if the initial positions and velocities are specified for every particle at some moment of time, determine the position of each particle at every future (or past) moment of time." For $n = 2$, the problem was solved by J. Bernoulli (1667-1748) in 1710. For $n > 2$, we generally have to rely on numerical integration techniques to determine the motion of the bodies. However, some general characteristics of the many-body problem can be derived. These characteristics are known as the *ten integrals of motion*, which will be derived in the following Section.

2.1. Integrals of motion

The position vector to the center of mass of the n bodies is given by

$$\bar{r}_{cm} = \frac{\sum_i m_i \bar{r}_i}{\sum_i m_i} \quad (2.4)$$

In celestial mechanics the center of mass of a system of bodies is called the *barycenter*. Summation of (2.3) for all i gives

$$\sum_i m_i \frac{d^2 \bar{r}_i}{dt^2} = G \sum_{i,j \neq i} \frac{m_i m_j}{r_{ij}^3} \bar{r}_{ij}$$

It is obvious that the right-hand side of this equation is equal to zero; hence,

$$\sum_i m_i \frac{d^2 \bar{r}_i}{dt^2} = \frac{d^2}{dt^2} \left(\sum_i m_i \bar{r}_i \right) = 0 \quad (2.5)$$

Integration of (2.5) leads to

$$\sum_i m_i \frac{d\bar{r}_i}{dt} = \bar{c}$$

where \bar{c} is an integration constant. This expression shows that the *total linear momentum* of the many-body system is constant in magnitude and direction. In rectangular coordinates this leads to the following three integrals of motion:

$$\sum_i m_i \frac{dx_i}{dt} = c_1 ; \quad \sum_i m_i \frac{dy_i}{dt} = c_2 ; \quad \sum_i m_i \frac{dz_i}{dt} = c_3 \quad (2.6)$$

where c_1, c_2, c_3 are integration constants.

Combination of (2.4) and (2.5) gives

$$\frac{d^2 \bar{r}_{cm}}{dt^2} = 0 \quad (2.7-1)$$

or, after integration,

$$\frac{d\bar{r}_{cm}}{dt} = \bar{a} ; \quad \bar{r}_{cm} = \bar{a}t + \bar{b} \quad (2.7-2)$$

where \bar{a} and \bar{b} are constant vectors. In rectangular coordinates this leads to the following three integrals of motion:

$$x_{cm} = a_1 t + b_1 ; \quad y_{cm} = a_2 t + b_2 ; \quad z_{cm} = a_3 t + b_3 \quad (2.8)$$

where $a_1, a_2, a_3, b_1, b_2, b_3$ are integration constants. These integrals of motion show that the barycenter of the n bodies does not experience an acceleration, but either remains at rest or performs a uniform rectilinear motion relative to the inertial reference frame. Thus, it can be concluded (Section 1.2) that a non-rotating reference frame with its origin at the center of mass of all bodies in the universe is the ‘primary’ inertial reference frame. All reference frames that

perform a uniform rectilinear motion with respect to this primary reference frame are inertial reference frames too.

Three more integrals of motion can be found by taking the vector product of (2.3) and \bar{r}_i and subsequently applying a summation for all i :

$$\sum_i \bar{r}_i \times \left(m_i \frac{d^2 \bar{r}_i}{dt^2} \right) = \sum_i \bar{r}_i \times \left(\sum_{j \neq i} G \frac{m_i m_j}{r_{ij}^3} \bar{r}_{ij} \right)$$

An evaluation of this expression yields, after substitution of (2.1),

$$\sum_i m_i \bar{r}_i \times \frac{d^2 \bar{r}_i}{dt^2} = G \sum_i \sum_{j \neq i} \frac{m_i m_j}{r_{ij}^3} \bar{r}_i \times \bar{r}_j$$

Due to its anti-symmetric properties, the right-hand side of this equation is equal to zero and we may write

$$\frac{d}{dt} \left(\sum_i m_i \bar{r}_i \times \frac{d\bar{r}_i}{dt} \right) = 0 \quad (2.9)$$

or

$$\bar{H} = \sum_i m_i \bar{r}_i \times \frac{d\bar{r}_i}{dt} = \bar{c} \quad (2.10)$$

where \bar{H} denotes the *total angular momentum* of the many-body system, which is constant in magnitude and direction. This constant vector defines a plane that passes through the barycenter of the n bodies and that is perpendicular to the angular momentum vector. This plane is called the *invariable plane of Laplace* and can be used as a reference plane for describing the motion of the n bodies. It was introduced by P.S. Laplace (1749-1827) in 1784. Three integrals of motion can be derived from (2.10), which indicate that the rectangular components of the angular momentum vector are constant:

$$\begin{aligned} H_x &= \sum_i m_i (y_i \frac{dz_i}{dt} - z_i \frac{dy_i}{dt}) = c_1 \\ H_y &= \sum_i m_i (z_i \frac{dx_i}{dt} - x_i \frac{dz_i}{dt}) = c_2 \\ H_z &= \sum_i m_i (x_i \frac{dy_i}{dt} - y_i \frac{dx_i}{dt}) = c_3 \end{aligned} \quad (2.11)$$

It is emphasized that the concept of the invariable plane of Laplace should be used carefully. The angular momentum of the system consists of contributions by the rotational motion of each body about its own axis and contributions by the motion of each body around the origin of the reference frame, is constant. We have assumed that all bodies are point masses (particles). Then, the angular momentum of rotation is zero for all bodies and there exists an invariable plane perpendicular to the angular momentum vector that is linked to the motion of the bodies about the origin of the reference frame (orbital angular momentum). If the bodies are not point masses or infinitely rigid bodies, precession phenomena and friction effects generated by tides will occur,

which will result in an exchange between the angular momentum of motion and the angular momentum of rotation of the bodies. In that case, the invariable plane as defined by (2.10) will not remain constant. However, in the solar system these effects are minute and the angle between the invariable plane and the plane in which the Earth moves about the Sun: the *ecliptic*¹, has an almost constant value of $1^{\circ}35'$.

From (2.3) it follows that the force on body i is given by

$$\bar{F}_i = \sum_{j \neq i} G \frac{\mathbf{m}_i \mathbf{m}_j}{\mathbf{r}_{ij}^3} \bar{r}_{ij}$$

The force per unit of mass of i is called the *field strength*, $\bar{\mathbf{g}}_i$, at the position of body i (Section 1.4). So,

$$\bar{\mathbf{g}}_i = \sum_{j \neq i} G \frac{\mathbf{m}_j}{\mathbf{r}_{ij}^3} \bar{r}_{ij} \quad (2.12)$$

Now, a scalar quantity U_i , which is a function of the spatial coordinates of all bodies, is introduced:

$$U_i = - \sum_{j \neq i} G \frac{\mathbf{m}_j}{\mathbf{r}_{ij}} + U_{i,0} \quad (2.13)$$

where $U_{i,0}$ is an arbitrary constant. Combination of (2.12) and (2.13) shows that the field strength can be written as

$$\bar{\mathbf{g}}_i = -\bar{\nabla}_i U_i \quad (2.14)$$

As already stated in Section 1.4, it is known from theoretical mechanics that if the local field strength can be found by partial differentiation of a scalar function of the spatial coordinates to the spatial coordinates, that scalar function is a *potential*. So, U_i is the potential of the force field at the position of body i and the *potential energy* of body i is given by $m_i U_i$. In Section 1.4, it was stated that it is common practice in celestial mechanics to set the potential at an infinitely large distance equal to zero. Then, $U_{i,0}$ in (2.13) is equal to zero and the potential has a negative value at any finite distance. Two remarks can be made about the force field at the position of i . First, the field certainly is not central, because U_i is a function of r_{ij} and not of r_i . Secondly, the value of the potential at a fixed position relative to the inertial reference frame will vary with time, because the bodies j are moving. For such a *time-varying potential* the sum of kinetic and potential energy of body m_i is not constant (Section 1.4). Therefore, we are dealing with a *non-central, non-conservative force field*.

It is possible to find an energy integral for the system of n bodies. Taking the scalar product of $d\bar{\mathbf{r}}_i/dt$ and (2.3), and subsequently applying a summation for all i , gives

$$\sum_i m_i \frac{d\bar{\mathbf{r}}_i}{dt} \cdot \frac{d^2\bar{\mathbf{r}}_i}{dt^2} = G \sum_i \sum_{j \neq i} \frac{\mathbf{m}_i \mathbf{m}_j}{\mathbf{r}_{ij}^3} \frac{d\bar{\mathbf{r}}_i}{dt} \cdot \bar{r}_{ij}$$

¹ The ‘ecliptic’ is described in more detail in Section 11.2.

or, with (2.1),

$$\frac{d}{dt} \left(\frac{1}{2} \sum_i m_i \frac{d\bar{r}_i}{dt} \cdot \frac{d\bar{r}_i}{dt} \right) = G \sum_i \sum_{j \neq i} \frac{m_i m_j}{r_{ij}^3} \frac{d\bar{r}_i}{dt} \cdot (\bar{r}_j - \bar{r}_i) \quad (2.15)$$

We now introduce the notation

$$K = \sum_i \sum_{j \neq i} \frac{m_i m_j}{r_{ij}^3} \frac{d\bar{r}_i}{dt} \cdot (\bar{r}_j - \bar{r}_i)$$

and write

$$K = - \sum_i \sum_{j \neq i} \frac{m_i m_j}{r_{ij}^3} \frac{d(\bar{r}_j - \bar{r}_i)}{dt} \cdot (\bar{r}_j - \bar{r}_i) + \sum_i \sum_{j \neq i} \frac{m_i m_j}{r_{ij}^3} \frac{d\bar{r}_j}{dt} \cdot (\bar{r}_j - \bar{r}_i)$$

When we use the fact that within the double summation the indices i and j may be interchanged, we can write this relation as

$$K = - \frac{1}{2} \sum_i \sum_{j \neq i} \frac{m_i m_j}{r_{ij}^3} \frac{d}{dt} ((\bar{r}_j - \bar{r}_i) \cdot (\bar{r}_j - \bar{r}_i)) - K$$

or

$$K = \frac{1}{2} \sum_i \sum_{j \neq i} m_i m_j \frac{d}{dt} \left(\frac{1}{r_{ij}} \right) \quad (2.16)$$

Substitution of (2.16) into (2.15) yields

$$\frac{d}{dt} \left(\sum_i \frac{1}{2} m_i V_i^2 \right) = \frac{d}{dt} \left(\frac{1}{2} G \sum_i \sum_{j \neq i} \frac{m_i m_j}{r_{ij}} \right)$$

where V_i is the velocity of body i . Integration of this equation leads to

$$\sum_i \frac{1}{2} m_i V_i^2 - \frac{1}{2} G \sum_i \sum_{j \neq i} \frac{m_i m_j}{r_{ij}} = C \quad (2.17)$$

The first term in (2.17) represents the *total kinetic energy*, \mathcal{E}_k , of the system of n bodies. The second term (including the minus sign), which actually expresses some kind of internal energy of the system, is called the *total potential energy*, \mathcal{E}_p , of the system. Consequently, (2.17) can be written as

$$\mathcal{E}_k + \mathcal{E}_p = C \quad (2.18)$$

This proves that, while the sum of kinetic energy and potential energy of an individual body is, generally, not constant, the sum of kinetic energy and potential energy of the entire system is constant.

Equation (2.17) shows that when two (or more) bodies collide, a singularity occurs while all distances r_i remain finite. This equation also shows that when two (or more) bodies approach each other very closely ($r_{ij} \rightarrow 0$), the velocity of at least one body will become very large ($V_i \rightarrow \infty$) and at least one of the bodies may escape from the system ($r_i \rightarrow \infty$). This would result in a so-called *non-collision singularity*. For a long time, an intriguing question was whether it could be

proved that without collisions one of the bodies could really be ejected to infinity in finite time? P. Painlevé (1863-1933) already conjectured that this was possible indeed for $n > 3$. The concern was finally resolved by Z. Xia (1962-) in 1988; he proved that three-dimensional examples of such ejections exist for all $n \geq 5$.

Above, we have proved that for a closed (isolated) system of n bodies, in which the only forces acting upon the bodies are the mutual gravitational forces according to Newton's law of gravity: 1) the total linear momentum is constant; 2) the barycenter remains at rest or performs a uniform rectilinear motion; 3) the total angular momentum is constant; 4) the total energy is constant. In fact, these conservation laws hold for any system that is not affected by external forces or torques and that does not experience an energy exchange with its surroundings, and for any type of internal forces. It is emphasized that in our analysis we have assumed that the total energy of the system of n bodies consists of gravitational potential energy and kinetic energy only, and we have excluded other forms of energy.

These conservation laws were expressed through ten independent algebraic integrals of the general n -body problem; i.e. ten integrals that are algebraic functions of position, velocity and time. The first general theorems on the dynamics of n -body systems were given by I. Newton (1643-1727) in his *Principia*; they relate to the motion of the barycenter. L. Euler (1707-1783) appears to have been the first to develop celestial mechanics much beyond the state in which Newton left it; the ten general integrals were known to him. E.H. Bruns (1848-1919) proved in 1887 that if rectangular coordinates are used as variables, besides these ten integrals there can be no other algebraic integrals independent of these ten integrals. This does not, of course, exclude the possibility of algebraic integrals when other variables are used. A generalization of Brun's theorem is due to Painlevé, who showed in 1898 that any integral of the n -body problem which is an algebraic function of the velocities and is analytic in the coordinates is a combination of the classical integrals. It is emphasized that these theorems do not prove that the n -body problem is unsolvable. Mathematically, they only show that a certain method, i.e. solving a system of differential equations by finding (first) integrals, fails to solve the problem! J.H. Poincaré (1854-1912) has demonstrated that the problem of three bodies admits no new uniform transcendental integrals, even when the masses of two of the bodies are very small compared to the mass of the third body. In his theorem, the dependent variables are the elements of the orbits of the bodies, which continually change under their mutual attractions. However, from this theorem it does not follow that integrals of the class considered by Poincaré do not exist when other dependent variables are employed. In fact, T. Levi-Civita (1873-1941) has shown the existence of this class of integrals in a special problem, which comes under Poincaré's theorem, when suitable variables are used. The practical importance of the theorems of Bruns, Painlevé and Poincaré have often been overrated by those who have forgotten the conditions under which they have been proved to hold.

To determine the orbits of the n bodies under the influence of their mutual attractions with respect to an inertial reference frame, a set of $3n$ scalar second-order differential equations has to be solved, which is equivalent to the determination of $6n$ integrals. Of these $6n$ integrals, only ten are known. Therefore, the many-body problem can be reduced to solving a set of first-order differential equations of the order $6n-10$. These reductions have actually been carried out, especially for the problem of three bodies, by various researchers, starting with J.L. Lagrange (1736-1813) in 1772, K.G.J. Jacobi (1804-1851) in 1842, J.C.R. Radau (1835-1911) in 1868 and Poincaré in 1896, and more recently by V.I. Arnol'd (1937-2010) in 1985. A further reduction by two orders is possible by the elimination of the time, i.e. by the use of one of the dependent variables as the independent variable, and by the so-called *elimination of the nodes*. So, an

ultimate reduction to the order $6n-12$ is possible. The method of the elimination of the nodes was originally developed for a three-body system by Jacobi in 1842. Since then it has been extended and applied to a system of four bodies up to a system of n bodies by T.L. Bennett (-) in 1904, Y. Hagihara (1897-1979) in 1970, F. Boigey (-) in 1981, A. Deprit (1926-2006) in 1983, and others. In essence, this quite complicated method comes down to an appropriate selection of the orientation of the barycentric inertial reference frame (or a series of reference frames) such that a coordinate transformation with preferred mathematical characteristics is possible. However, these further reductions have little practical value, since the resulting expressions are rather complicated. In 1912, the Finnish astronomer K.F. Sundman (1873-1949) published a series expansion in powers of $t^{1/3}$ for the coordinates of each of the bodies of the three-body problem. That series expansion is convergent for all t , except for the case that the angular momentum of the system is zero. Q. Wang (-) has generalized this kind of analysis and published in 1991 similar convergent series expansions for any $n \geq 3$. His method excludes the case of solutions leading to singularities—collisions in particular. These series expansions, however, have only very limited practical meaning. Although convergent, they show a slow convergence. One would have to sum up very many terms to determine the motion of the point masses with moderate accuracy, even for short intervals of time. Then, the round-off errors make these series unusable in numerical analyses.

The ten integrals found for the many-body problem, and especially the angular momentum and energy integrals, are very useful in classical celestial mechanics to check numerical calculations. They enable us to get an idea of the inevitable accumulation of numerical errors. They are also important for the physical interpretation of various problems in celestial mechanics. However, we have to realize that when we consider the orbit of an Earth satellite or the trajectory of an interplanetary spacecraft, the application of the conservation laws for angular momentum and energy has little practical value. The reason is that the mass of a spacecraft is very small in comparison to the mass of a celestial body. Even if the positions of the Moon and the planets would be known up to twenty significant digits, which they are not, a characteristic ratio in the order of 10^{22} between the mass of the planets and the mass of a spacecraft would still cause errors in the calculation of the total energy of the system that are a hundred times larger than the total energy of the spacecraft. In other words: the orbit of the spacecraft can deviate very much without any noticeable change in the total energy of the system.

2.2. Motion relative to the barycenter

Equation (2.3) holds for the motion of body i (point mass) relative to an arbitrary inertial reference frame, where it is assumed that only the gravitational attraction forces between body i and a number of bodies j ($j \neq i$) act on body i . We now consider the motion of body i relative to a non-rotating reference frame with its origin at the barycenter (center of mass) of the system of n bodies. As shown before, this reference frame is an inertial reference frame. Hence, the integrals of motion are also valid with respect to this reference frame.

Since the equation of motion (2.3) holds for any inertial reference frame, we may write for the motion of body i relative to the barycentric inertial reference frame:

$$\frac{d^2\bar{r}_i}{dt^2} = \sum_{j \neq i} G \frac{m_j}{r_{ij}^3} \bar{r}_{ij} \quad (2.19)$$

where the vector \bar{r}_i now denotes the position vector of body i in the barycentric reference frame. According to (2.1) we can write

$$\mathbf{m}_j \bar{\mathbf{r}}_{ij} = \mathbf{m}_j \bar{\mathbf{r}}_j - \mathbf{m}_j \bar{\mathbf{r}}_i$$

Since $\bar{\mathbf{r}}_{ii} = 0$, we thus can write

$$\sum_{j \neq i} \mathbf{m}_j \bar{\mathbf{r}}_{ij} = \sum_j \mathbf{m}_j \bar{\mathbf{r}}_{ij} = \sum_j \mathbf{m}_j \bar{\mathbf{r}}_j - \sum_j \mathbf{m}_j \bar{\mathbf{r}}_i$$

Since the barycenter is the origin of the reference frame, we conclude from (2.4) that the first term on the right-hand side of the expression given above is zero, and we obtain

$$\sum_{j \neq i} \mathbf{m}_j \bar{\mathbf{r}}_{ij} = -\sum_j \mathbf{m}_j \bar{\mathbf{r}}_i$$

or

$$G \sum_{j \neq i} \frac{\mathbf{m}_j}{\mathbf{r}_i^3} \bar{\mathbf{r}}_{ij} + G \sum_j \frac{\mathbf{m}_j}{\mathbf{r}_i^3} \bar{\mathbf{r}}_i = 0 \quad (2.20)$$

Subtraction of (2.20) from the right-hand side of (2.19) leads to

$$\frac{d^2 \bar{\mathbf{r}}_i}{dt^2} = G \sum_{j \neq i} \frac{\mathbf{m}_j}{\mathbf{r}_{ij}^3} \bar{\mathbf{r}}_{ij} - G \sum_{j \neq i} \frac{\mathbf{m}_j}{\mathbf{r}_i^3} \bar{\mathbf{r}}_{ij} - G \sum_j \frac{\mathbf{m}_j}{\mathbf{r}_i^3} \bar{\mathbf{r}}_i$$

or

$$\frac{d^2 \bar{\mathbf{r}}_i}{dt^2} = -G \frac{M}{\mathbf{r}_i^3} \bar{\mathbf{r}}_i + G \sum_{j \neq i} \mathbf{m}_j \left(\frac{1}{\mathbf{r}_{ij}^3} - \frac{1}{\mathbf{r}_i^3} \right) \bar{\mathbf{r}}_{ij} \quad (2.21)$$

where M is the total mass of the system of n bodies. This remarkable expression is known as the *barycentric form of the equation of motion*, and shows that the motion of body i relative to the barycentric inertial reference frame can be described as a superposition of two components. The first term on the right-hand side describes the so-called *two-body motion* of body i about the barycenter, where the entire mass of the system of n bodies is assumed to be concentrated at the barycenter and only the gravitational attraction between that fictitious mass and body i is accounted for; the second term describes a motion component that is the result of the attractions between body i and all bodies $j \neq i$. In Sections 5.2 and 5.3 it will be shown that the two-body motion leads to an orbit that has the shape of a conic section (circle, ellipse, parabola, hyperbola), and that this orbit can be computed in a closed analytical way. It is emphasized that (2.19) and (2.21) are equivalent; both describe the motion of body i relative to a non-rotating barycentric reference frame. In general, (2.19) is used for numerical analyses, because of its simpler structure. However, in cases where the second term on the right-hand side of (2.21) is small compared to the first term on the right-hand side, which implies that the gravitational attraction by the bodies j only produce a perturbation of the two-body motion of body i , then (2.21) certainly has to be preferred over (2.19). The reason is that then only a perturbative term, i.e. the second term on the right-hand side of (2.21), has to be integrated numerically, which generally allows larger integration steps and yields smaller integration errors. In Section 4.2 it will be proved that, both for the computation of the orbits of planets about the Sun and of the orbits of satellites about the Earth, all bodies j produce only small perturbing accelerations superimposed on the main acceleration that results from the gravitational attraction between Sun and planet or Earth and satellite, respectively.

There is an interesting aspect associated with (2.21) when we use it to compute the orbit of

the Earth, or of any other planet of our solar system. Then, the Sun is one of the bodies j in the second term on the right-hand side of (2.21), and so the mass of the Sun is not only included in the parameter M , but also appears in the perturbative term. Since the mass of the Sun is more than thousand times the mass of any other body of our solar system, $\sum m_j \approx M$ and it might seem that (2.21) would have little value for solar-system problems. This is, however, not the case, as will be shown below. When, in the second term on the right-hand side of (2.21), we separate the effect of the gravitational attraction by the Sun on the motion of body i from the effects of the attraction by all other bodies j , we can write (2.21) as

$$\frac{d^2 \bar{r}_i}{dt^2} = -G \frac{M}{r_i^3} \bar{r}_i + G m_S \left(\frac{1}{r_{iS}^3} - \frac{1}{r_i^3} \right) \bar{r}_{iS} + G \sum_{j \neq i, S} m_j \left(\frac{1}{r_{ij}^3} - \frac{1}{r_i^3} \right) \bar{r}_{ij} \quad (2.22)$$

where the index S refers to the Sun. Since the Sun makes up about 99.86% of the total mass of the solar system, the barycenter of the solar system is near the center of the Sun. In the most unfavorable case, namely when all planets would line up on one side of the Sun, the barycenter would still be only at a distance of about two solar radii, or about one percent of the distance between Sun and Earth, measured from the center of the Sun. Consequently, we may write $r_{iS} = r_i + \Delta$, where Δ is of the order of x_S, y_S, z_S , i.e. the coordinates of the Sun in the barycentric reference frame; these quantities are small compared to r_i for all planets. We then find:

$$\left| \frac{1}{r_{iS}^3} - \frac{1}{r_i^3} \right| = \left| \frac{1}{(r_i + \Delta)^3} - \frac{1}{r_i^3} \right| \approx 3 \frac{|\Delta|}{r_i^4} = O\left(\frac{|x_S|}{r_i^4}\right) \quad (2.23)$$

where the notation O denotes ‘of the order of’. Since we describe positions relative to the barycentric reference frame, we find from (2.4):

$$m_S x_S + \sum_{j \neq S} m_j x_j = 0 \quad (2.24)$$

Substitution of (2.23) and (2.24) into the second term on the right-hand side of (2.22) leads to

$$G m_S \left| \left(\frac{1}{r_{iS}^3} - \frac{1}{r_i^3} \right) \right| = O\left(G \frac{1}{r_i^4} \sum_{j \neq S} m_j |x_j|\right)$$

which shows that m_S effectively cancels out. Consequently, the entire second term on the right-hand side of (2.21) produces only a perturbation of the two-body orbit of a planet about the barycenter of the solar system. In Section 4.1, the application of (2.21) for perturbed trajectories will be discussed briefly.

2.3. Polar moment of inertia, angular momentum and energy

In this Section we continue our analysis of the motion of the n bodies relative to a non-rotating (inertial) reference frame with its origin at the barycenter of the system of the n bodies. The position of a body (point mass) with respect to this reference frame will be indicated by x, y en z . The polar moment of inertia, I , of the system is then given by

$$I = \sum_i m_i \bar{r}_i \cdot \bar{r}_i = \sum_i m_i r_i^2 \quad (2.25)$$

We may derive an interesting equation where I is expressed in the mutual distances between the

bodies. For that, we consider again a particular body i and denote an other body by the index j . Then,

$$\begin{aligned}\sum_j \mathbf{m}_j \mathbf{r}_{ij}^2 &= \sum_j \mathbf{m}_j (\bar{\mathbf{r}}_j - \bar{\mathbf{r}}_i) \cdot (\bar{\mathbf{r}}_j - \bar{\mathbf{r}}_i) = \sum_j \mathbf{m}_j (\mathbf{r}_j^2 - 2\bar{\mathbf{r}}_i \cdot \bar{\mathbf{r}}_j + \mathbf{r}_i^2) \\ &= \sum_j \mathbf{m}_j \mathbf{r}_j^2 - 2\bar{\mathbf{r}}_i \cdot \sum_j \mathbf{m}_j \bar{\mathbf{r}}_j + \mathbf{r}_i^2 \sum_j \mathbf{m}_j\end{aligned}\quad (2.26)$$

where the summation is taken over all bodies j of the n -body system. Since the motion is considered relative to an inertial reference frame attached to the barycenter of the n -body system, we can write with (2.4):

$$\sum_j \mathbf{m}_j \bar{\mathbf{r}}_j = 0$$

Hence, (2.26) simplifies to

$$\sum_j \mathbf{m}_j \mathbf{r}_{ij}^2 = I + M \mathbf{r}_i^2 \quad (2.27)$$

where M is the total mass of the system. Multiplication of (2.27) by m_i and subsequent summation for all i , yields with (2.25)

$$\sum_i \sum_j \mathbf{m}_i \mathbf{m}_j \mathbf{r}_{ij}^2 = I \sum_i \mathbf{m}_i + M \sum_i \mathbf{m}_i \mathbf{r}_i^2 = 2M I$$

or

$$I = \frac{\sum_i \sum_j \mathbf{m}_i \mathbf{m}_j \mathbf{r}_{ij}^2}{2M} \quad (2.28)$$

We will use this relation later.

Differentiating (2.15) with respect to time yields

$$\begin{aligned}\frac{dI}{dt} &= 2 \sum_i \mathbf{m}_i \bar{\mathbf{r}}_i \cdot \frac{d\bar{\mathbf{r}}_i}{dt} \\ \frac{d^2I}{dt^2} &= 2 \sum_i \mathbf{m}_i \left(\frac{d\mathbf{r}_i}{dt} \right)^2 + 2 \sum_i \mathbf{m}_i \bar{\mathbf{r}}_i \cdot \frac{d^2\bar{\mathbf{r}}_i}{dt^2}\end{aligned}\quad (2.29)$$

Substitution of (2.3) into (2.29-2) gives

$$\frac{d^2I}{dt^2} = 2 \sum_i \mathbf{m}_i V_i^2 + 2G \sum_i \sum_{j \neq i} \frac{\mathbf{m}_i \mathbf{m}_j}{\mathbf{r}_{ij}^3} \bar{\mathbf{r}}_i \cdot \bar{\mathbf{r}}_{ij} \quad (2.30)$$

We now introduce an auxiliary variable

$$K = \sum_i \sum_{j \neq i} \frac{\mathbf{m}_i \mathbf{m}_j}{\mathbf{r}_{ij}^3} \bar{\mathbf{r}}_i \cdot \bar{\mathbf{r}}_{ij} = \sum_i \sum_{j \neq i} \frac{\mathbf{m}_i \mathbf{m}_j}{\mathbf{r}_{ij}^3} (\bar{\mathbf{r}}_j \cdot \bar{\mathbf{r}}_{ij} - \bar{\mathbf{r}}_{ij} \cdot \bar{\mathbf{r}}_{ij})$$

Because in a double summation the indices i and j may be interchanged, we find from this relation

$$2K = - \sum_i \sum_{j \neq i} \frac{\mathbf{m}_i \mathbf{m}_j}{\mathbf{r}_{ij}} \quad (2.31)$$

Substitution of (2.31) into (2.30) gives

$$\frac{d^2I}{dt^2} = 4\mathcal{E}_K - G \sum_i \sum_{j \neq i} \frac{\mathbf{m}_i \mathbf{m}_j}{\mathbf{r}_{ij}}$$

or, with (2.17),

$$\frac{d^2I}{dt^2} = 4\mathcal{E}_k + 2\mathcal{E}_p \quad (2.32-1)$$

or, with (2.18),

$$\frac{d^2I}{dt^2} = 4C - 2\mathcal{E}_p \quad ; \quad \frac{d^2I}{dt^2} = 2C + 2\mathcal{E}_k \quad (2.32-2)$$

The relations (2.32) are known as the *Lagrange-Jacobi identity* and were derived by J.L. Lagrange (1736-1813) in 1772, and extended by Jacobi. From the definition of \mathcal{E}_p it follows that the value of \mathcal{E}_p is always negative. Of course, \mathcal{E}_k is always positive. Hence, from (2.32-2) follows that if $C \geq 0$, the quantity d^2I/dt^2 is always positive. Thus, even if dI/dt would be negative initially, dI/dt will become positive after some time and I will increase unboundedly with time. This means that at least one of the bodies will move unboundedly far away from the origin and will escape from the system, and we are dealing with an *unstable system*. So, a necessary, but not sufficient, condition for a *stable system* is $C < 0$; i.e. the sum of kinetic and potential energy should be negative. That this is a non-sufficient condition follows directly from the fact that, according to (2.32-2), for certain values of \mathcal{E}_p and \mathcal{E}_k , negative values of C can exist that still yield $d^2I/dt^2 > 0$. It should be noted that stability does not necessarily demand that $d^2I/dt^2 < 0$ holds continuously; it is sufficient that dI/dt and d^2I/dt^2 vary in such a way that I never becomes unboundedly large. However, no general statement about this required variation can be made.

The (constant) angular momentum of the n -body system may be written with (2.10) as

$$\bar{H} = \sum_i \mathbf{m}_i \bar{\mathbf{r}}_i \times \bar{\mathbf{V}}_i$$

So, the magnitude of the angular momentum is given by

$$H = \sum_i \mathbf{m}_i \mathbf{r}_i \cdot \mathbf{V}_i |\cos \gamma_i| = \sum_i \left[(\sqrt{m_i} r_i) (\sqrt{m_i} V_i) |\cos \gamma_i| \right] \quad (2.33)$$

where γ_i is the angle between \mathbf{V}_i and the normal to $\bar{\mathbf{r}}_i$. In the two-body problem (Chapter 5) this angle is called the *flight path angle*; it was already introduced in Section 1.7. When we apply *Cauchy's inequality*, which states that for arbitrary a and b :

$$|\sum a b|^2 \leq (\sum a^2)(\sum b^2)$$

to (2.33), we find

$$H^2 \leq \left(\sum_i m_i r_i^2 \right) \left(\sum_i m_i V_i^2 \right) \cos^2 \gamma_i \quad (2.34)$$

We may write (2.29-1) as

$$\frac{dI}{dt} = 2 \sum_i \mathbf{m}_i \bar{\mathbf{r}}_i \cdot \bar{\mathbf{V}}_i = 2 \sum_i \mathbf{m}_i \mathbf{r}_i \mathbf{V}_i \sin \gamma_i$$

or

$$\frac{dI}{dt} = 2 \sum_i [\sqrt{\mathbf{m}_i} \mathbf{r}_i] [\sqrt{\mathbf{m}_i} \mathbf{V}_i] \sin \gamma_i$$

Application of Cauchy's inequality to this expression yields

$$\frac{1}{4} \left(\frac{dI}{dt} \right)^2 \leq \left(\sum_i \mathbf{m}_i \mathbf{r}_i^2 \right) \left(\sum_i \mathbf{m}_i \mathbf{V}_i^2 \right) \sin^2 \gamma_i \quad (2.35)$$

Combining (2.34) and (2.35) yields

$$H^2 \leq \left(\sum_i \mathbf{m}_i \mathbf{r}_i^2 \right) \left(\sum_i \mathbf{m}_i \mathbf{V}_i^2 \right) - \frac{1}{4} \left(\frac{dI}{dt} \right)^2$$

or

$$H^2 \leq 2I\mathcal{E}_k - \frac{1}{4} \left(\frac{dI}{dt} \right)^2$$

Substitution of (2.32-2) into this relation gives

$$H^2 \leq I \left\{ \frac{d^2 I}{dt^2} - 2C \right\} - \frac{1}{4} \left(\frac{dI}{dt} \right)^2 \quad (2.36)$$

This so-called *inequality of Sundman* provides a relation between the angular momentum of the system of n bodies, its polar moment of inertia and the first- and second-derivatives of the polar moment of inertia, and its total energy. Often, a somewhat weaker form of this inequality is used:

$$H^2 \leq I \left\{ \frac{d^2 I}{dt^2} - 2C \right\} \quad (2.37)$$

We will use this form in Sections 2.4 and 2.5.

2.4. Evolution of n -body systems

In Section 2.2 we have found that if $C \geq 0$, the system of n bodies is always unstable; if $C < 0$, the system may be stable or unstable. In this Section some interesting relations that provide information about the evolution of the system will be derived.

Systems with $C > 0$

We know that for $C > 0$ the system is unstable, which implies that it will expand unboundedly. Now the question may be asked: "How fast will the system expand?" To answer this question, we consider (2.32-2), which, because $\mathcal{E}_p < 0$, $\mathcal{E}_k > 0$, $C > 0$, and both expressions (2.32-2) have to be satisfied, leads to the condition

$$\frac{d^2 I}{dt^2} > 4C \quad (2.38)$$

Since $C > 0$, and consequently $d^2I/dt^2 > 0$, integration of (2.38) yields

$$I > 2Ct^2 + A_1 t + A_2 \quad (2.39)$$

where A_1 and A_2 are integration constants; the values of these constants are, in general, not equal to zero. The right-hand side of (2.39) represents a set of second-degree polynomials, which have their minimum values: $-A_1^2/8C + A_2$ at $t = -A_1/4C$. For an expanding system, $I > 0$ for all values of $t \geq 0$. With these conditions, we obtain

$$A_2 > 0 \quad ; \quad -\frac{A_1^2}{8C} + A_2 > 0$$

Because $C > 0$ and $A_2 > 0$, the second inequality can be written as

$$C - \frac{A_1^2}{8A_2} > 0 \quad (2.40)$$

Now, it is postulated that, irrespective of the values of C , A_1 and A_2 , it is always possible to find a positive constant E , for which the following expression holds:

$$2Ct^2 + A_1 t + A_2 > 2Et^2 \quad (2.41)$$

This assertion can be proved as follows. From (2.41) follows that

$$2(C - E)t^2 + A_1 t + A_2 > 0$$

This second-degree equation in t has no roots and consequently

$$A_1^2 < 8(C - E)A_2$$

Since $A_2 > 0$, we may write this inequality as

$$8E < 8C - \frac{A_1^2}{A_2}$$

Because E is assumed to be positive, this relation shows that the inequality (2.41) can only be true if

$$0 < E < C - \frac{A_1^2}{8A_2} \quad (2.42)$$

From (2.40) and (2.42) follows that indeed there exists a value of $E > 0$ that satisfies (2.41). So, (2.39) may be written as

$$I > 2Et^2 \quad (2.43)$$

According to (2.28), we may write for the polar moment of inertia:

$$I = \frac{\sum_i \sum_j m_i m_j r_{ij}^2}{2M} < \frac{r_{max}^2 \sum_i \sum_j m_i m_j}{2M} \quad (2.44)$$

where $r_{max} = \text{MAX}(r_{ij})$, i.e. r_{max} is the maximum value of all distances between two bodies at a certain time. When we consider that stage of the expanding system for which the total mass is

constant, we can write according to (2.44)

$$I < \frac{1}{2}Mr_{\max}^2 \quad (2.45)$$

Combining this relation with (2.43) yields

$$2Et^2 < I < \frac{1}{2}Mr_{\max}^2$$

So,

$$r_{\max}^2 > \frac{4E}{M}t^2 \quad (2.46)$$

Since E and M are positive constants for a given system of n bodies, this relation shows that r_{\max} has to increase more than linearly with time. Hence, the expansion of the system occurs according to a kind of accelerating process, in which the rate of expansion increases with time.

Systems with $C = 0$

For this case, the system is also unstable and integration of (2.38) leads to

$$I > A_1 t + A_2 \quad (2.47)$$

where the values of the integration constants are, in general, not equal to zero. The right-hand side of (2.47) represents a set of lines. Because for an expanding system $I > 0$ for all values of $t \geq 0$, we find as constraints for the values of A_1 and A_2 : $A_1 > 0, A_2 > 0$. Now, it is postulated that, irrespective of the values of A_1 and A_2 , it is always possible to find a positive constant E , for which the following expression holds:

$$A_1 t + A_2 > Et \quad (2.48)$$

This assertion can be proved as follows. From (2.48) we find

$$(A_1 - E)t + A_2 > 0$$

This inequality can, for all values of t , and for $A_1 > 0, A_2 > 0$, only be satisfied if

$$0 < E < A_1$$

So, indeed there exists a value of $E > 0$ that satisfies (2.48). Combination of (2.47) and (2.48) yields

$$I > Et \quad (2.49)$$

Combination of (2.45) and (2.49) yields

$$Et < I < \frac{1}{2}Mr_{\max}^2$$

So,

$$r_{\max}^2 > \frac{2E}{M}t \quad (2.50)$$

Since $E > 0$ and $M > 0$, this relation shows that r_{\max} has to increase faster than $t^{1/2}$. Hence, the system is still expanding, but cases may occur in which the rate of expansion decreases with time.

Systems with $C < 0$

As explained in Section 2.2, these systems can be stable or unstable. Because $\mathcal{E}_p < 0$, $\mathcal{E}_k > 0$, and $C < 0$, and both expressions (2.32-2) have to be satisfied, the following condition holds:

$$\frac{d^2I}{dt^2} > 2C \quad (2.51)$$

which indicates that, because $C < 0$, there is a negative lower bound for d^2I/dt^2 . In this case, the simplified form of Sundman's inequality (2.37) can be written as

$$\frac{H^2}{2|C|} - I \leq \frac{I}{2|C|} \frac{d^2I}{dt^2} \quad (2.52)$$

Because H and C are constants, (2.52) shows that the sign of d^2I/dt^2 is determined solely by the instantaneous value of I . When a so-called *critical value* for the moment of inertia is introduced:

$$I_{cr} = \frac{H^2}{2|C|} \quad (2.53)$$

we find that for $I < I_{cr}$: $d^2I/dt^2 > 0$, which corresponds to an unbounded expansion. If $I > I_{cr}$, it may only be concluded from (2.52) that d^2I/dt^2 is larger than some negative value, which tells us nothing about the sign of d^2I/dt^2 .

Stable systems

We now consider the case that the system is stable, which means that the n bodies remain in a bounded region of space and that no body escapes from the system. We have already found that this requires that the distance between two or more bodies does not become too small. In the following, we will derive an expression for the upper bound of the minimum distances between the bodies in a stable system.

From the definition of \mathcal{E}_p follows:

$$-\mathcal{E}_p = \frac{1}{2} G \sum_i \sum_{j \neq i} \frac{\mathbf{m}_i \mathbf{m}_j}{r_{ij}} \leq \frac{1}{2} \frac{G}{r_{min}} \sum_i \sum_{j \neq i} \mathbf{m}_i \mathbf{m}_j \quad (2.54)$$

where r_{min} is the minimum distance between two bodies in the system at a certain time. When the notation

$$A = \frac{1}{2} G \sum_i \sum_{j \neq i} \mathbf{m}_i \mathbf{m}_j < \frac{1}{2} G M^2 \quad (2.55)$$

is introduced, where A is a known positive constant for a given system, (2.54) may be written as

$$-\mathcal{E}_p \leq \frac{A}{r_{min}} \quad (2.56)$$

For $C < 0$, (2.18) yields

$$\mathcal{E}_k = -\mathcal{E}_p - |C|$$

Since $\mathcal{E}_k \geq 0$, we find

$$-\mathcal{E}_p \geq |C| \quad (2.57)$$

Combining (2.56) and (2.57) yields

$$\frac{A}{r_{\min}} \geq |C|$$

Because r_{\min} has a positive value, we find with (2.55):

$$r_{\min} \leq \frac{A}{|C|} < \frac{1}{2} G \frac{M^2}{|C|} \quad (2.58)$$

Thus, for a given number of bodies with given masses and a given value of $C < 0$, we can always find an upper bound for the minimum distance between two bodies during the evolution of the system. The value of the upper bound increases when the total mass of system increases and/or when the absolute value of the constant C decreases.

When we average (2.32-1) over a time interval $0 - t_e$, we may write

$$\frac{1}{t_e} \int_0^{t_e} \frac{d^2 I}{dt^2} dt = \frac{4}{t_e} \int_0^{t_e} \bar{\mathcal{E}}_k dt + \frac{2}{t_e} \int_0^{t_e} \bar{\mathcal{E}}_p dt$$

or

$$\frac{1}{t_e} \left(\frac{dI}{dt} \right)_0^{t_e} = 4 \bar{\mathcal{E}}_k + 2 \bar{\mathcal{E}}_p$$

where the ‘bar’ indicates average values. Substitution of (2.29-1) into this expression yields

$$\frac{1}{t_e} \left[\sum_i m_i \bar{r}_i \cdot \frac{d\bar{r}_i}{dt} \right]_0^{t_e} = 2 \bar{\mathcal{E}}_k + \bar{\mathcal{E}}_p \quad (2.59)$$

In a stable system no collisions and no escapes occur. In other words: all bodies stay within a finite distance from the origin and the velocities of all bodies remain finite. In that case, the value of the expression between brackets in (2.59) will remain finite. Therefore, if the time interval $0 - t_e$ is chosen large enough, the left-hand side of (2.59) will approach zero. So, for a sufficiently long averaging period, we find for a stable system:

$$2 \bar{\mathcal{E}}_k + \bar{\mathcal{E}}_p = 0 \quad (2.60)$$

For each system we have

$$\mathcal{E}_k + \mathcal{E}_p = C$$

which implies

$$\bar{\mathcal{E}}_k + \bar{\mathcal{E}}_p = C \quad (2.61)$$

Combination of (2.60) and (2.61) yields

$$\bar{\mathcal{E}}_k = -\frac{1}{2} \bar{\mathcal{E}}_p = -C \quad (2.62)$$

This relation is known as the *virial theorem*. It shows that for a stable system the value of the total kinetic energy averaged over a long period of time is equal to minus half the average value of the total potential energy, and is equal to the negative value of the integration constant from relation (2.18). In essence, a similar analysis was used by R.J.E. Clausius (1822-1888) in his

studies on the mechanical nature of heat, and on the relation between pressure, volume and temperature in a non-ideal gas. He used the term ‘virial’ (derived from the Latin word ‘virias’ which means ‘forces’) for the first time around 1870 to denote what we would call today the internal energy of gas in a container. The virial theorem is therefore also known as the *Clausius’ theorem*. The virial theorem plays an important role in astrophysics. For instance, Poincaré applied a form of this theorem in 1911 to the problem of determining cosmological stability, and F. Zwicky (1898-1974) was the first to use the theorem to deduce the existence of ‘unseen matter’, what is nowadays called *dark matter*. S. Chandrasekhar (1910-1995) and E. Fermi (1901-1954) extended the virial theorem in 1953 for astrophysical applications to include the presence of magnetic fields. A simple application of the virial theorem concerns galaxy clusters. If a region of space is unusually full of galaxies, it is safe to assume that they have been together for a long time, and the theorem can be applied. Doppler measurements give lower bounds for the relative velocities of the galaxies, and the virial theorem then gives a lower bound for the total mass of the cluster, including any dark matter.

2.5. Total collision

When the n bodies all collide at the same time, we speak of a *total collision*. For such a collision, we know that at the last stage of the process: $r_{ij} \rightarrow 0$ for all i, j -combinations, $r_i \rightarrow 0$ and $I \rightarrow 0$. This indicates that the total collision occurs in the center of mass of the system. In this Section, the positions of the bodies are again described with respect to a non-rotating reference frame with its origin at the center of mass of the system of n bodies. First, it will be proved that if such a total collision occurs, it can never occur after an infinitely long time. To prove this statement, it will be shown that it is impossible that $I \rightarrow 0$ for $t \rightarrow \infty$.

For a total collision, $r_{ij} \rightarrow 0$ and we find from (2.17) and (2.18): $\mathcal{E}_p \rightarrow -\infty$. Since C is a finite constant, we find from (2.32-2): $d^2I/dt^2 \rightarrow \infty$. Consequently, for $t > t_1$, where t_1 is some point in time, the following relation holds:

$$\frac{d^2I}{dt^2} > A_0$$

where A_0 is a positive constant. Integration of this relation yields for $t > t_1$:

$$I > \frac{1}{2}A_0 t^2 + A_1 t + A_2 \quad (2.63)$$

where A_1 and A_2 are constants. From (2.63) follows that if $t \rightarrow \infty$: $I \rightarrow \infty$, which contradicts with the assumption: $t \rightarrow \infty$, $I \rightarrow 0$. So, we have proved that a total collision can only occur within a finite interval of time.

A theorem by Sundman states that a total collision can only occur if for the total system: $\bar{H} = 0$. This can be proved in the following way. For a system where none of the bodies escape, we know that $C < 0$. Now, it is assumed that the total collision occurs at $t = t_1$, where $0 - t_1$ indicates a finite time interval. Then, we have for $t \rightarrow t_1$: $I \rightarrow 0$, $\mathcal{E}_p \rightarrow -\infty$, $d^2I/dt^2 \rightarrow \infty$. So, for $t_2 \leq t \leq t_1$: $d^2I/dt^2 > 0$, where t_2 is an appropriately chosen point in time. However, the system can only undergo a total collision if $dI/dt < 0$ for $t_2 \leq t \leq t_1$. Multiplying the simplified form of Sundman’s inequality (2.37) by the positive quantity $-(dI/dt)/I$ yields

$$-\frac{H^2}{I} \frac{dI}{dt} \leq -\frac{dI}{dt} \left(\frac{d^2 I}{dt^2} - 2C \right)$$

Integration of this equation for $t > t_2$ yields

$$H^2 \ln \frac{1}{I} \leq 2CI - \frac{1}{2} \left(\frac{dI}{dt} \right)^2 + D$$

where D is a constant. Because $(dI/dt)^2 > 0$, we obtain

$$H^2 \leq \frac{2CI + D}{\ln(1/I)} \quad (2.64)$$

For $t \rightarrow t_1$: $I \rightarrow 0$ and $\ln(1/I) \rightarrow \infty$. So, according to (2.64): $H^2 \rightarrow 0$. But since H is constant, $H = 0$ should hold during the entire period of motion, which proves Sundman's theorem. Of course, $H = 0$ is a necessary, but not sufficient, condition for the total collision. Summarizing, it may be stated that in the final phase of a total collision the following conditions hold: $I \rightarrow 0$; $dI/dt < 0$; $d^2I/dt^2 \rightarrow \infty$.

2.6. Pseudo-inertial reference frames

Until now, we have considered the motion of celestial bodies (point masses) with respect to inertial reference frames. Although from a theoretical point of view the application of inertial reference frames is of fundamental importance, in practice we cannot use 'true' inertial reference frames and we are always forced to work with so-called *pseudo-inertial reference frames*. This implies that one deliberately neglects the accelerations and rotations of the pseudo-inertial reference frame relative to a 'true' inertial reference frame. As an example, consider the computation of the trajectory of a bullet. In that case, we may neglect the rotation of the Earth because the bullet's flight time is short. Although we know that the Earth's rotation gives rise to centrifugal and Coriolis accelerations, these accelerations can be neglected for the calculation of the trajectory of the bullet and therefore a pseudo-inertial reference frame may be chosen that is fixed to the Earth's surface. For the calculation of the trajectory of a ballistic missile, the rotation of the Earth may certainly not be neglected. However, in that case the motion of the Earth about the Sun may be neglected, and a non-rotating reference frame with its origin at the center of mass of the Earth can be chosen as pseudo-inertial reference frame. To describe the motion of bodies in our solar system, the origin of the pseudo-inertial reference frame should be chosen at the barycenter of the solar system. We then neglect the effects of bodies outside the solar system. The invariable plane of Laplace goes through this barycenter; the angle between this plane and the orbital plane of the Earth (ecliptic) is about $1^\circ 35'$ (Section 2.1). Because of the enormously large mass of the Sun with respect to the other masses in the solar system, the barycenter is near the center of the Sun and we often adopt a non-rotating reference frame with its origin at the center of the Sun as a pseudo-inertial reference frame. When one analyzes a two-body problem or a three-body problem, and describes the motions relative to a non-rotating reference frame with its origin at the barycenter of the two or three bodies, this reference frame is in fact a pseudo-inertial reference frame, because we know that the universe contains more bodies. In all the cases mentioned, we often omit the phrase 'pseudo', and speak for simplicity of an 'inertial reference frame'.

2.7. Angular momentum in the two-body problem

In the special case that the many-body problem reduces to a two-body problem (Chapter 5), an important relation for the orbital angular momentum of the bodies, which will already be used in Chapter 3, can be derived. When we describe the motion of the two bodies with respect to a (pseudo-)inertial reference frame $X'Y'Z'$ with its origin at the barycenter of the two bodies (Figure 2.2) that moves with a constant velocity relative to the inertial reference frame XYZ , we can write, according to (2.4),

$$\mathbf{m}_1 \bar{\mathbf{r}}_1 + \mathbf{m}_2 \bar{\mathbf{r}}_2 = 0$$

or

$$\bar{\mathbf{r}}_1 = -\frac{\mathbf{m}_2}{\mathbf{m}_1} \bar{\mathbf{r}}_2 \quad (2.65)$$

where $\bar{\mathbf{r}}_1$ and $\bar{\mathbf{r}}_2$ are the position vectors of the two bodies.

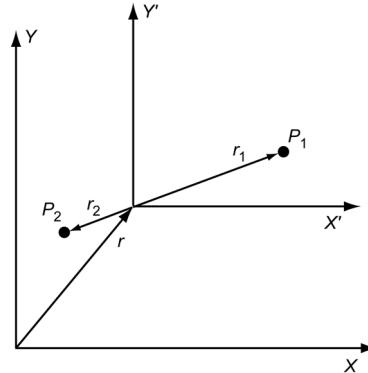


Figure 2.2: A pseudo-inertial reference frame and the positions of the bodies of the two-body problem relative to this reference frame.

For the two-body problem (2.10) reduces to

$$\bar{H} = \mathbf{m}_1 \bar{\mathbf{r}}_1 \times \frac{d\bar{\mathbf{r}}_1}{dt} + \mathbf{m}_2 \bar{\mathbf{r}}_2 \times \frac{d\bar{\mathbf{r}}_2}{dt} = c \quad (2.66)$$

Substitution of (2.65) into (2.66) yields

$$\begin{aligned} \bar{H} &= \mathbf{m}_1 \left(1 + \frac{\mathbf{m}_1}{\mathbf{m}_2} \right) \left(\bar{\mathbf{r}}_1 \times \frac{d\bar{\mathbf{r}}_1}{dt} \right) = \left(1 + \frac{\mathbf{m}_1}{\mathbf{m}_2} \right) \bar{H}_1 \\ \bar{H} &= \mathbf{m}_2 \left(1 + \frac{\mathbf{m}_2}{\mathbf{m}_1} \right) \left(\bar{\mathbf{r}}_2 \times \frac{d\bar{\mathbf{r}}_2}{dt} \right) = \left(1 + \frac{\mathbf{m}_2}{\mathbf{m}_1} \right) \bar{H}_2 \end{aligned} \quad (2.67)$$

Equations (2.67) indicate that the orbital angular momentum vectors of bodies 1 and 2 in their motion about the barycenter of the system have the same direction; namely the direction of \bar{H} . Hence, both bodies move in the same fixed plane about the barycenter, and the motion of the bodies about each other also occurs within that plane. The magnitude of the orbital angular momentum of each body is constant. Furthermore, from (2.65) we conclude that the two bodies will always be positioned diametrically opposite to each other and that the orbits described by

both bodies about the barycenter of the system are of the same shape (similar). As an example, Figure 2.3 shows elliptical orbits of the two bodies about the barycenter (absolute motion) and about each other (relative motion).

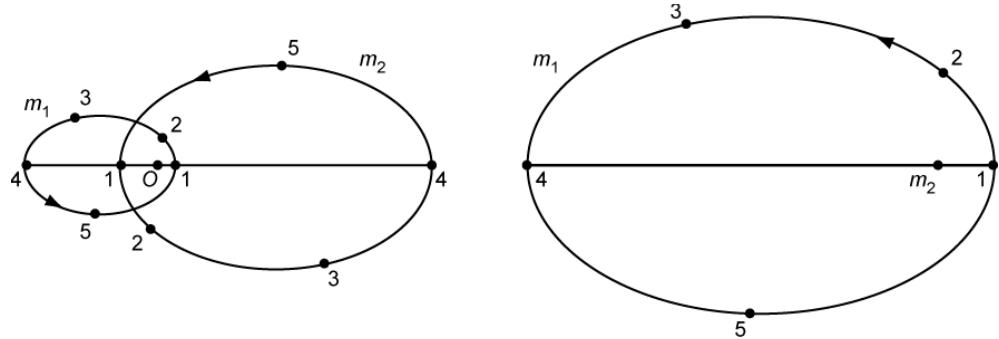


Figure 2.3: Absolute (left) and relative (right) motion of two bodies under the influence of their mutual gravitational attraction ($m_1/m_2 = 2$).

3. THREE-BODY PROBLEM

In modern terminology, the three-body problem may be stated as: “What are the motions of three given bodies (point masses) moving under the influence of their mutual gravitational attraction?” As such, knowledge of the three-body problem is of great importance in classical dynamical astronomy, e.g. to determine the motion of the Moon in the Sun-Earth-Moon system, and in astrodynamics, e.g. to determine the motion of a spacecraft in the Earth-Moon system. Throughout the last three centuries, the three-body problem has played a major role in the development of natural sciences. It has triggered many mathematical studies, methods and theories by L. Euler (1707-1783), J.L. Lagrange (1736-1813), P.S. Laplace (1749-1827), K.G.J. Jacobi (1804-1851), W.R. Hamilton (1805-1865), S. Newcomb (1835-1909), C.L. Siegel (1896-1981), and many others. The development of an accurate theory to describe the motion of the Moon about the Earth was competing during the eighteenth century with the progress in timekeeping systems for measuring longitudes on Earth.

It was Euler who stated the general three-body problem for the first time (as soon as 1727, in his diary), and who recognized the great difficulties arising in the solution of this problem. He realized that in the first instance the simplest cases have to be solved, e.g. the collinear case (Section 3.2), where the three bodies are always on a straight line; and the two-center problem, where the position of two bodies remains fixed. He also mentioned the triangular case (Section 3.2), where the three bodies always form an equilateral triangle; and the restricted three-body problem (Section 3.3), where the mass of one of the bodies being very small when compared with that of either of the other two bodies. The difficulties experienced in finding solutions for the three-body problem were the reason for the introduction of new qualitative analysis methods by J.H. Poincaré (1854-1912), G.D. Birkhoff (1884-1944) and others; methods which have since then been extended to many other branches of science. It is interesting to note that while in the early 1960’s meteorologists, like E.N. Lorenz (1917-2008), studied the chaotic behavior in atmospheric systems and have discovered the *strange attractor* (Section 1.3) by using modern computers in their analyses, such chaotic behavior had already been discovered by astronomers in numerical studies of the three-body system. These discoveries, in fact, are closely related to theoretical analyses of Poincaré and Birkhoff on the so-called *ergodic theorem* and that of A.N. Kolmogorov (1903-1987), V.I. Arnol’d (1937-2010) and J.K. Moser (1928-1999) on the behavior of orbits close to periodic motions. But, only during the last decades we start to understand how general these types of motions are in all kinds of dynamical systems.

The centuries of investigations in the three-body problem have reached the point where most theoreticians believe that, when each of the three masses is non-zero, all solutions are basically unstable, in the sense that at any place in the three-body space escape solutions are the most probable on the long run. However, this assertion is contested by most numerical analysts. If the theoreticians are correct, a new type of instability will appear: a kind of very-long-term stability, which is finally destroyed by very small and long-lasting resonance effects. Perhaps, this would even imply that the ultimate future of our solar system is dispersion (Section 5.8)!

In this Chapter, we will start with some general aspects of the three-body problem and with a discussion on a couple of famous solutions. Then, we will discuss in more detail the special case of the so-called *circular restricted three-body problem* and some practical applications of that problem.

3.1. Equations of motion

We assume that the forces on three bodies P_1 , P_2 and P_3 , with masses m_1 , m_2 , m_3 , are solely due

to the gravitational attractions between the bodies and that the bodies can be considered as point masses. In Figure 3.1, the geometry of the system is sketched; the inertial reference frame XYZ has its origin O at the mass center of the system of the three bodies. It is emphasized that the general situation is considered, where the motions of the bodies are not confined to a single plane.

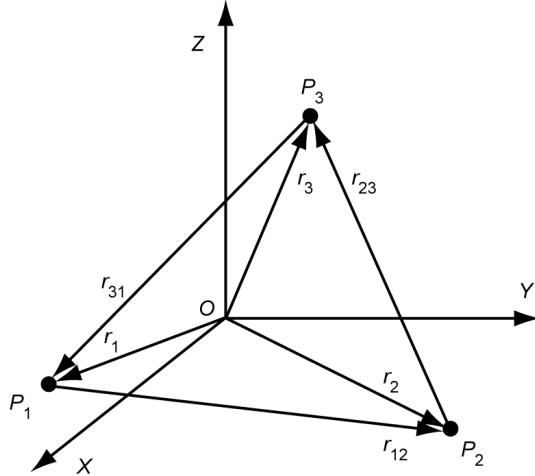


Figure 3.1: Geometry of the system of three bodies and the vectors used in the Euler and Lagrange formulation.

Using Newton's second law of motion and his law of gravitation (Chapter 1), we may write for the motion of the bodies:

$$\frac{d^2\bar{r}_i}{dt^2} = G \frac{m_j}{r_{ij}^3} \bar{r}_{ij} + G \frac{m_k}{r_{ik}^3} \bar{r}_{ik} \quad \{i,j,k\} = \{1,2,3\} \quad (3.1)$$

where

$$\bar{r}_{ij} = \bar{r}_j - \bar{r}_i \quad ; \quad r_{ij} = |\bar{r}_{ij}| \quad ; \quad \bar{r}_{ik} = \bar{r}_k - \bar{r}_i \quad ; \quad r_{ik} = |\bar{r}_{ik}| \quad (3.2)$$

The set of three second-order differential equations (3.1) represents the *classical* or *Euler formulation* of the three-body problem. When the position of the bodies is written in the rectangular coordinates x, y, z , we arrive at a set of first-order differential equations of the order eighteen. In Section 2.1 it was mentioned that by using the integrals of motion and applying the methods of the elimination of the time and the elimination of the nodes we can reduce the order of the set of differential equations. No general solutions have been obtained for the set of equations. Only some series expansion types of solutions and partial solutions for very special cases are known.

In the *Lagrange formulation* of the three-body problem, the variables are \bar{r}_{12} , \bar{r}_{23} , and \bar{r}_{31} (cyclic set of parameters). When we start with the vector \bar{r}_{12} , we may write

$$\bar{r}_{12} = \bar{r}_2 - \bar{r}_1$$

or

$$\frac{d^2\bar{r}_{12}}{dt^2} = \frac{d^2\bar{r}_2}{dt^2} - \frac{d^2\bar{r}_1}{dt^2} \quad (3.3)$$

Combination of (3.1) and (3.3) yields

$$\frac{d^2\bar{r}_{12}}{dt^2} = G \frac{m_3}{r_{23}^3} \bar{r}_{23} + G \frac{m_1}{r_{21}^3} \bar{r}_{21} - G \frac{m_2}{r_{12}^3} \bar{r}_{12} - G \frac{m_3}{r_{13}^3} \bar{r}_{13}$$

or

$$\frac{d^2\bar{r}_{12}}{dt^2} = G \left[m_3 \left(\frac{\bar{r}_{23}}{r_{23}^3} + \frac{\bar{r}_{31}}{r_{31}^3} \right) - (m_1 + m_2) \frac{\bar{r}_{12}}{r_{12}^3} \right] \quad (3.4-1)$$

In a similar way, we find

$$\frac{d^2\bar{r}_{23}}{dt^2} = G \left[m_1 \left(\frac{\bar{r}_{31}}{r_{31}^3} + \frac{\bar{r}_{12}}{r_{12}^3} \right) - (m_2 + m_3) \frac{\bar{r}_{23}}{r_{23}^3} \right] \quad (3.4-2)$$

$$\frac{d^2\bar{r}_{31}}{dt^2} = G \left[m_2 \left(\frac{\bar{r}_{12}}{r_{12}^3} + \frac{\bar{r}_{23}}{r_{23}^3} \right) - (m_3 + m_1) \frac{\bar{r}_{31}}{r_{31}^3} \right] \quad (3.4-3)$$

with, of course,

$$\bar{r}_{12} + \bar{r}_{23} + \bar{r}_{31} = 0 \quad (3.5)$$

This form of the equations of motion is often used as a starting point for analytical studies on the three-body problem. Note that each equation consists of a *two-body part* (the second term in brackets) and a part accounting for the attraction by the third body (the first term in brackets).

Jacobi has proposed to use other parameters to describe the positions of the three bodies and it turned out that his decomposition of the three-body problem is the most powerful for detailed analyses and e.g. forms the basis of *lunar theory* and triple stellar system problems. His decomposition is non-symmetric and he used two main vectors to describe the relative positions of the bodies (Figure 3.2): the vector \bar{r}_{12} and the vector \bar{R} from the center of mass of P_1 and P_2 to P_3 . This vector \bar{R} , of course, passes through the barycenter of the entire system. In Figure 3.2 these centers of mass are indicated by O_{12} and O , respectively.

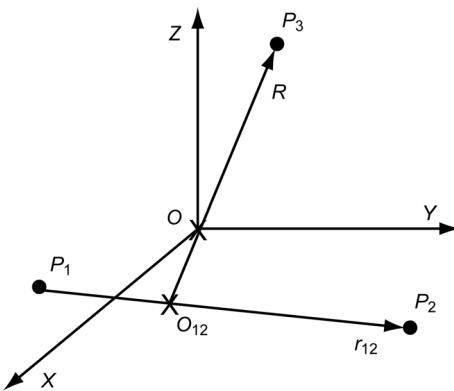


Figure 3.2: Jacobi method of describing the position of the three bodies.

Let

$$\alpha = \frac{m_1}{m_1 + m_2} \quad (3.6)$$

then

$$\begin{aligned}\bar{r}_{13} &= \bar{R} + (1 - \alpha) \bar{r}_{12} \\ \bar{r}_{23} &= \bar{R} - \alpha \bar{r}_{12}\end{aligned}\quad (3.7)$$

Multiplication of (3.7-1) with α and (3.7-2) with $(1-\alpha)$, and subsequently adding both relations gives

$$\bar{R} = \alpha \bar{r}_{13} + (1 - \alpha) \bar{r}_{23}$$

Differentiation of this equation results in

$$\frac{d^2 \bar{R}}{dt^2} = (1 - \alpha) \frac{d^2 \bar{r}_{23}}{dt^2} - \alpha \frac{d^2 \bar{r}_{31}}{dt^2}$$

Substitution of the Lagrange equations (3.4-2) and (3.4-3) gives

$$\begin{aligned}\frac{d^2 \bar{R}}{dt^2} &= (1 - \alpha) G \left[m_1 \left(\frac{\bar{r}_{31}}{r_{31}^3} + \frac{\bar{r}_{12}}{r_{12}^3} \right) - (m_2 + m_3) \frac{\bar{r}_{23}}{r_{23}^3} \right] \\ &\quad - \alpha G \left[m_2 \left(\frac{\bar{r}_{12}}{r_{12}^3} + \frac{\bar{r}_{23}}{r_{23}^3} \right) - (m_3 + m_1) \frac{\bar{r}_{31}}{r_{31}^3} \right]\end{aligned}$$

or

$$\frac{d^2 \bar{R}}{dt^2} = G \left[(m_1 - \alpha m_1 - \alpha m_2) \frac{\bar{r}_{12}}{r_{12}^3} + (-m_2 - m_3 + \alpha m_3) \frac{\bar{r}_{23}}{r_{23}^3} - (m_1 + \alpha m_3) \frac{\bar{r}_{31}}{r_{31}^3} \right] \quad (3.8)$$

From (3.6) we find

$$m_1 - \alpha m_1 - \alpha m_2 = 0$$

With this relation we may write

$$m_2 + m_3 - \alpha m_3 = m_1 + m_2 + m_3 - \alpha m_1 - \alpha m_2 - \alpha m_3 = (1 - \alpha) M$$

and

$$m_1 + \alpha m_3 = \alpha m_1 + \alpha m_2 + \alpha m_3 = \alpha M$$

where M is the total mass of the three bodies. Substitution of these relations into (3.8) finally yields

$$\frac{d^2 \bar{R}}{dt^2} = -G M \left[\alpha \frac{\bar{r}_{13}}{r_{13}^3} + (1 - \alpha) \frac{\bar{r}_{23}}{r_{23}^3} \right] \quad (3.9-1)$$

This equation and a slightly modified version of Lagrange equation (3.4-1):

$$\frac{d^2\bar{r}_{12}}{dt^2} = -G \left[(m_1 + m_2) \frac{\bar{r}_{12}}{r_{12}^3} + m_3 \left(\frac{\bar{r}_{13}}{r_{13}^3} - \frac{\bar{r}_{23}}{r_{23}^3} \right) \right] \quad (3.9-2)$$

form the *Jacobi set* of equations for the three-body problem. It is emphasized that these equations constitute a twelfth-order system.

As an application of the Jacobi set of equations, we consider the so-called *lunar case* and *planetary case*. In the lunar case, where P_1 is the Earth, P_2 the Moon and P_3 the Sun, we know that

$$\alpha \approx 1 \quad ; \quad \bar{r}_{13} \approx \bar{r}_{23} \approx \bar{R}$$

With these approximations, (3.9) can be simplified to

$$\begin{aligned} \frac{d^2\bar{r}_{12}}{dt^2} &= -G(m_1 + m_2) \frac{\bar{r}_{12}}{r_{12}^3} \\ \frac{d^2\bar{R}}{dt^2} &= -GM \frac{\bar{R}}{R^3} \end{aligned} \quad (3.10)$$

For the planetary case, with P_1 the Sun, P_2 the Earth and P_3 a planet, we have:

$$\alpha \approx 1 \quad ; \quad \frac{m_3}{m_1 + m_2} \ll 1 \quad ; \quad \bar{r}_{13} \approx \bar{R}$$

and we arrive at the same approximative equations of motion. Equation (3.10-1) describes the relative motion of bodies P_1 and P_2 ; (3.10-2) describes the motion of body P_3 relative to the center of mass of bodies P_1 and P_2 . So, we conclude that both for the lunar case and the planetary case the (relative) motions of the bodies may be approximated by a superposition of two two-body trajectories. It will be shown in Sections 5.2 and 5.3 that a two-body trajectory has the shape of a conic section; such orbits are generally called *Keplerian orbits*.

3.2. Central configuration solutions

Lagrange has found a particular case of three-body motion in which the mutual distances between the bodies remain constant, and Euler has extended this class of motion and has found solutions in which the ratios of the mutual distances remain constant. These classes of solutions refer to cases where the geometric shape of the three-body configuration does not change with time, although the scale may change and the configuration may rotate. Lagrange and Euler showed that for three bodies of arbitrary mass such solutions are possible if:

- The resultant force on each body passes through the barycenter of the system.
- The resultant force is proportional to the distance of a body from the barycenter of the system.
- The magnitudes of the initial velocity vectors are proportional to the respective distances of the bodies from the barycenter of the system, and these velocity vectors make equal angles with the radius vectors to the bodies from the barycenter of the system.

Because of these requirements, the solutions are generally referred to as *central configurations*. We will determine the possible configurations that satisfy these requirements.

Because in Figure 3.1 O is the barycenter of the system of three bodies, we may write

$$\mathbf{m}_1 \bar{\mathbf{r}}_1 + \mathbf{m}_2 \bar{\mathbf{r}}_2 + \mathbf{m}_3 \bar{\mathbf{r}}_3 = 0 \quad (3.11)$$

This relation can also be written as

$$M \bar{\mathbf{r}}_1 + \mathbf{m}_2 (\bar{\mathbf{r}}_2 - \bar{\mathbf{r}}_1) + \mathbf{m}_3 (\bar{\mathbf{r}}_3 - \bar{\mathbf{r}}_1) = 0$$

or

$$M \bar{\mathbf{r}}_1 = -\mathbf{m}_2 \bar{\mathbf{r}}_{12} - \mathbf{m}_3 \bar{\mathbf{r}}_{13} \quad (3.12)$$

where, again, M is the total mass of the system of three bodies. Scalar multiplication of this relation with itself yields

$$M^2 r_1^2 = \mathbf{m}_2^2 \mathbf{r}_{12}^2 + \mathbf{m}_3^2 \mathbf{r}_{13}^2 + 2\mathbf{m}_2 \mathbf{m}_3 \bar{\mathbf{r}}_{12} \cdot \bar{\mathbf{r}}_{13} \quad (3.13)$$

If the shape of the configuration does not alter, the distances r_{12} , r_{23} and r_{31} are given by

$$\frac{\mathbf{r}_{12}}{(\mathbf{r}_{12})_0} = \frac{\mathbf{r}_{23}}{(\mathbf{r}_{23})_0} = \frac{\mathbf{r}_{31}}{(\mathbf{r}_{31})_0} = f(t) \quad (3.14)$$

where the index 0 denotes the value of the distance at t_0 , i.e. the epoch¹ when the bodies are placed in the required configuration, and $f(t_0) = 1$. Combining (3.13) and (3.14), we obtain

$$M^2 r_1^2 = \{f(t)\}^2 \left[\mathbf{m}_2^2 (\mathbf{r}_{12})_0^2 + \mathbf{m}_3^2 (\mathbf{r}_{13})_0^2 + 2\mathbf{m}_2 \mathbf{m}_3 (\mathbf{r}_{12})_0 (\mathbf{r}_{13})_0 \cos \varphi \right]$$

where φ , the angle between $\bar{\mathbf{r}}_{12}$ and $\bar{\mathbf{r}}_{13}$, is constant. Since the term in brackets is constant, we may write $r_i = c f(t)$, where c is a constant. Because $f(t_0) = 1$, we find

$$\mathbf{r}_1 = (\mathbf{r}_1)_0 f(t)$$

or, generalizing this result,

$$\mathbf{r}_i = (\mathbf{r}_i)_0 f(t) \quad (3.15)$$

Since φ is constant, we can write for the angular velocity, ω_i , of body P_i about the barycenter

$$\omega_1 = \omega_2 = \omega_3 = \omega(t) \quad (3.16)$$

We know that the total angular momentum of the system about the origin (barycenter), \bar{H} , is constant, which means

$$\mathbf{m}_1 \mathbf{r}_1^2 \omega_1 + \mathbf{m}_2 \mathbf{r}_2^2 \omega_2 + \mathbf{m}_3 \mathbf{r}_3^2 \omega_3 = H \quad (3.17)$$

Combining (3.15), (3.16) and (3.17), we obtain

$$\left[\mathbf{m}_1 (\mathbf{r}_1)_0^2 + \mathbf{m}_2 (\mathbf{r}_2)_0^2 + \mathbf{m}_3 (\mathbf{r}_3)_0^2 \right] \{f(t)\}^2 \omega(t) = H \quad (3.18)$$

Because H is constant and the term in brackets is constant, also the product $\{f(t)\}^2 \omega(t)$ must be constant.

