

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.

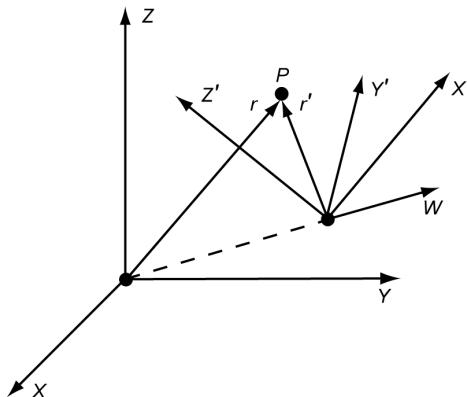


Figure 1.1: Inertial reference frame XYZ and a reference frame $X'Y'Z'$ that moves with constant velocity relative to the XYZ frame.

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

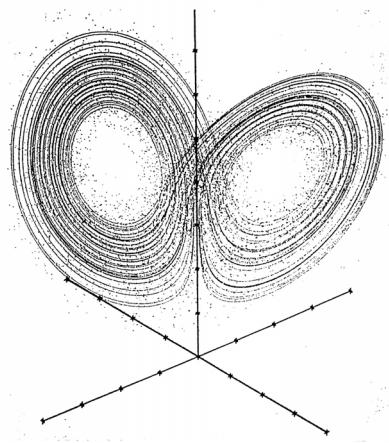


Figure 1.2: Lorenz strange attractor.

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

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

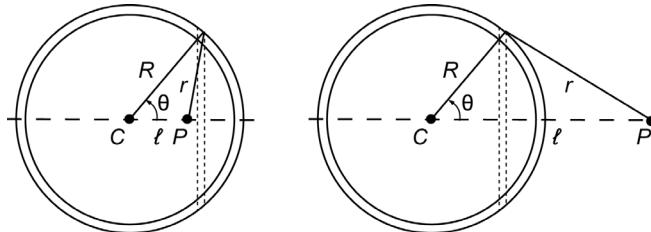


Figure 1.3: Geometry for the computation of the internal (left) and external (right) gravity field of a thin spherical shell.

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 .

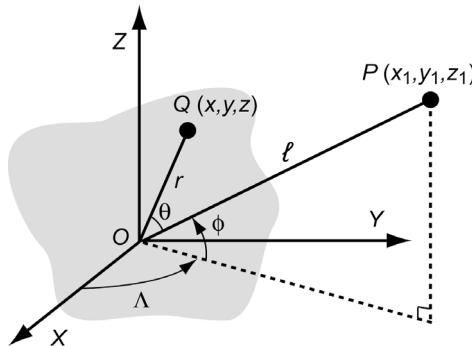


Figure 1.4: Arbitrary body L and an external point P . The origin of reference frame XYZ coincides with the center of mass of L .

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.

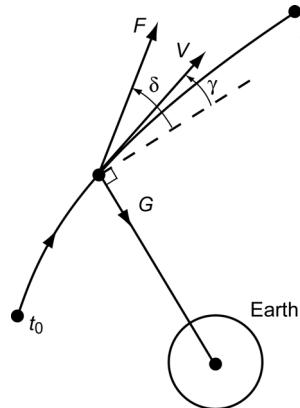


Figure 1.5: Geometry of powered flight and a definition of the flight path angle, γ , and the thrust angle, δ .

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

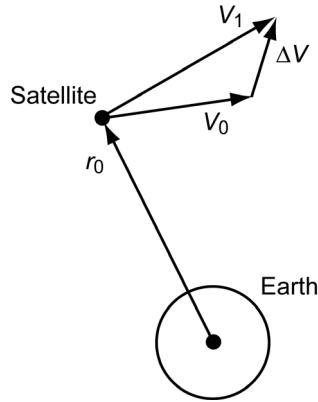


Figure 1.6: Geometry of an impulsive shot.

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

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

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

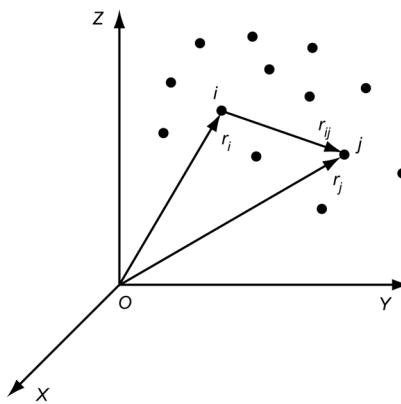


Figure 2.1: Position of n point masses relative to an inertial reference frame XYZ.

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

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

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:

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

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,

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

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

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

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:

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

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

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

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:

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

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

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

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.

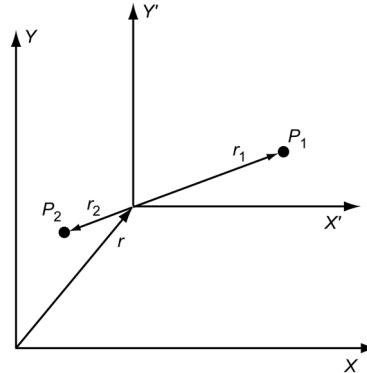


Figure 2.2: A pseudo-inertial reference frame and the positions of the bodies of the two-body problem relative to this reference frame.

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).

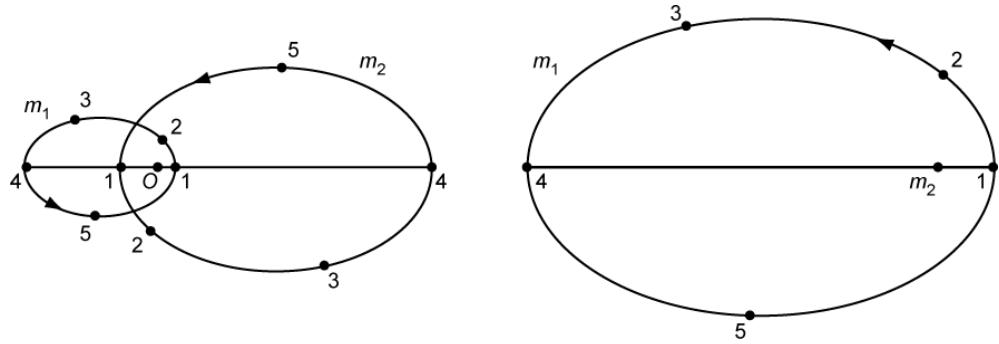


Figure 2.3: Absolute (left) and relative (right) motion of two bodies under the influence of their mutual gravitational attraction ($m_1/m_2 = 2$).

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.

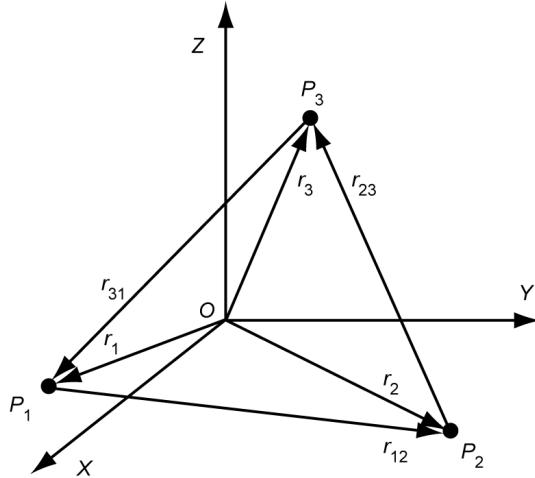


Figure 3.1: Geometry of the system of three bodies and the vectors used in the Euler and Lagrange formulation.

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

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

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.

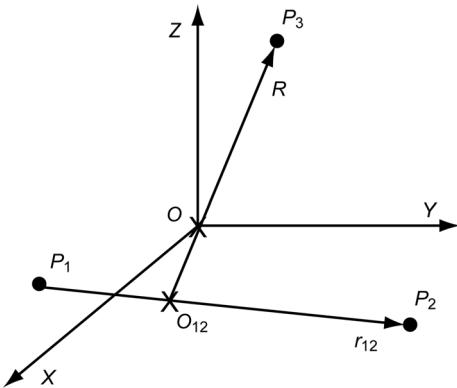


Figure 3.2: Jacobi method of describing the position of the three bodies.

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

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

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

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

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

$$r_i = (r_i)_0 f(t)$$

or, generalizing this result,

$$r_i = (r_i)_0 f(t) \quad (3.15)$$

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

$$\mathbf{m}_1 r_1^2 \omega_1 + \mathbf{m}_2 r_2^2 \omega_2 + \mathbf{m}_3 r_3^2 \omega_3 = H \quad (3.17)$$

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

$$\left[\mathbf{m}_1 (r_1)_0^2 + \mathbf{m}_2 (r_2)_0^2 + \mathbf{m}_3 (r_3)_0^2 \right] \{f(t)\}^2 \omega(t) = H \quad (3.18)$$

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

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

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.

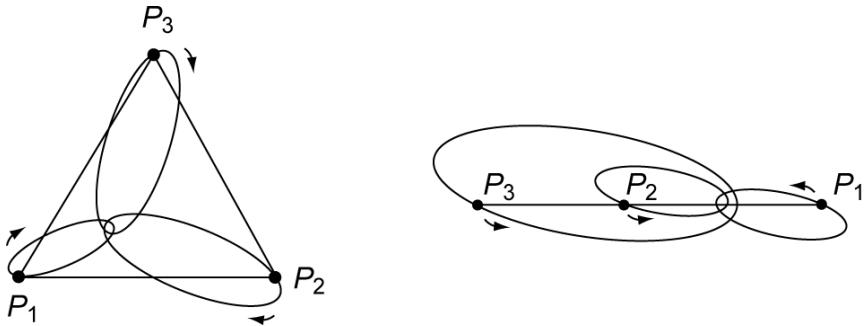


Figure 3.3: Elliptical Lagrange case (left) and elliptical Euler case (right) of central configuration solutions.

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 :

$$\mathbf{g}_1 = G \frac{\mathbf{m}_2}{\mathbf{r}_{12}^2} + G \frac{\mathbf{m}_3}{\mathbf{r}_{13}^2}$$

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

$$\mathbf{g}_1 = \frac{G}{\{f(t)\}^2} \left[\frac{\mathbf{m}_2}{(\mathbf{r}_{12})_0^2} + \frac{\mathbf{m}_3}{(\mathbf{r}_{13})_0^2} \right] = \frac{\mathbf{c}_1}{\{f(t)\}^2} = \frac{\mathbf{c}_2}{\mathbf{r}_1^2} \quad (3.27)$$

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.

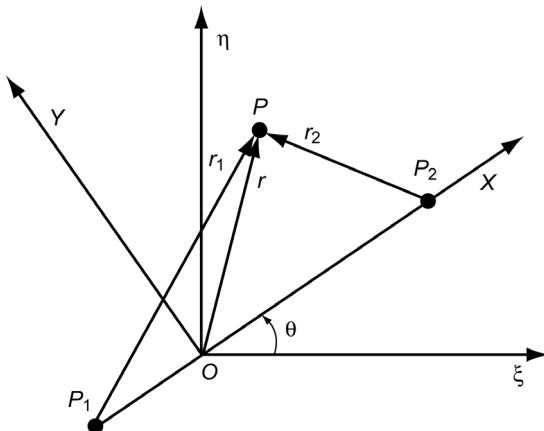


Figure 3.4: Inertial and rotating reference frames in the circular restricted three-body problem.

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 .

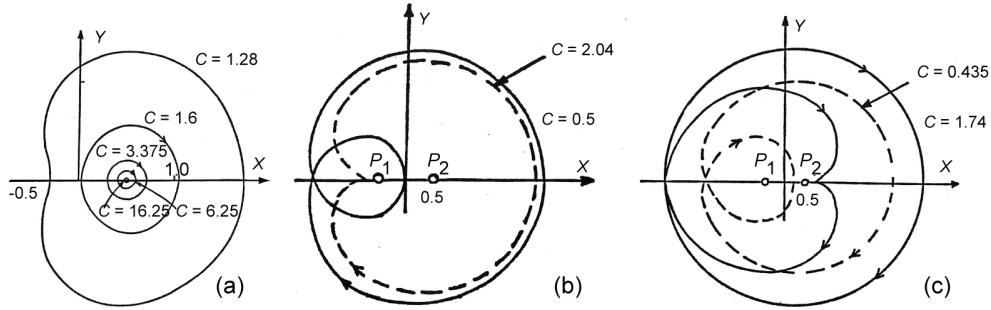


Figure 3.5: Evolution of orbits about P_2 for the Class F family of the Copenhagen problem at decreasing values of C .

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.

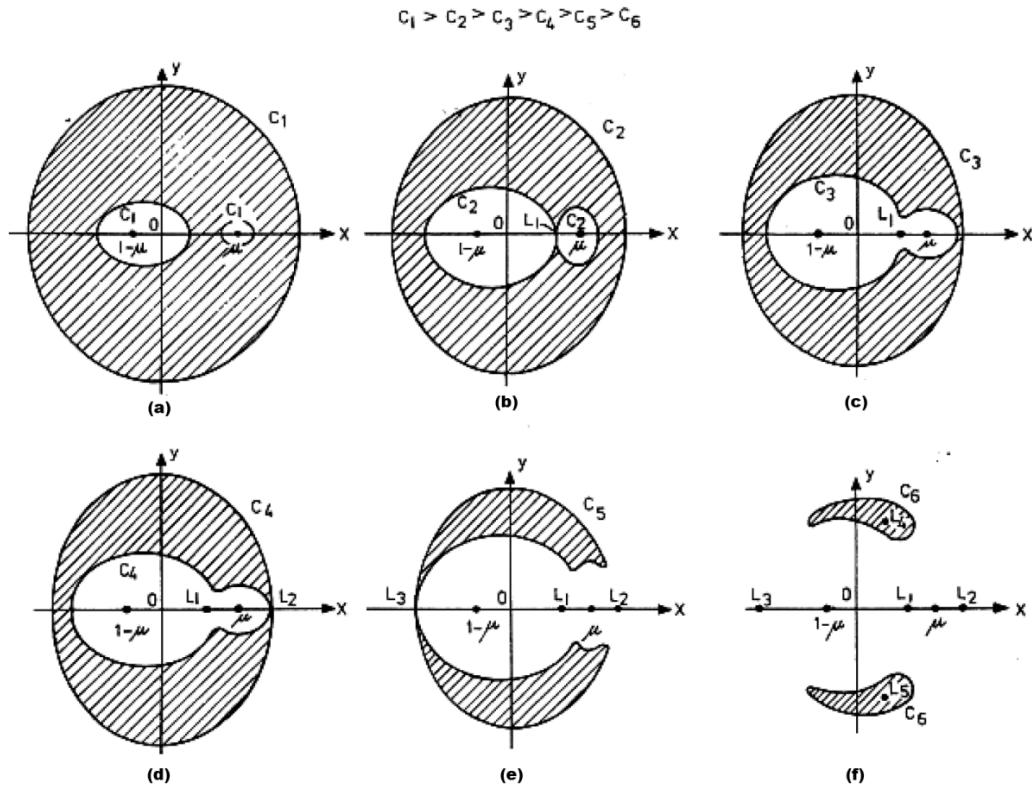


Figure 3.6: Schematic picture of the surfaces of Hill for decreasing values of C ($C_{n+1} < C_n$).

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

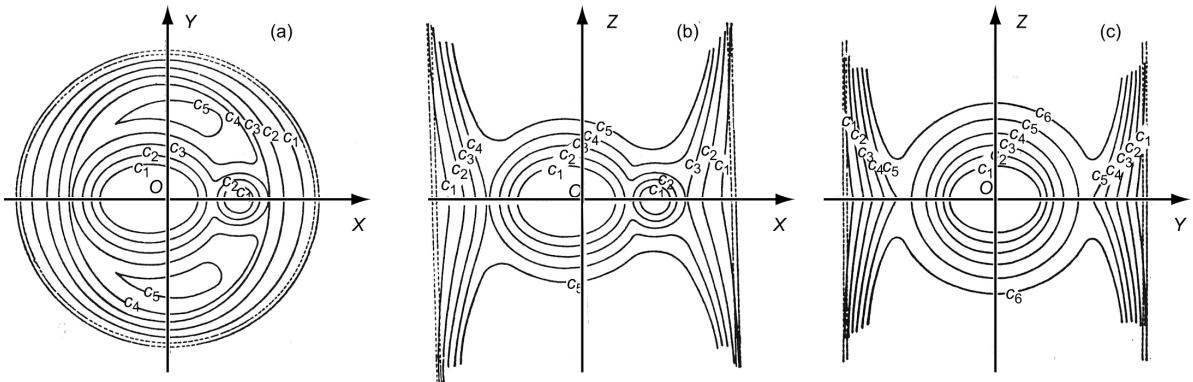


Figure 3.7: Cross sections of the surfaces of Hill with the XY-, XZ- and YZ-plane ($\mu = 0.27$).

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.

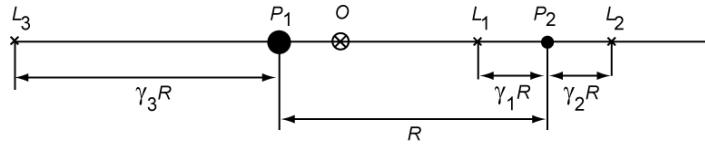
Substitution of (3.62) into (3.49) gives

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

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.

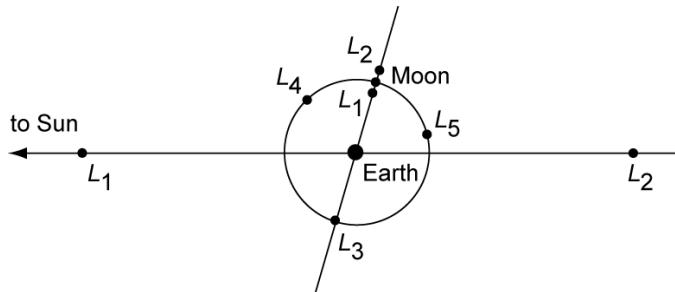


Figure 3.9: Positions of all Earth-Moon L points and of two Sun-Earth L points (to scale).

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?"

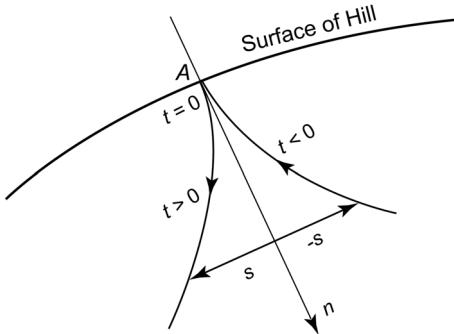


Figure 3.10: Cusp at a curve of zero velocity.

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

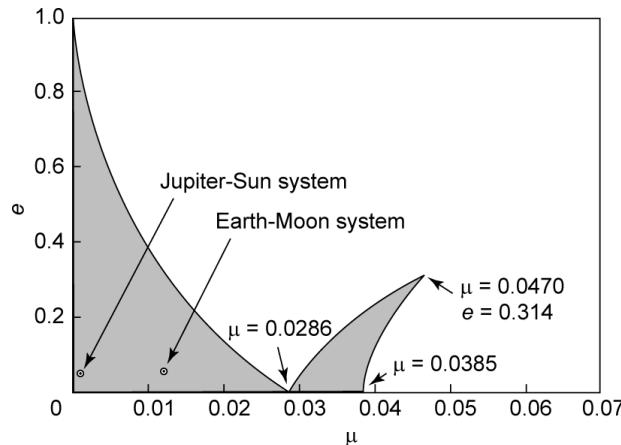


Figure 3.11: Regions of first-order stability in the L_4 and L_5 points for the case that the two main bodies move in elliptical orbits.

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

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.

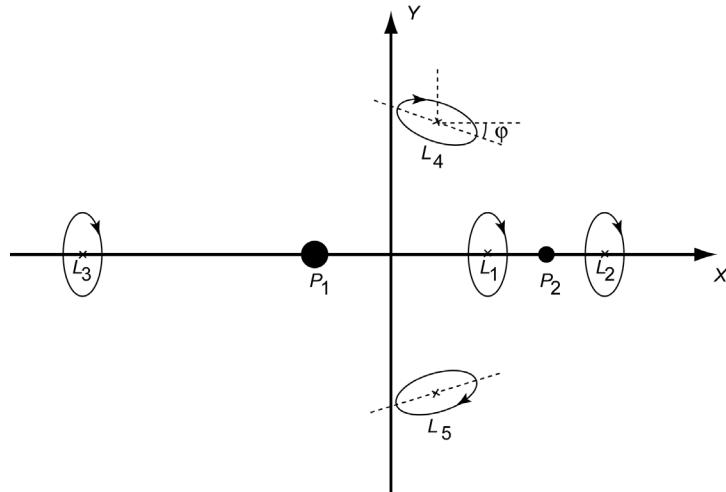


Figure 3.12: Periodic $x'y'$ -motion about the libration points (not to scale).

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.

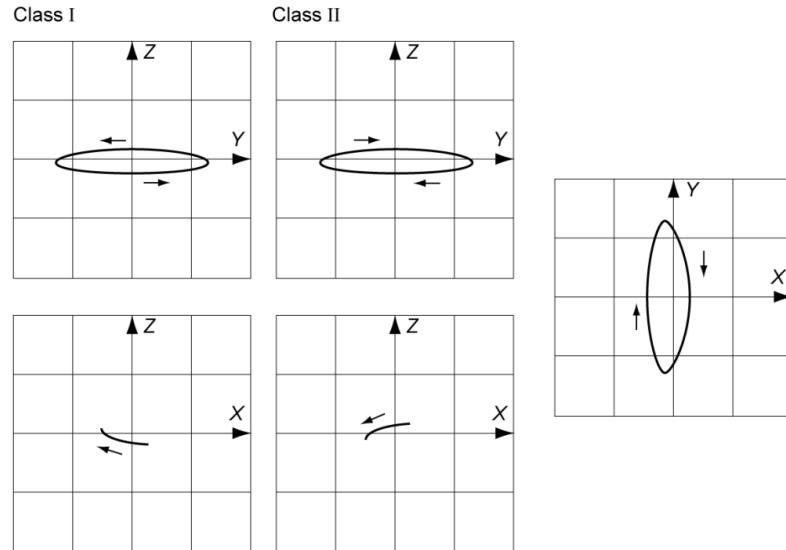


Figure 3.13: Two classes of L_2 halo orbits for the Sun-Earth(+Moon) system with x -, y -, and z -amplitudes of 215,000 km, 686,000 km, and 125,000 km, respectively.

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.

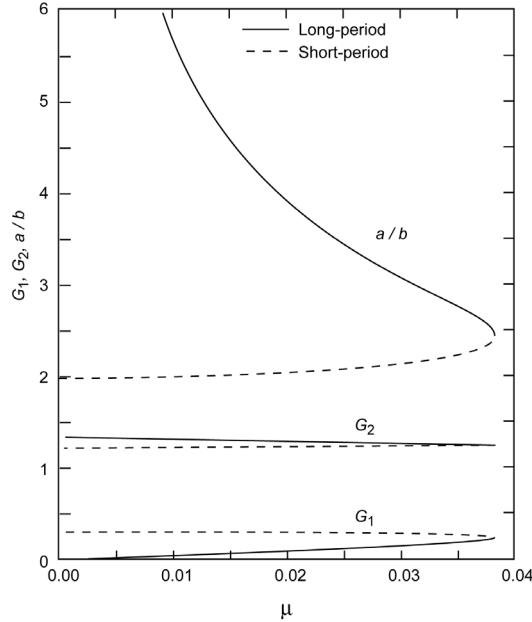


Figure 3.14: Values of G_1 , G_2 and a/b as a function of μ , both for the long-period and the short-period components of the motion about the L_4 and L_5 points.

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

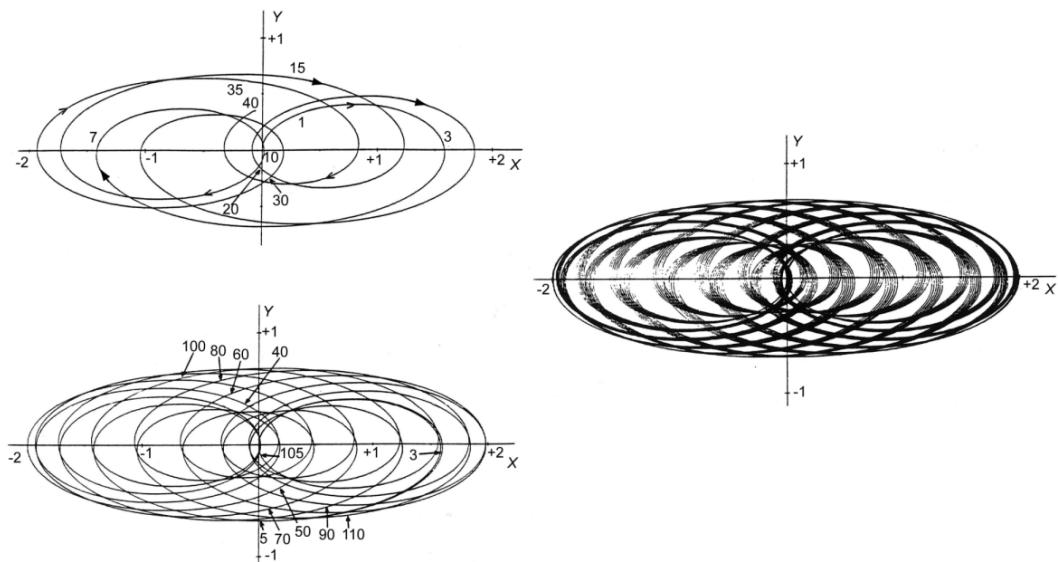


Figure 3.15: Linearized solution for the trajectory about an Earth-Moon triangular libration point for 0.5 year (top, left), one recurrence (≈ 1.3 year, bottom, left) and five recurrences (≈ 6.3 year, right). Numbers on curves denote the elapsed dimensionless time, t^* .

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

$$V_{rot}^2 = n_M^2(x^2 + y^2) + 2Gm_E\left(\frac{1}{r_1} + \frac{m_M}{m_E r_2}\right) - n_M^2 r_M^2 C \quad (3.160)$$

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.

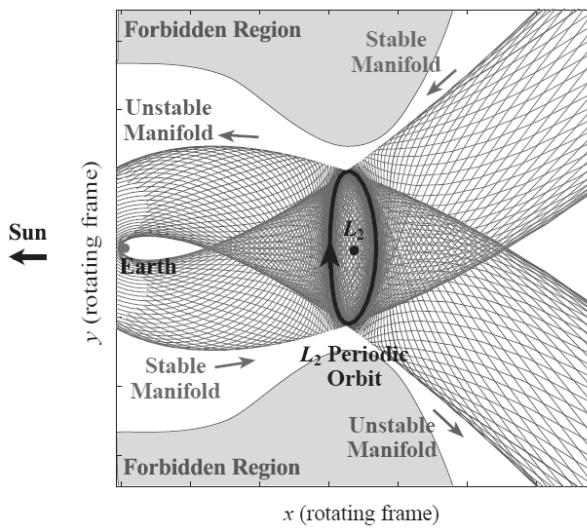


Figure 3.16: A periodic Lyapunov orbit about the Sun-Earth L_2 point and its associated stable and unstable invariant manifolds in position space. The two manifolds ‘twist’ as they wrap around the region near the Earth. [copied from: W.S. Koon et al., *Dynamical Systems and Space Mission Design*, presentation 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

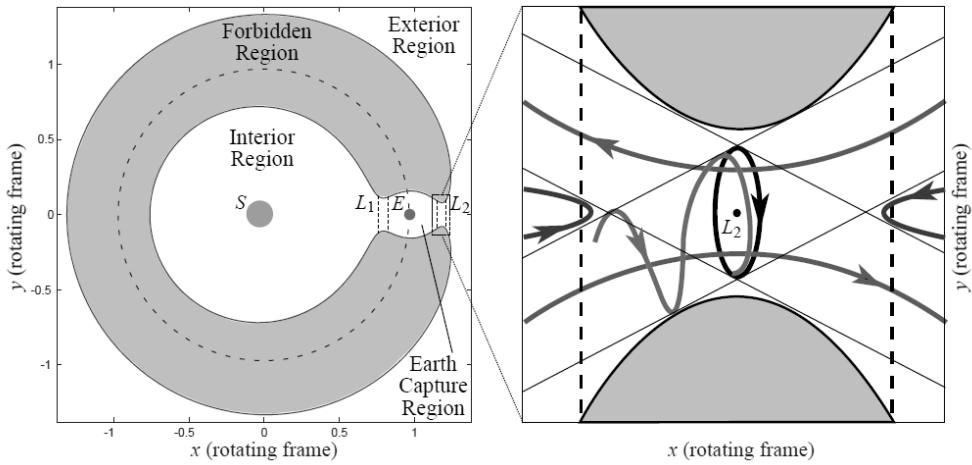


Figure 3.17: (left) Projection of the three-dimensional surfaces of Hill (schematic) on the XY-plane for the Sun-Earth system and for a value of C for which 'necks' exist about the L_1 and L_2 points. (right) Motion in the region near L_2 projected on the XY-plane, showing a bounded orbit about L_2 , 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

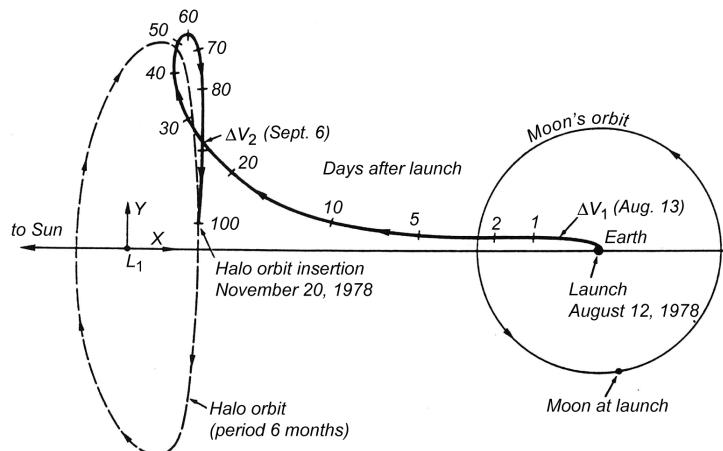


Figure 3.18: Transfer trajectory of the ISEE-3 spacecraft to its halo orbit.

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

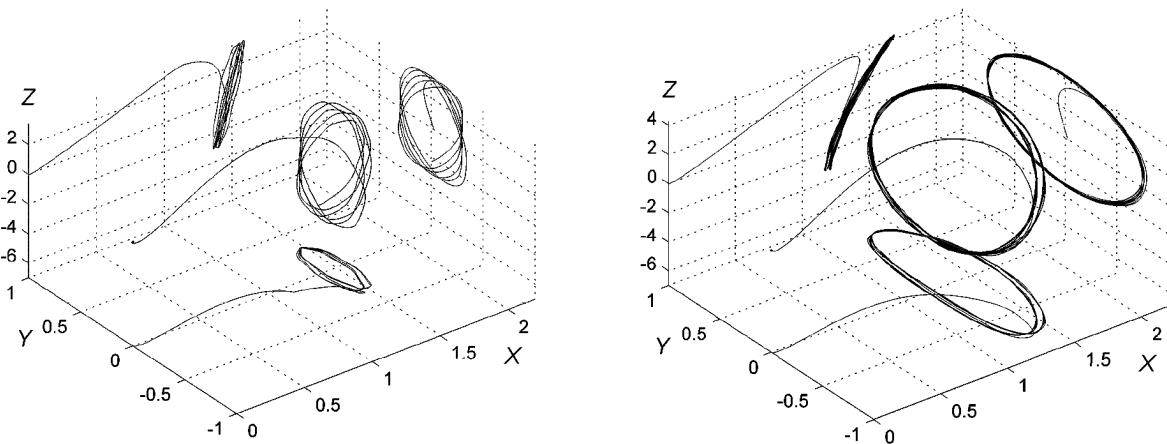


Figure 3.19: Three-dimensional picture of the transfer trajectory to and Lissajous orbit about the Sun-Earth L_2 point of the Planck spacecraft (left) and of the transfer trajectory to and quasi-halo orbit of the Herschel spacecraft about this L point (right). The unit of scale in the X- and Y-direction 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.

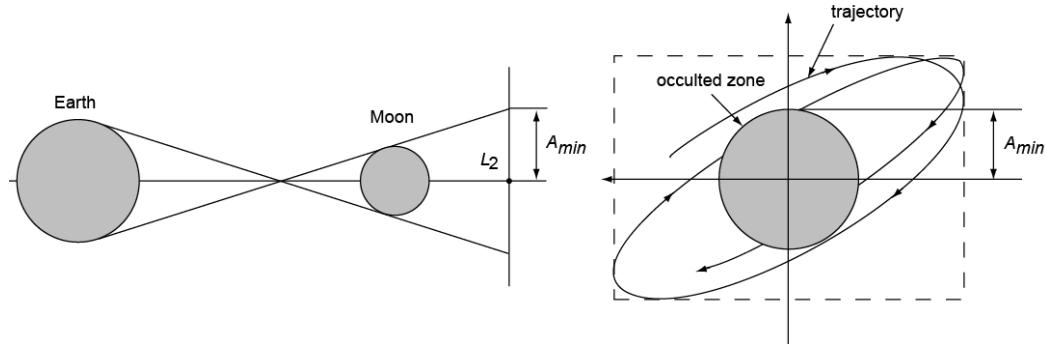


Figure 3.20: An uncontrolled trajectory about the L_2 point in the Earth-Moon system. The Figure shows the occultation geometry (left, not to scale), and the trajectory as viewed from the Earth (right).

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):

$$\begin{aligned} m_i \frac{d^2 \bar{r}_i}{dt^2} &= \sum_{j \neq i} G \frac{\mathbf{m}_i \mathbf{m}_j}{\bar{r}_{ij}^3} \bar{r}_{ij} \\ m_k \frac{d^2 \bar{r}_k}{dt^2} &= \sum_{j \neq k} G \frac{\mathbf{m}_k \mathbf{m}_j}{\bar{r}_{kj}^3} \bar{r}_{kj} \end{aligned} \quad (4.1)$$

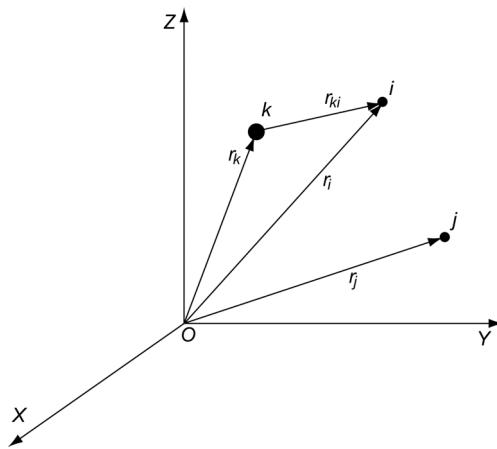


Figure 4.1: Geometry of the bodies i , j and k relative to an inertial reference frame XYZ .

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):

$$U_i = -G \frac{\mathbf{m}_k}{\mathbf{r}_i}$$

Now, a slightly modified function, defined as

$$\hat{U}_i = -G \frac{\mathbf{m}_k + \mathbf{m}_i}{\mathbf{r}_i} \quad (4.5)$$

and a scalar function, defined as

$$R_i = -G \sum_{j \neq i, k} \mathbf{m}_j \left(\frac{1}{\mathbf{r}_{ij}} - \frac{\bar{\mathbf{r}}_i \cdot \bar{\mathbf{r}}_j}{\mathbf{r}_j^3} \right) = -G \sum_{j \neq i, k} \mathbf{m}_j \left(\frac{1}{\mathbf{r}_{ij}} - \frac{x_i x_j + y_i y_j + z_i z_j}{\mathbf{r}_j^3} \right) \quad (4.6)$$

are introduced. With (4.5) and (4.6), and the expressions

$$\mathbf{r}_i = (x_i^2 + y_i^2 + z_i^2)^{1/2} \quad ; \quad \mathbf{r}_{ij} = \{(x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2\}^{1/2}$$

we can write for the partial derivatives of \hat{U}_i and R_i to x_i :

$$\frac{\partial \hat{U}_i}{\partial \mathbf{x}_i} = G(\mathbf{m}_k + \mathbf{m}_i) \frac{\mathbf{x}_i}{\mathbf{r}_i^3} \quad ; \quad \frac{\partial R_i}{\partial \mathbf{x}_i} = -G \sum_{j \neq i, k} \mathbf{m}_j \left(\frac{\mathbf{x}_j - \mathbf{x}_i}{\mathbf{r}_{ij}^3} - \frac{\mathbf{x}_j}{\mathbf{r}_j^3} \right)$$

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

$$\frac{d^2 \bar{\mathbf{r}}_i}{dt^2} = -\bar{\nabla}_i (\hat{U}_i + R_i) \quad (4.7)$$

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.

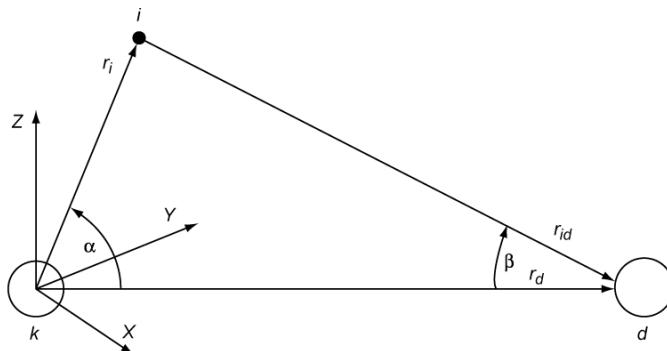


Figure 4.2: Relative positions of bodies k , i , and d .

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

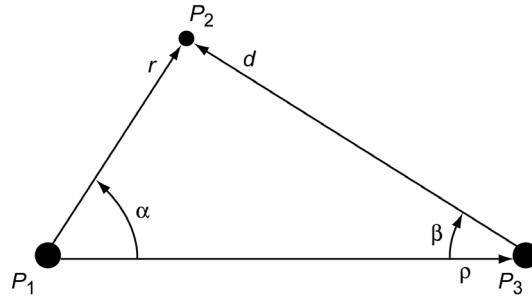


Figure 4.3: Relative positions of three bodies and the notation used in the analysis of the sphere of influence concept.

$$\frac{d^2 \bar{d}}{dt^2} + \frac{G(m_3 + m_2)}{d^3} \bar{d} = -G m_1 \left(\frac{\bar{r}}{r^3} - \frac{\bar{\rho}}{\rho^3} \right) \quad (4.19)$$

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

$$a_d = G m_3 \left[\left(\frac{\bar{d}}{d^3} + \frac{\bar{\rho}}{\rho^3} \right) \cdot \left(\frac{\bar{d}}{d^3} + \frac{\bar{\rho}}{\rho^3} \right) \right]^{1/2} = G m_3 \left[\frac{1}{d^4} + \frac{1}{\rho^4} - \frac{2 \cos \beta}{d^2 \rho^2} \right]^{1/2}$$

Consequently, we find for the ratio a_d/a_m :

$$\frac{a_d}{a_m} = \frac{m_3}{m_1 + m_2} \frac{r^2}{d^2} \left[1 - 2 \frac{d^2}{\rho^2} \cos \beta + \frac{d^4}{\rho^4} \right]^{1/2} \quad (4.20)$$

According to Figure 4.3, the following relation holds:

$$\cos \beta = \frac{\rho - r \cos \alpha}{d} = \frac{\rho}{d} \left(1 - \frac{r}{\rho} \cos \alpha \right)$$

Substitution of this expression into (4.20) yields

$$\frac{a_d}{a_m} = \frac{m_3}{m_1 + m_2} \left(\frac{r/\rho}{d/\rho} \right)^2 \left[1 - 2 \frac{d}{\rho} \left(1 - \frac{r}{\rho} \cos \alpha \right) + \left(\frac{d}{\rho} \right)^4 \right]^{1/2} \quad (4.21)$$

where, according to Figure 4.3,

$$\frac{d}{\rho} = \left[1 - 2 \frac{r}{\rho} \cos \alpha + \left(\frac{r}{\rho} \right)^2 \right]^{1/2} \quad (4.22)$$

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

$$\begin{aligned} \left(\frac{d}{\rho} \right)^4 &= 1 - 4 \frac{r}{\rho} \cos \alpha + 2 \left(\frac{r}{\rho} \right)^2 (1 + 2 \cos^2 \alpha) - 4 \left(\frac{r}{\rho} \right)^3 \cos \alpha + \left(\frac{r}{\rho} \right)^4 \\ \left[1 - 2 \left(\frac{r}{\rho} \right)^2 \cos \alpha + \left(\frac{r}{\rho} \right)^4 \right]^{1/2} &= 1 - \left(\frac{r}{\rho} \right)^2 \cos \alpha + \frac{1}{2} \left(\frac{r}{\rho} \right)^4 (1 - \cos^2 \alpha) + O \left(\frac{r}{\rho} \right)^6 \\ \left[1 - 2 \frac{d}{\rho} \left(1 - \frac{r}{\rho} \cos \alpha \right) + \left(\frac{d}{\rho} \right)^4 \right]^{-1/2} &= \left(\frac{r}{\rho} \right)^{-1} (1 + 3 \cos^2 \alpha)^{-1/2} . \\ \left[1 + 2 \left(\frac{r}{\rho} \right) \frac{\cos \alpha}{1 + 3 \cos^2 \alpha} - \frac{1}{8} \left(\frac{r}{\rho} \right)^2 \left(\frac{5 - 35 \cos^2 \alpha - 5 \cos^4 \alpha + 3 \cos^6 \alpha}{(1 + 3 \cos^2 \alpha)^2} \right) + O \left(\frac{r}{\rho} \right)^3 \right] \end{aligned}$$

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

$$\bar{r}_{Bk} = -\frac{\mathbf{m}_i}{\mathbf{m}_k} \bar{r}_{Bi} \quad ; \quad \bar{r}_{ik} = -\left(1 + \frac{\mathbf{m}_i}{\mathbf{m}_k}\right) \bar{r}_{Bi} \quad ; \quad \bar{r}_{ik}^3 = \left(1 + \frac{\mathbf{m}_i}{\mathbf{m}_k}\right)^3 \bar{r}_{Bi}^3$$

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 Bi 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.

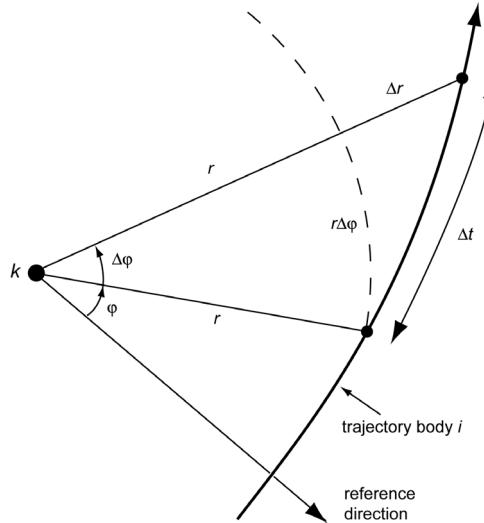


Figure 5.1: Motion of body i about body k .

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

$$\mathbf{r} = \frac{H^2/\mu}{1 + c_3 \cos(\varphi - \omega)} \quad (5.16-1)$$

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

$$H^2 = \bar{H} \cdot \bar{H} = \left(\bar{r} \times \frac{d\bar{r}}{dt} \right) \cdot \bar{H} = \left(\frac{d\bar{r}}{dt} \times \bar{H} \right) \cdot \bar{r}$$

finally gives

$$H^2 = \mu (r + \bar{r} \cdot \bar{c}_4) \quad (5.18)$$

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):

$$H^2 = \mu [r + r c_4 \cos(\varphi - \omega)]$$

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.

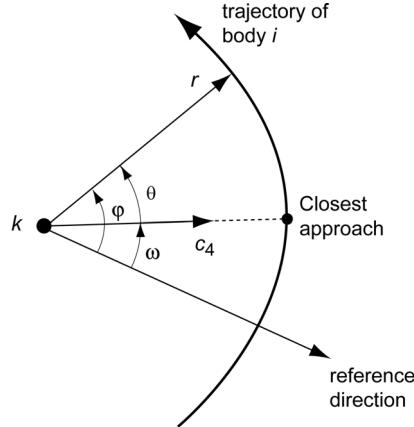


Figure 5.2: Direction of the vector \bar{c}_4 and the definition of the angles φ , ω and θ .