¹ Epoch is a moment in time used as a reference point for some time-varying astronomical quantity. In the older astronomical literature it was customary to denote as epoch not the reference time, but rather the values at that date and time of those time-varying quantities themselves.

The angular momentum of body i is given by $H_i = m_i \mathbf{r}_i^2 \omega_i$, or, with (3.15) and (3.16),

$$H_i = m_i (\mathbf{r}_i)_0^2 \{f(t)\}^2 \omega(t)$$

This proves that the angular momentum of each body about the barycenter is constant, which means that the resulting force acting on each body passes through the barycenter of the system. When the force per unit of mass acting on m_i is indicated by \mathbf{g}_i , we know from classical mechanics that, if the force acts along the radius vector, the equation of motion of m_i is

$$m_i \mathbf{g}_i = m_i (\ddot{\mathbf{r}}_i - \mathbf{r}_i \omega_i^2)$$

Substitution of (3.15) and (3.16) into this equation yields

$$m_i \mathbf{g}_i = m_i \left[(\mathbf{r}_i)_0 \frac{d^2 f(t)}{dt^2} - \mathbf{r}_i \{\omega(t)\}^2 \right]$$

or

$$m_i \mathbf{g}_i = m_i \mathbf{r}_i \left[\frac{1}{f(t)} \frac{d^2 f(t)}{dt^2} - \{\omega(t)\}^2 \right]$$

For each body the term in brackets is identical, which means that

$$g_1 : g_2 : g_3 = r_1 : r_2 : r_3 \quad (3.19)$$

Because the force acting on a body passes through the barycenter of the system, we have:

$$\bar{\mathbf{r}}_i \times \bar{\mathbf{g}}_i = 0$$

or, with Newton's second law of motion,

$$\bar{\mathbf{r}}_i \times \frac{d^2 \bar{\mathbf{r}}_i}{dt^2} = 0 \quad (3.20)$$

We now go back to (3.1) and take the vector product of $\bar{\mathbf{r}}_1$ with the left- and right-hand sides of (3.1) for $i = 1$. With (3.2) we then obtain

$$\bar{\mathbf{r}}_1 \times \frac{d^2 \mathbf{r}_1}{dt^2} = G \bar{\mathbf{r}}_1 \times \left(m_2 \frac{\bar{\mathbf{r}}_2}{\mathbf{r}_{12}^3} + m_3 \frac{\bar{\mathbf{r}}_3}{\mathbf{r}_{13}^3} \right)$$

or, with (3.20),

$$\bar{\mathbf{r}}_1 \times \left(m_2 \frac{\bar{\mathbf{r}}_2}{\mathbf{r}_{12}^3} + m_3 \frac{\bar{\mathbf{r}}_3}{\mathbf{r}_{13}^3} \right) = 0$$

Substitution of (3.11) into this equation gives

$$\bar{\mathbf{r}}_1 \times \left(m_2 \frac{\bar{\mathbf{r}}_2}{\mathbf{r}_{12}^3} - \frac{m_1 \bar{\mathbf{r}}_1 + m_2 \bar{\mathbf{r}}_2}{\mathbf{r}_{13}^3} \right) = 0$$

or

$$m_2 \bar{r}_1 \times \bar{r}_2 \left(\frac{1}{r_{12}^3} - \frac{1}{r_{13}^3} \right) = 0 \quad (3.21)$$

There are, of course, two similar equations for the other bodies:

$$\begin{aligned} m_3 \bar{r}_2 \times \bar{r}_3 \left(\frac{1}{r_{23}^3} - \frac{1}{r_{12}^3} \right) &= 0 \\ m_1 \bar{r}_3 \times \bar{r}_1 \left(\frac{1}{r_{13}^3} - \frac{1}{r_{23}^3} \right) &= 0 \end{aligned} \quad (3.21)$$

The set of relations (3.21) reveals immediately the two possible solutions:

$$\bar{r}_{12} = \bar{r}_{23} = \bar{r}_{13} = \bar{r} \quad (3.22)$$

which gives the *equilateral triangle solution*, and

$$\bar{r}_1 \times \bar{r}_2 = \bar{r}_2 \times \bar{r}_3 = \bar{r}_3 \times \bar{r}_1 = 0 \quad (3.23)$$

which puts the three bodies on a *straight line*. These two cases are the only ones possible. The first case is often called the *Lagrange case*, while the second case is usually called the *Euler case*. These solutions were published by Lagrange and Euler in 1772 and 1767, respectively.

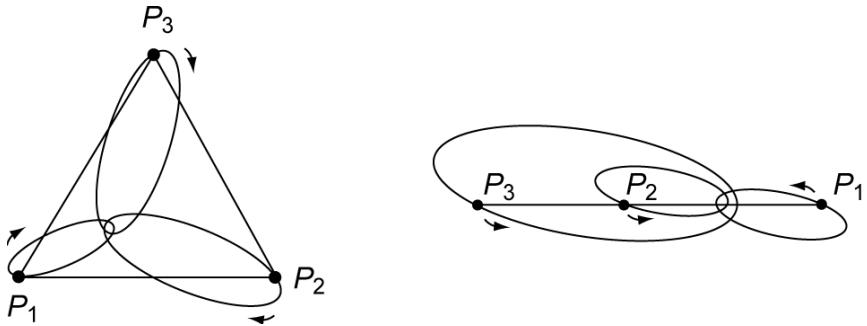


Figure 3.3: Elliptical Lagrange case (left) and elliptical Euler case (right) of central configuration solutions.

In the Lagrange case (Figure 3.3, left), we can write (3.1) for $i = 1$:

$$\frac{d^2 \bar{r}_1}{dt^2} = \frac{G}{r^3} (m_2 \bar{r}_{12} + m_3 \bar{r}_{13})$$

With (3.12) this relation can be written as

$$\frac{d^2 \bar{r}_1}{dt^2} + \frac{GM}{r^3} \bar{r}_1 = 0 \quad (3.24)$$

Because in this case the angle between \bar{r}_{12} and \bar{r}_{13} is 60° , we may write (3.13) as

$$M^2 r_1^2 = m_2^2 r_{12}^2 + m_3^2 r_{13}^2 + m_2 m_3 r_{12} r_{13}$$

or, with (3.22),

$$\frac{1}{r^3} = \frac{(m_2^2 + m_3^2 + m_2 m_3)^{3/2}}{M^3 r_1^3}$$

When we define

$$M^* = \frac{(m_2^2 + m_3^2 + m_2 m_3)^{3/2}}{(m_1 + m_2 + m_3)^2} \quad (3.25)$$

where M^* is constant, we can write (3.24) as

$$\frac{d^2 \bar{r}_1}{dt^2} + \frac{G M^*}{r_1^3} \bar{r}_1 = 0 \quad (3.26)$$

This is the *two-body* equation of motion (Chapter 5), which means that body P_1 moves about the barycenter of the three-body system in a conic section (ellipse, parabola or hyperbola, depending upon the initial velocity) as if a mass M^* was placed in the barycenter of the three-body system. Of course, a corresponding result is obtained for each of the two other bodies. As long as the initial conditions stated before are satisfied, the configuration of the three bodies remains an equilateral triangle, though its size may oscillate or grow indefinitely and its orientation may change. Relative to the *XYZ* frame with origin at the mass center of the system, the three bodies describe three Keplerian orbits with the same eccentricity and the same attraction center (Figure 3.3, left). However, because the value of the ‘artificial’ mass at the attraction center is different for each body, the semi-major axis of the three Keplerian orbits is not the same.

In the collinear (Euler) situation (Figure 3.3, right), where the three bodies are placed in the order $P_1-P_2-P_3$ with the barycenter, O , between P_1 and P_2 , we can write for the gravitational force per unit of mass acting on P_1 :

$$g_1 = G \frac{m_2}{r_{12}^2} + G \frac{m_3}{r_{13}^2}$$

With (3.14) and (3.15) we may write this equation as

$$g_1 = \frac{G}{\{f(t)\}^2} \left[\frac{m_2}{(r_{12})_0^2} + \frac{m_3}{(r_{13})_0^2} \right] = \frac{c_1}{\{f(t)\}^2} = \frac{c_2}{r_1^2} \quad (3.27)$$

where c_1 and c_2 are constants. So, P_1 is acted upon by a central force with a magnitude that is proportional to the inverse-square of the distance of P_1 from the origin of the reference frame. In Chapter 5 it will be shown that for such a force field the trajectory is a conic section, as are the orbits of the other two bodies. So, just as for the Lagrange case, the three bodies describe three Keplerian orbits, with the same eccentricity but with different periods, about the barycenter.

The line connecting the three bodies is rotating with angular velocity ω about the center of mass, such that the collinear configuration is always maintained. This means that we want solutions that satisfy (Figure 3.3, right)

$$\begin{aligned}\omega^2 \mathbf{r}_1 &= G \left(\frac{\mathbf{m}_2}{\mathbf{r}_{12}^2} + \frac{\mathbf{m}_3}{\mathbf{r}_{13}^2} \right) \\ \omega^2 \mathbf{r}_2 &= -G \left(\frac{\mathbf{m}_1}{\mathbf{r}_{12}^2} - \frac{\mathbf{m}_3}{\mathbf{r}_{23}^2} \right) \\ \omega^2 \mathbf{r}_3 &= -G \left(\frac{\mathbf{m}_1}{\mathbf{r}_{13}^2} + \frac{\mathbf{m}_2}{\mathbf{r}_{23}^2} \right)\end{aligned}\quad (3.28)$$

where the value of the constant ω depends upon the initial conditions. As the shape of the configuration does not alter, we may write according to (3.14)

$$\frac{\mathbf{r}_{23}}{\mathbf{r}_{12}} = \frac{(\mathbf{r}_{23})_0}{(\mathbf{r}_{12})_0} = \alpha \quad ; \quad \frac{\mathbf{r}_{13}}{\mathbf{r}_{12}} = \frac{\mathbf{r}_{12} + \mathbf{r}_{23}}{\mathbf{r}_{12}} = 1 + \alpha \quad (3.29)$$

where α is a constant. It is emphasized that this auxiliary parameter α is different from the one introduced in Section 3.1. Subtraction of (3.28-1) from (3.28-2) and (3.28-2) from (3.28-3) gives

$$\begin{aligned}\frac{\omega^2}{G} \mathbf{r}_{12} &= -\frac{\mathbf{m}_1 + \mathbf{m}_2}{\mathbf{r}_{12}^2} + \mathbf{m}_3 \left(\frac{1}{\mathbf{r}_{23}^2} - \frac{1}{\mathbf{r}_{13}^2} \right) \\ \frac{\omega^2}{G} \mathbf{r}_{23} &= -\frac{\mathbf{m}_2 + \mathbf{m}_3}{\mathbf{r}_{23}^2} + \mathbf{m}_1 \left(\frac{1}{\mathbf{r}_{12}^2} - \frac{1}{\mathbf{r}_{13}^2} \right)\end{aligned}\quad (3.30)$$

Substitution of the relations (3.29) into (3.30) gives

$$\begin{aligned}\frac{\omega^2}{G} \mathbf{r}_{12}^3 &= -(\mathbf{m}_1 + \mathbf{m}_2) + \mathbf{m}_3 \left(\frac{1}{\alpha^2} - \frac{1}{(1+\alpha)^2} \right) \\ \frac{\omega^2}{G} \alpha \mathbf{r}_{12}^3 &= -\frac{(\mathbf{m}_2 + \mathbf{m}_3)}{\alpha^2} + \mathbf{m}_1 \left(1 - \frac{1}{(1+\alpha)^2} \right)\end{aligned}$$

From these relations we find

$$-\alpha(\mathbf{m}_1 + \mathbf{m}_2) + \alpha \mathbf{m}_3 \left(\frac{1}{\alpha^2} - \frac{1}{(1+\alpha)^2} \right) = -\frac{(\mathbf{m}_2 + \mathbf{m}_3)}{\alpha^2} + \mathbf{m}_1 \left(1 - \frac{1}{(1+\alpha)^2} \right)$$

or

$$-\alpha^3(1+\alpha)^2(\mathbf{m}_1 + \mathbf{m}_2) + \alpha(1+\alpha)^2 \mathbf{m}_3 - \alpha^3 \mathbf{m}_3 = -(\mathbf{m}_2 + \mathbf{m}_3)(1+\alpha)^2 + \mathbf{m}_1 \alpha^2(1+\alpha)^2 - \mathbf{m}_1 \alpha^2$$

Re-ordering this equation in powers of α results in the quintic equation:

$$\begin{aligned}(\mathbf{m}_1 + \mathbf{m}_2) \alpha^5 + (3\mathbf{m}_1 + 2\mathbf{m}_2) \alpha^4 + (3\mathbf{m}_1 + \mathbf{m}_2) \alpha^3 \\ - (\mathbf{m}_2 + 3\mathbf{m}_3) \alpha^2 - (2\mathbf{m}_2 + 3\mathbf{m}_3) \alpha - (\mathbf{m}_2 + \mathbf{m}_3) = 0\end{aligned}\quad (3.31)$$

Since the coefficients of the powers of α change sign only once, we may conclude from *Descartes' rule of signs* that this equation has only one positive (real) root. Hence, from (3.29)

we conclude that for three given masses there are always three, and only three, collinear central configurations, according to which mass is between the other two.

From (3.31) we conclude that the mutual positions of the three bodies in the collinear configuration are a function of the masses of the three bodies. To indicate the position of the barycenter of the system, O , relative to the three bodies we may use the quantity r_1/r_{12} . Because O is the center of mass, we may write according to Figure 3.3 (right):

$$m_1 r_1 = m_2 r_2 + m_3 r_3$$

or

$$m_1 r_1 = m_2 (r_{12} - r_1) + m_3 (r_{13} - r_1)$$

or

$$(m_1 + m_2 + m_3) r_1 = m_2 r_{12} + m_3 r_{13} \quad (3.32)$$

Substitution of (3.29) into (3.32) leads to

$$\frac{r_1}{r_{12}} = \frac{m_2 + m_3 (1 + \alpha)}{M} \quad (3.33)$$

From (3.31) we find

$$\begin{aligned} m_2 + m_3 (1 + \alpha) &= (m_1 + m_2) \alpha^5 + (3m_1 + 2m_2) \alpha^4 + (3m_1 + m_2) \alpha^3 \\ &\quad - (m_2 + 3m_3) \alpha^2 - (2m_2 + 2m_3) \alpha \end{aligned}$$

or, after re-ordering of the various terms

$$m_2 + m_3 (1 + \alpha) = m_1 (\alpha^5 + 3\alpha^4 + 3\alpha^3) + m_2 (\alpha^5 + 2\alpha^4 + \alpha^3 - \alpha^2 - 2\alpha) - m_3 (3\alpha^2 + 2\alpha)$$

After some algebraic manipulation, we find

$$\{m_2 + m_3 (1 + \alpha)\} \{\alpha^4 + 2\alpha^3 + \alpha^2 + 2\alpha + 1\} = M(\alpha^5 + 3\alpha^4 + 3\alpha^3)$$

Substitution of this relation into (3.33) finally yields

$$\frac{r_1}{r_{12}} = \frac{\alpha^5 + 3\alpha^4 + 3\alpha^3}{\alpha^4 + 2\alpha^3 + \alpha^2 + 2\alpha + 1} \quad (3.34)$$

This equation shows that in a collinear Euler three-body configuration the position of the barycenter is a function of the positions of the three bodies only, and is independent of the masses of the bodies.

3.3. Circular restricted three-body problem

In an effort to gain insight into the characteristics of the three-body problem, Euler, Lagrange, Jacobi, G.W. Hill (1828-1914), Poincaré, and others, have made a large number of studies of the so-called *circular restricted three-body problem*. For this special three-body problem, the following assumptions are made:

- The mass of two bodies is much larger than the mass of the third body. Then, the third body moves in the gravity field of the two massive bodies, but the effect of the gravitational

attraction by the third body on the motion of these massive bodies can be neglected.

– The two massive bodies move in circular orbits about the barycenter of the system.

The orbits of the two massive bodies being known, the problem is to determine the motion of the third body. The general three-body problem is thus reduced from nine second-order differential equations to three second-order ones. This means a reduction from order eighteen to order six. Since the mass of the third body is assumed to be negligible, the two main bodies move as if they form a two-body system. In Section 2.7 it was shown that in such a system both bodies move in a single plane and that the two bodies are always positioned diametrically opposite to each other. It is emphasized that the angular momentum and energy integrals of motion found in Section 2.1 for the n -body problem cannot be used for the analysis of the motion of the third body in the circular restricted three-body problem. The reason is that we have assumed that the two massive bodies move in circular orbits in a single plane, which can only be true if the mass of body P is zero. In that case the angular momentum and total energy of that body is zero, and the integrals of motion hold for the system of the two main bodies. For any mass of P not equal to zero the motion of the two main bodies will not be confined to a single plane anymore. Of course, when the mass of P is small the assumption that the two massive bodies move in a single plane is a good approximation. However, the consequence of this assumption is that the conservation of energy and angular momentum is violated.

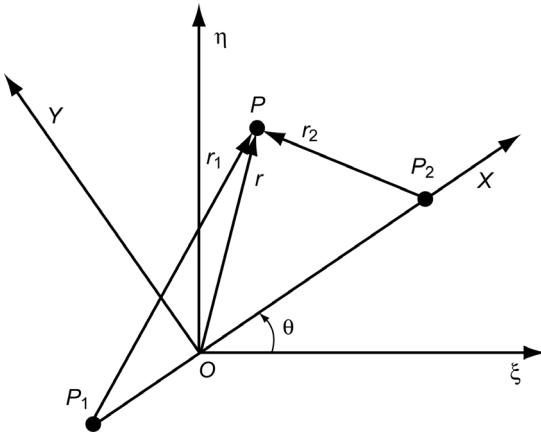


Figure 3.4: Inertial and rotating reference frames in the circular restricted three-body problem.

Now, a (pseudo-) inertial reference frame $\xi\eta\zeta$ is chosen with its origin, O , at the barycenter of the system of three bodies and with the ζ -axis perpendicular to the plane in which the two main bodies are moving (Figure 3.4). Of course, the barycenter is located on the line connecting the two massive bodies. The coordinates of the main bodies P_1 and P_2 are $\xi_1, \eta_1, 0$ and $\xi_2, \eta_2, 0$, respectively; the coordinates of the third body are ξ, η, ζ . The motion of the third body P is not restricted to the $\xi\eta$ -plane and hence the equation of motion with respect to the inertial reference frame is

$$\frac{d^2\bar{r}}{dt^2} = -G \frac{\mathbf{m}_1}{r_1^3} \bar{r}_1 - G \frac{\mathbf{m}_2}{r_2^3} \bar{r}_2 \quad (3.35)$$

where

$$r_1^2 = (\xi - \xi_1)^2 + (\eta - \eta_1)^2 + \zeta^2 \quad ; \quad r_2^2 = (\xi - \xi_2)^2 + (\eta - \eta_2)^2 + \zeta^2 \quad (3.36)$$

Since both massive bodies move in circular orbits about O , we conclude that:

- The distances OP_1 and OP_2 are constant.
- The line segment P_1P_2 rotates about O with a constant angular velocity.

Now, a new reference frame XYZ is chosen, which again has its origin at O and of which the X -axis coincides with P_1P_2 (Figure 3.4). The XY -plane coincides with the $\xi\eta$ -plane. This reference frame rotates about the ζ -axis (Z -axis) with a constant angular velocity ω ($= d\theta/dt$). When the velocity of P with respect to the inertial reference frame is indicated by $d\bar{r}/dt$ and with respect to the rotating reference frame by $\delta\bar{r}/\delta t$, the following expression holds:

$$\frac{d\bar{r}}{dt} = \frac{\delta\bar{r}}{\delta t} + \bar{\omega} \times \bar{r} \quad (3.37)$$

where $\bar{\omega}$ has the magnitude ω and is directed along the Z -axis. This relation between the time derivatives of a vector in both reference frames is generally applicable. So, we may also write

$$\frac{d}{dt} \left(\frac{\delta\bar{r}}{\delta t} \right) = \frac{\delta^2\bar{r}}{\delta t^2} + \bar{\omega} \times \frac{\delta\bar{r}}{\delta t} \quad (3.38)$$

Differentiation of (3.37) gives for the acceleration with respect to the inertial reference frame:

$$\frac{d^2\bar{r}}{dt^2} = \frac{d}{dt} \left(\frac{\delta\bar{r}}{\delta t} \right) + \bar{\omega} \times \frac{d\bar{r}}{dt} \quad (3.39)$$

where we have used the fact that $\bar{\omega}$ is constant. Substitution of (3.37) and (3.38) into (3.39) gives

$$\frac{d^2\bar{r}}{dt^2} = \frac{\delta^2\bar{r}}{\delta t^2} + 2\bar{\omega} \times \frac{\delta\bar{r}}{\delta t} + \bar{\omega} \times (\bar{\omega} \times \bar{r}) \quad (3.40)$$

Substitution of (3.40) into (3.35) yields for the equation of motion of P with respect to the rotating reference frame

$$\frac{\delta^2\bar{r}}{\delta t^2} = -G \left(\frac{m_1}{r_1^3} \bar{r}_1 + \frac{m_2}{r_2^3} \bar{r}_2 \right) - 2\bar{\omega} \times \frac{\delta\bar{r}}{\delta t} - \bar{\omega} \times (\bar{\omega} \times \bar{r}) \quad (3.41)$$

The second and third term on the right-hand side of this equation are the *Coriolis* and *centrifugal acceleration*, respectively. The equation illustrates the fact, which was already mentioned in Section 1.2, that we can apply Newton's second law of motion with respect to a non-inertial reference frame, provided that in addition to the *natural forces* we introduce suitably selected *apparent forces*.

To simplify (3.41), we now introduce new units of mass, length and time. As unit of mass we take $(m_1 + m_2)$. Then, the masses of the main bodies can be expressed as

$$m_1 = 1 - \mu \quad ; \quad m_2 = \mu \quad (3.42-1)$$

We now require that $\mu \leq 1/2$, which means that if the masses of both bodies are not equal, body P_1 has the larger mass. As unit of length, the distance P_1P_2 is selected. Since O is the barycenter of the system, the following expression holds:

$$\frac{OP_1}{OP_2} = \frac{m_2}{m_1} = \frac{\mu}{1 - \mu}$$

or

$$\mu(OP_1 + OP_2) = OP_1$$

Because $OP_1 + OP_2$ has with the new unit of length the value 1, we find

$$OP_1 = \mu ; \quad OP_2 = 1 - \mu \quad (3.42-2)$$

As unit of time we choose $1/\omega$. Using these new non-dimensional units, (3.41) can be written as

$$\begin{aligned} \omega^2(P_1P_2) \frac{\delta^2(\bar{r}/P_1P_2)}{\delta(t^2\omega^2)} &= -G \left[\frac{\frac{m_1}{m_1+m_2}}{\left(\frac{r_1}{P_1P_2}\right)^3} \frac{\bar{r}_1}{P_1P_2} + \frac{\frac{m_2}{m_1+m_2}}{\left(\frac{r_2}{P_1P_2}\right)^3} \frac{\bar{r}_2}{P_1P_2} \right] \frac{(m_1+m_2)}{(P_1P_2)^2} \\ &\quad - 2\omega \bar{e}_z \times \frac{\delta(\bar{r}/P_1P_2)}{\delta(t\omega)} \omega(P_1P_2) - \omega \bar{e}_z \times \left(\omega \bar{e}_z \times \frac{\bar{r}}{P_1P_2} \right) P_1P_2 \end{aligned} \quad (3.43)$$

where \bar{e}_z is the unit vector along the Z-axis. When the quantities expressed in the new units are indicated by $*$, we obtain from (3.43)

$$\frac{\delta^2\bar{r}^*}{\delta t^{*2}} = -\frac{G}{\omega^2} \left[\frac{1-\mu}{r_1^{*3}} \bar{r}_1^* + \frac{\mu}{r_2^{*3}} \bar{r}_2^* \right] - 2\bar{e}_z \times \frac{\delta\bar{r}^*}{\delta t^*} - \bar{e}_z \times (\bar{e}_z \times \bar{r}^*) \quad (3.44)$$

From theoretical mechanics we know that if P_2 (and thus also P_1) moves in a circular orbit about O , then the motion of P_2 is given by

$$m_2 \omega^2(OP_2) = G \frac{m_1 m_2}{(P_1P_2)^2}$$

or, with (3.42),

$$\frac{G}{\omega^2} = \frac{(OP_2)(P_1P_2)^2}{m_1} = \frac{1-\mu}{1-\mu} = 1$$

When, for simplicity, we now omit the index $*$, (3.44) can thus be written in non-dimensional units as

$$\frac{\delta^2\bar{r}}{\delta t^2} = -\left(\frac{1-\mu}{r_1^3} \bar{r}_1 + \frac{\mu}{r_2^3} \bar{r}_2 \right) - 2\bar{e}_z \times \frac{\delta\bar{r}}{\delta t} - \bar{e}_z \times (\bar{e}_z \times \bar{r}) \quad (3.45)$$

Using

$$\bar{r}_1 = (\mu + x) \bar{e}_x + y \bar{e}_y + z \bar{e}_z ; \quad \bar{r}_2 = -(1 - \mu - x) \bar{e}_x + y \bar{e}_y + z \bar{e}_z$$

$$\bar{r} = x \bar{e}_x + y \bar{e}_y + z \bar{e}_z ; \quad \frac{\delta\bar{r}}{\delta t} = \dot{x} \bar{e}_x + \dot{y} \bar{e}_y + \dot{z} \bar{e}_z$$

$$\bar{e}_z \times \frac{\delta\bar{r}}{\delta t} = \dot{x} \bar{e}_y - \dot{y} \bar{e}_x ; \quad \bar{e}_z \times (\bar{e}_z \times \bar{r}) = -x \bar{e}_x - y \bar{e}_y$$

(3.45) can be written as three scalar equations:

$$\begin{aligned}
 \ddot{x} - 2\dot{y} &= x - \frac{1-\mu}{r_1^3}(\mu + x) + \frac{\mu}{r_2^3}(1-\mu-x) \\
 \ddot{y} + 2\dot{x} &= y - \frac{1-\mu}{r_1^3}y - \frac{\mu}{r_2^3}y \\
 \ddot{z} &= -\frac{1-\mu}{r_1^3}z - \frac{\mu}{r_2^3}z
 \end{aligned} \tag{3.46}$$

where the notations \cdot and $\cdot\cdot$ are used to indicate velocities and accelerations, respectively, and

$$r_1^2 = (\mu + x)^2 + y^2 + z^2 ; \quad r_2^2 = (1 - \mu - x)^2 + y^2 + z^2 \tag{3.47}$$

Now, a scalar function, U , of spatial coordinates is introduced:

$$U = \frac{1}{2}(x^2 + y^2) + \frac{1-\mu}{r_1} + \frac{\mu}{r_2} \tag{3.48}$$

Partial differentiation of (3.48) yields

$$\begin{aligned}
 \frac{\partial U}{\partial x} &= x - \frac{1-\mu}{r_1^3}(\mu + x) + \frac{\mu}{r_2^3}(1-\mu-x) \\
 \frac{\partial U}{\partial y} &= y - \frac{1-\mu}{r_1^3}y - \frac{\mu}{r_2^3}y \\
 \frac{\partial U}{\partial z} &= -\frac{1-\mu}{r_1^3}z - \frac{\mu}{r_2^3}z
 \end{aligned} \tag{3.49}$$

Combination of (3.46) and (3.49) gives

$$\begin{aligned}
 \ddot{x} - 2\dot{y} &= \frac{\partial U}{\partial x} \\
 \ddot{y} + 2\dot{x} &= \frac{\partial U}{\partial y} \\
 \ddot{z} &= \frac{\partial U}{\partial z}
 \end{aligned} \tag{3.50}$$

In accordance with the definition of potential functions as given in Section 1.4, we conclude that U is a potential function that accounts for both the gravitational forces and the centrifugal force. The potential function can, of course, not account for the Coriolis force, because this force is a function of velocity components. The force field described by the potential U is clearly *non-central*. Because the bodies P_1 and P_2 have fixed positions with respect to the rotating reference frame, U is not explicitly a function of time, which means that the force field is *conservative*.

3.4. Jacobi's integral

Multiplication of (3.50-1) with \dot{x} , of (3.50-2) with \dot{y} and of (3.50-3) with \dot{z} , and subsequent summation of the resulting equations, gives

$$\dot{x}\ddot{x} + \dot{y}\ddot{y} + \dot{z}\ddot{z} = \dot{x}\frac{\partial U}{\partial x} + \dot{y}\frac{\partial U}{\partial y} + \dot{z}\frac{\partial U}{\partial z} \quad (3.51)$$

Since U is only a function of the spatial coordinates x, y, z and not explicitly of time, we may write

$$\frac{dU}{dt} = \frac{\partial U}{\partial x}\dot{x} + \frac{\partial U}{\partial y}\dot{y} + \frac{\partial U}{\partial z}\dot{z} \quad (3.52)$$

Combination of (3.51) and (3.52) yields, after integration,

$$\dot{x}^2 + \dot{y}^2 + \dot{z}^2 = 2U - C \quad (3.53-1)$$

or

$$V^2 = 2U - C \quad (3.53-2)$$

where the value of the integration constant C is determined by the position and velocity of body P at time $t=0$. In these relations the integration constant is written as $-C$; this is attractive for the analysis in Section 3.6. It is emphasized that V indicates the velocity of P with respect to the rotating reference frame. Equation (3.53) is known as *Jacobi's integral* and was discovered by Jacobi around 1836. In 1899, Poincaré proved that Jacobi's integral is the only algebraic integral of motion that exists in the circular restricted three-body problem. Any other integral would not be an analytical function of the system coordinates, momenta, and time. This total energy integral gives the relation between the velocity and position of the body with negligible mass with respect to a rotating reference frame XYZ , of which the X -axis coincides with the line connecting the two main bodies and of which the XY -plane coincides with the orbital plane of the two main bodies. It shows that if the third body approaches one of the main bodies closely, then it passes that body with high velocity. The constant C is referred to as *Jacobi's constant*, and may, according to (3.48) and (3.53-2), be expressed as

$$C = x^2 + y^2 + \frac{2(1-\mu)}{r_1} + \frac{2\mu}{r_2} - V^2 \quad (3.54)$$

However, as already stated in Section 3.3, we have to be careful in the physical interpretation of this result. The mass of body P may be very small, but is certainly not zero. So, in reality, bodies P_1 and P_2 will not move exactly in circular orbits in a single plane. Therefore, (3.53) can only be considered as a good approximation for the motion of a real body with negligible mass. Numerical calculations performed in the 1960s by various researches have indicated the existence of another integral of motion, besides Jacobi's integral, in the planar restricted three-body problem. That new integral is not analytical and will not be discussed in this book.

Jacobi's integral be rewritten in a form in which position and velocity are expressed relative to the inertial reference frame $\xi\eta\zeta$. To this end, (3.37) is written as

$$\frac{\delta\bar{r}}{\delta t} = \frac{d\bar{r}}{dt} - \bar{\omega} \times \bar{r}$$

Evaluation of this expression and rewriting the result in the dimensionless units introduced in the

previous Section leads to

$$\frac{\delta \bar{r}}{\delta t} = \dot{\xi} \bar{e}_\xi + \dot{\eta} \bar{e}_\eta + \dot{\zeta} \bar{e}_\zeta - (\xi \dot{\eta} - \eta \dot{\xi}) \quad (3.55)$$

where $\bar{e}_\xi, \bar{e}_\eta, \bar{e}_\zeta$ are unit vectors along the ξ -, η - and ζ -axis. Scalar multiplication of $\delta \bar{r}/\delta t$ with $\delta \bar{r}/\delta t$ and using (3.55) results for body P in

$$\left(\frac{\delta r}{\delta t} \right)^2 = (\dot{\xi}^2 + \dot{\eta}^2 + \dot{\zeta}^2) - 2(\xi \dot{\eta} - \eta \dot{\xi}) + (\xi^2 + \eta^2) \quad (3.56)$$

When we realize that $\delta r/\delta t = V$ and $x^2 + y^2 = \xi^2 + \eta^2$, then combination of (3.48), (3.53-2) and (3.56) results in

$$\frac{1}{2}(\dot{\xi}^2 + \dot{\eta}^2 + \dot{\zeta}^2) - \left(\frac{1-\mu}{r_1} + \frac{\mu}{r_2} \right) - (\xi \dot{\eta} - \eta \dot{\xi}) = -\frac{1}{2}C \quad (3.57)$$

where expressions for r_1 and r_2 are given by (3.36). The terms on the left-hand side of (3.57) express the kinetic energy, the potential energy and the angular momentum about the ζ -axis, respectively, all per unit of mass and with respect to the inertial reference frame $\xi\eta\zeta$. So, although neither the total energy nor the angular momentum about the ζ -axis of body P are constant, (3.57) shows that their sum is constant.

3.5. Copenhagen problem

The non-existence of uniform integrals apart from Jacobi's integral, makes it impossible to obtain the totality of solutions of the restricted three-body problem, and therefore already early in the study of the three-body problem attention was directed towards *periodic orbits*. It was hoped that their study would be sufficient for a qualitative description of all possible solutions, while their periodicity made the study of their properties easier. The total number of periodic orbits discovered and studied today is enormous and in this Section only reference is given to an exhaustive study between 1913 and 1939 that was made by S.E. Strömgren (1870-1947) and a group of researchers at Copenhagen. They considered the coplanar circular restricted case, where the three bodies move in a single plane, with $\mu = 1/2$. This group is known as the *Copenhagen school* and their special problem is commonly called the *Copenhagen problem*. Because of the assumptions, the configuration of periodic orbits in the Copenhagen problem is symmetric about the X -axis of the rotating reference frame.

In the Copenhagen problem, there are many families of periodic orbits. Only one, class F in the notation of the Copenhagen school, will be considered in this Section. The analysis of this class will give at least an understanding of what is meant by a 'family', and what is meant by the evolution of orbits within a family. The Class F family of the Copenhagen problem can be described by starting with a small circular orbit around P_2 (Figure 3.5). The body with infinitesimal mass is placed at a certain distance from P_2 on the X -axis, and the velocity is chosen such that the body moves initially in a retrograde circular orbit about P_2 , relative to the rotating reference frame. When the dimension of the orbit increases by starting the motion at greater and greater distances from P_2 , the orbits evolve from oval to kidney-shaped (Figure 3.5^a). For even greater distance the orbits become more and more distorted, until a collision orbit is reached and the body collides with P_1 (Figure 3.5^b). This orbit is, of course, also an ejection orbit and ends the first phase of the development. In the second phase (Figure 3.5^c), orbits develop showing a

loop about P_1 . This loop grows and distorts from orbit to orbit until the second phase ends with a collision at P_2 . A new oval appears, grows, and a new collision occurs; this process is repeated indefinitely. The calculation of Jacobi's constant, C , from orbit to orbit shows that, as expected, it falls in value rapidly at first from its very-high value for the tiny circular orbit about P_2 , reaching a value of 2.04 at the first collision with P_1 , and a value of 1.74 when collision occurs at P_2 .

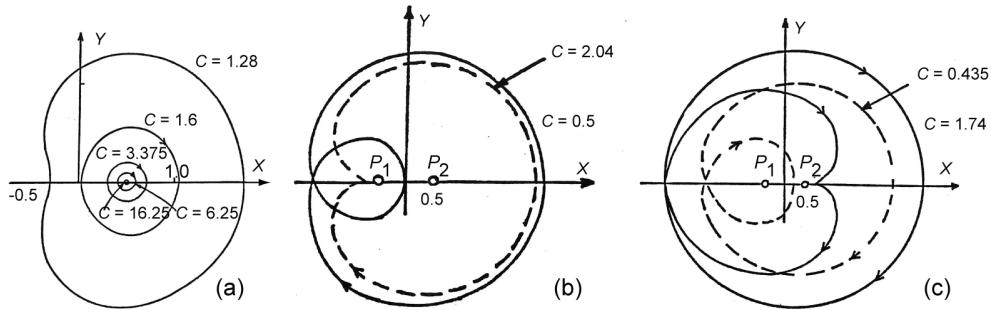


Figure 3.5: Evolution of orbits about P_2 for the Class F family of the Copenhagen problem at decreasing values of C .

The American and Russian lunar space programs inspired extensive numerical and analytical searches for periodic orbits in the Earth-Moon system. These investigations have led to the design of so-called free-return trajectories (Section 3.11) in the US Apollo program. During the 1960's, M. Hénon (1931-) came back to the Copenhagen problem using modern computers and found an unexpected phenomenon: the profusion of chaotic motion, which he called *semi-ergodic motion* in his early papers. In fact, these chaotic motions have already been known since the works of Poincaré, but they were considered merely as a curiosity and not at all as an essential phenomenon. We now know that chaotic motion occur in almost all domains of science (Section 1.3). In fact, it also plays a major role in astrodynamics in the design of minimum-energy trajectories to the Moon and the planets (Sections 3.12, 17.5 and 18.12).

3.6. Surfaces of Hill

We now return to the analysis of Jacobi's integral. A special case occurs when the velocity of the small body P is zero. According to (3.53), then

$$2U = C \quad (3.58)$$

or, with (3.48),

$$x^2 + y^2 + \frac{2(1-\mu)}{r_1} + \frac{2\mu}{r_2} = C \quad (3.59)$$

where expressions for r_1 and r_2 are given by (3.47). This equation describes the *surfaces of Hill*. These are surfaces in XYZ-space on which the velocity of the third body is zero. Inspection of (3.47) and (3.59) shows that, because only the squares of y and z occur, the Surfaces of Hill are symmetric with respect to the XY- and XZ-planes, and, when $\mu = \frac{1}{2}$, with respect to the YZ-plane too. Moreover, the surfaces are contained within a cylinder whose axis is the Z-axis and whose radius is \sqrt{C} , to which certain of the folds are asymptotic at $z^2 = \infty$; for, as z^2 increases, r_1 and r_2 increase and (3.59) approaches as a limit:

$$x^2 + y^2 = C$$

Since for any real body $V^2 \geq 0$, the region in space where the third body can move is given by

$$2U = x^2 + y^2 + \frac{2(1-\mu)}{r_1} + \frac{2\mu}{r_2} \geq C \quad (3.60)$$

So, although we cannot determine the orbit of the third body, with (3.60) we can determine which part of the XYZ -space is accessible to the third body for a given value of C ; i.e. for given initial conditions.

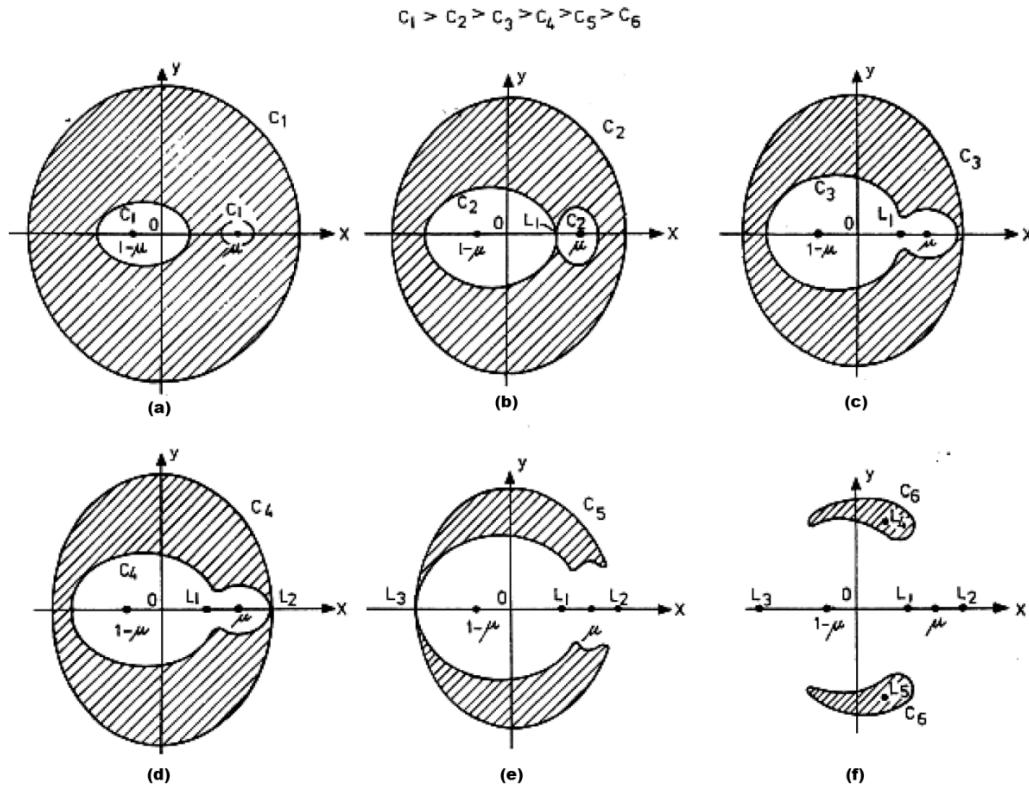


Figure 3.6: Schematic picture of the surfaces of Hill for decreasing values of C ($C_{n+1} < C_n$).

For a qualitative discussion about the shape of the surfaces of Hill as a function of C and a fixed value of μ , we now consider the intersection of the surfaces of Hill with the XY -plane (Figure 3.6). Using (3.59), we find the following expression for this intersection:

$$z = 0 \quad ; \quad r^2 + \frac{2(1-\mu)}{r_1} + \frac{2\mu}{r_2} = C \quad (3.61)$$

where r denotes the radius measured from the origin O . From this relation we conclude that for (very) large values of C the following solutions exist: r (very) large, or r_1 (very) small, or r_2 (very) small. So, for large values of C the intersection of the surfaces of Hill with the XY -plane consist of three (more or less) circles: one with a large radius (\sqrt{C}) around the origin, one with a small radius around P_1 and one with a small radius around P_2 (Figure 3.6^a). In the hatched area of Figure 3.6^a, which is bounded by the three curves, the square of the velocity would be negative. Hence, this part of the XYZ -space is not accessible to the third body. When the body is originally located near P_1 , it will remain forever in the close surroundings of P_1 , because it

cannot cross the hatched area. When the third body is originally located outside the larger circle surrounding O , it can never reach the neighborhood of P_1 or P_2 . Figure 3.6 shows that when the value of C decreases, the smaller ‘circles’ become oval and swell up, and that the ‘outer circle’ becomes smaller, until two ‘inner curves’ touch each other in point L_1 (Figure 3.6^b). A further decrease of the value of C causes the two curves around P_1 and P_2 to merge. Only then, it is possible for a body that is originally located near P_1 to reach the neighborhood of P_2 . The question whether and when this will happen, and which orbit the third body will follow, cannot be answered. A further decrease of the value of C will cause the ‘inner curve’ to touch the ‘outer curve’ in point L_2 (Figure 3.6^d). When the value of C is further decreased, then a body that is originally located near P_1 can not only reach the neighborhood of P_2 , but can also move away from the two main bodies to an unbounded distance. So, the body can escape from the gravity fields of P_1 and P_2 . With a further decreasing value of C , a new point L_3 shows up (Figure 3.6^c), after which for even smaller values of C the hatched area shrinks further, until it finally ‘dissolves’ into L_4 and L_5 . Hence, for all values of $C \leq C_{L_{4,5}}$ the entire XYZ -space is accessible for a body that is originally located near P_1 or P_2 .

The fact that for a certain value of C a small body that is originally located near P_1 can never reach an infinite distance from P_1 , was first discovered by Hill in 1878, hence the name *surfaces of Hill*. Hill considered the motion of the Moon in the Sun-Earth-Moon three-body system. He discovered that when the eccentricity of the Earth’s orbit and the mass of the Moon with respect to the mass of the Earth are neglected, the constant C of the motion of the Moon has such a value that the surface on which the velocity of the Moon would be zero, is closed. This surface is located at a distance of 100-110 Earth radii from the center of the Earth. The Moon, which is located at a distance of about 60 Earth radii, moves within this surface. This, Hill argued, was proof that the Moon has always been close to the Earth. However, it may be shown that his conclusion does not have to be valid when the applied simplifications are abandoned. In 1952, Y. Higihara (1897-1979) proved that all natural satellites in the solar system, with the exception of four moons of Jupiter (VIII, IX, XI, XII, which are in retrograde orbits), move in orbits within closed surfaces of Hill. This means that their orbits are stable in the sense of Hill.

The above discussion was based on the value of the integration constant C . To clarify the physical meaning of the ‘forbidden region’ we now assume that body P is close to body P_1 . If the position of P is known, we may compute with (3.48) the value of U . For a given value of U , (3.53) shows that a large value of C means a small value of the velocity V . So, for a small value of V the motion of the third body is restricted to a bounded region about P_1 . If, at the same position, the velocity of P would increase (smaller value of C) the ‘forbidden region’ shrinks until, for a sufficiently large value of V , the entire space is accessible to body P . The same kind of conclusion can be drawn if body P is at a (very) large distance from P_1 (and P_2).

3.7. Lagrange libration points

In Figure 3.7, the intersections of the surfaces of Hill with the XY -plane, XZ -plane and YZ -plane, respectively, are sketched for $\mu = 0.27$. All curves drawn in this Figure satisfy the relation $2U = C$, where the curves closer to P_1 and P_2 correspond to higher values of C and U . Figure 3.7^a shows that if $y = z = 0$, the equation for the surfaces of Hill has six real roots on the X -axis for $C = C_1$. For an appropriate value of C within the range $C_3 < C < C_2$, for which the two inner curves touch in L_1 , two of these x -roots coincide. Inspection of Figure 3.7^a reveals that therefore at point L_1 : $\partial U / \partial x = 0$. For $z = 0$, $x = x_{L_1}$, Figure 3.7^a shows that there are even four coinciding y -roots on the line $x = x_{L_1}$ for this value of C . This means that at point L_1 at least $\partial U / \partial y = 0$. Figure 3.7^b shows that for $x = x_{L_1}$, $y = 0$, there are also four coinciding z -roots. That means that

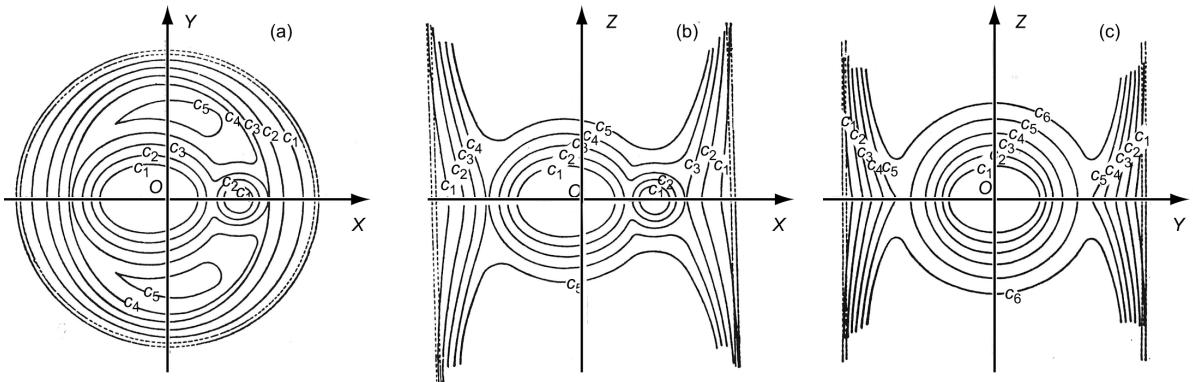


Figure 3.7: Cross sections of the surfaces of Hill with the XY-, XZ- and YZ-plane ($\mu = 0.27$).

at point L_1 at least $\partial U / \partial z = 0$. So, at point L_1 the following condition holds:

$$\frac{\partial U}{\partial x} = \frac{\partial U}{\partial y} = \frac{\partial U}{\partial z} = 0 \quad (3.62)$$

A further detailed analysis of the geometry of the surfaces of Hill shows that expression (3.62) holds for all L points.

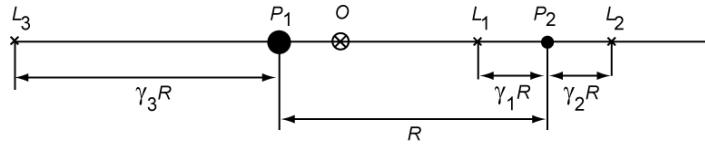
Substitution of (3.62) into (3.49) gives

$$\begin{aligned} x - \frac{1-\mu}{r_1^3}(\mu+x) + \frac{\mu}{r_2^3}(1-\mu-x) &= 0 \\ y \left(1 - \frac{1-\mu}{r_1^3} - \frac{\mu}{r_2^3} \right) &= 0 \\ z \left(\frac{1-\mu}{r_1^3} + \frac{\mu}{r_2^3} \right) &= 0 \end{aligned} \quad (3.63)$$

Because r_1 and r_2 are positive and $0 < \mu \leq \frac{1}{2}$, (3.63-3) yields $z = 0$. This implies that the five points L_1 to L_5 are all located in the XY-plane. Combination of (3.47), (3.63-1) and (3.63-2) yields as a first solution:

$$\begin{aligned} y &= 0 \\ x - (1-\mu) \frac{\mu+x}{|\mu+x|^3} + \mu \frac{1-\mu-x}{|1-\mu-x|^3} &= 0 \end{aligned} \quad (3.64)$$

This equation cannot be solved in a closed analytical way. However, it can be shown that the fifth-degree equation (3.64-2) has three real roots, corresponding to the x -coordinates of the points L_1 , L_2 and L_3 . Thus, these points are located on the X -axis. Point L_2 is located to the right of P_2 ; point L_3 is located to the left of P_1 . For point L_1 : $r_2 \leq r_1$, whereby L_1 is located closer to P_2 when μ is smaller. When we apply the notation shown in Figure 3.8, an extensive analytical elaboration of (3.64) yields the following series expansions for the dimensionless distances between the points L_1 , L_2 , L_3 , and the main bodies:

Figure 3.8: Definition of the non-dimensional distances γ_1 , γ_2 , and γ_3 .

$$\alpha = \frac{\mu}{(1-\mu)} ; \quad \beta = \left(\frac{1}{3}\alpha\right)^{1/3}$$

$$\gamma_1 = \beta - \frac{1}{3}\beta^2 - \frac{1}{9}\beta^3 - \frac{23}{81}\beta^4 + O(\beta^5)$$

$$\gamma_2 = \beta + \frac{1}{3}\beta^2 - \frac{1}{9}\beta^3 - \frac{31}{81}\beta^4 + O(\beta^5)$$

$$\gamma_3 = 1 - \frac{7}{12}\alpha + \frac{7}{12}\alpha^2 - \frac{13223}{20736}\alpha^3 + O(\alpha^4)$$

The second solution of (3.63) can be found by solving the two equations:

$$\begin{aligned} x - \frac{1-\mu}{r_1^3}(\mu+x) + \frac{\mu}{r_2^3}(1-\mu-x) &= 0 \\ 1 - \frac{1-\mu}{r_1^3} - \frac{\mu}{r_2^3} &= 0 \end{aligned} \tag{3.65}$$

Multiplication of (3.65-2) by $-(\mu+x)$ and adding this result to (3.65-1) yields $r_2 = 1$. Multiplication of (3.65-2) by $(1-\mu-x)$ and adding this result to (3.65-1) yields $r_1 = 1$. So, the second solution of (3.63) is

$$r_1 = r_2 = 1 \tag{3.66}$$

This solution corresponds to the points L_4 and L_5 . Thus, these points form an equilateral triangle with the two main bodies and the coordinates of the points L_4 and L_5 are

$$x = \frac{1}{2} - \mu ; \quad y = \pm \frac{1}{2}\sqrt{3} \tag{3.67}$$

Substitution of (3.66) and (3.67) into (3.48) gives

$$U_{L_4, L_5} = \frac{1}{2}(\mu^2 - \mu + 3) \tag{3.68}$$

The minimum value of C for which the surfaces of Hill exist, and thus for which the space in which the third body can move is bounded, can be found from (3.58) and (3.68):

$$C_{min} = \mu^2 - \mu + 3 = 2.75 + (\mu - \frac{1}{2})^2$$

Since $0 < \mu \leq \frac{1}{2}$, we find for the value of C_{min} :

$$2.75 \leq C_{min} < 3 \tag{3.69}$$

In Table 3.1 the positions of the points L_1 to L_5 are presented for some three-body systems with the main bodies indicated in the first column. The data on the mass of the celestial bodies were taken from Appendix B. Note that the points L_1 and L_2 are located relatively close to the main

body with the smaller mass (P_2). The distance between point L_3 and the main body with the larger mass (P_1) is almost equal to the distance between the two main bodies.

Table 3.1: Position of the points L_1 , L_2 , L_3 for various three-body systems.

| System | μ | γ_1 | γ_2 | γ_3 |
|------------------|------------------------|------------------------|------------------------|------------|
| Sun-Venus | 2.448×10^{-6} | 9.315×10^{-3} | 9.373×10^{-3} | 1.00000 |
| Sun-Earth+Moon | 3.040×10^{-6} | 1.001×10^{-2} | 1.008×10^{-2} | 1.00000 |
| Sun-Mars | 3.227×10^{-7} | 4.748×10^{-3} | 4.763×10^{-3} | 1.00000 |
| Sun-Jupiter | 9.537×10^{-4} | 6.668×10^{-2} | 6.978×10^{-2} | 0.99944 |
| Earth-Moon | 1.215×10^{-2} | 1.509×10^{-1} | 1.679×10^{-1} | 0.99291 |
| Jupiter-Ganymede | 7.804×10^{-5} | 2.934×10^{-2} | 2.992×10^{-2} | 0.99995 |

Figure 3.9 shows the location of all Earth-Moon L points and of two Sun-Earth L points to scale. In this projection, the Moon and the Earth-Moon L points, of course, rotate relative to the line connecting the Sun-Earth L_1 and L_2 points.

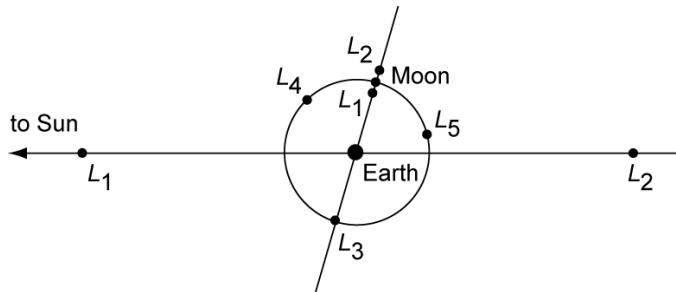


Figure 3.9: Positions of all Earth-Moon L points and of two Sun-Earth L points (to scale).

In the literature, point L_1 of the Earth-Moon system is sometimes confused with the *neutral point*, N , of this system. This neutral point is defined as the point on the line between Earth and Moon where the gravitational forces generated by the Earth and the Moon cancel. With the notation used in Table 3.1, the location of point N follows from

$$\frac{1 - \mu}{R^2 (1 - \gamma_N)^2} = \frac{\mu}{R^2 \gamma_N^2}$$

where γ_N is measured from body P_2 . For $\mu = 0.01215$, this equation has the solution $\gamma_N = 0.09983$. Comparing this value with the value of γ_1 listed in Table 3.1 for the Earth-Moon system, we conclude that point N is closer to the Moon than point L_1 . The reason, of course, is that in the computation of the location of point N , we did not account for the rotation of the Earth-Moon system, which will lead to a centrifugal force exerted on a body positioned in N .

Substitution of (3.62), which formulates the conditions in the L points, into the equations of motion (3.50) leads to

$$\begin{aligned} \ddot{x} - 2\dot{y} &= 0 \\ \ddot{y} + 2\dot{x} &= 0 \\ \ddot{z} &= 0 \end{aligned} \tag{3.70}$$

When a body with zero velocity is located at an L point, then, according to (3.70),

$$\ddot{x} = \ddot{y} = \ddot{z} = 0 \quad (3.71)$$

So, the body does not experience an acceleration with respect to the rotating reference frame; in other words: these points are *equilibrium points* and are called *Lagrange libration points* ('libration' means 'oscillation'). The three collinear libration points were discovered by Euler around 1750; the two equilateral libration points by Lagrange around 1760. However, it is common in the literature to refer to all points as the 'Lagrange libration points'.

3.8. Motion after leaving a surface of Hill

By definition, a small body located on a surface of Hill has a zero velocity. In general, the acceleration of the body will not be zero, unless it is at one of the five libration points. An interesting question that will be addressed in this Section is: "in what direction will the body leave the surface of zero velocity?"

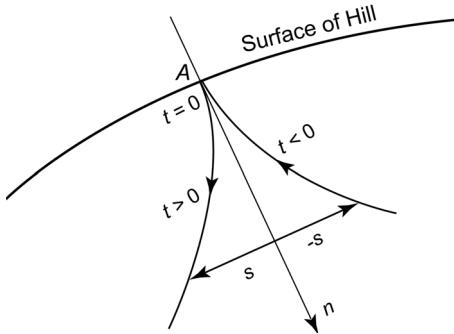


Figure 3.10: Cusp at a curve of zero velocity.

We assume that at $t = 0$ the small body is at point A on a surface of Hill (Figure 3.10), while A does not coincide with a libration point. The initial conditions of the body are specified as $x_0, y_0, z_0, \dot{x}_0 = \dot{y}_0 = \dot{z}_0 = 0$. When we apply a series expansion of the coordinates of the body around the starting point A , we may write

$$\begin{aligned} x &= x_0 + \dot{x}_0 t + \frac{1}{2} \ddot{x}_0 t^2 + \frac{1}{6} \ddot{\ddot{x}}_0 t^3 + \dots \\ y &= y_0 + \dot{y}_0 t + \frac{1}{2} \ddot{y}_0 t^2 + \frac{1}{6} \ddot{\ddot{y}}_0 t^3 + \dots \\ z &= z_0 + \dot{z}_0 t + \frac{1}{2} \ddot{z}_0 t^2 + \frac{1}{6} \ddot{\ddot{z}}_0 t^3 + \dots \end{aligned} \quad (3.72)$$

We know that $\dot{x}_0 = \dot{y}_0 = \dot{z}_0 = 0$, so the first coefficients for which we have to find suitable expressions are $\ddot{x}_0, \ddot{y}_0, \ddot{z}_0$. These expressions can be found from (3.50). When we introduce the notation

$$U_x = \left(\frac{\partial U}{\partial x} \right)_0, \quad \text{etc.} \quad (3.73-1)$$

we find

$$\ddot{x}_0 = U_x \quad ; \quad \ddot{y}_0 = U_y \quad ; \quad \ddot{z}_0 = U_z \quad (3.74)$$

The next coefficients in the series expansion can be obtained by differentiation of (3.50), using (3.52):

$$\begin{aligned}\ddot{x} &= 2\ddot{y} + U_{xx}\dot{x} + U_{xy}\dot{y} + U_{xz}\dot{z} \\ \ddot{y} &= -2\ddot{x} + U_{xy}\dot{x} + U_{yy}\dot{y} + U_{yz}\dot{z} \\ \ddot{z} &= U_{xz}\dot{x} + U_{yz}\dot{y} + U_{zz}\dot{z}\end{aligned}\quad (3.75)$$

where the notation

$$U_{xy} = \left(\frac{\partial^2 U}{\partial x \partial y} \right)_0, \quad \text{etc.} \quad (3.73-2)$$

has been used. Since $\dot{x}_0 = \dot{y}_0 = \dot{z}_0 = 0$, we find from (3.75)

$$\ddot{x}_0 = 2\ddot{y}_0 \quad ; \quad \ddot{y}_0 = -2\ddot{x}_0 \quad ; \quad \ddot{z}_0 = 0 \quad (3.76)$$

Substitution of (3.50) into (3.76) gives

$$\ddot{x}_0 = 2U_y \quad ; \quad \ddot{y}_0 = -2U_x \quad ; \quad \ddot{z}_0 = 0 \quad (3.77)$$

Substitution of (3.74) and (3.77) into (3.72) yields

$$\begin{aligned}x &= x_0 + \frac{1}{2}U_x t^2 + \frac{1}{3}U_y t^3 + O(t^4) \\ y &= y_0 + \frac{1}{2}U_y t^2 - \frac{1}{3}U_x t^3 + O(t^4) \\ z &= z_0 + \frac{1}{2}U_z t^2 + O(t^4)\end{aligned}\quad (3.78)$$

When we introduce the position vector

$$\bar{r} = x \bar{e}_x + y \bar{e}_y + z \bar{e}_z$$

where \bar{e}_x , \bar{e}_y and \bar{e}_z are unit vectors along the X-, Y- and Z-axis, we can write (3.78) as

$$\bar{r} = \bar{r}_0 + \frac{1}{2}(U_x \bar{e}_x + U_y \bar{e}_y + U_z \bar{e}_z)t^2 + \frac{1}{3}(U_y \bar{e}_x - U_x \bar{e}_y)t^3 + O(t^4) \quad (3.79)$$

We know that

$$U_x \bar{e}_x + U_y \bar{e}_y + U_z \bar{e}_z = \bar{\nabla}U = \bar{n} \quad ; \quad U_y \bar{e}_x - U_x \bar{e}_y = \bar{s}$$

where \bar{n} is a vector normal to the surface of Hill at A , and \bar{s} is a vector tangential to the surface of Hill at A and parallel to the XY -plane. After truncation of the series expressions (3.79) at terms of the order t^3 , we therefore can write

$$\begin{aligned}\bar{r} &= \bar{r}_0 + \frac{1}{2}\bar{n}t^2 + \frac{1}{3}\bar{s}t^3 \\ \bar{V} &= \bar{n}t + \bar{s}t^2\end{aligned}\quad (3.80)$$

These relations show that the motion starts along the normal to the surface of zero velocity and that the transverse component becomes only significant as t increases. In our analysis, $t < 0$ before the body is at point A . From (3.80-1) we note that the second term on the right-hand side

has an equal value for $t = t_1$ and for $t = -t_1$, while the third term on the right-hand side has opposite signs for both cases. This means that if at $t = 0$ the small mass is on the surface of Hill at A , then the transverse position component changes sign before and after $t = 0$. Consequently, the trajectory forms a cusp at x_0, y_0, z_0 , as shown in Figure 3.10.

3.9. Stability in the libration points

One may wonder whether the equilibrium in the libration points is stable or not. Here, we define the motion as stable if a small body positioned at a libration point and with zero velocity does not move away from that libration point unboundedly when a small perturbing force acts upon that body. To analyze the motion, we start from (3.50) and investigate the behavior of $\partial U/\partial x$, $\partial U/\partial y$, and $\partial U/\partial z$ in the neighborhood of a libration point. Therefore, we write with a Taylor series expansion for $\partial U/\partial x$:

$$\begin{aligned} \frac{\partial U}{\partial x} &= \left(\frac{\partial U}{\partial x} \right)_0 + (x - x_0) \left(\frac{\partial^2 U}{\partial x^2} \right)_0 + (y - y_0) \left(\frac{\partial^2 U}{\partial y \partial x} \right)_0 + (z - z_0) \left(\frac{\partial^2 U}{\partial z \partial x} \right)_0 \\ &\quad + O[(x - x_0)^2, (x - x_0)(y - y_0), \text{etc.}] \end{aligned} \quad (3.81)$$

where the index 0 refers to the conditions at a libration point. According to (3.62), the first term on the right-hand side of (3.81) is zero for all L points. If only small deviations are considered, higher-order terms are neglected, and the notation

$$x' = x - x_0 \quad ; \quad y' = y - y_0 \quad ; \quad z' = z - z_0 \quad (3.82)$$

is applied, (3.81) can be written as

$$\frac{\partial U}{\partial x} = x' U_{xx} + y' U_{xy} + z' U_{xz} \quad (3.83-1)$$

where the notation (3.73-2) has been used. Since U_0 is continuous:

$$U_{xy} = U_{yx} \quad ; \quad \text{etc.}$$

Similarly, we may derive the following relations:

$$\begin{aligned} \frac{\partial U}{\partial y} &= x' U_{xy} + y' U_{yy} + z' U_{yz} \\ \frac{\partial U}{\partial z} &= x' U_{xz} + y' U_{yz} + z' U_{zz} \end{aligned} \quad (3.83-2)$$

Substitution of (3.83) into (3.50) leads, with (3.82), for the motion of a body in the neighborhood of a libration point to

$$\begin{aligned} \ddot{x}' - 2\dot{y}' - x' U_{xx} - y' U_{xy} - z' U_{xz} &= 0 \\ \ddot{y}' + 2\dot{x}' - x' U_{xy} - y' U_{yy} - z' U_{yz} &= 0 \\ \ddot{z}' - x' U_{xz} - y' U_{yz} - z' U_{zz} &= 0 \end{aligned} \quad (3.84)$$

This is a set of simultaneous linear differential equations with constant coefficients. For each of the libration points, the values of the coefficients $U_{xx}, U_{yy}, U_{zz}, U_{xy}, U_{xz}, U_{yz}$ can be found by partial

differentiation of the expressions (3.49) and by subsequent substitution of the coordinates of the particular libration point. The system of equations (3.84) can be simplified by using the fact that all libration points are located in the XY -plane; so, $z = 0$. Equation (3.49-3) can be written as

$$\frac{\partial U}{\partial z} = -z \left(\frac{1-\mu}{r_1^3} + \frac{\mu}{r_2^3} \right)$$

From this equation, we find for a libration point ($z = 0$)

$$U_{xz} = U_{yz} = 0 \quad ; \quad U_{zz} = -\left(\frac{1-\mu}{r_1^3} + \frac{\mu}{r_2^3} \right) < 0 \quad (3.85)$$

Substitution of these expressions into (3.84) gives

$$\begin{aligned} \ddot{x}' - 2\dot{y}' - x' U_{xx} - y' U_{xy} &= 0 \\ \ddot{y}' + 2\dot{x}' - x' U_{xy} - y' U_{yy} &= 0 \\ \ddot{z}' - z' U_{zz} &= 0 \end{aligned} \quad (3.86)$$

Note that the motion in the Z -direction is completely decoupled from the motion in the X - and Y -direction. Because $U_{zz} < 0$, the solution of (3.86-3) is

$$z' = C_1 \cos(\sqrt{|U_{zz}|} t) + C_2 \sin(\sqrt{|U_{zz}|} t) \quad (3.87)$$

Hence, for all libration points the motion in the Z -direction is purely periodic, i.e. an undamped non-diverging oscillation, which, according to our definition, is considered as stable. The period of this motion is independent of the characteristics of the motion in the X - and Y -direction.

The solution of the first two simultaneous differential equations with constant coefficients ((3.86-1) and (3.86-2)) can be found by applying a classical technique. First, the expressions

$$x' = A e^{\lambda t} \quad ; \quad y' = B e^{\lambda t} \quad (3.88)$$

are substituted into these equations. We then obtain

$$\begin{aligned} (\lambda^2 - U_{xx})A - (2\lambda + U_{xy})B &= 0 \\ (2\lambda - U_{xy})A + (\lambda^2 - U_{yy})B &= 0 \end{aligned} \quad (3.89)$$

This set of homogeneous linear equations in A and B can only be solved for arbitrary values of A and B , if the coefficient-determinant is zero:

$$(\lambda^2 - U_{xx})(\lambda^2 - U_{yy}) + (2\lambda + U_{xy})(2\lambda - U_{xy}) = 0$$

or

$$\lambda^4 + (4 - U_{xx} - U_{yy})\lambda^2 + U_{xx}U_{yy} - U_{xy}^2 = 0 \quad (3.90)$$

The solution of (3.90) will consist of four values of λ that will, generally, have complex values. However, the equation is quadratic in λ^2 and thus the values of λ will consist of conjugate pairs. For example, $\lambda_2 = -\lambda_1$, $\lambda_4 = -\lambda_3$. If the four roots are all different, the motion of a small body with respect to a libration point can be written as

$$\begin{aligned}x' &= A_1 e^{\lambda_1 t} + A_2 e^{-\lambda_1 t} + A_3 e^{\lambda_3 t} + A_4 e^{-\lambda_3 t} \\y' &= B_1 e^{\lambda_1 t} + B_2 e^{-\lambda_1 t} + B_3 e^{\lambda_3 t} + B_4 e^{-\lambda_3 t}\end{aligned}\quad (3.91)$$

If λ_2 and λ_4 , and hence λ_1 and λ_3 , are equal, the solution is

$$\begin{aligned}x' &= A_1 e^{\lambda_1 t} + A_2 e^{-\lambda_1 t} + A_3 t e^{\lambda_1 t} + A_4 t e^{-\lambda_1 t} \\y' &= B_1 e^{\lambda_1 t} + B_2 e^{-\lambda_1 t} + B_3 t e^{\lambda_1 t} + B_4 t e^{-\lambda_1 t}\end{aligned}\quad (3.92)$$

From (3.91) and (3.92) we conclude that the motion described by (3.86) is, according to our definition, stable if all λ 's are different and if their real part is smaller than or equal to zero. The last requirement can be explained as follows. In general, we may write $\lambda = \alpha + i\beta$, where α and β are real values. So, $e^{\lambda t} = e^{\alpha t} e^{i\beta t}$. The term $e^{i\beta t}$ is associated with sine- and cosine-functions. We therefore conclude that if $\alpha = 0$ the term $e^{\lambda t}$ will lead to pure sinusoidal oscillations, if $\alpha < 0$ to damped oscillations, and if $\alpha > 0$ to diverging oscillations. So, if $\alpha \leq 0$, x' and y' cannot increase arbitrarily. Since the λ 's occur in conjugate pairs, the requirement for stability reduces to λ_i 's different and purely imaginary, which results in λ_i^2 being real and negative. In that case, the motion described by (3.86) is a pure oscillation.

First, the equilibrium of the point L_1 , L_2 and L_3 will be investigated. For these points, we write with (3.47)

$$y = z = 0 \quad ; \quad r_1^2 = (\mu + x)^2 \quad ; \quad r_2^2 = (1 - \mu - x)^2$$

Differentiation of (3.49) yields, after substitution of these relations,

$$U_{xx} = 1 + 2K \quad ; \quad U_{xy} = 0 \quad ; \quad U_{yy} = 1 - K \quad (3.93)$$

where

$$K = \frac{1 - \mu}{r_1^3} + \frac{\mu}{r_2^3} \quad (3.94)$$

Substitution of (3.93) into (3.90) leads to

$$\lambda^4 + (2 - K)\lambda^2 + (1 + 2K)(1 - K) = 0$$

We have found that for a stable equilibrium this equation should have two real negative roots for λ^2 . Then, the product of these roots is positive, which, since $K > 0$, requires

$$1 - K > 0 \quad (3.95)$$

The location of the points L_1 , L_2 , L_3 is described by (3.63-1), which can also be written as

$$x \left(1 - \frac{1 - \mu}{r_1^3} - \frac{\mu}{r_2^3} \right) - \frac{\mu(1 - \mu)}{r_1^3} + \frac{\mu(1 - \mu)}{r_2^3} = 0$$

or, with (3.94),

$$1 - K = \frac{\mu(1 - \mu)}{x} \left(\frac{1}{r_1^3} - \frac{1}{r_2^3} \right) \quad (3.96)$$

Inspection of Figure 3.4 and Figure 3.8 reveals that (3.96) yields for the points L_1, L_2, L_3 :

$$1 - K < 0 \quad (3.97)$$

From (3.95) and (3.97) we conclude that the equilibrium at the points L_1, L_2 en L_3 is *unstable*.