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:

$$P \text{ to the left of } l: \quad \mathbf{r} = e(\mathbf{k} - \mathbf{r} \cos \theta)$$

$$P \text{ to the right of } l: \quad \mathbf{r} = e(\mathbf{r} \cos \theta - \mathbf{k})$$

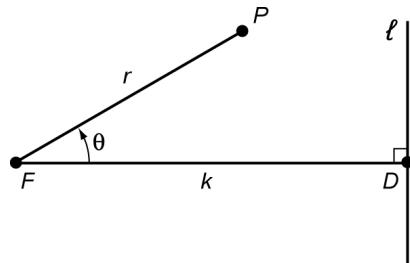


Figure 5.3: Geometric definition of a conic section.

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) Bernouilli (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) Bernouilli (1654-1705). Newton presented a similar analysis in the second edition of his *Principia* in 1713. J. (Johann) Bernouilli 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

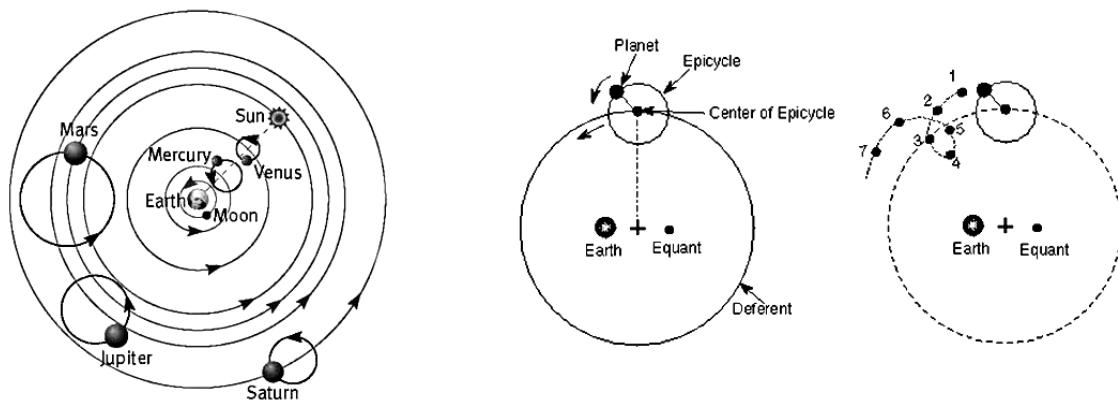


Figure 5.4: (left) Ptolemaic model for the motion of celestial bodies; Sun, Moon and planets move about the Earth. (center) A planet moves along an epicycle, while the center of this epicycle moves along the deferent. Epicycle speed is uniform with respect to the equant. (right) Deferent motion is in 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

$$\dot{r} = V \sin \gamma \quad (5.23)$$

$$r \dot{\theta} = V \cos \gamma \quad (5.24)$$

$$H = r V \cos \gamma \quad (5.25)$$

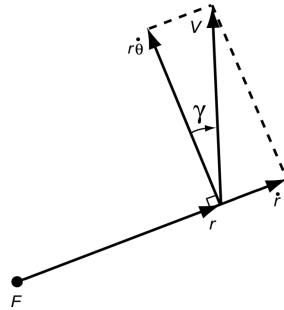


Figure 5.5: The flight path angle, γ , and the radial and normal velocity components.

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):

$$r \dot{\theta} = \frac{H}{r} = \frac{\mu}{H} (1 + e \cos \theta) \quad (5.27)$$

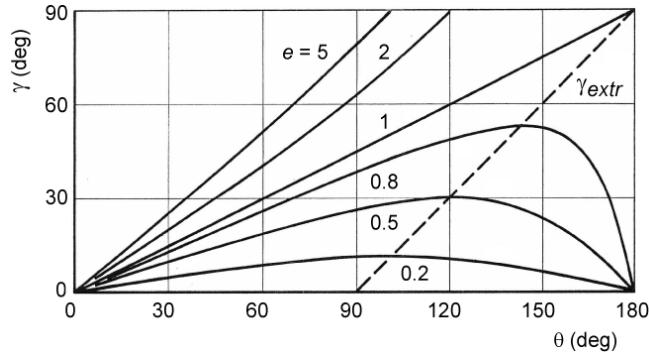
From (5.23), (5.24), (5.26) and (5.27) we obtain for the flight path angle

$$\tan \gamma = \frac{\dot{r}}{r \dot{\theta}} = \frac{e \sin \theta}{1 + e \cos \theta} \quad (5.28)$$

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 ($\theta_{\gamma,extr}$):

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

$$\dot{r}^2 + \left(r \dot{\theta} - \frac{\mu}{H} \right)^2 = \left(\frac{\mu e}{H} \right)^2 \quad (5.29)$$

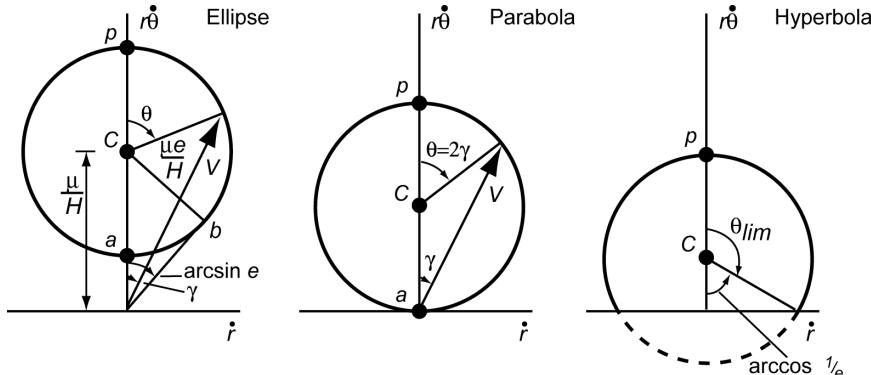


Figure 5.7: Velocity hodographs for elliptical, parabolic and hyperbolic motion.

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:

$$V_l = \frac{\dot{r}}{\sin(\pi - \theta)} = \frac{\dot{r}}{\sin \theta} ; \quad V_n = r \dot{\theta} + \frac{\dot{r}}{\tan(\pi - \theta)} = r \dot{\theta} - \frac{\dot{r}}{\tan \theta}$$

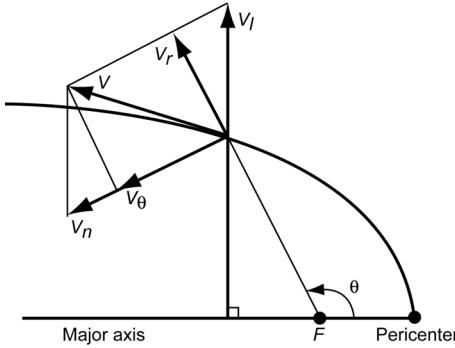


Figure 5.8: Definition of the velocity components V_l and V_n .

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).

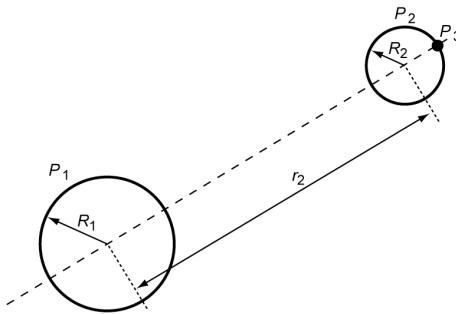


Figure 5.9: Definition of the parameters used in the analysis of the Roche limit.

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

$$\cos \beta \varphi \approx 1 ; \sin \beta \varphi \approx \beta \varphi$$

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).

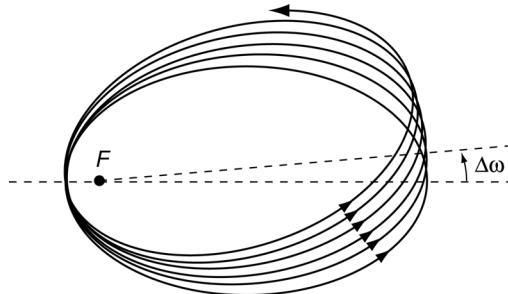


Figure 5.10: Relativistic precession of the orbit's major axis.

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.

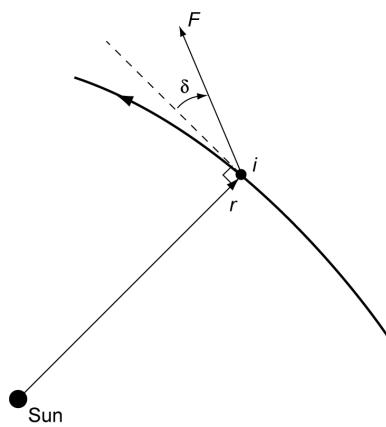
in orbit mechanics, the relativistic effect is very small and we can therefore certainly use Newton's law of gravitation and his laws of motion to compute the trajectories of these bodies.

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):

$$\begin{aligned} \ddot{\vec{r}} - r \dot{\phi}^2 \vec{r} &= -\frac{\mu}{r^2} + \frac{F}{m} \sin \delta \\ \frac{d}{dt}(r^2 \dot{\phi}) &= \frac{F}{m} r \cos \delta \end{aligned} \quad (5.58)$$

where on the right-hand side of the equations the two components of the radiation force, \vec{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 \vec{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

$$W' = W \left(1 - \frac{\dot{r}}{c} \right) \quad (5.70)$$

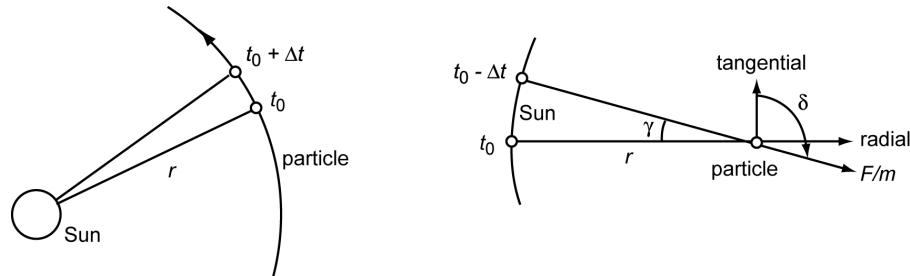


Figure 5.12: Motion of a particle relative to a heliocentric reference frame during the time interval Δt (left) and a visualization of the aberration effect (right).

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

$$H = H_0 - \frac{\alpha \Phi}{c} \quad (5.74)$$

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

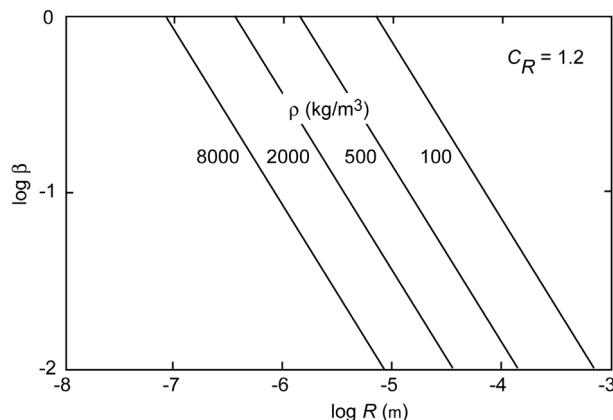


Figure 5.13: Relative solar radiation pressure as a function of the radius and the mean mass density of a particle.

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 $2b$.

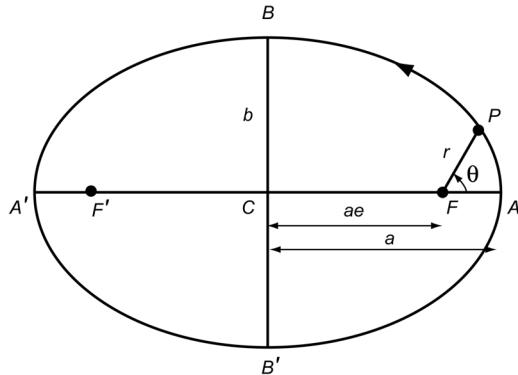


Figure 6.1: Geometry of an elliptical orbit.

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

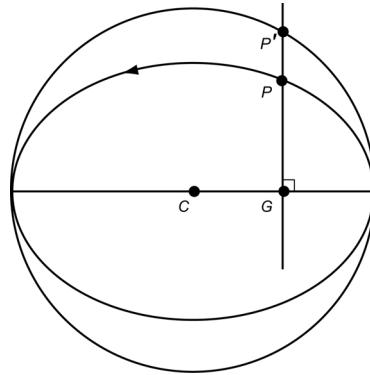


Figure 6.2: Projection of a point P on an ellipse onto the circumferential circle.

$$a = \frac{p}{1 - e^2} = \frac{r^2 V^2 \cos^2 \gamma}{r V^2 (2 - \frac{r V^2}{\mu}) \cos^2 \gamma} = \frac{\mu/2}{\frac{\mu}{r} - \frac{V^2}{2}} \quad (6.11)$$

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.

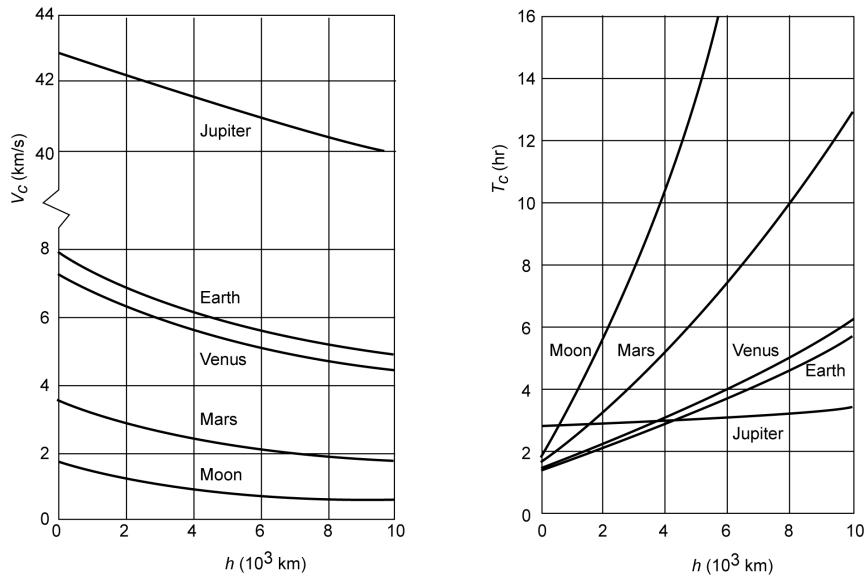


Figure 6.3: Circular velocity as a function of altitude for a number of celestial bodies (left) and the period of a circular orbit as a function of altitude for the same celestial bodies (right).

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 $+15^\circ$.

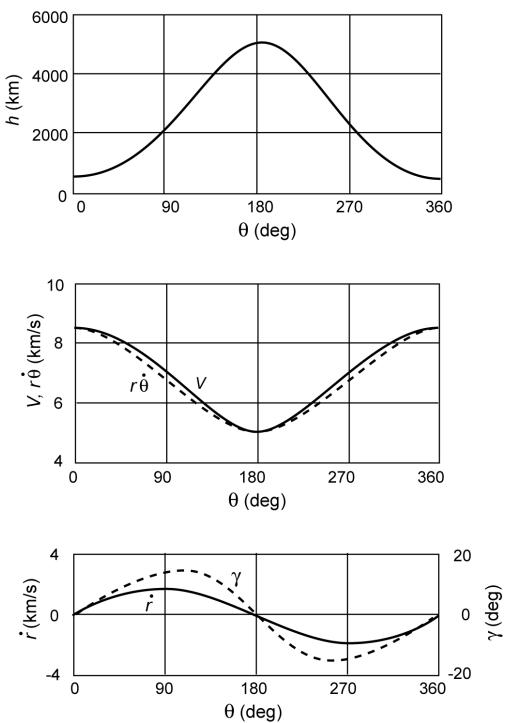


Figure 6.4: Variation of altitude, velocity and flight path angle in an elliptical orbit about the Earth ($h_p = 500$ km, $e = 0.25$).

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

$$n = \sqrt{\frac{\mu}{a^3}} \quad (6.27)$$

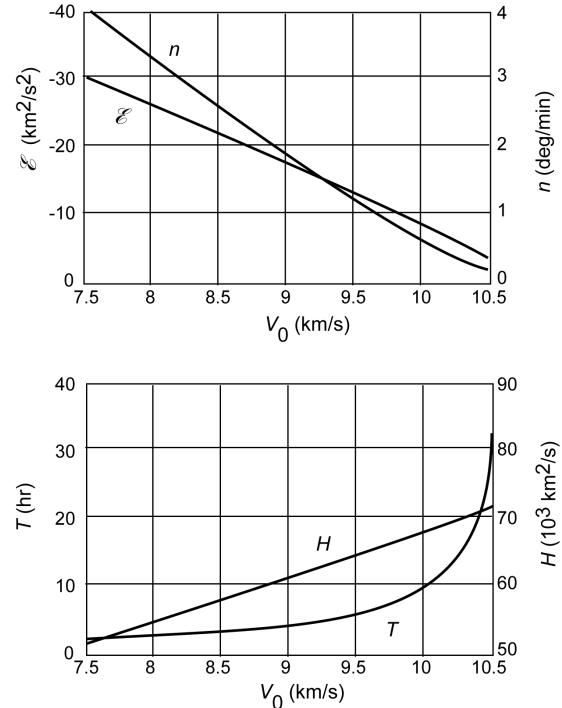


Figure 6.5: Variation of mean angular velocity, orbital energy, orbital angular momentum, and orbital period as a function of the initial velocity of a satellite about the Earth ($t = 0$: $r = 6800$ km, $\gamma = 0^\circ$).

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.

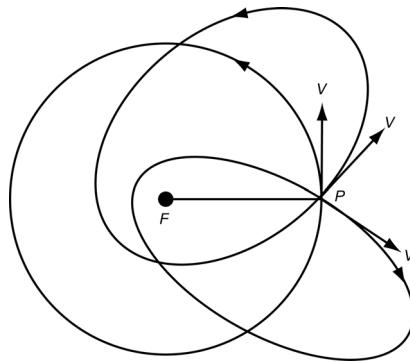


Figure 6.6: Three orbits with the same values of the major axis and of the orbital period.

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

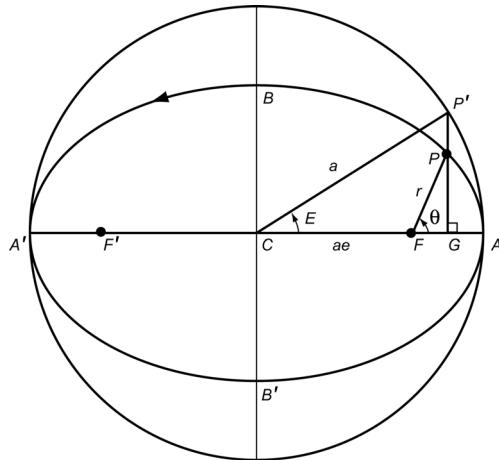


Figure 6.7: Definition of the eccentric anomaly.

$$r = \pm a(1 - e \cos E)$$

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 $1 < h/h_p < 250$.

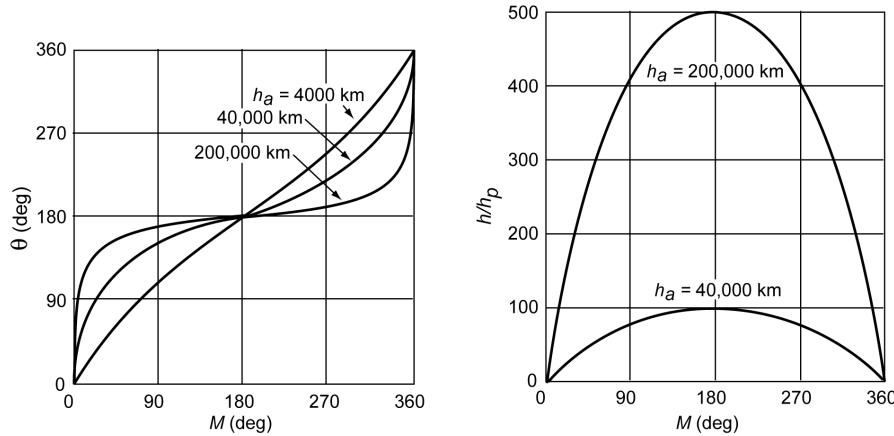


Figure 6.8: Variation of altitude and true anomaly in elliptical orbits about the Earth as a function of the time after perigee passage ($h_p = 400$ km).

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

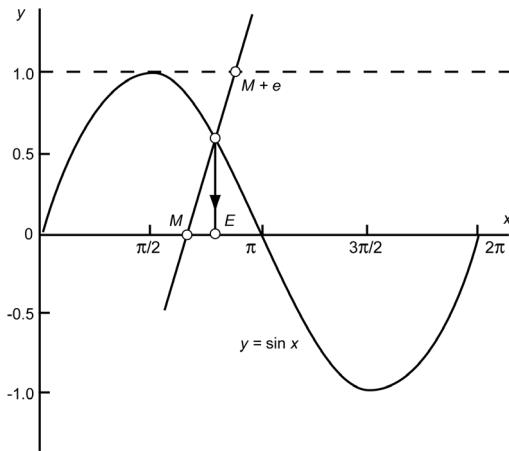


Figure 6.9: Graphical solution of Kepler's equation.

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

$$\alpha_1 e + \alpha_2 e^2 + \alpha_3 e^3 + \dots = e \sin M + e^2 (\alpha_1 \cos M) + e^3 (-\frac{1}{2} \alpha_1^2 \sin M + \alpha_2 \cos M) + \dots$$

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.

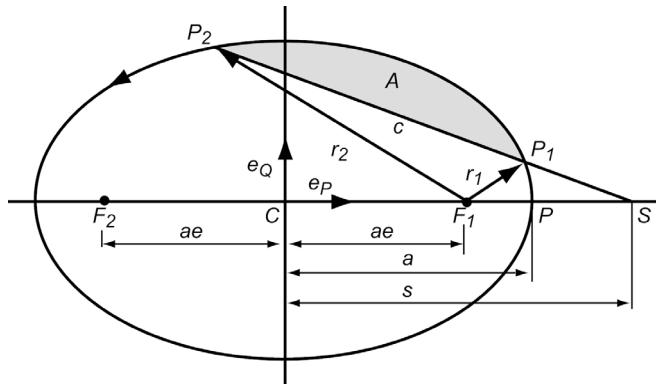


Figure 6.10: Geometry used for the derivation of Lambert's equation for elliptical motion.

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

$$\cos f = 1 - \frac{c + r_1 + r_2}{2a} ; \quad \cos g = 1 + \frac{c - r_1 - r_2}{2a}$$

or

$$\sin^2 \frac{1}{2}f = \frac{r_1 + r_2 + c}{4a} ; \quad \sin^2 \frac{1}{2}g = \frac{r_1 + r_2 - c}{4a} \quad (6.53)$$

Summing (6.51) and (6.52), we find

$$a = \frac{r_1 + r_2 + c}{2(1 - \cos f)}$$

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

$$a_{min} = \frac{1}{4}(r_1 + r_2 + c) \quad (6.54)$$

Further, define a parameter K as

$$K = 1 - \frac{c}{2a_{min}} = \frac{r_1 + r_2 - c}{r_1 + r_2 + c} \quad (6.55)$$

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

$$\bar{r} = e a \bar{e}_P + \lambda \bar{r}_1 + (1 - \lambda) \bar{r}_2 \quad ; \quad -\infty < \lambda < \infty \quad (6.57)$$

This relation, of course, also holds for point S , when $\bar{r} = \bar{r}_S$. The distance s between C and S , measured positive from C in the direction of the pericenter, can also be determined from the equation

$$\bar{r}_S = s \bar{e}_P$$

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).

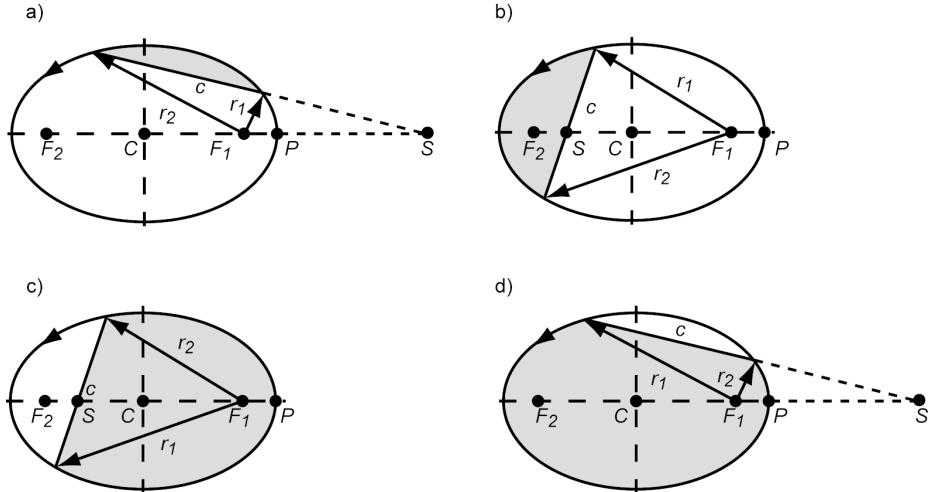


Figure 6.11: Four different cases for Lambert's equation.

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}):

$$V_{esc} = \sqrt{\frac{2\mu}{r}} = \sqrt{2} V_e \quad (7.3-2)$$

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.

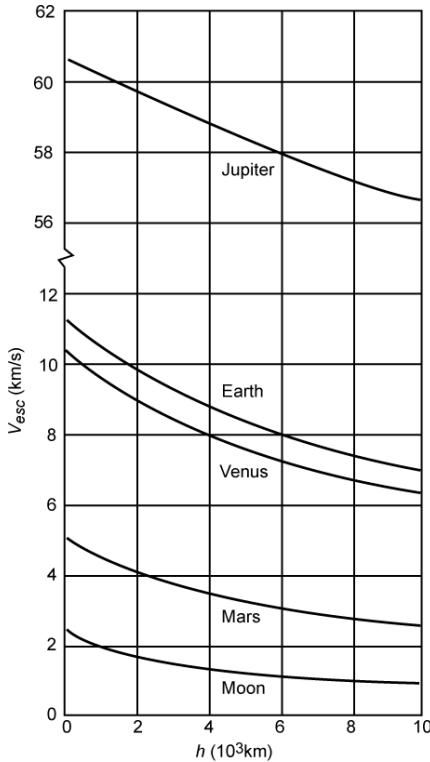


Figure 7.1: Escape velocity as a function of altitude for a number of celestial bodies.

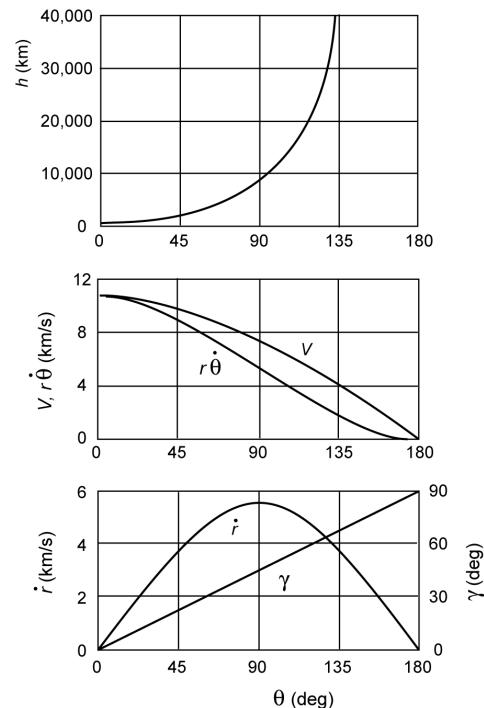


Figure 7.2: Variation of altitude, velocity and flight path angle in a parabolic orbit about the Earth ($h_p = 500$ km).

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

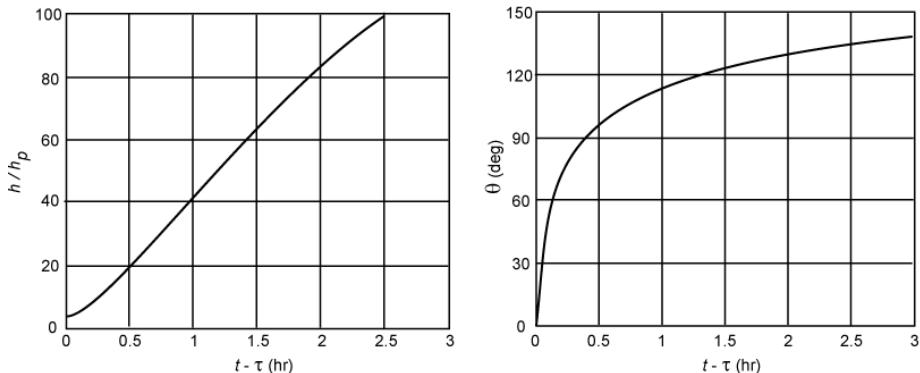


Figure 7.3: Variation of altitude and true anomaly in a parabolic orbit about the Earth as a function of the time after perigee passage ($h_p = 400$ km).

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

$$2\sqrt{\frac{\mu}{p^3}} t_f = \tan \frac{1}{2}\theta_2 - \tan \frac{1}{2}\theta_1 + \frac{1}{3} \left(\tan^3 \frac{1}{2}\theta_2 - \tan^3 \frac{1}{2}\theta_1 \right)$$

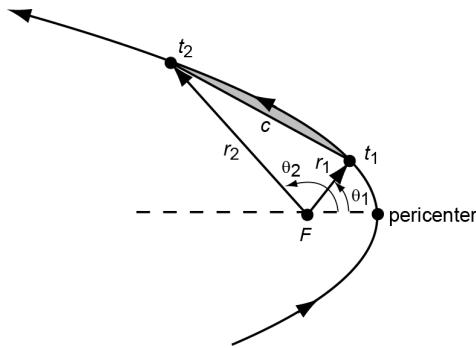


Figure 7.4: Geometry for the derivation of Euler's equation.

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

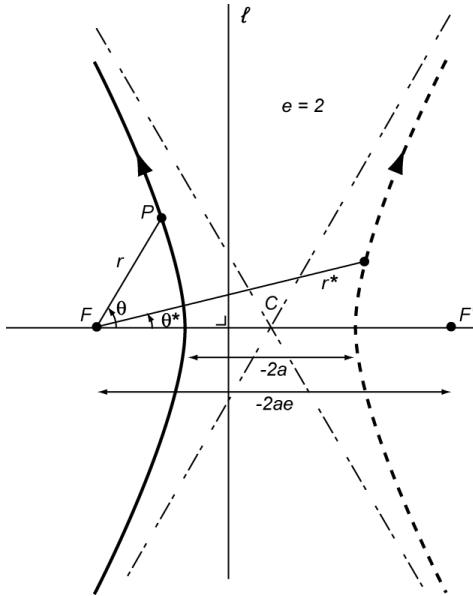


Figure 8.1: Geometry of a hyperbolic orbit.

$$r = \frac{p}{1 + e \cos \theta} \quad (8.1)$$

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.

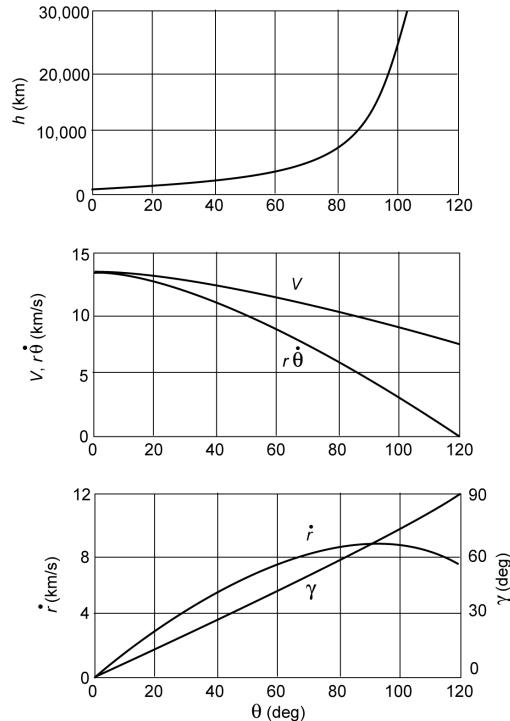


Figure 8.2: Variation of altitude, velocity and flight path angle in a hyperbolic orbit about the Earth ($h_p = 500$ km, $e = 2$).

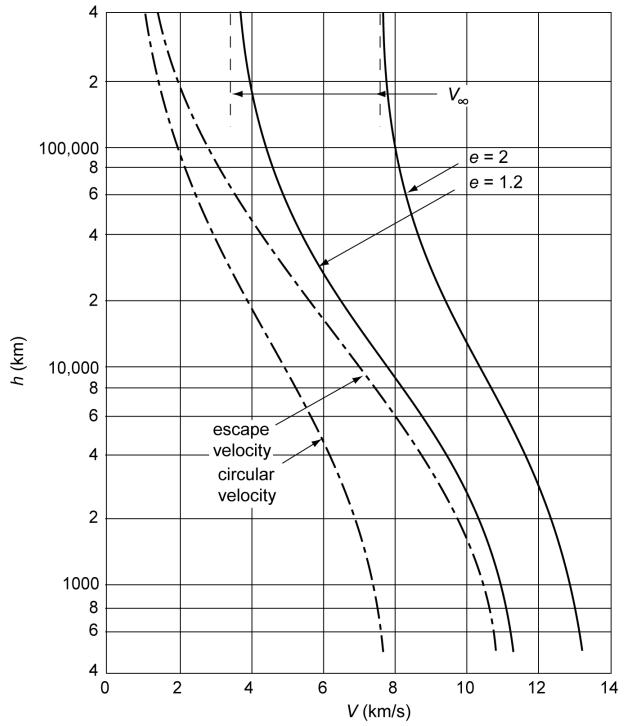


Figure 8.3: Variation of velocity versus altitude for two hyperbolic orbits about the Earth ($h_p = 500$ km), and curves for the local circular velocity and local escape velocity.

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

$$-\infty < \sinh x < \infty \quad ; \quad \cosh x \geq 1 \quad ; \quad -1 < \tanh x < 1$$

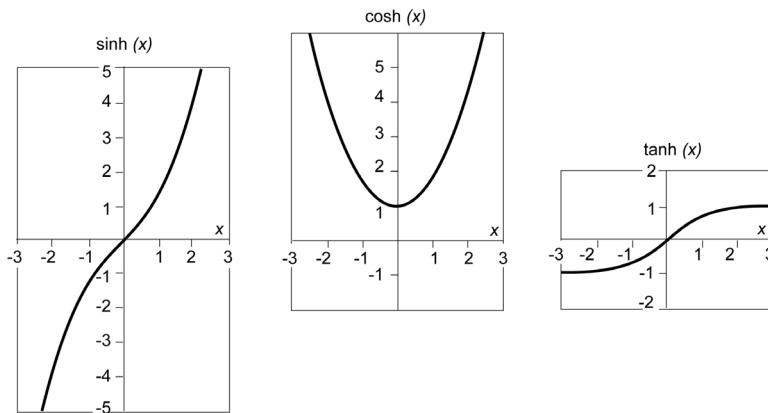


Figure 8.4: The hyperbolic functions $\sinh(x)$, $\cosh(x)$ and $\tanh(x)$.

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.

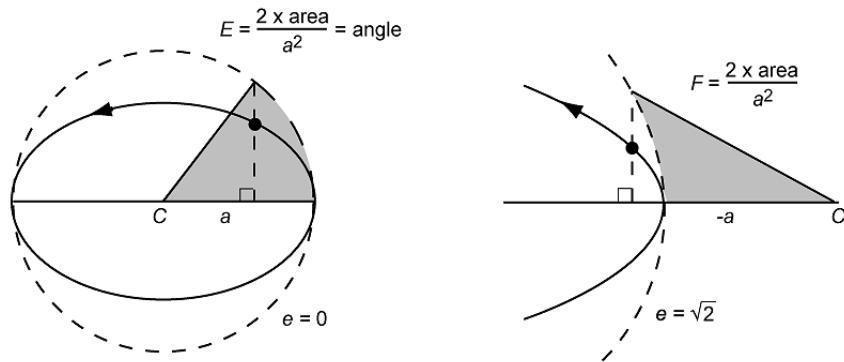


Figure 8.5: Geometric interpretation of the eccentric anomaly and hyperbolic anomaly.

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 .

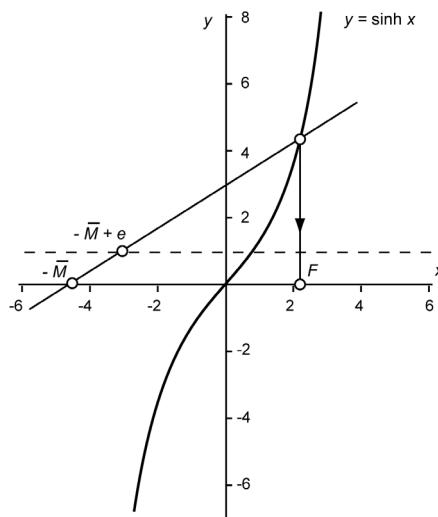


Figure 8.6: Graphical solution of the hyperbolic anomaly.

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.

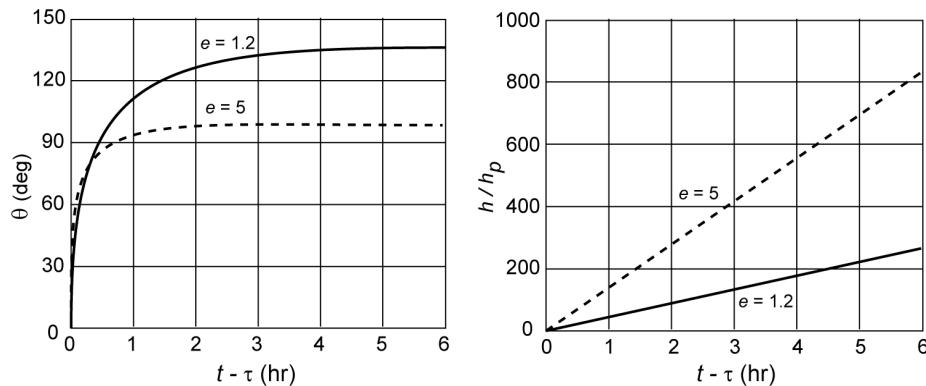


Figure 8.7: Variation of true anomaly and altitude in a hyperbolic orbit about the Earth as a function of time after perigee passage ($h_p = 400$ km).

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.

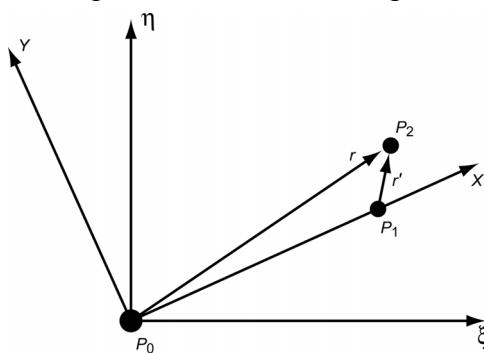


Figure 9.1: Geometry of relative motion. The XYZ reference frame rotates with uniform angular velocity n , about the ζ -axis of the (quasi-)inertial reference frame $\xi\eta\zeta$. The origin of both frames coincides with the Earth's center of mass; P_0 denotes the Earth, P_1 and P_2 the satellites 1 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

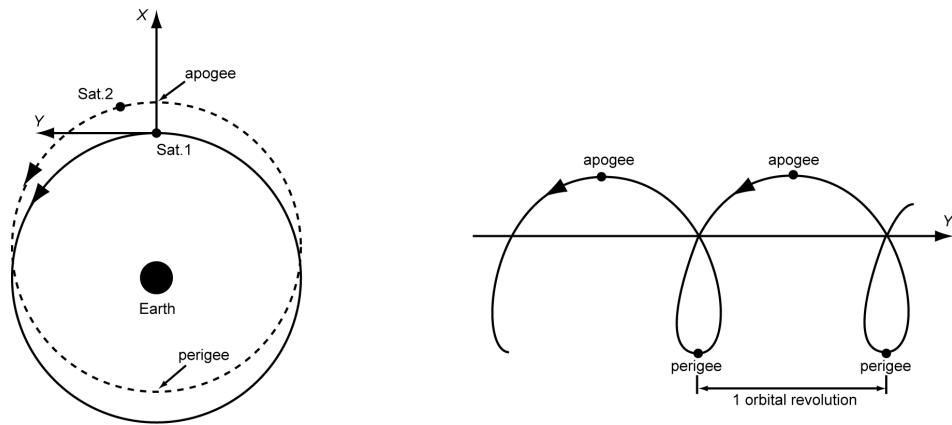


Figure 9.2: Sketch of a characteristic example of the absolute (left) and relative (right) motion of satellite 2 in the XY-plane.

$$\begin{aligned} x_C &= C_1 = 4x_0 + \frac{2\dot{y}_0}{n} \\ y_C &= C_2 - \frac{3}{2}C_1 nt = y_0 - 2\frac{\dot{x}_0}{n} - \frac{3}{2}\left(4x_0 + \frac{2\dot{y}_0}{n}\right)nt \end{aligned} \quad (9.17)$$

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, $z_0 = 0$, $\dot{x}_0 = \dot{z}_0 = 0$, $\dot{y}_0 = -1.5 n x_0$.

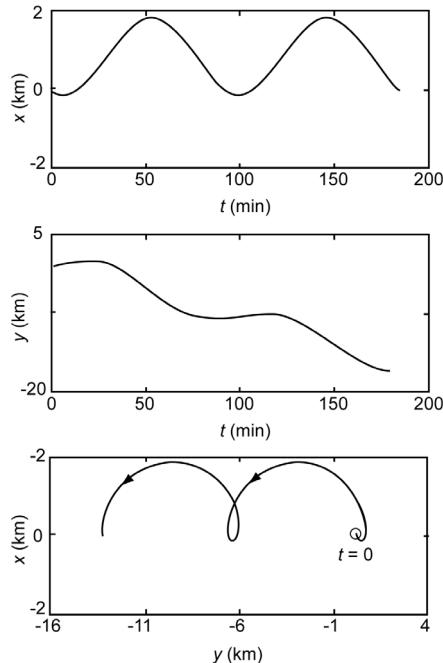


Figure 9.3: In-plane motion of satellite 2 relative to satellite 1 during two orbital revolutions of satellite 1. Assumptions: $h_1 = 400$ km, $x_0 = y_0 = z_0 = 0$, $\dot{x}_0 = -0.5$ m/s, $\dot{y}_0 = 0.5$ m/s, $\dot{z}_0 = 0$.

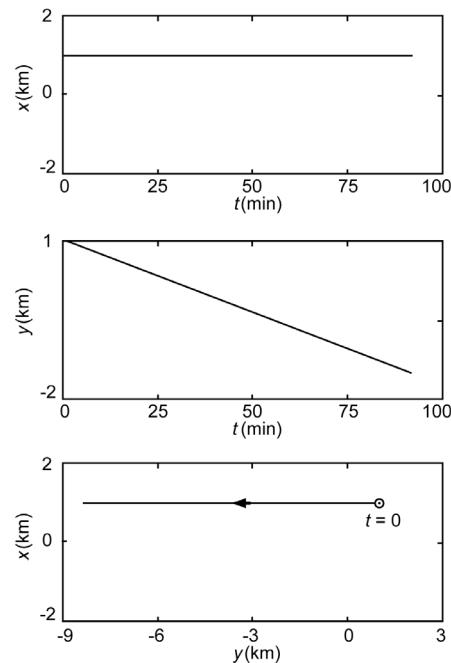


Figure 9.4: In-plane motion of satellite 2 relative to satellite 1 during one orbital revolution of satellite 1. Assumptions: $h_1 = 400$ km, $x_0 = y_0 = 1$ km, $z_0 = 0$, $\dot{x}_0 = \dot{z}_0 = 0$, $\dot{y}_0 = -1.5 n x_0$.