For the libration points L_4 and L_5 we write with (3.66) and (3.67)

$$r_1 = r_2 = 1 \quad ; \quad x = \frac{1}{2} - \mu \quad ; \quad y = \pm \frac{1}{2}\sqrt{3} \quad ; \quad z = 0 \quad ; \quad K = 1$$

Differentiation of (3.49) yields, using these relations and (3.47),

$$U_{xx} = \frac{3}{4} \quad ; \quad U_{xy} = \pm \frac{3}{4}\sqrt{3}(1 - 2\mu) \quad ; \quad U_{yy} = \frac{9}{4} \quad (3.98)$$

Substitution of these values into (3.90) results in

$$\lambda^4 + \lambda^2 + \frac{27}{4}\mu(1 - \mu) = 0$$

The two roots λ^2 of this equation are given by

$$\lambda_{1,2}^2 = \frac{-1 \pm \sqrt{1 - 27\mu(1 - \mu)}}{2}$$

For a stable equilibrium we must require that these two roots are different and real, and have a negative value. This leads to the condition $27\mu(1-\mu) < 1$. Since $0 < \mu \leq \frac{1}{2}$: $27\mu(1-\mu) > 0$, which means that if the roots satisfy the conditions of being real and different, they are negative. So, the problem can be reduced to finding the values of μ for which

$$\mathcal{F} = 27\mu^2 - 27\mu + 1 > 0 \quad (3.99)$$

For the roots of \mathcal{F} we find

$$\mu_{1,2} = \frac{27 \pm \sqrt{27^2 - 4 \cdot 27}}{54} = \frac{1}{2} \pm \sqrt{\frac{23}{108}}$$

Substitution of $\mu = \frac{1}{2}$ into (3.99) yields $\mathcal{F} < 0$ and substitution of $\mu = 0$ yields $\mathcal{F} > 0$. Since always $\mu \leq \frac{1}{2}$, the requirement for a *stable* equilibrium in the libration points L_4, L_5 becomes

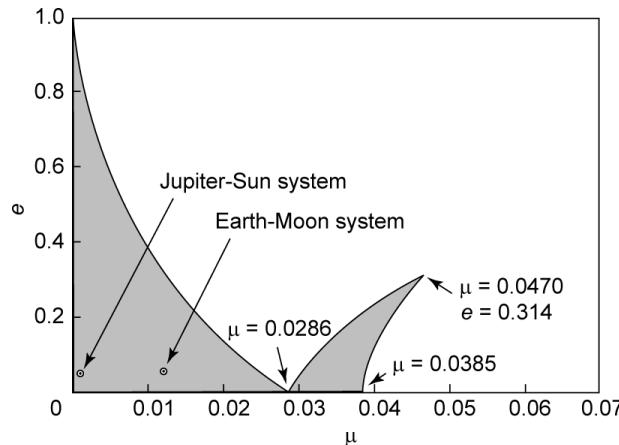


Figure 3.11: Regions of first-order stability in the L_4 and L_5 points for the case that the two main bodies move in elliptical orbits.

$$\mu < \frac{1}{2} - \sqrt{\frac{23}{108}} \approx 0.03852 \quad (3.100)$$

So, if $\mu < 0.03852$, then a small body located at the libration points L_4 or L_5 will perform, after a small perturbation, a pure oscillation about these points, which represents, according to our definition, stable motion. Of course, this motion is *infinitesimally stable*, since we are dealing with a linearized theory. For $\mu > 0.03852$, the motion of the body will diverge from the libration points L_4 or L_5 . We have to realize that in reality other forces may be present that can drive the body from the L_4 or L_5 point. In addition, it is emphasized that our stability analysis only holds for the case when the orbits of P_1 and P_2 are circular. In case the orbits of P_1 and P_2 are elliptical, the equilibrium at the points L_4 and L_5 may also be stable, although then the stability is not just a function of μ , but also of the eccentricity of the elliptical orbits. In Figure 3.11, the stability area (hatched area) is presented for the L_4 and L_5 points for elliptical orbits of the main bodies.

3.10. Motion about the libration points

Equations (3.86) describe the linearized motion of a small body about a libration point of the circular restricted three-body problem. In this Section, we will analyze some characteristics of that motion.

We already have found in Section 3.9 that for all five libration points the motion in the Z -direction is an undamped oscillation and is uncoupled from the motion in the X - and Y -direction. From (3.87) we find that the period of the motion in the Z -direction is given by

$$T_z = \frac{2\pi}{\sqrt{|U_{zz}|}}$$

or, with (3.85) and (3.94),

$$T_z = \frac{2\pi}{\sqrt{K}} \quad (3.101)$$

In Section 3.9 we have found:

$$L_1, L_2, L_3 : K > 1 \quad ; \quad L_4, L_5 : K = 1 \quad (3.102)$$

For a given three-body problem and a selected libration point, the value of K is known. In the following, the motion in the X - and Y -direction will be analyzed.

Collinear libration points

Combination of (3.93) and (3.102) gives for the three collinear libration points

$$K > 1 \quad ; \quad U_{xx} = 1 + 2K > 3 \quad ; \quad U_{xy} = 0 \quad ; \quad U_{yy} = 1 - K < 0 \quad (3.103)$$

With (3.103), the equations of motion (3.86) can be written as

$$\begin{aligned} \ddot{x}' - 2\dot{y}' - (2K + 1)x' &= 0 \\ \ddot{y}' + 2\dot{x}' + (K - 1)y' &= 0 \end{aligned} \quad (3.104)$$

and the characteristic equation (3.90) as

$$\lambda^4 + (2 - K)\lambda^2 - (2K + 1)(K - 1) = 0 \quad (3.105)$$

Defining

$$\alpha = 1 - \frac{1}{2}K < \frac{1}{2} \quad ; \quad \beta^2 = (2K+1)(K-1) > 0 \quad ; \quad \lambda^2 = \Lambda \quad (3.106)$$

we can write (3.105) as

$$\Lambda^2 + 2\alpha\Lambda - \beta^2 = 0$$

The roots of this equation are

$$\Lambda_1 = -\alpha + \sqrt{\alpha^2 + \beta^2} > 0 \quad ; \quad \Lambda_2 = -\alpha - \sqrt{\alpha^2 + \beta^2} < 0 \quad (3.107)$$

So, both roots are real and of opposite sign. When we write

$$\lambda_{1,2} = \pm \Lambda_1^{\frac{1}{2}} \quad ; \quad \lambda_{3,4} = \pm \Lambda_2^{\frac{1}{2}} \quad (3.108)$$

we find that λ_1 and λ_2 are real, while λ_3 and λ_4 are pure imaginary. In the previous Section, it was already concluded that the four roots λ_i are two by two equal in magnitude but of opposite sign, resulting in the equations of motion (3.91). The coefficients A_i and B_i in these equations are not independent. In fact, substitution of (3.93) into (3.89-1) gives

$$B_i = \frac{\lambda_i^2 - 2K - 1}{2\lambda_i} A_i = \gamma_i A_i \quad (3.109)$$

where $i = 1, \dots, 4$. From (3.106) to (3.109) we conclude that the values of λ_i and γ_i are a function of K only, and thus are known for each collinear libration point. Consequently, the four initial conditions $x'_0, y'_0, \dot{x}'_0, \dot{y}'_0$ will completely determine the four coefficients A_i and B_i . When, just as in Section 3.9, we assume that $\lambda_2 = -\lambda_1$ and $\lambda_4 = -\lambda_3$, then (3.109) yields $\gamma_2 = -\gamma_1$ and $\gamma_4 = -\gamma_3$, and we can therefore write (3.91) as

$$\begin{aligned} x' &= A_1 e^{\lambda_1 t} + A_2 e^{-\lambda_1 t} + A_3 e^{\lambda_3 t} + A_4 e^{-\lambda_3 t} \\ y' &= \gamma_1 A_1 e^{\lambda_1 t} - \gamma_1 A_2 e^{-\lambda_1 t} + \gamma_3 A_3 e^{\lambda_3 t} - \gamma_3 A_4 e^{-\lambda_3 t} \end{aligned} \quad (3.110)$$

Because λ_1 is real and λ_3 is purely imaginary, the first two terms on the right-hand side of (3.110) represent exponentially increasing or decreasing motion, and the last two terms represent periodic bounded motion about the libration point. The actual type of motion depends on the values of the integration constants A_i . A few families of trajectories can be discerned:

- If $A_1 = A_2 = 0$, the solution consists of periodic terms only, leading to oscillations about the equilibrium position.
- If $\lambda_1 > 0$ and $A_1 \neq 0$, or if $\lambda_1 < 0$ and $A_2 \neq 0$, then for $t \rightarrow \infty$ the small body moves unboundedly far from the equilibrium position.
- If $\lambda_1 > 0$ and $A_1 = 0$, or if $\lambda_1 < 0$ and $A_2 = 0$, then for $t \rightarrow \infty$ the trajectories asymptotically approach periodic oscillations about the equilibrium position.
- If $\lambda_1 > 0$ and $A_1 = A_3 = A_4 = 0$, or if $\lambda_1 < 0$ and $A_2 = A_3 = A_4 = 0$, then for $t \rightarrow \infty$ the trajectory approaches the equilibrium position.

We now continue our analysis for the case that $A_1 = A_2 = 0$ (pure oscillatory motion). Then, (3.110) can be written as

$$\begin{aligned}x' &= A_3 e^{\lambda_3 t} + A_4 e^{-\lambda_3 t} \\y' &= \gamma_3 A_3 e^{\lambda_3 t} - \gamma_3 A_4 e^{-\lambda_3 t}\end{aligned}\quad (3.111)$$

Substituting the initial conditions:

$$t = 0 : \quad x' = x'_0 \quad ; \quad y' = y'_0$$

we find

$$x'_0 = A_3 + A_4 \quad ; \quad y'_0 = \gamma_3 A_3 - \gamma_3 A_4$$

Solving for A_3 and A_4 , we find

$$A_3 = \frac{\gamma_3 x'_0 + y'_0}{2\gamma_3} \quad ; \quad A_4 = \frac{\gamma_3 x'_0 - y'_0}{2\gamma_3}$$

Substitution of these expressions into (3.111), leads, after some algebraic manipulation, to

$$\begin{aligned}x' &= x'_0 \left[\frac{1}{2}(e^{\lambda_3 t} + e^{-\lambda_3 t}) \right] + \frac{y'_0}{\gamma_3} \left[\frac{1}{2}(e^{\lambda_3 t} - e^{-\lambda_3 t}) \right] \\y' &= y'_0 \left[\frac{1}{2}(e^{\lambda_3 t} + e^{-\lambda_3 t}) \right] + \gamma_3 x'_0 \left[\frac{1}{2}(e^{\lambda_3 t} - e^{-\lambda_3 t}) \right]\end{aligned}\quad (3.112)$$

Because λ_3 and γ_3 are both imaginary, we define

$$\lambda_3 = is \quad ; \quad \gamma_3 = iv \quad (3.113)$$

where s and v are real, and find with (3.107) to (3.109)

$$s = \sqrt{\alpha + \sqrt{\alpha^2 + \beta^2}} > 0 \quad ; \quad v = \frac{s^2 + 2K + 1}{2s} > 0 \quad (3.114)$$

Substituting (3.113) into (3.112), and using the classical relation between trigonometric functions and exponential functions

$$\sin p = \frac{e^{ip} - e^{-ip}}{2i} \quad ; \quad \cos p = \frac{e^{ip} + e^{-ip}}{2}$$

we find

$$x' = x'_0 \cos st + \frac{y'_0}{v} \sin st \quad ; \quad y' = y'_0 \cos st - v x'_0 \sin st \quad (3.115)$$

Differentiation of (3.115) with respect to time, and subsequent substitution of $t = 0$, yields

$$\dot{x}'_0 = \frac{s}{v} y'_0 \quad ; \quad \dot{y}'_0 = -s v x'_0 \quad (3.116)$$

These expressions indicate that if the initial conditions x'_0 and y'_0 have been selected, the initial velocities required for the bounded motion cannot be chosen arbitrarily, but have to satisfy (3.116). For $x'_0 > 0, y'_0 = 0$, (3.116) yields $\dot{x}'_0 = 0, \dot{y}'_0 < 0$. This demonstrates that the direction of motion about the libration point is *opposite* to the direction of rotation of the XYZ-reference

frame.

Squaring both equations (3.115) and subsequently adding the resulting relations, leads to

$$x'^2 + \frac{y'^2}{v^2} = x'^2_0 + \frac{y'^2_0}{v^2} \quad (3.117)$$

So, in the XY -plane the trajectory of a small body about a collinear libration point is an *ellipse* centered at the libration point. Since $K > 1$, we find from (3.114):

$$v = \frac{1}{2} \left(s + \frac{2K+1}{s} \right) > \frac{1}{2} \left(s + \frac{3}{s} \right)$$

Since $s > 0$, we find that the minimum value of the right-hand side of this inequality occurs at $s = \sqrt{3}$ and takes the value $v = \sqrt{3}$. So, we conclude that always $v^2 \geq 3$. This means that the major axis of the elliptical trajectory described by (3.117) is parallel to the Y -axis, and that the minor axis is along the X -axis. When the semi-major axis and semi-minor axis are indicated by a and b , respectively, we find from (3.117):

$$a = \sqrt{v^2 x'^2_0 + y'^2_0} ; \quad b = \sqrt{x'^2_0 + \frac{y'^2_0}{v^2}} ; \quad \frac{a}{b} = v \geq \sqrt{3} \quad (3.118)$$

Note that for a particular collinear libration point the factor a/b is independent of the initial conditions of the body that performs the bounded motion about that libration point. These two-dimensional periodic orbits in the orbital plane of the primary bodies are generally called *Lyapunov orbits* after A.M. Lyapunov (1857-1918), who pioneered the analysis of stability in dynamical systems. Note that, because the libration point does not coincide with one of the two massive bodies of the three-body problem, the small body moves in an elliptical orbit about an ‘empty mathematical point’. This is in contrast to ordinary Keplerian motion (Chapter 5), where the small body moves in a conic section about a massive body. Another remarkable difference is that in Keplerian orbits velocity decreases with increasing distance, while, according to (3.116), in a Lyapunov orbit velocity increases with increasing distance. Combination of the results from (3.110) and (3.118) leads to the conclusion that the trajectory of a small body relative to a collinear libration point is, generally, a Lyapunov periodic orbit superimposed on an exponentially increasing or decreasing component, and that the trajectory is fully determined by the initial conditions. This result is a crucial element of the theory of *invariant manifolds*, which is described in Section 3.12.

For the period of the motion in the XY -plane, we find according to (3.115):

$$T_{xy} = \frac{2\pi}{s} \quad (3.119)$$

where the time is expressed in the unit of time introduced in Section 3.3: $1/\omega$, where ω is the angular velocity of the reference frame. The period of the circular motion of body P_2 about body P_1 is in physical time expressed by

$$T_2 = \frac{2\pi}{\omega} \quad (3.120)$$

So, in physical time we may write

$$T_{xy} = \frac{2\pi}{s\omega} = \frac{T_2}{s} \quad (3.121)$$

and with (3.101) for the motion in the Z-direction:

$$T_z = \frac{T_2}{\sqrt{K}} \quad (3.122)$$

These expressions show that the period of the motion is independent of the initial conditions of the body that moves about the libration point, and is thus independent of the dimension of the periodic orbit. This is an important result for *formation flying* missions, where two or more spacecraft move close together. It will be shown in Section 5.8 and Section 10.1 that two satellites in an orbit about the Earth (or about any other celestial body) will drift apart if they move in different, even slightly different, orbits. A constellation of spacecraft flying about an L point will, according to our first-order analysis, remain together, because the orbital periods of all spacecraft involved are the same.

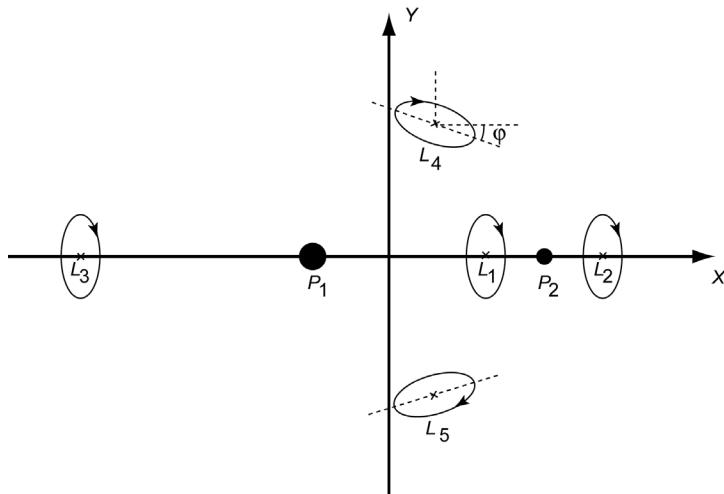


Figure 3.12: Periodic $x'y'$ -motion about the libration points (not to scale).

Figure 3.12 shows a cartoon of the $x'y'$ -motion about the three collinear libration points. Note that, as was already mentioned before, the direction of motion in the elliptical orbits about the L points is opposite to the direction of rotation of the XYZ-reference frame, which means in our case a clockwise motion in the elliptical orbits. The reason for this clockwise motion can be understood as follows. Consider in Figure 3.12 the elliptical orbit about the L_2 point. At the crossing point of this orbit and the X -axis to the right of point L_2 : $\ddot{x}' < 0$ in the elliptical orbit. However, at that point on the X -axis the inward gravitational attraction is slightly smaller than at point L_2 , and the outward centrifugal acceleration is slightly larger than at point L_2 , leading to a residual outward acceleration. The only way to force \ddot{x}' to be negative is to decrease the centrifugal acceleration, which means that at that crossing point \dot{y}' has to be negative. This leads to a clockwise motion about L_2 . Similar arguments can be given for the other crossing point and for the other L points. Table 3.2 summarizes the values of some characteristic parameters of the motion about the collinear libration points of the Sun-Earth(+Moon), Sun-Jupiter and Earth-Moon circular restricted three-body systems. Note that for the L_1 and L_2 libration points the value of the ratio a/b is in each system about 3, while for L_3 this value is 2 in each system. Both for the motion in the XY -plane and for the motion in the Z -direction we note that the period of the motion

Table 3.2: Characteristics of the motion of a small body about the collinear libration points for three three-body systems.

| | Sun-Earth+Moon | | | Sun-Jupiter | | | Earth-Moon | | |
|-----------------|----------------|--------|--------|-------------|--------|--------|------------|--------|--------|
| | L_1 | L_2 | L_3 | L_1 | L_2 | L_3 | L_1 | L_2 | L_3 |
| K | 4.062 | 3.939 | 1.000 | 4.446 | 3.623 | 1.001 | 5.150 | 3.192 | 1.011 |
| U_{xx} | 9.123 | 8.877 | 3.000 | 9.893 | 8.247 | 3.002 | 11.299 | 7.384 | 3.021 |
| U_{yy} | -3.062 | -2.939 | 0.000 | -3.446 | -2.623 | -0.001 | -4.150 | -2.192 | -0.011 |
| U_{zz} | -4.062 | -3.939 | -1.000 | -4.446 | -3.623 | -1.001 | -5.150 | -3.192 | -1.011 |
| s | 2.087 | 2.057 | 1.000 | 2.178 | 1.977 | 1.001 | 2.335 | 1.863 | 1.010 |
| v | 3.229 | 3.187 | 2.000 | 3.360 | 3.074 | 2.000 | 3.587 | 2.913 | 2.000 |
| a/b | 3.229 | 3.187 | 2.000 | 3.360 | 3.074 | 2.000 | 3.587 | 2.913 | 2.000 |
| T_2 (yr/d) | 1.000 | 1.000 | 1.000 | 11.857 | 11.857 | 11.857 | 27.322 | 27.322 | 27.322 |
| T_{xy} | 3.011 | 3.055 | 6.283 | 2.885 | 3.178 | 6.278 | 2.691 | 3.373 | 6.218 |
| T_{xy} (yr/d) | 0.479 | 0.486 | 1.000 | 5.447 | 5.999 | 11.852 | 11.702 | 14.665 | 27.040 |
| T_z | 3.118 | 3.166 | 6.283 | 2.980 | 3.301 | 6.281 | 2.769 | 3.517 | 6.250 |
| T_z (yr/d) | 0.496 | 0.504 | 1.000 | 5.625 | 6.232 | 11.857 | 12.040 | 15.293 | 27.177 |

T_2 , T_{xy} and T_z are given in years for the Sun-Earth+Moon and the Sun-Jupiter systems, and in days for the Earth-Moon system.

is for the L_3 point about twice the period of the motion about the L_1 and L_2 points. To give an impression of the velocity (relative to the rotating reference frame) of the small body, consider the insertion of a spacecraft into a Lyapunov orbit about the L_1 point of the Sun-Earth system. If we assume that injection takes place at $x'_0 = 5000$ km, $y'_0 = 0$, then we find with (3.116) and Table 3.2 that the spacecraft should have velocity components of $\dot{x}'_0 = 0$, $\dot{y}'_0 = -6.7$ m/s.

The three-dimensional motion of the body can be found by superposition of the motion in the XY -plane and the motion in the Z -direction. In general, $s^2 \neq K$, which means that the period of the motion in the Z -direction is different from the period of the motion in the XY -plane. Consequently, the trajectory does not lie in a fixed plane and constitutes a three-dimensional *Lissajous orbit*, named after J.A. Lissajous (1822-1880) who pioneered the study of (acoustic) waves and vibrations. However, since the difference between T_{xy} and T_z is rather small for realistic cases (Table 3.2), the three-dimensional trajectory can be viewed as a *slowly-changing elliptical path*. If the small body is a spacecraft, it is possible to correct for the slow drift of the quasi-elliptical trajectory by firing rocket thrusters periodically. This leads to a quasi-periodic elliptical orbit about the libration point. It is remarkable that only very small rocket pulses after relatively long time intervals (some months) have to be applied to keep the spacecraft in the, basically unstable, Lissajous orbit, which makes this trajectory an attractive concept for various missions. The total maneuver ΔV , generally, is of the order of only 10 m/s per year.

One should keep in mind that all results discussed above follow from linearized theory. Within the constraints of such a theory, the existence of special initial conditions giving trigonometric functions as solutions of the equations of motion means that the collinear libration points, while yielding in general unstable motion, possess *conditional stability* in the linear sense for motions with infinitesimal amplitudes. In reality, of course, second-order effects and forces that are neglected in the circular restricted three-body problem will affect the finite-amplitude motion of the body. However, also in that case, rocket pulses may be applied to keep a spacecraft in a quasi-elliptical orbit about a libration point.

We now consider the L_1 and L_2 points in the Sun-Earth system. Then the XY -plane is the ecliptic, i.e. the plane in which the Earth moves about the Sun. A spacecraft orbiting the L_1 or L_2

point describes a Lissajous pattern centered at that libration point. For both libration points this can have adverse consequences for communications, since at times the line of sight from the Earth to the spacecraft (L_1) or from the spacecraft to the Earth (L_2) comes quite close to the Sun, which introduces radio noise on the downlink signal (L_1) or uplink signal (L_2). A solution is to inject the spacecraft into the Lissajous orbit such that it crosses the Earth-Sun line very late. Before the crossing, an orbit change has to be executed to set the spacecraft on a trajectory that crosses the Earth-Sun line later. However, one would, of course, prefer a situation where the spacecraft circulates about the libration point in a closed loop orbit of fixed geometry and size, which is called a *halo orbit*, and that this orbit is oriented such that the communication problems mentioned before do not occur. It can be shown that halo-type periodic motion is possible if the amplitude of the in-plane motion of the Lissajous orbit is of sufficient magnitude. Then, second-order effects induce a coupling between the motion in the XY -plane and the motion in the Z -direction. This results in a situation where the period of the motion about the libration point is no longer independent of the size of the trajectory, and where the period of the motion in the XY -plane becomes equal to the period of motion in the Z -direction. Consequently, the motion is in a plane that is inclined to the XY -plane. For the Sun-Earth collinear libration points, such an orbit needs a minimum amplitude of about 200,000 km in the X -direction, 650,000 km in the Y -direction, and 120,000 km in the Z -direction.

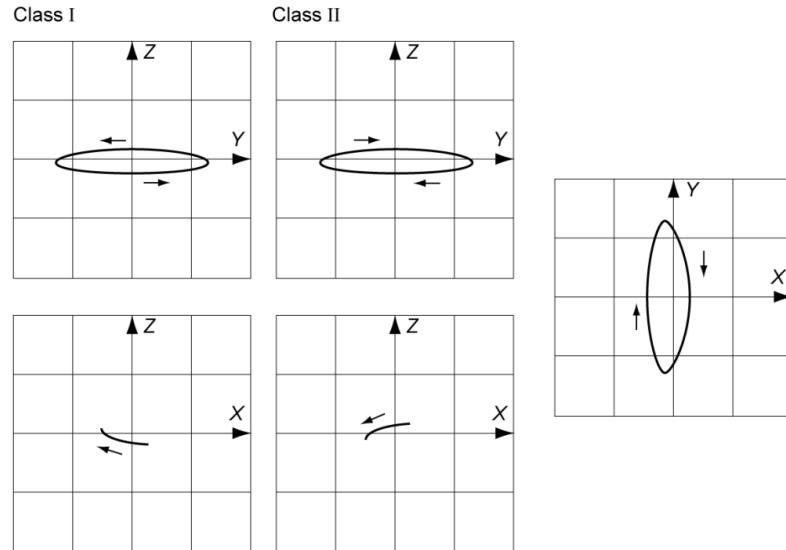


Figure 3.13: Two classes of L_2 halo orbits for the Sun-Earth(+Moon) system with x -, y -, and z -amplitudes of 215,000 km, 686,000 km, and 125,000 km, respectively.

Halo orbits comprise two classes symmetrical to the XY -plane as shown in Figure 3.13. The orbits of *Class 1* have the point of maximum distance from the Earth (apogee) above the ecliptic; the orbits of *Class 2* below the ecliptic. As an example, Figure 3.13 shows for the two classes a L_2 halo orbit for the Sun-Earth(+Moon) system having x -, y -, and z -amplitudes of approximately 215,000 km, 686,000 km, and 125,000 km, respectively. The orbital period of these orbits is about six months and the elongation from the solar direction varies from about 4° to 25° .

Equilateral libration points

For these points $r_1 = r_2 = 1$, and according to (3.102):

$$K = 1 \quad (3.123)$$

which means, according to (3.101), that the period of the motion in the Z -direction is

$$T_z = 2\pi \quad (3.124)$$

So, the period of the motion in the Z -direction is equal to the period of the rotation of the XYZ reference frame relative to inertial space.

Equation (3.98) gives the second-order derivatives of the potential U in these libration points, where the plus-sign in the expression for U_{xy} holds for the libration point L_4 and the minus-sign for the libration point L_5 . With these relations, the equations of motion (3.86) for a small body describing a trajectory about the points L_4 or L_5 can be written as

$$\begin{aligned} \ddot{x}' - 2\dot{y}' - \frac{3}{4}x' &\mp \frac{3}{4}\sqrt{3}(1-2\mu)y' = 0 \\ \ddot{y}' + 2\dot{x}' &\mp \frac{3}{4}\sqrt{3}(1-2\mu)x' - \frac{9}{4}y' = 0 \end{aligned} \quad (3.125)$$

and the characteristic equation (3.90) can be written as

$$\lambda^4 + \lambda^2 + \frac{27}{4}\mu(1-\mu) = 0 \quad (3.126)$$

Defining again

$$\Lambda^2 = \Lambda \quad (3.127)$$

we can write (3.126) as

$$\Lambda^2 + \Lambda + \frac{27}{4}\mu(1-\mu) = 0$$

The roots of this equation are

$$\Lambda_1 = -\frac{1}{2}[1 - \sqrt{1 - 27\mu(1-\mu)}] ; \quad \Lambda_2 = -\frac{1}{2}[1 + \sqrt{1 - 27\mu(1-\mu)}] \quad (3.128)$$

Note that Λ_1 and Λ_2 have either real or complex values, depending on the value of

$$\mathcal{F} = 27\mu(1-\mu) \quad (3.129)$$

In Section 3.9 we have found that if $0 < \mu < 0.03852$ the motion of a small body about the libration points L_4 and L_5 is (unconditional) stable; if $0.03852 < \mu \leq 0.5$ the motion will be unstable. Because for most existing three-body systems $0 < \mu < 0.03852$, we will confine the following discussion to the case of stable motion.

In that case:

$$0 < \mathcal{F} < 1 \quad (3.130)$$

which means, according to (3.128), that both Λ_1 and Λ_2 are real and negative. Consequently, the original four roots λ_i are pure imaginary:

$$\lambda_{1,2} = \pm i\sqrt{-\Lambda_1} ; \quad \lambda_{3,4} = \pm i\sqrt{-\Lambda_2} \quad (3.131)$$

With the notation

$$s_1 = \sqrt{-\Lambda_1} ; \quad s_2 = \sqrt{-\Lambda_2} \quad (3.132)$$

where s and v are real, we can write (3.131) as

$$\lambda_{1,2} = \pm is_1 ; \quad \lambda_{3,4} = \pm is_2 \quad (3.133)$$

where, according to (3.128),

$$s_1 = \sqrt{\frac{1}{2}[1 - \sqrt{1 - 27\mu(1 - \mu)}]} ; \quad s_2 = \sqrt{\frac{1}{2}[1 + \sqrt{1 - 27\mu(1 - \mu)}]} \quad (3.134)$$

Combining (3.129), (3.130) and (3.134), we find

$$0 < s_1 < \frac{1}{2}\sqrt{2} ; \quad \frac{1}{2}\sqrt{2} < s_2 < 1 \quad (3.135)$$

Substitution of (3.134) into (3.91) yields

$$\begin{aligned} x' &= A_1 e^{is_1 t} + A_2 e^{-is_1 t} + A_3 e^{is_2 t} + A_4 e^{-is_2 t} \\ y' &= B_1 e^{is_1 t} + B_2 e^{-is_1 t} + B_3 e^{is_2 t} + B_4 e^{-is_2 t} \end{aligned} \quad (3.136)$$

where, according to (3.89-1), the relation between A_i and B_i is

$$B_i = \frac{\lambda_i^2 - U_{xx}}{2\lambda_i + U_{xy}} A_i \quad (3.137)$$

With (3.133) we can write (3.137), after some algebraic manipulation, as

$$\begin{aligned} B_1 &= \Gamma_1 (2is_1 - U_{xy}) A_1 ; \quad B_2 = -\Gamma_1 (2is_1 + U_{xy}) A_2 \\ B_3 &= \Gamma_2 (2is_2 - U_{xy}) A_3 ; \quad B_4 = -\Gamma_2 (2is_2 + U_{xy}) A_4 \end{aligned} \quad (3.138)$$

where

$$\Gamma_i = \frac{s_i^2 + U_{xx}}{4s_i^2 + U_{xy}^2} \quad (3.139)$$

and s_1, s_2, U_{xy} are known functions of μ , while U_{xx} is a constant. Substitution of (3.98) into (3.139) yields

$$\Gamma_i = \frac{s_i^2 + \frac{3}{4}}{4s_i^2 + \frac{27}{16}(1 - 2\mu)^2} \quad (3.140)$$

This equation can be simplified further. From (3.134) we obtain

$$2s_i^2 = 1 \mp \sqrt{1 - 27\mu(1 - \mu)}$$

or, after some algebraic manipulation,

$$\frac{27}{16}(1 - 2\mu)^2 = s_i^4 - s_i^2 + \frac{27}{16}$$

Substitution of this result into (3.140) yields

$$\Gamma_i = \frac{s_i^2 + \frac{3}{4}}{s_i^4 + 3s_i^2 + \frac{27}{16}} = \frac{1}{s_i^2 + \frac{9}{4}} > 0 \quad (3.141)$$

With the classical relations

$$e^{ip} = \cos p + i \sin p ; \quad e^{-ip} = \cos p - i \sin p$$

we may write (3.136) in the form

$$\begin{aligned} x' &= (A_1 + A_2) \cos s_1 t + i(A_1 - A_2) \sin s_1 t + (A_3 + A_4) \cos s_2 t + i(A_3 - A_4) \sin s_2 t \\ y' &= -\Gamma_1 \{U_{xy}(A_1 + A_2) - 2is_1(A_1 - A_2)\} \cos s_1 t \\ &\quad -\Gamma_1 \{2s_1(A_1 + A_2) + iU_{xy}(A_1 - A_2)\} \sin s_1 t \\ &\quad -\Gamma_2 \{U_{xy}(A_3 + A_4) - 2is_2(A_3 - A_4)\} \cos s_2 t \\ &\quad -\Gamma_2 \{2s_2(A_3 + A_4) + iU_{xy}(A_3 - A_4)\} \sin s_2 t \end{aligned}$$

Defining

$$\begin{aligned} C_1 &= A_1 + A_2 ; \quad S_1 = i(A_1 - A_2) \\ C_2 &= A_3 + A_4 ; \quad S_2 = i(A_3 - A_4) \\ \bar{C}_1 &= -\Gamma_1(U_{xy}C_1 - 2s_1S_1) ; \quad \bar{S}_1 = -\Gamma_1(2s_1C_1 + U_{xy}S_1) \\ \bar{C}_2 &= -\Gamma_2(U_{xy}C_2 - 2s_2S_2) ; \quad \bar{S}_2 = -\Gamma_2(2s_2C_2 + U_{xy}S_2) \end{aligned} \quad (3.142)$$

we finally obtain

$$\begin{aligned} x' &= C_1 \cos s_1 t + S_1 \sin s_1 t + C_2 \cos s_2 t + S_2 \sin s_2 t \\ y' &= \bar{C}_1 \cos s_1 t + \bar{S}_1 \sin s_1 t + \bar{C}_2 \cos s_2 t + \bar{S}_2 \sin s_2 t \end{aligned} \quad (3.143)$$

The four initial ($t = 0$) conditions $x'_0, y'_0, \dot{x}'_0, \dot{y}'_0$ are linearly related to the four independent coefficients C_1, S_1, C_2, S_2 appearing in the solutions (3.143). This can be shown by substituting these initial conditions into (3.143) and into the time-derivatives of these equations:

$$\begin{aligned} x'_0 &= C_1 + C_2 \\ y'_0 &= \bar{C}_1 + \bar{C}_2 = -\Gamma_1(U_{xy}C_1 - 2s_1S_1) - \Gamma_2(U_{xy}C_2 - 2s_2S_2) \\ \dot{x}'_0 &= s_1S_1 + s_2S_2 \\ \dot{y}'_0 &= s_1\bar{S}_1 + s_2\bar{S}_2 = -\Gamma_1s_1(2s_1C_1 + U_{xy}S_1) - \Gamma_2s_2(2s_2C_2 + U_{xy}S_2) \end{aligned} \quad (3.144)$$

where it is recalled that $\Gamma_1, \Gamma_2, s_1, s_2$ and U_{xy} are known functions of μ .

Equations (3.143) show that, in general, the $x'y'$ -motion of a small body about libration point L_4 or L_5 is a superposition of two periodic motions with frequencies s_1 and s_2 . According to (3.135): $s_1 < s_2$, which means that the terms containing s_2 describe *short-period* motion components and the terms containing s_1 describe *long-period* motion components. The periods of both motions are given by

$$T_{xy_l} = \frac{2\pi}{s_1} ; \quad T_{xy_s} = \frac{2\pi}{s_2} \quad (3.145)$$

or, using (3.120) and (3.121), in physical time:

$$T_{xy_l} = \frac{T_2}{s_1} ; \quad T_{xy_s} = \frac{T_2}{s_2} \quad (3.146)$$

Either the short- or the long-period terms in the motion of the body can be eliminated by selecting the initial conditions properly. We will consider the case of the elimination of short-period terms, but the same approach can be applied to eliminate long-period terms. In fact, the equations that will be derived may be used directly for the short-period motion component, if all terms with index 1 are replaced by corresponding terms with index 2. The reason being that (3.143) are symmetrical in terms containing s_1 and terms containing s_2 .

According to (3.143), the elimination of short-period terms requires

$$C_2 = S_2 = \bar{C}_2 = \bar{S}_2 = 0 \quad (3.147)$$

Substitution of (3.147) into (3.144) leads to

$$x'_0 = C_1 ; \quad \dot{x}'_0 = s_1 S_1$$

$$y'_0 = -\Gamma_1 (U_{xy} C_1 - 2s_1 S_1) ; \quad \dot{y}'_0 = -\Gamma_1 s_1 (2s_1 C_1 + U_{xy} S_1)$$

From these equations the parameters C_1 and S_1 can be eliminated, leading to expressions for the initial velocity required to perform the prescribed long-period motion, starting at the initial position components x'_0, y'_0 :

$$\dot{x}'_0 = \frac{1}{2} \left(x'_0 U_{xy} + \frac{y'_0}{\Gamma_1} \right) ; \quad \dot{y}'_0 = -\frac{1}{2} \left\{ \Gamma_1 x'_0 (4s_1^2 + U_{xy}^2) \right\} - U_{xy} y'_0 \quad (3.148)$$

Because s_1, U_{xy} and Γ_1 are known functions of μ , the required initial velocity components can be computed from (3.148). Since $\Gamma_1 > 0$, and $U_{xy} > 0$ for libration point L_4 and $U_{xy} < 0$ for libration point L_5 , we conclude from (3.148) that at $x'_0 > 0, y'_0 = 0$ the following conditions hold:

$$L_4: \quad \dot{x}'_0 > 0 , \quad \dot{y}'_0 < 0 ; \quad L_5: \quad \dot{x}'_0 < 0 , \quad \dot{y}'_0 < 0$$

This means that in both cases the small body traverses the trajectory about the libration point in a direction *opposite* to the direction of rotation of the reference frame relative to inertial space, just as in the collinear libration points case.

To find the characteristics of the long-period motion, we start from (3.143) and write

$$\begin{aligned} x' &= C_1 \cos s_1 t + S_1 \sin s_1 t \\ y' &= \bar{C}_1 \cos s_1 t + \bar{S}_1 \sin s_1 t \end{aligned} \quad (3.149)$$

Multiplying the first of these equations by \bar{C}_1 , the second by $-C_1$, and subsequently adding both resulting expressions gives

$$x' \bar{C}_1 - y' C_1 = (S_1 \bar{C}_1 - C_1 \bar{S}_1) \sin s_1 t \quad (3.150-1)$$

Multiplying the first of (3.149) by \bar{S}_1 and the second by $-S_1$, and subsequently adding the resulting expressions gives

$$x' \bar{S}_1 - y' S_1 = -(S_1 \bar{C}_1 - C_1 \bar{S}_1) \cos s_1 t \quad (3.150-2)$$

Squaring both equations (3.150) and subsequently adding the results yields

$$(\bar{C}_1^2 + \bar{S}_1^2)x'^2 - 2(C_1 \bar{C}_1 + S_1 \bar{S}_1)x'y' + (C_1^2 + S_1^2)y'^2 = (S_1 \bar{C}_1 C_1 - C_1 \bar{S}_1 S_1)^2 \quad (3.151)$$

With (3.142), the various combinations of coefficients in this equation may be written as

$$\bar{C}_1^2 + \bar{S}_1^2 = \Gamma_1^2 (C_1^2 + S_1^2) (4s_1^2 + U_{xy}^2)$$

$$C_1 \bar{C}_1 + S_1 \bar{S}_1 = -\Gamma_1 (C_1^2 + S_1^2) U_{xy}$$

$$S_1 \bar{C}_1 - C_1 \bar{S}_1 = 2\Gamma_1 (C_1^2 + S_1^2) s_1$$

Substitution of these relations into (3.151) yields

$$\Gamma_1^2 (4s_1^2 + U_{xy}^2)x'^2 + 2\Gamma_1 U_{xy}x'y' + y'^2 = 4\Gamma_1^2 s_1^2 (C_1^2 + S_1^2) \quad (3.152)$$

If the term with $x'y'$ would be absent in this equation, then the ‘reduced’ equation indicates that the body would move in an elliptical orbit about the libration point. In an attempt to eliminate the term containing the product $x'y'$, we introduce the coordinates \vec{x}' and \vec{y}' , defined by

$$x' = \vec{x}' \cos \varphi - \vec{y}' \sin \varphi ; \quad y' = \vec{x}' \sin \varphi + \vec{y}' \cos \varphi \quad (3.153)$$

The introduction of these new coordinates effectively means a rotation of the set of coordinates $x'y'z'$ about the Z' -axis over an angle φ in the direction of the angular velocity of the XYZ reference frame relative to inertial space. Substitution of (3.153) into (3.152) yields, after some algebraic manipulation,

$$\begin{aligned} & [\{\Gamma_1^2 (4s_1^2 + U_{xy}^2) - 1\} \cos^2 \varphi + \Gamma_1 U_{xy} \sin 2\varphi + 1] \vec{x}'^2 + \\ & [-\{\Gamma_1^2 (4s_1^2 + U_{xy}^2) - 1\} \sin 2\varphi + 2\Gamma_1 U_{xy} \cos 2\varphi] \vec{x}' \vec{y}' + \\ & [\{\Gamma_1^2 (4s_1^2 + U_{xy}^2) - 1\} \sin^2 \varphi - \Gamma_1 U_{xy} \sin 2\varphi + 1] \vec{y}'^2 = 4\Gamma_1^2 s_1^2 (C_1^2 + S_1^2) \end{aligned} \quad (3.154)$$

Note that the mixed $\vec{x}'\vec{y}'$ -term becomes zero, if

$$\tan 2\varphi = \frac{2\Gamma_1 U_{xy}}{\Gamma_1^2 (4s_1^2 + U_{xy}^2) - 1}$$

With (3.98), (3.134) and (3.141), we can write this expression as

$$\tan 2\varphi = \mp \sqrt{3} (1 - 2\mu) \quad (3.155)$$

where the minus-sign holds for the point L_4 and the plus-sign for the point L_5 . So, for $\mu < 0.0385$: $\varphi \approx -30^\circ$ at L_4 and $\varphi \approx 30^\circ$ at L_5 . A sketch of the periodic $x'y'$ -motion is given in Figure 3.12.

When φ is specified by (3.155), we may write (3.154) as

$$\frac{\vec{x}'^2}{G_3^2/G_1} + \frac{\vec{y}'^2}{G_3^2/G_2} = 1 \quad (3.156)$$

where

$$\begin{aligned}
 G_1 &= \left\{ \Gamma_1^2 (4s_1^2 + U_{xy}^2) - 1 \right\} \cos^2 \varphi + \Gamma_1 U_{xy} \sin 2\varphi + 1 \\
 G_2 &= \left\{ \Gamma_1^2 (4s_1^2 + U_{xy}^2) - 1 \right\} \sin^2 \varphi - \Gamma_1 U_{xy} \sin 2\varphi + 1 \\
 G_3^2 &= 4\Gamma_1^2 s_1^2 (C_1^2 + S_1^2)
 \end{aligned} \tag{3.157}$$

and G_1 , G_2 and G_3 are functions of μ only. The functions G_1 and G_2 are plotted in Figure 3.14. This Figure shows that $G_1 \geq 0$, $G_2 > G_1 > 1$, which means that (3.156) represents an ellipse with its major axis parallel to the \vec{x} direction. The magnitudes of the semi-major axis, a , of the semi-minor axis, b , and of the ratio a/b are given by

$$a = \frac{G_3}{\sqrt{G_1}} ; \quad b = \frac{G_3}{\sqrt{G_2}} ; \quad \frac{a}{b} = \sqrt{\frac{G_2}{G_1}} \tag{3.158}$$

The ratio a/b is shown in Figure 3.14 as a function of μ . A numerical analysis has shown that if $0.0002 < \mu < 0.0385$ then $2.43 < a/b < 41$. Figure 3.14 also shows similar results for the short-period motion. In that case a numerical analysis has shown that if $0.0002 < \mu < 0.0385$ then $2.00 < a/b < 2.40$. Just as in the case of elliptical motion about the collinear libration points, the period of the motion in the Z -direction will, in general, differ from the period of motion in the XY -plane. Consequently, the trajectory does not lie in a fixed plane and is shaped as a Lissajous figure.

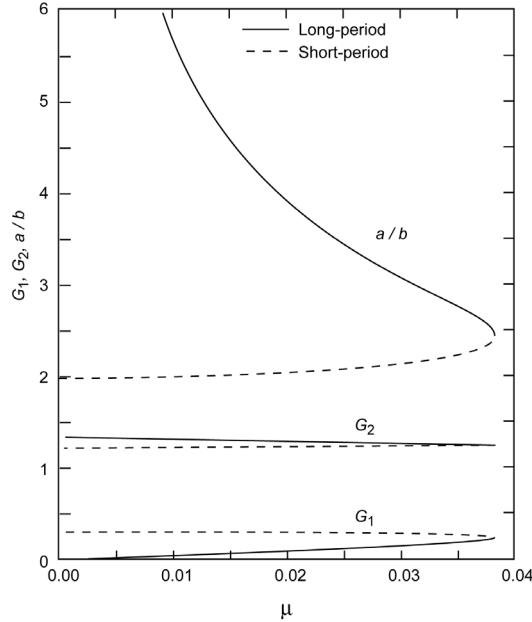


Figure 3.14: Values of G_1 , G_2 and a/b as a function of μ , both for the long-period and the short-period components of the motion about the L_4 and L_5 points.

Table 3.3 summarizes the values of some characteristic parameters of the motion about the L_4 and L_5 points for the Sun-Earth(+Moon), Sun-Jupiter and Earth-Moon circular restricted three-body problems. Note that for the short-period component of the motion (index s), the periods in the XY -plane and in the Z -direction are about equal. This means that this motion may be visualized as a slowly-changing elliptical path. By firing small thrusters periodically it is possible to correct for the slow drift of the quasi-elliptical trajectory, leading to a near-halo orbit. For the long-period

Table. 3.3: Characteristics of the motion of a small body about the triangular libration points for three three-body systems.

| | Sun-Earth+Moon | | Sun-Jupiter | | Earth-Moon | |
|---------------------|----------------|---------|-------------|---------|------------|---------|
| | L_4 | L_5 | L_4 | L_5 | L_4 | L_5 |
| U_{xx} | 0.7500 | 0.7500 | 0.7500 | 0.7500 | 0.7500 | 0.7500 |
| U_{xy} | 1.2990 | -1.2990 | 1.2966 | -1.2966 | 1.2675 | -1.2675 |
| U_{yy} | 2.2500 | 2.2500 | 2.2500 | 2.2500 | 2.2500 | 2.2500 |
| U_{zz} | -1.0000 | -1.0000 | -1.0000 | -1.0000 | -1.0000 | -1.0000 |
| s_1 | 0.00453 | 0.00453 | 0.0805 | 0.0805 | 0.2982 | 0.2982 |
| s_2 | 0.9999 | 0.9999 | 0.9968 | 0.9968 | 0.9545 | 0.9545 |
| $(a/b)_l$ | 331.27 | 331.27 | 18.668 | 18.668 | 5.1316 | 5.1316 |
| $(a/b)_s$ | 2.0000 | 2.0000 | 2.0022 | 2.0022 | 2.0344 | 2.0344 |
| T_2 (yr/d) | 1.000 | 1.000 | 11.857 | 11.857 | 27.322 | 27.322 |
| $(T_{xy})_l$ | 1387.04 | 1387.04 | 78.0860 | 78.0860 | 21.0704 | 21.0704 |
| $(T_{xy})_s$ (yr/d) | 220.75 | 220.75 | 147.42 | 147.42 | 91.623 | 91.623 |
| $(T_{xy})_s$ | 6.2832 | 6.2832 | 6.3036 | 6.3036 | 6.5827 | 5.5827 |
| $(T_{xy})_s$ (yr/d) | 1.0000 | 1.0000 | 11.901 | 11.901 | 28.624 | 28.624 |
| T_z | 6.2832 | 6.2832 | 6.2832 | 6.2832 | 6.2832 | 6.2832 |
| T_z (yr/d) | 1.0000 | 1.0000 | 11.862 | 11.862 | 27.322 | 27.322 |
| φ (°) | -30.00 | 30.00 | -30.00 | 30.00 | -30.00 | 30.00 |

T_2 , T_{xy} and T_z are given in years for the Sun-Earth+Moon and the Sun-Jupiter systems, and in days for the Earth-Moon system.

motion component (index l), we observe that the period in the XY -plane is much larger than the period in the Z -direction. The oscillation with the long period is generally called *libration*. For the Sun-Earth and Sun-Jupiter cases the values of the ratio a/b are very large, meaning that the motion in the XY -plane primarily is a motion along the X' -axis.

For arbitrary initial conditions, the motion of the body about the equilateral libration points

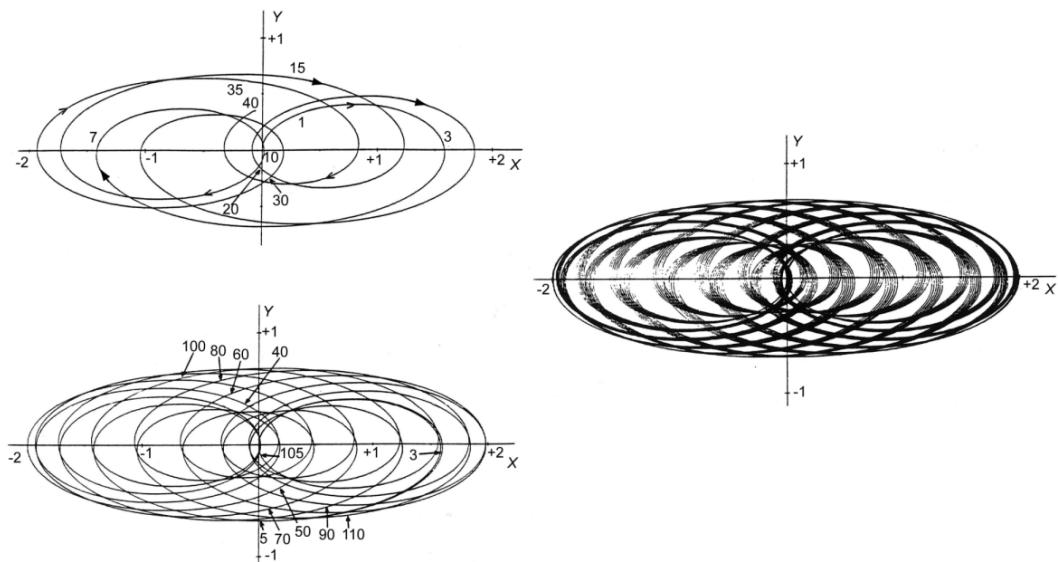


Figure 3.15: Linearized solution for the trajectory about an Earth-Moon triangular libration point for 0.5 year (top, left), one recurrency (≈ 1.3 year, bottom, left) and five recurrences (≈ 6.3 year, right). Numbers on curves denote the elapsed dimensionless time, t^* .

will be a superposition of the short-period (s_2) and the long-period (s_1) solutions. Figure 3.15 shows an example of a complete linearized solution for the $\vec{x}'\vec{y}'$ -motion about the Earth-Moon libration point L_4 . The numbers along the orbit represent the elapsed non-dimensional time $t^* = 2\pi/27.32 t$, where t is measured in days. The orbit is non-periodic, since s_1/s_2 is an irrational number. The motion starts at $\vec{x}' = \vec{y}' = 0$, with $\dot{\vec{x}}' = 0, \dot{\vec{y}}' > 0$.

We should keep in mind, however, that the analysis presented in this Section only yields a linearized solution for the actual motion about the L_4 and L_5 points. In reality, second-order terms and other perturbing forces will produce effects that will disturb this simple motion.

3.11. Application of Jacobi's integral for lunar trajectories

The circular restricted three-body problem is of great importance for a qualitative analysis of trajectories to the Moon (Chapter 17). In Earth-Moon space it are primarily the Earth and the Moon that determine the trajectory of a spacecraft traveling to the Moon. Since the Moon moves in an almost circular orbit about the Earth, one can obtain a first-order approximation of such a trajectory by regarding the motion as a circular restricted three-body problem.

In the years 1968-1972 the USA has flown a number of manned missions to the Moon in the Apollo program (Chapter 17). The Apollo spacecraft consisted of a Command Module (CM) that housed the three astronauts, a Service Module (SM) that carried the propulsion system, electrical power system and life-support system, and a Lunar Module (LM), in which two astronauts performed a landing on the Moon; the combination of CM and SM was indicated by CSM. The Apollo spacecraft was launched by the giant Saturn V rocket. After injection into a lunar trajectory, the CSM separated from the third stage of the launcher, the S-IVB stage, turned in space by 180° and docked with the LM that was still attached to the third stage. Then, the complete Apollo spacecraft separated from the third stage. In this Section, we take an Apollo trajectory as a characteristic example of a classical trajectory to the Moon.

We consider the flight of the Apollo spacecraft after it has been injected into a trajectory to the Moon by the third stage of the Saturn V rocket, and we will perform a simplified analysis for determining the required injection velocity. In Figure 17.12, a picture of the mission geometry is shown. For our simplified analysis we first consider the motion of the spacecraft relative to the rotating reference frame XYZ defined in Section 3.3, and assume that: 1) the trajectory of the spacecraft lies in the plane in which the Moon moves about the Earth ($z = 0$); 2) the spacecraft is injected by an impulsive shot at an altitude of 200 km above the Earth into a lunar trajectory (Section 1.7); 3) injection occurs above that part of the Earth's surface that is turned away from the Moon, and on the line connecting Earth and Moon; 4) the spacecraft's velocity vector at injection is directed perpendicular to the line connecting Earth and Moon, and in the negative Y -direction (Figure 3.4). These assumptions are quite realistic. With the expressions given in Section 3.7 for $\gamma_1, \gamma_2, \gamma_3$, and for the location of the L_4 and L_5 libration points, and the geometry depicted in Figure 3.8, we compute the x - and y -coordinates of the five libration points in the Earth-Moon system ($\mu = 0.0121506$) in the non-dimensional units applied in Section 3.7. Subsequently, we compute from (3.59) the value of Jacobi's constant, C , for the case that a surface of Hill passes through a libration point, also in non-dimensional units. These values of x, y , and C are listed in Table 3.4 for each libration point. Note that if the Apollo spacecraft has to reach the Moon, we must require $C < 3.18834$. If $C < 2.98799$, any place in Earth-Moon space is accessible to the Apollo spacecraft. This means that when no braking would occur near the Moon, the spacecraft would probably escape from the Earth-Moon system.

Table. 3.4: Position of the libration points in the Earth-Moon system, boundary values of C and the corresponding injection velocities for a simplified Apollo-type trajectory.

| | L_1 | L_2 | L_3 | L_4, L_5 |
|--------------------|----------|----------|-----------|----------------|
| γ | 0.150871 | 0.167884 | 0.992912 | 1 |
| x | 0.836978 | 1.155734 | -1.005063 | 0.487849 |
| y | 0 | 0 | 0 | ± 0.866025 |
| C | 3.188341 | 3.172160 | 3.012147 | 2.987997 |
| V_{rot} (km/s) | 10.85715 | 10.85793 | 10.86564 | 10.86680 |
| V_{inert} (km/s) | 10.88709 | 10.88787 | 10.89558 | 10.89675 |

To compute the injection velocity, relative to the rotating reference frame, that corresponds to the value of C for which the surface of Hill passes through a libration point, we substitute (3-48) into (3.53-2) and write the resulting expression in the conventional physical units. We then find

$$V_{rot}^2 = \omega^2(x^2 + y^2) + 2G\left(\frac{m_1}{r_1} + \frac{m_2}{r_2}\right) - (P_1 P_2)^2 \omega^2 C \quad (3.159)$$

where V_{rot} is the velocity of the spacecraft relative to the rotating reference frame, and the other parameters are defined in Section 3.3. It is emphasized that C is still expressed in the non-dimensional units applied in Section 3.3. We now apply (3.159) for the Earth-Moon system and write $m_1 = m_E$, $m_2 = m_M$, $P_1 P_2 = r_M$, $\omega = n_M$, where m_E is the mass of the Earth, m_M is the mass of the Moon, r_M the mean distance between Earth and Moon; $n_M = 2\pi/T_{sid}$, where T_{sid} is the length of a sidereal month, is the mean motion of the Moon in its orbit about the Earth. With this notation we can write (3.159) as

$$V_{rot}^2 = n_M^2(x^2 + y^2) + 2G m_E \left(\frac{1}{r_1} + \frac{m_M}{m_E r_2} \right) - n_M^2 r_M^2 C \quad (3.160)$$

From the adopted injection geometry relative to the rotating reference frame, we find

$$x = -\frac{m_M r_M}{m_E + m_M} - R - h \quad ; \quad y = 0 \quad ; \quad r_1 = R + h \quad ; \quad r_2 = r_M + r_1 \quad (3.161)$$

where R is the mean equatorial Earth's radius and h is the altitude of the injection point ($h = 200$ km). Combination of (3.160) and (3.161) yields the injection velocity, V_{rot} , for each value of C . The values of Gm_E , $m_E/m_M = (1 - \mu)/\mu$, r_M , T_{sid} , and R are given in Appendix B. Note that (3.160) only yields the magnitude of the injection velocity and not its direction, which means that the trajectory after injection cannot be computed. We can only say that if C is slightly smaller than the value for which a surface of Hill passes through libration point L_2 , then a spacecraft may escape from the Earth-Moon system, will then fly through the ‘neck’ region around L_1 and L_2 (Section 3.12) and will pass the Moon at a relatively small distance.

We have assumed that the injection velocity is perpendicular to the X -axis and in the negative Y -direction and we therefore can compute the injection velocity relative to the inertial reference frame, V_{inert} , from

$$V_{inert} = V_{rot} + n_M \left(\frac{m_M r_M}{m_E + m_M} + R + h \right)$$

The values of V_{rot} and V_{inert} are listed in Table 3.4 for each value of C . We conclude that in order to reach the Moon, the spacecraft has to be accelerated to a velocity of at least 10.88709 km/s with respect to the inertial reference frame. A velocity that is only 0.78 m/s larger is already sufficient to escape from the Earth-Moon system. This leads to the important conclusion that a lunar trajectory requires almost as much energy as an escape trajectory from the Earth-Moon system, which is a minimum requirement for interplanetary flights. In reality, the injection of an Apollo spacecraft was certainly not an impulsive shot: the rocket engine of the Saturn third stage that was used for leaving the initial near-circular low-altitude parking orbit about the Earth and for injection into the translunar trajectory burned for 5.3–5.8 min. During this powered flight the spacecraft covered an appreciable horizontal and vertical distance, resulting in significantly higher injection altitudes. Table 17.2 presents a summary of trajectory parameters for a series of Apollo flights. From that Table we conclude that: 1) injection altitude actually varied from 314 km to 369 km; 3) injection velocity actually varied from 10.79 km/s to 10.84 km/s. When the differences in injection altitude are accounted for, the value of 10.887 km/s computed above with a simplified theory compares quite well with the real injection velocities of the Apollo spacecraft. In Section 17.3, results from a simplified two-dimensional analysis of a lunar mission that starts from a 200 km circular parking orbit about the Earth and ends in a 500 km altitude orbit about the Moon are presented. For that analysis it was assumed that both the acceleration maneuver to leave the parking orbit about the Earth and the deceleration maneuver to enter an orbit about the Moon can be modeled as an impulsive shot. The minimum-energy trajectory for that mission was found to require an injection velocity of 10.914 km/s. As expected, this value is slightly larger than the ‘absolute’ minimum injection velocity of 10.887 km/s found in this Section.

In Section 7.1 an expression for the escape velocity, V_{esc} , is derived ((7.3)), for the case that the motion of the spacecraft is considered as a pure two-body problem. From that expression we find for an altitude of 200 km above the Earth: $V_{esc} = 11.0086$ km/s. From Table 3.4 we conclude that for the three-body approximation an injection velocity of at least 10.8879 km/s is required to escape from the Earth-Moon system. Consequently, for the two-body approximation the injection velocity required to escape from the Earth is about 0.12 km/s higher than the minimum escape injection velocity computed for the three-body model. So, we conclude that the gravity field of the Moon may (in certain cases) reduce the injection impulse required to escape from the Earth. This is, in fact, an example of the *swingby effect* that will be discussed in Section 18.11.

Upon arrival at the Moon, the engine of the SM was ignited such that the value of C was increased to $C > 3.1883$. The applied deceleration impulse was about 0.9 km/s. As a result, the Apollo spacecraft entered an orbit about the Moon. The value of C was, albeit in a somewhat different form, calculated by the onboard computer of the Apollo spacecraft and was, in addition to other parameters, used to determine the burning period of the engine.

The first Apollo flights have used so-called *free-return trajectories*. These trajectories were designed such that after the S-IVB engine stopped thrusting the spacecraft was able to reach the Moon, but could not escape from the Earth-Moon system ($C > 3.1722$), even if the SM engine that had to provide the braking to enter an orbit about the Moon would fail. Moreover, in this trajectory the spacecraft would approach the Earth again after a reasonable period of time, providing an option for a save return. From Apollo-12 on, a so-called *hybrid trajectory* was flown in order to save propellants and to satisfy a number of operational constraints. In these missions, the flight started out in a free-return trajectory. After the CSM had docked with the LM during its translunar trajectory, a mid-course correction of about 4.5 m/s was applied that brought the spacecraft into a *non-free-return trajectory* to the Moon. If for these trajectories the spacecraft was not decelerated near the Moon, it would irrevocably be lost in space. The non-free-return trajectories were designed such that if the SM engine would fail, the spacecraft’s reaction control

system or the LM engine could provide the impulse required to go back to a free-return trajectory. As we know, a serious problem has occurred with Apollo-13, launched in April 1970. That spacecraft moved in such a non-free-return trajectory, when at a distance of about 322,000 km from Earth an explosion occurred within the SM, destroying the SM engine. In order to save the crew, the descent engine of the LM was used to increase the value of C of the spacecraft such that it returned to a free-return trajectory that, after passing the Moon, would bring the spacecraft back to Earth. This maneuver was successful and the crew was saved. In later Apollo flights the shift from a free-return trajectory to a non-free-return trajectory was executed progressively earlier in the mission, and during the last Apollo 15 to 17 missions the spacecraft was even directly injected into a non-free-return trajectory.

3.12. Ballistic capture, weak stability boundary and invariant manifold

For a numerical computation of the trajectory of a spacecraft in Earth-Moon space, in interplanetary space, or within the natural satellite systems of the outer planets, the gravitational attractions by all relevant celestial bodies are taken into account. So, the motion of the spacecraft is considered as a many-body problem. For the analytical analysis of such trajectories, traditionally the motion is described by a series of patched two-body problems (Sections 17.3 and 18.1). In Section 5.3 it will be shown that in a two-body problem the trajectories are conic sections. So, in the classical approach the trajectory is described by a series of patched conics, where for each conic section only the gravitational attraction by one celestial body is taken into account. However, in the analytical approach the motion can more accurately be described by a, or by a series of, (circular) restricted three-body problem(s); in particular when it concerns flights to the Moon or within natural satellite systems. When the spacecraft is close to a celestial body, the motion may be considered again as a perturbed two-body problem. In this analysis, use is made of the known characteristics of two-body orbits or of (circular) restricted three-body trajectories. This dualism of describing a spacecraft trajectory as a series of two-body trajectories or as a, or a series of, three-body trajectory(ies), is very helpful for understanding modern transfer trajectories in Earth-Moon space or within the planetary natural satellite systems.

In this Section, three fundamental concepts: *ballistic capture*, *weak stability boundary* and *invariant manifold*, which are nowadays often applied for the design of spacecraft trajectories to the Moon, the planets, and the moons of the outer planets, will be introduced. A full treatment of these concepts is beyond the scope of this book and therefore a qualitative analysis addressing only the main features of these concepts will be presented.

Ballistic capture

To understand the concept of *ballistic capture*, also called *gravitational capture*, it is necessary to discriminate between a *closed* and an *open* two-body orbit. A spacecraft is in a closed orbit about a celestial body if its velocity is not large enough to escape from that body; a spacecraft is in an open orbit about a celestial body if its velocity is large enough to escape from that body. To identify the type of two-body orbit about the celestial body, we may use the definition of the total orbital energy in the two-body problem: $\mathcal{E} = \frac{1}{2}V^2 - \mu/r$ (Section 5.1), where V is the velocity of the spacecraft relative to a non-rotating reference frame centered at the attracting body, and r is the distance of the spacecraft from that body. In Chapters 6 to 8 it will be shown that this total energy is negative for an elliptical orbit, zero for a parabolic orbit and positive for a hyperbolic orbit. So, we may say that the spacecraft is in an open orbit if its energy is zero or positive, and that it is in a closed orbit if its energy is negative. In the two-body problem this energy remains constant. However, when we consider the motion of the spacecraft as a three-body problem, the

total energy of the spacecraft relative to one of the two main bodies is no longer constant, and can change sign from positive to negative or from negative to positive. When the variation is from positive to negative, a *ballistic capture* has taken place. The opposite situation, when the energy changes from negative to positive is called a *ballistic escape*.

Now, consider a spacecraft that was launched from the Earth and is approaching the Moon. If at a certain position its velocity is too high, then the Moon's gravity field will only bend the trajectory and the spacecraft will escape again from the Moon's gravity field. If the velocity is too low, the trajectory will be bent very much and the spacecraft may even crash on the lunar surface. Therefore, the question may be raised: "Is it possible that the spacecraft arrives at a certain position above the lunar surface with a special magnitude and direction of its velocity relative to the Moon such that it is just captured by the gravity field of the Moon"? Then, the spacecraft will be in a phase between escape from and capture by the lunar gravity field. This is basically a chaotic phase, and only a tiny amount of ΔV at some moment is needed to change the motion of the spacecraft from an escape trajectory to a closed orbit about the Moon, or vice versa. If the spacecraft is in a (temporary) closed orbit and this ΔV is not applied, the spacecraft will perform several orbits around the Moon and then it will escape again. Therefore, this process is generally called *weak capture*. E.A. Belbruno (1951-) has addressed the concept of weak ballistic capture around 1987 in his pioneering research on propellant-efficient trajectories for the LGAS (Lunar Get Away Special) and Hiten (Section 17.5) spacecraft.

Weak stability boundary

As stated above, weak ballistic capture may occur when a spacecraft arrives at a given location near the Moon with special values of the direction and magnitude of its velocity. A combination of all possible cases yields a catalog of velocity values a spacecraft would need as a function of altitude, latitude and longitude relative to the Moon. This catalog forms what was called by Belbruno a *weak stability boundary* around the Moon. The weak stability boundary is a very useful concept for designing low-energy (propellant-efficient) trajectories from the Earth to the Moon, or for designing low-energy trajectories through the multi-moon systems of the large outer planets. In Sections 17.5 and 18.12 applications for lunar and planetary missions will be described.

In Section 3.7, it was shown that at the five Lagrange libration points the gravitational forces of Earth and Moon together with the centrifugal force exactly balance on a spacecraft, provided that the spacecraft keeps a fixed position relative to the rotating reference frame and so relative to the Moon. The weak stability boundary about the Moon can be viewed as a region where these three forces approximately balance when the spacecraft is, more generally, in motion with respect to the Moon. Belbruno has demonstrated that the motion associated with this region is not only chaotic in nature but also resonant in nature. This means that if a trajectory started in this region as an orbit about the Moon, then the motion would quickly evolve into an elliptical orbit about the Earth, in resonance with the Moon. Moreover, when the trajectory returned to the Moon, it would again interact with the weak stability boundary, and go into another resonant elliptical orbit about the Earth, of a different resonance type. Thus, the spacecraft would perform a resonance transition, from one resonant orbit to another via a weak capture phase. In principle, the same definition can be applied to define weak stability regions about the Lagrange points. In this book, however, we will not do that.

A typical Belbruno low-energy trajectory to the Moon consists of the following phases: 1) spacecraft injection from an initial circular orbit about the Earth into an elliptical orbit that crosses the Moon's orbit; 2) swingby maneuver (Section 18.11) about the Moon to increase the apocenter distance of the initial elliptical orbit; 3) use of solar gravitational attraction, and

sometimes a small rocket impulse, when the spacecraft is in the apocenter region, in order to rise the pericenter to the distance of the Moon's orbit from the Earth; 4) ballistic capture of the spacecraft by the Moon. The apocenter lies in the region about the Sun-Earth L_1 or L_2 libration points, which are located at a distance of about 1.5×10^6 km from the Earth. There, the velocity and acceleration of the spacecraft are small. So, the spacecraft stays for a relatively long period of time in a region where the gravitational attraction by the Sun may drastically change the spacecraft's trajectory. Typically, the total ΔV required for such a transfer from a low-altitude orbit about the Earth to a low-altitude orbit about the Moon is about 0.3 km/s less than for a classical direct transfer. However, the transfer takes 100-300 days, versus about 5 days for a direct transfer.

An analytical analysis of this type of transfer trajectories is difficult, since the motion is basically characterized by an Earth-Moon-Sun-spacecraft restricted four-body problem. However, a qualitative analysis can be given when the motion is considered as two coupled (circular) restricted three-body problems: a Sun-Earth-spacecraft and an Earth-Moon-spacecraft problem. In Section 3.6 it was shown that for small enough values of Jacobi's constant, C , the surface of Hill 'opens' at the L_2 libration point, and a spacecraft that originally moves about main body P_1 (Earth) can leave the three-body system, or a spacecraft from 'outside' this system can reach P_1 or P_2 (Moon). A spacecraft that has to arrive in the region about the Sun-Earth L_1 or L_2 libration points has to leave the Earth-Moon system. For designing propellant-efficient (low-energy) Belbruno transfer trajectories from an orbit about the Earth to an orbit about the Moon, we are therefore particularly interested in trajectories with a value of Jacobi's constant just below that corresponding to the Earth-Moon L_2 point ($C = 3.172$; Table 3.4). For this case, the three-dimensional Hill's region contains a 'neck' about L_1 and L_2 (Figures 3.6^d and 3.6^e). The spacecraft then crosses three regions of unstable dynamical equilibrium: the L_1 or L_2 region of the Sun-Earth-spacecraft circular restricted three-body problem, and the L_1 and L_2 regions of the Earth-Moon-spacecraft circular restricted three-body problem. Furthermore, the passage through the Earth-Moon L_1 - L_2 region guides the spacecraft to a dynamical state close to the corresponding zero velocity curves of the Earth-Moon system. For this type of exterior trajectories the angle between the Earth-apocenter line and the Earth-Sun line is crucial. If this angle lies in the appropriate quadrant the solar gravitational attraction will shape the trajectory such that the spacecraft can be weakly captured by the Moon. The solar effect must be strong enough to raise the pericenter to the lunar orbit, but small enough to allow the spacecraft to reach the Moon with the minimum possible energy in order to be (weakly) captured by the lunar gravity field. A small ΔV is then required to lower the apolune, since further Earth perturbations could again send the spacecraft into a higher energy escape orbit.

It is noted that in contrast to the exterior transfer discussed above, there also exist interior low-energy lunar transfers. In these interior transfers, the semi-major axis of the spacecraft's orbit about the Earth is steadily increased by small ΔV 's provided by a chemical propulsion system or by electric propulsion thrusting arcs (Chapter 19), and resulting from lunar flybys, until the spacecraft is weakly captured by the Moon. Then the Moon centered semi-major axis is decreased by the same method. In this transfer technique, the spacecraft-Moon synchronization is very important, since the spacecraft has to meet the Moon properly and several times at several apogee distances. The ΔV saving and flight time are about equal to the values obtained for an exterior low-energy lunar transfer. Of course, in designing interior transfers, the solar gravitational field must be taken into account, but this is a perturbation effect, which does not change the general scheme.