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

$$\begin{aligned} x &= x_0 \cos nt + \frac{\dot{x}_0}{n} \sin nt \\ y &= y_0 - 2x_0 \sin nt + 2 \frac{\dot{x}_0}{n} (\cos nt - 1) \end{aligned} \quad (9.33)$$

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, $\dot{y}_0 = -2nx_0$, $\dot{z}_0 = 0$.

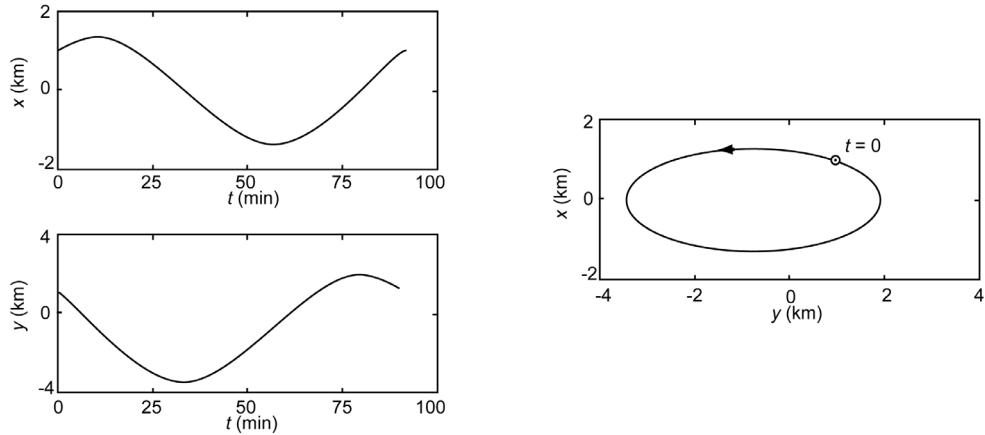


Figure 9.5: In-plane motion of satellite 2 relative to satellite 1 during one orbital revolution of satellite 1. Assumptions: $h_1 = 400 \text{ km}$, $x_0 = y_0 = 1 \text{ km}$, $z_0 = 0$, $\dot{x}_0 = 1 \text{ m/s}$, $\dot{y}_0 = -2nx_0$, $\dot{z}_0 = 0$.

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

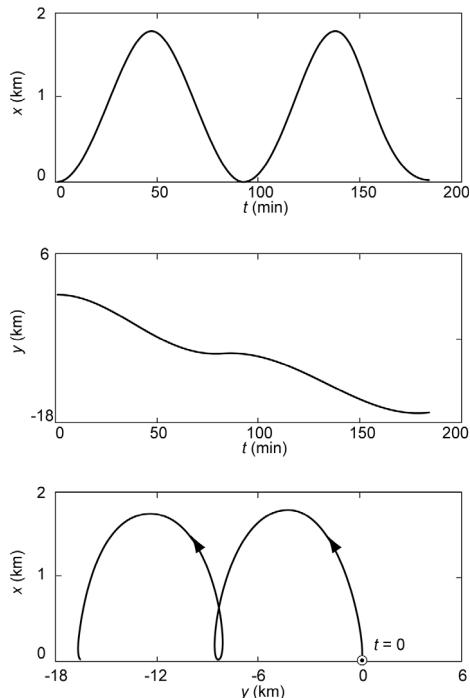


Figure 9.6: In-plane motion of a satellite after a tangential impulsive shot of $\Delta V = 0.5$ m/s relative to the fictitious position of that satellite in its original unperturbed circular orbit at an altitude of 400 km.

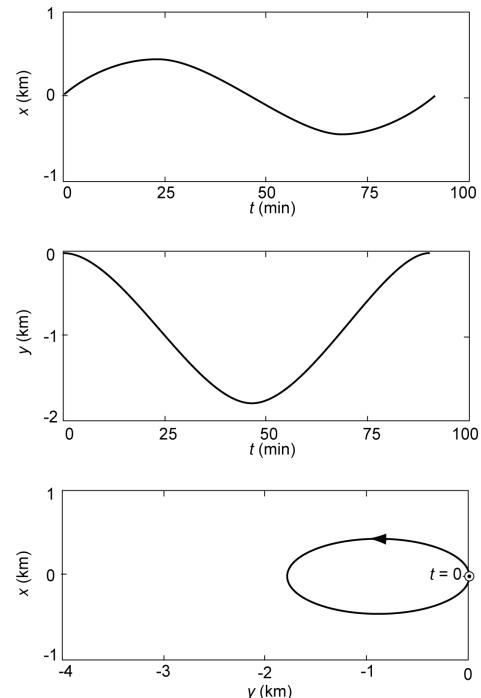


Figure 9.7: In-plane motion of a satellite after a radial impulsive shot of $\Delta V = 0.5$ m/s relative to the fictitious position of that satellite in its original unperturbed circular orbit at an altitude of 400 km.

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.

10. REGULARIZATION

In Chapters 6, 7 and 8 elliptical, parabolic and hyperbolic orbits were analyzed in detail. That are classical topics, where the only difficulty is the efficient manipulation of the various relevant formulas. In this Chapter, some attention will be paid to the more modern and much more difficult concept of *regularization*.

In the discussion on the two-body problem (Chapter 5), it was found that the motion of body i with respect to a non-rotating reference frame with body k at the origin can be described by

$$\frac{d^2\bar{r}}{dt^2} + \frac{\mu}{r^3}\bar{r} = 0 \quad (10.1)$$

In reality, other forces will also act on body i . In general, these additional forces are small and therefore can be considered as *perturbing forces*. The equation of motion for body i then becomes

$$\frac{d^2\bar{r}}{dt^2} + \frac{\mu}{r^3}\bar{r} = \bar{f} \quad (10.2)$$

where \bar{f} is the perturbing force per unit of mass, i.e. the *perturbing acceleration*. Generally, this system of three scalar coupled non-linear second-order differential equations cannot be solved in a closed analytical form, and one is forced to use numerical integration techniques or approximative analytical methods. The techniques used for the computation of perturbed orbits will be discussed further in Chapters 20 through 23.

10.1. Singularity and numerical instability

From (10.2) we conclude that the differential equation describing the motion of body i becomes singular for $r \rightarrow 0$. Of course, in reality r never becomes zero, but one may imagine that this singularity becomes important when body i approaches body k very closely. Classical celestial mechanics had not to worry about this, since close approaches of planetary bodies do not occur. However, in astrodynamics, a close approach often occurs, e.g. in a very-eccentric elliptical orbit. For a given semi-major axis, (6.2) and (6.4) show that if $e \uparrow 1$, then $p \downarrow 0$, $r_p \downarrow 0$ and $r_a \uparrow 2a$, and the trajectory approaches a so-called rectilinear ellipse (Section 5.1). If we consider an orbit with perigee at 500 km altitude and apogee at 200,000 km altitude above the Earth's surface and draw such an orbit to scale, it turns out to be a very elongated orbit in which the magnitude and direction of the satellite's velocity changes quickly near perigee (Figure 10.1). This causes a number of problems, which are directly related to the technique of numerical integration. For an understanding of these problems, it is therefore necessary to briefly discuss the essence of numerical integration processes.

Equation (10.2) can be written as a system of three second-order or six first-order coupled scalar differential equations. Several methods have been developed to numerically integrate differential equations. The best-known are those belonging to the Runge-Kutta family and the methods of Nordsieck, Adams-Bashforth, Gauss-Jackson and Adams-Moulton. When, for simplicity, we write a first-order differential equation as

$$\frac{dy}{dx} = f(x, y) \quad (10.3)$$

then we may state that numerical integration techniques are founded on the application of an

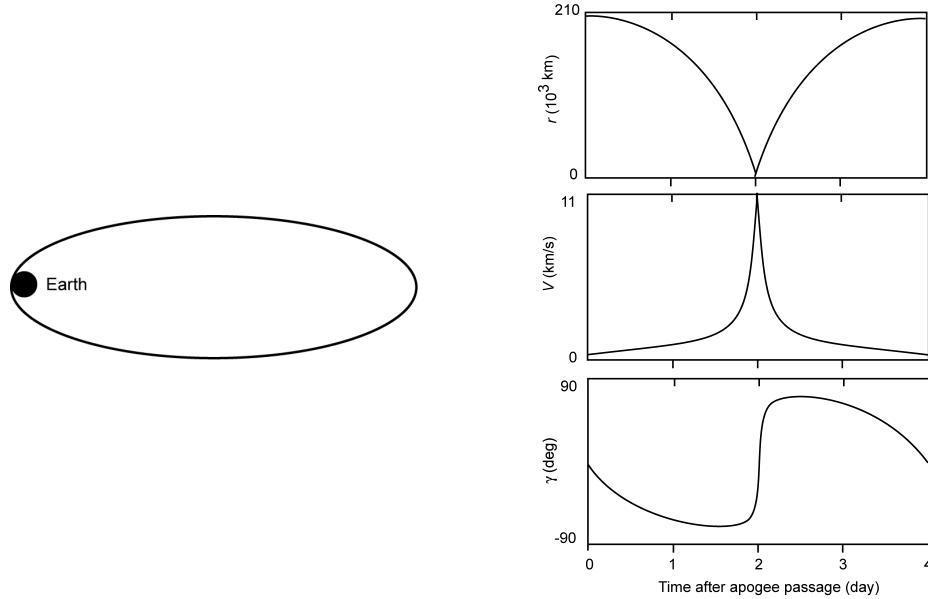


Figure 10.1: A highly-eccentric elliptical orbit with perigee and apogee at 500 km and 200,000 km altitude, respectively (left), and the variation of distance, velocity and flight path angle as a function of the time after apogee passage.

expression of the type

$$y_{n+1} = y_n + \mathcal{F}(h, f_{n+1}, f_n, f_{n-1}, \dots, f_m) \quad (10.4)$$

where for each integration technique the structure of the function \mathcal{F} is different. In this expression, h indicates the stepsize of the integration process, $f_{n+1}, f_n, f_{n-1}, \dots, f_m$ symbolically indicate a series of function evaluations at $x_n + h, x_n, x_n - h, \dots, x_m$. When the value of y_n is known, a value of h is chosen and the function on the right-hand side of (10.3) is evaluated at the selected evaluation points. Then, from (10.4) follows the value of y_{n+1} , with which the process can be repeated until the end of the integration interval has been reached. It is emphasized that this is a highly-simplified description of the integration process; however, it is sufficient for the present discussion.

Because for very-eccentric elliptical orbits velocity and acceleration of body i become large near perigee (close to the singularity), the accuracy with which the function \mathcal{F} in (10.4) can approximate the true course of y will decrease in that region. Integration techniques that make use of a large number of function evaluations in previous integration points, or intermediate points, will be less susceptible, but this phenomenon will always be present. This leads to the so-called *truncation error* in the integration process. Generally, this type of error increases with an increasing stepsize. Therefore, one would like to decrease the stepsize of the integration process. However, for a given integration interval this would mean more integration steps, causing the *round-off error* to increase. This round-off error is a consequence of the fact that in a computer a number is represented by a limited amount of significant digits, which is related to the so-called wordlength used by the computer. Such an error occurs every integration step; it is often cumulative and the accumulated round-off error thus increases with the number of integration steps, i.e. with a decreasing stepsize. This discussion not only holds for very-eccentric orbits, but also for every elliptical orbit, although to a lesser extent.

The occurrence of these two types of integration errors requires a careful selection of the

stepsize to be applied in order to obtain high-accuracy results. Several techniques are known with which one can determine the local optimum stepsize to be selected. For a non-circular orbit, the optimum stepsize near perigee will be (much) smaller than near apogee. This, however, leads to another problem, namely that most efficient integration techniques lose their efficiency when the stepsize has to be changed too often during the integration process. To give an impression of the effect of the applied stepsize on the accuracy of the computed orbit, Figure 10.2 shows some results of numerical integrations. It was assumed that the satellite moves in a circular orbit

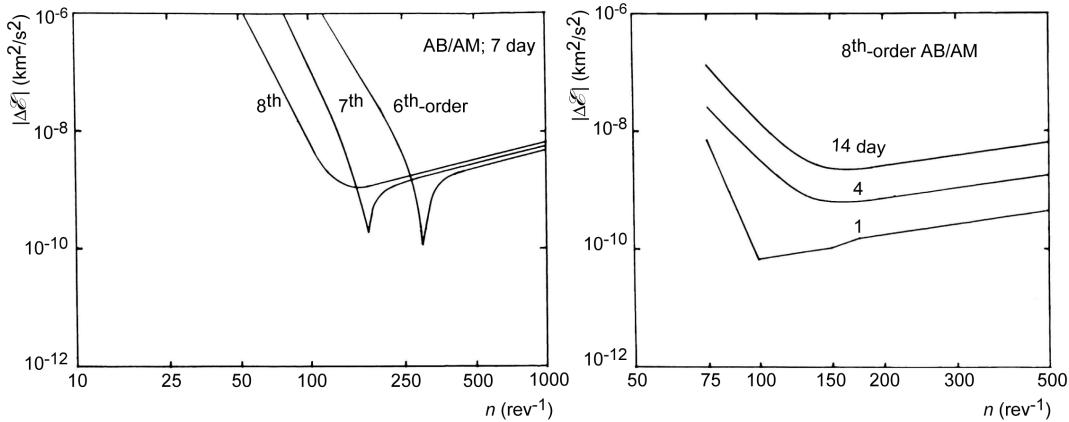


Figure 10.2: Variation of $|\Delta E|$ as a function of the number of integration steps per orbital revolution, n , for a total integration period of 7 days and using a 6th-, 7th- or 8th-order Adams-Bashforth/Adams-Moulton integration scheme (left); and for integration periods of 1, 4 or 14 days and using an 8th-order Adams-Bashforth/Adams-Moulton integration scheme (right).

with an altitude of 900 km and an inclination (Section 11.5) of 45° . The orbit is supposed to be perturbed only by zonal harmonics (Section 20.1) of the Earth's gravity field. In that case, the gravity field model is axially-symmetric about the Earth's polar axis. In Section 23.1, it will be shown that for such a gravity field the total energy per unit of mass of the satellite, \mathcal{E} , is constant. This means that if the deviation of the local total energy from its original value, $\Delta\mathcal{E}$, is computed at each integration step from the position and velocity vectors at that time, as obtained in the numerical integration process, the deviation $\Delta\mathcal{E}$ can be attributed to numerical integration errors. The numerical integration was performed on an IBM 370/158 machine in double precision, which is equivalent to a wordlength of seventeen significant digits. Only the first twenty zonal harmonics, i.e. the terms with the coefficients J_2 up to J_{20} (Section 20.1) were taken into account. The integration was performed with a sixth-, seventh- or eighth-order Adams-Bashforth/Adam-Moulton (AB/AM) predictor/corrector scheme that is often applied in astrodynamics; this multi-step integration procedure was started with an eighth-order Runge-Kutta scheme. Figure 10.2 (left) shows the values of $|\Delta\mathcal{E}|$ as a function of the number of integration steps per orbital revolution for a total integration period of 7 days, and the sixth-, seventh- or eighth-order integration scheme. This Figure clearly demonstrates that there is an optimum integration stepsize. For the sixth- and eighth-order methods, the optimum number of integration steps per revolution is about 300 and 160, respectively, which is for the orbital altitude of 900 km equivalent to a stepsize of about 20 s and 40 s, respectively. Interesting to note is that for the larger number of steps per revolution (smaller stepsize) the sixth-order method provides slightly more-accurate results, but that for stepsizes larger than 25 s the eighth-order method provides much more accurate results. Figure 10.2 (right) shows the values of $|\Delta\mathcal{E}|$ as a function of the number of integration steps per revolution, for integration periods of 1, 4 and 14 days, and for

the eighth-order integration scheme. This Figure shows that the optimum integration stepsize does not vary much with the total integration period, but that the accuracy of the integrated orbit certainly degrades with increasing integration periods. For all curves, the increase of $|\Delta\mathcal{E}|$ at stepsizes smaller than the optimum stepsize is due to the increasing round-off error resulting from the larger number of steps within the integration period; the increase of $|\Delta\mathcal{E}|$ at stepsizes larger than the optimum stepsize is due to the increasing truncation error.

As a result of the problems mentioned above, a need was felt to transform (10.2) into a form that is no longer singular for $r \rightarrow 0$. Such a process is called *regularization*. The credit for introducing *regularizing variables* is usually given to K.F. Sundman (1873-1949). For the regularization of the two-body problem he used transformations of the independent (time) variable. It is recognized today that such transformations do not offer the best formulation for numerical analyses. Transformations involving both the dependent and the independent variables are preferred to put the equations of motion in their optimum form for numerical integration. Earlier, L. Euler (1707-1783) used regularization to study the straight-line motion in the problem of two bodies. T. Levi-Civita (1873-1941) regularized the restricted problem of three bodies in 1903. K. Stumpff (1895-1970) and S. Herrick (1911-1974) used basically the same ideas when they introduced the concept of *universal variables* (Section 11.11). The popularity of these regularization techniques only came after 1968 and presently several regularization techniques are known. The application of these techniques is even mandatory when one deals with close approaches of celestial bodies in interplanetary missions (Chapter 18), and, in general, for the accurate computation of highly-eccentric orbits.

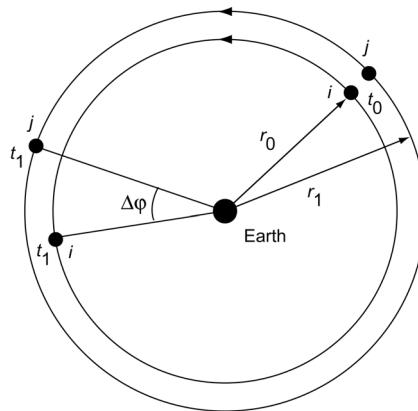


Figure 10.3: Two satellites in circular orbits with slightly different radii.

There is, however, another problem with the differential equation (10.2). That is the fact that even the unperturbed solution is *dynamically unstable*. This subject was already mentioned in Section 5.8 for a circular orbit. In the following, a different treatment of this phenomenon will be given. We consider the case that body i describes a circular orbit about the Earth, where the radius of the orbit is r_0 (Figure 10.3). The angular velocity of body i then is $n_0 = (\mu/r_0^3)^{1/2}$. Now, imagine that another body j is at $t = 0$ near body i , but in a circular orbit with a slightly larger radius r_1 . The angular velocity of body j is $n_1 = (\mu/r_1^3)^{1/2}$. As a result of the different angular velocities, the relative angular position, $\Delta\phi$, of the bodies i en j will increase after $t = 0$, according to

$$\Delta\phi = \left(\sqrt{\frac{\mu}{r_0^3}} - \sqrt{\frac{\mu}{r_1^3}} \right) (t - t_0)$$

No matter how small $|r_1 - r_0|$, the angular distance between the bodies will become large when $(t - t_0)$ is large enough. This implies that when a small error is made in the computation of a satellite orbit, for instance due to the inevitable round-off and truncation errors related to the numerical integration of the equations of motion, the computed position of the satellite in its orbit will show increasing along-track errors. So, although the calculation of the orbit itself is a stable process, the computation of the position of the satellite in its orbit (along track) shows an instability. The second advantage of regularization is that a large portion of this *dynamic instability* disappears.

From the discussion given above it will be clear that both the singularity and the dynamic instability have nothing to do with the perturbing acceleration itself, but are characteristics of the two-body problem. For an introduction to regularization techniques, we therefore will assume in Section 10.2 that the orbit is a pure Keplerian one. The method of regularization that will be used is that of C.A. Burdet (-), which was published in 1968.

10.2. Method of Burdet for a Keplerian orbit

The starting point for the analysis is (5.32), which can be written as

$$(V^2 - \frac{\mu}{r})\bar{r} - (\bar{r} \cdot \bar{V})\bar{V} = \mu \bar{e} = \frac{\mu e}{r_p} \bar{r}_p \quad (10.5)$$

where \bar{e} is the eccentricity vector or first Laplace vector (Section 5.7). When we subtract the term $\mu \bar{e}/r^2$ on both sides of (10.1), we obtain

$$\frac{d^2\bar{r}}{dt^2} - \frac{1}{r^2} \mu \bar{e} + \frac{\mu}{r^3} \bar{r} = -\frac{1}{r^2} \mu \bar{e} \quad (10.6)$$

Substitution of (10.5) into the left-hand side of (10.6) yields

$$\frac{d^2\bar{r}}{dt^2} + \frac{1}{r} \frac{dr}{dt} \frac{d\bar{r}}{dt} - (V^2 - 2\frac{\mu}{r}) \frac{1}{r^2} \bar{r} = -\frac{1}{r^2} \mu \bar{e}$$

Substitution of the classical relation for Keplerian orbits:

$$V^2 - \frac{2\mu}{r} = -\frac{\mu}{a} = 2\mathcal{E} \quad (10.7)$$

results in

$$\frac{d^2\bar{r}}{dt^2} + \frac{1}{r} \frac{dr}{dt} \frac{d\bar{r}}{dt} + \frac{\mu}{ar^2} \bar{r} = -\frac{1}{r^2} \mu \bar{e} \quad (10.8)$$

Note that we already have made some progress in the process of regularization: the singularity has weakened, because instead of r^3 now r^2 appears as the highest power in the denominator.

Subsequently, a *fictitious time*, s , is introduced, which is defined by

$$\Delta t = \frac{1}{\sqrt{\mu}} \int r ds \quad \text{or} \quad dt = \frac{r}{\sqrt{\mu}} ds \quad \text{or} \quad ds = \frac{\sqrt{\mu}}{r} dt \quad (10.9)$$

This relation shows that when a constant stepsize Δs is applied, near perigee much smaller steps in physical time t are taken than near apogee (Figure 10.4). From the discussion in Section 10.1,

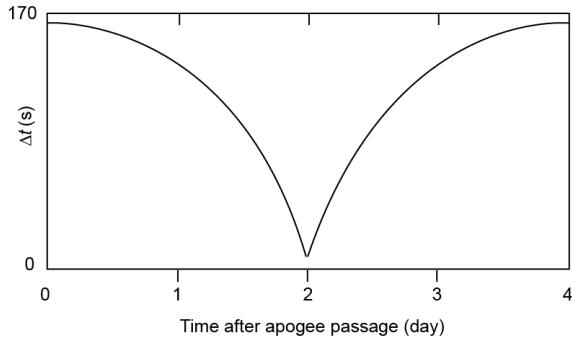


Figure 10.4: Variation of the stepsize Δt in physical time as a function of the time after apogee passage and for a constant stepsize $\Delta s = 0.5 \text{ km}^{1/2}$ in fictitious time. Orbital perigee and apogee altitudes are 500 km and 200,000 km, respectively.

we conclude that a transformation of the physical time, t , in the differential equation (10.1) to the independent variable s will already have a positive effect on the accuracy of the numerical integration process. Therefore, this transformation is always recommended, even when no regularization is aimed for. Note that the transition to the new time variable s implies that seven instead of six first-order differential equations have to be integrated for the computation of position and velocity of a satellite at a certain time t . This is no problem, because in the present age of high-speed computers, the number of differential equations is not as important as the stability and the numerical behavior of the system of differential equations.

From (10.9) we obtain

$$\begin{aligned}\frac{d\bar{r}}{dt} &= \frac{d\bar{r}}{ds} \frac{ds}{dt} = \frac{\sqrt{\mu}}{r} \frac{d\bar{r}}{ds} \quad ; \quad \frac{dr}{dt} = \frac{\sqrt{\mu}}{r} \frac{dr}{ds} \\ \frac{d^2\bar{r}}{dt^2} &= \frac{d}{ds} \left(\frac{\sqrt{\mu}}{r} \frac{d\bar{r}}{ds} \right) \frac{\sqrt{\mu}}{r} = -\frac{\mu}{r^3} \frac{dr}{ds} \frac{d\bar{r}}{ds} + \frac{\mu}{r^2} \frac{d^2\bar{r}}{ds^2}\end{aligned}$$

Substitution of these relations into (10.8) yields

$$\frac{d^2\bar{r}}{ds^2} + \frac{1}{a} \bar{r} = -\bar{e} \tag{10.10}$$

in which a and \bar{e} are constants, which are determined by the initial conditions. The equation of motion (10.10) can be written as three scalar *uncoupled linear* second-order differential equations, which display for $r \downarrow 0$ no singularity anymore. Comparing this equation with (5.14), we note that they look somewhat similar. This means that the introduction of $u = 1/r$ may be considered as some form of smoothing and regularization. However, this concerns only the in-orbit position, while the regularization technique described in this Section provides a regularization of the full three-dimensional position vector.

When we consider elliptical orbits, for which $0 < a < \infty$, then the solution of (10.10) is

$$\bar{r} = -a\bar{e} + \bar{A} \sin\left(\sqrt{\frac{1}{a}}s\right) + \bar{B} \cos\left(\sqrt{\frac{1}{a}}s\right) \tag{10.11-1}$$

where \bar{A} and \bar{B} are integration constants (vectors), which follow from the initial conditions. When we assume that at $s = 0$ the satellite passes perigee, we find from (10.11-1)

$$\bar{r}_p = -\alpha \bar{e} + \bar{B} \quad ; \quad \left(\frac{d\bar{r}}{ds} \right)_p \equiv \bar{V}_p = \sqrt{\frac{1}{\alpha}} \bar{A}$$

where the index p refers to perigee and \bar{V}_p is the velocity at perigee. Since \bar{e} points towards perigee, we conclude that \bar{B} also points towards perigee and that \bar{A} points along the velocity vector at perigee and thus is normal to \bar{r}_p and \bar{B} . Equation (10.11-1) describes a three-dimensional harmonic oscillator with a constant frequency $\sqrt{1/\alpha}/2\pi$ and with its center at $-\alpha \bar{e}$. The frequency is fully determined by the initial conditions; α and \bar{e} are constants and do not have to be computed over and over again during the numerical integration process. When it is assumed that the initial conditions are perfectly known, it is obvious that numerical integration errors have no influence on the frequency and thus that the computation of the along-track position is dynamically and numerically stable.

Scalar multiplication of (10.10) by \bar{r} results, after substitution of (10.5), in

$$\bar{r} \cdot \frac{d^2\bar{r}}{ds^2} + \frac{1}{\alpha} r^2 + \frac{1}{\mu} \left[(V^2 - \frac{\mu}{r}) r^2 - (\bar{r} \cdot \bar{V}) (\bar{r} \cdot \bar{V}) \right] = 0 \quad (10.12)$$

We may write

$$\begin{aligned} \bar{V} &= \frac{d\bar{r}}{dt} = \frac{d\bar{r}}{ds} \frac{ds}{dt} = \frac{\sqrt{\mu}}{r} \frac{d\bar{r}}{ds} \quad ; \quad V^2 = \frac{\mu}{r^2} \left(\frac{dr}{ds} \right)^2 \\ \bar{r} \cdot \bar{V} &= r \frac{dr}{dt} = \sqrt{\mu} \frac{dr}{ds} \quad ; \quad \bar{r} \cdot \frac{d\bar{r}}{ds} = r \frac{dr}{ds} \quad ; \quad \bar{r} \cdot \frac{d^2\bar{r}}{ds^2} = r \frac{d^2r}{ds^2} \end{aligned} \quad (10.13)$$

Substitution of these relations into (10.12) results in a simple differential equation for the distance:

$$\frac{d^2r}{ds^2} + \frac{r}{\alpha} = 1 \quad (10.14)$$

For an elliptical orbit ($0 < \alpha < \infty$), the general solution of (10.14) is

$$r = \alpha + C \sin\left(\sqrt{\frac{1}{\alpha}} s\right) + D \cos\left(\sqrt{\frac{1}{\alpha}} s\right) \quad (10.11-2)$$

where C and D are integration constants, which vanish both for circular orbits. Substitution of (10.11-2) into (10.9) yields for the physical time

$$\Delta t = \frac{1}{\sqrt{\mu}} \int \left\{ \alpha + C \sin\left(\sqrt{\frac{1}{\alpha}} s\right) + D \cos\left(\sqrt{\frac{1}{\alpha}} s\right) \right\} ds \quad (10.15)$$

or

$$\Delta t = \frac{\alpha}{\sqrt{\mu}} \Delta s + \text{periodic terms}$$

This relation clearly shows that it is important to know the value of α , which is determined by the initial conditions, very precisely. Any inaccuracy in α generates a continuously growing error in

the computation of the physical time. So, due to regularization the dynamic instability has been removed from the equations of motion, but it reappears in the relation for the physical time. It is logical that the dynamic instability has not disappeared completely; after all, no mathematical ‘trick’ can eliminate a physical reality! However, the advantage of the regularization of the equations of motion is that during the numerical integration of these equations in fictitious time no stability problem occurs, resulting in a minimum accumulation of computation errors during the numerical integration process.

To clarify the physical meaning of the fictitious time, s , we assume that $t = 0$ at a satellite perigee passage. So,

$$t = 0: \quad s = 0 \quad ; \quad r = a(1 - e) \quad ; \quad \frac{dr}{ds} = 0$$

Combination of these conditions with (10.11-2) yields $C = 0, D = -ae$. With these values, (10.11-2) and (10.15) can be written as

$$r = a \left[1 - e \cos\left(\sqrt{\frac{1}{a}}s\right) \right] \quad ; \quad t = \sqrt{\frac{a^3}{\mu}} \left[\sqrt{\frac{1}{a}}s - e \sin\left(\sqrt{\frac{1}{a}}s\right) \right] \quad (10.16)$$

A comparison between (10.16) and the equations for elliptical motion presented in Section 6.5:

$$r = a(1 - e \cos E) \quad ; \quad E - e \sin E = \sqrt{\frac{\mu}{a^3}}(t - \tau)$$

shows that the fictitious time, s , is directly related to the eccentric anomaly, E , according to

$$\sqrt{\frac{1}{a}}s = E \quad (10.17)$$

This suggests that the eccentric anomaly, and thus also the true anomaly, which are the usual parameters to describe the position of a satellite in an elliptical orbit, are ‘regularizing parameters’. This may be proved by using equations (5.21), (6.33) and (6.36-2) to derive the following relations:

$$d\theta = H \frac{dt}{r^2} \quad ; \quad dE = a n \frac{dt}{r} \quad (10.18)$$

Comparing these transformation relations with (10.9), we conclude that the transformation of time, t , to true anomaly, θ , or eccentric anomaly, E , will have a positive effect on the accuracy of the computation process. We may even say that the origin of regularization could be contributed to Kepler. He certainly had no idea of the concept of regularization, especially since to him calculus and differential equations were not known. Furthermore, Kepler’s interest in the dynamics of the solar system excluded close approaches. Nevertheless, the use of the eccentric anomaly introduced in his equation (6.36) does result in a uniform distribution of the local truncation errors along the orbit, which certainly benefits the integration process (Section 10.3).

10.3. Generalized time transformation

The transformation of the independent variable as given by (10.9) and (10.18) can be generalized and written as

$$ds = C \frac{dt}{r^m} \quad (10.19)$$

where C and m are constants. Note that $m = 1$ corresponds to the introduction of the eccentric anomaly, E , and $m = 2$ to the introduction of the true anomaly, θ , with proper selection of the constant A . As already stated in Section 10.2, such a transformation should preferably always be applied, even if no regularization of the equations of motion is pursued. However, it is pointed out that there exists an essential difference between the integration of the original equations of motion and applying (10.19), and regularization. In the latter case, the differential equations are rewritten in terms of a new independent variable. In the first case, we integrate the original equations of motion with a variable stepsize in physical time, which does not eliminate the singularity. Therefore, such a technique must necessarily fail when a near-encounter takes place, since then the required number of integration steps tends to increase enormously, leading to large integration errors.

In the period 1968 to 1975 many studies have been performed to determine the optimum value of m in case a transformation of the type (10.19) is applied to integrate the original (non-regularized) equations of motion. Various authors (e.g. J. Baumgarte, V.R. Bond, C.A. Burdet, J.M.A. Danby, T. Feagin, R.P. Mikkilineni, P. Nacozy, C.E. Velez, P. Wong) have published the results of their analyses. Some of their major conclusions can be summarized as follows:

- Time transformations of the type (10.19) often reduce significantly the overall integration errors.
- For $m = 1$, the dynamic instability of a Keplerian orbit is significantly reduced; as m is increased beyond $m = 1$, the dynamic instability further decreases. For $m = 2$, the unperturbed two-body system is completely stabilized in the Liapunov sense.
- However, as m increases, the numerical instability of the applied integration algorithms may increase, leading to the possibility of a larger global integration error by overriding the reduced dynamic instability.
- Transformations of the type (10.19) may provide *uniformization* of local truncation errors, also called *analytical stepsize regulation*. This means that for a proper value of m in (10.19) and using a constant stepsize, each step will contribute about the same local truncation error. Then, fixed-stepsize integration methods, which are in many situations more efficient than variable-stepsize methods, can be used.
- For the unperturbed two-body problem, $m = 1$ provides approximately uniformization of local truncation errors throughout the orbit; for a two-body problem with perturbations due to the second zonal harmonic of the Earth's gravity field (Sections 20.1 and 21.2), $m = 1.5$ provides uniformization.
- Time transformations of the type (10.19) may reduce the local truncation error, allowing a larger integration step. At perigee, $m = 2$ provides a smaller local truncation error than $m = 1$ when a fourth-order Runge-Kutta integration scheme is applied; at apogee, $m = 1$ is better than $m = 2$.
- When a seventh-order Runge-Kutta scheme is applied to integrate an eccentric orbit perturbed by the Earth's second zonal harmonic, for a minimum local truncation error at perigee m should be between 1.5 and 2 and for a minimum error at apogee between 1 and 1.5. In general, it turns out that in many occasions a best comprise is to use $m = 1.5$ for the entire orbit.

10.4. Method of Burdet for a perturbed Keplerian orbit

The relations (10.10) and (10.14) are valid for an unperturbed orbit. In a similar way, for a

perturbed orbit we find

$$\frac{d^2\bar{r}}{ds^2} + \frac{1}{a(s)}\bar{r} + \bar{e}(s) = \frac{r^2}{\mu}\bar{f} \quad (10.20)$$

$$\frac{d^2r}{ds^2} + \frac{1}{a(s)}r - 1 = \frac{r}{\mu}(\bar{f} \cdot \bar{r}) \quad (10.21)$$

It is emphasized that the parameter a has to be interpreted as the semi-major axis as defined by (10.7). This means that it is related to the total energy of a Keplerian orbit, in which the potential energy (per unit of mass) is given by $-\mu/r$. These equations are also non-singular for $r \downarrow 0$. However, in this case a and \bar{e} are no longer constants, but have to be computed again after each integration step. But, they only vary as a result of the perturbing acceleration \bar{f} and not as a result of the main central acceleration $-\mu/r^2$. This means that for perturbed orbits regularization will not fully eliminate the effects of the dynamic instability on the computation of the along-track position of a satellite, but these effects will be much smaller than when (10.2) is integrated directly. Moreover, when one realizes that a is related to the total energy of the motion, it will be clear that regularization is particularly useful when the total energy can be computed separately from the integration process, which is the case when the perturbing forces are described by conservative force fields.

In principle, the computation of the quantities a and \bar{e} can take place by differentiation of (10.7) and (10.5), which hold for Keplerian orbits, provided that they are written as

$$\begin{aligned} \frac{1}{a} &= \frac{2}{r} - \frac{V^2}{\mu} \\ \bar{e} &= \left(\frac{V^2}{\mu} - \frac{1}{r} \right) \bar{r} - \frac{1}{\mu} (\bar{r} \cdot \bar{V}) \bar{V} \end{aligned} \quad (10.22)$$

Differentiation of (10.22-1) yields

$$\frac{d}{ds} \left(\frac{1}{a} \right) = -\frac{2}{\mu} \left(\frac{\mu}{r^2} \frac{dr}{ds} + V \frac{dV}{ds} \right) \quad (10.23)$$

Now, the scalar product of (10.20) and $d\bar{r}/ds$ is taken:

$$\frac{d\bar{r}}{ds} \cdot \frac{d^2\bar{r}}{ds^2} + \frac{1}{a} \bar{r} \cdot \frac{d\bar{r}}{ds} + \bar{e} \cdot \frac{d\bar{r}}{ds} = \frac{r^2}{\mu} \bar{f} \cdot \frac{d\bar{r}}{ds}$$

Substitution of (10.22-1) and (10.22-2) into this expression results in

$$\frac{d\bar{r}}{ds} \cdot \frac{d^2\bar{r}}{ds^2} + \frac{dr}{ds} - \frac{1}{\mu} (\bar{r} \cdot \bar{V}) (\bar{V} \cdot \frac{d\bar{r}}{ds}) = \frac{r^2}{\mu} \bar{f} \cdot \frac{d\bar{r}}{ds} \quad (10.24)$$

From the relations (10.13) we obtain

$$\frac{d\bar{r}}{ds} = \frac{r}{\sqrt{\mu}} \bar{V} ; \quad \frac{d^2\bar{r}}{ds^2} = \frac{1}{\sqrt{\mu}} \frac{dr}{ds} \bar{V} + \frac{r}{\sqrt{\mu}} \frac{d\bar{V}}{ds} ; \quad \bar{V} \cdot \frac{d\bar{r}}{ds} = \frac{r}{\sqrt{\mu}} V^2 \quad (10.25)$$

Substitution of these relations into (10.24) gives

$$\frac{\mu}{r^2} \frac{dr}{ds} + V \frac{dV}{ds} = \bar{f} \cdot \frac{d\bar{r}}{ds} \quad (10.26)$$

Substitution of (10.26) into (10.23) gives

$$\frac{d}{ds} \left(\frac{1}{a} \right) = -\frac{2}{\mu} \bar{f} \cdot \frac{d\bar{r}}{ds} \quad (10.27)$$

In a similar way, an expression for $d\bar{e}/ds$ can be derived. For that purpose, (10.22-2) is differentiated and we obtain

$$\frac{d\bar{e}}{ds} = \left(\frac{2V}{\mu} \frac{dV}{ds} + \frac{1}{r^2} \frac{dr}{ds} \right) \bar{r} + \left(\frac{V^2}{\mu} - \frac{1}{r} \right) \frac{d\bar{r}}{ds} - \frac{1}{\mu} \left\{ \left(\frac{d\bar{r}}{ds} \cdot \bar{V} + \bar{r} \cdot \frac{d\bar{V}}{ds} \right) \bar{V} + (\bar{r} \cdot \bar{V}) \frac{d\bar{V}}{ds} \right\}$$

Evaluation of this relation, using (10.13) and (10.25), results in

$$\begin{aligned} \frac{d\bar{e}}{ds} &= \left(\frac{2V}{\mu} \frac{dV}{ds} + \frac{1}{r^2} \frac{dr}{ds} \right) \bar{r} + \left\{ \frac{2}{r^2} \left(\frac{dr}{ds} \right)^2 - \frac{1}{r} \right\} \frac{d\bar{r}}{ds} \\ &\quad - \frac{1}{r^2} \left\{ (\bar{r} \cdot \frac{d^2\bar{r}}{ds^2}) \frac{d\bar{r}}{ds} + (\bar{r} \cdot \frac{d\bar{r}}{ds}) \frac{d^2\bar{r}}{ds^2} \right\} \end{aligned} \quad (10.28)$$

From (10.20) and (10.22-2) follows

$$\begin{aligned} \frac{d^2\bar{r}}{ds^2} &= -\frac{\bar{r}}{a} - \bar{e} + \frac{r^2}{\mu} \bar{f} = -\left(\frac{V^2}{\mu} + \frac{1}{a} - \frac{1}{r} \right) \bar{r} + \frac{1}{r} \frac{dr}{ds} \frac{d\bar{r}}{ds} + \frac{r^2}{\mu} \bar{f} \\ \bar{r} \cdot \frac{d^2\bar{r}}{ds^2} &= -r^2 \left(\frac{V^2}{\mu} + \frac{1}{a} - \frac{1}{r} - \frac{1}{r^2} \left(\frac{dr}{ds} \right)^2 \right) + \frac{r^2}{\mu} (\bar{f} \cdot \bar{r}) \end{aligned}$$

From (10.7) follows

$$\frac{V^2}{\mu} + \frac{1}{a} - \frac{1}{r} = \frac{1}{r}$$

which allows the previous relations to be written as

$$\begin{aligned} \frac{d^2\bar{r}}{ds^2} &= -\frac{\bar{r}}{r} + \frac{1}{r} \frac{dr}{ds} \frac{d\bar{r}}{ds} + \frac{r^2}{\mu} \bar{f} \\ \bar{r} \cdot \frac{d^2\bar{r}}{ds^2} &= \left(\left(\frac{dr}{ds} \right)^2 - r \right) + \frac{r^2}{\mu} (\bar{f} \cdot \bar{r}) \end{aligned} \quad (10.29)$$

Substitution of (10.26) and (10.29) into (10.28), finally, leads to

$$\frac{d\bar{e}}{ds} = \frac{2}{\mu} (\bar{f} \cdot \frac{d\bar{r}}{ds}) \bar{r} - \frac{1}{\mu} (\bar{f} \cdot \bar{r}) \frac{d\bar{r}}{ds} - \frac{1}{\mu} (\bar{r} \cdot \frac{d\bar{r}}{ds}) \bar{f} \quad (10.30)$$

When the perturbing acceleration \bar{f} is known, the set of differential equations (10.9), (10.20), (10.27) and (10.30) can be integrated numerically to compute the perturbed orbit; this set is of the order eleven. The equations of motion contain no singularity for $r \downarrow 0$ and therefore form a regularized set. In the following Sections, another classical regularization method is discussed.

10.5. Method of Stiefel for perturbed two-dimensional motion

In this Section, we will first analyze a regularization technique for the motion in a plane, i.e. two-dimensional motion. The analysis is based on the work of E.L. Stiefel (1909-1987) and G. Scheifele (1940-)¹, who followed some basic ideas of T. Levi-Civita, who has studied extensively the regularization of elliptical motion in a plane.

The starting point for this analysis is (10.2) and the introduction of a fictitious time, s , now defined as

$$dt = r ds \quad (10.31)$$

From this definition, we find

$$\begin{aligned} \frac{d\bar{r}}{dt} &= \frac{d\bar{r}}{ds} \frac{ds}{dt} = \frac{1}{r} \frac{d\bar{r}}{ds} \\ \frac{d^2\bar{r}}{dt^2} &= \frac{d}{ds} \left(\frac{1}{r} \frac{d\bar{r}}{ds} \right) \frac{ds}{dt} = -\frac{1}{r^3} \frac{dr}{ds} \frac{d\bar{r}}{ds} + \frac{1}{r^2} \frac{d^2\bar{r}}{ds^2} \end{aligned} \quad (10.32)$$

Substitution of (10.32) into (10.2) yields

$$\frac{d^2\bar{r}}{ds^2} - \frac{1}{r} \frac{dr}{ds} \frac{d\bar{r}}{ds} + \frac{\mu}{r} \bar{r} = r^2 \bar{f}$$

To avoid any confusion, in this Section the position vector in physical space will be indicated by \bar{x} , with components x_1, x_2, x_3 . The scalar distance r is indicated by r or $|\bar{x}|$, where appropriate. In this notation, the equation of motion can be written as

$$\frac{d^2\bar{x}}{ds^2} - \frac{1}{r} \frac{dr}{ds} \frac{d\bar{x}}{ds} + \frac{\mu}{r} \bar{x} = r^2 \bar{f} \quad (10.33)$$

Obviously, this equation is still singular for $r \downarrow 0$.