Invariant manifold

In the late nineteenth century, Poincaré searched for mathematical theories that would allow an understanding of the dynamical stability of systems. His investigations resulted in the development of *dynamical systems theory*. This theory is based on a geometric view for the set of all possible states of a system in the six-dimensional position-velocity space. In his analysis on the motion of the small body P in the circular restricted three-body problem (Section 3.3), Poincaré introduced the concept of organizing similar trajectories into ‘manifolds’; such a manifold can be any smooth surface. In Section 3.10 we have found that a family of planar periodic Lyapunov orbits exists about the collinear libration points. Since the dynamics at these Lagrange points is unstable (Section 3.9), this implies that the Lyapunov periodic orbits about these points are also unstable. Poincaré noted that if the (unstable) motion of the third body P is periodic, it generates special surfaces in the six-dimensional position-velocity space, that the trajectories lie on. These are the so-called *invariant manifolds*, which means that a body P with negligible mass that starts on a particular surface will remain on that surface forever unless a force is applied to make it to leave that surface. These manifolds appear as two-dimensional tube-shaped surfaces when projected onto the three-dimensional position space. So, if we plot the path of a spacecraft drifting out of a particular Lyapunov orbit with a specific orbital energy (Jacobi’s constant C) around the Sun-Earth L_2 point, for example, we will see it slowly unwind into a spiral wrapped along the surface of the tube corresponding to that energy level. This tube is called the *unstable manifold* of the initial orbit. Furthermore, another manifold contains all the paths with the same orbital energy that wind onto the original orbit; that manifold is called the *stable manifold* (Figure 3.16). It should be noted, however, that the ‘stable’ manifold actually refers to an ‘unstable’ Lyapunov orbit! If, for instance, a spacecraft moves on a stable manifold near the Earth, it can travel all the way to L_2 and go into orbit about it for no ΔV , thus saving substantial propellant. An indication of the existence of these families of trajectories was already given by (3.110), which describes the linearized motion of a body with negligible mass about a collinear libration point.

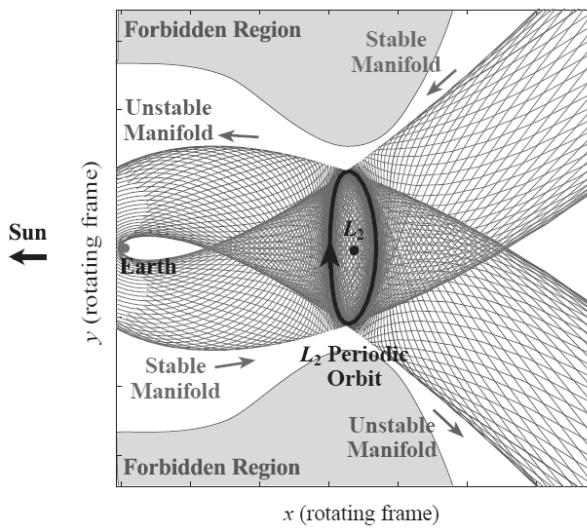


Figure 3.16: A periodic Lyapunov orbit about the Sun-Earth L_2 point and its associated stable and unstable invariant manifolds in position space. The two manifolds ‘twist’ as they wrap around the region near the Earth. [copied from: W.S. Koon et al., *Dynamical Systems and Space Mission Design*, presentation at California Institute of Technology, Pasadena, June 2000.]

In the late 1960s, C.C. Conley (1933-1984) and R. McGehee (-) extended the work of Poincaré

on manifolds and analyzed in detail the orbits winding onto and off a quasi-periodic three-dimensional halo or Lissajous orbit about the Sun-Earth L_1 point. In 1968, Conley constructed a low-energy lunar transfer trajectory based on the dynamics of the three-body problem. Around the same time, R.W. Farquhar (-) and A.A. Kamel (-) constructed large halo orbits about the lunar L_1 and L_2 points using a series-expansion technique. In the 1980s, various groups in the USA and Spain reintroduced Poincaré's tube theory for spaceflight applications. The central question was: "Is it possible to ride the stable and unstable manifolds to travel in a propellant-efficient way between the L_1 and L_2 points of a single three-body system or of different three-body systems?". In the latter case, a four-body system, consisting e.g. of Sun, Earth, Moon, and spacecraft, is considered as a superposition of two three-body systems, in this case: Sun-Earth-spacecraft, and Earth-Moon-spacecraft, where for one of the systems the unstable manifold is considered and for the other the stable manifold. This, of course, requires that these unstable and stable manifolds of equal energy intersect. However, even if only manifolds of slightly different energies intersect, this would be useful as the energy difference can be bridged by executing a small rocket engine impulse.

It is important to note that a transport between the L_1 and L_2 points requires that the spacecraft initially just misses the manifold. If a spacecraft rides the manifold in, it will be trapped in a Lissajous orbit around the Lagrange point, but if the spacecraft is maneuvered to a point that lies inside the tube, it will plunge towards the planet. What happens next gets very complicated—chaotic, in fact—but the spacecraft can emerge at either Lagrange point, or can wind up in orbit around the planet. Figure 3.17 (right) shows the four types of trajectories that exist for the Sun-Earth system in the region around L_2 : 1) an unstable periodic orbit (oval); 2) two cylinders of asymptotic trajectories that wind on to or off this periodic orbit; they form pieces of the stable and unstable manifolds; 3) transit trajectories which the spacecraft can use to make a transition from one realm to the other; for example, passing from the exterior region (interplanetary space) into the inner space surrounding the Earth via the neck region; 4) non-transit trajectories where the spacecraft bounces back to its original realm. So, the tubes act as separatrices for the trajectories through the L_2 region: those inside the tubes are transit orbits and

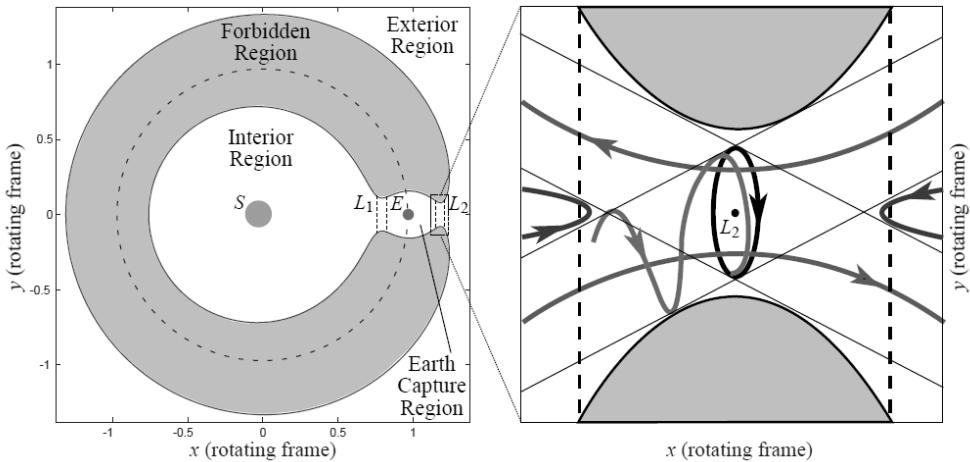


Figure 3.17: (left) Projection of the three-dimensional surfaces of Hill (schematic) on the XY-plane for the Sun-Earth system and for a value of C for which 'necks' exist about the L_1 and L_2 points. (right) Motion in the region near L_2 projected on the XY-plane, showing a bounded orbit about L_2 , an asymptotic orbit winding onto this bounded orbit, two transit trajectories, and two non-transit trajectories (schematic). A similar sketch holds for the region around L_1 . [copied from: G. Gómez et al., Connecting orbits and invariant manifolds in the spatial restricted three-body problem, Nonlinearity, Vol. 17, 2004.]

those outside the tubes are non-transit orbits.

During the last decades, many scientists have worked on the application of the stable and unstable manifolds of three-body systems for designing low-energy (propellant-efficient) spacecraft trajectories. In 1990, the Japanese Hiten mission was sent to the Moon through the implementation of a low-energy transfer, constructed by Belbruno using his weak stability boundary theory. In 2000, W.S. Koon (-) et al. developed one of the first algorithms that could be used to reproduce a Hiten-like trajectory using invariant manifold theory, although it was limited to planar transfers. In 2005, J.S. Parker (-) and M.W. Lo (-) developed a method to construct three-dimensional low-energy lunar transfers using invariant manifold theory. Nowadays, for the design of efficient transfer trajectories to the Moon and the planets, or within planetary satellite systems, extensive use is made of the concepts of weak stability boundaries and invariant manifold tubes.

We should realize that any three-body system has this kind of tube systems. The tubes are global objects and extend far beyond the vicinity of the L_1 and L_2 points. In case the manifolds of two three-body systems intersect in position-velocity phase space, a ballistic transit from a manifold to the other is easily achievable, whereas if intersections only occur in position space multiple burns of a chemical propulsion system or low-thrust arcs by an electric propulsion system are mandatory. Fortunately, the manifolds for the Sun-Earth and for the Earth-Moon systems intersect in phase space, making a propellant-efficient transfer possible. At the intersection only a (very) small trajectory correction has to be executed by rocket thrusters to let the spacecraft move from the Sun-Earth manifold to the Earth-Moon manifold. For interplanetary missions we can imagine a large number of Sun-planet and planet-moon tube systems that rotate with various angular velocities in inertial space. Sometimes, they intersect but it may take up to thousands of years before they do that again. When a series of these tube systems intersect, one can visualize the motion of the spacecraft as a travel along an interplanetary network of tubes. During that travel no significant ΔV maneuver is required, except every now and then a small rocket impulse to ‘jump’ to another manifold tube. That network of tubes in space is sometimes referred to as the system of *interplanetary superhighways*. It is noted, however, that no intersections between Sun-planet manifolds for the inner planets of the solar system exist. A disadvantage of this kind of trajectories is that the overall flight time to the Moon, the target planet or a moon of that target planet may be significantly longer than for a Hohmann transfer, which is the minimum-energy transfer in a two-body system (Section 12.1).

The manifold tubes also are an important concept for analyzing the dynamics of asteroids and comets. In the classical picture of the solar system, it is visualized as a series of planets isolated in stately, concentric, nearly circular orbits. In that picture, it is surprising that asteroids and comets periodically intrude the inner solar system. But the invariant manifolds theory provides an easy way to explain that. As a tube sweeps through the outer reaches of our solar system, every now and then some debris will fall into it and will be whisked in toward the center. Referring back to the schematic diagram in Figure 3.17 (right), we notice that the transit trajectories pass through the oval of the periodic orbit about L_2 . This e.g. explains why Jupiter comets always seem to pass by the Sun-Jupiter L_1 and L_2 points. In fact, it was noticed that the shattered comet Shoemaker-Levy 9, which consisted of 21 major fragments, passed by the Sun-Jupiter L_2 point before the fragments crashed into Jupiter from July 16 through 22, 1994 (Section 18.11).

It is interesting to note that while the concepts of weak stability boundaries and invariant manifolds have been developed independently, they are, in fact, interrelated. Low-energy lunar

trajectories were originally found by applying the weak stability boundary concept via a trial-and-error approach, before the concept of tube dynamics in the system was known. However, weak stability boundary theory may be explained heuristically using invariant manifold theory, although the relevant algorithms do not require any knowledge or computation of invariant manifolds. In fact, the Lissajous orbits about the Earth-Moon L_1 and L_2 points together with the manifold pathways that lead to them intersect in an infinitely complex manner in a region about the Moon, forming a weak stability boundary.

3.13. Phenomena at and use of libration points

In astronomy, there are certain phenomena that are closely connected to the existence of libration points and in this Section we will mention a few.

We know that for the Earth-Moon system the L_4 and L_5 points provide a stable equilibrium ($\mu = 0.01215 < 0.0385$). This means that material may be trapped near these points. Indeed, near these points clouds of particles have been observed. These clouds can only be seen with sensitive equipment at night, because the matter concentrated in these clouds reflects sunlight.

Also in the L_2 point of the Sun-Earth system, which is observable from the dark side (night side) of the Earth, there appears to be a concentration of material. With sensitive equipment, and under very favorable conditions also with the naked eye, a faint glow diametrically opposite to the Sun can be seen in the sky at night, called the *Gegenschein*. Because the L_2 point in the Sun-Earth system is located in the penumbra region of the Earth (Section 20.4), in the older literature this light phenomenon is explained as the reflection of sunlight from material that is temporarily trapped at the L_2 point. Although the equilibrium at this point is unstable, there might be a concentration of material at this point if the material density in space is sufficiently high. During the last Apollo flights, pictures were taken of this phenomenon against the background of the stars when the spacecraft was far from the Sun-Earth line. From these pictures it was possible to determine the distance of the region where the phenomenon is produced through triangulation techniques. It was found that the region is much farther away from the Earth than the L_2 point. Since the Gegenschein was still present in the observations by the Pioneer 10 spacecraft out to 1.86 AU, it is now widely accepted that the Gegenschein is the enhancement of scattered sunlight in the backward direction by interplanetary dust grains.

The famous and classical proof of the theory of motion about the L_4 and L_5 libration points is the existence of the Trojan asteroids; a class of asteroids moving in the orbit of Jupiter and oscillating about the L_4 and L_5 points of the Sun-Jupiter system. These were predicted by Lagrange in 1772. Observational verification came in 1906 when the first of the Trojan group, 588 Achilles, was discovered by M.F.J.C. Wolf (1863-1932) near the equilateral libration point L_4 of the Sun-Jupiter system. A total of 4,076 Jupiter Trojans have been identified near the L_4 and L_5 point as of February 2010. The total number of Jupiter Trojans larger than 1 km in diameter is believed to be about 1 million. The names given to the larger Trojans are Greek or Trojan and are chosen from the Iliad. As in the mythology, the Greeks are ahead of Jupiter and the Trojans are behind Jupiter, with the exception of Hector and Patroclus: two unfortunate warriors who are surrounded by their enemies and more than one billion kilometers away from their friends! The average period of their libration motion is about 150 years (compare the values listed in Table 3.3). The amplitude of the libration motion, as seen from the Sun, varies from 0.6° to 88° , with the average being about 33° . The Voyager-1 and -2 spacecraft have discovered two small asteroids in the Saturnian system close to an equilateral libration point of the Saturn-Thetys system and another two asteroids at an equilateral libration point of the system Saturn-Dione. Since then more asteroids in the L_4 and L_5 points of other systems have been found, and these are

generally also referred to as Trojans. By mid 2011, seven Trojans were identified in the Sun-Neptune system, four in the Sun-Mars system, and one in the Sun-Earth system. The latter Trojan is moving about the Sun-Earth L_4 point, leading the Earth in its motion about the Sun. That no Trojans have yet been found in the Sun-Saturn system is most likely the result of the perturbing forces produced by Jupiter's gravity field which have removed them from their orbits about the libration points.

The libration points also provide important applications for space missions. In 1950, A.C. Clark (1917-2008) suggested that the L_2 point of the Earth-Moon system would be an ideal position for a spacecraft to broadcast radio and TV signals to colonies on the backside of the Moon. Of course, that satellite would have to fly a trajectory about this point, because at the L_2 point the spacecraft would be invisible from Earth. In 1966, Farquhar proposed the Lissajous path, mentioned in Section 3.10, around this point to keep maneuvering costs low and allowing visibility from Earth most of the time. To allow continuous communication with Earth, periodic out-of-plane maneuvers would be needed. A data-relay satellite that would fly such a path was considered for the Apollo 17 mission, when that mission was planned to be the first manned landing on the far side of the Moon. But that idea was dropped when the Apollo program was shortened and Apollo 17 was redefined as a near-side mission. In 1973, Farquhar and Kamel discovered that when the in-plane oscillation about the Earth-Moon L_2 point is larger than 32,380 km, the out-of-plane oscillation has the same period, producing a halo orbit (Section 3.10). With the end of the Apollo program, interest in lunar missions waned until they returned high on the agenda of the space agencies around 2004.

As early as 1964, Farquhar recognized that the Sun-Earth L_1 point would be an ideal location to continuously monitor Earth-Sun interactions, in a first step of what now is known as 'space weather' monitoring. In 1972, NASA decided to include a spacecraft near the Sun-Earth L_1 point in a three-spacecraft program that became known as the International Sun-Earth Explorer (ISEE) program. ISEE-1 and ISEE-2 would stay in a highly-elliptical Earth orbit with an apogee distance of about 24 Earth radii. ISEE-3 would be located in a halo orbit about the Sun-Earth L_1 point to monitor high-energy particles ejected by the Sun about one hour before they reached the Earth's magnetosphere, and ISEE-1 and 2. A relatively small-amplitude Lissajous path was ruled out, because of the resulting frequent crossings of the solar radio interference exclusion zone: a 3° radius centered at the Sun (as seen from the Earth) where communication would be difficult or impossible, because the radio signals transmitted by the spacecraft would be buried in the

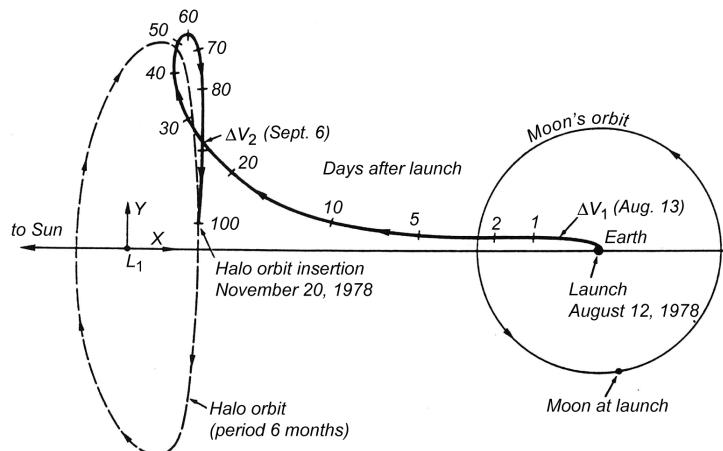


Figure 3.18: Transfer trajectory of the ISEE-3 spacecraft to its halo orbit.

background solar radio noise and therefore can hardly be detected by radio telescopes on Earth. The Z-amplitude selected for ISEE-3's orbit was 120,000 km, subtending 4.5° at the distance of the Sun-Earth L_1 point; the corresponding Y-amplitude was 666,670 km. The spacecraft was launched by a Delta 2000 rocket on August 12, 1978; its 100 days transfer trajectory is shown in Figure 3.18, which is a rotating ecliptic-plane view with the Sun-Earth line fixed (horizontal). The design of the trajectory was largely based on dynamical systems theory and the application of the characteristics of invariant manifolds (Section 3.12). Three ΔV maneuvers, totaling 57 m/s, were applied to remove launch injection errors and to insert the spacecraft into the halo orbit on November 20, 1978. It was the first spacecraft to enter this type of orbit. During the 3.5 years it remained in this orbit (Section 18.12), less than 10 m/s per year was needed to maintain the orbit.

In recent years, Lissajous and halo orbits around the Sun-Earth L_1 and L_2 points have become a popular concept for scientific space missions. The SOHO (1995, halo), ACE (1997, Lissajous), and Genesis (2001, halo) spacecraft have orbited the Sun-Earth L_1 point, while the MAP (2001, Lissajous) spacecraft was the first to orbit the Sun-Earth L_2 point. The LISA Pathfinder spacecraft, which will pave the way for the LISA gravitational waves detection mission by testing in flight the measurement concept, will be placed in a halo orbit about the Sun-Earth L_1 point in 2015. In particular for advanced astrophysical missions, the Sun-Earth L_2 point is very attractive. When the satellite carries cryogenic sensors it needs a cold and stable environment. Then, orbits about the Earth are less suitable, because of the radiation by the Earth and the Moon and the thermal cycling from frequent encounters with the Earth's shadow cone. In Lissajous or halo orbits about the Sun-Earth L_2 point these problems do not occur; these orbits provide a constant geometry for observation with half the celestial sphere available at all times, and a nearly constant communication range of roughly 1.5×10^6 km. The Sun, Earth and Moon are always 'behind' the spacecraft, thereby providing a stable observation environment, making observation planning much simpler, and allowing the application of a single heatshield to isolate the spacecraft from the heat inputs from Sun, Earth and Moon. For a space interferometer mission, involving a number of spacecraft flying in formation, a halo orbit about the L_2 point is also very attractive. In the region around an L point the gradient of the force field is very small (Section 3.10), which certainly helps to achieve the extreme relative position accuracies of the spacecraft

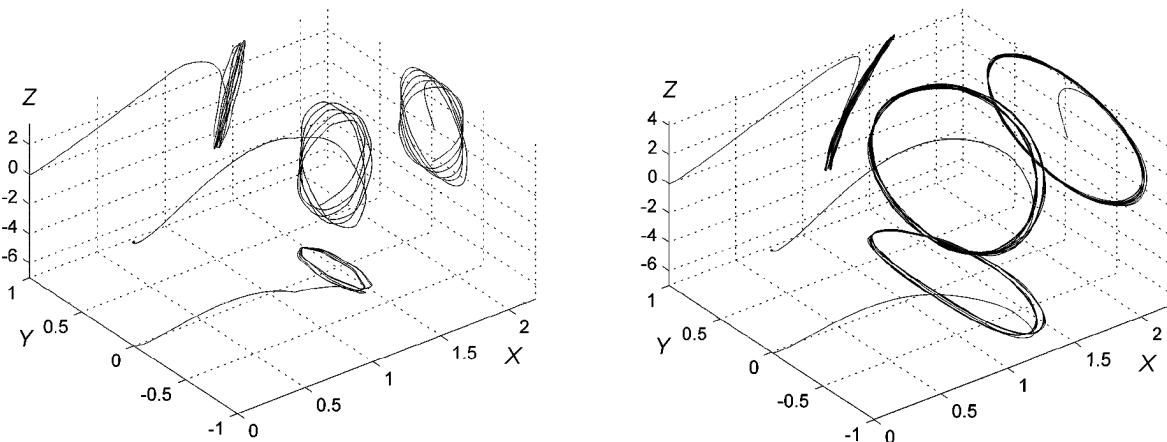


Figure 3.19: Three-dimensional picture of the transfer trajectory to and Lissajous orbit about the Sun-Earth L_2 point of the Planck spacecraft (left) and of the transfer trajectory to and quasi-halo orbit of the Herschel spacecraft about this L point (right). The unit of scale in the X- and Y-direction is 10^6 km; in the Z-direction 10^5 km. [copied from: M. Hechler, Orbiting L2, Herschel-Planck Mission Analysis, ESOC, March 2009.]

required for space interferometry. On May 14, 2009, an Ariane 5 ECA launcher injected the ESA Herschel and Planck spacecraft into two different trajectories to the Sun-Earth L_2 point. On July 3, 2009, Planck was injected into a medium-amplitude Lissajous orbit about the L_2 point, with the Sun-spacecraft-Earth angle limited to 15° . On July 13, 2009, Herschel slipped into a large-amplitude Lissajous, quasi-halo, orbit about this point, with a maximum Sun-spacecraft-Earth angle of 40° . Figure 3.19 shows a three-dimensional picture of the transfer trajectory to and the medium-amplitude Lissajous orbit (2.5 yr propagation) about the L_2 point of the Planck spacecraft (left) and of the transfer trajectory to and the quasi-halo orbit (4 yr propagation) of the Herschel spacecraft about this L point (right). In addition, orthogonal projections of the entire trajectories onto the XY -, XZ - and YZ -plane are shown. A number of other astronomy missions are already flying or are planned to fly in (quasi-)halo or Lissajous orbits about the Sun-Earth L_2 point.

An interesting feature of some large-amplitude Lissajous orbits, which are close to halo orbits, about the Sun-Earth L_2 point is that they can be reached from a low-altitude Earth parking orbit by a single injection into a transfer trajectory. This injection requires an impulsive shot of about 3.2 km/s. No further insertion maneuver is needed for these Lissajous orbits. The inclination of the initial parking orbit should be selected such that the transfer trajectory has an out-of-ecliptic component in order to induce an out-of-plane amplitude of the orbit above 120,000 km, below which no halo-like orbits exist. The distance of the L_2 point is so large that the spacecraft is always outside the Earth's shadow cone (umbra), but it may periodically enter the Earth's penumbra (Section 20.4). This can be avoided by performing a relatively small maneuver of 15 m/s in order to jump to another orbit, where the in- and out-of-plane motion have the same amplitudes but a different relative phase. This results in a continuous and constant illumination by the Sun, which is ideal for the generation of electrical power. A disadvantage of free-insertion Lissajous orbits is that they have relatively large amplitudes above 600,000 km and thus there can be large angular separations between the positions of the Sun and the Earth as seen from the spacecraft. This leads to a more-complex and larger thermal shielding system. The solution to this problem is to invest some ΔV for an insertion maneuver that puts the spacecraft on a smaller-amplitude Lissajous orbit. However, on such Lissajous orbits it cannot be avoided that the spacecraft comes close to the Earth's penumbra. This requires a periodical adjustment of the libration orbit.

In the future, man will undoubtedly build lunar bases. As was mentioned before, a spacecraft in an orbit about the Earth-Moon L_2 point can be very attractive for a lunar base at the far side of the Moon. That spacecraft will, in general, describe a three-dimensional Lissajous curve. However, because the frequency difference between the periodic motion in the XY -plane and the Z -axis oscillation is small (Table 3.2), the trajectory can be viewed as a slowly-changing elliptical path. For communication purposes, the spacecraft should perform oscillations that are large enough to make the spacecraft visible from any point on the part of the Earth that faces the Moon. Figure 3.20 shows an example of such an uncontrolled trajectory, as seen from the Earth. Unfortunately, the spacecraft will periodically enter the lunar occultation zone. To avoid that situation, the oscillation can be adverted by using a rocket thrusters control technique. In this scheme, a single-axis control is used to synchronize the fundamental Y -axis and Z -axis oscillations. That control will produce a closed elliptical path in the YZ -plane that always avoids the occultation zone, and a continuous and uninterrupted communication link between the Earth and the far side of the Moon can be established.

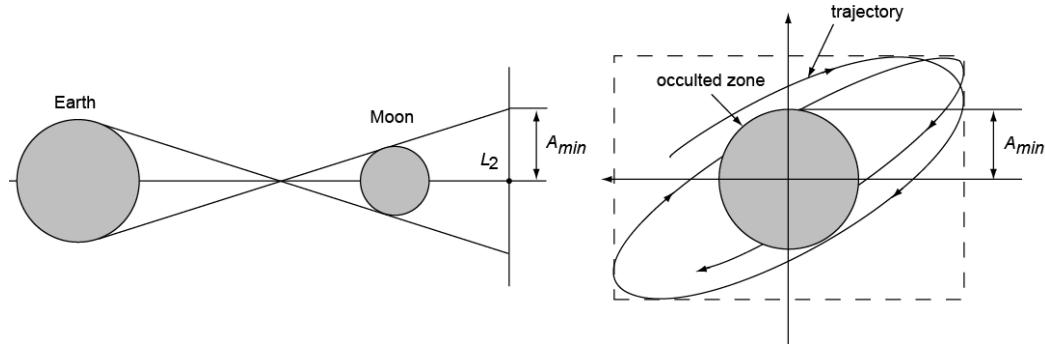


Figure 3.20: An uncontrolled trajectory about the L_2 point in the Earth-Moon system. The Figure shows the occultation geometry (left, not to scale), and the trajectory as viewed from the Earth (right).

As was mentioned in Section 3.10, in the Apollo project the combined CSM/LM was inserted into a low-altitude lunar orbit, from which the LM with two astronauts descended to the surface of the Moon. An alternative method would use the Earth-Moon L_1 point for the parking position of the CSM/LM, instead of the lunar parking orbit. The LM would then descend from the L_1 point to the lunar surface and would have to return to this L_1 point again after the visit to the Moon. It can be shown that this L_1 point rendez-vous concept requires less propellant than the lunar parking concept for landing sites at higher lunar latitudes. Because the L_1 point is stationary with respect to the lunar surface, the libration point rendez-vous technique has the important operational advantage of an infinite launch window of the lunar lander to and from the lunar surface. This timing advantage also makes the L_1 point an ideal location for a lunar logistics staging depot or for a depot of lunar materials that have to be transported to a space station in orbit about the Earth.

A modified version of this libration point rendez-vous technique can be used for future missions to the outer planets. One of the options is to operate an interplanetary shuttle vehicle between the L_1 point of the Sun-Earth system and the L_2 point of a Sun-planet system. That vehicle would use a combination of a high-thrust and a low-thrust (Chapter 19) propulsion system. The transfer is initiated by applying a small impulse (high thrust) at the Sun-Earth L_1 point and then starting the low-thrust engine. As the spacecraft passes close to the Earth, a much larger impulse (high thrust) is applied. During the heliocentric part of the transfer, only low thrust is applied. At the closest approach to the target planet, another large impulse (high thrust) is applied, and the planetary landing vehicle is separated from the interplanetary shuttle and lands on the planet. The shuttle then proceeds to the Sun-planet L_2 point, where capture is effected by another small impulse (high thrust). A reverse procedure is used for the inbound transfer back to Earth. From the total propellant consumption standpoint, this libration point rendez-vous technique probably does not have any significant advantage over the classical interplanetary flights (Chapter 18). However, the increased flexibility in the timing of the various operations (e.g. rendez-vous, abort, and landing) may justify the application of this concept.

4. RELATIVE MOTION IN THE MANY-BODY PROBLEM

The previous Chapters dealt with the motion of bodies with respect to a (quasi-)inertial reference frame. However, in many practical cases one will hardly be interested in the motion of a body with respect to the center of mass of a system of n bodies, but one wants to know the motion with respect to one of the other bodies. For example, it is not that interesting to know the orbit of an Earth satellite about the center of mass of the solar system, but it is much more interesting to know its orbit about the Earth. Then, we have to describe the motion of the satellite relative to a non-rotating reference frame with its origin at the center of the Earth. However, such a reference frame experiences translational accelerations and thus is not an inertial one. Consequently, we cannot use (2.3) or (2.21), which describe the motion of body i relative to an inertial reference frame. In this Chapter, we will therefore derive expressions for the motion of body i relative to one of the other bodies; that body will be referred to as body k .

4.1. Equations of motion

When we consider a system of n bodies, where all bodies attract each other according to Newton's law of gravitation, we can write for the motion of bodies i and k with respect to a non-rotating reference frame XYZ (Figure 4.1), with its origin at the center of mass of the n -body system (inertial reference frame), according to (2.3):

$$\begin{aligned} m_i \frac{d^2 \bar{r}_i}{dt^2} &= \sum_{j \neq i} G \frac{\mathbf{m}_i \mathbf{m}_j}{\bar{r}_{ij}^3} \bar{r}_{ij} \\ m_k \frac{d^2 \bar{r}_k}{dt^2} &= \sum_{j \neq k} G \frac{\mathbf{m}_k \mathbf{m}_j}{\bar{r}_{kj}^3} \bar{r}_{kj} \end{aligned} \quad (4.1)$$

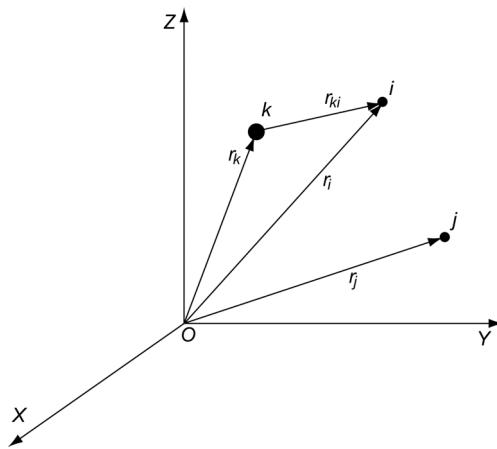


Figure 4.1: Geometry of the bodies i , j and k relative to an inertial reference frame XYZ .

or

$$\frac{d^2 \bar{r}_i}{dt^2} = G \frac{\mathbf{m}_k}{\bar{r}_{ik}^3} \bar{r}_{ik} + \sum_{j \neq i, k} G \frac{\mathbf{m}_j}{\bar{r}_{ij}^3} \bar{r}_{ij} \quad (4.2-1)$$

$$\frac{d^2\bar{r}_k}{dt^2} = G \frac{\mathbf{m}_i}{r_{ki}^3} \bar{r}_{ki} + \sum_{j \neq i, k} G \frac{\mathbf{m}_j}{r_{kj}^3} \bar{r}_{kj} \quad (4.2-2)$$

Of course, the following identities hold:

$$\bar{r}_{ik} = -\bar{r}_{ki} ; \quad \bar{r}_{ki} = \bar{r}_i - \bar{r}_k ; \quad \bar{r}_{ij} = \bar{r}_j - \bar{r}_i = \bar{r}_{kj} - \bar{r}_{ki} \quad (4.3)$$

Subtracting (4.2-2) from (4.2-1) and using (4.3) we arrive at

$$\frac{d^2\bar{r}_{ki}}{dt^2} = -G \frac{\mathbf{m}_i + \mathbf{m}_k}{r_{ki}^3} \bar{r}_{ki} + G \sum_{j \neq i, k} \mathbf{m}_j \left(\frac{\bar{r}_{kj} - \bar{r}_{ki}}{r_{ij}^3} - \frac{\bar{r}_{kj}}{r_{kj}^3} \right)$$

All vectors in this equation originate from body k . When we consider the motion of body i with respect to a non-rotating reference frame fixed to body k , the index k may be omitted and the equation can be written as

$$\frac{d^2\bar{r}_i}{dt^2} = -G \frac{\mathbf{m}_i + \mathbf{m}_k}{r_i^3} \bar{r}_i + G \sum_{j \neq i, k} \mathbf{m}_j \left(\frac{\bar{r}_j - \bar{r}_i}{r_{ij}^3} - \frac{\bar{r}_j}{r_j^3} \right) \quad (4.4)$$

This equation can be integrated numerically to find the trajectory of body i relative to the non-rotating (non-inertial) reference frame with its origin at body k . In Section 2.2 we have derived equation (2.21) that describes the motion of body i relative to a non-rotating (inertial) reference frame with its origin at the barycenter of the system. Note that expressions (2.21) and (4.4) both contain a two-body term and a term expressing the effects of all other bodies. In general, only the second term has to be integrated numerically as the two-body term leads to an orbit that can be computed in a closed analytical way (Section 5.2). When we only look at the efficiency and accuracy of the numerical integration process, the choice whether (4.4) or (2.21) should be used in a particular numerical situation depends upon the relative magnitude of the second term on the right-hand side of both equations. We should preferably use the formulation where the ratio of the magnitude of the second term to the magnitude of the first term is smallest. The effectiveness of the barycentric form for a planet superior to Jupiter was first demonstrated by S.B. Nicholson (1891-1963) and N.U. Mayall (1906-1993) in 1931, when a computed unperturbed barycentric orbit of Pluto gave results nearly as good as a computed heliocentric orbit including perturbations. In that same year, E.C. Bower (1890-1964) numerically integrated a perturbed barycentric orbit of Pluto, finding that the smoothness of the barycentric motion enabled a lengthening of the integration interval to 320 days.

Nowadays, there are also other reasons to compute the orbits of planets and spacecraft relative to the barycentric reference frame. The original approaches to construct reference frames in astronomy were completely based on the concepts of Newtonian gravitation and Euclidean (Euclid; ~300 B.C.) absolute space and time. Modern astronomy, astrodynamics and navigation, however, are based on very accurate geometric and dynamical models, timing systems and measurements (Sections 11.2 and 11.4). At this level, the primary gravitational theory must be Einstein's theory (A. Einstein; 1879-1955) of general relativity with a corresponding replacement of the Euclidean space and time by the four-dimensional Riemannian (G.F.B. Riemann; 1826-1866) space-time manifold. In general relativity theory the concept of inertial reference frames loses its meaning. In astronomy, official transition from Newtonian to relativistic concepts commenced in 1991, when a few recommendations were adopted by the International Astronomical Union (IAU). Nowadays, it is widely accepted that in order to fully exploit the

modern high-precision astronomical observations and to compute very accurate orbits of celestial bodies and spacecraft, one has to use several relativistic reference systems. The solar system *Barycentric Celestial Reference System* (BCRS), in combination with the relativistic *Barycentric Dynamical Time* (*Temps Dynamique Barycentrique*, TDB) time scale, can be used to model light propagation from distant celestial objects and thus to describe accurately the angular positions of these objects, as well as to model the motion of bodies within the solar system. The *Geocentric Celestial Reference System* is physically adequate to describe processes occurring in the vicinity of the Earth (Earth rotation, motion of Earth satellites, etc.). Further local reference systems centered at the Moon, Mars, Jupiter, etc. are defined for specific applications, where physical phenomena in the vicinity of the corresponding body play a role. In the vicinity of an observer one may construct an observer-centered local reference system and use that for modeling phenomena in the neighborhood of the observer. Coordinate transformations between BCRS and the other relativistic reference systems require complicated four-dimensional space-time transformations that also contain acceleration terms and gravitational potentials. In this book we base our analyses on Newtonian mechanics and will not apply these advanced concepts for the analysis of orbits.

We now return to our analysis of the motion of body i relative to a non-rotating reference frame centered at body k . In celestial mechanics it is common practice to express the gravitational influence of bodies through potential functions. To this end, we start with the expression for the gravitational potential of body k at the position of body i (Section 1.4):

$$U_i = -G \frac{m_k}{r_i}$$

Now, a slightly modified function, defined as

$$\hat{U}_i = -G \frac{m_k + m_i}{r_i} \quad (4.5)$$

and a scalar function, defined as

$$R_i = -G \sum_{j \neq i, k} m_j \left(\frac{1}{r_{ij}} - \frac{\bar{r}_i \cdot \bar{r}_j}{r_j^3} \right) = -G \sum_{j \neq i, k} m_j \left(\frac{1}{r_{ij}} - \frac{x_i x_j + y_i y_j + z_i z_j}{r_j^3} \right) \quad (4.6)$$

are introduced. With (4.5) and (4.6), and the expressions

$$r_i = (x_i^2 + y_i^2 + z_i^2)^{1/2} \quad ; \quad r_{ij} = \{(x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2\}^{1/2}$$

we can write for the partial derivatives of \hat{U}_i and R_i to x_i :

$$\frac{\partial \hat{U}_i}{\partial x_i} = G(m_k + m_i) \frac{x_i}{r_i^3} \quad ; \quad \frac{\partial R_i}{\partial x_i} = -G \sum_{j \neq i, k} m_j \left(\frac{x_j - x_i}{r_{ij}^3} - \frac{x_j}{r_j^3} \right)$$

Similar expressions can be derived for the partial derivatives of \hat{U}_i and R_i to y and z . Substitution of these expressions into (4.4) yields

$$\frac{d^2 \bar{r}_i}{dt^2} = -\bar{\nabla}_i (\hat{U}_i + R_i) \quad (4.7)$$

Analogous to the discussion in Section 2.1, we conclude that body i moves in a force field that is described by the potential $\hat{U}_i + \mathbf{R}_i$. This force field is clearly non-central and non-conservative. The value of the potential \hat{U}_i is only determined by bodies k and i and is therefore called the *primary potential*; R_i expresses the influence of the perturbing bodies j and is called the *perturbing potential*. This perturbing potential plays a major role in the analysis of the perturbations of satellite orbits due to the gravitational attraction by the Sun, Moon and planets.

Returning to (4.4), we note that the influence of the bodies j on the motion of body i with respect to body k is expressed by the term

$$G \sum_{j \neq i, k} m_j \left(\frac{\bar{\mathbf{r}}_j - \bar{\mathbf{r}}_i}{r_{ij}^3} - \frac{\bar{\mathbf{r}}_j}{r_j^3} \right) \quad (4.8)$$

The first part of this term expresses the acceleration of body i as a result of the gravitational attraction between body i and body j ; the second part expresses the acceleration of body k , the origin of the reference frame, as a result of the gravitational attraction between bodies k and j . Therefore, the first part is generally called the *principal part* and the second part the *indirect part*. To illustrate the physical meaning of (4.8), we consider the Sun-Earth-Moon system. Using the numerical data presented in Appendix B, we find for the magnitude of the gravitational forces between the Sun (index S) and the Moon (index M), and between the Earth (index E) and the Moon

$$F_{S-M} = G \frac{m_S m_M}{r_{SM}^2} = 4.4 * 10^{20} \text{ N} \quad ; \quad F_{E-M} = G \frac{m_E m_M}{r_{EM}^2} = 2.0 * 10^{20} \text{ N}$$

So, the force between the Sun and the Moon is about twice the force between the Earth and the Moon. This seems to contradict the fact that the Moon keeps orbiting the Earth! The explanation for this apparent contradiction is given by (4.8). This equation shows that the influence of the Sun on the motion of the Moon about the Earth is not determined by the pure gravitational attraction between the Sun and the Moon, but by the *difference* between the acceleration of the Moon due to the gravitational attraction by the Sun and the acceleration of the Earth due to the gravitational attraction by the Sun. Because this difference is small in comparison to the acceleration of the Moon due to the gravitational attraction by the Earth, the gravitational attraction by the Sun produces only a perturbation on the motion of the Moon about the Earth.

We will use (4.4) as the starting point for the analysis of the acceleration of the Earth (body i) produced by the gravitational attraction by the Sun (body k) and by the ‘differential attraction’ by the Moon, one of the other planets, a near-by star or a near-by galaxy (body j); as well as for the analysis of the acceleration of a satellite (body i) produced by the gravitational attraction by the Earth (body k) and by the ‘differential attraction’ by the Sun, the Moon, one of the other planets, a near-by star or a near-by galaxy (body j). For simplicity, we will assume that the planets move in circular orbits about the Sun and that the satellite moves in a circular orbit about the Earth. It will be shown that for both cases the acceleration of body i produced by each body j is (very) small when compared to the acceleration due to the gravitational attraction between bodies i and k . Therefore, in the following we will replace the index j with d to indicate that it is a disturbing body, and we will call the first term on the right-hand side of (4.4) the *main acceleration*, a_m , of body i and the second term the *perturbing acceleration*, a_d , of body i .

4.2. Relative perturbing acceleration of the Earth and of an Earth satellite

The magnitude and direction of the perturbing acceleration of body i is, of course, determined by the relative positions of the bodies i, j and k . At any time, these relative positions determine a plane. That plane and the notation used for the analysis are shown in Figure 4.2. A qualitative analysis of (4.8) and the geometry shown in Figure 4.2 shows that, for specified values of r_i and r_j , the term $(\bar{r}_j - \bar{r}_i)/r_{ij}^3 - \bar{r}_j/r_j^3$ expresses a vector that rotates over an angle of 180° when α increases from $\alpha = 0^\circ$ to $\alpha = 180^\circ$, and of which the magnitude first decreases and later increases again but never becomes zero. According to the definition given above, the magnitude of the main acceleration of body i is

$$a_m = G \frac{m_k}{r_i^2} \quad (4.9)$$

where the mass of body i is neglected with respect to the mass of body k . This is justified by the fact that the mass of the Earth is much smaller than the mass of the Sun, and that the mass of a satellite is much smaller than the mass of the Earth.

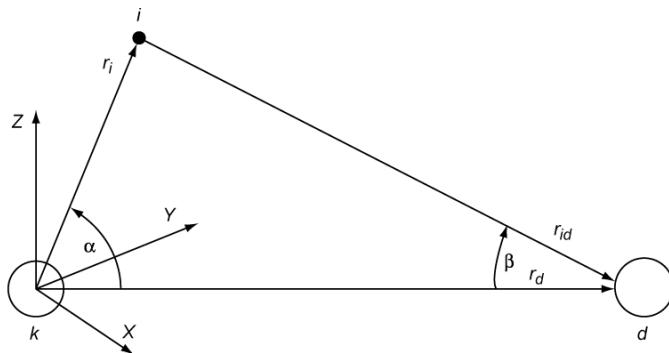


Figure 4.2: Relative positions of bodies k , i , and d .

For the magnitude of the perturbing acceleration follows from (4.4) with the present notation:

$$a_d = G m_d \sqrt{\left(\frac{\bar{r}_{id}}{r_{id}^3} - \frac{\bar{r}_d}{r_d^3} \right) \cdot \left(\frac{\bar{r}_{id}}{r_{id}^3} - \frac{\bar{r}_d}{r_d^3} \right)}$$

which gives

$$a_d = G m_d \sqrt{\frac{1}{r_{id}^4} + \frac{1}{r_d^4} - \frac{2 \cos \beta}{r_{id}^2 r_d^2}} \quad (4.10)$$

According to Figure 4.2, the following relations hold:

$$\cos \beta = \frac{r_d - r_i \cos \alpha}{r_{id}} \quad ; \quad r_{id}^2 = r_i^2 + r_d^2 - 2 r_i r_d \cos \alpha$$

With these expressions, $\cos \beta$ and r_{id} can be eliminated from (4.10) and we obtain

$$a_d = G \frac{m_d}{r_d^2} \sqrt{1 + \frac{1}{(1 - 2\gamma \cos \alpha + \gamma^2)^2} - \frac{2(1 - \gamma \cos \alpha)}{(1 - 2\gamma \cos \alpha + \gamma^2)^{3/2}}} \quad (4.11)$$

where $\gamma = r_i/r_d$. Note that this equation holds for $r_d > r_i$ and for $r_d < r_i$; the latter case only occurs when computing the perturbation of the Earth's orbit by Mercury and Venus. An analysis of the square-root function reveals that:

- The value of the function under square-root is positive for all values of $\gamma > 0$, irrespective of the value of α .
- For an arbitrary value of γ the square-root function takes a maximum value at $\alpha = 0^\circ$, which corresponds to the situation of a minimum distance between the bodies i and d . This result could be expected from a qualitative analysis of (4.8).
- For $\gamma < 1.74$ a second maximum occurs at $\alpha = 180^\circ$; for $\gamma < 0.01$ the value of that maximum is (about) equal to the value of the maximum at $\alpha = 0^\circ$. For larger values of γ the value of the maximum at $\alpha = 180^\circ$ steadily decreases relative to the value of the maximum at $\alpha = 0^\circ$.
- For $\gamma < 0.5$ the function takes a minimum value at $\alpha \approx 90^\circ, 270^\circ$, while for larger values of γ the minimum occurs at values of α that slowly shift away from 90° and 270° into the direction of 180° , until for $\gamma > 1.74$ the minimum occurs at $\alpha = 180^\circ$.
- For $\gamma < 0.1$ the minimum value of the square-root function is about equal to γ and the maximum value is about equal to 2γ .

For the maximum perturbing acceleration ($\alpha = 0^\circ$) we find from (4.11)

$$a_{d_{max}} = G \frac{m_d}{r_d^2} \left| \left(\frac{1}{1 - \gamma} \right)^2 - 1 \right| \quad (4.12)$$

Relative perturbing acceleration of the Earth

In this case, $\gamma = r_E/r_d$, where r_E and r_d are the distances of the Earth and the perturbing body from the center of the Sun, respectively. The value of γ ranges from 10^{-10} (if the perturbing body is a near-by galaxy) to 2.6 (if the perturbing body is Mercury). Combining (4.9) and (4.12), and substituting the mass of the Sun, m_S , for m_k , we find for the maximum relative perturbing acceleration

$$\left(\frac{a_d}{a_m} \right)_{max} = \frac{m_d}{m_S} \left(\frac{r_E}{r_d} \right)^2 \left| \left(\frac{1}{1 - r_E/r_d} \right)^2 - 1 \right| \quad (4.13)$$

If the perturbing body is the Moon or a planet, the term m_d/m_S is always very small; even for the largest planet (Jupiter) this term is $m_d/m_S \approx 9.5 \cdot 10^{-4}$. The product of the second and third term on the right-hand side of (4.13), which are functions of r_E/r_d , does not have to be small. For distant planets it is small: for Jupiter its magnitude is about 0.02. But e.g. for Venus the magnitude of this term is about 11.1. However, because for Venus m_d/m_S is very small ($2.4 \cdot 10^{-6}$) also for this planet the maximum relative perturbing acceleration is small.

Table 4.1 lists the maximum value of the relative perturbing accelerations by the Moon, the planets¹, a near-by star (Proxima Centauri) and a near-by galaxy (Large Magellanic Cloud) on

¹ In this book, Pluto is still considered as a planet, although the International Astronomical Union has decided on August 24, 2006, that Pluto should be considered as a 'dwarf planet' and therefore has added this body to the class of dwarf planets.

Table 4.1: Maximum relative perturbing acceleration of the Earth in its motion about the Sun due to the gravitational attraction by celestial bodies.

| Perturbing body | m_d/m_s | r_E/r_d | $(a_d/a_m)_{max}$ |
|------------------------|-----------------------|-----------------------|-----------------------|
| Moon | 3.69×10^{-8} | 9.97×10^{-1} | 5.6×10^{-3} |
| Mercury | 1.66×10^{-7} | 2.58 | 6.7×10^{-7} |
| Venus | 2.45×10^{-6} | 1.38 | 2.7×10^{-5} |
| Mars | 3.23×10^{-7} | 6.56×10^{-1} | 1.0×10^{-6} |
| Jupiter | 9.55×10^{-4} | 1.92×10^{-1} | 1.9×10^{-5} |
| Saturn | 2.86×10^{-4} | 1.05×10^{-1} | 7.8×10^{-7} |
| Uranus | 4.37×10^{-5} | 5.21×10^{-2} | 1.3×10^{-8} |
| Neptune | 5.15×10^{-5} | 3.33×10^{-2} | 4.0×10^{-9} |
| Pluto | 6.58×10^{-9} | 2.53×10^{-2} | 2.2×10^{-13} |
| Proxima Centauri | 1.23×10^{-1} | 3.75×10^{-6} | 1.3×10^{-17} |
| Large Magellanic Cloud | $\approx 10^{10}$ | $\approx 10^{-10}$ | $\approx 10^{-20}$ |

the Earth in its orbit about the Sun. The relevant data on the masses and distances of the celestial bodies were taken from Appendix B. This Table shows that the Moon produces the largest maximum relative perturbation. The planet Venus comes in second place, due to its relatively large mass in combination with its short distance from the Earth; next the planet Jupiter. Despite its short distance from the Earth, Mars comes in fourth place because of its relatively small mass. Of course, we can apply the same analysis method for the perturbing acceleration experienced by another planet. For example, when consider the planet Saturn, we find that the largest relative perturbing acceleration is produced by the planet Jupiter and amounts to 1.4×10^{-3} ; the Earth produces a relative perturbing acceleration of Saturn of 2.7×10^{-4} . Table 4.1 shows that the perturbations produced by the star Proxima Centauri (distance 4.3 light-year) and the Large Magellanic Cloud (distance 160,000 light-year) on the motion of the Earth are extremely small. In fact, this holds for all bodies of the solar system. We thus may neglect these perturbations and conclude that the solar system may be considered as an isolated n -body system. This fact is used implicitly in the analysis of planetary orbits and interplanetary trajectories (Chapter 18).

Relative perturbing acceleration of a satellite

In this case, $\gamma = r_s/r_d$, where r_s and r_d are the distances of the satellite (or the Moon) and the perturbing body from the center of the Earth, respectively. Combining (4.9) and (4.12), and substituting the mass of the Earth, m_E , for m_k , we find for the maximum relative perturbing acceleration

$$\left(\frac{a_d}{a_m} \right)_{max} = \frac{m_d}{m_E} \left(\frac{r_s}{r_d} \right)^2 \left| \left(\frac{1}{1 - r_s/r_d} \right)^2 - 1 \right| \quad (4.14)$$

Note that, in contrast to the analysis of the relative perturbation of the Earth, in this case the mass of perturbing bodies in our solar system is not always small when compared to the mass of the body at the origin of the reference frame: the Earth. For example, selecting the Sun as the perturbing body, we have $m_s/m_E \approx 3.3 \times 10^5$; for Jupiter as perturbing body, we find $m_j/m_E \approx 3.2 \times 10^2$. However, in these cases the value of r_s/r_d is (very) small, and the accelerations produced by the Sun or Jupiter are still small. Note that for a given disturbing body the relative perturbing acceleration of the satellite increases with increasing orbital altitudes.

Table 4.2 lists the maximum value of the relative perturbing acceleration of a satellite that moves in a circular orbit with a radius of 42,164 km about the Earth, as produced by other celestial bodies². Again, the relevant data on the masses and distances of the celestial bodies were taken from Appendix B. The Table shows that for this satellite, the Moon produces the largest relative perturbing acceleration; next come the Sun, Venus, Jupiter and Mars, respectively. Note that the relative perturbation produced by the Moon is about twice that produced by the Sun. Moon and the Sun produce relative perturbing accelerations that are at least a factor 10^4 larger than the perturbations by all other celestial bodies. Therefore, in first-order perturbations analysis it is often sufficient to consider only the perturbations by the Sun and the Moon. Again, it is clear that perturbations due to bodies outside the solar system are completely negligible.

Table 4.2: Maximum relative perturbing acceleration of a geostationary satellite due to the gravitational attraction by celestial bodies.

| Perturbing body | m_d/m_E | r_s/r_d | $(a_d/a_m)_{max}$ |
|------------------------|-------------------|--------------------|--------------------|
| Sun | $3.33*10^5$ | $2.82*10^{-4}$ | $1.5*10^{-5}$ |
| Moon | $1.23*10^{-2}$ | $1.10*10^{-1}$ | $3.9*10^{-5}$ |
| Mercury | $5.53*10^{-2}$ | $4.60*10^{-4}$ | $1.1*10^{-11}$ |
| Venus | $8.15*10^{-1}$ | $1.02*10^{-3}$ | $1.7*10^{-9}$ |
| Mars | $1.07*10^{-1}$ | $5.38*10^{-4}$ | $3.4*10^{-11}$ |
| Jupiter | $3.18*10^2$ | $6.71*10^{-5}$ | $1.9*10^{-10}$ |
| Saturn | $9.52*10^1$ | $3.30*10^{-5}$ | $6.8*10^{-12}$ |
| Uranus | $1.45*10^1$ | $1.55*10^{-5}$ | $1.1*10^{-13}$ |
| Neptune | $1.71*10^1$ | $9.70*10^{-6}$ | $3.1*10^{-14}$ |
| Pluto | $2.10*10^{-3}$ | $7.32*10^{-6}$ | $1.7*10^{-18}$ |
| Proxima Centauri | $4.09*10^4$ | $1.06*10^{-9}$ | $9.6*10^{-23}$ |
| Large Magellanic Cloud | $\approx 10^{15}$ | $\approx 10^{-14}$ | $\approx 10^{-25}$ |

For a satellite altitude below 36,000 km, we find for the Moon as disturbing body $r_s/r_d \leq 1.1*10^{-1}$, and for all other disturbing bodies $r_s/r_d \leq 10^{-3}$. Therefore, in this case we can apply a series expansion of (4.11) in powers of γ and find after some algebraic manipulation

$$a_d = G \frac{m_d}{r_d^2} \gamma \sqrt{1 + 3 \cos^2 \alpha + 12 \gamma \cos^3 \alpha + O(\gamma^2)}$$

The first and second term of the expression under square-root describe a contribution that is rotationally symmetric about \bar{r}_d and symmetric about a plane through m_d and perpendicular to \bar{r}_d ; the third term describes an asymmetric contribution that is the result of the gradient of the gravity field of the disturbing body. Because for all disturbing bodies, excluding the Moon, $\gamma = r_s/r_d$ is very small, we can further approximate this relation by

$$a_d \approx G m_d \frac{r_s}{r_d^3} \sqrt{1 + 3 \cos^2 \alpha} \quad (4.15)$$

For the Moon as disturbing body, this approximation is less accurate and leads to errors in the

² This is the so-called *geostationary orbit*, which will be defined in Section 6.2.

computed value of a_d , depending on the value of α , of up to 30% for a geostationary satellite and of up to 4% for a satellite at 1000 km altitude. Combining (4.9) and (4.15), and substituting the mass of the Earth, m_E , for m_k , we find for the relative perturbing acceleration of the satellite

$$\frac{a_d}{a_m} \approx \frac{m_d}{m_E} \left(\frac{r_s}{r_d} \right)^3 \sqrt{1 + 3 \cos^2 \alpha} \quad (4.16)$$

This expression shows that: 1) the relative perturbing acceleration is maximum if $\alpha = 0^\circ$ and $\alpha = 180^\circ$; 2) the minimum relative perturbing acceleration occurs at $\alpha = 90^\circ$ and $\alpha = 270^\circ$; 3) the maximum value is twice the minimum value. These results agree with the conclusions of the analysis of (4.11) given above, for $\gamma < 0.01$. It is interesting to note that this analysis can also be used to analyze the tidal phenomenon. If we assume that the entire Earth would be covered by an ocean and that we restrict ourselves to ocean tides, we may conclude from (4.16) that the gravity field of the Moon and the Sun each produce two tidal bulges of the sea surface; one on the side of the Earth that is turned towards the Sun or the Moon ($\alpha = 0^\circ$), and the other on the side that is turned away from the Sun or the Moon ($\alpha = 180^\circ$). As the Earth rotates, the tidal bulges move around the Earth. When Earth, Moon, and Sun line up, the combined gravitational effects of the Moon and the Sun reinforce each other and produce very high tides (spring tide). When the Earth-Moon and Earth-Sun lines are at right angles to each other the lowest tides (neap tide) occur. In reality, the observed tides at a certain location on Earth are affected by the distribution of the land masses, bathymetry, sea currents, winds, and other factors.

From (4.16), we conclude that the maximum relative perturbing acceleration is approximately given by

$$\left(\frac{a_d}{a_m} \right)_{\max} \approx 2 \frac{m_d}{m_E} \left(\frac{r_s}{r_d} \right)^3 \quad (4.17)$$

This relation, which can also be obtained by linearizing (4.14) for $r_s/r_d \ll 1$, shows that, for a satellite moving about the Earth at an altitude below 36,000 km, the relative perturbing acceleration increases with the cube of the distance from the center of the Earth. When we substitute the values of the relevant parameters of the Sun or the Moon listed in Appendix B into (4.17), we find that the maximum relative perturbing acceleration produced by the Moon is 2.2 times the maximum relative perturbing acceleration produced by the Sun, for any satellite orbital altitude. Note that when we compare the values listed in Table 4.2 for a geostationary satellite, we find that the ratio of the maximum relative perturbing accelerations produced by Moon and Sun is 2.6. The reason for this somewhat larger value for this high orbit is that the numerical values listed in Table 4.2 are computed from the full expression (4.14), while the value of 2.2 is obtained from the approximative relation (4.17).

When we consider the motion of the Moon about the Earth with the Sun as perturbing body, we find $\gamma \approx 2.6 \cdot 10^{-3}$. Therefore, the approximative relation (4.17) can also be applied for the motion of the Earth's natural satellite, and we find for the maximum relative perturbing acceleration caused by the Sun: $(a_d/a_m)_{\max} \approx 1.1 \cdot 10^{-2}$. Although this value is still small enough to make the influence of the Sun on the motion of the Moon about the Earth a perturbation, it shows that the perturbation of the motion of the Earth about the Sun caused by, for example, Venus or Jupiter (Table 4.1) is much smaller than the perturbation of the motion of the Moon about the Earth caused by the Sun. This is the reason why an analytical analysis of the orbit of the Moon about the Earth belongs to the most difficult problems in celestial mechanics.

From the examples treated in this Section a very important conclusion can be drawn: for the motion of the Earth about the Sun as well as for the motion of a satellite about the Earth, the influence of all other celestial bodies is (very) small. So, for first-order analyses the influence of these perturbing bodies can be neglected and the motion of the Earth about the Sun, or of a satellite about the Earth, may be considered as a pure *two-body problem*. Chapters 5 to 8 deal with such two-body motion.

4.3. Sphere of influence

In Section 4.2 we have found that for a satellite moving about the Earth, the relative perturbing acceleration due to the gravitational attraction by other celestial bodies increases with increasing distance from the center of the Earth. Now, consider the motion of an interplanetary spacecraft (Chapter 18). That motion starts close to the Earth, where the gravitational attraction by the Earth dominates the spacecraft's motion. Gradually, the spacecraft's distance from the Earth increases and the relative perturbing acceleration produced by the gravitational attraction of the Sun increases, and at large distances from the Earth the gravitational attraction by the Sun will dominate the motion of the spacecraft. For first-order analyses, one usually approximates the interplanetary trajectory of the spacecraft by a series of two-body trajectories (Chapter 18). Close to the Earth, one then describes the motion of the spacecraft relative to a non-rotating reference frame with its origin at the center of the Earth. For that part of the trajectory the gravitational attraction by the Sun can be considered as a perturbing force. Far from the Earth, one describes the motion of the spacecraft relative to a non-rotating reference frame with its origin at the center of the Sun. For that part of the trajectory the gravitational attraction by the Earth can be considered as a perturbing force. The question now is: "Within what volume of space around the Earth can we describe the motion of the spacecraft as a perturbed two-body trajectory about the Earth?" P.S. Laplace (1749-1827) has addressed a similar question around 1805 in his studies on the motion of comets that pass Jupiter at a relatively short distance. He was interested to know within what volume of space around Jupiter the motion of the comet could be described as a perturbed two-body trajectory about Jupiter. He proved that that volume is approximately a sphere about Jupiter, which is called the *sphere of influence*. That concept can also be used to answer the question formulated above concerning the motion of a spacecraft about the Earth (or any other celestial body).

In this Section, we consider the general case where three bodies are indicated by P_1 , P_2 and P_3 . We address the question whether the motion of body P_2 should be expressed with respect to a non-rotating reference frame attached to body P_1 or with respect to a non-rotating reference frame attached to body P_3 , while body P_3 or body P_1 , respectively, is considered as the perturbing body. Figure 4.3 shows the relative positions of the three bodies in the instantaneous plane through the bodies, and the notation used in the analysis. Note that at any time a plane can be constructed through the three bodies; so the geometry indicated can be used to describe the three-dimensional position of P_2 relative to the line connecting bodies P_1 and P_2 . With (4.4) and using the notation indicated in Figure 4.3, we can write for the motion of body P_2 with respect to P_1

$$\frac{d^2\bar{r}}{dt^2} + \frac{G(m_1 + m_2)}{r^3}\bar{r} = -Gm_3\left(\frac{\bar{d}}{d^3} + \frac{\bar{p}}{\rho^3}\right) \quad (4.18)$$

while the motion of P_2 with respect to P_3 is described by

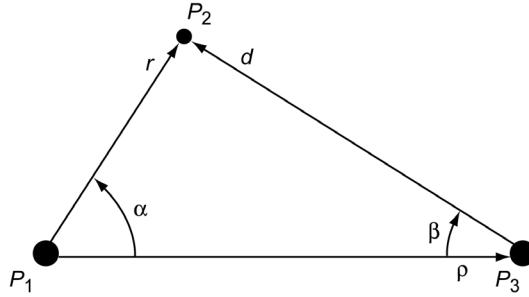


Figure 4.3: Relative positions of three bodies and the notation used in the analysis of the sphere of influence concept.

$$\frac{d^2 \bar{d}}{dt^2} + \frac{G(m_3 + m_2)}{d^3} \bar{d} = -G m_1 \left(\frac{\bar{r}}{r^3} - \frac{\bar{p}}{p^3} \right) \quad (4.19)$$

According to Laplace, the advantage of either form depends on the ratio of the perturbing force to the corresponding main force. Whichever form provides the smaller ratio is the one to be preferred.

We first consider the motion of body P_2 with respect to body P_1 . According to (4.18), the magnitude of the main acceleration is

$$a_m = G \frac{(m_1 + m_2)}{r^2}$$

and the magnitude of the perturbing acceleration is

$$a_d = G m_3 \left[\left(\frac{\bar{d}}{d^3} + \frac{\bar{p}}{p^3} \right) \cdot \left(\frac{\bar{d}}{d^3} + \frac{\bar{p}}{p^3} \right) \right]^{1/2} = G m_3 \left[\frac{1}{d^4} + \frac{1}{p^4} - \frac{2 \cos \beta}{d^2 p^2} \right]^{1/2}$$

Consequently, we find for the ratio a_d/a_m :

$$\frac{a_d}{a_m} = \frac{m_3}{m_1 + m_2} \frac{r^2}{d^2} \left[1 - 2 \frac{d^2}{p^2} \cos \beta + \frac{d^4}{p^4} \right]^{1/2} \quad (4.20)$$

According to Figure 4.3, the following relation holds:

$$\cos \beta = \frac{p - r \cos \alpha}{d} = \frac{p}{d} \left(1 - \frac{r}{p} \cos \alpha \right)$$

Substitution of this expression into (4.20) yields

$$\frac{a_d}{a_m} = \frac{m_3}{m_1 + m_2} \left(\frac{r/p}{d/p} \right)^2 \left[1 - 2 \frac{d}{p} \left(1 - \frac{r}{p} \cos \alpha \right) + \left(\frac{d}{p} \right)^4 \right]^{1/2} \quad (4.21)$$

where, according to Figure 4.3,

$$\frac{d}{p} = \left[1 - 2 \frac{r}{p} \cos \alpha + \left(\frac{r}{p} \right)^2 \right]^{1/2} \quad (4.22)$$

Now, we consider the orbit of P_2 with respect to P_3 . According to (4.19), the magnitude of

the main acceleration then is

$$a_m = \frac{G(m_3 + m_2)}{d^2}$$

and the magnitude of the perturbing acceleration is

$$a_d = G m_1 \left[\left(\frac{\bar{r}}{r^3} - \frac{\bar{\rho}}{\rho^3} \right) \cdot \left(\frac{\bar{r}}{r^3} - \frac{\bar{\rho}}{\rho^3} \right) \right]^{1/2} = G m_1 \left[\frac{1}{r^4} + \frac{1}{\rho^4} - \frac{2 \cos \alpha}{r^2 \rho^2} \right]^{1/2}$$

Consequently, we now find for the ratio a_d/a_m :

$$\frac{a_d}{a_m} = \frac{m_1}{m_3 + m_2} \left(\frac{d/\rho}{r/\rho} \right)^2 \left[1 - 2 \frac{r^2}{\rho^2} \cos \alpha + \left(\frac{r}{\rho} \right)^4 \right]^{1/2} \quad (4.23)$$

By definition, on the sphere of influence the ratios a_d/a_m found for both cases are equal. Equating (4.21) and (4.23) gives

$$\left(\frac{r}{\rho} \right)^4 = \frac{m_1(m_1 + m_2)}{m_3(m_3 + m_2)} \left(\frac{d}{\rho} \right)^4 \left[\frac{1 - 2 \left(\frac{r}{\rho} \right)^2 \cos \alpha + \left(\frac{r}{\rho} \right)^4}{1 - 2 \frac{d}{\rho} \left(1 - \frac{r}{\rho} \cos \alpha \right) + \left(\frac{d}{\rho} \right)^4} \right]^{1/2} \quad (4.24)$$

where the ratio d/ρ is given by (4.22).

We now select P_1 as the body relative to which we want to describe the motion of body P_2 , and we consider P_3 as the perturbing body. Both for the motion of an interplanetary spacecraft receding from the Earth and for the motion of a comet close to Jupiter we may assume $r \ll \rho, m_2 \ll m_1, m_2 \ll m_3$. In that case, we can obtain an approximate explicit solution for r/ρ . Evaluation of (4.22) and a series expansion of two terms in (4.24) yields

$$\begin{aligned} \left(\frac{d}{\rho} \right)^4 &= 1 - 4 \frac{r}{\rho} \cos \alpha + 2 \left(\frac{r}{\rho} \right)^2 (1 + 2 \cos^2 \alpha) - 4 \left(\frac{r}{\rho} \right)^3 \cos \alpha + \left(\frac{r}{\rho} \right)^4 \\ \left[1 - 2 \left(\frac{r}{\rho} \right)^2 \cos \alpha + \left(\frac{r}{\rho} \right)^4 \right]^{1/2} &= 1 - \left(\frac{r}{\rho} \right)^2 \cos \alpha + \frac{1}{2} \left(\frac{r}{\rho} \right)^4 (1 - \cos^2 \alpha) + O \left(\frac{r}{\rho} \right)^6 \\ \left[1 - 2 \frac{d}{\rho} \left(1 - \frac{r}{\rho} \cos \alpha \right) + \left(\frac{d}{\rho} \right)^4 \right]^{-1/2} &= \left(\frac{r}{\rho} \right)^{-1} (1 + 3 \cos^2 \alpha)^{-1/2} \cdot \\ \left[1 + 2 \left(\frac{r}{\rho} \right) \frac{\cos \alpha}{1 + 3 \cos^2 \alpha} - \frac{1}{8} \left(\frac{r}{\rho} \right)^2 \left(\frac{5 - 35 \cos^2 \alpha - 5 \cos^4 \alpha + 3 \cos^6 \alpha}{(1 + 3 \cos^2 \alpha)^2} \right) + O \left(\frac{r}{\rho} \right)^3 \right] \end{aligned}$$

Substitution of these expressions into (4.24) leads to the relation

$$\left(\frac{r}{\rho} \right)^5 = \left(\frac{m_1}{m_3} \right)^2 (1 + 3 \cos^2 \alpha)^{-1/2} \left[1 - 2 \left(\frac{r}{\rho} \right) \cos \alpha \left(\frac{1 + 6 \cos^2 \alpha}{1 + 3 \cos^2 \alpha} \right) + O \left(\frac{r}{\rho} \right)^2 \right]$$

or, in first-order approximation,

$$\frac{r}{\rho} \approx \left(\frac{m_1}{m_3} \right)^{2/5} (1 + 3 \cos^2 \alpha)^{-1/10} \quad (4.25)$$

This expression describes a three-dimensional surface about P_1 . This surface is rotationally symmetric about the line connecting the bodies P_1 and P_3 , and symmetric about a plane through P_1 and perpendicular to the line connecting the bodies P_1 and P_3 . The minimum radius, r_{min} , occurs in the direction to P_3 and opposite to the direction to P_3 ; the maximum radius, r_{max} , occurs perpendicular to the direction to P_3 . For these radii we find from (4.25):

$$\frac{r_{min}}{\rho} \approx 0.87 \left(\frac{m_1}{m_3} \right)^{2/5} ; \quad \frac{r_{max}}{\rho} \approx \left(\frac{m_1}{m_3} \right)^{2/5}$$

If we neglect this difference in radii and substitute $R_{s,i}$ for r_{max} , we can further approximate (4.25) by

$$\frac{R_{s,i}}{\rho} \approx \left(\frac{m_1}{m_3} \right)^{2/5} \quad (4.26)$$

This equation describes a sphere about P_1 with radius $R_{s,i}$; this sphere is referred to as the *sphere of influence* of P_1 with respect to P_3 . Inside this sphere it is appropriate to describe the motion of P_2 with respect to a non-rotating reference frame with P_1 as origin and to consider P_3 as the perturbing body, while outside this sphere we should use P_3 as the origin of the non-rotating reference frame and consider P_1 as the perturbing body.

For a first-order analysis of the motion of an interplanetary spacecraft about the planets of our solar system, P_1 is the Earth or another planet, P_2 is the spacecraft, and P_3 is the Sun. We can then compute the radii of the sphere of influence of the various planets with respect to the Sun from (4.26); these radii are listed in Table 4.3. Again, the relevant data on the masses and distances of the Sun and the planets were taken from Appendix B. Note that for the Earth $R_{s,i} \approx 10^6$ km, while for the giant outer planets $R_{s,i} \approx 48*10^6 - 87*10^6$ km. For each planet, the radius of the sphere of influence is large when compared to the radius of the planet (R), but small when compared to the (mean) distance of the planet from the Sun (a). This is a very important result for the computation of interplanetary trajectories (Chapter 18). In fact, it allows us to apply a relatively simple analysis scheme for the computation of these trajectories.

Table 4.3: The radii of the sphere of influence of the planets with respect to the Sun, both in kilometers and in terms of the mean distance of the planet from the Sun (a) and of the planet's radius (R).

| Planet | $R_{s,i} (10^6 \text{ km})$ | $R_{s,i} (10^{-2} a)$ | $R_{s,i} (10^2 R)$ |
|---------|-----------------------------|-----------------------|--------------------|
| Mercury | 0.11 | 0.19 | 0.46 |
| Venus | 0.62 | 0.57 | 1.02 |
| Earth | 0.92 | 0.62 | 1.45 |
| Mars | 0.58 | 0.25 | 1.70 |
| Jupiter | 48.2 | 6.19 | 6.74 |
| Saturn | 54.5 | 3.82 | 9.05 |
| Uranus | 51.8 | 1.80 | 20.2 |
| Neptune | 86.6 | 1.93 | 35.0 |
| Pluto | 3.10 | 0.052 | 25.9 |

When we consider a lunar trajectory, we are, of course, interested in the sphere of influence of the Moon with respect to the Earth in the system Moon (P_1), spacecraft (P_2), Earth (P_3). We then find from (4.26): $R_{s,i}/\rho \approx 0.172$, which leads with $\rho = 384,401$ km to $R_{s,i} \approx 66,183$ km. Although $R_{s,i}/\rho < 0.2$, this value is significantly larger than the corresponding values for the planetary spheres of influence. This means that the series expansions leading to (4-25) are less accurate. It can also be concluded that the sphere of influence of the Moon with respect to the Earth is located within the sphere of influence of the Earth with respect to the Sun. So, for a first-order analysis of trajectories to the Moon (Chapter 17) only the gravitational attraction by the Earth and the Moon have to be taken into account. During an interplanetary flight (Chapter 18) the spacecraft moves within the sphere of influence of the Earth, of the target planet, possibly of the Moon, and possibly of other planets that are passed at a short distance. Therefore, at least the gravitational forces of these celestial bodies have to be taken into account.

5. TWO-BODY PROBLEM

Equation (4.4) describes the motion of body i with respect to a non-rotating reference frame with body k as origin, under the influence of all gravitational forces between the bodies i, j and k . In Section 4.2, it was shown that for the motion of the Earth about the Sun and of satellites about the Earth (and about other planets) in first-order approximation the effects of the gravitational attraction between the bodies j and i can be neglected with respect to the effect of the gravitational attraction between the bodies i and k . In that case, the relative motion of body i is to good approximation described by

$$\frac{d^2 \bar{r}_{ki}}{dt^2} = -G \frac{\mathbf{m}_k + \mathbf{m}_i}{\bar{r}_{ki}^3} \bar{r}_{ki} \quad (5.1-1)$$

where the index ki indicates that the motion of body i is described relative to a non-rotating reference frame with body k as origin. When we would have started from the barycentric form of the equation of motion ((2.21)), then the equation for two-body motion would read

$$\frac{d^2 \bar{r}_{Bi}}{dt^2} = -G \frac{\mathbf{m}_k + \mathbf{m}_i}{\bar{r}_{Bi}^3} \bar{r}_{Bi} + G \mathbf{m}_k \left(\frac{1}{\bar{r}_{ik}^3} - \frac{1}{\bar{r}_{Bi}^3} \right) \bar{r}_{ik} \quad (5.1-2)$$

where the index Bi indicates that the motion of body i is described relative to a non-rotating reference frame with the barycenter as origin. With (2.65), we may write

$$\bar{r}_{Bk} = -\frac{\mathbf{m}_i}{\mathbf{m}_k} \bar{r}_{Bi} ; \quad \bar{r}_{ik} = -\left(1 + \frac{\mathbf{m}_i}{\mathbf{m}_k}\right) \bar{r}_{Bi} ; \quad \bar{r}_{ik}^3 = \left(1 + \frac{\mathbf{m}_i}{\mathbf{m}_k}\right)^3 \bar{r}_{Bi}^3$$

Substitution of these expressions into (5.1-2) leads, after some algebraic manipulation, to

$$\frac{d^2 \bar{r}_{Bi}}{dt^2} = -G \frac{\mathbf{m}_k + \mathbf{m}_i}{\left(1 + \frac{\mathbf{m}_i}{\mathbf{m}_k}\right)^3 \bar{r}_{Bi}^3} \bar{r}_{Bi} \quad (5.1-3)$$

A comparison of (5.1-1) and (5.1-3) shows that the orbits described by both equations are similar; this result was already found in Section 2.7.

If the notation

$$\mu = G \mathbf{m}_k \left(1 + \frac{\mathbf{m}_i}{\mathbf{m}_k}\right) \quad (5.2-1)$$

is introduced and the index ki is omitted, (5.1-1) can be written as

$$\frac{d^2 \bar{r}}{dt^2} = -\frac{\mu}{\bar{r}^3} \bar{r} \quad (5.3)$$

where μ is a constant that depends on the universal gravitational constant, G , and the masses of both bodies. If the notation

$$\mu = G \mathbf{m}_k \left(1 + \frac{\mathbf{m}_i}{\mathbf{m}_k}\right)^{-2} \quad (5.2-2)$$

is introduced and the index i is omitted, (5.1-3) also can be written as (5.3). So, (5.3) holds both for the motion of body i relative to a non-rotating reference frame connected to body k and for the motion of body i relative to a non-rotating reference frame connected to the barycenter of the two-body system. However, the parameter μ is defined differently in both cases. In many practical cases $m_i \ll m_k$. For example, for the motion of the Earth about the Sun: $m_i/m_k \approx 3.0 \cdot 10^{-6}$; for the motion of the largest planet (Jupiter) about the Sun: $m_i/m_k \approx 9.5 \cdot 10^{-4}$; for a satellite with a mass of 10 ton in an orbit about the Earth: $m_i/m_k \approx 1.7 \cdot 10^{-21}$. Therefore, as a good approximation we may write (5.2-1) and (5.2-2) as

$$\mu = G m_k \quad (5.4)$$

in particular for computing orbits of satellites about the Earth (or other planets). The parameter μ , as defined by (5.4), is only dependent on the mass of the main body k (i.e. the Sun or the Earth). This makes it a characteristic parameter of the main body and it is called the *gravitational parameter* of that body. For the Sun, $\mu \approx 1.32712 \cdot 10^{11} \text{ km}^3/\text{s}^2$; for the Earth, $\mu \approx 398,600 \text{ km}^3/\text{s}^2$. The use of (5.4) as a definition of μ has a fundamental consequence. Equation (5.3) then describes the motion of body i about body k if we assume that $m_i = 0$. In that case, body k does not experience an acceleration, and the reference frame with body k as origin is an inertial reference frame. We then deal with the so-called *one-body problem*.

In the following discussion, we will use the general equation (5.3) that holds irrespective as to whether μ is defined according to (5.2-1), (5.2-2) or (5.4).

5.1. Conservation laws

Two conservation laws can be derived for the motion of body i . To derive the first conservation law, we take the scalar product of (5.3) and $d\bar{r}/dt$, and obtain

$$\frac{d\bar{r}}{dt} \cdot \frac{d^2\bar{r}}{dt^2} + \frac{\mu}{r^3} \frac{d\bar{r}}{dt} \cdot \bar{r} = 0$$

or

$$\frac{1}{2} \frac{d}{dt} \left(\frac{d\bar{r}}{dt} \cdot \frac{d\bar{r}}{dt} \right) + \frac{1}{2} \frac{\mu}{r^3} \frac{d}{dt} (\bar{r} \cdot \bar{r}) = \frac{1}{2} \frac{d}{dt} (V^2) - \frac{d}{dt} \left(\frac{\mu}{r} \right) = 0$$

Integration of this equation gives

$$\frac{1}{2} V^2 - \frac{\mu}{r} = \mathcal{E} \quad (5.5)$$

where \mathcal{E} is an integration constant. The first term of (5.5) indicates the *kinetic energy* per unit of mass of body i . In Section 4.1, it was found that if the motion of body i is described relative to a non-rotating reference frame fixed to body k , the gravitational potential of body k at the position of body i is given by

$$\hat{U}_i = -G \frac{m_k + m_i}{r_i}$$

or, in the notation used here,

$$\hat{U} = -\frac{\mu}{r}$$

So, the term $-\mu/r$ in (5.5) indicates the *potential energy* per unit of mass of body i , and \mathcal{E} indicates the *total energy* per unit of mass of body i . According to (5.5), this total energy is constant.

To find the second conservation law, we take the vector product of (5.3) and \bar{r} , and obtain:

$$\bar{r} \times \frac{d^2\bar{r}}{dt^2} = \frac{d}{dt} \left(\bar{r} \times \frac{d\bar{r}}{dt} \right) = 0$$

Integration gives

$$\bar{r} \times \bar{V} = \bar{H} \quad (5.6)$$

where \bar{H} is an integration constant. Equation (5.6) is the *angular momentum integral*, which shows that the angular momentum of the motion of body i (per unit of mass) remains constant. In Section 2.1 it was proved that also for the many-body problem the total angular momentum is constant and defines the *invariable plane of Laplace*. In classical celestial mechanics, \bar{H} is therefore also called the *second Laplace vector*. This term implies that there also is a *first Laplace vector*; this vector will be introduced in Section 5.7.

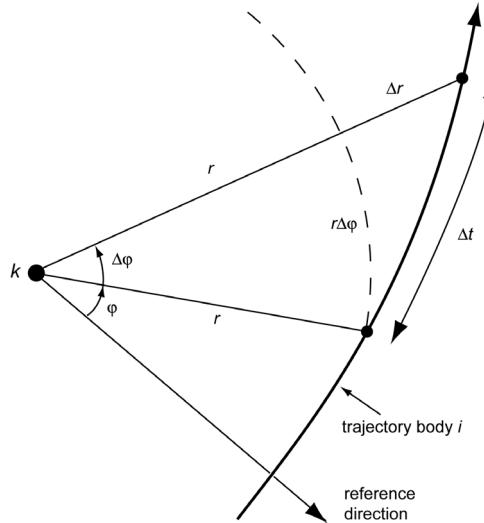


Figure 5.1: Motion of body i about body k .

Because the motion of body i occurs in a fixed plane perpendicular to \bar{H} , a non-rotating reference frame is introduced of which the origin coincides with body k and of which the XY -plane coincides with the plane of motion (Figure 5.1). At t_0 the distance between bodies i and k is r ; at some time interval Δt later $r + \Delta r$. When φ is a polar coordinate measured from an arbitrary fixed reference direction in this plane, the following expression holds according to (5.6):

$$r^2 \dot{\varphi} = \text{constant} = H \quad (5.7)$$

where H is the magnitude of the angular momentum vector (per unit of mass). From (5.7) we observe that the angular velocity, $\dot{\varphi}$, with which body i moves about the origin, is larger when the body is closer to the origin. Of course, in all practical cases $H \neq 0$. Otherwise, we would have

$$r^2 \dot{\varphi} = 0$$

which, for $r > 0$, would mean $\varphi = \text{constant}$. So, in that case the motion degenerates to a *rectilinear orbit* through body k . Most textbooks discuss rectilinear orbits and then distinguish between rectilinear elliptical, parabolic and hyperbolic orbits. However, these types of degenerated orbits have only limited practical value; an exception is the motion of some comets that can very well be approximated by a rectilinear ellipse or hyperbola. In this book only the case $H \neq 0$ is considered, except in the next paragraph where the case $H = 0$ is analyzed to illustrate the enormous effect that a continuously acting gravitational force may have on the motion of a body.

We assume that at time t_0 a body is at a distance r_0 from the center of the Sun and that it then has a velocity $V_0 = 0$ relative to a non-rotating reference frame with origin at the mass center of the Sun. We further assume that the gravitational force between that body and the Sun is the only force acting on the body. In that case, the body will move in a rectilinear orbit towards the center of the Sun.

For the acceleration of the body at a distance r_1 from the center of the Sun, we write according to (5.3)

$$\mathbf{g}_1 = \frac{\mu_s}{r_1^2}$$

where μ_s is the gravitational parameter of the Sun. This acceleration is, of course, directed towards the Sun. Because $V_0 = 0$, we find from (5.5) for the (inward) velocity of the body at a distance r_1 :

$$V_1 = \sqrt{\frac{2\mu_s}{r_0} \left(\frac{r_0}{r_1} - 1 \right)}$$

Since for this type of rectilinear motion: $V = -dr/dt$, we can write for the local velocity

$$\frac{dr}{dt} = - \sqrt{\frac{2\mu_s}{r_0} \left(\frac{r_0}{r} - 1 \right)}$$

From this relation we obtain

$$dt = - \sqrt{\frac{r_0^3}{2\mu_s}} \sqrt{\frac{r/r_0}{1 - r/r_0}} d(r/r_0)$$

with $0 \leq r/r_0 \leq 1$. Integration of this equation from $r = r_0$ to $r = r_1$ yields, after some algebraic manipulation, for the flight time $t_f = t_1 - t_0$, where t_1 is the time that the body reaches the distance r_1 from the Sun,

$$t_f = \sqrt{\frac{r_0^3}{2\mu_s}} \left[\frac{\pi}{2} - \arcsin \sqrt{\frac{r_1}{r_0}} + \sqrt{\frac{r_1}{r_0} \left(1 - \frac{r_1}{r_0} \right)} \right]$$

From this equation we obtain for the hypothetical flight time to the center of the Sun ($r_1 = 0$):

$$t_{center} = \frac{\pi}{2} \sqrt{\frac{r_0^3}{2\mu_s}}$$

If the acceleration along the trajectory would have been constant: $\mathbf{g} = \mu_s/r_0^2$, then the flight time to the center of the Sun would be given by $\sqrt{2r_0^3/\mu_s}$. So, the increasing acceleration with decreasing distance results in a (true) flight time to the center of the Sun that is equal to $\pi/4 \approx 0.78$ times the flight time in case of a hypothetical constant acceleration.

For a numerical example, we assume that at t_0 : $r_0 = 227.936 \times 10^6$ km; this is about the distance of Mars from the Sun. At that time the velocity of the body is zero and its (inward) acceleration is only 0.255 cm/s^2 . With the relations given above, we can compute that after 28.7 days the body is still at a distance of 220×10^6 km from the center of the Sun; at that time its (inward) velocity is already 6.48 km/s and its acceleration is still only 0.274 cm/s^2 . About 56.5 days later, the body crosses the orbit of the Earth ($r_1 = 149.598 \times 10^6$ km) with a velocity of 24.7 km/s and an acceleration of 0.593 cm/s^2 . About 16.1 days later it crosses the orbit of Venus ($r_1 = 108.209 \times 10^6$ km) with a velocity of 35.9 km/s and an acceleration of 1.13 cm/s^2 . About 13.0 days later it crosses the orbit of Mercury ($r_1 = 57.909 \times 10^6$ km) with a velocity of 58.5 km/s and an acceleration of 3.96 cm/s^2 , and about 7.2 days later it plunges into the outer layers of the Sun ($r_1 = 0.6955 \times 10^6$ km) with a velocity of 617 km/s and an acceleration of 274 m/s^2 . The entire trip from the orbit of Mars to the Sun lasts only 121.4 days. From this analysis we conclude that the continuously acting gravitational force speeds up the body to very high velocities. We also conclude that any body in the solar system with an orbital angular momentum $H = 0$ will be removed by the Sun within a relatively short period of time and that only bodies with $H \neq 0$ may survive and live in the solar system for extended periods of time.

We now continue with our analysis for the case $H \neq 0$, and consider a small surface element that is defined by the distances r and $r + \Delta r$ (Figure 5.1). For the area of this surface element we may write

$$\Delta A = \frac{1}{2} r^2 \Delta\phi + O(r \Delta r \Delta\phi) = \frac{1}{2} r^2 \Delta\phi + O(r \dot{\phi} \Delta r \Delta t)$$

When we take the limit $\Delta t \rightarrow 0$, we obtain

$$\frac{dA}{dt} = \frac{1}{2} r^2 \frac{d\phi}{dt}$$

or, with (5.7),

$$\frac{dA}{dt} = \frac{1}{2} H$$

which, after integration, yields

$$A = \frac{1}{2} H(t - t_0) = \frac{1}{2} H \Delta t \quad (5.8)$$

This result shows that body i sweeps out equal segments A in equal intervals of time Δt . It is noted that in this derivation we only have used the fact that the angular momentum of body i is constant. As the angular momentum is constant when only a radially directed force acts on body i , (5.8) holds for any central force field and not only for a force field that is inversely proportional to the square of the distance.

5.2. Shape of the orbit

To determine the shape of the orbit described by body i , we form the scalar product of (5.3) with \bar{r} :

$$\bar{r} \cdot \frac{d^2\bar{r}}{dt^2} + \frac{\mu}{r^3} \bar{r} \cdot \bar{r} = \frac{d}{dt} \left(\bar{r} \cdot \frac{d\bar{r}}{dt} \right) - \frac{d\bar{r}}{dt} \cdot \frac{d\bar{r}}{dt} + \frac{\mu}{r} = 0$$

Substitution of

$$\bar{r} \cdot \frac{d\bar{r}}{dt} = r \frac{dr}{dt} ; \quad \frac{d\bar{r}}{dt} \cdot \frac{d\bar{r}}{dt} = V^2 ; \quad V^2 = \dot{r}^2 + r^2 \dot{\varphi}^2 \quad (5.9)$$

results in

$$\ddot{r} - r \dot{\varphi}^2 = -\frac{\mu}{r^2} \quad (5.10)$$

Equations (5.7) and (5.10) form a system of two coupled non-linear differential equations. To solve these equations, we write

$$\dot{r} = \frac{dr}{d\varphi} \frac{d\varphi}{dt} ; \quad \ddot{r} = \frac{d\dot{r}}{d\varphi} \frac{d\varphi}{dt}$$

Substitution of (5.7) into these equations gives

$$\dot{r} = \frac{H}{r^2} \frac{dr}{d\varphi} ; \quad \ddot{r} = \frac{H}{r^2} \frac{d\dot{r}}{d\varphi} \quad (5.11)$$

For simplicity, we introduce a new variable u , defined as

$$u = \frac{1}{r} \quad (5.12)$$

Then, (5.11) can be written as

$$\begin{aligned} \dot{r} &= Hu^2 \left(-\frac{1}{u^2} \frac{du}{d\varphi} \right) = -H \frac{du}{d\varphi} \\ \ddot{r} &= Hu^2 \left(-H \frac{d^2u}{d\varphi^2} \right) = -H^2 u^2 \frac{d^2u}{d\varphi^2} \end{aligned} \quad (5.13)$$

Substitution of (5.7), (5.12) and (5.13-2) into (5.10) yields, since $H \neq 0$,

$$\frac{d^2u}{d\varphi^2} + u = \frac{\mu}{H^2} \quad (5.14)$$

The general solution of this second-order differential equation reads

$$u = \frac{\mu}{H^2} + c_1 \cos \varphi + c_2 \sin \varphi = \frac{\mu}{H^2} (1 + c_3 \cos(\varphi - \omega)) \quad (5.15)$$

where c_1, c_2, c_3 and ω are integration constants. When we transform (5.15) back to the variable r , we obtain

$$r = \frac{H^2/\mu}{1 + c_3 \cos(\varphi - \omega)} \quad (5.16-1)$$

This is the *orbital equation* that describes the relation between r and φ .

A mathematically more-elegant method for finding this orbital equation is given below. First, the vector product of (5.3) with \bar{H} is taken:

$$\frac{d^2\bar{r}}{dt^2} \times \bar{H} = -\frac{\mu}{r^3} \bar{r} \times \bar{H} = -\frac{\mu}{r^3} \bar{r} \times \left(\bar{r} \times \frac{d\bar{r}}{dt} \right)$$

Evaluation of the vector triple-product gives

$$\frac{d^2\bar{r}}{dt^2} \times \bar{H} = -\frac{\mu}{r^3} \left[\left(\bar{r} \cdot \frac{d\bar{r}}{dt} \right) \bar{r} - (\bar{r} \cdot \bar{r}) \frac{d\bar{r}}{dt} \right]$$

or, with (5.9),

$$\frac{d^2\bar{r}}{dt^2} \times \bar{H} = -\frac{\mu}{r^2} \left(\frac{dr}{dt} \bar{r} - r \frac{d\bar{r}}{dt} \right) = \mu \frac{d}{dt} \left(\frac{\bar{r}}{r} \right) \quad (5.17)$$

Since \bar{H} is constant, integration of this equation leads to

$$\frac{d\bar{r}}{dt} \times \bar{H} = \frac{\mu}{r} (\bar{r} + r \bar{c}_4)$$

where \bar{c}_4 is a constant vector. Scalar multiplication with \bar{r} yields

$$\bar{r} \cdot \left(\frac{d\bar{r}}{dt} \times \bar{H} \right) = \mu (r + \bar{r} \cdot \bar{c}_4)$$

Substitution of

$$\bar{H}^2 = \bar{H} \cdot \bar{H} = \left(\bar{r} \times \frac{d\bar{r}}{dt} \right) \cdot \bar{H} = \left(\frac{d\bar{r}}{dt} \times \bar{H} \right) \cdot \bar{r}$$

finally gives

$$\bar{H}^2 = \mu (r + \bar{r} \cdot \bar{c}_4) \quad (5.18)$$

Now, we define \bar{c}_4 to be directed towards the point in the orbit where the distance between body i and body k is a minimum. We already have introduced the angle φ between a reference direction in the orbital plane and \bar{r} ; we now introduce ω to indicate the angle between the reference direction and \bar{c}_4 (Figure 5.2). With this notation, we obtain from (5.18):

$$\bar{H}^2 = \mu [r + r c_4 \cos(\varphi - \omega)]$$

or

$$r = \frac{H^2/\mu}{1 + c_4 \cos(\varphi - \omega)} = \frac{H^2/\mu}{1 + c_4 \cos \theta} \quad (5.16-2)$$

Comparison of (5.16-1) and (5.16-2) shows that both equations are, of course, identical, with $c_3 = c_4$ and $\theta = \varphi - \omega$. In the next Section, we will prove that (5.16) describes a conic section in polar coordinates, where the origin of the reference frame (body k) is located at one of the foci

of that conic section.

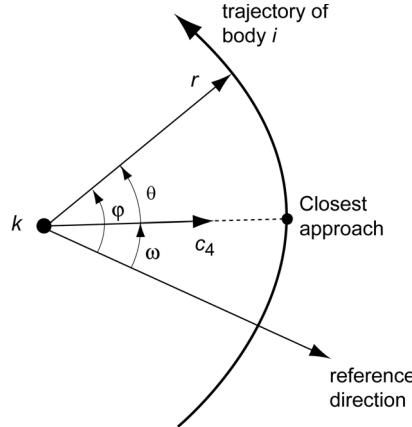


Figure 5.2: Direction of the vector \bar{c}_4 and the definition of the angles φ , ω and θ .

5.3. Conic sections

A conic section can be defined as the geometric collection of all points P for which the ratio of the distance to a fixed point F and the distance to a fixed line l is constant (Figure 5.3). The fixed point is called the *focus*, the fixed line is called the *directrix* and the ratio of the distances is called the *eccentricity*, e , of the conic section. With this definition and the notation shown in Figure 5.3, we find the following equations for a conic section:

$$P \text{ to the left of } l: \quad \mathbf{r} = e(\mathbf{k} - \mathbf{r} \cos \theta)$$

$$P \text{ to the right of } l: \quad \mathbf{r} = e(\mathbf{r} \cos \theta - \mathbf{k})$$

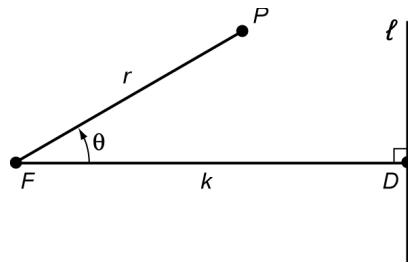


Figure 5.3: Geometric definition of a conic section.

When we introduce the *focal parameter*, p , which is defined as

$$p = e k \tag{5.19}$$

then we can write these equations as

$$P \text{ to the left of } l: \quad \mathbf{r} = \frac{\mathbf{p}}{1 + e \cos \theta} \tag{5.20-1}$$

$$P \text{ to the right of } l: \quad \mathbf{r} = \frac{-\mathbf{p}}{1 - e \cos \theta} \tag{5.20-2}$$

Because the cosine function is symmetric with respect to $\theta = 0^\circ$, the line FD is an axis of symmetry of the conic section. What type of conic section is described by (5.20) depends on the

value of the eccentricity, e . The following cases can be distinguished:

- $e = 0$: circle
- $0 < e < 1$: ellipse
- $e = 1$: parabola
- $e > 1$: hyperbola

For a circle: $k = \infty$. Since $e \geq 0$ and $k > 0$, p has a finite positive value for all conic sections. Of course, also $r \geq 0$. With (5.20) we then obtain

$$P \text{ to the left of } l: \quad \cos \theta \geq -1/e$$

$$P \text{ to the right of } l: \quad \cos \theta \geq 1/e$$

Because $e \geq 0$ and $\cos \theta \leq 1$ it is clear that the curve to the right of l can only be realized for $e > 1$; in other words: for a hyperbola. The circle, the ellipse and the parabola are entirely located to the left of l ; the hyperbola has a branch to the left as well as to the right of l .

Comparison of (5.16) and (5.20) demonstrates that body i moves in a conic section about body k , with body k located at a focal point, and that the following relations must hold:

$$H^2/\mu = p \quad ; \quad c_3 = c_4 = e \quad ; \quad \theta = \varphi - \omega \quad ; \quad H = r^2 \dot{\varphi} = r^2 \dot{\theta} \quad (5.21)$$

We know that H^2 , μ and p have positive values and therefore in physical reality (5.16) can only describe a conic section to the left of l . So, if the orbit is hyperbolic, (5.16) describes only the left branch. The reason is that the gravitational force is an attracting one. Therefore, for all problems in celestial mechanics the trajectory is uniquely described by

$$r = \frac{H^2/\mu}{1 + e \cos(\varphi - \omega)} = \frac{H^2/\mu}{1 + e \cos \theta} = \frac{p}{1 + e \cos \theta} \quad (5.22)$$

The chord of the conic section that passes through F and is perpendicular to FD ($\theta = 90^\circ$), has, according to (5.22), a length of $2p$. This chord is generally referred to as the *latus rectum*. It is interesting to note that Newton already has shown that elliptical motion is possible when the attractive central force is described by either $F = c/r^2$ or $F = cr$, where c is a positive constant. In the first case, the attracting body k is located at a focus of the ellipse; in the second case at the center of the ellipse.

For a circle and an ellipse: $e < 1$, and (5.22) shows that for each value of θ there exists a finite positive value of r . Hence, circles and ellipses are closed curves. For a parabola: $e = 1$, which, according to (5.22), also yields a finite value of r for each value of θ , except for $\theta = 180^\circ$, where $r = \infty$. For a hyperbola: $e > 1$. According to (5.22), for this type of conic section only finite values of r exist if θ satisfies the relation

$$-\pi + \arccos(1/e) < \theta < \pi - \arccos(1/e)$$

The integration constants p , e and ω describe the size, shape and orientation of the conic section in the orbital plane of m_i . Because these are constants, we conclude that the size, shape and orientation of the conic section are independent of time. The angle ω indicates the position of the point where r is a minimum. When we do not refer to a particular celestial body, this point of closest approach is called the *pericenter* or *periapsis*; for an elliptical orbit the point of farthest excursion is then called the *apocenter* or *apoapsis*. For a trajectory about the Sun these points are called *aphelion*¹ and *perihelion*; for a trajectory about the Earth *apogee* and *perigee*; for a

¹ The term *aphelion* is often used for euphony.

trajectory about the Moon *aposelene* and *periselene*, or *apolune* and *perilune*. The angle ω is then called *argument of pericenter*, *argument of perihelion*, *argument of perigee* or *argument of perilune*, respectively. The angle $\theta = \varphi - \omega$, which is called *true anomaly*, is an angle in the orbital plane, measured from the pericenter to the orbiting body in the direction of motion. The name ‘true anomaly’ is based on the ancient belief that for celestial bodies uniform circular motion was considered to be ‘normal’; a deviation from this motion was therefore considered an ‘anomaly’.

We now return to (5.10) and write this in the form

$$\ddot{\mathbf{r}} = \mathbf{r} \dot{\varphi}^2 - \frac{\mu}{r^2}$$

From this relation we conclude that the radial acceleration of body i is equal to the difference between the centrifugal acceleration and the gravitational acceleration. Now, suppose that body i moves in a circular orbit about body k . Then, $\ddot{\mathbf{r}} = \mathbf{0}$ and, consequently, the centrifugal force equals in magnitude the (centripetal) gravitational force, but acts in an opposite direction. So, the two forces cancel but this should not let us conclude that the body will move along a straight line. We should realize that the expression, in fact, holds relative to a rotating reference frame and for that reason the centrifugal force, which is an apparent force, appears in the expression. Relative to the rotating reference frame, body i keeps a fixed position. This is in agreement with our conclusion that the two forces acting on the body cancel. Relative to the inertial reference frame, only the centripetal gravitational force (natural force) acts on the body, leading to the circular orbit.

5.4. Kepler's laws

In the previous Sections, two important properties have been derived:

- Body i sweeps out sectors of equal area in equal intervals of time.
- The orbit of body i about body k is a conic section, of which body k is located at one of the foci.

These results are the general form of *Kepler's second* and *first law*, respectively. In 1609, J. Kepler (1571-1630) published *Astronomia nova seu physica coelestis* (usually referred to as *Astronomia nova*), in which he formulated two empirical laws about the orbits of the planets around the Sun:

Kepler's first law: The orbit of a planet is an ellipse with the Sun at one of the two foci.

Kepler's second law: A line joining a planet and the Sun sweeps out equal areas during equal intervals of time.

Kepler deduced these laws from observations of Mars taken by Tycho Brahe (Tyge Ottesen Brahe; 1546-1601) over a long period of time (Section 5.5), but he immediately concluded that these laws must hold for all planets. His laws broke with the Ptolemaic model (Claudius Ptolemy; latin: Ptolemaeus; ~85-165), in which the Earth was the center of the universe, and also with Copernicus' model (N. Copernicus; 1473-1543), in which the planets moved in circular orbits about the Sun (Section 5.5). In 1619, Kepler published *Harmonices mundi*, in which he formulated, without proof, his third empirical law:

Kepler's third law: The square of the orbital period of a planet is directly proportional to the cube of the semi-major axis of its orbit.

This law will be proved in Section 6.4.

Kepler's first law describes only one case (ellipse) of the orbits in which celestial bodies can

move. The other types of conic sections occur too. Some comets move in (near-)parabolic or hyperbolic orbits about the Sun, and hyperbolic orbits with respect to the Earth and other planets are flown in interplanetary missions (Chapter 18). Some natural satellites and many artificial satellites move in (near-)circular orbits about the Earth and other planets. We call all these conic sections *Keplerian orbits*.

In this book, we have derived Kepler's laws from Newton's laws of motion and Newton's law of gravitation. Historically, a different development has taken place. I. Newton (1643-1727) published his three laws of motion in his *Principia* in 1687. He proved by geometric methods that if his three laws of motion are valid to describe the elliptical motion of a planet about the Sun, and if the planetary motion can be described by the three empirical laws of Kepler, then there must be an attractive force between the Sun and a planet that is proportional to the mass of both bodies and inversely proportional to the square of the distance between them. Since Newton realized that, as a result of the enormous distance between the Sun and a planet, both bodies can be viewed as point masses, he postulated that such an attractive force is present between all point masses: *Newton's law of gravitation*. It is not widely known, however, that Newton thoroughly analyzed a variety of power-law central attracting forces of the type $F = -\alpha r^n$, where α is a positive constant, and showed that different integer exponents n yield circular, conic, spiral, and other orbits in an often surprising manner. Only the exponents $n = -2$ and $n = 1$ lead to conic sections; for $n = -2$ a conic section with its focus at the center of the Sun, and for $n = 1$ a conic section with its center at the center of the Sun. In continental Europe the methods of calculus were cultivated with ardor at the beginning of the eighteenth century, and Newton's system of mechanics did not find immediate acceptance; indeed, the French clung to the *vortex theory* of R. Descartes (1596-1650) until Voltaire (F.M. Arouet, 1694-1778), after his visit to London in 1727, vigorously supported the Newtonian theory. This, with the fact that the English continued to employ the geometric methods of the *Principia*, delayed the analytical solution of the two-body problem. J. (Johann) Bernoulli (1667-1748) solved the inverse problem in 1710. He used analytical methods to show that a central attracting force with magnitude inversely proportional to the square of the distance between a planet and the Sun always leads to an orbit about the Sun that can be represented by a conic section. There are, however, indications that this solution was actually found by J. Hermann (1678-1733), a pupil of J. (Jacob) Bernoulli (1654-1705). Newton presented a similar analysis in the second edition of his *Principia* in 1713. J. (Johann) Bernoulli proved that a central attracting force with magnitude inversely proportional to the cube of the distance between a body and the Sun leads to a particular Cotes' spiral (R. Cotes, 1682-1716) about the Sun, depending on the choice of the initial conditions. The two-body problem was solved in detail by L. Euler (1707-1783) in 1744 in his *Theoria Motuum Planetarum et Cometarum*. Since that time, the modifications have been mainly in the choice of variables in which the problem is expressed. By the end of the nineteenth century, various studies were performed to determine by analytical methods which types of central forces would make particles to move in conic sections. J.L.F. Bertrand (1822-1900) has shown in 1873 that the only laws of central attraction force under the action of which a particle will describe a conic section for all initial conditions are $F = -\alpha/r^2$ and $F = -\alpha r$, where α is a positive constant. This confirmed the results obtained by Newton who used geometric methods. G.X.P. Koenigs (1858-1931) has proved in 1889 that the only laws of central attraction force depending upon the distance alone, for which the curves described by the particle are algebraic for all initial conditions are $F = -\alpha/r^2$ and $F = -\alpha r$. F.L. Griffin (1854-1933) has shown in 1905 that the only law, where the attraction force is a function of the distance alone, where it does not vanish at the center of force, and where it is real throughout the plane, giving an elliptical orbit is the Newtonian law.

5.5 From geocentrism to heliocentrism

To anyone on Earth looking at the sky, it seems clear that the Earth is stationary, while all celestial bodies rise in the east and set in the west once a day. Observing over a longer time, one observes that some bodies exhibit more complicated motions: planets and comets follow their individual trajectories relative to the invariable stellar configurations, and sometimes meteoroids flash across the sky when they burn up in the atmosphere. Ancient cultures already observed these phenomena and were in particular interested in the motion of the five planets that can be seen with the naked eye (Mercury, Venus, Mars, Jupiter, Saturn). These objects were called by the Greek ‘*astēr planētēs*’ (‘wandering star’). Scientists of ancient antiquity all over the world started to develop kinematic models of the universe to describe and predict the motions of the observed celestial bodies and to find explanations for recurring celestial phenomena. Examples are the development of cosmology in ancient Egypt and Mesopotamia. Since in Egypt the principal deities were heavenly bodies, a great deal of effort was made by the priesthood to calculate and predict the time and place of their god’s appearances. Babylonian astronomy is noted for the detailed, and continuous, records of astronomical phenomena, such as eclipses, positions of the planets and risings and settings of the Moon; these records date back to 800 B.C. While their record keeping was a novel technology for the time, and their system of stellar names and measurement system was passed on to later civilizations, the Babylonians never developed a cosmological model in which to interpret their observations. It were the early Greek astronomers that have achieved this goal using the Babylonian data.

According to the early Babylonian cosmology (~3000 B.C.) the Earth and the heavens form a unit within infinite ‘waters of chaos’. The Earth is assumed to be flat and circular, and a solid dome (Latin: ‘firmamentum’) keeps out the outer ‘chaos ocean’. This conception was adopted in the early Jewish cosmology and is reflected in the book Genesis of the Bible. However, the flat Earth model conflicted with various kinds of observations, and already in the sixth century B.C., Greek, Egyptian and Babylonian astronomers knew that the Earth is a sphere. Central to Greek cosmology is the belief that the underlying order of the universe can be expressed in mathematical form. In Greek antiquity the ideas of celestial spheres first appeared in the cosmology of Anaximander (~610-546 B.C.). After him, Pythagoras of Samos (~570-495 B.C.), Xenophanes of Colophon (~570-475 B.C.) and Parmenides of Elea (~515-450) held that the universe is spherical and that celestial bodies are placed on concentric celestial spheres. According to Plato (~427-347 B.C.), the Earth is a sphere, stationary at the center of the universe. The celestial bodies are housed in spheres about the Earth and the motion of these bodies is uniform along circles. His argument was that the heavens are the place of the gods and thus represent perfection: a ‘*kosmos*’ (Greek for ‘order’ and ‘harmony’), and spheres and circles are of ‘divine uniformity’. He proposed that the seemingly irregular motions of the planets could be explained by combinations of uniform circular motions centered on a spherical Earth. In essence, this is an early formulation of a Fourier analysis by which we may approximate any function, e.g. the trajectory of a planet, by a series of trigonometric functions. One should realize that the adoption of a geocentric model was largely based on several astronomical observations; e.g., ancient cultures could not measure any stellar parallax, and observed an apparent consistency of Venus’ luminosity. In this context, it is noted that: 1) the parallax of a star (61 Cygni) was first measured by F.W. Bessel (1784-1846) in 1839; 2) the loss of light caused by Venus’ phases largely compensates for the increase in apparent size caused by its varying distance from Earth. The idea that the Earth is at the center of the universe was further supported by the belief that the Earth could not rotate or move through space, because we otherwise would experience huge winds, objects would be spun out into space, objects dropped from a tower would fall behind that tower,

etc. Eudoxus of Cnidus (~409-355 B.C.), who worked with Plato, developed a more mathematical explanation of the planet's motions based on Plato's dictum that all phenomena in the heavens can be explained with uniform circular motions. He developed a system involving some 27 geocentric 'crystalline' spheres. Of these spheres, the outermost bore the stars and accounted for their daily motion across the sky, three each were used to account for the motion of the Sun and the Moon, and four each were necessary for the more complicated motions of the planets known at that time. Each sphere turns uniformly about an axis, and each planet is attached to a sphere. So, the motion of the planet is a circle about the Earth. By tilting the axes of the spheres, and assigning each a different period of revolution, he was able to approximate the celestial 'appearances'. Note that the concept of an outer sphere to which the stars are attached has survived the centuries and is still used to describe angular positions of objects in space (Section 11.2). A great achievement of Eudoxus was his concept that the rotations of two of the spheres for a given planet could be tuned in such a way that the phenomenon of planetary retrogression could be modeled. Aristotle (Aristotélēs; ~384-322 B.C.), a student of Plato, and Callippus of Cyzicus (~370-300 B.C.), who was a student of Eudoxus and worked with Aristotle, found that 27 spheres was insufficient to accurately account for the planetary movements, and so they added more spheres. In Aristotle's fully developed model, the spherical Earth is at the center of the universe and the planets are moved by either 47 or 55 interconnected spheres. Each of these concentric spheres is moved by its own god. While Eudoxus thought of his spheres as being purely mathematical concepts, Aristotle and Callippus have thought of them as having material existence. Aristotle believed that these spheres are made of an unchanging fifth element², aether. Heraclides Ponticus (~387-312 B.C.) explained the apparent daily motion of the celestial bodies through the rotation of the Earth and proposed to account for the retrograde motions of Mercury and Venus by circular orbits centered in the Sun and carried by it in a circular orbit around the Earth. Ambrosius Theodosius Macrobius (395-423) later described this as the 'Egyptian system', suggesting that the model originally came from ancient Egypt. Thus may have been born the Greek concepts of *epicycle* and *deferent* (Figure 5.4), which were developed by Appollonius of

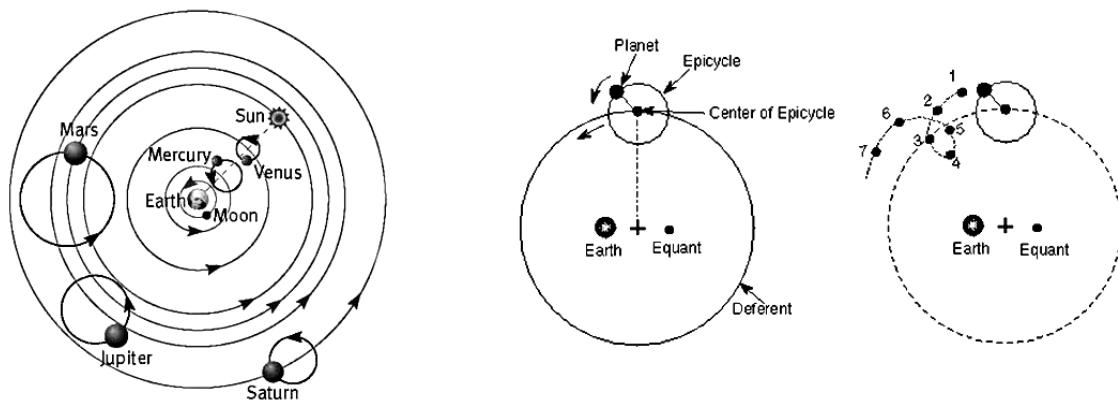


Figure 5.4: (left) Ptolemaic model for the motion of celestial bodies; Sun, Moon and planets move about the Earth. (center) A planet moves along an epicycle, while the center of this epicycle moves along the deferent. Epicycle speed is uniform with respect to the equant. (right) Deferent motion is in direction of point 1 to 7, but the planet's epicycle carries it on a cycloid path so that from points 3 through 5 the planet moves backwards (retrograde).

² The ancient Greeks believed that everything was made up of four elements: *earth*, *water*, *air*, and *fire*. Aristotle suggested that there was a fifth element, *aether*, because it seemed strange that heavenly bodies would be made out of earthly elements.

Perga (~265-190 B.C.) and Hipparchos (~190-120 B.C.) into a system that accounted quite well for observed variations in the motion of planets. The system was modified and improved by Ptolemy and published in his *Almagest* (*Mathematike syntaxis*; *Syntaxis mathematica*; *Almagestum*) in about 150. The basic Ptolemaic model is based upon a system consisting of a deferent and an epicycle for each of the five planets known to him and only a deferent for the Sun and the Moon (Figure 5.4). The deferent is a circle centered near the Earth; because the Earth is positioned outside the center of the deferent this center is referred to as the *eccentric*. Although this construction violated the rule that the Earth is the center of the cosmos and all planetary motions, this displacement was minimal and was considered a slight bending of the rule rather than a violation. The eccentric in Figure 5.4 is fixed; it could also be made movable. In this case, the center of the deferent is a point that rotates around the Earth in a small circle centered on the Earth. In some constructions this little circle was even not centered in the Earth. The epicycle is a circle centered on a point on the deferent which moves counterclockwise (seen from the celestial north pole) along the deferent at an angular rate that appears constant when viewed, not from the center of the circle, but from the *equant*. By introducing this equant, Ptolemy could account for the velocity variations in the motions of the planets. The equant is placed directly opposite the Earth from the center of the deferent and at the same distance from the center as the Earth. A planet moves counterclockwise along an epicycle with a constant angular rate. The planes of the deferents of the planets are inclined at small angles to the plane of the Sun about the Earth, while the planes of their epicycles are inclined at small angles to the deferents. By selecting appropriate values of these angles and of the radii of deferent and epicycle, Ptolemy succeeded in fitting the model with reasonably accuracy through the existing observations. To account for discrepancies between his model and observations, he later added more epicycles to the model. His model, with some further modifications introduced by later scholars, came down to the sixteenth and seventeenth centuries as the accepted Ptolemaic system.

Already long before Ptolemy there were scholars in the ancient Arabic, Chinese, Greek, Indian and Latin world who expressed ideas that not the Earth but the Sun is at the center of the universe and that some or all of the planets revolve around the Sun. These early sources, however, do not provide techniques to compute any observational consequences of their proposed heliocentric ideas, and most of the ancient literature has been lost. The first complete and detailed non-geocentric model of the universe was proposed by Philolaus of Croton (~470-385 B.C.), a Pythagorean. He taught that the Earth, Sun, Moon and planets revolve in uniform circular motion about a ‘central fire’ at the center of the universe, and that a ‘counter-Earth’, which is positioned on the other side of the central fire and collinear with the Earth and the central fire, revolves about the central fire with the same period of revolution as the Earth. The central fire, the five known planets, Sun, Moon and Earth add up to nine bodies. In his days, 10 was considered as the ‘perfect’ number and therefore it was believed that the total number of bodies should be ten; hence the introduction of the counter-Earth. The Sun revolves around the central fire once a year, and the stars are stationary. The Earth maintains the same hidden face towards the central fire, rendering both it and the counter-Earth invisible from Earth. The first person known to have proposed a true heliocentric system is Aristarchus of Samos (~310-240 B.C.). His writings are lost, but some information is known from surviving descriptions and critical commentary by his contemporaries. Aristarchus stated that the Earth rotates daily about its axis and revolves annually about the Sun in a circular orbit. Hipparchos also has studied a heliocentric system, but he abandoned his work because the calculations showed that the orbits were not perfectly circular as believed to be mandatory by the science of his time. There were likely other astronomers in the classical period who also espoused heliocentrism, but whose work is now lost to us. The only

other astronomer from antiquity who is known to have supported Aristarchus' heliocentric model was Seleucus of Seleucia (~190-150 B.C.). A fragment of his work has survived in Arabic translation. At that time some ideas about the existence of a force attracting the planets towards the Sun had already been formulated (Section 1.4), but the motion of celestial bodies was still considered as a kinematic problem.

The European cultural elite of the first century A.D. were fully aware of the existence of a heliocentric model of planetary motions. However, the geocentric model of Ptolemy was generally used and remained in use. This was partly the result of the rise of the Christian Church, which gave the Earth a special and privileged position. In Roman Carthage, Martianus Capella (~365-440) expressed the opinion that the planets Venus and Mercury did not go about the Earth but instead circle the Sun. His model was discussed in the Early Middle Ages by various anonymous commentators. Medieval astronomers and philosophers developed diverse theories about the causes of the celestial sphere's motions. By the end of the Middle Ages, the common opinion in Europe was that the outermost sphere carrying the stars was moved by the Prime Mover, who was identified with God. Each of the lower spheres was moved by a subordinate spiritual mover, identified with the angels of Revelation.

In India, Aryabhata (476-550) propounded a model in which the Earth is spinning about its axis and the periods of the planets are given with respect to a stationary Sun. Although this may be seen as a sign of an underlying heliocentric concept, he developed a geocentric model of the solar system, in which Sun and Moon are each carried by epicycles which revolve around the Earth. The motions of the planets are each governed by two epicycles; the positions of the planets was calculated relative to uniformly moving points. In the case of Mercury and Venus, they move around the Earth at the same mean speed as the Sun. In the case of Mars, Jupiter, and Saturn, they move around the Earth at specific speeds. This two-epicycle model reflects elements of pre-Ptolemaic Greek astronomy. Early followers of Aryabhata's model include Varahamihira (505-587) and Brahmagupta (598-668). Bhaskara (1114-1185) also made reference to heliocentrism and he accurately calculated astronomical constants based on a heliocentric system. Nilakantha Somayaji (1444-1544) developed around 1500 a partially heliocentric planetary model, in which the planets orbit the Sun, which in turn orbits the Earth, similar to the Tychonic system (see below). Nilakantha's system also incorporated the Earth's rotation about its axis and elliptical orbits.

Little is known about the development of cosmological theories in ancient China, but we know that Chinese astronomers were the most persistent and accurate observers of celestial phenomena anywhere in the world before the Arabs. Early maps of stellar configurations were found that date back to 6000 years ago. The ancient Chinese developed three different cosmological models: 1) a hemispherical dome model conceived the heavens as a hemisphere lying over a dome-shaped Earth; 2) a model that saw heavens as a celestial sphere quite similar to the spherical models developed in the Greek and Hellenistic traditions; 3) a model that viewed heavens as infinite in extent and the celestial bodies as floating about at rare intervals. Detailed records of astronomical observations began in the fourth century B.C. Shen Kuo (Shen Gua; 1031-1095) published cosmological hypotheses explaining the variations of planetary motions, including retrogradation. His hypotheses were similar to the concept of the epicycle in the Greco-Roman tradition.

In the Medieval Islamic world, the Muslim astronomers accepted unanimously the Ptolemaic geocentric model. However, several Muslim scholars questioned the Earth's apparent immobility and centrality within the universe. Abu Ali al-Hasan (Alhacen; 965-1039) proposed that the Earth is rotating about its axis. Abu Rayhan Biruni (973-1048) discussed the possibility of whether the

Earth rotates about its own axis and around the Sun, but he considered this a philosophical problem rather than a mathematical one. Mu'ayyad al-Din al-'Urdi (-1266) developed a non-Ptolemaic model, in particular for the motion of the Moon. Muhammad ibn Muhammad ibn Hasan Tusi (1201-1274), better known as Tusi in the West, made very accurate tables of planetary movements. He invented a geometric technique called a Tusi-couple, which generates linear motion from the sum of two circular motions. He used this technique to replace Ptolemy's equant. Tusi criticized Ptolemy's use of observational evidence to show that the Earth was at rest, and he was the first to present empirical observational evidence of the Earth's rotation. Najm al-Din al-Qazwini al-Katibi (-1276) wrote an argument for a heliocentric model, but later abandoned the idea. Qotb al-Din Shirazi (1236-1311) discussed the possibility of heliocentrism, but also later rejected it. Ibn al-Shatir (1304-1375) applied the Tusi-couple concept and developed a geocentric system that employed mathematical techniques, which were almost identical to those Copernicus (see below) later employed in his heliocentric system. However, the Islamic scientists never made the big leap to heliocentrism. By the end of the fourteenth century Arabian astronomy practically ceased to make any further progress.

Astronomy began to revive in Europe toward the end of the fifteenth century in the labors of G. von Peuerbach (1423-1461), B. Walther (Waltherus; 1430-1504) and J. Müller von Königsberg (Regiomontanus; 1436-1476). It was given a great impetus by Copernicus. In 1543, he published his *De revolutionibus orbium coelestium*, in which he proposed that the Sun is the center of the universe and that all planets move about the Sun. He discussed the philosophical implications of his proposed system, used selected astronomical observations to derive the parameters of his model, and wrote astronomical tables which enabled one to compute the past and future positions of the stars and planets. In doing so, Copernicus moved heliocentrism from philosophical speculation to predictive geometrical astronomy. His theory resolved the issue of planetary retrograde motion by arguing that such motion was only perceived and apparent, rather than real. It is emphasized that Copernicus' heliocentric theory was still based on circular orbits. Therefore, he was forced to use epicycles to account for deviations in the observed planetary motion. He eliminated Ptolemy's concept of the equant at the cost of additional epicycles. In developing his theories of planetary motion, Copernicus was probably indebted to the earlier work of Martianus Capella, Aryabhata, Tusi and Ibn al-Shatir; they had resolved significant problems in the Ptolemaic system, though retaining an essentially geocentric arrangement.

The first information about the heliocentric views of Copernicus were circulated already some time before 1514 in manuscript and became well known among astronomers and others. His ideas contradicted the then-prevailing understanding of the Bible. In 1533, J.A. Widmannstetter (1506-1557) delivered in Rome a series of lectures outlining Copernicus' theory. The lectures were heard with interest by Pope Clement VII and several Catholic cardinals. In 1536, the Archbishop of Capua, N. von Schönberg, wrote a letter to Copernicus from Rome encouraging him to publish a full version of his theory. Copernicus finished his *De revolutionibus orbium coelestium* in 1530, but did not publish it until the year of his death. The book, which he dedicated to Pope Paul III, contained an unsigned preface by A. Osiander (1498-1552) defending the system and arguing that it was useful for computation even if its hypotheses were not necessarily true. Possibly because of that preface, the work of Copernicus inspired very little debate on whether it might be heretical during the next sixty years.

Tycho Brahe advocated around 1585 an alternative to the Ptolemaic geocentric system. It was a geo-heliocentric system, now known as the Tychonic system, in which the five then known planets orbit the Sun, while the Sun and the Moon orbit the Earth. He rejected the system of Copernicus because he could not observe any parallax in the fixed stars. Tycho was an

indefatigable and most painstaking observer. He is credited with the most accurate astronomical observations of his time; no one before Tycho had attempted to make so many planetary observations. He was the last major astronomer to work without the aid of a telescope. From 1600 until his death in 1601, he had Kepler as assistant.

After Tycho's death, Kepler was appointed his successor and imperial mathematician at the court of the Holy Roman Emperor Rudolph II at Prague. In that position he had access to all of Tycho's observations. Kepler lived in an era where there was no clear distinction between astronomy and astrology, but there was a strong division between astronomy and physics. He was the first who tried to combine both disciplines and to treat astronomy as part of a universal mathematical physics. With this approach he broke with the ancient picture that the motion of celestial bodies was the result of 'divine powers' and that the cause of that motion was therefore out of the reach of human beings. Within his religious view of the cosmos, he considered the Sun as the source of 'motive force' in the solar system and supposed that the 'motive power' radiated by the Sun weakens with distance, causing faster or slower motion as planets move closer or farther from it. His *Mysterium cosmographicum* (1596) was the first published defense of the Copernican system. During his studies he became very interested in the shape of the orbits of the planets and spent a lot of effort in trying to fit polygons, three-dimensional polyhedra and ovoids through the existing observations of the motion of the planets, in particular of Mars. After these efforts had failed, in early 1605 he at last hit upon the idea of an ellipse, which he had previously assumed to be too simple a solution for earlier astronomers to have overlooked. This time he had success. Finding that an elliptical orbit fitted the Mars data, he immediately concluded that all planets move in elliptical orbits about the Sun. The introduction of elliptical orbits eliminated the need of epicycles. In this context, it is interesting to note that any ellipse can be described by a deferent and an epicycle with appropriately selected radii! Kepler's research culminated in the publication of *Astronomia nova seu physica coelestis* (1609), in which he formulated his first two laws of planetary motion (Section 5.4). His ideas, however, were not immediately accepted and many well-known astronomers and mathematicians of his time objected Kepler's introduction of physics into astronomy, and completely ignored his early works. In *Epitome astronomia Copernicanae*, consisting of three books published between 1615 and 1621, he explained heavenly motions through physical causes and he explicitly extended his first two laws of planetary motion to all the planets as well as the Moon and the satellites of Jupiter, which were discovered by Galileo Galilei (1564-1642) in 1610 (see below). In 1619 he published *Harmonices mundi*, in which he attempted to explain the proportions of the natural world—particularly the astronomical and astrological aspects—in terms of music. The central set of 'harmonies' was the *musica universalis*, which had been studied by Pythagoras, Ptolemy and many others before Kepler. In book V, dealing with planetary motion, Kepler articulated his harmonic law, which became known as his third law of planetary motion (Sections 5.4 and 6.4). He found that law empirically after trying many combinations of the mean distance of planets and the periods of their motion about the Sun. Following Kepler's death in 1630 his *Epitome astronomia Copernicanae* was the main vehicle for spreading his ideas. It was read by astronomers and mathematicians throughout Europe. In the late seventeenth century, a number of physical astronomical theories draw on Kepler's work. This culminated in Newton's *Principia*, published in 1687, in which Newton derived Kepler's laws of planetary motion from a force-based theory of universal gravitation.

Galileo Galilei, a contemporary of Kepler, applied the newly invented telescope to look at celestial objects. In 1610 he discovered four satellites encircling Jupiter, the rings of Saturn and spots on the Sun; he also found that Venus exhibited a full range of phases. These discoveries were not consistent with the Ptolemaic model of the solar system. He therefore became an ardent

supporter of the heliocentric theory, but completely ignored Kepler's work. Galileo claimed that heliocentrism was not contrary to the Bible. In 1616, Pope Paul V summoned Galileo to Rome to defend his position. The Church then decided to accept the use of heliocentrism as a calculating device of great mathematical simplicity, but opposed it as a literal description of the solar system. Galileo received the papal command not to 'hold or defend' the heliocentric idea in public. Pope Urban VIII encouraged Galileo to publish the pros and cons of heliocentrism. Galileo's *Dialogue concerning the two chief world systems*, published in 1632, clearly advocated heliocentrism and appeared to ridicule the Ptolemaic model and the Pope. Therefore, Urban VIII became hostile to Galileo and he was again summoned to Rome and was put on trial before the Inquisition in 1633. For advancing heliocentric theory he was accused of heresy and was sentenced to house arrest for life. In 1664, Pope Alexander VII published a new version of the *Index Librorum prohibitorum* (*Index of prohibited books*), which included all previous condemnations of heliocentric books. An annotated copy of Newton's *Principia* was published in 1742 by two Catholic mathematicians with a preface stating that the author's work assumed heliocentrism and could not be explained without that theory. In 1758 the Catholic Church dropped the general prohibition of books advocating heliocentrism from the *Index librorum prohibitorum*.

It is emphasized that all ancient models for the motion of celestial bodies were kinematic. The early scholars looked for regularities and repetitions of observed astronomical phenomena and tried to model these in a kinematic way such that they could predict future phenomena, like solar eclipses, etc. These theories lacked an analysis of the cause of the motion of celestial bodies, i.e. the nature of dynamics. That cause was generally considered as 'divine power' and outside the domain of human investigation. However, already before the second century B.C. various scientist had proposed that the planets are attracted towards the Sun and that this force shapes the trajectories of the celestial bodies (Section 1.4). The link between observed motion of a celestial body and the physical laws that govern this motion was established in the seventeenth century. The development of dynamical theories started with Newton's *Principia* in 1687 (Section 1.1).

In the view of modern science, Kepler's laws of planetary motion were used as arguments in favor of the heliocentric hypothesis. An apparent proof of the heliocentric hypothesis was provided by F.W. Bessel (1784-1846) in 1838. He measured a parallax of $0.314''$ of the star 61 Cygni and by this proved that the Earth moves in space. In the same year F.G.W. von Struve (1793-1864) and T. Henderson (1798-1844) measured the parallaxes of the stars Vega and Alpha Centauri. The thinking that the heliocentric view is also not true in a strict sense was achieved in steps. That the Sun was not the center of the universe, but one of innumerable stars, was already advocated by the Giordano Bruno (1548-1600). Over the course of the eighteenth and nineteenth centuries, the status of the Sun as merely one star among many became increasingly obvious. With the observations of F.W. Herschel (1738-1822) astronomers realized that the solar system is moving through space, and by the 1920s E.P. Hubble (1889-1953) had shown that it is part of a galaxy that is only one of many billions. We know today that the Sun is not at the center of mass of the solar system because the masses of the planets cannot be neglected in comparison to the Sun's mass (Section 2.2). In addition, the concept of an absolute velocity, including being 'at rest' as a particular case, is ruled out by the principle of relativity, eliminating any obvious 'center' of the universe as a natural origin of coordinates. In modern calculations, the origin and orientation of a coordinate system often are selected for practical reasons, and we may select the origin in the center of mass of the Earth, of the Earth-Moon system, of the Sun, of the Sun plus the major planets, or of the entire solar system, as long as the laws of dynamics are applied correctly. For us, the selection of 'geocentric' or 'heliocentric' reference frames has

only practical implications and not philosophical or physical ones. Today, we know that there is no real ‘truth’ in describing the orbits of celestial bodies. We may think the planets to move about the Sun as well as the Sun and the planets to move about the Earth. However, there is a major difference: the concept of planets moving about the Sun, in combination with Newton’s law of gravitation, leads to the simplest dynamical model and representation of the orbits.

5.6. Velocity components

When the *flight path angle*, γ , is introduced (Figure 5.5), the radial and normal velocity components and the angular momentum, per unit of mass, of the motion can be expressed as

$$\dot{r} = V \sin \gamma \quad (5.23)$$

$$r \dot{\theta} = V \cos \gamma \quad (5.24)$$

$$H = r V \cos \gamma \quad (5.25)$$

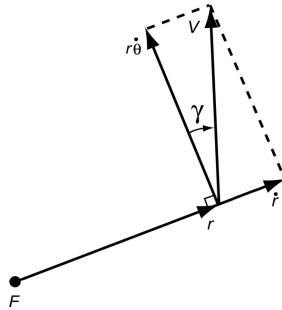


Figure 5.5: The flight path angle, γ , and the radial and normal velocity components.

Differentiation of (5.22) to time yields another expression for the radial velocity component:

$$\dot{r} = \frac{(H^2/\mu) e \dot{\theta} \sin \theta}{(1 + e \cos \theta)^2} = r^2 \dot{\theta} \frac{\mu e}{H^2} \sin \theta = \frac{\mu}{H} e \sin \theta \quad (5.26)$$

An expression for the normal velocity component follows directly from (5.7) and (5.22):

$$r \dot{\theta} = \frac{H}{r} = \frac{\mu}{H} (1 + e \cos \theta) \quad (5.27)$$

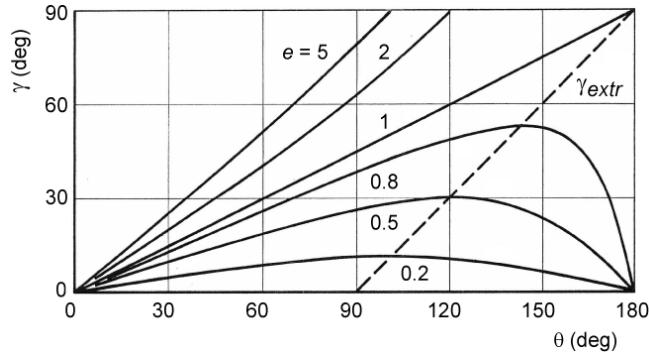
From (5.23), (5.24), (5.26) and (5.27) we obtain for the flight path angle

$$\tan \gamma = \frac{\dot{r}}{r \dot{\theta}} = \frac{e \sin \theta}{1 + e \cos \theta} \quad (5.28)$$

Since $-90^\circ \leq \gamma \leq 90^\circ$, γ is unambiguously defined by (5.28). This relation is depicted in Figure 5.6 for $0^\circ \leq \theta \leq 180^\circ$; for $180^\circ \leq \theta \leq 360^\circ$ a mirror image of the curve is obtained. The extreme value of γ (γ_{extr}) for a particular value of $e < 1$ is found from

$$\frac{\partial}{\partial \theta} \left[\arctan \left(\frac{e \sin \theta}{1 + e \cos \theta} \right) \right] = 0$$

We then obtain, after some algebraic manipulation, for γ_{extr} and the true anomaly for which γ_{extr} occurs ($\theta_{\gamma_{\text{extr}}}$):

Figure 5.6: The flight path angle, γ , as a function of the true anomaly, θ .

$$\theta_{\gamma_{extr}} = \arccos(-e) ; \quad \gamma_{extr} = \pm \arctan \left[\frac{e}{\sqrt{1 - e^2}} \right]$$

As $e \geq 0$, values of γ_{extr} and $\theta_{\gamma,extr}$ are only found for circular, elliptical and parabolic orbits. We then find $-90^\circ \leq \gamma_{extr} \leq 90^\circ$ and $90^\circ \leq \theta_{\gamma,extr} \leq 270^\circ$. Meaningful values of γ_{extr} and $\theta_{\gamma,extr}$ only exist for elliptical and parabolic orbits. For parabolic orbits: $\gamma_{extr} = 90^\circ, 270^\circ$; $\theta_{\gamma,extr} = 180^\circ$. For elliptical orbits we find that if $90^\circ < \theta_{\gamma,extr} < 180^\circ$ then $\gamma_{extr} = \theta_{\gamma,extr} - 90^\circ$; if $180^\circ < \theta_{\gamma,extr} < 270^\circ$ then $\gamma_{extr} = \theta_{\gamma,extr} - 270^\circ$. The first of these relations is plotted as a dashed line in Figure 5.6.

Elimination of θ from (5.26) and (5.27) yields

$$\dot{r}^2 + \left(r \dot{\theta} - \frac{\mu}{H} \right)^2 = \left(\frac{\mu e}{H} \right)^2 \quad (5.29)$$

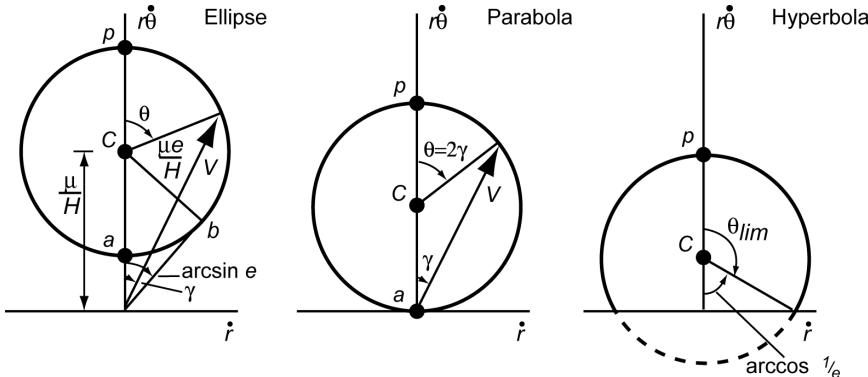


Figure 5.7: Velocity hodographs for elliptical, parabolic and hyperbolic motion.

We now draw a *velocity hodograph*, in which the radial velocity \dot{r} is plotted versus the normal velocity $r\dot{\theta}$ (Figure 5.7). According to (5.29), such a hodograph is a circle with radius $\mu e/H$ and with its center on the $r\dot{\theta}$ -axis at a distance μ/H from the origin. In these hodographs a number of quantities can be indicated: position of pericenter (p) and apocenter (a), true anomaly, θ , γ_{extr} and e , and the limit value θ_{lim} for a hyperbola. From Figure 5.7 we conclude that for a parabola: $\theta = 2\gamma$; this characteristic will be proved analytically in Chapter 7. The straight line corresponding to $e = 1$ in Figure 5.6 also suggests this linear relationship between γ and θ . These hodographs, and also *acceleration hodographs*, in which the acceleration components are plotted versus each other, have found application for a qualitative analysis of *optimal transfer orbits* and *rendez-vous orbits*.

The velocity of body i can also be resolved into a component perpendicular to the radius vector, V_n , and a component perpendicular to the axis of symmetry of the conic section, V_l . According to Figure 5.8, then the following relations hold:

$$V_l = \frac{\dot{r}}{\sin(\pi - \theta)} = \frac{\dot{r}}{\sin \theta} ; \quad V_n = r \dot{\theta} + \frac{\dot{r}}{\tan(\pi - \theta)} = r \dot{\theta} - \frac{\dot{r}}{\tan \theta}$$

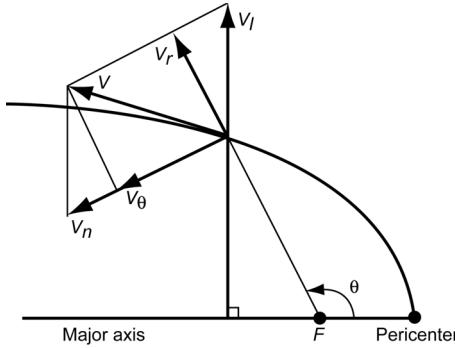


Figure 5.8: Definition of the velocity components V_l and V_n .

Substitution of (5.26) and (5.27) into these relations yields

$$V_l = \frac{\mu e}{H} ; \quad V_n = \frac{\mu}{H} \quad (5.30)$$

The expressions (5.30) show that for a Keplerian orbit the magnitude of both velocity components V_l and V_n is constant; the velocity component V_l also has a constant direction. At pericenter passage both components have the same direction, leading to a maximum value of the velocity; at apocenter passage V_n is directed opposite to V_l , leading to a minimum value of the velocity. The existence of these two constant velocity components is known as *Whittaker's theorem*, which was formulated by E.T. Whittaker (1873-1956) around 1904.

5.7. Eccentricity vector

In Section 5.2, a constant vector \bar{e}_4 was introduced that is directed from the focus of the conic section towards pericenter (Figure 5.2). In Section 5.3 it was shown that the magnitude of this vector is equal to the eccentricity, e , of the conic section. This vector is therefore generally called the *eccentricity vector*, \bar{e} .

An interesting expression can be derived from (5.17). Since \bar{H} is constant, we can write that equation as

$$\frac{d}{dt} \left[\frac{d\bar{r}}{dt} \times \left(\bar{r} \times \frac{d\bar{r}}{dt} \right) \right] = \mu \frac{d}{dt} \left(\frac{\bar{r}}{r} \right)$$

Evaluation of the vector triple-product gives

$$\frac{d}{dt} \left[\left(\frac{d\bar{r}}{dt} \cdot \frac{d\bar{r}}{dt} \right) \bar{r} - \left(\frac{d\bar{r}}{dt} \cdot \bar{r} \right) \frac{d\bar{r}}{dt} \right] = \mu \frac{d}{dt} \left(\frac{\bar{r}}{r} \right)$$

or

$$\frac{d}{dt} \left[\left(V^2 - \frac{\mu}{r} \right) \bar{r} - (\bar{r} \cdot \bar{V}) \bar{V} \right] = 0$$

Integration yields

$$\left(V^2 - \frac{\mu}{r} \right) \bar{r} - (\bar{r} \cdot \bar{V}) \bar{V} = \mu \bar{A} \quad (5.31)$$

where \bar{A} is a constant vector. This equation shows that in a Keplerian orbit velocity and position vary in such a way that if the local position and velocity vectors, and their magnitudes, are inserted into (5.31) a constant vector $\mu \bar{A}$ results. That vector lies in the orbital plane and its direction and magnitude can be determined as follows. At pericenter:

$$\bar{r}_p \cdot \bar{V}_p = 0 \quad ; \quad r_p V_p = H \quad ; \quad r_p = \frac{H^2/\mu}{1+e}$$

Substitution of these relations into (5.31) gives

$$e \frac{\bar{r}_p}{r_p} = \bar{A}$$

This means that \bar{A} points towards pericenter and has a magnitude e ; so, the vector \bar{A} is identical to the eccentricity vector, \bar{e} , mentioned above. In the literature, this vector is called the *first Laplace vector* or the *Laplace-Runge-Lenz vector*, which illustrates the fact that this vector was ‘rediscovered’ several times. J. Hermann was the first to show in 1710 that \bar{A} is conserved for an inverse-square central force field and worked out its connection to the eccentricity of an elliptical orbit. Hermann’s work was generalized by J. (Johann) Bernouilli in 1710. P.S. Laplace (1749-1827) rediscovered the conservation of \bar{A} in 1799, deriving it analytically rather than geometrically. W.R. Hamilton (1805-1865) derived the equivalent eccentricity vector in 1847, using it to show that, for motion in an inverse-square central force field, the velocity vector moves on a circle in a velocity hodograph (Figure 5.7). J.W. Gibbs (1839-1903) derived the eccentricity vector by vector analysis in 1901. Gibb’s derivation was used as an example by C.D.T. Runge (1856-1927) in a textbook published in 1919, which was referenced by W. Lenz (1888-1957) in an old quantum theory analysis of the hydrogen atom in 1924. In 1926, the eccentricity vector was used by W.E. Pauli (1900-1958) to derive the spectrum of the hydrogen atom from modern quantum mechanics theory. After this publication, the vector became generally known as the Runge-Lenz vector.

According to (5.31), we thus can write for the eccentricity vector:

$$\bar{e} = \frac{1}{\mu} \left[\left(V^2 - \frac{\mu}{r} \right) \bar{r} - (\bar{r} \cdot \bar{V}) \bar{V} \right] \quad (5.32)$$

This interesting expression for the motion in a Keplerian orbit will be used in Section 10.2 for an analysis on *regularization*.

From (5.32) we obtain

$$e^2 = \bar{e} \cdot \bar{e} = \frac{1}{\mu^2} \left[\left(V^2 - \frac{\mu}{r} \right)^2 r^2 - 2 \left(V^2 - \frac{\mu}{r} \right) (\bar{r} \cdot \bar{V})^2 + (\bar{r} \cdot \bar{V})^2 V^2 \right]$$

Substitution of $\bar{r} \cdot \bar{V} = r V \sin \gamma$ (Figure 5.5) leads after some algebraic manipulation to

$$e^2 = 1 - \frac{rV^2}{\mu} \left(2 - \frac{rV^2}{\mu} \right) \cos^2 \gamma \quad (5.33)$$

This is a very important relation between the (constant) eccentricity of a Keplerian orbit and the continuously varying quantities r , V and γ that describe the instantaneous position and velocity of body i in that orbit. We will use this relation several times in other Chapters.