Levi-Civita noticed that as the eccentricity of an ellipse with a fixed value of the semi-major axis approaches one, the shape of the orbit degenerates into a straight line segment along which the moving body oscillates (Section 10.1). In Figure 10.5 (left) this motion is depicted; x_1 and x_2 are the components of \bar{x} in the plane considered. Note that the position vector makes a sharp bend over an angle 2π at the origin, which corresponds to a singularity for the equation of motion. In order to remove that singularity, Levi-Civita proposed to map the physical x -plane onto a parametric u -plane, in such a way that the image of the moving body passes through the origin

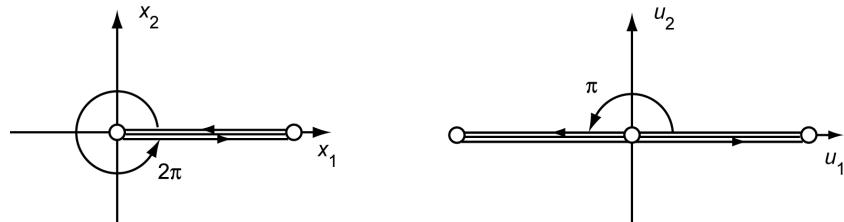


Figure 10.5: Sketch of planar elliptical motion if the eccentricity approaches the value 1 at a fixed value of the semi-major axis, for the physical x -plane (left) and the parametric u -plane (right).

¹ E.L. Stiefel and G. Scheifele, *Linear and regular celestial mechanics*, Springer Verlag, 1971.

after ‘collision’. Hence, the angle 2π in the physical plane must be transformed to the angle π in the parametric plane.

This goal can be achieved by the conformal mapping:

$$\mathbf{u}_1 + i\mathbf{u}_2 = \sqrt{\mathbf{x}_1 + i\mathbf{x}_2} \quad (10.34)$$

where \mathbf{u}_1 and \mathbf{u}_2 are the new coordinates in the parametric plane (Figure 10.5, right) and are the components of the vector $\bar{\mathbf{u}}$. Squaring the transformation relation (10.34) and separating into real and imaginary parts leads to

$$\mathbf{x}_1 = \mathbf{u}_1^2 - \mathbf{u}_2^2 ; \quad \mathbf{x}_2 = 2\mathbf{u}_1 \mathbf{u}_2 \quad (10.35)$$

or, in matrix notation,

$$\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{u}_1 & -\mathbf{u}_2 \\ \mathbf{u}_2 & \mathbf{u}_1 \end{pmatrix} \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{pmatrix} \quad (10.36)$$

Differentiation of (10.35) yields

$$\dot{\mathbf{x}}_1 = 2(\mathbf{u}_1 \dot{\mathbf{u}}_1 - \mathbf{u}_2 \dot{\mathbf{u}}_2) ; \quad \dot{\mathbf{x}}_2 = 2(\mathbf{u}_1 \dot{\mathbf{u}}_2 + \mathbf{u}_2 \dot{\mathbf{u}}_1) \quad (10.37)$$

where the dot-notation indicates in this Section a differentiation with respect to the fictitious time, s . The relations (10.37) can be written in matrix notation as

$$\begin{pmatrix} \dot{\mathbf{x}}_1 \\ \dot{\mathbf{x}}_2 \end{pmatrix} = 2 \begin{pmatrix} \mathbf{u}_1 & -\mathbf{u}_2 \\ \mathbf{u}_2 & \mathbf{u}_1 \end{pmatrix} \begin{pmatrix} \dot{\mathbf{u}}_1 \\ \dot{\mathbf{u}}_2 \end{pmatrix} \quad (10.38)$$

The two-by-two matrix appearing on the right-hand sides of (10.36) and (10.38) is known as the *Levi-Civita matrix*, abbreviated in the following as *LC-matrix*:

$$L(\bar{\mathbf{u}}) = \begin{pmatrix} \mathbf{u}_1 & -\mathbf{u}_2 \\ \mathbf{u}_2 & \mathbf{u}_1 \end{pmatrix} \quad (10.39)$$

Using this matrix, (10.36) and (10.38) can be written as

$$\begin{aligned} \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} &= L(\bar{\mathbf{u}}) \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{pmatrix} \quad \text{or} \quad \bar{\mathbf{x}} = L(\bar{\mathbf{u}}) \bar{\mathbf{u}} \\ \begin{pmatrix} \dot{\mathbf{x}}_1 \\ \dot{\mathbf{x}}_2 \end{pmatrix} &= 2L(\bar{\mathbf{u}}) \begin{pmatrix} \dot{\mathbf{u}}_1 \\ \dot{\mathbf{u}}_2 \end{pmatrix} \quad \text{or} \quad \dot{\bar{\mathbf{x}}} = 2L(\bar{\mathbf{u}}) \dot{\bar{\mathbf{u}}} \end{aligned} \quad (10.40)$$

From (10.35) follows the important relation:

$$r = |\bar{\mathbf{x}}| = \sqrt{\mathbf{x}_1^2 + \mathbf{x}_2^2} = \mathbf{u}_1^2 + \mathbf{u}_2^2 = |\bar{\mathbf{u}}|^2 = (\bar{\mathbf{u}} \cdot \bar{\mathbf{u}}) \quad (10.41)$$

From these expressions we conclude that:

- The mapping of the physical plane onto the parametric plane ((10.34)) results in the square-root of the original distance from the origin.
- That mapping also yields angles at the origin that are halved when compared to the original situation.

Hence, the image of a point (x_1, x_2) is most easily found by taking the square-root of its original distance and by halving the original polar angle.

The *LC*-matrix, $L(\bar{\mathbf{u}})$, has a number of interesting properties. First, it may be stated that the matrix is orthogonal, since

$$L^T(\bar{\mathbf{u}}) L(\bar{\mathbf{u}}) = L(\bar{\mathbf{u}}) L^T(\bar{\mathbf{u}}) = (\mathbf{u}_1^2 + \mathbf{u}_2^2) I = (\bar{\mathbf{u}} \cdot \bar{\mathbf{u}}) I = r I \quad (10.42)$$

where I is the unit matrix (identity matrix). Note that this definition of orthogonality is a generalization of the definition $L^T L = I$ that is usually adopted. For the inverse of $L(\bar{\mathbf{u}})$ we may write

$$L^{-1}(\bar{\mathbf{u}}) L(\bar{\mathbf{u}}) = I \quad (10.43)$$

Combination of (10.42) and (10.43) yields

$$L^{-1}(\bar{\mathbf{u}}) = \frac{1}{r} L^T(\bar{\mathbf{u}}) \quad (10.44)$$

This permits solving $\dot{\bar{\mathbf{u}}}$ from (10.40):

$$\dot{\bar{\mathbf{u}}} = \frac{1}{2} L^{-1}(\bar{\mathbf{u}}) \dot{\bar{\mathbf{x}}} = \frac{1}{2r} L^T(\bar{\mathbf{u}}) \dot{\bar{\mathbf{x}}} \quad (10.45)$$

A second property of the *LC*-matrix is that its elements are linear and homogeneous functions of the parameters u_1, u_2 . Therefore, we may write

$$\frac{d}{ds} \left(L(\bar{\mathbf{u}}) \right) = L \left(\frac{d\bar{\mathbf{u}}}{ds} \right) \quad (10.46)$$

A third property is that the first column of $L(\bar{\mathbf{u}})$ is just the position vector $\bar{\mathbf{u}}$.

For the subsequent analysis we need two rules, which are valid for two arbitrary vectors $\bar{\mathbf{u}}, \bar{\mathbf{v}}$ in the parametric plane. The first one is

$$L(\bar{\mathbf{u}}) \bar{\mathbf{v}} = L(\bar{\mathbf{v}}) \bar{\mathbf{u}} \quad (10.47)$$

where $L(\bar{\mathbf{v}})$ denotes the *LC*-matrix for the components of $\bar{\mathbf{v}}$. This rule may easily be verified, using (10.39). The second rule is

$$(\bar{\mathbf{u}} \cdot \bar{\mathbf{u}}) L(\bar{\mathbf{v}}) \bar{\mathbf{v}} - 2(\bar{\mathbf{u}} \cdot \bar{\mathbf{v}}) L(\bar{\mathbf{u}}) \bar{\mathbf{v}} + (\bar{\mathbf{v}} \cdot \bar{\mathbf{v}}) L(\bar{\mathbf{u}}) \bar{\mathbf{u}} = 0 \quad (10.48)$$

which can be proved as follows. Multiplication of (10.48) with $L^{-1}(\bar{\mathbf{u}})$ and substitution of (10.41) and (10.44) yields

$$[L^T(\bar{\mathbf{u}}) L(\bar{\mathbf{v}}) - 2(\bar{\mathbf{u}} \cdot \bar{\mathbf{v}}) I] \bar{\mathbf{v}} = -(\bar{\mathbf{v}} \cdot \bar{\mathbf{v}}) I \bar{\mathbf{u}} \quad (10.49)$$

After some manipulation, the matrix in brackets turns out to be

$$\begin{pmatrix} -\mathbf{u}_1 \mathbf{v}_1 - \mathbf{u}_2 \mathbf{v}_2 & -\mathbf{u}_1 \mathbf{v}_2 + \mathbf{u}_2 \mathbf{v}_1 \\ -\mathbf{u}_2 \mathbf{v}_1 + \mathbf{u}_1 \mathbf{v}_2 & -\mathbf{u}_2 \mathbf{v}_2 - \mathbf{u}_1 \mathbf{v}_1 \end{pmatrix}$$

which means that the left-hand side of (10.49) becomes

$$\begin{pmatrix} -\mathbf{u}_1 v_1^2 - \mathbf{u}_1 v_2^2 \\ -\mathbf{u}_2 v_1^2 - \mathbf{u}_2 v_2^2 \end{pmatrix} = -(\mathbf{v}_1^2 + \mathbf{v}_2^2) \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{pmatrix} = -(\bar{\mathbf{v}} \cdot \bar{\mathbf{v}}) I \bar{\mathbf{u}}$$

This is identical to the right-hand side of (10.49), which proves the rule (10.48).

We now have all the relations needed to transform the equations of motion (10.33) into the $\bar{\mathbf{u}}$ -plane. Differentiation of the second of equations (10.40) yields

$$\frac{d^2 \bar{x}}{ds^2} = 2L(\bar{\mathbf{u}}) \frac{d^2 \bar{\mathbf{u}}}{ds^2} + 2 \frac{dL(\bar{\mathbf{u}})}{ds} \frac{d\bar{\mathbf{u}}}{ds}$$

or, with (10.46),

$$\frac{d^2 \bar{x}}{ds^2} = 2L(\bar{\mathbf{u}}) \frac{d^2 \bar{\mathbf{u}}}{ds^2} + 2L\left(\frac{d\bar{\mathbf{u}}}{ds}\right) \frac{d\bar{\mathbf{u}}}{ds}$$

Substitution of this relation into (10.33) and using (10.41) leads to

$$2(\bar{\mathbf{u}} \cdot \bar{\mathbf{u}}) L(\bar{\mathbf{u}}) \frac{d^2 \bar{\mathbf{u}}}{ds^2} + 2(\bar{\mathbf{u}} \cdot \bar{\mathbf{u}}) L\left(\frac{d\bar{\mathbf{u}}}{ds}\right) \frac{d\bar{\mathbf{u}}}{ds} - \frac{dr}{ds} \frac{d\bar{x}}{ds} + \mu \bar{x} = (\bar{\mathbf{u}} \cdot \bar{\mathbf{u}})^3 \bar{f}$$

Substitution of both equations (10.40) yields

$$2(\bar{\mathbf{u}} \cdot \bar{\mathbf{u}}) L(\bar{\mathbf{u}}) \frac{d^2 \bar{\mathbf{u}}}{ds^2} + 2(\bar{\mathbf{u}} \cdot \bar{\mathbf{u}}) L\left(\frac{d\bar{\mathbf{u}}}{ds}\right) \frac{d\bar{\mathbf{u}}}{ds} - 2 \frac{dr}{ds} L(\bar{\mathbf{u}}) \frac{d\bar{\mathbf{u}}}{ds} + \mu L(\bar{\mathbf{u}}) \bar{\mathbf{u}} = (\bar{\mathbf{u}} \cdot \bar{\mathbf{u}})^3 \bar{f}$$

From (10.41) we obtain

$$\frac{dr}{ds} = 2(\bar{\mathbf{u}} \cdot \frac{d\bar{\mathbf{u}}}{ds})$$

Substitution of this relation in the previous one leads to

$$\begin{aligned} 2(\bar{\mathbf{u}} \cdot \bar{\mathbf{u}}) L(\bar{\mathbf{u}}) \frac{d^2 \bar{\mathbf{u}}}{ds^2} + 2(\bar{\mathbf{u}} \cdot \bar{\mathbf{u}}) L\left(\frac{d\bar{\mathbf{u}}}{ds}\right) \frac{d\bar{\mathbf{u}}}{ds} - 4(\bar{\mathbf{u}} \cdot \frac{d\bar{\mathbf{u}}}{ds}) L(\bar{\mathbf{u}}) \frac{d\bar{\mathbf{u}}}{ds} \\ + \mu L(\bar{\mathbf{u}}) \bar{\mathbf{u}} = (\bar{\mathbf{u}} \cdot \bar{\mathbf{u}})^3 \bar{f} \end{aligned} \quad (10.50)$$

Equation (10.50) can be simplified by inserting after the first term minus two times the identity (10.48) applied to the vectors $\bar{\mathbf{u}}$ and $d\bar{\mathbf{u}}/ds$. We then find

$$2(\bar{\mathbf{u}} \cdot \bar{\mathbf{u}}) L(\bar{\mathbf{u}}) \frac{d^2 \bar{\mathbf{u}}}{ds^2} - 2\left(\frac{d\bar{\mathbf{u}}}{ds} \cdot \frac{d\bar{\mathbf{u}}}{ds}\right) L(\bar{\mathbf{u}}) \bar{\mathbf{u}} + \mu L(\bar{\mathbf{u}}) \bar{\mathbf{u}} = (\bar{\mathbf{u}} \cdot \bar{\mathbf{u}})^3 \bar{f}$$

Multiplying this relation with $L^{-1}(\bar{\mathbf{u}})$ and using (10.44) reduces it to

$$2(\bar{\mathbf{u}} \cdot \bar{\mathbf{u}}) \frac{d^2 \bar{\mathbf{u}}}{ds^2} + \left[\mu - 2\left(\frac{d\bar{\mathbf{u}}}{ds} \cdot \frac{d\bar{\mathbf{u}}}{ds}\right) \right] \bar{\mathbf{u}} = (\bar{\mathbf{u}} \cdot \bar{\mathbf{u}})^2 L^T(\bar{\mathbf{u}}) \bar{f}$$

Finally, this relation is rewritten as

$$\frac{d^2\bar{\mathbf{u}}}{ds^2} + \frac{\mu}{2} - \left(\frac{d\bar{\mathbf{u}}}{ds} \cdot \frac{d\bar{\mathbf{u}}}{ds} \right) \bar{\mathbf{u}} = \frac{(\bar{\mathbf{u}} \cdot \bar{\mathbf{u}})}{2} L^T(\bar{\mathbf{u}}) \bar{f} \quad (10.51)$$

We will not discuss this result here, but will first prove in Section 10.6 that the same set of differential equations also holds for the motion in three-dimensional space, provided that $\bar{\mathbf{u}}$ is then considered a four-dimensional vector and special measures are taken.

10.6. Method of Stiefel for perturbed three-dimensional motion

The natural approach to extend the method of Stiefel discussed in Section 10.5 to the case of three-dimensional motion would be to construct a square matrix of three rows and columns having the characteristic properties of the two-by-two *LC*-matrix defined by (10.39). However, it was shown by A. Hurwitz (1859-1919) and published in 1933 that the generalization of Levi-Civit  's transformation to three dimensions is not possible, but that the transformation may be extended to four dimensions. P.E. Kustaanheimo (1924-1997) proposed in 1964 to take advantage of the ideas used in the theory of spinors by employing a pair of complex numbers, thus generalizing the single complex variable of Levi-Civit  's theory. Working in real variable analysis, this is equivalent to the introduction of four parameters u_1, u_2, u_3, u_4 . Such an artificial increase of the degrees of freedom does, of course, introduce some difficulties. However, they turn out to be harmless if appropriate measures are taken.

In 1965, Kustaanheimo and Stiefel proposed the following four-by-four matrix:

$$L(\bar{\mathbf{u}}) = \begin{pmatrix} u_1 & -u_2 & -u_3 & u_4 \\ u_2 & u_1 & -u_4 & -u_3 \\ u_3 & u_4 & u_1 & u_2 \\ u_4 & -u_3 & u_2 & -u_1 \end{pmatrix} \quad (10.52)$$

which is now known as the *Kustaanheimo-Stiefel matrix*, or shortly the *KS-matrix*. This matrix, however, already occurs in a letter by L. Euler to Ch. Goldbach (1690-1764) in 1748 by way of a special case of the so-called Euler Identity. Note that the *LC*-matrix appears in the upper-left corner of the *KS*-matrix. Bearing in mind the *LC* transformation, we now define the transformation of a point in four-dimensional u -space to the corresponding point in three-dimensional x -space by

$$\bar{x} = L(\bar{\mathbf{u}}) \bar{\mathbf{u}} \quad (10.53)$$

where the vector \bar{x} is augmented by a fourth component that should vanish; i.e. has the value zero. We will generalize this concept and adopt the convention that any vector in three-dimensional physical space may be supplemented to a four-dimensional vector by adding a fourth component of zero value. The transformation (10.53) is called the *KS-transformation*. Combination of (10.52) and (10.53) yields the scalar relations

$$\begin{aligned} x_1 &= u_1^2 - u_2^2 - u_3^2 + u_4^2 & ; \quad x_2 &= 2(u_1u_2 - u_3u_4) \\ x_3 &= 2(u_1u_3 + u_2u_4) & ; \quad x_4 &= 0 \end{aligned} \quad (10.54)$$

and it is seen that by the appropriate selection of the *KS*-matrix the fourth component of \bar{x} does

indeed automatically vanish.

We may easily verify that also for the *KS*-matrix holds:

$$L^T(\bar{u}) L(\bar{u}) = L(\bar{u}) L^T(\bar{u}) = (\bar{u} \cdot \bar{u}) I = r I \quad (10.55)$$

which means that also the *KS*-matrix is orthogonal. Just as with the *LC*-matrix, we also find for the *KS*-matrix the relation

$$L^{-1}(\bar{u}) = \frac{1}{r} L^T(\bar{u}) \quad (10.44)$$

It will be obvious that the *KS*-matrix also has the additional two properties found earlier for the *LC*-matrix, thus,

$$\frac{d}{ds} (L(\bar{u})) = L\left(\frac{d\bar{u}}{ds}\right) \quad (10.46)$$

and the first column of $L(\bar{u})$ is just the position vector \bar{u} . It is stressed that the two rules (10.47) and (10.48) may not be applied unrestrictedly anymore. Applying the definition (10.52) we find that the first three components of the vectors $L(\bar{u})\bar{v}$ and $L(\bar{v})\bar{u}$ of (10.47) are identical, but that the fourth components have opposite signs. That means that we may establish a theorem:

Theorem 1: When two vectors \bar{u} and \bar{v} satisfy the relation

$$u_4 v_1 - u_3 v_2 + u_2 v_3 - u_1 v_4 = 0 \quad (10.56)$$

then

$$L(\bar{u})\bar{v} = L(\bar{v})\bar{u} \quad (10.47)$$

The relation (10.56) is called the *bilinear relation*. It plays a fundamental role in the method described in this Section.

We now formulate a second theorem.

Theorem 2: If two vectors \bar{u} and \bar{v} satisfy the bilinear relation (10.56), then

$$(\bar{u} \cdot \bar{u}) L(\bar{v})\bar{v} - 2(\bar{u} \cdot \bar{v}) L(\bar{u})\bar{v} + (\bar{v} \cdot \bar{v}) L(\bar{u})\bar{u} = 0 \quad (10.48)$$

This theorem may be proved by writing (10.48) first in its equivalent form (10.49)

$$[L^T(\bar{u}) L(\bar{v}) - 2(\bar{u} \cdot \bar{v}) I]\bar{v} = -(\bar{v} \cdot \bar{v}) I\bar{u}$$

Evaluating the first row of the matrix in brackets and multiplying that with the vector \bar{v} gives

$$-u_1(v_1^2 + v_2^2 + v_3^2 + v_4^2) + 2v_4(u_1v_4 - u_2v_3 + u_3v_2 - u_4v_1)$$

Using (10.56) this yields

$$-(\bar{v} \cdot \bar{v}) u_1$$

which is the first component of the vector on the right-hand side of (10.49). The analysis for the remaining three components yields similar results. We thus conclude that the two rules (10.47) and (10.48) hold both for the motion in a plane and for the motion in space; but in the latter case only if the bilinear relation (10.56) is satisfied.

In the planar case, discussed in Section 10.5, we proceeded by rewriting the equation of motion for the vector \bar{x} into an equation of motion for the vector \bar{u} . However, this direct approach

cannot be followed in the three-dimensional case analyzed in this Section. The reason is that the *KS*-transformation is ambiguous if applied in the direction from \bar{x} to \bar{u} . There is a whole one-dimensional set of vectors \bar{u} corresponding to a given vector \bar{x} . In fact, all points lying on a given circle about the origin in u -space lead to the same points in physical x -space. The only way to proceed therefore is to *postulate* that (10.51) also holds for the three-dimensional case and to verify that the original equations of motion (10.33) are still satisfied. That proof will be given below, using a number of auxiliary mathematical theorems originally stated and proven in the book by Stiefel and Scheifele¹.