5.8. Stability of Keplerian orbits

In Sections 5.2 and 5.3 it has been shown that the trajectory of a body in a central force field with potential $-\mu/r$ is a conic section. One usually assumes implicitly that this trajectory is stable. However, when discussing the topic of ‘stability’, one should be aware that there exist numerous definitions of various kinds of stability, often leading to contradictory conclusions.

An elliptical two-body orbit is generally considered stable, since a small change of the initial conditions will not significantly change the shape and orientation of the orbit. This kind of stability, known as *orbital stability*, is intuitively clear. Consider, on the other hand, the same elliptical orbit and let us again change the initial conditions slightly so that the semi-major axis will change ever so slightly. The orbital stability is still valid, but the change of the semi-major axis will result in a change of the orbital period (Section 6.3). The resulting small change in the mean angular motion (Section 6.3) will displace the body along the orbit. After a sufficiently large number of revolutions, the perturbed body might be close to apocenter at the time the body on its original orbit will be at pericenter (Section 10.1). The two orbits will be very close, but the distance between the body on the original orbit and the body on the slightly changed orbit will be the length of the major axis. This behavior certainly cannot be considered stable, in spite of the fact that the original and the new orbits are close. J.H. Poincaré (1854-1912), using the first, geometric idea, calls the motion stable, while A.M. Lyapunov (1857-1918), using the second, kinematic idea, considers the motion unstable. Poincaré’s definition refers to *orbital stability*; Lyapunov’s is usually referred to as *isochronous (equal time or simultaneous) stability*; it is also called *dynamical stability*.

The question on the stability of our solar system is an important issue in celestial mechanics. In Newton’s mind the solar system was unstable. He thought that gravitation alone could not account for its stability and that the intervention of divine forces was necessary for preserving the regular planetary motion around the Sun. Newton, however, never approached the stability question with mathematical tools. Euler was the first to apply perturbations theory to the motion of planets. His works opened the way for the research of P.S. Laplace (1749-1827) and J.L. Lagrange (1736-1813) on the stability of the solar system. According to Laplace, stability requires that the semi-major axes of the planetary orbits show no secular changes but only small periodic changes, so that the orbits do not intersect. Another, similar definition connects planetary stability with no collisions and no escapes. It took until 1950 before D. Brouwer (1902-1966) and A.J.J. van Woerkom (1915-1991) proved that if the semi-major axes of the planetary orbits are assumed bounded and without secular trends, then the analytical solution for the long-term behavior of the other orbital elements of all major planets is also bounded and represented by purely oscillatory functions. In 1773, Laplace proved that, up to the second powers of the eccentricities, the major axes, and consequently the mean motions of the planets, have no secular terms. This theorem was extended by Lagrange in 1774 and 1776 to all powers of the eccentricities and of the sine of the mutual inclinations, for perturbations of the first order with respect to the masses. S.D. Poisson (1781-1840) proved in 1809 that the major axes have no purely secular terms in the perturbations of the second order with respect to the masses. The

computations of Poisson were considerably simplified by J. Liouville (1809-1882) and V.A. Puiseux (1820-1883) in 1841, and then by F.F. Tisserand (1845-1896) in 1876. However, S.C. Haretu (1851-1912) proved in 1877 that there are secular variations in the expressions for the major axes to third-order approximation. It is interesting to note that Poisson himself discussed the third-order approximation in a paper published in 1816. He ignored, however, several aspects of the problem and concluded that no secular terms show up. Moreover, he trusted that this was true for any order of approximation and aimed to find a general proof of this fact. Haretu's result may point towards instability and eventual disruption of the solar system, but most experts conjectured that the third-order secular term is just one of the many other secular terms in the higher orders, which, all taken together, sum to a periodic function. Using a new analytical approach, P.J. Message (1931-2008) has shown in 1976 that no secular terms exist in the semi-major axes to any order. S. Newcomb (1835-1909) has established by 1876 that it is possible, in the case of the planetary perturbations, to represent the elements by purely periodic functions of time that formally satisfy the differential equations of motion. If these series were convergent, the stability of the solar system would be assured; but Poincaré has shown in 1892 that they are in general divergent. A. Lindstedt (1854-1939) and H. Gyldén (1841-1896) have also succeeded in integrating the equations of the motion of n bodies in periodic series, which were shown to be in general divergent. However, more recently, other researchers have shown that if certain non-resonant conditions are satisfied and if the perturbing masses are small enough, some of the series are actually convergent and give rise to a rigorous description of solutions of the planetary problem valid for all time.

In the last decades, J. Laskar (1955-) and others have performed extensive numerical simulations of the long-term behavior of our solar system. These simulations have indicated that over the next 5 billion years (5×10^9 yr), which is a period that is about equal to the present age of our solar system, the motion of the large planets is very regular; e.g. the escape time of Uranus was found to be substantially longer than the lifetime of our Sun. However, the change of the inner planets' orbits is significantly larger for this period. The results indicate that the orbits of Venus and Earth may change by more than 0.03 in eccentricity over the next 5 billion years; for Mars the changes may be about 0.08 in eccentricity. For Mercury the change in eccentricity can be so large that ejection of this planet out of the solar system due to a close encounter with Venus is possible in less than 3.5 billion years. Fortunately, this is not fatal to the global stability of the whole planetary system owing to the small mass of Mercury. It was also found that the Earth plays a crucial role in the long-term stability of the orbits of the inner planets. In the absence of Earth the orbits of Venus and Mercury would be heavily exposed to strong destabilizing resonances with giant planets. The main conclusion from these simulations is that the solar system seems to be in a state of 'relative' stability, in the sense that we should not expect major changes in the motion of the planets for the next few hundred million years and that strong instabilities (collision or escape) can only occur on a timescale of some billions of years. However, the final question of the stability of the solar system remains still unsolved.

In the early stage of the solar system, some extra inner planets may have existed. However, if this were the case then their existence would have resulted in a much more unstable planetary system, leading to close encounters or collisions between planets and to the escape of these extra planets. This then has led to the present much more stable solar system. The organization of the inner planetary system is thus most probably due to its long-term orbital evolution, and not uniquely to its relatively rapid (less than 100 million years) formation process. This means that the solar system at the end of its formation process may have been significantly different from the present one, and has since then evolved toward the present configuration because of the gravitational instabilities. It is even quite possible that such quasi-stable planetary systems are

the natural ‘end product’ in the development of any planetary system from the disk of gas and dust surrounding a newly born star.

As an example, we will now analyze the stability of a circular orbit in a central force field, which produces a force (per unit of mass) on the spacecraft that can be expressed by

$$F(r) = -\alpha r^n \quad (5.34)$$

where α is an arbitrary positive constant and n is also a constant. The minus-sign indicates that we are dealing with an attractive force. Note that for the Newtonian gravity field $\alpha = \mu$, $n = -2$, where μ is the gravitational parameter defined by (5.4). With (5.34), the equations of motion (5.7) and (5.10) can be written as

$$\begin{aligned} \ddot{r} - r \dot{\varphi}^2 &= F(r) \\ r^2 \dot{\varphi} &= H \end{aligned} \quad (5.35)$$

Suppose that the spacecraft is initially moving in a circular orbit with radius r_0 . In this orbit $\ddot{r}_0 = 0$; so, according to (5.35-1),

$$r_0 \dot{\varphi}_0^2 = -F(r_0)$$

where the index 0 indicates the state in the initial orbit. Combination of this relation and (5.35-2) yields

$$H_0^2 = r_0^4 \dot{\varphi}_0^2 = -r_0^3 F(r_0) \quad (5.36)$$

Now, assume that the body experiences a small perturbation in the radial direction, for example due to a small rocket impulsive shot. Just after that impulsive shot the spacecraft’s position has not changed (Section 1.7), but a radial velocity component has been added to the body’s original velocity vector; this original velocity vector is directed normal to the position vector (circular orbit). So, the radial impulsive shot has not affected the spacecraft position and normal velocity component and thus the body’s angular momentum has not changed by the impulsive shot. When the angular momentum, the radial position and the position in along-track direction after the perturbation are indicated by H , r and φ , then we can write

$$H = H_0 \quad ; \quad r = r_0 + \Delta r \quad ; \quad \varphi = \varphi_0 + \Delta\varphi \quad (5.37)$$

where r_0 and φ_0 denote the radial and along-track position components in the unperturbed orbit.

First, the motion in the along-track direction is considered. Combining (5.35-2) and (5.37) yields

$$\dot{\varphi} = \frac{H_0}{r^2}$$

If second- and higher-order terms are neglected (linearization), substitution of (5.37) into this relation gives

$$\dot{\varphi}_0 + \Delta\dot{\varphi} = \frac{H_0}{r_0^2} \left(1 - 2 \frac{\Delta r}{r_0} \right)$$

With (5.36), this equation can be written as

$$\Delta\dot{\phi} = -2 \left\{ -\frac{F(r_0)}{r_0^3} \right\}^{1/2} \Delta r$$

from which, after integration, follows:

$$\Delta\phi = -2 \left\{ -\frac{F(r_0)}{r_0^3} \right\}^{1/2} (\Delta r) t$$

Substitution of (5.34) finally yields

$$\Delta\phi = -2 \left[\alpha r_0^{n-3} \right]^{1/2} (\Delta r) t \quad (5.38)$$

This equation shows that for any value of n , a value of $\Delta r \neq 0$ will always lead to a linear increase of $|\Delta\phi|$ with time. For $\Delta r > 0$, the spacecraft will be behind the position it would have had in its original orbit at the same instant of time; for $\Delta r < 0$ the spacecraft will be ahead of the corresponding position in its original orbit. So, the motion in the along-track direction is unstable in the sense that a radial perturbation yields an ever-increasing value of $|\Delta\phi|$. For a Newtonian gravity field with potential $-\mu/r$, (5.38) reads

$$\Delta\phi = -2 \left(\frac{\mu}{r_0^5} \right)^{1/2} (\Delta r) t \quad (5.39)$$

This dynamical instability leads to a *numerical instability* in orbit computations (Section 10.1).

Now, we will analyze the motion in the radial direction. Substitution of (5.35-2) and (5.37) into (5.35-1) gives

$$\ddot{r} - \frac{H_0^2}{r^3} = F(r)$$

Substitution of (5.37) and application of a Taylor series expansion:

$$F(r) = F(r_0) + F'(r_0) \Delta r + \dots$$

where $F'(r_0)$ is defined as

$$F'(r_0) = \left[\frac{dF(r)}{dr} \right]_{r_0}$$

lead, after linearization, to

$$\ddot{r}_0 + \Delta\ddot{r} - \frac{H_0^2}{r_0^3} \left(1 - 3 \frac{\Delta r}{r_0} \right) = F(r_0) + F'(r_0) \Delta r$$

Subsequent substitution of (5.36) leads to

$$\Delta\ddot{r} - \left\{ 3 \frac{F(r_0)}{r_0} + F'(r_0) \right\} \Delta r = 0 \quad (5.40)$$

This is a well-known type of a second-order differential equations. For

$$3 \frac{F(r_0)}{r_0} + F'(r_0) < 0 \quad (5.41)$$

the solution for Δr is a pure harmonic oscillation, which corresponds to stable motion. If the left-hand side of the inequality (5.41) is larger than zero, the radial motion after the perturbation will be unstable and $|\Delta r|$ will increase exponentially with time. Substitution of (5.34) into (5.41) gives as a condition for stable motion:

$$(3 + n) \alpha r_0^{n-1} > 0$$

Since α is a positive constant, this condition leads to the requirement

$$n > -3 \quad (5.42)$$

So, a circular orbit is only ‘geometrically’ stable if $n > -3$. In Section 5.4 it was mentioned that J. (Johann) Bernouilli proved that for $n = -3$ the orbit is a spiral, which certainly is ‘geometrically’ unstable. For a Newtonian gravity field $n = -2$. Hence, a circular orbit in a Newtonian gravity field is a stable orbit; however, the motion along this orbit is unstable! In Section 10.1 the topic of orbital stability will be addressed in a more general way.

5.9. Roche limit

Suppose that a spherical body P_2 with mass m_2 and radius R_2 is moving in a circular orbit about a spherical body P_1 with mass m_1 and radius R_1 , where $m_2 \ll m_1$ (Figure 5.9). Furthermore, assume that both bodies are rigid and have a radially-symmetric mass distribution. This means that they do not experience tidal deformations and that, as far as their gravitational attraction is concerned, they may be considered as point masses (Section 1.5). On the surface of body P_2 a point mass P_3 with mass $m_3 \ll m_2$ is located on the line connecting the centers of bodies P_1 and P_2 (Figure 5.9).

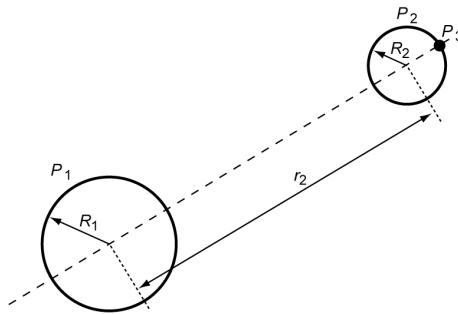


Figure 5.9: Definition of the parameters used in the analysis of the Roche limit.

Because of its small mass, P_3 has no effect on the motion of P_2 about P_1 . Since we have assumed that P_2 moves in a circular orbit about P_1 , and $m_2 \ll m_1$, we may write according to (5.4) and (5.10)

$$\dot{\varphi}_2^2 = \frac{G m_1}{r_2^3} \quad (5.43)$$

We assume that P_3 maintains its position on the line connecting the centers of P_1 and P_2 . This implies that we assume that P_2 completes one revolution about its axis in the same period in

which it completes one revolution about P_1 . This is the case for the Moon and for any moon that is tidally locked to its primary (Section 17.1). So, the angular velocity of P_3 about P_1 equals the orbital angular velocity of P_2 as expressed by (5.43), and the (rotational) angular velocity of P_2 about its axis is also given by (5.43). We now will derive an expression for the radial acceleration of P_3 relative to P_2 .

For the centrifugal acceleration of P_3 due to the rotation of P_2 about its axis we may write

$$a_{\text{rot}} = \frac{G m_1}{r_2^3} R_2 \quad (5.44-1)$$

At the position of P_3 the gravitational attraction by P_1 is somewhat less than at the center of P_2 . This difference produces a tidal force on P_3 , as explained in Section 4.2 and Section 17.1. This force is, for the geometry depicted in Figure 5.9, directed outward. The associated acceleration can be computed by applying the method discussed in Section 4.2. Now, body k indicated in Figure 4.2 is P_2 , body d is P_1 and body i is P_3 . Using the notation from Figure 5.9, we then obtain for the differential gravitational (tide inducing) acceleration from (4.9) and (4.17):

$$a_{\text{tide}} = 2 \frac{G m_1}{r_2^3} R_2 \quad (5.44-2)$$

For the acceleration of P_3 due to the gravitational attraction between body P_2 and point mass P_3 we may write

$$a_{\text{grav}} = \frac{G m_2}{R_2^2} \quad (5.44-3)$$

For the geometry depicted in Figure 5.9, this acceleration is directed inward. We now define the relative radial acceleration of P_3 with respect to P_2 as $a_{\text{rel}} = a_{\text{rot}} + a_{\text{tide}} - a_{\text{grav}}$. Substitution of (5.44) into this relation yields

$$a_{\text{rel}} = G \left(3 \frac{m_1 R_2}{r_2^3} - \frac{m_2}{R_2^2} \right) \quad (5.45)$$

If $a_{\text{rel}} < 0$, P_3 will be pushed against P_2 and will remain on the surface of P_2 . However, if $a_{\text{rel}} > 0$, then the resulting force on P_3 will be directed away from the surface of P_2 . Consequently, when P_3 is e.g. a boulder on the surface of a moon of a planet, $a_{\text{rel}} > 0$ means that the boulder may leave the surface of that moon. Or, in general, when P_2 consists of material with little cohesion and therefore with a low internal tensile strength, this may result in a gradual disintegration of that body. The transit from ‘stability’ to ‘disintegration’, which is defined as the *Roche limit*, occurs at $a_{\text{rel}} = 0$. For this Roche limit we find from (5.45)

$$(r_2)_{\text{Ro}} = \sqrt[3]{3 \frac{m_1}{m_2} R_2}$$

If the mean densities of bodies P_1 and P_2 are indicated by ρ_1 and ρ_2 , respectively, we find

$$(r_2)_{\text{Ro}} = \sqrt[3]{3 \frac{\rho_1}{\rho_2} R_1} = 1.44 \sqrt[3]{\frac{\rho_1}{\rho_2} R_1} \quad (5.46)$$

This expression was derived using a number of simplifying assumptions. E. Roche (1820-1883) applied around 1846 a more-complex physical model where body P_2 consists of an incompressible fluid with negligible bulk tensile strength. He then arrived at the following expression:

$$(r_2)_{Ro} = 2.44 \sqrt[3]{\frac{\rho_1}{\rho_2}} R_1 \quad (5.47)$$

Note that, for a given planet, a moon with a lower mean density will have a larger Roche limit. If $r_2 < (r_2)_{Ro}$, then $a_{rel} > 0$, which means that body P_2 is *Roche-unstable*; if $r_2 > (r_2)_{Ro}$, then body P_2 is *Roche-stable*. The fluid solution (5.47) is appropriate for bodies that are only loosely held together, such as comets. For instance, comet Shoemaker-Levy 9 was first observed in 1993, and its orbit indicated that it had been captured by Jupiter a few decades prior. Its decaying orbit around Jupiter passed within Jupiter's Roche limit in July 1992, causing it to fragment into a number of smaller pieces. On its next approach in July 1994 the fragments crashed into the planet's atmosphere (Section 18.11). In the Earth-Moon system the bodies may be considered solid. For this system holds: $R_1 = 6371$ km, $\rho_1 = 5515$ kg/m³, $\rho_2 = 3341$ kg/m³, and we find from (5.46): $(r_2)_{Ro} = 10,859$ km. The Moon is located at an average distance of $r_2 = 384,401$ km; so, $r_2 \gg (r_2)_{Ro}$ and thus the Moon is Roche-stable and will not disintegrate. It turns out that most moons in the solar system are Roche-stable. Some moons, such as Mars' moon Phobos, Jupiter's moon Amalthea, Saturn's moon Pan, Uranus' moon Cordelia and Neptune's moon Naiad, move within their Roche limit but are held together because of their tensile strength. Within the Roche limit, no satellite can coalesce out of smaller particles. Indeed, almost all known planetary rings are located within their Roche limit; Saturn's E ring and Phoebe ring being notable exceptions. They could either be remnants from the planet's proto-planetary accretion disc that failed to coalesce into moonlets, or conversely have formed when a moon passed within its Roche limit and broke apart.

Let us consider a hypothetical spherical space station in orbit about the Earth, tidally locked to the Earth and with a mass and a volume equal to the mass and volume of the European Columbus module that forms an essential element of the International Space Station (ISS; Section 15.1). We then find for the mean mass density of that space station: $\rho_2 = 180$ kg/m³, and obtain from (5.46) for the Roche limit: $(r_2)_{Ro} = 28,752$ km. This means that only when that space station would encircle the Earth at a distance larger than 28,752 km from the Earth's center, loose objects on that station (and astronauts) would have the tendency to remain near the station. Since most spacecraft orbit the Earth at much lower altitudes, usually loose objects on spacecraft will move away from that spacecraft and will become space debris.

5.10. Relativistic effects

According to the *general relativity theory*, which was formulated by A. Einstein (1879-1955) around 1905, Newton's law of gravitation is but a first, although very good, approximation and the external gravity field of a body k with spherically symmetric mass distribution is described by the so-called *Schwarzschild metric* (K. Schwarzschild (1873-1916)). When this metric is used, we find that, if $m_i \ll m_k$, a good approximation for the relativistic motion of body i about body k is given by

$$\frac{d^2 u}{d\varphi^2} + u = \frac{\mu}{H^2} + 3 \frac{\mu}{c^2} u^2 \quad (5.48)$$

where $u = 1/r$, c is the speed of light and H denotes the classical angular momentum of body i .

This equation only holds when the orbit of body i is not exactly circular. By comparing (5.14) and (5.48) we conclude that the first-order *relativistic effect* is expressed by the term $3\mu u^2/c^2$. The right-hand side of (5.48) can be written as

$$\frac{\mu}{H^2} \left(1 + 3 \frac{H^2 u^2}{c^2} \right) = \frac{\mu}{H^2} \left(1 + 3 \frac{V_\varphi^2}{c^2} \right)$$

where V_φ is the normal velocity component of body i . In celestial mechanics, this component is always much smaller than the speed of light, and we therefore conclude that the relativistic effect on the motion of planets and spacecraft is always very small. If a quantity α , which is defined as

$$\alpha = 3 \frac{\mu}{c^2} \quad (5.49)$$

is introduced, (5.48) can be written as

$$\frac{d^2 u}{d\varphi^2} + u = \frac{\mu}{H^2} + \alpha u^2 \quad (5.50)$$

In contrast to (5.14), this weakly non-linear second-order differential equation cannot be solved analytically in a closed form. However, it is possible to obtain an approximative solution. In this Section, we will derive that solution by applying the *method of successive approximations* that is often used in celestial mechanics.

A zeroth-order approximation of the solution of (5.50) can be found by neglecting the small term αu^2 with respect to μ/H^2 . In that case, the solution is the equation for Keplerian motion (Section 5.2):

$$u = \frac{\mu}{H^2} [1 + e \cos(\varphi - \omega)] \quad (5.51)$$

which describes a conic section with eccentricity e and argument of pericenter ω . Subsequently, this expression for u is substituted into the right-hand side of (5.50), which yields the differential equation:

$$\frac{d^2 u}{d\varphi^2} + u = \frac{\mu}{H^2} + \alpha \frac{\mu^2}{H^4} \left[1 + \frac{1}{2} e^2 + 2e \cos(\varphi - \omega) + \frac{1}{2} e^2 \cos 2(\varphi - \omega) \right] \quad (5.52)$$

It is emphasized that substitution of (5.51) into the second term on the left-hand side of (5.50) is not allowed, because that term does not contain a small multiplier. The differential equation (5.52) can be solved analytically and yields a first-order approximation of the solution of (5.50). The homogeneous equation associated with (5.52) reads

$$\frac{d^2 u}{d\varphi^2} + u = 0$$

and has the solution

$$u = c_1 \cos(\varphi - c_2)$$

where c_1 and c_2 are arbitrary constants. A particular solution of (5.52) is

$$u = \frac{\mu}{H^2} + \alpha \frac{\mu^2}{H^4} \left[1 + \frac{1}{2}e^2 + e\varphi \sin(\varphi - \omega) - \frac{1}{6}e^2 \cos 2(\varphi - \omega) \right]$$

Hence, the complete solution of (5.52) is

$$u = c_1 \cos(\varphi - c_2) + \frac{\mu}{H^2} + \alpha \frac{\mu^2}{H^4} \left[1 + \frac{1}{2}e^2 + e\varphi \sin(\varphi - \omega) - \frac{1}{6}e^2 \cos 2(\varphi - \omega) \right]$$

In case $\alpha = 0$, this solution, of course, has to be identical to (5.51). With this requirement we can determine the values of the constants c_1 and c_2 , and find for the first-order approximation of the solution of (5.50)

$$u = \frac{\mu}{H^2} [1 + e \cos(\varphi - \omega)] + \alpha \frac{\mu^2}{H^4} \left[1 + \frac{1}{2}e^2 + e\varphi \sin(\varphi - \omega) - \frac{1}{6}e^2 \cos 2(\varphi - \omega) \right] \quad (5.53)$$

If we compare this solution with (5.51) we conclude that the relativistic effect consists of:

- An increase of u by a constant value of $\alpha\mu^2(1+e^2/2)/H^4$.
- A fluctuating term, of which the amplitude continuously increases with increasing values of φ .
- A pure oscillation with a constant amplitude of $\alpha\mu^2e^2/6H^4$.

The first contribution essentially comes down to a change of the scale on which distances are measured, and will be hardly noticeable, if at all. The third contribution is a pure oscillation with a small amplitude. The term with the ever-increasing amplitude will therefore dominate the long-term relativistic effect. So, for an analysis of the long-term relativistic effect we may approximate (5.53) by

$$u = \frac{\mu}{H^2} [1 + e \cos(\varphi - \omega) + \beta e \varphi \sin(\varphi - \omega)] \quad (5.54)$$

where

$$\beta = \alpha \frac{\mu}{H^2} \quad (5.55)$$

In celestial mechanics, the value of β is always very small (for the planets: $\beta < 9.6 \cdot 10^{-8}$; for Earth satellites: $\beta < 2.1 \cdot 10^{-9}$). So, even for large values of φ : $\beta\varphi \ll 1$. For example, if we consider the planet Mercury then we find that after 48,000 years: $\beta\varphi = 0.1$; for a low-altitude Earth satellite this value is reached after about 1300 years. So, for long periods of time it is allowed to write

$$\cos \beta \varphi \approx 1 \quad ; \quad \sin \beta \varphi \approx \beta \varphi$$

and (5.54) can, in good approximation, be written as

$$u = \frac{\mu}{H^2} [1 + e \cos(\varphi - \omega - \beta \varphi)] \quad (5.56-1)$$

or

$$r = \frac{H^2/\mu}{1 + e \cos(\varphi - \omega - \beta \varphi)} \quad (5.56-2)$$

This equation demonstrates that the motion of body i about body k may be interpreted as a conic

section of which the instantaneous argument of pericenter is given by $\omega + \beta\varphi$. Hence, it is a conic section of which the major axis slowly rotates. This motion resembles a rosette (Figure 5.10).

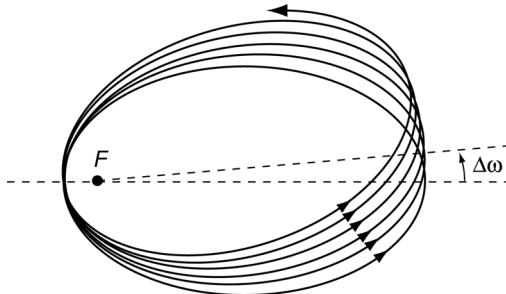


Figure 5.10: Relativistic precession of the orbit's major axis.

Note that the spatial orientation of the orbital plane, and the size and shape of the orbit in the orbital plane do not change. After a full revolution of body i in its orbit about body k , i.e. $\varphi_1 = \varphi_0 + 2\pi$, the major axis has rotated over an angle of $2\beta\pi$ in the direction of motion of body i . Hence, the change in the argument of pericenter per revolution can be approximated by

$$\Delta\omega = 2\pi\alpha \frac{\mu}{H^2} = 6\pi \frac{\mu^2}{H^2 c^2} \quad (5.57)$$

This relation shows that the relativistic effect is larger when the gravity field of body k is stronger and when the angular momentum of the motion of body i is smaller. For low-eccentricity orbits we find $H^2 \approx \mu r$, where r is the mean radius of the orbit. In that case, the right-hand side of (5.57) is about equal to $6\pi\mu/(rc^2)$, which shows that for near-circular orbits about body k the relativistic effect is larger if the mean radius of the orbit of body i is smaller.

The result (5.57) is a famous verification of the correctness of Einstein's *general relativity theory*. If the perturbations of all planets on the motion of Mercury about the Sun are taken into account according to Newton's law of gravitation, it is found that the perihelion of Mercury should precess by an amount of $532''$ per century. This 'modern' value is close to the value of $527''$ per century that was already determined by U.J.J. Le Verrier (1811-1877) around 1865. However, it appeared from observation that in reality the perihelion of Mercury precesses $574''$ per century; the difference of $42''$ per century could not be explained by classical mechanics. When the parameters of the Sun and Mercury are inserted into (5.57), we obtain a value of $43''$ per century for the relativistic precession of the perihelion; this is just the difference that could not be explained by Newtonian mechanics. Such a relativistic precession, of course, also occurs for the motion of the other planets about the Sun, of natural satellites about planets, and of satellites about the Earth (Table 5.1). The relativistic precession of e.g. the perihelion of the Earth's orbit is equal to $3.8''$ per century; the relativistic pericenter precession of a satellite with perigee at 800 km and apogee at 1000 km altitude above the Earth's surface is $1210''$ per century, and of Jupiter's inner moon Amalthea $2211''$ per century. Of all natural bodies of the solar system, Mercury has the largest relativistic pericenter precession per orbital revolution: $0.104''$. Although the relativistic perigee precession for the satellite about the Earth is much larger than that for Mercury, the relativistic perigee precession of Earth satellites is hardly observable. The reason is that numerous perturbing forces act on Earth satellites (Chapters 20 to 23) and that these forces can only be modeled to a limited accuracy. This limited modeling accuracy and the inherent dynamical instability of Keplerian orbits (Sections 5.8 and 10.1) will result in argument of perigee errors that are much larger than the relativistic perigee precession.

From these results we conclude that in classical celestial mechanics as well as in satellite

Table 5.1: Relativistic pericenter precession for some planets, the Moon and an Earth satellite.

| Body | $\Delta\omega$ ("/century) | Body | $\Delta\omega$ ("/century) |
|---------|-------------------------------|------------------|-------------------------------|
| Mercury | 42.98 | Jupiter | 0.062 |
| Venus | 8.62 | Saturn | 0.014 |
| Earth | 3.84 | Moon | 0.06 |
| Mars | 1.35 | Earth satellite* | 1210 |

* perigee altitude 800 km, apogee altitude 1000 km.

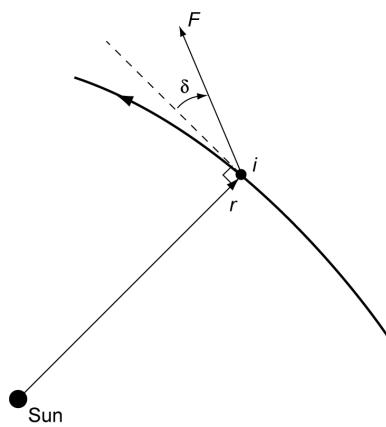
in orbit mechanics, the relativistic effect is very small and we can therefore certainly use Newton's law of gravitation and his laws of motion to compute the trajectories of these bodies.

5.11. Solar radiation pressure and the Poynting-Robertson effect

An interesting classical problem in celestial mechanics is the motion of a body i moving about the Sun, under the combined action of the Sun's gravitational force and the radiation force produced by sunlight. The starting point for this analysis are again the equations of motion (5.7) and (5.10):

$$\begin{aligned} \ddot{\vec{r}} - r \dot{\phi}^2 \vec{r} &= -\frac{\mu}{r^2} + \frac{\bar{F}}{m} \sin \delta \\ \frac{d}{dt}(r^2 \dot{\phi}) &= \frac{\bar{F}}{m} r \cos \delta \end{aligned} \quad (5.58)$$

where on the right-hand side of the equations the two components of the radiation force, \bar{F} , acting on the body have been added; δ is the angle between this force and the normal to the radius vector \vec{r} (Figure 5.11). By using (5.58), it is assumed implicitly that the force \bar{F} lies in the plane of motion of body i .

Figure 5.11: Definition of δ in the analysis of the Poynting-Robertson effect.

From quantum mechanics, it is known that light may be considered as a stream of photons, each moving with the speed of light ($c = 299,792.5$ km/s) and possessing an energy

$$\mathcal{E}_{ph} = h v \quad (5.59)$$

where h is *Planck's constant* ($h = 6.626 \times 10^{-34}$ kg m²/s) and ν is the frequency of the radiation. This frequency is related to the wavelength, λ , of the radiation by

$$\nu = \frac{c}{\lambda}$$

The momentum carried by each photon is

$$p_{ph} = \frac{h\nu}{c}$$

or, with (5.59),

$$p_{ph} = \frac{\mathcal{E}_{ph}}{c} \quad (5.60)$$

Assume that a flat plate with area A is placed perpendicular to a parallel lightbeam and that the reflected photons also move perpendicular to the plate. Per time interval Δt , n photons will hit the plate and n^* photons will be reflected back. The total impulse transferred by the photons to the plate is then, according to (5.60), given by

$$\Delta p = (n + n^*) \frac{\mathcal{E}_{ph}}{c} \quad (5.61)$$

where it is assumed that the energy of the photons (wavelength of the radiation) has not changed by the interaction with the plate. If the radiation flux (also called power density) of the incoming beam (in W/m²) is expressed by W , the amount of energy transferred to the plate by the incoming photons in the time interval Δt is

$$n \mathcal{E}_{ph} = WA \Delta t$$

In the same way, we find for the energy leaving the plate in the same time interval:

$$n^* \mathcal{E}_{ph} = W^* A \Delta t$$

where W^* is the radiation flux of the reflected beam. Substitution of these relations into (5.61) gives

$$\Delta p = (W + W^*) \frac{A}{c} \Delta t \quad (5.62)$$

When we define the *reflection coefficient*, R , by

$$R = \frac{W^*}{W} \quad (5.63)$$

we can write (5.62) as

$$\Delta p = (1 + R) \frac{WA}{c} \Delta t \quad (5.64)$$

For a perfect *black body*: $R = 0$, while for an *ideal reflector*: $R = 1$.

According to Newton's second law, we may write for the force experienced by the plate:

$$\mathbf{F} = \frac{d\mathbf{p}}{dt}$$

or, with (5.64),

$$\mathbf{F} = \lim_{\Delta t \rightarrow 0} \frac{\Delta \mathbf{p}}{\Delta t} = (1 + R) \frac{WA}{c} \quad (5.65)$$

This force is oriented along the Sun-Earth vector. The term $(1+R)W/c$ is called *radiation pressure*. The fact that electromagnetic radiation exerts a pressure upon any surface exposed to it was deduced theoretically by J.C. Maxwell (1831-1879) in 1871 and was proven experimentally by P.N. Lebedev (1866-1912) in 1900. The radiation pressure produced by sunlight is very feeble and decreases with the square of the distance from the Sun; at the distance of Mercury from the Sun and $R = 0$ its magnitude is $30 \mu\text{N/m}^2$, at the distance of the Earth $4.6 \mu\text{N/m}^2$ and at the distance of Jupiter $0.17 \mu\text{N/m}^2$.

Real satellites are no flat plates. However, the outer surface of a satellite can be modeled by a large number of small flat surface elements. Each element has specific surface properties and solar radiation reflection characteristics, and the orientation of each element relative to the incoming solar radiation is known at any time. In principle, it is then possible to compute the solar radiation force on each surface element and to determine the total force acting on the satellite by adding the forces on all elements. In this analysis the reflection of solar radiation by a surface element towards another element, and the shadowing of an element by other elements have to be taken into account. It is noted that the total force does not necessarily act along the Sun-Earth vector. This implies the possibility of a solar sail (Section 21.5), where the solar radiation force is used to propel a spacecraft.

The detailed modeling of the overall solar radiation force is quite complicated and therefore often a very simple approximative model is used. The satellite is then represented by an equivalent flat plate perpendicular to the incoming solar radiation and the force is written as

$$\mathbf{F} = (1 + R^*) \frac{WA^*}{c} \quad (5.66)$$

where R^* is the *mean reflection coefficient* and A^* is the *effective cross-sectional area* of the satellite, and the force acts along the Sun-Earth vector. This approximation is especially valid when the satellite is equipped with large solar panels. When we introduce the satellite's *reflectivity*: $C_R = 1 + R^*$, we can write (5.66) as

$$\mathbf{F} = C_R \frac{WA^*}{c} \quad (5.67)$$

This relation can be used to compute the force produced by any electromagnetic radiation source. The reflectivity depends on the spectrum of the radiation, the shape of the object and the photon reflection mechanism (type of surface, roughness of surface, etc.). A value of $C_R = 0$ means that the object is transparent to the incoming radiation; a value of 1 means that all the radiation is absorbed (black body), and all the radiation pressure is transferred to the body; a value of 2 means that all the radiation is reflected and twice the radiation pressure is transferred to the body. The computation of C_R for a specified radiation spectrum and a particular body is quite difficult. This is especially true for satellites with a complex shape, with a constantly changing orientation, and of which the surface consists of various materials. Often, a value of $C_R = 1.2 - 1.4$ is used, if no additional information is available.

Just as the gravitational force, the power density of the Sun's radiation varies inversely proportional to the square of the distance from the Sun. We therefore can write (5.67) as

$$F = C_R W_S \left(\frac{R_S}{r} \right)^2 \frac{A^*}{c} \quad (5.68)$$

where R_S is the Sun's radius and W_S is the radiation flux of sunlight at $r = R_S$. We now assume that body i is spherical with radius R and mass density ρ . Then, $A^* = \pi R^2$, $m = 4\pi R^3 \rho / 3$, and we find for the quantity F/m in (5.58):

$$\frac{F}{m} = \frac{3}{4} \frac{C_R W_S R_S^2}{c \rho R} \frac{1}{r^2} = \frac{\alpha}{r^2} \quad (5.69)$$

The quantity α is a function of parameters of the Sun, and of the reflectivity, radius and density of body i ; for a given body, α is a constant.

When applying (5.58) and (5.69) for analyzing the motion of body i , two phenomena should be taken into account. First, if the body has a radial velocity component relative to the Sun, then the frequency v of the incoming sunlight is shifted through the *Doppler effect* to a frequency v' , given by

$$v' = v \left(1 - \frac{\dot{r}}{c} \right)$$

This effect, which was proposed by C.A. Doppler (1803-1853) in 1842, implies that the radiation flux intercepted by body i becomes

$$W' = W \left(1 - \frac{\dot{r}}{c} \right) \quad (5.70)$$

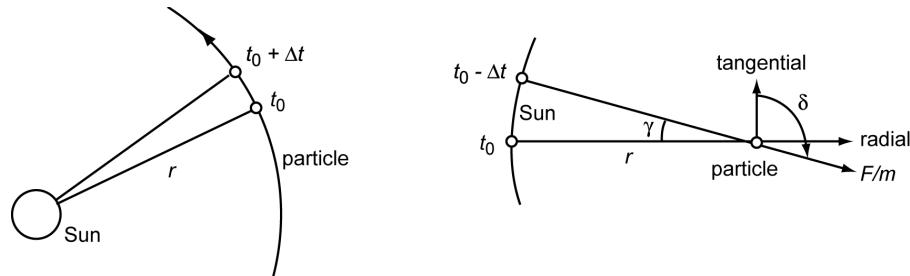


Figure 5.12: Motion of a particle relative to a heliocentric reference frame during the time interval Δt (left) and a visualization of the aberration effect (right).

A second phenomenon is related to the finite speed of light. This means that sunlight intercepted by body i at t_0 is actually emitted by the Sun at $t_0 - \Delta t$. During the time interval Δt light travels the distance $c\Delta t$, while the body has moved over a distance $r\dot{\phi}\Delta t$ in the direction normal to the direction to the Sun (Figure 5.12). This leads to the so-called *aberration* of the incoming sunlight. Light emitted by the Sun requires about 500 s (8.3 min) to arrive on the Earth, during which time the Sun moves about the Earth through an angle of about $20''$. The arriving sunlight shows us where the Sun was 8.3 min ago. The true, instantaneous position of the Sun is about $20''$ ahead of its visible position. In the same way, star positions are displaced from their yearly average position by up to $20''$, depending on the relative direction of the Earth's motion around the Sun. This aberration phenomenon was discovered by J. Bradley (1692-1762) in 1728. Since the

aberration of sunlight is very small, the aberration angle γ is approximately given by

$$\gamma \approx \frac{r \dot{\phi} \Delta t}{c \Delta t} = \frac{r \dot{\phi}}{c} \quad (5.71)$$

For the angle δ in Figure 5.12 we may write

$$\sin \delta = \sin(\frac{1}{2}\pi + \gamma) = \cos \gamma \quad ; \quad \cos \delta = \cos(\frac{1}{2}\pi + \gamma) = -\sin \gamma$$

or, with (5.71),

$$\sin \delta \approx 1 \quad ; \quad \cos \delta \approx -\gamma = -\frac{r \dot{\phi}}{c} \quad (5.72)$$

Combination of (5.69) and (5.70), and subsequent substitution of the result and (5.72) into (5.58) leads to

$$\begin{aligned} \ddot{r} - r \dot{\phi}^2 &= -\frac{\mu}{r^2} + \frac{\alpha}{r^2} \left(1 - \frac{\dot{r}}{c} \right) \\ \frac{1}{r} \frac{d}{dt}(r^2 \dot{\phi}) &= -\frac{\alpha}{r^2} \left(1 - \frac{\dot{r}}{c} \right) \frac{r \dot{\phi}}{c} \end{aligned}$$

Since for all bodies in our solar system $\dot{r} \ll c$, we may approximate these equations by

$$\begin{aligned} \ddot{r} - r \dot{\phi}^2 &= -\frac{\mu - \alpha}{r^2} \\ \frac{d}{dt}(r^2 \dot{\phi}) &= -\frac{\alpha \dot{\phi}}{c} \end{aligned} \quad (5.73)$$

These equations show that the action of sunlight effectively reduces the gravitational attraction force by the Sun, but also produces an additional term proportional to the circumferential velocity of the body. Because of the minus-sign, the latter term corresponds to a drag-type force. In connection with his investigations on radiation pressure, J. H. Poynting (1852-1914) established the existence of this force in 1903, using non-relativistic physics and Newton's theory of gravitation. He found that as a result of this force a spherical particle with a radius of 1 cm and a density of 5.5 gr/cm³, starting at the Earth's distance from the Sun and with the Earth's orbital velocity, could make at most some 10^8 revolutions about the Sun before falling into the Sun, while a particle with the same density but a radius of 10 μm could survive at most some 10^5 revolutions. These numbers correspond to periods of 39 million years and 39 thousand years, respectively. For the Earth, this effect shortens the length of the year by about $3 \cdot 10^{-8}$ s per year. A complete analysis of the effects of solar radiation force on the motion of small bodies was first performed by H.P. Robertson (1903-1961) in 1937. He re-analyzed the phenomenon discovered by Poynting, now using Einstein's special theory of relativity and Newton's theory of gravitation, and found that the solar radiation force also leads to an effective reduction of the solar gravitational attraction acting on the particle. The Newtonian approximation of the expressions obtained by Robertson are for $\dot{r} \ll c$ identical to (5.73). Below, the general characteristics of the motion described by (5.73) will be discussed.

With the definition of the angular momentum of motion (per unit of mass, (5.7)), integration of (5.73-2) leads for a constant value of α to

$$H = H_0 - \frac{\alpha \varphi}{c} \quad (5.74)$$

where H_0 is the angular momentum at $t = 0$, when $\varphi = 0^\circ$. This relation confirms that the body gradually loses angular momentum of motion and by that slowly spirals in towards the Sun and gains velocity. So, this mechanism leads to a gradual removal of particles from the solar system. Substitution of the values of all constants in the definition of α ((5.69)), we find

$$\alpha = 7.66 * 10^{16} \frac{C_R}{\rho R} \quad (5.75)$$

where R should be expressed in meter and ρ in kilogram per cubic meter to obtain α in cubic meter per second squared. This means that highly reflective low-density bodies, like ice crystals, are much more affected than e.g. iron particles, and that small particles are much more affected than planets.

Equation (5.73-1) can be written as

$$\ddot{r} - r \dot{\varphi}^2 = -\frac{\mu}{r^2} (1 - \beta) \quad (5.76)$$

where $\beta = \alpha/\mu$. This equation confirms that the radial force acting on the body is the difference between a gravitational attraction force directed towards the Sun and a radiation force directed away from the Sun. With (5.75) we conclude that the radiation force is larger for increasing values of the body's reflectivity, and for decreasing values of the radius and the density of the body. If $\beta < 1$, then the spiraling-in effect mentioned above will dominate the motion of the body. If $\beta > 1$, then the right-hand side of (5.76) will become positive and this repulsive force will dominate the motion. In that case, the particle will spiral out and will be blown out of the solar system eventually. So, this second mechanism also leads to a gradual removal of particles from the solar system. In Figure 5.13 the value of β is plotted as a function of the particle's radius, R , and its mass density, ρ , for $C_R = 1.2$. Note that for realistic values of $\rho = 400 - 6,000 \text{ kg/m}^3$, a value $\beta = 1$ is reached for $R = 0.2 - 2 \mu\text{m}$. We know that visible light corresponds to a wavelength interval of $0.38 - 0.75 \mu\text{m}$, and that the Sun radiates 93% of its power in the wavelength band $0.2 - 2 \mu\text{m}$ and 55% in the band $0.3 - 0.8 \mu\text{m}$. So, we conclude that low-density, highly reflective particles with a dimension of the order of the wavelength of solar visible radiation or smaller are blown out of the solar system by radiation pressure. It should be realized, however, that for these

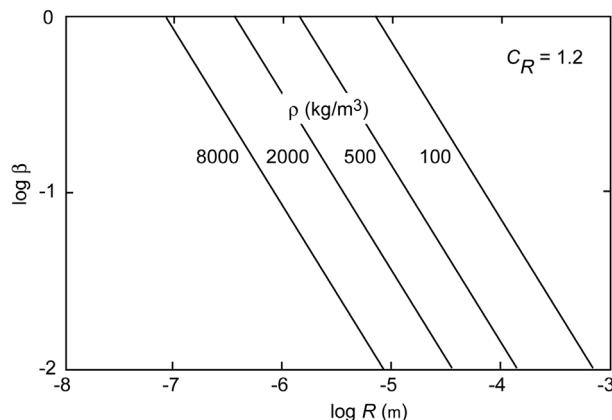


Figure 5.13: Relative solar radiation pressure as a function of the radius and the mean mass density of a particle.

particles α cannot be computed correctly by classical optical methods and that more-complicated light-scattering computation models must be applied. In fact, the phenomenon described above indicates that in the long history of the solar system the pressure of sunlight has effectively eliminated those dust particles that scatter sunlight too much. This means that this phenomenon is responsible for the fact that we see the Sun as a clear source of illumination and not as a diffuse ('misty') illumination source.

Now, assume that body i moves in the orbit of a planet about the Sun and that this orbit is (about) circular. In Section 6.2 it will be shown that the velocity of a body moving in a circular orbit in a Newtonian gravity field is given by

$$V = \sqrt{\frac{\mu}{r}}$$

In the case that the body is also subjected to the solar radiation force, it will be clear from (5.73-1) that the above equation should be modified to

$$V = \sqrt{\frac{\mu - \alpha}{r}}$$

This relation demonstrates that the radiation force reduces the orbital velocity of the body below that for the purely gravitational case. It was already mentioned that a small body will experience the effect of solar radiation pressure much stronger than a large body, e.g. a planet. This means that a particle moving in (about) the same orbit as a planet will have a non-zero circumferential velocity relative to that planet. Consequently, the planet will gradually clean up its region of motion from small particles orbiting the Sun. This effect was noted by Poynting in 1912, soon after the existence of the phenomenon of radiation pressure had been experimentally verified.

Because Poynting and Robertson were the first to analyze in detail all consequences of solar radiation pressure for the motion of a body about the Sun, the drag-type circumferential force, the effective reduction of the central gravitational force, and the relative velocity of a small body moving in the orbit of a planet with respect to that planet are nowadays collectively called the *Poynting-Robertson effect*.

6. ELLIPTICAL AND CIRCULAR ORBITS

In Section 5.3 it was shown that the orbit of body i about body k is a conic section with body k at a focal point, and that the equation for the orbit is

$$r = \frac{p}{1 + e \cos \theta} \quad (6.1)$$

For elliptical and circular orbits, which will be analyzed and discussed in this Chapter: $0 \leq e < 1$. From analytical geometry we know a number of characteristic parameters for an ellipse, which describe its shape and size, and a number of relations between these parameters, which will prove to be useful for our analysis. These will be briefly discussed in the next Section.

6.1. Geometry, energy and angular momentum

Figure 6.1 shows an ellipse and a number of useful parameters; the line AA' is an axis of symmetry of the ellipse. This axis is called the *major axis* of the ellipse and the length AA' is assigned a value of $2a$. The ellipse has a second axis of symmetry, which is perpendicular to AA' and passes through the middle of AA' . This axis is called the *minor axis* of the ellipse and the length BB' is assigned a value of $2b$.

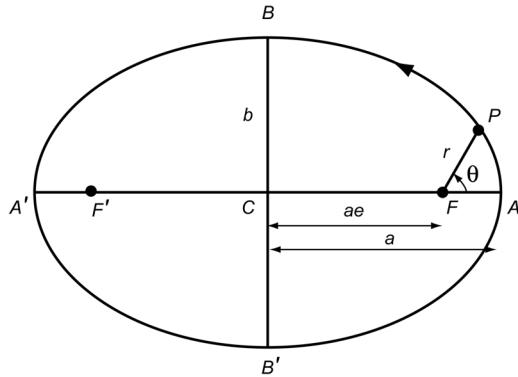


Figure 6.1: Geometry of an elliptical orbit.

The angle θ , measured from pericenter in the direction of motion, is the *true anomaly*, which was introduced in Section 5.3. From Figure 6.1 follows

$$2a = r_{\theta=0} + r_{\theta=\pi} = \frac{p}{1+e} + \frac{p}{1-e} = \frac{2p}{1-e^2}$$

or

$$p = a(1 - e^2) \quad (6.2)$$

Substitution of this relation into (6.1) gives

$$r = \frac{a(1 - e^2)}{1 + e \cos \theta} \quad (6.3)$$

This form of the equation for the orbit will be used frequently. The reason is that in (6.3) the semi-major axis, a , which indicates the size of the ellipse, and the eccentricity, e , which indicates the shape or ‘flattening’ of the ellipse, occur separately, while in (6.1) these quantities are com-

bined in the parameter p . The polar equation of elliptical motion ((6.3)) probably was first published by L. Euler (1707-1783) in 1740. In fact, Euler wrote the polar equation as $r = (a + b)(a - b)/(a + b \cos \theta)$, with a the semi-major axis, $b = ea$, and θ the true anomaly. A. Fontaine des Bertins (1704-1771) in 1739 seems to have arrived at a similar representation of the ellipse, independent of Euler.

For the pericenter and apocenter distance follows from (6.3):

$$r_p = r_{\theta=0} = a(1 - e) ; \quad r_a = r_{\theta=\pi} = a(1 + e) \quad (6.4)$$

from which the following relations can be derived:

$$a = \frac{r_a + r_p}{2} ; \quad e = \frac{r_a - r_p}{r_a + r_p} \quad (6.5)$$

The distance CF , i.e. the distance between the center of the ellipse and a focal point of the ellipse, can be written as

$$CF = a - r_p = ae \quad (6.6)$$

The distance of body i from the major axis varies with θ according to

$$s = r \sin \theta = a(1 - e^2) \frac{\sin \theta}{1 + e \cos \theta}$$

Differentiation of this equation with respect to θ and setting the result equal to zero, yields for the maximum distance and the correspondig value of θ :

$$s_{max} = b = a\sqrt{1 - e^2} ; \quad \cos \theta = -e \quad (6.7)$$

So, the crossing points of the orbit and the minor axis are specified by $\theta = \arccos(-e)$. In Section 5.6 we have found that at these values of θ the flight path angle, γ , takes its maximum or minimum value. For the distance FB , we find with (6.6) and (6.7):

$$FB^2 = b^2 + a^2 e^2 = a^2$$

or

$$FB = a \quad (6.8)$$

Combination of (6.6) and (6.8) yields

$$e = \sin(FBC) \quad (6.9)$$

To conclude, three characteristics of an ellipse are mentioned without proof:

- For the second focal point F' holds $CF' = CF$.
- If a circle with center C and radius a is drawn around an ellipse, then for each line through a point P on the ellipse and perpendicular to a (Figure 6.2) holds

$$\frac{PG}{P'G} = \frac{b}{a} = \sqrt{1 - e^2} \quad (6.10)$$

- The area enclosed by an ellipse is equal to πab .

Substitution of (5.21), (5.25) and (5.33) into (6.2) gives

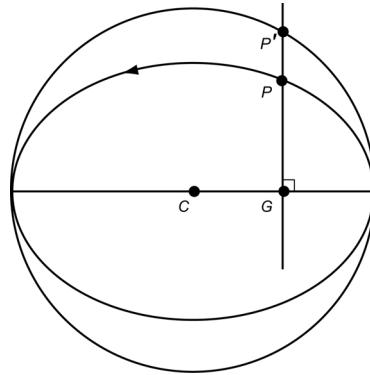


Figure 6.2: Projection of a point P on an ellipse onto the circumferential circle.

$$a = \frac{p}{1 - e^2} = \frac{r^2 V^2 \cos^2 \gamma}{r V^2 (2 - \frac{r V^2}{\mu}) \cos^2 \gamma} = \frac{\mu/2}{\frac{\mu}{r} - \frac{V^2}{2}} \quad (6.11)$$

Since a and μ are positive constants:

$$\frac{V^2}{2} - \frac{\mu}{r} < 0 \quad (6.12)$$

In Section 5.1 it was shown that the term $-\mu/r$ indicates the potential energy per unit of mass of body i . Hence, (5.5) and (6.12) show that for a body in an elliptical orbit:

$$\mathcal{E} = \frac{V^2}{2} - \frac{\mu}{r} = \mathcal{E}_k + \mathcal{E}_p = \text{constant} < 0 \quad (6.13)$$

where \mathcal{E} denotes the total energy per unit of mass of body i . Note that this result is in agreement with our finding in Section 2.3 that for a stable system the total energy (per unit of mass) is negative. Elliptical motion satisfies our criterium of stability, in the sense that body i will not move unboundedly far from body k . In Chapters 7 and 8 we will see that for parabolic and hyperbolic orbits, where body i does move unboundedly far from body k , the total energy is zero or positive. Note that the result in Section 2.3 was found for the motion relative to an inertial reference frame, while in this Section the motion refers to a non-rotating reference frame connected to body k . However, this difference has been accommodated by the introduction of the parameter $\mu = G(m_k + m_i)$.

Substitution of (6.13) into (6.11) gives

$$a = -\frac{\mu}{2\mathcal{E}} \quad (6.14)$$

Equation (5.33) can also be written as

$$e^2 = 1 + 2 \frac{r^2 V^2}{\mu^2} \left(\frac{V^2}{2} - \frac{\mu}{r} \right) \cos^2 \gamma$$

Substitution of (5.25) and (6.13) into this relation yields

$$e^2 = 1 + 2 \frac{H^2 \mathcal{E}}{\mu^2} \quad (6.15)$$

As $0 \leq e < 1$, we find

$$-\frac{\mu^2}{2} \leq H^2 \mathcal{E} < 0 \quad (6.16)$$

From (5.21), (6.14) and (6.15) we conclude that the parameters a , p and e of the ellipse can be expressed by the integrals of motion H and \mathcal{E} . The major axis is completely determined by the orbital energy (per unit of mass) of body i ; the larger (less negative) this energy, the larger the major axis of the orbit. The latus rectum is completely determined by the angular momentum of the motion; the larger the angular momentum, the larger the value of p . The eccentricity, and thus the magnitude of the Laplace-Runge-Lenz vector, A , introduced in Section 5.7 and pointing towards pericenter, is a function of both the angular momentum and the orbital energy; the larger the angular momentum and the larger the orbital energy, the larger the value of e .

6.2. Circular orbit

A special case of elliptical orbits is a circular orbit, in which, of course, the velocity is constant. That velocity is called the *circular velocity*, V_c . To find an expression for that velocity, we write (5.33) as

$$e^2 = \sin^2 \gamma + \cos^2 \gamma \left(1 - \frac{r V^2}{\mu} \right)^2$$

All terms on the right-hand side of this equation are larger than or equal to zero. This means that the minimum value of the eccentricity ($e = 0$) is obtained if

$$\gamma = 0 \quad \wedge \quad V^2 = \frac{\mu}{r} \quad (6.17)$$

In other words, in a circular orbit the flight path angle is zero and the circular velocity is given by

$$V_c = \sqrt{\frac{\mu}{r}} \quad (6.18)$$

We now generalize the concept of circular velocity and consider it as a field parameter; i.e. we postulate that at any point in space the local circular velocity is defined by (6.18). From this expression we learn that when we want to launch a satellite in a circular orbit, it is not sufficient to accelerate the satellite to a velocity equal to the local circular velocity, but the direction of the velocity vector should also be oriented perpendicular to the radius vector. In Figure 6.3 (left), the circular velocity is plotted as a function of the altitude above the surface of the Earth, Moon, Venus, Mars and Jupiter. This Figure shows that satellites moving in (near) circular orbits at low altitudes above the Earth's surface have a velocity of about 7.9 km/s or 28,500 km/hr. For a satellite around Venus the circular velocity at low altitudes is slightly lower; for the Moon and Mars it is much lower and about 1.7 km/s and 3.5 km/s, respectively. For the giant planet Jupiter the circular velocity at low altitude is very large and about 42.8 km/s.

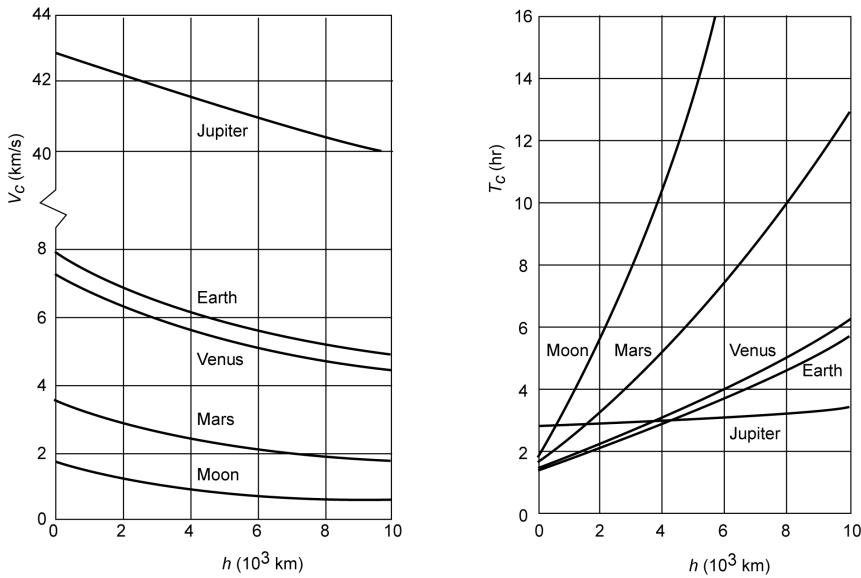


Figure 6.3: Circular velocity as a function of altitude for a number of celestial bodies (left) and the period of a circular orbit as a function of altitude for the same celestial bodies (right).

The time it takes a satellite to complete one orbital revolution about a planet is called the *period* of the orbit. In celestial mechanics, this period is defined in various ways. When it is defined as the time it takes a spacecraft in a Keplerian orbit to return to the same location with respect to the (pseudo-)inertial reference frame, then we speak of the *sidereal period*. The *synodic period* is the time that passes between two consecutive transits of the satellite through a certain meridian (Section 11.1) on the Earth's surface. In interplanetary spaceflight, the concept of a synodic period is also in use. There, it indicates the time it takes the Sun, the Earth and another planet to return to their original relative positions (Section 18.7). For perturbed satellite orbits, there are still other definitions in use, such as the *anomalistic period* and the *draconic period* (Section 23.3). When in this book we speak of the ‘period’ of an orbit, we will always mean the *sidereal* period.

For a circular orbit, we find with (6.18) for the period:

$$T_c = \frac{2\pi r}{V_c} = 2\pi \sqrt{\frac{r^3}{\mu}} \quad (6.19)$$

This period is plotted in Figure 6.3 (right) as a function of the altitude above the surface of the Earth, Moon, Venus, Mars and Jupiter. For low-altitude orbits this period is about 1.5 hr for the Earth, Moon, Venus and Mars; for Jupiter it is about 3 hr. It seems strange that although the velocities of a satellite in a low-altitude orbit about the Earth, Moon and Mars differ considerably, the period of these orbits does not differ much. There is a simple explanation for this result. Since the planets and the Moon are approximately spherical, we may write

$$\mu \approx GM \approx G\rho \frac{4}{3}\pi R^3 \quad (6.20)$$

where G is the universal gravitational constant, M is the mass of the celestial body, ρ its mean mass density, and R its radius. Substitution of (6.20) into (6.19) yields for low-altitude orbits ($r \approx R$)

$$T_e \approx \sqrt{\frac{3\pi}{G\rho}} = C \sqrt{\frac{1}{\rho}}$$

where C is a constant. Hence, the period of a low-altitude circular orbit is only dependent on the mean mass density of the celestial body and not, for example, on its radius. The densities of Earth, Moon, Mercury, Venus and Mars do not differ much (Appendix B) and are $3.3\text{-}5.5 \text{ gr/cm}^3$. Consequently, the periods of low-altitude orbits about these planets do not differ much. The giant planets Jupiter, Saturn, Uranus and Neptune have a considerably lower mean density ($0.7\text{-}1.6 \text{ gr/cm}^3$), which results in a significantly larger value for the orbital period at low altitudes.

When we consider a moon that orbits a planet at a distance equal to its Roche limit (Section 5.9), and substitute (5.47) and (6.20) into (6.19) we find

$$T_{Ro} \approx \sqrt{\frac{2.44^3 3\pi}{G\rho_2}} = C \sqrt{\frac{1}{\rho_2}}$$

where T_{Ro} is the *Roche period*, C is a constant, and ρ_2 is the mean mass density of the moon. Note that T_{Ro} is a function of the density of the moon only and not of the characteristics of its primary. The Moon's mean mass density is $\rho = 3341 \text{ kg/m}^3$, which results in a Roche period of $T_{Ro} = 6.88 \text{ hr}$. The actual Moon's orbital period is 27.32 days (Section 11.4), which confirms that the Moon is Roche-stable, as was already shown in Section 5.9.

Now, consider the situation that a satellite encircles the Earth in an easterly direction, and in an equatorial orbit with a period of $23^{\text{h}}56^{\text{m}}4^{\text{s}}$ (mean sidereal day (Section 11.4)). Because in this period the Earth completes precisely one rotation about its axis, for an observer on the surface of the rotating Earth the satellite will appear stationary above a fixed location on the Earth's equator. Therefore, such a satellite is called a *geostationary satellite*. The radius of the geostationary orbit can be computed from (6.19) and we find $r = 42,164 \text{ km} \approx 6.6 R$, where R is the mean equatorial radius of the Earth. So, the altitude of this orbit is $h = 35,786 \text{ km}$. It will be shown in Section 23.7 that the actual radius of the geostationary orbit is about 2 km larger, due to the effect of the J_2 zonal harmonic in the model of the Earth's gravity field (Section 20.1). However, in this book, except in Chapter 23, we will use the value of 42,164 km. Nowadays, the geostationary orbit is used for virtually all communication satellites. The primary reason is that a satellite in such an orbit is continuously visible for ground stations in the satellite Earth coverage region (footprint) and that it is therefore possible to have an uninterrupted communication link between these ground stations. Because the distance between the satellite and the Earth is so large, it is theoretically possible to realize a global telecommunication network using only three satellites in geostationary orbit, each separated by 120° in longitude. However, in practice, many more satellites are providing communication services from geostationary orbit. The geostationary orbit is also used intensively for monitoring the Earth's surface and atmosphere. The reason is that satellites in this orbit can observe large parts of the Earth continuously and thereby can easily detect temporal variations over these regions.

6.3. Velocity and orbital period

From (6.11) we obtain the following relation for the variation of the velocity along an elliptical orbit:

$$V^2 = \mu \left(\frac{2}{r} - \frac{1}{a} \right) \quad (6.21)$$

This relation is called the *vis-viva integral* and gives a relation between distance and velocity. The term ‘vis viva’ (Latin for ‘living force’) was introduced by G.W. von Leibniz (1646-1716) around 1685 in the context of his now obsolete theory that served as a limited early formulation of the principle of conservation of energy. Leibniz noticed that in many mechanical systems of several bodies, each with mass m_i and velocity V_i , the quantity $\sum m_i V_i^2$ is conserved. He called this quantity the vis viva of the system. In present-day terminology we would call it twice the kinetic energy of the system.

In Section 5.6, we have derived (5.26), (5.27) and (5.28), which express the radial and normal velocity components, \dot{r} and $r\dot{\theta}$, and the flight path angle, γ , as a function of the true anomaly, θ . From these relations and (6.21) a number of conclusions can be drawn about the variation of the velocity and its components along an elliptical orbit:

- The velocity reaches a minimum value when the distance is maximum, i.e. at apocenter. Then,

$$V_a^2 = \mu \left(\frac{2}{a(1+e)} - \frac{1}{a} \right) = \frac{\mu}{a} \left(\frac{1-e}{1+e} \right) = V_{c,a}^2 (1-e) \quad (6.22)$$

where $V_{c,a}$ is the circular velocity at apocenter. Hence, at apocenter the velocity is smaller than the local circular velocity.

- The velocity reaches a maximum value at pericenter:

$$V_p^2 = \frac{\mu}{a} \left(\frac{1+e}{1-e} \right) = V_{c,p}^2 (1+e) \quad (6.23)$$

So, at pericenter the velocity is always larger than the local circular velocity.

- The ratio between the maximum and minimum velocity in an elliptical orbit is given by

$$\frac{V_p}{V_a} = \frac{1+e}{1-e} \quad (6.24)$$

This ratio is a function of eccentricity only and not, for example, of a or p . Equation (6.24) shows that the range of variation of the velocity during an orbital revolution rapidly increases for increasing values of e . For $e = 0.3$, the pericenter velocity is already about twice the apocenter velocity; for $e = 0.9$, the velocity at apocenter is only about 5% of that at pericenter.

- The velocity is equal in magnitude (not in direction!) to the local circular velocity for $a = r$. From the geometry of the ellipse (Section 6.1) follows that the spacecraft is then located on the minor axis.
- The radial velocity is zero at pericenter and apocenter. It reaches a maximum value at $\theta = 90^\circ$, 270° ; i.e. in the points where the latus rectum intersects the orbit.
- The normal velocity is always positive. It reaches a maximum value at pericenter and a minimum value at apocenter.
- The flight path angle is zero at pericenter and apocenter. The extreme values of the flight path angle occur at $\cos \theta = -e$, i.e. at the intersection points of the orbit with the minor axis (Sections 5.6 and 6.1); these extreme values increase for increasing values of e . For example, for $e = 0.5$: $\gamma_{extr} = \pm 30^\circ$; when e approaches 1, γ_{extr} approaches $\pm 90^\circ$.

Figure 6.4 gives an impression of the variation of altitude, velocity and flight path angle in a rather eccentric elliptical orbit about the Earth ($e = 0.25$) with a perigee altitude of 500 km. Note that the satellite reaches a maximum altitude of about 5000 km. The maximum radial velocity is,

even for this high eccentricity, only about 1.5 km/s. As a result, the maximum difference between the velocity and its normal component is rather small; the flight path angle varies between -15° and $+15^\circ$.

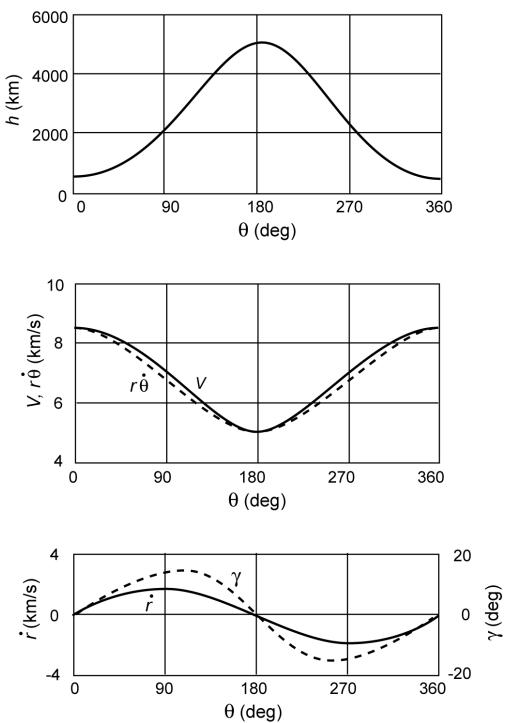


Figure 6.4: Variation of altitude, velocity and flight path angle in an elliptical orbit about the Earth ($h_p = 500$ km, $e = 0.25$).

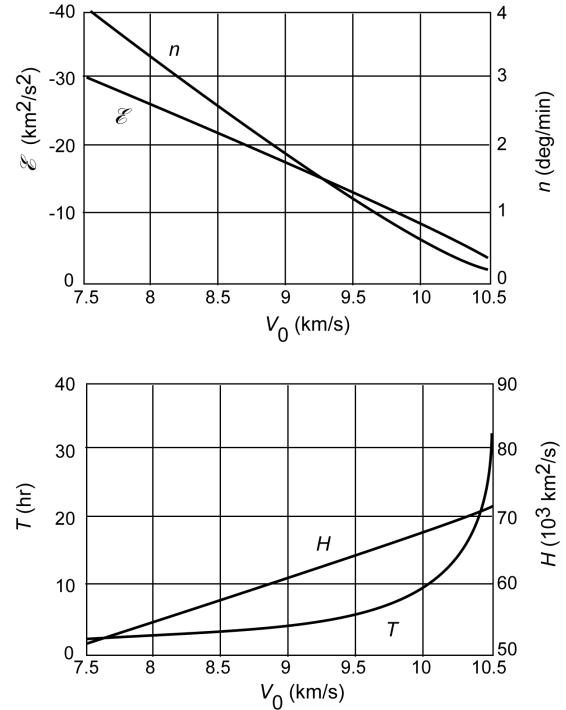


Figure 6.5: Variation of mean angular velocity, orbital energy, orbital angular momentum, and orbital period as a function of the initial velocity of a satellite about the Earth ($t = 0$: $r = 6800$ km, $\gamma = 0^\circ$).

An expression for the period of an elliptical orbit can be derived from (5.8), when we make use of the fact that in one orbital revolution body i sweeps out the entire area of the ellipse, which is given by πab . So, we may write

$$\pi a b = \frac{1}{2} H T$$

where T is the period of the elliptical orbit. Substitution of (5.21), (6.2) and (6.7) into this expression yields

$$T = 2\pi \sqrt{\frac{a^3}{\mu}} \quad (6.25)$$

If the *mean angular motion*, n , of the satellite in its orbit is defined as

$$n = \frac{2\pi}{T} \quad (6.26)$$

then (6.25) yields

$$n = \sqrt{\frac{\mu}{a^3}} \quad (6.27)$$

To gain some insight in the order of magnitude of the quantities H , \mathcal{E} , n and T of a satellite in an elliptical orbit, these quantities are plotted in Figure 6.5 as a function of the velocity of the satellite at $t = 0$. It is assumed that the satellite is then located at a distance of 6800 km from the center of the Earth (about 425 km altitude) and that at that time $\gamma = 0^\circ$. The Figure shows that the mean angular motion of the satellite has a maximum value for low orbits ($V_0 = 7.5\text{-}8 \text{ km/s}$) and then equals about $4^\circ/\text{min}$. The total energy (per unit of mass) in low orbits is approximately $-30 \text{ km}^2/\text{s}^2$. The period of the orbit increases rapidly above $V_0 \approx 9 \text{ km/s}$ for increasing values of V_0 and is already about 33 hr for $V_0 = 10.5 \text{ km/s}$. The angular momentum (per unit of mass) varies between 50,000 and 70,000 km^2/s .

6.4. Kepler's third law

In Chapter 5 we have discussed Kepler's laws and it was announced that Kepler's third law would be proved in this Section. To do that, we start from (6.25) and write

$$\frac{a^3}{T^2} = \frac{\mu}{4\pi^2}$$

When the definition of μ according to (5.2-1): $\mu = G(m_k + m_i)$ is substituted into this expression, we obtain

$$\frac{a^3}{T^2} = \frac{G m_k}{4\pi^2} \left(1 + \frac{m_i}{m_k} \right) \quad (6.28)$$

This is the 'improved' version of *Kepler's third law*. If m_i is negligible with respect to m_k , as is the case for the motion of planets about the Sun as well as to an even higher accuracy for the motion of satellites about the Earth, (6.28) can be approximated by

$$\frac{a^3}{T^2} = \frac{G m_k}{4\pi^2} = \text{constant} \quad (6.29)$$

which describes Kepler's third law in its original form (Section 5.4). It states that the ratio between the cube of the semi-major axis and the square of the orbital period is a constant for all elliptical orbits in the gravity field of m_k . For satellite orbits about the Earth, this constant has a value of roughly $10^4 \text{ km}^3/\text{s}^2$.

A useful application of Kepler's third law is found in astronomy for determining the mass of a planet. This is an essential parameter for interplanetary spaceflight, because the trajectory of a spacecraft close to a planet is determined by the gravity field of that planet and in particular by the gravitational parameter of that planet. The value of that parameter can be computed if the mass of the planet is known. The classical astronomical method to determine the mass of a planet uses the observed motion of a natural satellite (moon) about a planet in combination with Kepler's third law. When we apply Kepler's third law to the orbit of the moon, we obtain

$$\frac{a_s^3}{T_s^2} = \frac{G}{4\pi^2} (m_p + m_m)$$

where the indices p and m refer to the planet and the moon, respectively. When we apply Kepler's third law to the orbit of that planet about the Sun, we obtain

$$\frac{a_p^3}{T_p^2} = \frac{G}{4\pi^2} (m_s + m_p)$$

where the index S refers to the Sun. From these two expressions follows

$$\frac{m_p + m_m}{m_s + m_p} = \left(\frac{a_m}{a_p} \right)^3 \left(\frac{T_p}{T_m} \right)^2$$

or, because $m_m \ll m_p$ and $m_p \ll m_s$,

$$\frac{m_p}{m_s} \approx \left(\frac{a_m}{a_p} \right)^3 \left(\frac{T_p}{T_m} \right)^2 \quad (6.30)$$

The periods of the orbit of the planet about the Sun and of the orbit of the moon about that planet can be measured quite easily. By means of standard astronomical measurements we can also determine the major axis of the planet's orbit and of the moon's orbit. Using these data, the ratio between the mass of the planet and the mass of the Sun can be obtained from (6.30). When we repeat the analysis for the motion of the Earth about the Sun and of the Moon about the Earth, we find the ratio of the mass of the Earth and the mass of the Sun. So, we can compute the ratio of the mass of a planet and the mass of the Earth. If we want to compute the mass of a planet in kilograms, we have to know the mass of the Earth in kilograms; this can be determined from the known values of the universal gravitational constant, the dimensions of the Earth, and the acceleration due to gravity on the Earth's surface. This method to determine the mass of a planet can, of course, only be applied if that planet has moons. Therefore, it cannot be applied for the planets Mercury and Venus.

The mass of a planet can also be determined from the perturbations caused by that planet on the orbits of the other planets. However, since these perturbations are very small and are complex in nature, this method yields only a relatively rough estimate of the planet's mass. A much better method is to accurately measure from the Earth the Doppler shift of the radio signals transmitted by a spacecraft that passes a planet at a relatively close distance and to use the measured Doppler shift to determine the gravitational perturbations of the trajectory of that spacecraft. From these perturbations the gravitational parameter and mass of the planet can be determined. In this way, it was possible to quite accurately determine the mass of the planets Mercury, Venus and Mars using the Mariner series of interplanetary spacecraft, which have flown in the period 1962 to 1973. Spacecraft from the Pioneer and Voyager series, which were launched in the period 1972 to 1977, were used to accurately determine the mass of the planets Jupiter, Saturn, Uranus and Neptune, and of many of their moons. Observations from later missions have further improved our knowledge of the masses of planets, moons and other bodies in our solar system.

If (6.21) and (6.29) are combined, we obtain an interesting result. For a satellite with a certain velocity V at a certain distance r , the value of the semi-major axis can be computed from (6.21). The direction of the velocity does not affect the value of a . Equation (6.29) shows that only the size of the major axis determines the period of the orbit. In Figure 6.6 three orbits are drawn, for which in point P the velocity is equal in magnitude. Hence, these orbits have the same orbital period. If several spacecraft are inserted at P simultaneously into these different orbits, they will meet each other again at point P after each revolution in their orbits.

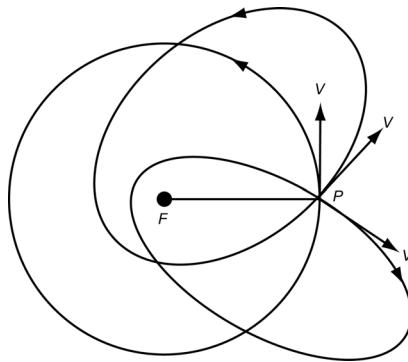


Figure 6.6: Three orbits with the same values of the major axis and of the orbital period.

6.5. Kepler's equation

In the previous Sections we have analyzed the variation of position and velocity of a body in an elliptical orbit as a function of the true anomaly, θ . Now, we will derive expressions for the relation between position in the orbit and time. In principle, one could start from (5.21) and (6.1) and write

$$dt = \frac{r^2}{\sqrt{\mu p}} d\theta = \frac{p^2}{\sqrt{\mu p}} \frac{d\theta}{(1 + e \cos \theta)^2}$$

or

$$\Delta t = \sqrt{\frac{p^3}{\mu}} \int_0^{\theta_1} \frac{d\theta}{(1 + e \cos \theta)^2}$$

where the integration is performed from $\theta = 0^\circ$ (pericenter) to $\theta = \theta_1$. This integral has different types of solutions for $e < 1$, $e = 1$ and $e > 1$, and the structure of the solution is rather complicated for $e < 1$ and $e > 1$. For example, evaluation of the integral for an elliptical orbit ($e < 1$) yields

$$t - \tau = \sqrt{\frac{a^3}{\mu}} \left[2 \arctan \left(\sqrt{\frac{1-e}{1+e}} \tan \frac{\theta}{2} \right) - e \sqrt{1-e^2} \frac{\sin \theta}{1+e \cos \theta} \right]$$

where τ is the time of pericenter passage ($\theta = 0^\circ$). Because of its complicated form, it is hard to use this expression for analytical analyses; it is also less suitable for numerical analyses. There exists a classical method to obtain a much simpler expression for the position-time relation. To this end, we draw a circle around the ellipse (Figure 6.7). From the position of body i (P), a line is drawn perpendicularly to the major axis. This line intersects the circle in P' . The angle ACP' is called the *eccentric anomaly*, E . From Figure 6.7 we obtain

$$r \cos \theta = a \cos E - a e \quad (6.31)$$

With the property of an ellipse expressed by (6.10), we find

$$r \sin \theta = a \sqrt{1-e^2} \sin E \quad (6.32)$$

Squaring (6.31) and (6.32), and subsequent summation of the results yields

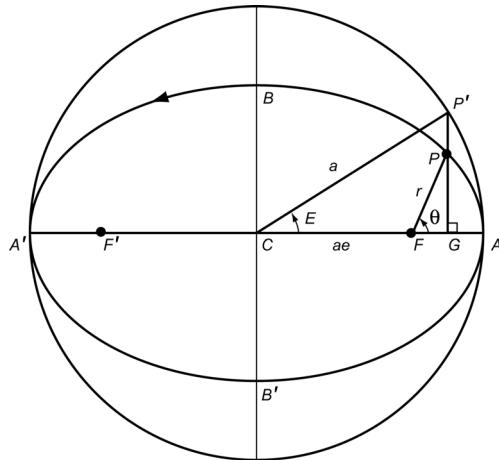


Figure 6.7: Definition of the eccentric anomaly.

$$r = \pm a(1 - e \cos E)$$

The minus-sign has no physical meaning, since $e < 1$, $a > 0$, $r > 0$. So, for the ellipse holds

$$r = a(1 - e \cos E) \quad (6.33)$$

When we compare (6.3) and (6.33), we note that in (6.33) the angular variable is in the numerator, while in (6.3) it is in the denominator. The presence of the angular variable in the numerator makes it possible to find a simple expression for the relation between position in the orbit (E) and time.

To find a relation between θ and E , we start with the trigonometric expression

$$\tan^2 \frac{\theta}{2} = \frac{1 - \cos \theta}{1 + \cos \theta} \quad (6.34)$$

Substitution of (6.31) and (6.33) yields

$$\tan^2 \frac{\theta}{2} = \left(\frac{1 + e}{1 - e} \right) \left(\frac{1 - \cos E}{1 + \cos E} \right)$$

which, according to (6.34), can be written as

$$\tan \frac{\theta}{2} = \pm \sqrt{\frac{1 + e}{1 - e}} \tan \frac{E}{2}$$

Because for an ellipse $a\sqrt{1-e^2}/r > 0$, we conclude from (6.32) that if $0^\circ \leq \theta/2 \leq 90^\circ$ then $0^\circ \leq E/2 \leq 90^\circ$, and if $90^\circ \leq \theta/2 \leq 180^\circ$ then $90^\circ \leq E/2 \leq 180^\circ$. This means that the angle $\theta/2$ always lies in the same quadrant as $E/2$. So, only the plus-sign is valid in the equation above and we obtain

$$\tan \frac{\theta}{2} = \sqrt{\frac{1 + e}{1 - e}} \tan \frac{E}{2} \quad (6.35)$$

This relation shows that when body i moves from pericenter to apocenter ($0^\circ \leq \theta \leq 180^\circ$): $\theta \geq E$; in the interval from apocenter to pericenter ($180^\circ \leq \theta \leq 360^\circ$): $\theta \leq E$. At pericenter: $\theta = E = 0^\circ$, and at apocenter: $\theta = E = 180^\circ$.

Differentiation of (6.33) to time yields

$$\dot{r} = a e \dot{E} \sin E$$

Substitution of (5.21), (5.26) and (6.2) gives

$$\frac{\mu e \sin \theta}{\sqrt{\mu a (1 - e^2)}} = a e \dot{E} \sin E$$

Combination with (6.32) and (6.33) leads, after some algebraic manipulation, to

$$\dot{E} (1 - e \cos E) = \sqrt{\frac{\mu}{a^3}}$$

which can be integrated to

$$E - e \sin E = \sqrt{\frac{\mu}{a^3}} (t - \tau) \quad (6.36-1)$$

where τ is an integration constant. The physical meaning of this integration constant becomes clear when the condition $E = 0$ is inserted into (6.36-1). We then find $t = \tau$, so τ is the time of (last) pericenter passage. With (6.27), (6.36-1) can also be written as

$$E - e \sin E = n(t - \tau) \quad (6.36-2)$$

We conclude that, while a , e and ω determine the size, shape and the orientation of the ellipse in the orbital plane, the integration constant τ is required to compute the position of body i in its orbit at a specified moment of time. The right-hand side of (6.36-2) is the product of the mean angular motion of body i and the time elapsed since the last pericenter passage. This quantity has the dimension of an angle and is called the *mean anomaly*, M . So, (6.36-2) can also be written as

$$E - e \sin E = M \quad (6.36-3)$$

The equations (6.36) are known as *Kepler's equation* that was published by J. Kepler (1571-1630) around 1618. They give the relation between the angular position of body i in its orbit and time. In this relation, the angular position is specified through the quantity E , which is linked to the true anomaly θ by (6.35). This position-time relation is of a much simpler form than the one presented for the relation between θ and t . Note that the mean anomaly changes by 360° during one orbital revolution but, in contrast to the true and eccentric anomalies, increases uniformly with time. Instead of specifying the time of perigee passage, τ , to describe the orbit, sometimes the mean anomaly at some reference epoch t_0 , M_0 , is used. The mean anomaly at an arbitrary instant of time can then be found from $M = M_0 + n(t - t_0)$.

Figure 6.8 presents the variation of altitude and true anomaly of an Earth satellite as a function of the time elapsed since the last perigee passage, for three (left) and two (right) elliptical orbits. The orbits have a perigee altitude of 400 km, while the apogee is located at 4000, 40,000 or 200,000 km altitude. From this Figure an important conclusion can be drawn: a satellite stays in the neighborhood of its apogee for a relatively long period of time. For example, the satellite in the orbit with an apogee altitude of 200,000 km, stays half of its orbital period ($T = 96.2$ hr) between a true anomaly of 170.5° and 189.5° . During this period, it moves within an altitude range of 165,260 - 200,000 km. For approximately 80% of the orbital period, the satellite moves within the altitude interval $250 < h/h_p < 500$ and for only 20% of the orbital period within the interval $1 < h/h_p < 250$.

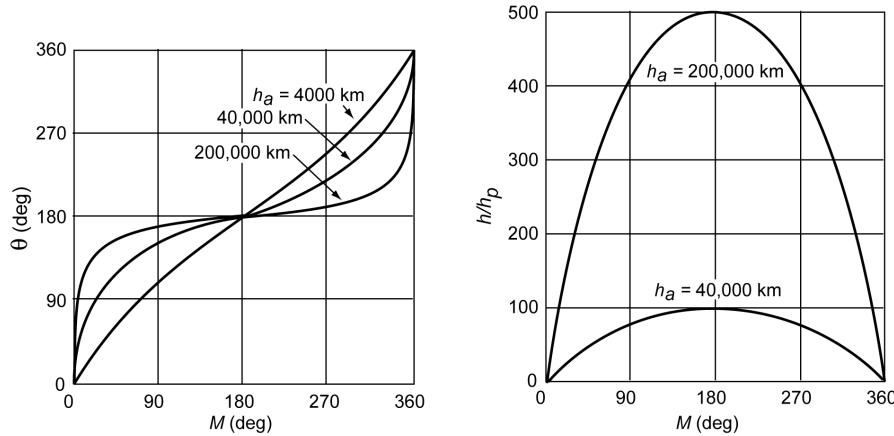


Figure 6.8: Variation of altitude and true anomaly in elliptical orbits about the Earth as a function of the time after perigee passage ($h_p = 400$ km).

The structure of Kepler's equation shows that if one wants to know *when* the satellite is at a *certain location* in its orbit, the solution of (6.36) presents no problem. However, usually one wants to know *where* the satellite is located at a *certain time*, and then a transcendental equation in E has to be solved. To prove that this transcendental equation has always one real solution, we consider the function

$$\mathcal{F}(E, M) = E - e \sin E - M$$

For a specified value of M , we have

$$\mathcal{F}(-\infty, M) = -\infty \quad ; \quad \mathcal{F}(\infty, M) = \infty \quad ; \quad \frac{d\mathcal{F}}{dE} = 1 - e \cos E > 0$$

Hence, for each value of M there is always one value of E for which $\mathcal{F}(E, M) = 0$.

Of course, Kepler was the first to solve (6.36); he did this graphically. I. Newton (1643-1727), in his *Principia*, presented a graphical solution based on a cycloid. Afterwards, until mid-nineteenth century almost all leading mathematicians and astronomers have developed efficient techniques to solve this transcendental equation and, as a result, dozens of methods are known today. For us, who have computers at our disposal, a numerical solution of this equation is very easy, for instance by applying the *Newton-Raphson method*:

$$E_{k+1} = E_k - \frac{\mathcal{F}(E_k, M)}{\frac{d}{dE}\{\mathcal{F}(E, M)\}|_{E=E_k}} = E_k - \frac{E_k - e \sin E_k - M}{1 - e \cos E_k}$$

Because $e < 1$, the very simple iterative method

$$E_{k+1} = M + e \sin E_k$$

also results in a converging iteration process. It requires relatively few iterations if $e < 0.1$; for larger values of e the Newton-Raphson method is preferable, because the simple method then requires many iterations. In both iteration schemes $E_k = M$ is an appropriate substitution to start the iteration process. Of the non-numerical methods, two will be discussed in the next Section: a simple graphical method and an analytical method that is particularly suited to low-eccentricity satellite orbits.

6.6. Graphical and analytical solution of Kepler's equation

To solve Kepler's equation in a graphical way, we first draw the function $y = \sin x$ (Figure 6.9). Then, the value of M is marked on the X -axis and the value of $M+e$ on the line $y = 1$. Subsequently, a straight line is drawn through these two points. The intersection of this line with the

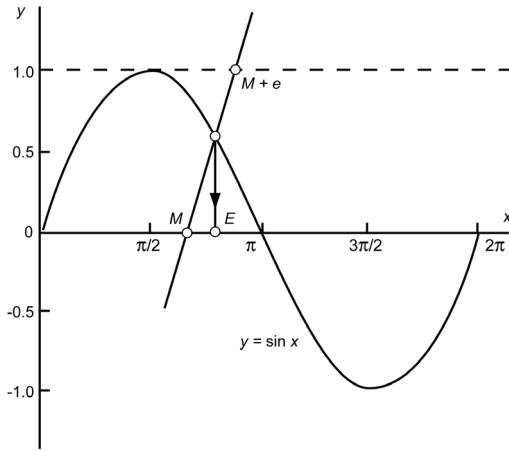


Figure 6.9: Graphical solution of Kepler's equation.

sine-function yields the solution for E . This can be proved as follows. The general equation of a straight line is

$$y = ax + b$$

This line should pass through the point $x = M, y = 0$, and the point $x = M+e, y = 1$. Substitution of these conditions in the equation of the line yields

$$0 = aM + b ; \quad 1 = a(M + e) + b$$

This system of equations has the solution $a = 1/e$, $b = -M/e$, and thus the equation of the line can be written as

$$y = \frac{E - M}{e}$$

For the intersection of the line and the sine-function, we find

$$\frac{E - M}{e} = \sin E$$

or

$$E - e \sin E = M$$

which proves the correctness of the method.

For the development of an analytical method to solve Kepler's transcendental equation, we start with the introduction of a parameter α :

$$\alpha = E - M \tag{6.37}$$

We then may write Kepler's equation as

$$\alpha = e \sin(M + \alpha) \quad (6.38)$$

Hence, the value of α is bounded between $-e$ and $+e$, where $e < 1$ and in most cases $e \ll 1$. Now, assume that we can find for α a converging power series of the type

$$\alpha = \alpha_1 e + \alpha_2 e^2 + \alpha_3 e^3 + \dots \quad (6.39)$$

where the coefficients α_n are functions of M . These functions can be found by equating (6.38) and (6.39):

$$\alpha_1 e + \alpha_2 e^2 + \alpha_3 e^3 + \dots = e \sin M \cos \alpha + e \cos M \sin \alpha \quad (6.40)$$

Since α is small, we may use standard series expansions for $\sin \alpha$ and $\cos \alpha$. Substitution of these series expansions and (6.39) into the right-hand side of (6.40) leads to

$$\begin{aligned} & e \sin M \left[1 - \frac{1}{2}(\alpha_1 e + \alpha_2 e^2 + \dots)^2 + \frac{1}{24}(\alpha_1 e + \alpha_2 e^2 + \dots)^4 + \dots \right] + \\ & e \cos M \left[(\alpha_1 e + \alpha_2 e^2 + \dots) - \frac{1}{6}(\alpha_1 e + \alpha_2 e^2 + \dots)^3 + \dots \right] \end{aligned}$$

So, (6.40) can be written as

$$\alpha_1 e + \alpha_2 e^2 + \alpha_3 e^3 + \dots = e \sin M + e^2 (\alpha_1 \cos M) + e^3 \left(-\frac{1}{2} \alpha_1^2 \sin M + \alpha_2 \cos M \right) + \dots$$

Equating the terms with equal powers of e on the left-hand and right-hand sides of this equation leads to

$$\begin{aligned} \alpha_1 &= \sin M \\ \alpha_2 &= \alpha_1 \cos M = \sin M \cos M = \frac{1}{2} \sin 2M \\ \alpha_3 &= -\frac{1}{2} \alpha_1^2 \sin M + \alpha_2 \cos M = -\frac{1}{2} \sin^3 M + \sin M \cos^2 M = \frac{3}{8} \sin 3M - \frac{1}{8} \sin M \end{aligned}$$

Substituting these expressions into (6.37) gives

$$E = M + e \left(1 - \frac{1}{8} e^2 \right) \sin M + \frac{1}{2} e^2 \sin 2M + \frac{3}{8} e^3 \sin 3M + O(e^4)$$

where we have neglected terms of order e^4 . If we would have expanded the series a little further, the result would have been:

$$\begin{aligned} E &= M + e \left(1 - \frac{1}{8} e^2 + \frac{1}{192} e^4 \right) \sin M + e^2 \left(\frac{1}{2} - \frac{1}{6} e^2 \right) \sin 2M + \\ &+ e^3 \left(\frac{3}{8} - \frac{27}{128} e^2 \right) \sin 3M + \frac{1}{3} e^4 \sin 4M + \frac{125}{384} e^5 \sin 5M + O(e^6) \end{aligned} \quad (6.41)$$

We can use this *Fourier series* analytical expression to directly compute E for any given time (M). A remarkable property of this series expansion is that the order of successive terms can be expressed by $e^k \sin kM$, i.e. the order of each coefficient in the Fourier series is equal to the order of the Fourier component. This property is characteristic for many series expansions in celestial mechanics and can be used as a rule of thumb to check the correctness of algebraic manipulations. Furthermore, for each Fourier component, successive terms in the series of e always decrease with a factor of e^2 . So, a coefficient of a sine-function of an odd argument only contains odd powers of e ; similarly, a sine-function of an even argument only contains even powers of e . This property was first emphasized by J.B. le Rond d'Alembert (1717-1783). For this reason it

was called by E.W. Brown (1866-1938) the *d'Alembert characteristic*. The two properties mentioned are closely related to the fact that (6.41) can be written as a series of *Bessel functions of the first kind* and to the properties of these functions:

$$E = M + 2 \sum_{l=1}^{\infty} J_l(l e) \frac{\sin(lM)}{l} ; \quad J_l(l e) = \frac{1}{\pi} \int_0^{\pi} \cos(lE - l e \sin E) dE$$

where l is the order of the functions. We will not examine this series expansion in Bessel functions any further; it is only mentioned here that it was just this astronomical application for which F.W. Bessel (1784-1846) introduced these functions in 1817.

A Fourier series expansion for the true anomaly can be found as follows. First, the derivative of the true anomaly to the mean anomaly is written as

$$\frac{d\theta}{dM} = \frac{d\theta}{dE} \frac{dE}{dM} \quad (6.42)$$

Subsequently, expressions for $d\theta/dE$ and dE/dM will be derived, which will then be substituted into (6.42). From (6.35) follows

$$\frac{1}{\cos^2 \frac{1}{2}\theta} \frac{d\theta}{dE} = \sqrt{\frac{1+e}{1-e}} \frac{1}{\cos^2 \frac{1}{2}E}$$

or

$$\frac{d\theta}{dE} = \sqrt{\frac{1+e}{1-e}} \frac{1+\cos\theta}{1+\cos E}$$

Substituting (6.31) and (6.33) into this relation leads, after some algebraic manipulation, to

$$\frac{d\theta}{dE} = \frac{\sqrt{1-e^2}}{1-e\cos E} \quad (6.43)$$

From (6.36-3) follows

$$\frac{dE}{dM} = \frac{1}{1-e\cos E} \quad (6.44)$$

Substitution of (6.43) and (6.44) into (6.42) leads to

$$\frac{d\theta}{dM} = \sqrt{1-e^2} \left(\frac{dE}{dM} \right)^2$$

An expression for the derivative dE/dM in this equation can be obtained by differentiation of (6.41). Squaring that expression and subsequent multiplication with the series expansion for $\sqrt{1-e^2}$ yields, after integration,

$$\begin{aligned} \theta = M + e &\left(2 - \frac{1}{4}e^2 + \frac{5}{96}e^4 \right) \sin M + e^2 \left(\frac{5}{4} - \frac{11}{24}e^2 \right) \sin 2M + \\ &+ e^3 \left(\frac{13}{12} - \frac{43}{64}e^2 \right) \sin 3M + \frac{103}{96}e^4 \sin 4M + \frac{1097}{960}e^5 \sin 5M + O(e^6) \end{aligned} \quad (6.45)$$

To obtain a Fourier series for the distance, we may proceed as follows. Substitution of (6.33)

into (6.44) gives

$$\frac{r}{a} = \left(\frac{dE}{dM} \right)^{-1}$$

Substitution of the series expansion for dE/dM already used above, and subsequently expanding the negative power results in

$$\begin{aligned} \frac{r}{a} &= \left(1 + \frac{1}{2}e^2 \right) - e \left(1 - \frac{3}{8}e^2 + \frac{5}{192}e^4 \right) \cos M - e^2 \left(\frac{1}{2} - \frac{1}{3}e^2 \right) \cos 2M \\ &\quad - e^3 \left(\frac{3}{8} - \frac{45}{128}e^2 \right) \cos 3M - \frac{1}{3}e^4 \cos 4M - \frac{125}{384}e^5 \cos 5M + O(e^6) \end{aligned} \quad (6.46)$$

P.S. Laplace (1749-1827) proved that the series expansions (6.41), (6.45) and (6.46) converge for all values of M , if $e < 0.663$. If $e > 0.663$, the series may diverge for certain values of M . Within their range of application, the M -series converge very slowly for larger values of e and they are in fact useless for $e > 0.5$. The astronomers of the nineteenth century were lucky insofar as the planets of the solar system move on small-eccentricity orbits and thus their mean anomaly expansions did converge in a satisfactory way. The larger eccentricities sometimes met in modern astrodynamics require the use of other independent variables, which accelerate the convergence process. Still, for most satellite orbits: $e \ll 0.3$, which make these relations very useful indeed.

Equation (6.46) is a good starting point to address the definition of the ‘mean distance’ during an orbital revolution. Usually, the mean of the perigee and apogee distances, which is just equal to the semi-major axis, a , is adopted for the mean distance. However, this is only one definition of the mean distance. Equation (6.46) shows that when we average the distance variation over time, i.e. over the mean anomaly, M , we find

$$r_{m,M} = \frac{1}{2\pi} \int_0^{2\pi} r dM = a \left(1 + \frac{1}{2}e^2 \right)$$

where the index m denotes ‘mean’ and the index M indicates that the average is taken over the mean anomaly. When we average over the eccentric anomaly, E , we find, using (6.33),

$$r_{m,E} = \frac{1}{2\pi} \int_0^{2\pi} r dE = \frac{1}{2\pi} \int_0^{2\pi} a (1 - e \cos E) dE = a$$

When we average over the true anomaly, θ , we find, using (6.33) and (6.43),

$$r_{m,\theta} = \frac{1}{2\pi} \int_0^{2\pi} r d\theta = \frac{1}{2\pi} \int_0^{2\pi} a (1 - e \cos E) \frac{\sqrt{1 - e^2}}{1 - e \cos E} dE = a \sqrt{1 - e^2} = b$$

where b is the semi-minor axis of the elliptical orbit. Of course, for small eccentricities the differences between the three definitions of the mean distance are small.

6.7. Lambert’s theorem

To determine the time-of-flight between two points in an elliptical orbit, we could compute the eccentric anomalies at both points and use Kepler’s equation (6.36) twice. In this Section, a more

convenient method will be developed. It is based on *Lambert's theorem* or *Lambert's equation*, which was discovered by J.H. Lambert (1728-1777) in 1761 using geometric arguments and was subsequently proved analytically by J.L. Lagrange (1736-1813) in 1778. The theorem, which is in essence a reformulation of Kepler's equation, states that the time needed to transverse an elliptical arc only depends on the semi-major axis of the ellipse, the sum of the distances from the attraction center to the initial and final points of the arc, and the length of the chord joining the initial and final points. The theorem plays a crucial role in many applications of astrodynamics, ranging from re-entry and rendez-vous missions to interplanetary flights.

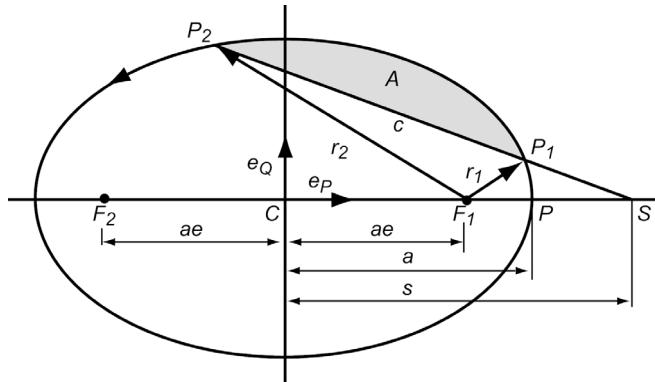


Figure 6.10: Geometry used for the derivation of Lambert's equation for elliptical motion.

Let P_1 and P_2 be the initial and final points of an elliptical arc (Figure 6.10) and suppose that the radius vector and eccentric anomaly of P_1 and P_2 are \bar{r}_1 and E_1 , and \bar{r}_2 and E_2 , respectively. Then, according to (6.36-1), the flight time, t_f , from P_1 to P_2 along the elliptical arc is given by

$$t_f = \sqrt{\frac{a^3}{\mu}} [E_2 - E_1 - e(\sin E_2 - \sin E_1)]$$

or

$$t_f = \sqrt{\frac{a^3}{\mu}} \left[E_2 - E_1 - 2e \cos \frac{1}{2}(E_2 + E_1) \sin \frac{1}{2}(E_2 - E_1) \right] \quad (6.47)$$

We now introduce the auxiliary variables f and g , defined by

$$\begin{aligned} \cos \frac{1}{2}(f+g) &= e \cos \frac{1}{2}(E_2 + E_1) \quad ; \quad 0 \leq f+g < 2\pi \\ f-g &= E_2 - E_1 \quad ; \quad 0 \leq f-g < 2\pi \end{aligned} \quad (6.48)$$

Substitution of (6.48) into (6.47) leads to

$$t_f = \sqrt{\frac{a^3}{\mu}} \left[f-g - 2 \cos \frac{1}{2}(f+g) \sin \frac{1}{2}(f-g) \right]$$

or, after some trigonometric manipulation,

$$t_f = \sqrt{\frac{a^3}{\mu}} [(f - \sin f) - (g - \sin g)] \quad (6.49)$$

We now have to express f and g in terms of the semi-major axis, a , chord, c , and r_1+r_2 . Therefore,

let \bar{e}_P be the unit vector pointing towards pericenter and \bar{e}_Q the unit vector in the plane of motion, 90° ahead of \bar{e}_P in the direction of motion. Then,

$$\bar{r}_1 = r_1 \cos \theta_1 \bar{e}_P + r_1 \sin \theta_1 \bar{e}_Q$$

where θ_1 is the true anomaly of P_1 . According to (6.31) and (6.32), this relation can also be written as

$$\bar{r}_1 = a(\cos E_1 - e) \bar{e}_P + a\sqrt{1-e^2} \sin E_1 \bar{e}_Q \quad (6.50-1)$$

A similar expression can be obtained for the position vector of P_2 :

$$\bar{r}_2 = a(\cos E_2 - e) \bar{e}_P + a\sqrt{1-e^2} \sin E_2 \bar{e}_Q \quad (6.50-2)$$

The length of the chord, c , joining P_1 and P_2 can be found from

$$c^2 = (\bar{r}_1 - \bar{r}_2) \cdot (\bar{r}_1 - \bar{r}_2)$$

Substitution of (6.50-1) and (6.50-2) into this relation yields

$$c^2 = 4a^2 [1 - e^2 \cos^2 \frac{1}{2}(E_2 + E_1)] \sin^2 \frac{1}{2}(E_2 - E_1)$$

or, using (6.48) and the appropriate trigonometric identities,

$$c = a(\cos g - \cos f) \quad (6.51)$$

From (6.33), we obtain

$$\bar{r}_1 + \bar{r}_2 = 2a[1 - \frac{1}{2}e(\cos E_1 + \cos E_2)] = 2a[1 - e \cos \frac{1}{2}(E_2 + E_1) \cos \frac{1}{2}(E_2 - E_1)]$$

Substitution of (6.48) into this relations leads, after some trigonometric manipulation, to

$$\bar{r}_1 + \bar{r}_2 = 2a[1 - \frac{1}{2}(\cos f + \cos g)] \quad (6.52)$$

Solving (6.51) and (6.52) for $\cos f$ and $\cos g$ yields

$$\cos f = 1 - \frac{c + \bar{r}_1 + \bar{r}_2}{2a} ; \quad \cos g = 1 + \frac{c - \bar{r}_1 - \bar{r}_2}{2a}$$

or

$$\sin^2 \frac{1}{2}f = \frac{\bar{r}_1 + \bar{r}_2 + c}{4a} ; \quad \sin^2 \frac{1}{2}g = \frac{\bar{r}_1 + \bar{r}_2 - c}{4a} \quad (6.53)$$

Summing (6.51) and (6.52), we find

$$a = \frac{\bar{r}_1 + \bar{r}_2 + c}{2(1 - \cos f)}$$

This equation shows that the shortest semi-major axis, a_{min} , of a conic section through P_1 and P_2 occurs at $f = 180^\circ$ and is given by

$$a_{min} = \frac{1}{4}(\bar{r}_1 + \bar{r}_2 + c) \quad (6.54)$$

Further, define a parameter K as

$$K = 1 - \frac{c}{2a_{min}} = \frac{\mathbf{r}_1 + \mathbf{r}_2 - c}{\mathbf{r}_1 + \mathbf{r}_2 + c} \quad (6.55)$$

then (6.53) can be written as

$$\sin^2 \frac{1}{2}f = \frac{a_{min}}{a} \quad ; \quad \sin^2 \frac{1}{2}g = K \frac{a_{min}}{a} \quad (6.56)$$

When a , r_1 , r_2 and c are known, formally we can compute from (6.49) and (6.54) to (6.56) the flight time, t_f , which is in essence *Lambert's equation* for elliptical motion. However, it is evident that (6.56) does not give an unambiguous solution of f and g . Therefore, a further analysis is required.

From the inequalities in (6.48) we conclude that $0^\circ \leq f < 360^\circ$ and $-180^\circ \leq g < 180^\circ$. Now, let S be the point where the line through P_1 and P_2 cuts the major axis (Figure 6.10). The position vector of an arbitrary point on the line through P_1 and P_2 with respect to the center C of the ellipse is given by

$$\bar{\mathbf{r}} = e a \bar{\mathbf{e}}_P + \lambda \bar{\mathbf{r}}_1 + (1 - \lambda) \bar{\mathbf{r}}_2 \quad ; \quad -\infty < \lambda < \infty \quad (6.57)$$

This relation, of course, also holds for point S , when $\bar{\mathbf{r}} = \bar{\mathbf{r}}_S$. The distance s between C and S , measured positive from C in the direction of the pericenter, can also be determined from the equation

$$\bar{\mathbf{r}}_S = s \bar{\mathbf{e}}_P$$

Substitution of (6.57) with (6.50) into this relation, and solution of the resulting equation for λ and s , yields, after some algebraic manipulation,

$$\lambda = \frac{\sin E_2}{\sin E_2 - \sin E_1} \quad ; \quad s = a \frac{\sin(E_2 - E_1)}{\sin E_2 - \sin E_1} = a \frac{\cos \frac{1}{2}(E_2 - E_1)}{\cos \frac{1}{2}(E_2 + E_1)}$$

We then find for the distances F_1S , F_2S and PS , measured positive to the right in Figure 6.10:

$$F_1S = s - ae = a \frac{\cos \frac{1}{2}(E_2 - E_1) - e \cos \frac{1}{2}(E_2 + E_1)}{\cos \frac{1}{2}(E_2 + E_1)}$$

$$F_2S = s + ae = a \frac{\cos \frac{1}{2}(E_2 - E_1) + e \cos \frac{1}{2}(E_2 + E_1)}{\cos \frac{1}{2}(E_2 + E_1)}$$

$$PS = s - a = a \frac{\cos \frac{1}{2}(E_2 - E_1) - \cos \frac{1}{2}(E_2 + E_1)}{\cos \frac{1}{2}(E_2 + E_1)}$$

or, using (6.48) and the appropriate trigonometric identities,

$$\frac{F_1 S}{P S} = \frac{\sin \frac{1}{2} f \sin \frac{1}{2} g}{\sin \frac{1}{2} E_1 \sin \frac{1}{2} E_2} \quad ; \quad \frac{F_2 S}{P S} = \frac{\cos \frac{1}{2} f \cos \frac{1}{2} g}{\sin \frac{1}{2} E_1 \sin \frac{1}{2} E_2} \quad (6.58)$$

Now, $\sin f/2$ and $\cos g/2$ are always positive, while, as we may take E_1 between 0° and 360° , $\sin E_1/2$ is positive too. We then may distinguish four cases (Figure 6.11).

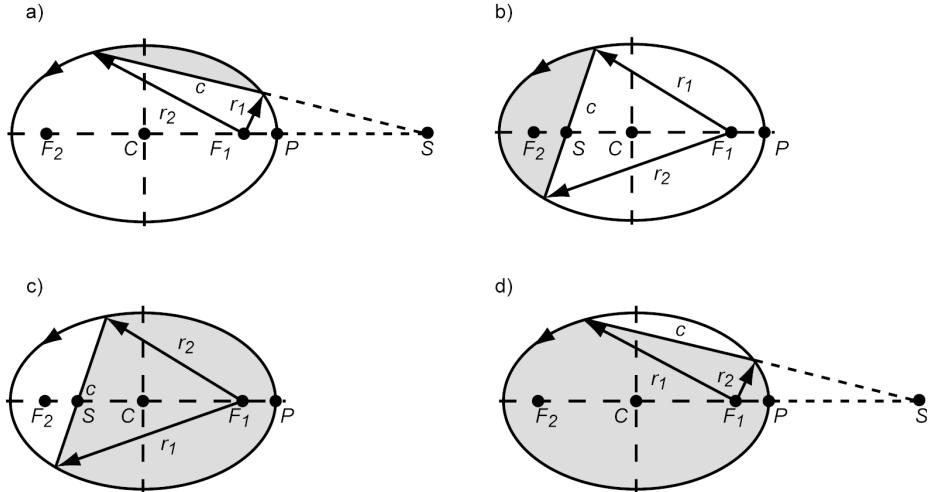


Figure 6.11: Four different cases for Lambert's equation.

Let A be the area bounded by the arc $P_1 P_2$ and the chord joining P_1 and P_2 (Figure 6.10). Then the following cases exist:

a) If A includes neither F_1 nor F_2 (Figure 6.11^a):

$$\sin \frac{1}{2} E_2 > 0 \quad ; \quad (F_1 S)/(P S) > 0 \quad ; \quad (F_2 S)/(P S) > 0$$

and we find

$$\sin \frac{1}{2} g > 0 \quad ; \quad \cos \frac{1}{2} f > 0 \quad (6.59-1)$$

b) If A does not include F_1 , but includes F_2 (Figure 6.11^b):

$$\sin \frac{1}{2} E_2 > 0 \quad ; \quad (F_1 S)/(P S) > 0 \quad ; \quad (F_2 S)/(P S) < 0$$

and we find

$$\sin \frac{1}{2} g > 0 \quad ; \quad \cos \frac{1}{2} f < 0 \quad (6.59-2)$$

c) If A includes F_1 , but not F_2 (Figure 6.11^c):

$$\sin \frac{1}{2} E_2 < 0 \quad ; \quad (F_1 S)/(P S) > 0 \quad ; \quad (F_2 S)/(P S) < 0$$

and we find

$$\sin \frac{1}{2} g < 0 \quad ; \quad \cos \frac{1}{2} f > 0 \quad (6.59-3)$$

d) If A includes both foci (Figure 6.11^d):

$$\sin \frac{1}{2}E_2 < 0 ; \quad (F_1 S)/(PS) > 0 ; \quad (F_2 S)/(PS) > 0$$

and we find

$$\sin \frac{1}{2}g < 0 ; \quad \cos \frac{1}{2}f < 0 \quad (6.59-4)$$

When the parameters α and β are defined as

$$\begin{aligned} \alpha &= 2 \arcsin \sqrt{\frac{a_{\min}}{a}} ; \quad 0 \leq \alpha \leq \pi \\ \beta &= 2 \arcsin \sqrt{K \frac{a_{\min}}{a}} ; \quad 0 \leq \beta \leq \pi \end{aligned} \quad (6.60)$$

we find, according to (6.56),

$$f = \alpha \quad \text{or} \quad f = 2\pi - \alpha ; \quad g = \beta \quad \text{or} \quad g = -\beta \quad (6.61)$$

The flight time for the four different cases (Figure 6.11) follows from (6.49) together with (6.59) to (6.61):

a) If A does not include F_1 and F_2 :

$$t_f = \sqrt{\frac{a^3}{\mu}} [(\alpha - \sin \alpha) - (\beta - \sin \beta)] \quad (6.62-1)$$

b) If A does not include F_1 , but includes F_2 :

$$t_f = \sqrt{\frac{a^3}{\mu}} [2\pi - (\alpha - \sin \alpha) - (\beta - \sin \beta)] \quad (6.62-2)$$

c) If A includes F_1 , but not F_2 :

$$t_f = \sqrt{\frac{a^3}{\mu}} [(\alpha - \sin \alpha) + (\beta - \sin \beta)] \quad (6.62-3)$$

d) If A includes both F_1 and F_2 :

$$t_f = \sqrt{\frac{a^3}{\mu}} [2\pi - (\alpha - \sin \alpha) + (\beta - \sin \beta)] \quad (6.62-4)$$

These relations constitute the ‘operational’ version of Lambert’s equation. The cases a) and b) correspond to a transfer angle $\Delta\theta = \theta_2 - \theta_1$ of less than 180° , while c) and d) correspond to a transfer angle $\Delta\theta$ of greater than 180° .

7. PARABOLIC ORBITS

For a parabolic orbit: $e = 1$, and we find from the general equation for a conic section (5.22):

$$r = \frac{p}{1 + \cos\theta} \quad (7.1)$$

where $p > 0$. In spaceflight as well as in classical astronomy a pure parabolic orbit does not exist; any variation in the value of e , no matter how small, is sufficient to change the orbit into an ellipse or a hyperbola. Still, the parabolic orbit does not only have a theoretical meaning; in computations we often approximate an elliptical or a hyperbolic orbit with $e \approx 1$ by a parabolic orbit. The reasons are that we sometimes do not know whether we are dealing with a highly-eccentric elliptical orbit or a low-eccentricity hyperbolic orbit, and that the computation process for orbits with $e \approx 1$ runs faster when we use the equations for a parabolic orbit. It is clear that the major axis of a parabola has an infinite length and so it is not appropriate to separate the parameter p in (7.1) into a and e , as was done in Section 6.1 for an elliptical orbit. For the pericenter distance holds, according to (7.1),

$$r_p = \frac{p}{2} \quad (7.2)$$

As at $\theta = 90^\circ$ the radius vector has the length p , we conclude that in the interval $0^\circ \leq \theta \leq 90^\circ$ the radius vector doubles in length.

7.1. Escape velocity

In Chapter 5, we have derived (5.33) that expresses the relation between the instantaneous quantities r , V and γ , and the eccentricity of the orbit. Since for a parabola $e = 1$, (5.33) yields

$$\frac{rV^2}{\mu} \left(2 - \frac{rV^2}{\mu} \right) \cos^2\gamma = 0$$

A solution of this equation is

$$\frac{rV^2 \cos^2\gamma}{\mu} = \frac{p}{r} = 0$$

where (5.21) and (5.25) have been used. For $r \neq \infty$, this represents a rectilinear motion through the center of attraction. Such motion, which was discussed in Section 5.1, has little physical meaning. Hence, the ‘real’ solution is

$$V = \sqrt{\frac{2\mu}{r}} \quad (7.3-1)$$

In other words, for each point in a parabolic orbit the velocity is given by (7.3-1). With the expression for the local circular velocity, (6.18), we can write (7.3-1) as

$$V = \sqrt{2} V_c$$

This relation shows that if we want to launch a spacecraft into a parabolic orbit, it has to be accelerated to a velocity equal to $\sqrt{2}$ times the local circular velocity. The direction of the velocity is of no importance; the spacecraft will always describe a parabolic orbit and conse-

quently recedes infinitely far away from the Earth. We thus may say that the spacecraft ‘escapes’ from the gravity field of the Earth, and the velocity defined by (7.3-1) is therefore called the *escape velocity* (V_{esc}):

$$V_{esc} = \sqrt{\frac{2\mu}{r}} = \sqrt{2} V_c \quad (7.3-2)$$

Combining (7.3-2) with (5.21), (5.24) and (5.27) results in

$$p = \frac{4\mu}{V_{esc}^2} \cos^2 \gamma \quad (7.4)$$

This relation shows that the size of the parabola that is flown by the spacecraft after it has been accelerated to escape velocity, is dependent on the direction of the velocity vector (γ). Just as we did for the circular velocity, we now generalize the concept of escape velocity and consider it as a field parameter; i.e. at any point in space the local escape velocity is defined by (7.3-2).

In Figure 7.1, the escape velocity is plotted as a function of the altitude above the surface of the Earth, Moon, Mars, Venus and Jupiter. Note that the escape velocity at the Earth’s surface is approximately 11.2 km/s; for Jupiter, this velocity is about 60.5 km/s. From the surface of the Moon, a rather small velocity increase of about 2.4 km/s is already sufficient to enter an escape trajectory. This illustrates the attractiveness of the Moon as a launch base for interplanetary spaceflight.

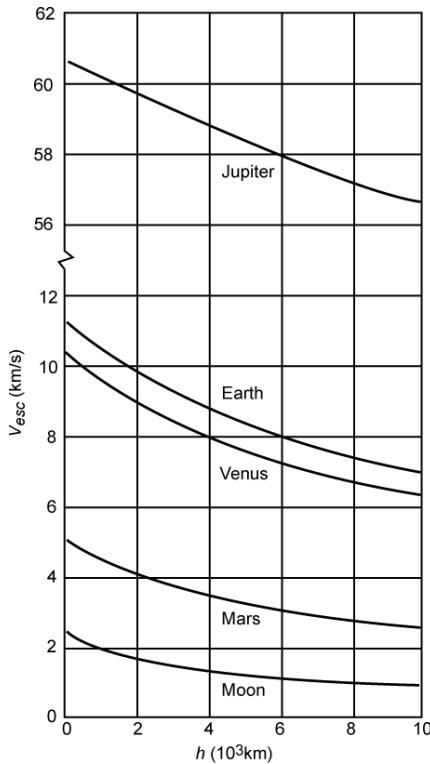


Figure 7.1: Escape velocity as a function of altitude for a number of celestial bodies.

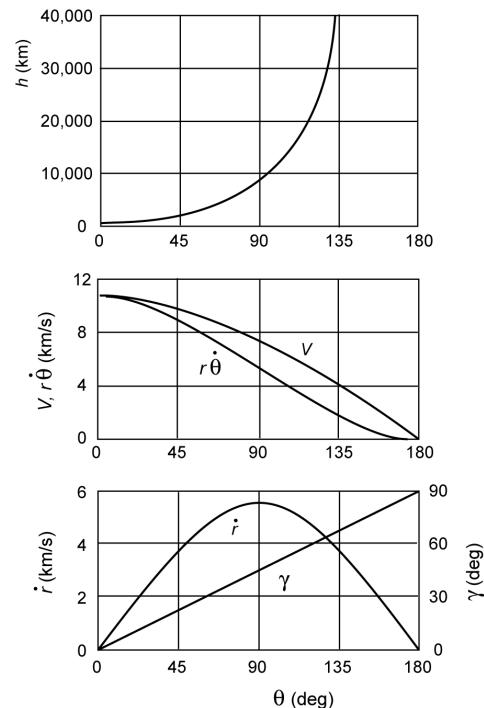


Figure 7.2: Variation of altitude, velocity and flight path angle in a parabolic orbit about the Earth ($h_p = 500$ km).

Table 7.1 presents some characteristic parameters of the Sun, a white dwarf, a neutron star and a black hole. In addition, the escape velocity at the surface of these stars, and the velocity and orbital period of a body moving about these stars are listed. The velocities are expressed in terms

Table 7.1: Comparison of the escape velocity on the surface of the Sun and on the surface of some degenerated types of stars, and of the velocity and orbital period of a body moving about these stars.

| | Sun | White dwarf | Neutron star | Black hole |
|----------------------------------|----------------------|----------------------|----------------------|----------------------|
| Radius (km) | 6.96×10^5 | 1.2×10^4 | 15 | 10 |
| Mass (M_\odot) | 1 | 0.5 | 1 | 5 |
| Mean density (kg/m^3) | 1.4×10^3 | 1.4×10^8 | 1.4×10^{17} | 2.4×10^{18} |
| Surface escape velocity (c) | 2.1×10^{-3} | 1.1×10^{-2} | 0.44 | 1.21 |
| Velocity (c)* | 1.0×10^{-3} | 5.5×10^{-3} | 0.22 | 0.61 |
| Orbital period (s)* | 2.8×10^4 | 91 | 2.8×10^{-3} | 6.9×10^{-4} |

* For a circular orbit with a radius of twice the radius of the star.

of the speed of light, c . For the black hole the surface escape velocity turns out to be larger than the speed of light. This was already found by J. Michell (1724-1793), who wrote in 1784 that a very heavy star can have an escape velocity that is larger than the speed of light. He reasoned that light particles shooting upward would be pulled back by the star. Hence, the star would not be visible. P.S. Laplace (1749-1827) repeated this reasoning in 1795 and calculated that a star with the density of the Earth but with a radius of 250 times the radius of the Sun, would be such a dark body. We now know that the reasoning by Michell and Laplace is not correct; after all, light is not slowed down by a gravity field! According to Einstein's *general theory of relativity*, mass influences the local curvature of space; the trajectory of light is deflected by the presence of this mass. If we consider a spherical body with a very large mass but with a small radius (thus a high density) that emits light, the so-called *Schwarzschild radius* (K. Schwarzschild (1873-1916)) can be defined. If the radius of the body is smaller than the Schwarzschild radius, the deflection of the light that is emitted by the body is too large for the light to escape that celestial body; in that case the body is a *black hole*. It is interesting to mention that the Schwarzschild radius is precisely equal to the radius of the celestial body (with the same mass) for which, according to the calculations by Michell and Laplace, the escape velocity at the surface of the celestial body is precisely equal to the speed of light. So, although their reasoning was wrong, the result of that reasoning was correct! Table 7.1 also shows that for compact bodies, such as white dwarfs, neutron stars and black holes, the velocity in a low-altitude circular orbit around the star is very high and the associated orbital period is extremely short.

7.2. Flight path angle, total energy and velocity

From (5.28) we obtain for $e = 1$:

$$\tan \gamma = \frac{\sin \theta}{1 + \cos \theta} = \tan \frac{1}{2} \theta$$

For $0^\circ \leq \theta \leq 180^\circ$, the value of $\tan(\theta/2)$ is positive (or zero). Since, by definition $-90^\circ \leq \gamma \leq 90^\circ$, we find for this range of true anomalies $0^\circ \leq \gamma \leq 90^\circ$, and hence for $0^\circ \leq \theta \leq 180^\circ$: $\gamma = \theta/2$. Similarly, for $180^\circ < \theta < 360^\circ$: $\gamma = -\theta/2$. Hence, the flight path angle increases linearly with θ and we can write

$$-180^\circ < \theta \leq 180^\circ : \quad \gamma = \frac{1}{2} \theta \quad (7.5)$$

This result has already been found graphically in Section 5.6.

Substituting (7.3) into (5.5), we find for the total energy per unit of mass:

$$\mathcal{E} = 0 \quad (7.6)$$

So, in a parabolic orbit the total energy of a body is always equal to zero.

To obtain the magnitude of the radial and normal velocity components in a parabolic orbit we substitute $e = 1$ into (5.26) and (5.27), and find

$$\dot{r} = \frac{\mu \sin \theta}{H} \quad (7.7)$$

$$r \dot{\theta} = \frac{\mu}{H} (1 + \cos \theta) \quad (7.8)$$

From expressions (7.3), (7.7) and (7.8) a number of conclusions can be drawn about the variation of the velocity along a parabolic orbit:

- The velocity reaches a minimum value of zero when r is maximum; i.e. $\theta = \pm 180^\circ$, $r = \infty$. It reaches a maximum value when r is minimum; i.e. at $\theta = 0^\circ$. The maximum velocity is given by

$$V_{\max} = 2 \sqrt{\frac{\mu}{p}} \quad (7.9)$$

- At any point along a parabolic orbit the local velocity is equal to $\sqrt{2}$ times the local circular velocity.
- The radial velocity is zero for $\theta = 0^\circ, \pm 180^\circ$; i.e. at pericenter and at $r = \infty$. It reaches a maximum value for $\theta = 90^\circ, 270^\circ$; i.e. in the point where the latus rectum intersects the parabola.
- The normal velocity decreases monotonously from a maximum value at pericenter to a minimum value of zero at $\theta = \pm 180^\circ$.

Figure 7.2 gives an impression of the variation of altitude, velocity and flight path angle along a parabolic orbit with a perigee altitude of 500 km above the Earth's surface. Note that for $\theta > 60^\circ$ the distance increases rapidly with increasing values of θ .

7.3. Barker's equation

To obtain a relation between position and time, we start from (5.21) and (7.1) and write

$$dt = \sqrt{\frac{p^3}{\mu} \frac{d\theta}{(1 + \cos \theta)^2}} \quad (7.10)$$

In Section 6.5, we have found a similar expression for the elliptical orbit and decided that because of the complicated form of the integral we should introduce a new variable E . Here, this is not necessary; the reason is that the denominator does not contain the eccentricity as was the case for the elliptical orbit. From trigonometry we know that

$$\frac{d\theta}{(1 + \cos \theta)^2} = \frac{d(\theta/2)}{2 \left(\frac{1 + \cos \theta}{2} \right)^2} = \frac{d(\theta/2)}{2 \cos^4(\theta/2)} = \frac{d(\tan(\theta/2))}{2 \cos^2(\theta/2)}$$

Substitution of this relation into (7.10) and subsequent integration yields

$$\tan \frac{\theta}{2} + \frac{1}{3} \tan^3 \frac{\theta}{2} = 2 \sqrt{\frac{\mu}{p^3}} (t - \tau) \quad (7.11-1)$$

where τ is an integration constant. The physical interpretation of τ becomes clear when $\theta = 0^\circ$ is inserted into (7.11-1). It then follows that $t = \tau$; so, τ is the time of (last) pericenter passage. This equation, which is known as *Barker's equation* as a tribute to T. Barker (1722-1809), who developed the first tables of solutions for parabolic orbits, gives the relation between true anomaly and time; it is the analogue of Kepler's equation for elliptical orbits.

Now, an angular velocity \bar{n} is introduced, defined as

$$\bar{n} = \sqrt{\frac{\mu}{p^3}} \quad (7.12)$$

According to (7.10), the orbital angular motion equals this value at $\theta = 90^\circ, 270^\circ$; i.e. in the intersections of the parabola with the latus rectum. Substitution of (7.12) into (7.11) gives

$$\tan \frac{1}{2} \theta + \frac{1}{3} \tan^3 \frac{1}{2} \theta = 2 \bar{n} (t - \tau) \quad (7.11-2)$$

or, when, somewhat analogous to the discussion on elliptical orbits, a kind of mean anomaly, $M = \bar{n}(t - \tau)$, is introduced,

$$\tan \frac{1}{2} \theta + \frac{1}{3} \tan^3 \frac{1}{2} \theta = 2 M \quad (7.11-3)$$

It is emphasized that, while for an elliptical orbit $0^\circ \leq M \leq 360^\circ$, for a parabolic orbit $-\infty < \bar{M} < \infty$. Figure 7.3 presents the variation of altitude and true anomaly as a function of time along a

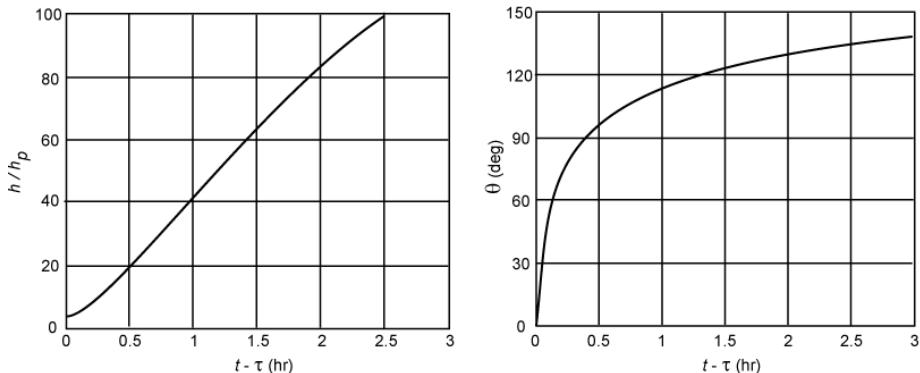


Figure 7.3: Variation of altitude and true anomaly in a parabolic orbit about the Earth as a function of the time after perigee passage ($h_p = 400$ km).

parabolic orbit with its perigee at 400 km altitude above the Earth's surface. Note that initially the true anomaly increases rapidly; so, the spacecraft has a large angular velocity. This rate of increase decreases strongly during the first hour, until after a few hours the spacecraft moves almost radially away from the Earth. As a result of the constant decrease in velocity, the rate of increase of altitude decreases gradually.

The time it takes a body in a parabolic orbit to complete the arc $-90^\circ \leq \theta \leq 90^\circ$, i.e. the part of the orbit between the two intersections of the parabola with the latus rectum, is according to (7.11-1)

$$T_p = t_2 - t_1 = \frac{4}{3} \sqrt{\frac{p^3}{\mu}}$$

This expression can also be written as

$$\frac{p^3}{T_p^2} = \frac{9}{16} \mu = \frac{9}{16} G m_k \left(1 + \frac{m_i}{m_k} \right) \quad (7.13)$$

where m_k is the mass of the central body and m_i the mass of the body that moves in the parabolic orbit about the central body. If m_i can be neglected with respect to m_k , then:

$$\frac{p^3}{T_p^2} = \frac{9}{16} G m_k = \text{constant} \quad (7.14)$$

This equation for parabolic orbits is the analogue of Kepler's third law for elliptical orbits ((6.29)), and states that, for all parabolic orbits in the gravity field of a central body, the ratio between the cube of the semi-latus rectum and the square of the flight time from $\theta = -90^\circ$ to $\theta = 90^\circ$ is constant.

Note that Barker's equation is a third-degree equation in $\tan(\theta/2)$ and can be solved analytically. To prove that this equation has but one real solution, we consider the function

$$\mathcal{F}(\theta, \bar{M}) = \tan \frac{1}{2}\theta + \frac{1}{3} \tan^3 \frac{1}{2}\theta - 2\bar{M}$$

For a specified value of \bar{M} , we have

$$\mathcal{F}(-\pi, \bar{M}) = -\infty ; \quad \mathcal{F}(\pi, \bar{M}) = \infty ; \quad \frac{d\mathcal{F}}{d\theta} = \frac{1}{2} (\cos \frac{1}{2}\theta)^{-4} > 0$$

which means that the equation has one solution of θ for any value of \bar{M} .

In the classical trigonometric method for solving (7.11-3) a new angular variable x is introduced, which is defined as

$$\tan \frac{1}{2}\theta = 2 \cot 2x \quad (7.15)$$

With the trigonometric relation

$$2 \cot 2x = \frac{1 - \tan^2 x}{\tan x} = \cot x - \tan x \quad (7.16)$$

follows

$$\tan \frac{1}{2}\theta = \cot x - \tan x ; \quad \tan^3 \frac{1}{2}\theta = \cot^3 x - 3(\cot x - \tan x) - \tan^3 x$$

Substituting these relations into (7.11-3) leads to

$$\cot^3 x - \tan^3 x = 6\bar{M}$$

If an angle y is defined as

$$\cot x = \left(\cot \frac{1}{2}y \right)^{1/3} \quad (7.17)$$

we obtain

$$\cot \frac{1}{2}y - \tan \frac{1}{2}y = 6\bar{M}$$

Using (7.16), this equation can be written as

$$\cot y = 3\bar{M} \quad (7.18)$$

With (7.15), (7.17) and (7.18) we can find the angular position θ as a function of \bar{M} , and thus of time. In the computation process the angles y , x and $\theta/2$ take values between -90° and 90° ; when \bar{M} is negative, y , $\cot(y/2)$, x and θ are negative.

In a more efficient algebraic solution scheme, we first set

$$C = 3\bar{M} ; \quad x = \tan \frac{1}{2}\theta \quad (7.19)$$

and write (7.11-3) as

$$x^3 + 3x - 2C = 0$$

As shown above, this equation has always one real root, which can be computed analytically:

$$x = \sqrt[3]{C + \sqrt{C^2 + 1}} + \sqrt[3]{C - \sqrt{C^2 + 1}} \quad (7.20)$$

Taking

$$y = \sqrt[3]{C + \sqrt{C^2 + 1}} \quad (7.21)$$

we find

$$\frac{1}{y} = -\sqrt[3]{C - \sqrt{C^2 + 1}}$$

and we obtain from (7.20)

$$x = y - \frac{1}{y} \quad (7.22)$$

Note that y is always positive, regardless of the sign of \bar{M} . With (7.19), (7.21) and (7.22) we can find θ as a function of \bar{M} , and thus of time.

7.4. Euler's equation

To compute the orbit of comets, L. Euler (1707-1783) published in 1743 an equation that may be considered as a version of Lambert's theorem (Section 6.7) for parabolic orbits. In this Section, that equation will be derived.

Consider the positions of a body in a parabolic orbit at the instances t_1 and t_2 (Figure 7.4). Let the corresponding radii be r_1 and r_2 , the chord joining their extremities c , and the corresponding true anomalies θ_1 and θ_2 . Then, it follows from (7.11-1) that

$$2\sqrt{\frac{\mu}{p^3}} t_f = \tan \frac{1}{2}\theta_2 - \tan \frac{1}{2}\theta_1 + \frac{1}{3} \left(\tan^3 \frac{1}{2}\theta_2 - \tan^3 \frac{1}{2}\theta_1 \right)$$

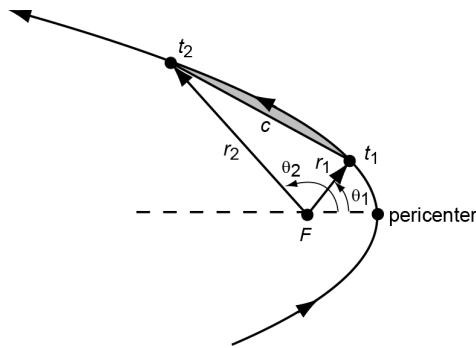


Figure 7.4: Geometry for the derivation of Euler's equation.

or

$$6\sqrt{\frac{\mu}{p^3}}t_f = \left(\tan \frac{1}{2}\theta_2 - \tan \frac{1}{2}\theta_1 \right) \left[3 \left(1 + \tan \frac{1}{2}\theta_1 \tan \frac{1}{2}\theta_2 \right) + \left(\tan \frac{1}{2}\theta_2 - \tan \frac{1}{2}\theta_1 \right)^2 \right] \quad (7.23)$$

where $t_f = t_2 - t_1$. The equation for the chord is

$$c^2 = r_1^2 + r_2^2 - 2r_1 r_2 \cos(\theta_2 - \theta_1) = (r_1 + r_2)^2 - 4r_1 r_2 \cos^2 \frac{1}{2}(\theta_2 - \theta_1)$$

From this equation, we find

$$2\sqrt{r_1 r_2} \cos \frac{1}{2}(\theta_2 - \theta_1) = \pm \sqrt{(r_1 + r_2 + c)(r_1 + r_2 - c)} \quad (7.24)$$

The plus-sign is to be taken before the radical if $\theta_2 - \theta_1 < 180^\circ$; the minus-sign if $\theta_2 - \theta_1 > 180^\circ$. From (7.1) follows

$$r_1 = \frac{p}{2 \cos^2 \frac{1}{2}\theta_1} \quad ; \quad r_2 = \frac{p}{2 \cos^2 \frac{1}{2}\theta_2} \quad (7.25)$$

and we obtain

$$r_1 + r_2 = \frac{p}{2} \left(\frac{1}{\cos^2 \frac{1}{2}\theta_1} + \frac{1}{\cos^2 \frac{1}{2}\theta_2} \right) \quad (7.26)$$

Substitution of (7.25) into the left-hand side of (7.24) yields, after applying some trigonometry,

$$p \left(1 + \tan \frac{1}{2}\theta_1 \tan \frac{1}{2}\theta_2 \right) = \pm \sqrt{(r_1 + r_2 + c)(r_1 + r_2 - c)} \quad (7.27)$$

With (7.26) and (7.27) we write

$$\begin{aligned} (r_1 + r_2 + c) + (r_1 + r_2 - c) &\mp 2\sqrt{(r_1 + r_2 + c)(r_1 + r_2 - c)} = \\ &= p \left(\frac{1}{\cos^2 \frac{1}{2}\theta_1} + \frac{1}{\cos^2 \frac{1}{2}\theta_2} \right) - 2p \left(1 + \tan \frac{1}{2}\theta_1 \tan \frac{1}{2}\theta_2 \right) \end{aligned}$$

With

$$\frac{1}{\cos^2 x} = \tan^2 x + 1$$

we finally find

$$\sqrt{r_1 + r_2 + c} \mp \sqrt{r_1 + r_2 - c} = \sqrt{p} \left(\tan \frac{1}{2} \theta_2 - \tan \frac{1}{2} \theta_1 \right) \quad (7.28)$$

Substitution of (7.27) and (7.28) into (7.23) yields

$$6 \sqrt{\frac{\mu}{p^3}} t_f = \frac{1}{\sqrt{p}} (\sqrt{r_1 + r_2 + c} \mp \sqrt{r_1 + r_2 - c}) \left[\pm \frac{3}{p} \sqrt{(r_1 + r_2 + c)(r_1 + r_2 - c)} + \frac{1}{p} (\sqrt{r_1 + r_2 + c} \mp \sqrt{r_1 + r_2 - c})^2 \right]$$

This equation can be simplified to

$$6\sqrt{\mu} t_f = (\sqrt{r_1 + r_2 + c} \mp \sqrt{r_1 + r_2 - c}) [2(r_1 + r_2) \pm \sqrt{(r_1 + r_2 + c)(r_1 + r_2 - c)}]$$

or, after some algebraic manipulation,

$$6\sqrt{\mu} t_f = (r_1 + r_2 + c)^{3/2} \mp (r_1 + r_2 - c)^{3/2} \quad (7.29)$$

This equation is known as *Euler's equation* and is remarkable in that it does not involve p . It yields the flight time between the points of the trajectory specified by $\theta_2 - \theta_1$ directly from the values of the radii r_1 and r_2 for these points. As mentioned before, the use of the plus- or minus-sign depends on the value of $\theta_2 - \theta_1$. The equation is of great importance in some methods for determining the elements of a parabolic orbit from (geocentric) observations.

Euler's equation can also be obtained as a limiting case of Lambert's equation derived in Section 6.7 for elliptical orbits. From (6.60) we conclude that if the semi-major axis becomes very large the angles α and β become very small, and we then find to first-order approximation:

$$\frac{1}{4} \alpha^2 = \frac{a_{min}}{a} \quad ; \quad \frac{1}{4} \beta^2 = K \frac{a_{min}}{a}$$

Substitution of (6.54) and (6.55) into these expressions gives

$$\alpha^2 = \frac{r_1 + r_2 + c}{a} \quad ; \quad \beta^2 = \frac{r_1 + r_2 - c}{a} \quad (7.30)$$

A series expansion of $\sin \alpha$ and $\sin \beta$ gives first-order approximations for the terms in (6.62):

$$\alpha - \sin \alpha = \frac{1}{6} \alpha^3 \quad ; \quad \beta - \sin \beta = \frac{1}{6} \beta^3 \quad (7.31)$$

From Figure 7.4 we conclude that only the cases *a*) and *c*) of the four cases identified for elliptical orbits exist for $a \rightarrow \infty$. Substitution of (7.30) and (7.31) into (6.62-1) and (6.62-3) yields

$$t_f = \frac{1}{6} \sqrt{\frac{a^3}{\mu}} \left[\left(\frac{r_1 + r_2 + c}{a} \right)^{3/2} \mp \left(\frac{r_1 + r_2 - c}{a} \right)^{3/2} \right]$$

or

$$6\sqrt{\mu} t_f = (r_1 + r_2 + c)^{3/2} \mp (r_1 + r_2 - c)^{3/2}$$

which is, of course, identical to (7.29).

8. HYPERBOLIC ORBITS

In Section 5.3 it was shown that for $e > 1$ the orbit of body i about body k is a hyperbola with body k at focus F (Figure 8.1). It was also shown that the equation for the orbit is

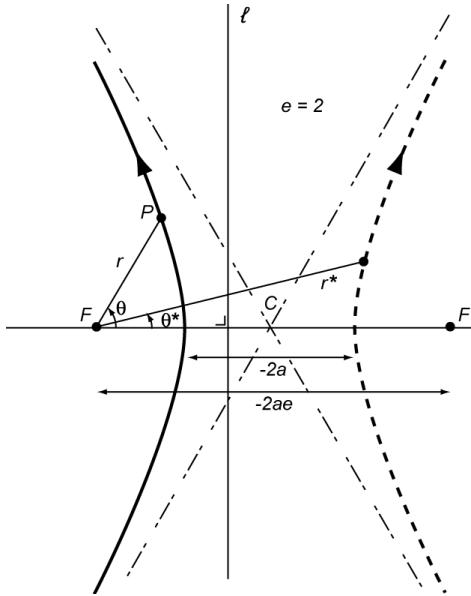


Figure 8.1: Geometry of a hyperbolic orbit.

$$r = \frac{p}{1 + e \cos \theta} \quad (8.1)$$

and that the second branch of the hyperbola (dashed curve in Figure 8.1), which has no physical meaning for celestial mechanics, is described by

$$r^* = \frac{-p}{1 - e \cos \theta^*} \quad (8.2)$$

For the important left branch in Figure 8.1, i.e. the branch that represents hyperbolic motion in a gravity field, the true anomaly is limited to

$$\cos \theta > -\frac{1}{e} \quad (8.3)$$

because distance, r , and semi-latus rectum, p , always have positive values.

8.1. Geometry, energy and angular momentum

For the ellipse, we could easily define the major axis. For a hyperbola, the concept of major axis is much less obvious. However, to simplify the computations for hyperbolic orbits, we define the major axis of the hyperbola as the negative value of the distance between the vertices of its two branches (Figure 8.1). This major axis is assigned a value of $2a$, which means that a has a negative value. This looks strange, but we should keep in mind that the sole purpose of this definition is to simplify the computation process. With this definition, we can write

$$2a = -(r_{2_{\theta^*}} - r_{1_{\theta=0}}) = \frac{p}{1-e} + \frac{p}{1+e} = \frac{2p}{1-e^2}$$

or

$$p = a(1-e^2) \quad (8.4)$$

Substitution of (8.4) into (8.1) yields for a hyperbolic orbit

$$r = \frac{a(1-e^2)}{1+e\cos\theta} \quad (8.5)$$

Thus, we conclude that because of our special definition of the major axis of a hyperbola, the equation in polar coordinates for a hyperbola is identical to the corresponding equation for an ellipse. Therefore, we may apply a number of equations derived for elliptical orbits directly to hyperbolic orbits. However, we should remember that for a hyperbola $1-e^2 < 0$, $a < 0$; while for an ellipse $1-e^2 > 0$, $a > 0$.

From analytical geometry we know two characteristics of a hyperbola that will be mentioned here without prove:

- The two branches of a hyperbola are each others mirror image with respect to the line through the center of the major axis (C) and perpendicular to the major axis.
- The asymptotes of the hyperbola cross point C and define an angle $\theta_{lim} = \arccos(-1/e)$ with the major axis. Since $e > 1$, this definition yields two solutions for θ_{lim} with $90^\circ < \theta_{lim} < 270^\circ$. For both values of θ_{lim} : $r = \infty$.

Just as for an elliptical orbit, we can write for the pericenter distance of a hyperbolic orbit:

$$r_p = a(1-e) \quad (8.6)$$

According to (6.21), the semi-major axis is given by

$$a = \frac{\mu/2}{\mu/r - V^2/2}$$

Because for a hyperbola the value of a is negative, we find for the total energy per unit of mass:

$$\mathcal{E} = \frac{V^2}{2} - \frac{\mu}{r} > 0 \quad (8.7)$$

So, the total energy of a spacecraft in a hyperbolic orbit is always positive. For comparison, in a parabolic orbit $\mathcal{E} = 0$; in an elliptical orbit $\mathcal{E} < 0$. Just as for the elliptical orbit, the relations between a , e and p and the quantities H and \mathcal{E} are given by

$$a = -\frac{\mu}{2\mathcal{E}} \quad (8.8)$$

$$e^2 = 1 + 2 \frac{H^2 \mathcal{E}}{\mu^2} \quad (8.9)$$

$$p = \frac{H^2}{\mu} \quad (8.10)$$

8.2. Velocity

In Section 5.6 we have derived equations for the radial velocity ((5.26)), normal velocity ((5.27)) and flight path angle ((5.28)) for any Keplerian orbit. In Section 6.3 we have derived an expression for the velocity ((6.21)) in elliptical orbits. With our definition of the major axis of a hyperbola, that equation equally applies to hyperbolic orbits. From the relations indicated, a number of conclusions can be drawn about the variation of the velocity and its components along a hyperbolic orbit:

- The velocity reaches a maximum value when the distance is minimum; i.e. at pericenter:

$$V_p^2 = \mu \left(\frac{2}{a(1-e)} - \frac{1}{a} \right) = \frac{\mu}{-a} \left(\frac{e+1}{e-1} \right) = V_{c_p}^2(e+1) \quad (8.11)$$

- The velocity reaches a minimum value for $r = \infty$:

$$V_\infty^2 = -\frac{\mu}{a} \quad (8.12)$$

So, at an infinitely large distance from body k , body i still has a finite velocity with respect to body k .

- The ratio between the maximum and minimum velocities in a hyperbolic orbit is given by

$$\frac{V_p}{V_\infty} = \sqrt{\frac{e+1}{e-1}} \quad (8.13)$$

This ratio is a function of the eccentricity, e , only and not of, e.g. a or p .

- The local velocity in a hyperbolic orbit is always larger than the local escape velocity (parabolic velocity) and is consequently also always larger than the local circular velocity. For the ratio between the velocity at pericenter and the circular velocity at pericenter follows

$$\frac{V_p}{V_{c_p}} = \sqrt{e+1} \quad (8.14)$$

- The radial velocity is only zero at pericenter. It reaches a maximum value for $\theta = 90^\circ, 270^\circ$; i.e. at the points of the orbit where the latus rectum intersects the hyperbola.
- The maximum normal velocity occurs at pericenter. The minimum value of the normal velocity component is zero and occurs at $\cos\theta = -1/e$.
- The flight path angle is zero at pericenter. It reaches a maximum value of $\pm 90^\circ$ at $\cos\theta = -1/e$. Then, body i is moving radially away from body k or towards body k .

A very interesting equation can be derived by substituting (7.3) and (8.12) into (6.21):

$$V^2 = V_{esc}^2 + V_\infty^2 \quad (8.15)$$

So, at every point along a hyperbolic orbit the instantaneous velocity is completely determined by the local escape velocity and by the velocity at infinity. This expression is very important for the analysis of interplanetary trajectories (Chapter 18). As will be explained in that Chapter, for a given interplanetary mission the value of V_∞ is known. Then, the required velocity V at the point where the spacecraft is injected into a hyperbolic escape trajectory can be computed directly from (8.15). For example, to inject a spacecraft at an altitude of 200 km above the Earth's surface into a hyperbolic orbit with $V_\infty = 4$ km/s, the spacecraft has to be accelerated to $V = 11.71$ km/s

($V_{esc} = 11.01$ km/s). To inject it into a hyperbolic orbit with V_∞ twice as large ($V_\infty = 8$ km/s), the spacecraft has to be accelerated to $V = 13.61$ km/s; or a velocity that is only 16% higher.

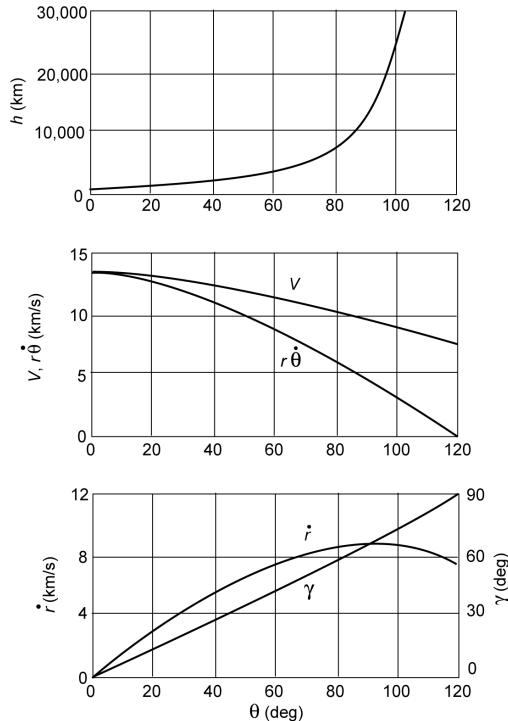


Figure 8.2: Variation of altitude, velocity and flight path angle in a hyperbolic orbit about the Earth ($h_p = 500$ km, $e = 2$).

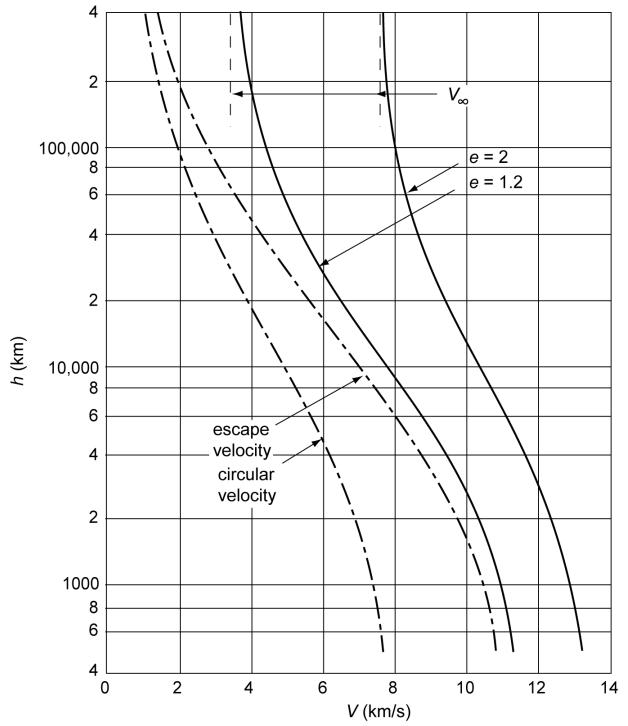


Figure 8.3: Variation of velocity versus altitude for two hyperbolic orbits about the Earth ($h_p = 500$ km), and curves for the local circular velocity and local escape velocity.

Figure 8.2 shows the variation of altitude, velocity and flight path angle along a hyperbolic orbit about the Earth. The eccentricity is chosen as $e = 2$ and the perigee height is set at $h_p = 500$ km. Note that after $\theta \approx 70^\circ$, the altitude rapidly increases with increasing values of θ ; for $\theta = 100^\circ$ the altitude is already about 25,000 km. The difference between the velocity and its normal component increases continuously. At an infinitely large altitude ($\theta_{lim} = 120^\circ$), the spacecraft still has a radial velocity of 7.6 km/s; the maximum radial velocity is approximately 8.8 km/s at $\theta = 90^\circ$. The flight path angle continuously increases to $\gamma = 90^\circ$ at $\theta = 120^\circ$. In Figure 8.3 the relation between velocity and altitude is presented for two hyperbolic orbits, both with a perigee altitude of 500 km. Also, curves representing the local circular velocity and the local escape velocity are plotted as a function of altitude. Note that a small velocity increase above the escape velocity at an altitude of 500 km causes a rather large value of V_∞ . From this Figure also another conclusion can be drawn: at an altitude of only about 400,000 km above the Earth's surface (approximately the distance to the Moon), the difference between the local velocity of the spacecraft and its velocity at an infinitely large distance from the Earth is very small. This property of hyperbolic orbits will turn out to be of fundamental importance for the analysis of interplanetary trajectories (Chapter 18).

8.3. Relation between position and time

To derive a relation between position and time for a hyperbolic orbit, we could start from (5.21)

and (8.1) and write

$$\Delta t = \sqrt{\frac{p^3}{\mu}} \int_0^{\theta_1} \frac{d\theta}{(1 + e \cos \theta)^2}$$

where the integration is performed from $\theta = 0^\circ$ (pericenter) to $\theta = \theta_1$. However, integration of this equation leads to the same kind of problems that were found in Chapter 6 for elliptical orbits. That is why we will follow an alternative method that is very similar to the method applied for elliptical orbits. To this end, a *hyperbolic anomaly*, F , defined as

$$r = a(1 - e \cosh F) \quad (8.16)$$

is introduced. Note that this equation contains a hyperbolic function. Because we do not use these functions very often, a summary of the definitions of a few hyperbolic functions and of their properties is presented below.

The sinh-, cosh-, and tanh-functions are defined as

$$\begin{aligned} \sinh x &= \frac{1}{2}(\exp(x) - \exp(-x)) & \cosh x &= \frac{1}{2}(\exp(x) + \exp(-x)) \\ \tanh x &= \frac{\sinh x}{\cosh x} = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)} = \frac{\exp(2x) - 1}{\exp(2x) + 1} \end{aligned} \quad (8.17)$$

where $\exp(x)$ denotes the exponential function of x , often indicated by e^x . Because the notation e is already used for the eccentricity of the orbit, the notation $\exp(x)$ is used here instead of e^x to avoid any confusion. Sketches of these functions are shown in Figure 8.4; they demonstrate that

$$-\infty < \sinh x < \infty \quad ; \quad \cosh x \geq 1 \quad ; \quad -1 < \tanh x < 1$$

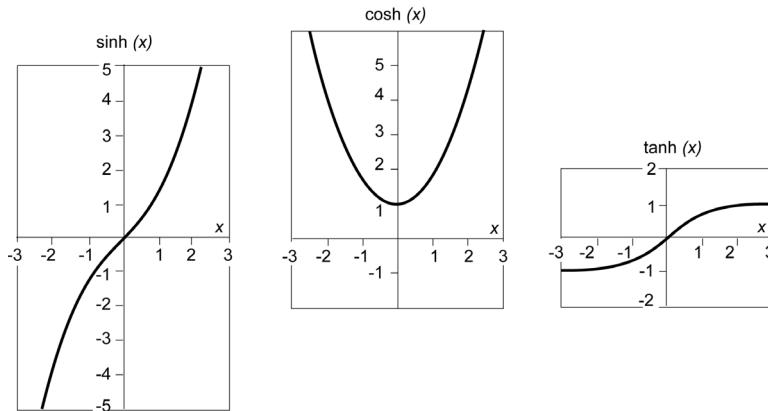


Figure 8.4: The hyperbolic functions $\sinh(x)$, $\cosh(x)$ and $\tanh(x)$.

For these hyperbolic functions there exist mutual relations and summation theorems, which show some similarity to those for trigonometric functions. Two important relations are

$$\cosh^2 x - \sinh^2 x = 1 \quad ; \quad \tanh^2\left(\frac{x}{2}\right) = \frac{\cosh x - 1}{\cosh x + 1} \quad (8.18)$$

From (8.17) we obtain the following expressions for the differential of two hyperbolic functions:

$$\frac{d}{dx}(\sinh x) = \cosh x \quad ; \quad \frac{d}{dx}(\cosh x) = \sinh x \quad (8.19)$$

We have defined the hyperbolic anomaly by means of (8.16). However, we still have to prove that (8.16) can be used to describe the entire hyperbolic orbit. That can be done as follows. Since $1 \leq \cosh F < \infty$, we conclude that (8.16) describes the distance in the hyperbolic orbit for the interval $a(1 - e) \leq r < \infty$, which is just the interval in which r varies for a hyperbola.

For an elliptical orbit, it was possible to directly interpret the eccentric anomaly as a measure of the angular position in the orbit. The hyperbolic anomaly, however, is not an angle (which is clear from the fact that F varies from $-\infty$ to ∞), but may be interpreted as the ratio of two areas. This is illustrated in Figure 8.5. Note that also the eccentric anomaly should be interpreted as the ratio of two areas, but this ratio corresponds to an angle due to the properties of a circle.

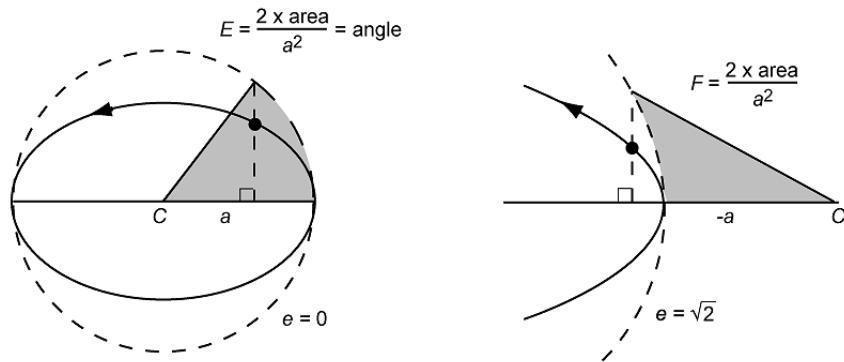


Figure 8.5: Geometric interpretation of the eccentric anomaly and hyperbolic anomaly.

Equating the right-hand sides of (8.5) and (8.16) yields a relation between the true anomaly and the hyperbolic anomaly:

$$\frac{1 - e^2}{1 + e \cos \theta} = 1 - e \cosh F$$

or

$$\cos \theta = \frac{1}{e} \left(\frac{1 - e^2}{1 - e \cosh F} - 1 \right)$$

Substitution of this equation into the trigonometric relation

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

leads, after some algebraic manipulation, to

$$\tan^2 \frac{\theta}{2} = \frac{(1 + e)(1 - \cosh F)}{(1 - e)(1 + \cosh F)}$$

Using (8.18-2), we obtain

$$\tan \frac{\theta}{2} = \pm \sqrt{\frac{e + 1}{e - 1}} \tanh \frac{F}{2}$$

This expression resembles the relation between the true anomaly and the eccentric anomaly for

elliptical orbits. Again, the relation is made unambiguous by selecting the plus-sign:

$$\tan \frac{\theta}{2} = \sqrt{\frac{e+1}{e-1}} \tanh \frac{F}{2} \quad (8.20)$$

which means that for $-90^\circ < \theta/2 < 0^\circ$ (fourth quadrant) F is negative; for $0^\circ \leq \theta/2 < 90^\circ$ (first quadrant) F is positive.

To derive a relation between hyperbolic anomaly and time, we write (6.21) as

$$\dot{r}^2 + r^2 \dot{\theta}^2 = \mu \left[\frac{2}{r} - \frac{1}{a} \right]$$

Using the expression for the angular momentum

$$H = r^2 \dot{\theta} = \sqrt{\mu p} = \sqrt{\mu a (1 - e^2)}$$

this expression can be written as

$$\dot{r}^2 = \left(\frac{\mu}{-a} \right) \left[\frac{a^2 (1 - e^2)}{r^2} - \frac{2a}{r} + 1 \right] \quad (8.21)$$

Differentiation of (8.16) yields

$$\dot{r} = -a e \dot{F} \sinh F \quad (8.22)$$

Substitution of (8.16) and (8.22) into (8.21) gives, after some analytical manipulation,

$$(e \cosh F - 1) \frac{dF}{dt} = \pm \sqrt{\frac{\mu}{-a^3}}$$

In our notation dF/dt is always positive; furthermore, $e \cosh F > 1$. This means that in the relation above the plus-sign should be used. Then, integration of the above equation gives

$$e \sinh F - F = \sqrt{\frac{\mu}{-a^3}} (t - \tau) \quad (8.23-1)$$

where τ is an integration constant. For $\theta = 0^\circ$ the spacecraft is at pericenter. According to (8.20), then $F = 0$, and according to (8.17) $\sinh F = 0$. Hence, the integration constant τ indicates the time of (last) pericenter passage. If an angular velocity \bar{n} is defined as

$$\bar{n} = \sqrt{\frac{\mu}{-a^3}} \quad (8.24)$$

then (8.23-1) can be written as

$$e \sinh F - F = \bar{n} (t - \tau) \quad (8.23-2)$$

When, just as we did for parabolic orbits, a kind of mean anomaly \bar{M} is introduced:

$$\bar{M} = \bar{n} (t - \tau) \quad (8.25)$$

then (8.23-2) can be written as

$$e \sinh F - F = \bar{M} \quad (8.23-3)$$

Note that this equation resembles Kepler's equation for elliptical orbits. Just as for an elliptical orbit, to determine *when* body i is at a certain position in its orbit, is no problem. However, if the question is asked *where* body i is at a certain time, a transcendental equation in F has to be solved. Like for Kepler's equation, it is not possible to solve this equation analytically in a closed form. In addition, it is not possible to formulate a series expansion that is applicable for the complete range of hyperbolic anomalies. This is a direct consequence of the fact that $1 < e < \infty$ and that the functions $\sinh F$ and $M - F$ are non-periodic. Therefore, one is forced to use graphical or numerical methods.

It is also possible to derive a version of Lambert's theorem (Section 6.7) for hyperbolic orbits by repeating every step in the derivation of that theorem, but now using the relevant expressions for hyperbolic orbits. This derivation will not be given in this Chapter and only the result is presented:

$$t_f = \sqrt{\frac{-\alpha^3}{\mu}} [(\sinh \alpha' - \alpha') \mp (\sinh \beta' - \beta')] \\ \sinh\left(\frac{\alpha'}{2}\right) = \frac{1}{2} \left(\frac{r_1 + r_2 + c}{-\alpha} \right)^{1/2} ; \quad \sinh\left(\frac{\beta'}{2}\right) = \frac{1}{2} \left(\frac{r_1 + r_2 - c}{-\alpha} \right)^{1/2}$$

where r_1 and r_2 are the radial distances at t_1 and t_2 , $t_f = t_2 - t_1$, and c is the chord between the points where the spacecraft is located at t_1 and t_2 . The plus-sign holds for $\theta_2 - \theta_1 < 180^\circ$ and the minus-sign holds for $\theta_2 - \theta_1 > 180^\circ$.

8.4. Numerical and graphical solution of the transcendental equation

To prove that the transcendental equation (8.23) has only one solution for F , we analyze the function

$$\mathcal{F}(F, \bar{M}) = e \sinh F - F - \bar{M}$$

A series expansion of $\sinh F$ reads

$$\sinh F = F + \frac{F^3}{3!} + \frac{F^5}{5!} + \frac{F^7}{7!} + \dots$$

These equations show that

$$\mathcal{F}(-\infty, \bar{M}) = -\infty ; \quad \mathcal{F}(\infty, \bar{M}) = \infty$$

Furthermore, $d\mathcal{F}/dF = e \cosh F - 1 > 0$. This means that there is only one value of F for which $\mathcal{F}(F, \bar{M}) = 0$ for an arbitrary value of \bar{M} .

To find the solution of (8.23), the simple method of successive approximations according to the scheme

$$F_{k+1} = e \sinh F_k - \bar{M}$$

which was used for elliptical orbits, does not converge for a hyperbolic orbit because $e > 1$. Instead, a *Newton-Raphson* procedure can be used:

$$F_{k+1} = F_k - \frac{e \sinh F_k - F_k - \bar{M}}{e \cosh F_k - 1}$$

For small values of \bar{M} we may take $F = 0$ as starting value for the iteration process. However, this may lead to a poor convergence of the iteration process and for $F > 6$ even to computational problems. Therefore, for a safe and quick convergence an appropriately selected starting value of F should be used. For relatively small values of F , an appropriate starting value can be found by using the Taylor series expansion up to order F^7 :

$$e \sinh F - F = e \left\{ \frac{1}{5040} F^7 + \frac{1}{120} F^5 + \frac{1}{6} F^3 \right\} + F(e - 1)$$

For all values of $e > 1$ and $|F| < 2.7$, we may write with an accuracy of better than 30%:

$$e \sinh F - F = \frac{1}{6} e F^3 + (e - 1) F$$

This leads to the third-degree equation

$$F^3 + \frac{6(e-1)}{e} F - \frac{6\bar{M}}{e} = 0$$

The real root of this equation can be found analytically from

$$x = \sqrt{\frac{8(e-1)}{e}} \quad ; \quad y = \frac{1}{3} \operatorname{arcsinh} \left(\frac{3\bar{M}}{x(e-1)} \right) \quad ; \quad F = x \sinhy \quad (8.26-1)$$

where the $\operatorname{arcsinh}$ -function can be written as

$$\operatorname{arcsinh} x = \ln \left[x + \sqrt{x^2 + 1} \right]$$

We can also apply the standard method for solving a third-degree equation and then find the real root from

$$x = \frac{3\bar{M}}{e} \quad ; \quad y = \sqrt{x^2 + \left(\frac{2(e-1)}{e} \right)^3} \quad ; \quad F = \sqrt[3]{x+y} + \sqrt[3]{x-y} \quad (8.26-2)$$

Note that $x - y$ is negative and so the second cube root yields a negative value.

For $|F| > 2.7$, an appropriate starting value of F can be found from

$$e \sinh F - F = \frac{1}{2} e [\exp(F) - \exp(-F)] - F$$

For all values of $e > 1$ and $F > 2.7$, we may write with an accuracy of better than 37%:

$$e \sinh F - F = \frac{1}{2} e \exp(F)$$

This leads to the expression

$$F = \ln \left(\frac{2\bar{M}}{e} \right) \quad (8.27)$$

Similarly, for $F < -2.7$ we may write

$$F = -\ln\left(\frac{-2\bar{M}}{e}\right) \quad (8.28)$$

However, at the start of the iteration process we do not know the value of F and, consequently, the criterion for the selection of the expression to be used for the computation of the starting value of F must be based on the known values of e and \bar{M} . A numerical analysis of (8.23-3), in combination with (8.26) to (8.28), shows that the quantity $\bar{M}=6e$ can be used to determine which expression has to be used to find an appropriate starting value of F for the iterative Newton-Raphson procedure:

$|\bar{M}| \leq 6e$: F should be computed with (8.26)

$\bar{M} > 6e$: F should be computed with (8.27)

$\bar{M} < -6e$: F should be computed with (8.28).

Various authors have proposed other iteration schemes to solve (8.23-3). These will not be presented in this book.

It is also possible to find a graphical solution for the transcendental equation (8.23), just like for elliptical orbits. For that purpose, first the hyperbolic function $y = \sinh x$ is drawn (Figure 8.6), after which the point $(-\bar{M}, 0)$ is marked on the X -axis. Subsequently, the point $(-\bar{M}+e, 1)$ is marked on the line $y = 1$ and a straight line is drawn through these two points. Since for a hyperbola $e > 1$, the slope of this line is always smaller than 45° . The intersection of the curve $\sinh x$ and this line yields the required value of F .

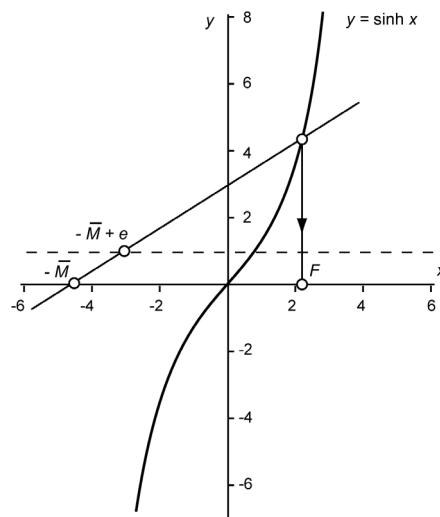


Figure 8.6: Graphical solution of the hyperbolic anomaly.

The correctness of this graphical method can be proved as follows. The general equation of the line is

$$y = ax + b$$

This line passes through the points $(-\bar{M}, 0)$ and $(-\bar{M}+e, 1)$, so

$$0 = -a\bar{M} + b ; 1 = -a(\bar{M} - e) + b$$

or $a = 1/e$, $b = \bar{M}/e$. As a result, the equation of the line is

$$y = \frac{F + \bar{M}}{e}$$

and for the intersection we find

$$e \sinh F - F = \bar{M}$$

which is identical to (8.23-3).

In Figure 8.7, true anomaly and altitude are plotted as a function of time after perigee passage for two hyperbolic orbits, both with their perigee at 400 km above the Earth's surface. Note that during the first half hour the true anomaly increases rapidly, which means that the spacecraft has a large angular velocity. After a few hours, the rate of increase of the true anomaly is only very small and the spacecraft is moving almost radially away from the Earth. The altitude increases almost linearly with time.

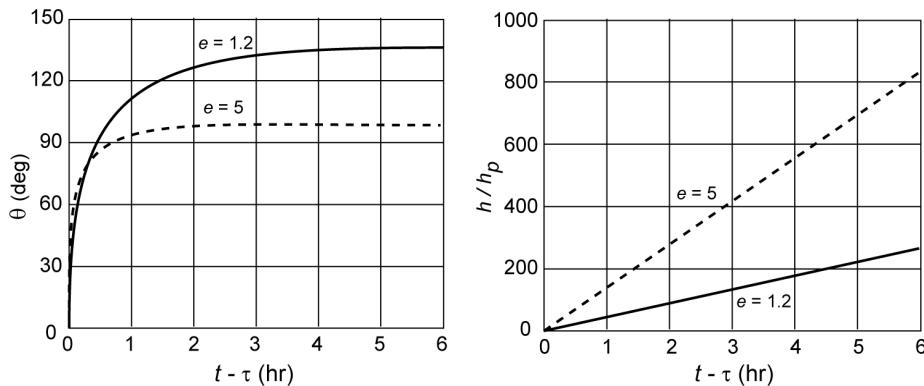


Figure 8.7: Variation of true anomaly and altitude in a hyperbolic orbit about the Earth as a function of time after perigee passage ($h_p = 400$ km).

8.5. Comparison of the expressions for elliptical and hyperbolic orbits

To conclude, it is interesting to point out that many expressions for elliptical orbits and for hyperbolic orbits can be derived from each other when the substitution

$$F = iE \tag{8.29}$$

is applied, where $i = \sqrt{-1}$, and E and F are the eccentric and hyperbolic anomalies. This can be shown as follows.

We know that the following relations hold for trigonometric and hyperbolic functions:

$$\sin x = \frac{1}{2i} (\exp(ix) - \exp(-ix)) ; \quad \cos x = \frac{1}{2} (\exp(ix) + \exp(-ix))$$

$$\sinh x = \frac{1}{2} (\exp(x) - \exp(-x)) ; \quad \cosh x = \frac{1}{2} (\exp(x) + \exp(-x))$$

So, we may write with (8.29):

$$\sinh F = \sinh(iE) = \frac{1}{2} (\exp(iE) - \exp(-iE)) = i \sin E$$

$$\cosh F = \cosh(iE) = \frac{1}{2} (\exp(iE) + \exp(-iE)) = \cos E$$

$$\tanh \frac{F}{2} = \frac{\sinh F/2}{\cosh F/2} = i \tan \frac{E}{2}$$

When these relations are substituted into the expressions for the relation between distance and hyperbolic anomaly, true anomaly and hyperbolic anomaly, and time and hyperbolic anomaly, the following expressions result:

$$r = a(1 - \cosh F) = a(1 - \cos E)$$

$$\tan \frac{\theta}{2} = \sqrt{\frac{e+1}{e-1}} \tanh \frac{F}{2} = \sqrt{\frac{e+1}{e-1}} i \tan \frac{E}{2} = \sqrt{\frac{1+e}{1-e}} \tan \frac{E}{2}$$

$$e \sinh F - F = \sqrt{\frac{\mu}{-a^3}} (t - \tau) = i e \sin E - i E$$

From the latter expression follows

$$E - e \sin E = \sqrt{\frac{\mu}{a^3}} (t - \tau)$$

So, when we start from the expressions for hyperbolic orbits and substitute $F = iE$, then the corresponding expressions for elliptical orbits are obtained. Similarly, from the expressions for elliptical orbits the corresponding ones for hyperbolic orbits can be derived.

A compilation of some expressions that have been derived for elliptical, parabolic and hyperbolic orbits is presented in Appendix C.

9. RELATIVE MOTION OF TWO SATELLITES

When it is assumed that the only force acting on a satellite is the mutual gravitational attraction between the Earth and the satellite, and that the gravitational potential of the Earth can be described by the Newton potential, $-\mu/r$, then the satellite moves in a Keplerian orbit about the Earth. In Chapters 5 through 8, these Keplerian orbits have been analyzed. However, sometimes one is not that much interested in the motion of the satellite relative to a (quasi-)inertial reference frame with origin at the center of the Earth, but more in the motion of a satellite relative to another satellite. Examples of such cases are *rendez-vous missions* (Chapter 15), the analysis of the motion of a body in the vicinity of the International Space Station, and the analysis of the motion of a cluster of satellites used for e.g. space interferometry. In addition, the theory of relative motion can be applied when the actual perturbed motion of a satellite is compared to the motion of the same satellite in its nominal unperturbed orbit. In that case, one, in fact, considers the motion of that satellite with respect to a fictitious second satellite.

In this Chapter, a linearized theory will be developed that describes the relative motion of two satellites for the case that the distance between the satellites will not become too large.

9.1. Clohessy-Wiltshire equations

It is assumed that a satellite (satellite 1) moves in a circular orbit about the Earth and that we want to analyze the motion of a second satellite (satellite 2) with respect to the first satellite. It is also assumed that the motion of both satellites is only affected by the gravity field of the Earth, which is described by the Newton potential, $-\mu/r$. In that case, we can use the equations that have been derived for the motion of the third body in the circular restricted three-body problem (Section 3.3); the geometry of that problem is shown in Figure 3.4.

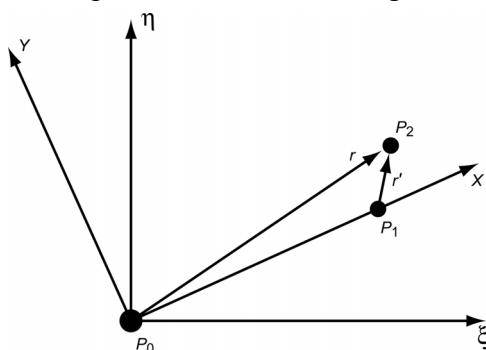


Figure 9.1: Geometry of relative motion. The XYZ reference frame rotates with uniform angular velocity n , about the ζ -axis of the (quasi-)inertial reference frame $\xi\eta\zeta$. The origin of both frames coincides with the Earth's center of mass; P_0 denotes the Earth, P_1 and P_2 the satellites 1 and 2.

In contrast to the situation in the circular restricted three-body problem, in the present case the mass of the ‘second main body’, i.e. satellite 1, is also negligible with respect to the mass of the ‘first main body’, i.e. the Earth. So, we can write for the mass ratio μ defined by (3.42): $\mu = 0$, and we may assume that the center of mass of the system coincides with the center of the Earth. The geometry of this problem and the notation used are indicated in Figure 9.1. The Earth is denoted by P_0 ; the satellites 1 and 2 by P_1 and P_2 , respectively. The motion of satellite 2 is described with respect to a rotating reference frame XYZ with its origin at the mass center of the Earth, and with the Z-axis perpendicular to the orbital plane of satellite 1. The X-axis points towards satellite 1 and, because that satellite moves in a circular orbit about the Earth, the

reference frame rotates with constant angular velocity about the ζ -axis of the inertial reference frame $\xi\eta\zeta$. The position of satellite 1 is fixed relative to this rotating reference frame and, in the dimensionless quantities introduced in Section 3.3, the position of satellite 1 is given by $x_1 = 1$, $y_1 = z_1 = 0$. The X -, Y - and Z -axis specify the *radial*, *along-track* and *cross-track* directions of the motion of satellite 2 with respect to satellite 1. The positive radial direction is outward, the positive along-track direction is along the direction of motion of satellite 1 and the positive cross-track direction complements a right-handed reference frame. The motion of satellite 2 is, according to (3.46) and using the simplifications mentioned above, described by the following equations:

$$\begin{aligned}\ddot{x} - 2\dot{y} &= \left(1 - \frac{1}{r^3}\right)x \\ \ddot{y} + 2\dot{x} &= \left(1 - \frac{1}{r^3}\right)y \\ \ddot{z} &= -\frac{1}{r^3}z\end{aligned}\tag{9.1}$$

where x, y, z are the components of the position vector \bar{r} (Figure 9.1). It is emphasized that (9.1) is written in the dimensionless quantities introduced in Section 3.3. For the relative coordinates of satellite 2, we may write

$$x = 1 + x' \quad ; \quad y = y' \quad ; \quad z = z'\tag{9.2}$$

where x', y' and z' are the components of vector \bar{r}' (Figure 9.1). For the term $1/r^3$ in (9.1) we then can write

$$\frac{1}{r^3} = [(1 + x')^2 + y'^2 + z'^2]^{-3/2} = [1 + 2x' + (x'^2 + y'^2 + z'^2)]^{-3/2}$$

Now, we assume that the distance between both satellites is at all times small when compared to the radius of the orbit of satellite 1; i.e. $r' \ll 1$. Then, the relation given above may be linearized and we find

$$\frac{1}{r^3} \approx 1 - 3x'\tag{9.3}$$

Substitution of (9.2) and (9.3) into (9.1) gives, neglecting terms that are small of the second order,

$$\begin{aligned}\ddot{x}' - 2\dot{y}' &= 3x' \\ \ddot{y}' + 2\dot{x}' &= 0 \\ \ddot{z}' &= -z'\end{aligned}\tag{9.4}$$

For the analysis of the relative motion of two satellites it is preferable to write the equations in the usual physical units. With the definition of the dimensionless quantities given in Section 3.3, we then find after some algebraic manipulation

$$\ddot{x}' - 2n_1\dot{y}' - 3n_1^2x' = 0\tag{9.5}$$

$$\begin{aligned}\ddot{y}' + 2n_1 \dot{x}' &= 0 \\ \ddot{z}' + n_1^2 z' &= 0\end{aligned}\tag{9.5}$$

where n_1 is the angular motion ((6.27)) of satellite 1 in its orbit about the Earth. We conclude that the motion in the Z' -direction is a pure harmonic oscillation that is uncoupled from the motion in the X' - and Y' -direction.

From these equations, an integral of motion can be found. When the three equations are multiplied by \dot{x}' , \dot{y}' and \dot{z}' respectively, and the results are subsequently added, we find

$$\dot{x}' \ddot{x}' + \dot{y}' \ddot{y}' + \dot{z}' \ddot{z}' = (3x' \dot{x}' - z' \dot{z}') n_1^2$$

or

$$\frac{d}{dt} \left(\frac{1}{2} \dot{x}'^2 + \frac{1}{2} \dot{y}'^2 + \frac{1}{2} \dot{z}'^2 \right) = n_1^2 \frac{d}{dt} \left(\frac{3}{2} x'^2 - \frac{1}{2} z'^2 \right)$$

Integration leads to

$$V'^2 - n_1^2 (3x'^2 - z'^2) = C\tag{9.6}$$

where V' is the velocity of satellite 2 relative to satellite 1. Note that $3x'^2 - z'^2 = C_1$, where C_1 is an arbitrary constant, represents hyperbolic surfaces. So, (9.6) indicates that at all points where the relative trajectory intersects such a hyperbolic surface for a particular value of C_1 , the relative velocity of satellite 2 has the same value.

When, in addition to the attracting force described by the Newton gravitational potential, also another force is acting on the satellite, the equations for the relative motion read

$$\begin{aligned}\ddot{x}' - 2n_1 \dot{y}' - 3n_1^2 x' &= f_{x'} \\ \ddot{y}' + 2n_1 \dot{x}' &= f_{y'} \\ \ddot{z}' + n_1^2 z' &= f_{z'}\end{aligned}\tag{9.7}$$

where $f_{x'}$, $f_{y'}$, $f_{z'}$ are the components of the additional force, per unit of mass. These equations were first found by G.W. Hill (1838-1914) around 1878 and are known in celestial mechanics as the *Hill equations*. They were re-discovered in the era of spaceflight for the analysis of rendez-vous missions by W.H. Clohessy (-) and R.S. Wiltshire (-) around 1960, and are therefore also known as the *Clohessy-Wiltshire equations*. The equations describe the linearized motion of satellite 2 with respect to satellite 1, which moves in a circular orbit about the Earth. The terms that contain the factors \dot{x}' and \dot{y}' represent *Coriolis accelerations*, while the terms that contain the factor n_1^2 describe *centrifugal accelerations*. In the following, for simplicity we drop the index ' $'$; however, the quantities x , y and z still indicate relative coordinates. So, in fact, we introduce an *XYZ* reference frame centered at satellite 1, with the X -axis directed radially outwards and the Y -axis directed in the along-track direction. Also the index 1 is left out; however, the quantity n still indicates the angular velocity of satellite 1 in its orbit about the Earth.

Equations (9.7) can be solved analytically in a closed form, if the (additional) accelerations on the right-hand side of (9.7) are constant (or zero), or if these accelerations can be expressed by suitable analytical functions, such as Fourier series. In the following Section, a solution will

be presented for the case that these accelerations are constant.

9.2. Analytical solution of the Clohessy-Wiltshire equations

We start with the third differential equation of (9.7). The solution of the homogeneous part of this equation is

$$z = A_1 \sin nt + B_1 \cos nt$$

and a particular solution is

$$z = \frac{f_z}{n^2}$$

So, the general solution can be written as

$$z = A_1 \sin nt + B_1 \cos nt + \frac{f_z}{n^2} \quad (9.8)$$

The solutions of the first two differential equations of (9.7) can be found as follows. From the second equation follows

$$\dot{y} = -2nx + f_y t + C_2 \quad (9.9)$$

Substitution of this result into the first equation results in

$$\ddot{x} + n^2 x = f_x + 2nf_y t + 2nC_2$$

The solution of the homogeneous part of this equation is

$$x = A_2 \sin nt + B_2 \cos nt$$

and a particular solution is

$$x = \frac{1}{n^2} (f_x + 2nf_y t + 2nC_2)$$

The general solution then becomes

$$x = A_2 \sin nt + B_2 \cos nt + \frac{f_x}{n^2} + \frac{2f_y t}{n} + \frac{2C_2}{n} \quad (9.10)$$

Substitution of this relation into (9.9) yields

$$\dot{y} = -2A_2 n \sin nt - 2B_2 n \cos nt - \frac{2f_x}{n} - 3f_y t - 3C_2$$

Integration leads to

$$y = 2A_2 \cos nt - 2B_2 \sin nt - \frac{2f_x}{n} t - \frac{3}{2} f_y t^2 - 3C_2 t + D_2 \quad (9.11)$$

The constants A_1 , A_2 , B_1 , B_2 , C_2 and D_2 in the expressions (9.8), (9.10) and (9.11) can be determined by substitution of the values of position and velocity at $t = 0$: $x = x_0$, $y = y_0$, $z = z_0$, $\dot{x} = \dot{x}_0$, $\dot{y} = \dot{y}_0$, $\dot{z} = \dot{z}_0$. This results in the relations

$$\begin{aligned}x_0 &= B_2 + \frac{f_x}{n^2} + \frac{2C_2}{n} ; \quad \dot{x}_0 = A_2 n + \frac{2f_y}{n} \\y_0 &= 2A_2 + D_2 ; \quad \dot{y}_0 = -2B_2 n - \frac{2f_x}{n} - 3C_2 \\z_0 &= B_1 + \frac{f_z}{n^2} ; \quad \dot{z}_0 = A_1 n\end{aligned}$$

from which the parameters A_1, A_2, B_1, B_2, C_2 and D_2 can be solved:

$$\begin{aligned}A_1 &= \frac{\dot{z}_0}{n} ; \quad A_2 = \frac{\dot{x}_0}{n} - \frac{2f_y}{n^2} \\B_1 &= z_0 - \frac{f_z}{n^2} ; \quad B_2 = -3x_0 - \frac{f_x}{n^2} - \frac{2\dot{y}_0}{n} \\C_2 &= 2nx_0 + \dot{y}_0 ; \quad D_2 = y_0 - 2\frac{\dot{x}_0}{n} + \frac{4f_y}{n^2}\end{aligned}$$

Substitution of these relations into (9.8), (9.10) and (9.11) yields, after some algebraic manipulation,

$$\begin{aligned}x &= x_0(4 - 3\cos nt) + \frac{\dot{x}_0}{n} \sin nt + \frac{2\dot{y}_0}{n}(1 - \cos nt) \\&\quad + \frac{f_x}{n^2}(1 - \cos nt) + \frac{2f_y}{n^2}(nt - \sin nt) \\y &= y_0 - \frac{\dot{y}_0}{n}(3nt - 4\sin nt) - 6x_0(nt - \sin nt) - 2\frac{\dot{x}_0}{n}(1 - \cos nt) \\&\quad - \frac{2f_x}{n^2}(nt - \sin nt) + \frac{2f_y}{n^2}(2 - \frac{3}{4}n^2t^2 - 2\cos nt) \\z &= z_0 \cos nt + \frac{\dot{z}_0}{n} \sin nt + \frac{f_z}{n^2}(1 - \cos nt)\end{aligned}\tag{9.12}$$

where the quantity nt indicates the central angle that satellite 1 has covered in the time interval $0 - t$. By differentiation, we obtain for the velocity components

$$\begin{aligned}\dot{x} &= 3x_0 n \sin nt + \dot{x}_0 \cos nt + 2\dot{y}_0 \sin nt + \frac{f_x}{n} \sin nt + 2\frac{f_y}{n}(1 - \cos nt) \\&\quad - \frac{2f_x}{n}(1 - \cos nt) - 2\frac{f_y}{n}(\frac{3}{2}nt - 2\sin nt) \\&= -\dot{y}_0(3 - 4\cos nt) - 6x_0 n(1 - \cos nt) - 2\dot{x}_0 \sin nt \\&\quad - \frac{2f_x}{n}(1 - \cos nt) - 2\frac{f_y}{n}(\frac{3}{2}nt - 2\sin nt)\end{aligned}\tag{9.13}$$

$$\dot{z} = -z_0 n \sin nt + \dot{z}_0 \cos nt + \frac{f_z}{n} \sin nt \quad (9.13)$$

Equations (9.12) show that, in the absence of the additional force f :

- The period of the periodic terms is equal to the orbital period of satellite 1.
- The motions in radial and cross-track directions are purely harmonic.
- The motion in the along-track direction includes a component that increases linearly with time (*drift*), if the values of x_0 and \dot{y}_0 are not equal to zero.
- The oscillation in the along-track direction is a quarter period ahead of the oscillation in the radial direction.
- The amplitude of the along-track oscillation is twice the amplitude of the radial oscillation.

When also the additional force is acting on satellite 2, we find that:

- The component f_x yields, in addition to periodic terms, a constant contribution in the radial direction and a linear drift in the along-track direction.
- The component f_y yields, in addition to periodic terms, a linear drift in the radial direction, and a constant term and a quadratic drift in the along-track direction.
- The component f_z yields, in addition to a periodic term, a constant contribution in the cross-track direction.

For short time intervals, we may write

$$\sin nt \approx nt \quad ; \quad \cos nt \approx 1 - \frac{1}{2}n^2 t^2$$

where terms of the order $(nt)^3$ are neglected. In this case, (9.12) can be simplified to

$$\begin{aligned} x &= x_0 + \dot{x}_0 t + \frac{1}{2} \left(3x_0 + 2 \frac{\dot{y}_0}{n} + \frac{f_x}{n^2} \right) (nt)^2 \\ y &= y_0 + \dot{y}_0 t - \left(\frac{\dot{x}_0}{n} - \frac{1}{2} \frac{f_y}{n^2} \right) (nt)^2 \\ z &= z_0 + \dot{z}_0 t - \frac{1}{2} \left(z_0 - \frac{f_z}{n^2} \right) (nt)^2 \end{aligned}$$

These relations show that when $nt \ll 1$, and when the initial values of the relative position and velocity components and of the additional force components are sufficiently small, the relative motion of satellite 2 is in first approximation linear.

Equations (9.12) and (9.13) were, although in another set of coordinates, programmed in the onboard computer of the manned Gemini capsule for the computation of the first rendez-vous trajectories (Section 15.1). The rapid developments in computer hardware made it possible for later missions to numerically integrate the complete set of equations of motion onboard.

It is recalled that the Clohessy-Wiltshire equations are linearized equations of motion relative to a spacecraft that moves in a circular orbit about the Earth, and therefore only generate accurate results for relatively short time intervals and relatively small distances between the two satellites. As rule of thumb, we may use the criterion that the equations are (reasonably) accurate for a period of up to one revolution of satellite 1 in its orbit about the Earth ($nt < 2\pi$) and for $|x|/r_1 < 8 \cdot 10^{-3}$, $|y|/r_1 < 6 \cdot 10^{-2}$, $|z|/r_1 < 6 \cdot 10^{-2}$, where r_1 is the radius of the orbit of satellite 1. Outside this range, the accuracy of the equations deteriorate rapidly. To increase the accuracy of

the analytical theory, a number of alternative sets of equations have been developed that contain higher-order terms. This topic will, however, not be discussed in this Chapter.

9.3. Characteristics of unperturbed relative motion

In case no additional force acts on satellite 2, the equations for the relative motion (9.12) can be written as

$$\begin{aligned} x &= \left(4x_0 + \frac{2\dot{y}_0}{n} \right) + \frac{\dot{x}_0}{n} \sin nt - \left(3x_0 + \frac{2\dot{y}_0}{n} \right) \cos nt \\ y &= \left(y_0 - \frac{2\dot{x}_0}{n} \right) - 3 \left(2x_0 + \frac{\dot{y}_0}{n} \right) nt + 2 \left(3x_0 + \frac{2\dot{y}_0}{n} \right) \sin nt + \frac{2\dot{x}_0}{n} \cos nt \\ z &= z_0 \cos nt + \frac{\dot{z}_0}{n} \sin nt \end{aligned} \quad (9.14)$$

The last equation shows that, as was already mentioned before, the motion in the cross-track direction is a pure harmonic oscillation, uncoupled from the motions in the radial and along-track directions. The first two equations (9.14) can be written in a more compact form. With the notation

$$C_1 = 4x_0 + \frac{2\dot{y}_0}{n} ; \quad C_2 = y_0 - \frac{2\dot{x}_0}{n} ; \quad C_3 = 3x_0 + \frac{2\dot{y}_0}{n} ; \quad C_4 = \frac{\dot{x}_0}{n} \quad (9.15)$$

they can be written as

$$\begin{aligned} x - C_1 &= C_4 \sin nt - C_3 \cos nt \\ \frac{1}{2}y - \frac{1}{2}C_2 + \frac{3}{4}C_1 nt &= C_3 \sin nt + C_4 \cos nt \end{aligned}$$

Squaring these relations and subsequently adding the resulting relations yields for $C_3^2 + C_4^2 \neq 0$:

$$\frac{(x - C_1)^2}{C_3^2 + C_4^2} + \frac{(y - C_2 + \frac{3}{2}C_1 nt)^2}{4(C_3^2 + C_4^2)} = 1 \quad (9.16)$$

This equation shows that in the XY -plane the motion of satellite 2 with respect to satellite 1 is an ellipse, where the y -coordinate of the center of the ellipse, C , changes linearly with time. Note the similarity with the Ptolemaic planetary model discussed in Section 5.5. In that model the (relative) motion of a planet about the Earth is described by a circle (deferent) with superimposed a second circle (epicycle). Here, the linearized motion of satellite 2 relative to satellite 1, and relatively close to satellite 1, is described by a straight line parallel to the Y -axis with superimposed an ellipse, while the absolute motion of satellite 1 is a circle about the Earth and of satellite 2, in general, an ellipse about the Earth. Figure 9.2 shows a sketch of the absolute (left) and the relative (right) motion of satellite 2 in the XY -plane.

The coordinates of the center of the ellipse, which describes the relative motion, are given by

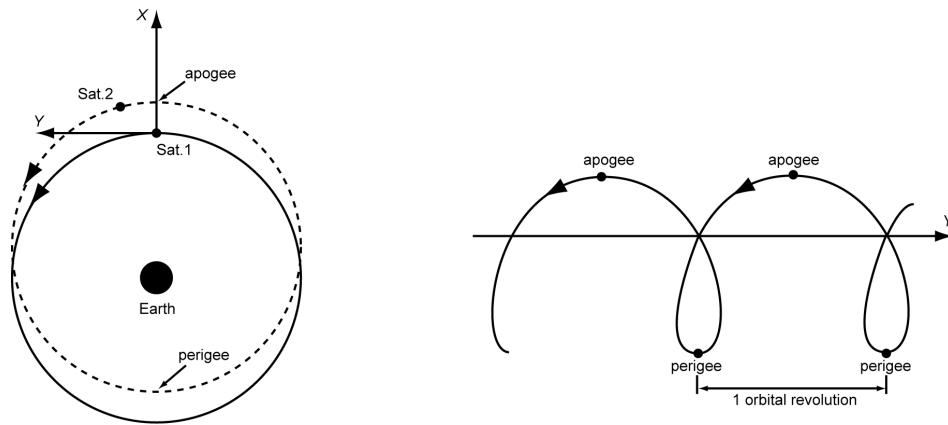


Figure 9.2: Sketch of a characteristic example of the absolute (left) and relative (right) motion of satellite 2 in the XY-plane.

$$\begin{aligned} x_C &= C_1 = 4x_0 + \frac{2\dot{y}_0}{n} \\ y_C &= C_2 - \frac{3}{2}C_1 nt = y_0 - 2\frac{\dot{x}_0}{n} - \frac{3}{2}\left(4x_0 + \frac{2\dot{y}_0}{n}\right)nt \end{aligned} \quad (9.17)$$

from which follows that the center of the ellipse moves with a constant velocity:

$$V_{y_c} = -3(2nx_0 + \dot{y}_0) \quad (9.18)$$

parallel to the Y -axis. The major axis of this ellipse is oriented parallel to the Y -axis; the magnitude of the semi-major axis is given by

$$a = 2\sqrt{C_3^2 + C_4^2} = 2\sqrt{\left(3x_0 + 2\frac{\dot{y}_0}{n}\right)^2 + \left(\frac{\dot{x}_0}{n}\right)^2} \quad (9.19)$$

The minor axis is parallel to the X -axis; the magnitude of the semi-minor axis is given by

$$b = \frac{1}{2}a \quad (9.20)$$

Combination of (6.7) and (9.20) yields for the eccentricity of the relative elliptical orbit $e = \frac{1}{2}\sqrt{3}$; this value is independent of the initial conditions.

As an example, Figure 9.3 shows the variations of x and y for the case that satellite 1 moves in a circular orbit at an altitude of 400 km about the Earth. The period of this orbit is 92.5 min and the angular velocity is $n = 3.892^\circ/\text{min}$. For satellite 2 the following initial conditions have been adopted: $x_0 = y_0 = z_0 = 0$; $\dot{x}_0 = -0.5 \text{ m/s}$, $\dot{y}_0 = 0.5 \text{ m/s}$, $\dot{z}_0 = 0$. Note that the relative motion consists of a periodic component superimposed on a drift component. Below, we will analyze three special cases of relative motion.

Concentric coplanar circular orbits

We now consider the case that $C_3^2 + C_4^2 = 0$ or $C_3 = C_4 = 0$. If the initial conditions satisfy the relations

$$\dot{x}_0 = 0 \quad ; \quad \dot{y}_0 = -\frac{3}{2}n x_0 \quad ; \quad z_0 = 0 \quad ; \quad \dot{z}_0 = 0 \quad (9.21)$$

then $C_3 = C_4 = 0$, the trigonometric terms in (9.14) are zero, and (9.14) simplifies to

$$x = x_0 \quad ; \quad y = y_0 - \frac{3}{2}n x_0 t \quad ; \quad z = 0 \quad (9.22)$$

So, for the initial conditions specified by (9.21) satellite 2 moves in a circular orbit about the Earth; this motion takes place in the XY -plane. When $x_0 > 0$, the radius of this circular orbit is larger than that of satellite 1, and satellite 2 will gradually lag behind satellite 1 in the along-track direction. When $x_0 < 0$, the radius of the circular orbit of satellite 2 is smaller than that of satellite 1, and satellite 2 will gradually lead satellite 1 in the along-track direction. In Figure 9.4 the relative motion of satellite 2 is plotted for the case that satellite 1 moves in a circular orbit at an altitude of 400 km about the Earth, and $x_0 = y_0 = 1$ km, $z_0 = 0$, $\dot{x}_0 = \dot{z}_0 = 0$, $\dot{y}_0 = -1.5 n x_0$.

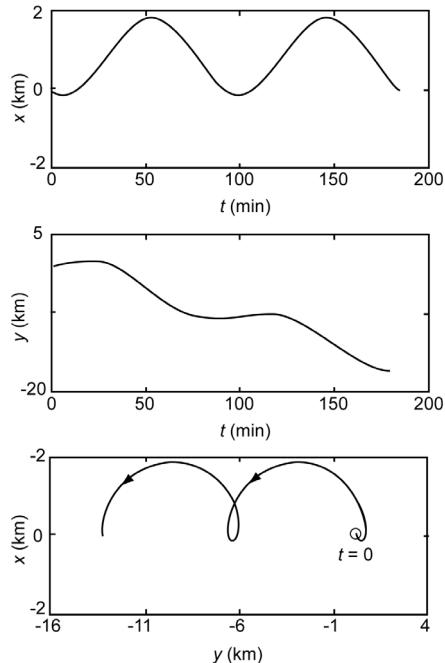


Figure 9.3: In-plane motion of satellite 2 relative to satellite 1 during two orbital revolutions of satellite 1. Assumptions: $h_1 = 400$ km, $x_0 = y_0 = z_0 = 0$, $\dot{x}_0 = -0.5$ m/s, $\dot{y}_0 = 0.5$ m/s, $\dot{z}_0 = 0$.

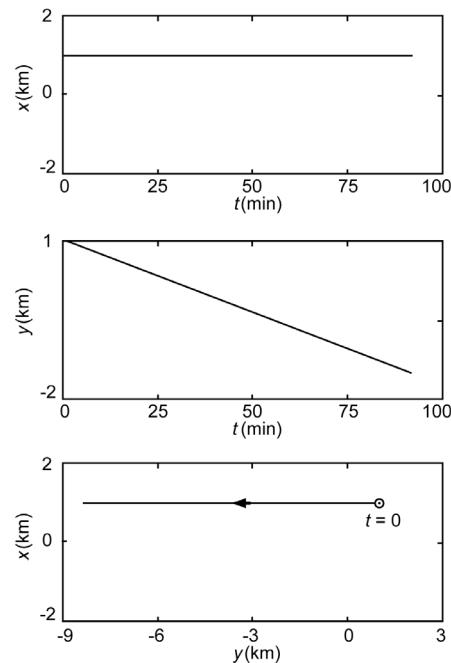


Figure 9.4: In-plane motion of satellite 2 relative to satellite 1 during one orbital revolution of satellite 1. Assumptions: $h_1 = 400$ km, $x_0 = y_0 = 1$ km, $z_0 = 0$, $\dot{x}_0 = \dot{z}_0 = 0$, $\dot{y}_0 = -1.5 n x_0$.

Three-dimensional circular relative orbits

According to (9.17), the elliptical relative orbit of satellite 2 in the XY -plane remains centered at satellite 1 ($x_C = y_C = 0$) if the following conditions hold:

$$\dot{x}_0 = \frac{1}{2}n y_0 \quad ; \quad \dot{y}_0 = -2n x_0 \quad (9.23)$$

which implies $C_1 = C_2 = 0$. For these conditions the ‘offset’ and drift terms in (9.14) vanish and (9.14) simplifies to

$$x = x_0 \cos nt + \frac{1}{2}y_0 \sin nt \quad (9.24)$$

$$\begin{aligned} y &= y_0 \cos nt - 2x_0 \sin nt \\ z &= z_0 \cos nt + \frac{\dot{z}_0}{n} \sin nt \end{aligned} \tag{9.24}$$

From (9.24), we find for the distance, d , between both satellites

$$\begin{aligned} d^2 &= \left(4x_0^2 + \frac{1}{4}y_0^2 + \left(\frac{\dot{z}_0}{n} \right)^2 \right) \sin^2 nt + (x_0^2 + y_0^2 + z_0^2) \cos^2 nt \\ &\quad - \left(3x_0 y_0 - 2z_0 \frac{\dot{z}_0}{n} \right) \sin nt \cos nt \end{aligned}$$

This distance is constant if the coefficient of $\sin^2 nt$ is equal to the coefficient of $\cos^2 nt$, and if the coefficient of $\sin nt \cos nt$ is zero. These requirements can be written as

$$\begin{aligned} z_0^2 - \left(\frac{\dot{z}_0}{n} \right)^2 &= 3(x_0^2 - \frac{1}{4}y_0^2) \\ 2z_0 \frac{\dot{z}_0}{n} &= 3x_0 y_0 \end{aligned}$$

Multiplication of the second relation by $i \equiv \sqrt{-1}$ and subsequent summation of the relations yields

$$\left(z_0 + i \frac{\dot{z}_0}{n} \right)^2 = 3 \left(x_0 + \frac{1}{2}i y_0 \right)^2$$

The solution of this equation is

$$z_0 = \pm \sqrt{3} x_0 \quad ; \quad \frac{\dot{z}_0}{n} = \pm \frac{1}{2} \sqrt{3} y_0 \tag{9.25}$$

Substitution of (9.25) into (9.24) gives for the relative motion:

$$\begin{aligned} x &= x_0 \cos nt + \frac{1}{2}y_0 \sin nt \\ y &= y_0 \cos nt - 2x_0 \sin nt \\ z &= \pm \frac{1}{2}\sqrt{3} y_0 \sin nt \pm \sqrt{3} x_0 \cos nt \end{aligned} \tag{9.26}$$

To analyze the component of the motion in the XY -plane, we multiply (9.26-1) by 2 and subsequently square the equation, square (9.26-2), and add both resulting relations. We then obtain

$$\frac{x^2}{\frac{1}{4}(4x_0^2 + y_0^2)} + \frac{y^2}{(4x_0^2 + y_0^2)} = 1 \tag{9.27}$$

This equation shows that in the XY -plane the motion is an ellipse, centered at satellite 1, with the

semi-major axis, a , along the Y -axis and the semi-minor axis, b , along the X -axis, with

$$a = 2b = \sqrt{4x_0^2 + y_0^2} \quad (9.28)$$

where the eccentricity of this ellipse is $e = \frac{1}{2}\sqrt{3}$. Combination of (9.26-1) and (9.26-3) gives

$$z = \pm\sqrt{3}x \quad (9.29)$$

From (9.27) and (9.29) we conclude that the three-dimensional circular motion of satellite 2 relative to satellite 1 takes place in a plane that is rotated about the Y -axis over an angle of 60° or 120° relative to the plane of motion of satellite 1. So, the elliptical relative motion of satellite 2 in the XY -plane is the result of the projection of the three-dimensional circular relative motion of satellite 2 onto the plane of motion of satellite 1.

Stationary coplanar elliptical relative orbits

If the initial conditions satisfy the relations:

$$\dot{y}_0 = -2n x_0 \quad ; \quad z_0 = 0 \quad ; \quad \dot{z}_0 = 0 \quad (9.30)$$

then $C_1 = 0$ and the motion takes place in the XY -plane with, according to (9.18) $V_{yC} = 0$. Consequently, the motion of satellite 2 relative to satellite 1 is an ellipse of which the center does not move relative to satellite 1. For the location of the center of the ellipse, we find from (9.17)

$$x_C = 0 \quad ; \quad y_C = y_0 - 2 \frac{\dot{x}_0}{n} \quad (9.31)$$

The center of this stationary ellipse is therefore always located along the orbit of satellite 1. For the semi-major axis and the semi-minor axis of this stationary ellipse, we find with (9.19), (9.20) and (9.30):

$$a = 2b = 2 \sqrt{x_0^2 + \left(\frac{\dot{x}_0}{n} \right)^2} \quad (9.32)$$

Substitution of (9.30) into (9.14) yields for the relative motion of satellite 2 in the XY -plane

$$x = x_0 \cos nt + \frac{\dot{x}_0}{n} \sin nt \quad (9.33)$$

$$y = y_0 - 2x_0 \sin nt + 2 \frac{\dot{x}_0}{n} (\cos nt - 1)$$

Differentiation of (9.33) yields for the velocity components

$$\begin{aligned} \dot{x} &= -n x_0 \sin nt + \dot{x}_0 \cos nt \\ \dot{y} &= -2n x_0 \cos nt - 2 \dot{x}_0 \sin nt \end{aligned} \quad (9.34)$$

The relations (9.33) show that the period of the relative motion of satellite 2 is equal to the period of the orbit of satellite 1. As an example, in Figure 9.5 the variation of x and y is shown for a characteristic stationary elliptical relative orbit. It is assumed that satellite 1 moves in a circular orbit at an altitude of 400 km about the Earth, and that $x_0 = y_0 = 1$ km, $z_0 = 0$; $\dot{x}_0 = 1$ m/s, $\dot{y}_0 = -2nx_0$, $\dot{z}_0 = 0$.

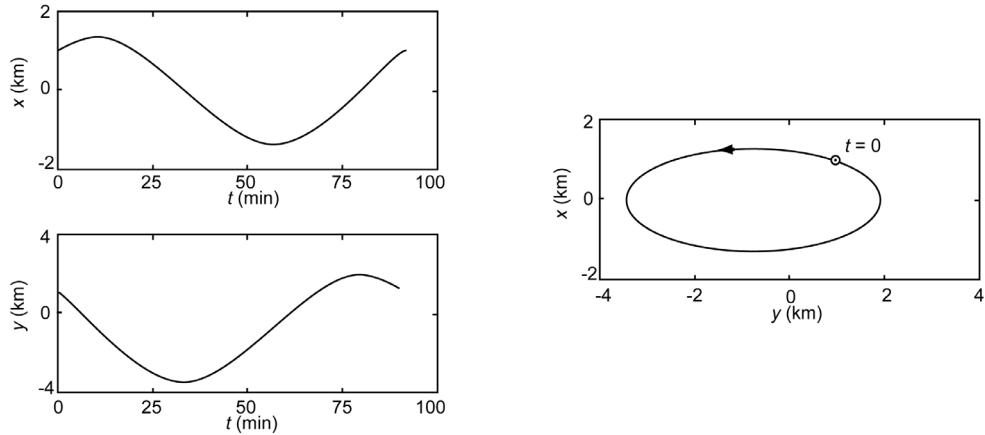


Figure 9.5: In-plane motion of satellite 2 relative to satellite 1 during one orbital revolution of satellite 1. Assumptions: $h_1 = 400 \text{ km}$, $x_0 = y_0 = 1 \text{ km}$, $z_0 = 0$, $\dot{x}_0 = 1 \text{ m/s}$, $\dot{y}_0 = -2nx_0$, $\dot{z}_0 = 0$.

When satellite 2 is initially at the position of satellite 1, then

$$x_0 = y_0 = 0 \quad (9.35)$$

In that case, we find from (9.30) to (9.32):

$$\dot{y}_0 = 0 \quad ; \quad x_C = 0 \quad ; \quad y_C = -2 \frac{\dot{x}_0}{n} \quad ; \quad a = 2b = 2 \frac{\dot{x}_0}{n} \quad (9.36)$$

and the equations of motion (9.33) become

$$x = \frac{\dot{x}_0}{n} \sin nt \quad ; \quad y = 2 \frac{\dot{x}_0}{n} (\cos nt - 1) \quad (9.37)$$

Obviously, satellite 2 approaches satellite 1 very closely again after each orbital revolution.

When we require that the coordinates of the center of the relative elliptical orbit are $x_C = y_C = 0$, which means that satellite 2 performs a stationary elliptical orbit centered at satellite 1, then the initial conditions (9.23) should be satisfied and the equations of relative motion in the XY-plane read according to (9.24):

$$\begin{aligned} x &= x_0 \cos nt + \frac{1}{2}y_0 \sin nt \\ y &= y_0 \cos nt - 2x_0 \sin nt \end{aligned} \quad (9.38)$$

These expressions are identical to the expressions derived for the motion in the X- and Y-direction in the case of three-dimensional circular relative orbits, which was discussed above. Consequently, the semi-major and semi-minor axis of this ellipse are given by (9.28).

9.4. Relative motion after an impulsive shot

The theory developed above can also be applied to analyze the changes in the orbit resulting from a short rocket thrusting period, i.e. an *impulsive shot* (Section 1.7). In this Section, we will analyze the motion of the satellite after an impulsive shot maneuver, relative to the orbit that the satellite would have followed if no impulsive shot would have been applied. So, in this case satellite 2 is the satellite that has experienced the impulsive shot and satellite 1 is a fictitious

satellite that follows the original orbit of satellite 2. Consequently, the state vector components of satellite 2 just after the impulsive shot are identical to the state vector components of satellite 1 at that time, except for the velocity component in the direction of the impulsive shot. In the following, three special cases will be analyzed.

Tangential impulsive shot

When the impulsive shot is applied in the direction of motion, i.e. tangential to the orbit (along track), we have

$$x_0 = y_0 = z_0 = \dot{x}_0 = \dot{z}_0 = 0 \quad ; \quad \dot{y}_0 = \Delta V \quad (9.39)$$

where ΔV is the magnitude of the impulsive shot. The equations of motion after the impulsive shot are found by substituting (9.39) into (9.14):

$$x = 2 \frac{\Delta V}{n} (1 - \cos nt) \quad ; \quad y = -\frac{\Delta V}{n} (3nt - 4 \sin nt) \quad ; \quad z = 0 \quad (9.40)$$

So, during the motion always $x \geq 0$. The extreme value of x is reached after half an orbital revolution ($nt = \pi$) and equals

$$x_{\text{extr}} = 4 \frac{\Delta V}{n}$$

From (9.40-2) we conclude that y exhibits a periodic variation superimposed on a linear variation with time (drift). In the first phase of the motion $y \geq 0$; after which a phase with $y < 0$ starts. In that second phase the satellite lags behind its corresponding position in the original orbit. After

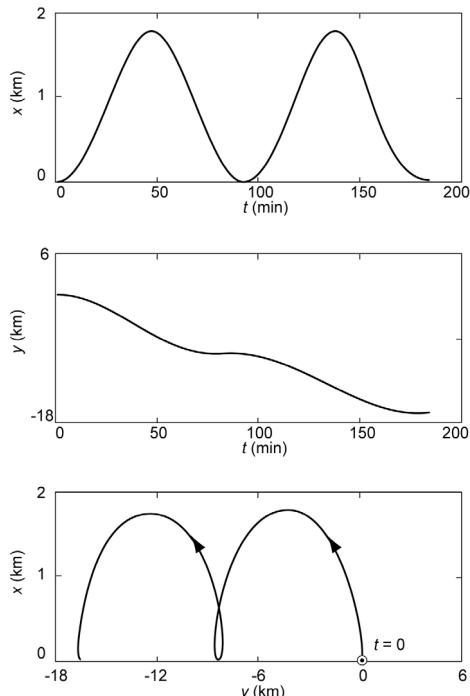


Figure 9.6: In-plane motion of a satellite after a tangential impulsive shot of $\Delta V = 0.5$ m/s relative to the fictitious position of that satellite in its original unperturbed circular orbit at an altitude of 400 km.

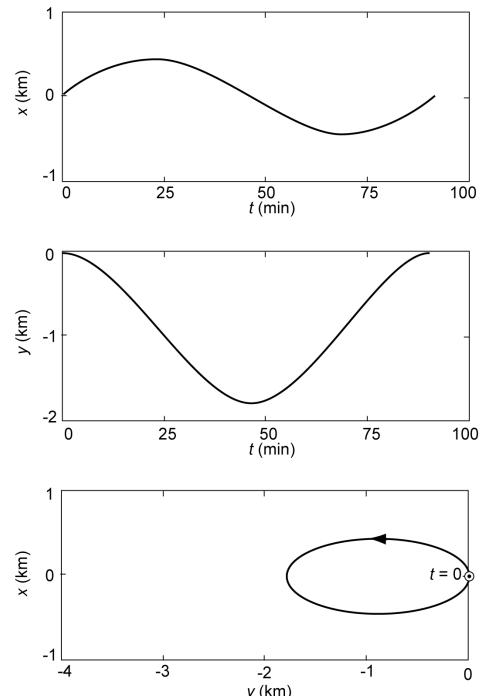


Figure 9.7: In-plane motion of a satellite after a radial impulsive shot of $\Delta V = 0.5$ m/s relative to the fictitious position of that satellite in its original unperturbed circular orbit at an altitude of 400 km.

the satellite would have completed a revolution in the nominal orbit ($nt = 2\pi$), its along-track position relative to the nominal along-track position is given by

$$y_{2\pi} = -6\pi \frac{\Delta V}{n}$$

When the pulse is applied in the opposite direction, we have

$$x_0 = y_0 = z_0 = \dot{x}_0 = \dot{z}_0 = 0 ; \quad \dot{y}_0 = -\Delta V$$

and the resulting orbit is a mirror image of the relative orbit discussed above. In Figure 9.6, the variation of x and y after a tangential impulse in the direction of motion is shown. It is assumed that the satellite originally moves in a circular orbit at an altitude of 400 km about the Earth and that the tangential impulsive shot has a magnitude of $\Delta V = 0.5$ m/s.

Radial impulsive shot

When the impulsive shot is directed radially outward, we have

$$x_0 = y_0 = z_0 = \dot{y}_0 = \dot{z}_0 = 0 ; \quad \dot{x}_0 = \Delta V \quad (9.41)$$

The equations of motion after the impulsive shot are found by substituting (9.41) into (9.14):

$$x = \frac{\Delta V}{n} \sin nt ; \quad y = -2 \frac{\Delta V}{n} (1 - \cos nt) ; \quad z = 0 \quad (9.42)$$

Note that the motion in X - and Y -direction is purely periodic, which means that an impulse in radial direction results in a closed relative orbit and does not introduce any drift effect. The extreme values of x and y relative to the satellite's position in the (nominal) unperturbed orbit are

$$x_{extr} = \pm \frac{\Delta V}{n} ; \quad y_{extr} = -4 \frac{\Delta V}{n}$$

When the radial pulse is directed inward, we have

$$x_0 = y_0 = z_0 = \dot{y}_0 = \dot{z}_0 = 0 ; \quad \dot{x}_0 = -\Delta V$$

and the resulting orbit is the mirror image of the relative orbit discussed above. In Figure 9.7, the variation of x and y after a outward radial impulsive shot is shown. It is assumed that the satellite originally moves in a circular orbit at an altitude of 400 km about the Earth and that the radial impulsive shot has a magnitude of $\Delta V = 0.5$ m/s.

Normal impulsive shot

When an impulsive shot is applied perpendicular to the orbital plane (cross track) in the $+Z$ -direction, we have

$$x_0 = y_0 = z_0 = \dot{x}_0 = \dot{y}_0 = 0 ; \quad \dot{z}_0 = \Delta V \quad (9.43)$$

The equations of motion after the impulsive shot are found by substituting (9.43) into (9.14):

$$x = 0 ; \quad y = 0 ; \quad z = \frac{\Delta V}{n} \sin nt$$

Note that there is no relative motion in the X - and Y -direction and that the relative motion in the Z -direction is purely periodic. The extreme values of z relative to the satellite's position in the unperturbed orbit are

$$z_{extr} = \pm \frac{\Delta V}{n}$$

When the cross-track impulsive shot is applied in the opposite direction, we find

$$x_0 = y_0 = z_0 = \dot{x}_0 = \dot{y}_0 = 0 \quad ; \quad \dot{z}_0 = -\Delta V$$

and the resulting orbit is the mirror image of the relative orbit discussed above.