However, first some comments are needed about the initial conditions of our differential equations. Assume that the initial values $\bar{x}(0)$ and $\dot{\bar{x}}(0)$ of the functions $\bar{x}(s)$ and $\dot{\bar{x}}(s)$ are given at the instant $s = 0$. To construct the corresponding initial values $\bar{u}(0)$ and $\dot{\bar{u}}(0)$ the following procedure should be followed. Basically, $\bar{u}(0)$ could be chosen arbitrarily among the vectors corresponding to $\bar{x}(0)$ by the *KS*-transformation (10.54). It is recommended, however, to compute $\bar{u}(0)$ from

$$\begin{aligned} u_1^2 + u_4^2 &= \frac{1}{2}(r + x_1) \quad ; \quad u_2 = \frac{x_2 u_1 + x_3 u_4}{r + x_1} \quad ; \quad u_3 = \frac{x_3 u_1 - x_2 u_4}{r + x_1} \\ \text{if } x_1 \geq 0, \text{ else} \end{aligned} \tag{10.57}$$

$$u_2^2 + u_3^2 = \frac{1}{2}(r - x_1) \quad ; \quad u_1 = \frac{x_2 u_2 + x_3 u_3}{r - x_1} \quad ; \quad u_4 = \frac{x_3 u_2 - x_2 u_3}{r - x_1}$$

In the first case, u_1 and u_4 should be chosen in such a way that the first relation is satisfied; in the second case, u_2 and u_3 should be chosen in such a way that the first relation is satisfied. For simplicity, we usually select in the first case either $u_1 = 0$ or $u_1 = u_4$; in the second case either $u_2 = 0$ or $u_2 = u_3$. It is emphasized that if at $s = 0$: $r = 0$, then also $x_1 = x_2 = x_3 = 0$, which means that for this physically very-unrealistic case (10.57) may not be used. Once the vector $\bar{u}(0)$ has been defined, and bearing in mind that we are concerned with a generalization of the two-dimensional Levi-Cevit  situation, we use (10.45) to find $\dot{\bar{u}}(0)$:

$$\dot{\bar{u}}(0) = \frac{1}{2|\bar{u}(0)|^2} L^T(\bar{u}(0)) \dot{\bar{x}}(0) \tag{10.58}$$

As has already been mentioned above, we now postulate that the motion in three-dimensional physical space is described by (10.51), where now, of course, \bar{u} is a four-dimensional vector.

We will now derive a number of relevant theorems.

Theorem 3: Consider any vector \bar{y} in physical space, which is considered as a four-dimensional vector with a fourth component of value zero. Define \bar{v} as

$$\bar{v} = L^T(\bar{u}) \bar{y} \tag{10.59}$$

Then, the vectors \bar{u} and \bar{v} satisfy the bilinear relation (10.56). This theorem may be proved by multiplying (10.59) by $L(\bar{u})$:

$$L(\bar{u}) \bar{v} = L(\bar{u}) L^T(\bar{u}) \bar{y} = (\bar{u} \cdot \bar{u}) \bar{y} \tag{10.60}$$

The fourth component of the quantity $L(\bar{u}) \bar{v}$ is given by

$$u_4 v_1 - u_3 v_2 + u_2 v_3 - u_1 v_4$$

which is the left-hand side of (10.56). Because the fourth component of \bar{y} in (10.60) is zero, the bilinear relation is satisfied.

Theorem 4: The initial vectors $\bar{u}(0)$ and $\dot{\bar{u}}(0)$ satisfy the bilinear relation (10.56):

$$u_4(0)\dot{u}_1(0) - u_3(0)\dot{u}_2(0) + u_2(0)\dot{u}_3(0) - u_1(0)\dot{u}_4(0) = 0 \quad (10.61)$$

To prove this theorem we apply Theorem 3 to $\bar{y} = \dot{\bar{x}}(0)$, $\bar{u} = \bar{u}(0)$ and $\bar{v} = \dot{\bar{u}}(0)$:

$$\dot{\bar{u}}(0) = L^T(\bar{u}(0))\dot{\bar{x}}(0)$$

Multiplication with $L(\bar{u}(0))$ leads to

$$L(\bar{u}(0))\dot{\bar{u}}(0) = (\bar{u}(0) \cdot \bar{u}(0))\dot{\bar{x}}(0)$$

Evaluating the fourth component of the left-hand side of this equation, and realizing that $\dot{x}_4(0) = 0$, yields (10.61) and proves the theorem.

Theorem 5: The right-hand side of (10.51):

$$\bar{q} = \frac{(\bar{u} \cdot \bar{u})}{2} L^T(\bar{u})\bar{f} \quad (10.62)$$

satisfies the bilinear relation

$$u_4 q_1 - u_3 q_2 + u_2 q_3 - u_1 q_4 = 0 \quad (10.63)$$

To prove this theorem, we start again from (10.59) and use $\bar{y} = \bar{f}$ and $\bar{v} = \bar{q}$:

$$\bar{q} = L^T(\bar{u})\bar{f}$$

Multiplication with $L(\bar{u})$ yields

$$L(\bar{u})\bar{q} = (\bar{u} \cdot \bar{u})\bar{f}$$

Evaluating the fourth component of the left-hand side of this equation and realizing that $f_4 = 0$ yields (10.63).

We now proceed with the set of differential equations (10.51), where we write the right-hand side according to (10.62) as

$$\frac{d^2\bar{u}}{ds^2} + \frac{\frac{\mu}{2} - \left(\frac{d\bar{u}}{ds} \cdot \frac{d\bar{u}}{ds}\right)}{(\bar{u} \cdot \bar{u})} \bar{u} = \bar{q} \quad (10.64)$$

We also write

$$g(\bar{u}, \dot{\bar{u}}) = u_4\dot{u}_1 - u_3\dot{u}_2 + u_2\dot{u}_3 - u_1\dot{u}_4 \quad (10.65)$$

for the left-hand side of the bilinear relation for $\bar{u}, \dot{\bar{u}}$. Again, we derive a number of relevant theorems.

Theorem 6: The function $g(\bar{u}, \dot{\bar{u}})$ is a first integral of the differential equation (10.64). This means that $g(\bar{u}, \dot{\bar{u}})$ is constant, provided any solution of (10.64) is inserted. To proof this theorem, we differentiate (10.65):

$$\frac{dg}{ds} = u_4 \ddot{u}_1 - u_3 \ddot{u}_2 + u_2 \ddot{u}_3 - u_1 \ddot{u}_4$$

We now insert (10.64) and find

$$\frac{dg}{ds} = u_4 q_1 - u_3 q_2 + u_2 q_3 - u_1 q_4$$

The right-hand side of this equation is according to (10.63) equal to zero, which means that g is constant.

Theorem 7: The solution $\bar{u}(s)$ of (10.64), which is subjected to the initial conditions $\bar{u}(0), \dot{\bar{u}}(0)$ stated above, satisfies the bilinear relation

$$u_4(s) \dot{u}_1(s) - u_3(s) \dot{u}_2(s) + u_2(s) \dot{u}_3(s) - u_1(s) \dot{u}_4(s) = 0 \quad (10.66)$$

where s takes an arbitrary value. The proof of this theorem is easy. Theorem 6 states that the function g is constant and therefore takes anywhere the same value as for $s = 0$. That value is zero according to Theorem 4. An important consequence of this result is that Theorems 1 and 2 are applicable to the vectors $\bar{u}(s)$ and $\dot{\bar{u}}(s)$. In particular, it follows from the definition of the KS-transformation (10.53) and from Theorem 1 that

$$\dot{\bar{x}} = L(\dot{\bar{u}}) \bar{u} + L(u) \dot{\bar{u}} = 2L(\bar{u}) \dot{\bar{u}}$$

showing that the Levi-Cevit   relation (10.40-2) also holds in three-dimensional space.

Theorem 8: Let $\bar{u}(s)$ be the solution of (10.64). Its KS-transform (10.53) satisfies the original equations of motion (10.31) and (10.33) as well as the given initial conditions $\bar{x}(0)$ and $\dot{\bar{x}}(0)$. To proof this theorem, we start with differentiating (10.41) and (10.40-2):

$$\dot{r} = 2(\bar{u} \cdot \dot{\bar{u}}) ; \quad \ddot{\bar{x}} = 2L(\bar{u}) \ddot{\bar{u}} + 2L(\dot{\bar{u}}) \dot{\bar{u}} \quad (10.67)$$

where (10.46) has been used. After multiplying the left- and right-hand side of (10.33) with r , the left-hand side may be written as

$$\bar{m} = r \ddot{\bar{x}} - \dot{r} \dot{\bar{x}} + \mu \bar{x}$$

Substitution of (10.41), (10.53), (10.40-2) and (10.67) leads to

$$\bar{m} = 2(\bar{u} \cdot \dot{\bar{u}}) L(\bar{u}) \ddot{\bar{u}} + 2(\bar{u} \cdot \dot{\bar{u}}) L(\dot{\bar{u}}) \dot{\bar{u}} - 4(\bar{u} \cdot \dot{\bar{u}}) L(\bar{u}) \dot{\bar{u}} + \mu L(\bar{u}) \bar{u} \quad (10.68)$$

Applying (10.48) to \bar{u} and $\dot{\bar{u}}$ gives

$$(\bar{u} \cdot \dot{\bar{u}}) L(\dot{\bar{u}}) \dot{\bar{u}} - 2(\bar{u} \cdot \dot{\bar{u}}) L(\bar{u}) \dot{\bar{u}} + (\dot{\bar{u}} \cdot \dot{\bar{u}}) L(\bar{u}) \bar{u} = 0$$

Substituting this relation into (10.68) yields

$$\bar{m} = L(\bar{u}) [2(\bar{u} \cdot \dot{\bar{u}}) \ddot{\bar{u}} - 2(\dot{\bar{u}} \cdot \dot{\bar{u}}) \bar{u} + \mu \bar{u}] \quad (10.69)$$

From (10.64) we obtain

$$(\bar{u} \cdot \dot{\bar{u}}) \ddot{\bar{u}} - (\dot{\bar{u}} \cdot \dot{\bar{u}}) \bar{u} + \frac{1}{\mu} \bar{u} \bar{u} = (\bar{u} \cdot \dot{\bar{u}}) \bar{q}$$

Substituting this relation into (10.69) gives

$$\bar{m} = 2(\bar{u} \cdot \bar{u}) L(\bar{u}) \bar{q}$$

or, with (10.62),

$$\bar{m} = (\bar{u} \cdot \bar{u})^2 L(\bar{u}) L^T(\bar{u}) \bar{f}$$

or, using the orthogonality of the $L(\bar{u})$ matrix (10.55),

$$\bar{m} = r^3 \bar{f}$$

which is just the right-hand side of the equations of motion (10.33) after multiplication with $(\bar{u} \cdot \bar{u})$. So, the first part of Theorem 8 has been proven.

Concerning the second part of Theorem 8, i.e. that the *KS*-transform (10.53) satisfies the initial conditions $\dot{\bar{x}}(0)$ and $\ddot{\bar{x}}(0)$, we note that from (10.40-2) follows

$$\dot{\bar{x}}(0) = 2L(\bar{u}(0)) \dot{\bar{u}}(0)$$

or, with (10.58),

$$\dot{\bar{x}}(0) = \frac{1}{|\bar{u}(0)|^2} L^T(\bar{u}(0)) L(\bar{u}(0)) \dot{\bar{u}}(0)$$

or, using the orthogonality of $L(\bar{u})$ again,

$$\dot{\bar{x}}(0) = \dot{\bar{u}}(0)$$

With this final theorem we have proved that the differential equations (10.51) for the parameters u_1, u_2, u_3, u_4 , together with starting rule (10.58), fully describe the three-dimensional motion of the body.

It is emphasized that (10.51) still contains a singularity for $\bar{u} \downarrow 0$. To eliminate this singularity, we introduce the total Keplerian energy per unit of mass:

$$\mathcal{E} = \frac{1}{2} \frac{d\bar{x}}{dt} \cdot \frac{d\bar{x}}{dt} - \frac{\mu}{|\bar{x}|} \quad (10.70)$$

Because a perturbing force acts on the body, the total energy, of course, is not constant but varies with time. In the parametric space the energy relation reads

$$\mathcal{E} = \frac{1}{2} \left(\frac{d\bar{x}}{ds} \frac{ds}{dt} \right) \cdot \left(\frac{d\bar{x}}{ds} \frac{ds}{dt} \right) - \frac{\mu}{|\bar{x}|}$$

or, with (10.31) and (10.40-2),

$$\mathcal{E} = \frac{2}{(\bar{u} \cdot \bar{u})^2} \left[(L(\bar{u}) \frac{d\bar{u}}{ds}) \cdot (L(\bar{u}) \frac{d\bar{u}}{ds}) \right] - \frac{\mu}{(\bar{u} \cdot \bar{u})}$$

The term in brackets may be written as

$$\left(L(\bar{u}) \frac{d\bar{u}}{ds} \right)^T \left(L(\bar{u}) \frac{d\bar{u}}{ds} \right) = \left(\frac{d\bar{u}}{ds} \right)^T L^T(\bar{u}) L(\bar{u}) \frac{d\bar{u}}{ds} = \frac{d\bar{u}}{ds} \cdot \left(L^T(\bar{u}) L(\bar{u}) \frac{d\bar{u}}{ds} \right)$$

Because $L(\bar{u})$ is orthogonal, we may write for this expression according to (10.55):

$$(\bar{u} \cdot \bar{u}) \left(\frac{d\bar{u}}{ds} \cdot \frac{d\bar{u}}{ds} \right)$$

and we find for the energy per unit of mass

$$\mathcal{E} = \frac{2}{(\bar{u} \cdot \bar{u})} \left[\left(\frac{d\bar{u}}{ds} \cdot \frac{d\bar{u}}{ds} \right) - \frac{\mu}{2} \right] \quad (10.71)$$

Substitution of (10.71) into (10.51) yields

$$\frac{d^2\bar{u}}{ds^2} - \frac{1}{2}\mathcal{E}\bar{u} = \frac{(\bar{u} \cdot \bar{u})}{2} L^T(\bar{u})\bar{f} \quad (10.72-1)$$

Using the Keplerian semi-major axis:

$$a = -\frac{\mu}{2\mathcal{E}} \quad (10.73)$$

we may also write (10.72-1) as

$$\frac{d^2\bar{u}}{ds^2} + \frac{\mu}{4a}\bar{u} = \frac{(\bar{u} \cdot \bar{u})}{2} L^T(\bar{u})\bar{f} \quad (10.72-2)$$

where one should realize that the semi-major axis is mathematically defined by (10.73) and therefore is strictly related to the definition of the potential energy (per unit of mass) by $-\mu/r$.

Because \mathcal{E} and a vary with time, the numerical integration of (10.72) requires an additional equation for the time variation of \mathcal{E} or a . To find such an equation, we start from (10.27), which reads, according to (10.40-2), in parametric u -space:

$$\frac{d}{ds} \left(\frac{1}{a} \right) = -\frac{2}{\mu} \bar{f} \cdot \left(2L(\bar{u}) \frac{d\bar{u}}{ds} \right)$$

This relation may also be written as

$$\frac{d}{ds} \left(\frac{1}{a} \right) = -\frac{4}{\mu} \left(L(\bar{u}) \frac{d\bar{u}}{ds} \cdot \bar{f} \right) = -\frac{4}{\mu} \left(L(\bar{u}) \frac{d\bar{u}}{ds} \right)^T \bar{f} = -\frac{4}{\mu} \left(\frac{d\bar{u}}{ds} \right)^T L^T(\bar{u}) \bar{f}$$

or

$$\frac{d}{ds} \left(\frac{1}{a} \right) = -\frac{4}{\mu} \frac{d\bar{u}}{ds} \cdot \left(L^T(\bar{u}) \bar{f} \right) = -\frac{4}{\mu} \left(L^T(\bar{u}) \bar{f} \cdot \frac{d\bar{u}}{ds} \right) \quad (10.74-1)$$

From (10.73) we find after differentiation

$$-\frac{2}{\mu} \frac{d\mathcal{E}}{ds} = \frac{d}{ds} \left(\frac{1}{a} \right)$$

Substitution of (10.74-1) yields

$$\frac{d\mathcal{E}}{ds} = 2 \left(L^T(\bar{u}) \bar{f} \cdot \frac{d\bar{u}}{ds} \right) \quad (10.74-2)$$

Finally, we write (10.31), using (10.41), as

$$\frac{dt}{ds} = (\bar{\mathbf{u}} \cdot \bar{\mathbf{u}}) \quad (10.75)$$

We now have all equations available to integrate the regularized equations of motion numerically.

Note that (10.72-1) and (10.72-2) remain regular when $\bar{\mathbf{u}} \downarrow \mathbf{0}$, provided that the perturbing force remains finite. So, with this method we have achieved the goal of regularization. In case there is no perturbing force ($\bar{\mathbf{f}} = \mathbf{0}$), then the equations (10.72-1) and (10.72-2) are reduced to the equations of a harmonic oscillator centered at $\bar{\mathbf{u}} = \mathbf{0}$:

$$\frac{d^2\bar{\mathbf{u}}}{ds^2} - \frac{\mathcal{E}}{2}\bar{\mathbf{u}} = \mathbf{0} \quad ; \quad \frac{d^2\bar{\mathbf{u}}}{ds^2} + \frac{\mu}{4a}\bar{\mathbf{u}} = \mathbf{0} \quad (10.76)$$

where \mathcal{E} and a are constants. The constant frequency of the harmonic motion is expressed by $\sqrt{-\mathcal{E}/2} = \sqrt{\mu/4a}$. The equations (10.76) are stable in the sense of Liapunov, in contrast to the original equations of motion (10.1).

For the method of Burdet (Section 10.2) we have derived a differential equation for the distance r . We can do the same for the method of Stiefel and start by differentiating the first of (10.67):

$$\frac{d^2r}{ds^2} = 2(\bar{\mathbf{u}} \cdot \frac{d^2\bar{\mathbf{u}}}{ds^2}) + 2(\frac{d\bar{\mathbf{u}}}{ds} \cdot \frac{d\bar{\mathbf{u}}}{ds})$$

Substitution of (10.72-1) into this relation leads to

$$\frac{d^2r}{ds^2} = \mathcal{E}(\bar{\mathbf{u}} \cdot \bar{\mathbf{u}}) + (\bar{\mathbf{u}} \cdot \bar{\mathbf{u}}) \left(L^T(\bar{\mathbf{u}}) \bar{\mathbf{f}} \right) \cdot \bar{\mathbf{u}} + 2(\frac{d\bar{\mathbf{u}}}{ds} \cdot \frac{d\bar{\mathbf{u}}}{ds})$$

Substitution of (10.41) and (10.71) yields

$$\frac{d^2r}{ds^2} - 2\mathcal{E}r = \mu + r \left(L^T(\bar{\mathbf{u}}) \bar{\mathbf{f}} \right) \cdot \bar{\mathbf{u}} \quad (10.77)$$

Also this equation remains regular when $r \downarrow 0$, provided that the perturbing force remains finite. In case $\bar{\mathbf{f}} = \mathbf{0}$, a is constant and (10.77) reduces to

$$\frac{1}{\mu} \frac{d^2r}{ds^2} + \frac{r}{a} = 1 \quad (10.78)$$

This equation is similar to (10.14), which was obtained with the method of Burdet (Section 10.2). The only difference is the appearance of the factor $1/\mu$ in the first term on the left-hand side of (10.78), which is a direct result of a different definition of the fictitious time.

After the regularized equations of motion have been integrated until the prescribed end of the integration interval, a transformation from u_1, u_2, u_3, u_4, s to x_1, x_2, x_3, t has to be performed. For the transformation from the parametric u -space to the physical x -space, we use (10.52), (10.53) and (10.40-2). The transformation of the fictitious time, s , to the physical time, t , requires the integration of (10.31). It is emphasized that, as was already stated in Section 10.2, the solution of (10.31) is unstable, which means that any error in the radius eventually yields arbitrarily large errors in time. The full equations (10.72-1) and (10.72-2) are of the type of a perturbed harmonic oscillator with slowly-varying frequency, if the perturbing force is small. As a final remark, it is noticed that the total set of equations that has to be integrated numerically consists of the four

second-order differential equations (10.72-1) or (10.72-2), the first-order differential equation (10.74-1) or (10.74-2), and the first-order differential equation (10.75). So, the order of the system of differential equations is ten, which is one less than in the Method of Burdet (Section 10.2), but four more than of the original set of equations in the physical space x . In practice, it is seldom necessary to carry out the KS -transformation during the numerical integration. Thus, the increase of computational labor produced by expressing the differential equations in four-dimensional u -space should not be overestimated; it is also in most cases compensated by the fact that these equations permit a larger stepsize during numerical integration.

In case the perturbing forces can be described partially by a perturbing potential, R , and partially by a perturbing acceleration, \bar{f} , basically the same regularization scheme may be applied. In that case, the original equations of motion ((10.33)) read

$$\frac{d^2\bar{x}}{ds^2} - \frac{1}{r} \frac{dr}{ds} \frac{d\bar{x}}{ds} + \frac{\mu}{r} \bar{x} = r^2 \left(-\frac{\partial R}{\partial \bar{x}} + \bar{f} \right) \quad (10.79)$$

where $\partial R/\partial \bar{x}$ is the gradient of R . Repeating the derivation given in Sections (10.5) and (10.6) we then find for the regularized equations of motion ((10.72-1) and (10.74-2)):

$$\begin{aligned} \frac{d^2\bar{u}}{ds^2} - \frac{1}{2} \mathcal{E} \bar{u} &= \frac{(\bar{u} \cdot \bar{u})}{2} \left[-\frac{1}{2} \frac{\partial R}{\partial \bar{u}} + L^T(\bar{u}) \bar{f} \right] \\ \frac{d\mathcal{E}}{ds} &= -\left(\frac{\partial R}{\partial \bar{u}} \cdot \frac{d\bar{u}}{ds} \right) + 2 \left(L^T(\bar{u}) \bar{f} \cdot \frac{d\bar{u}}{ds} \right) \end{aligned} \quad (10.80)$$

and similar relations when \mathcal{E} is replaced by a in the regularized equations of motion. In case the perturbing potential is conservative, i.e. not explicitly depending on time, and no additional perturbing force is present ($\bar{f} = 0$), (10.80) shows that the energy \mathcal{E} is constant and consequently the frequency of the oscillator (10.76) is constant. This is an important property of the equations, e.g. in situations where the orbit is computed under the assumption that only perturbing forces due to the so-called *zonal harmonics* of the Earth's gravity field are acting on the satellite (Section 21.2 and Chapter 23).

11. REFERENCE FRAMES, COORDINATES, TIME AND ORBITAL ELEMENTS

For determining the position and velocity of a spacecraft at a certain moment of time, we require a reference frame with respect to which position and velocity are described, an appropriately chosen set of coordinates and a system for time measurement. For computing the spacecraft's motion from measurements taken by instruments on Earth, in addition we need a model describing the rotational motion of the Earth relative to the adopted reference frame, and for very-precise analyses even a model for the horizontal and vertical motions of the Earth's surface. In this context, it is emphasized that, formally, one should distinguish between a 'reference system' and a 'reference frame'. A *reference system* then is the complete specification of how a celestial coordinate system is to be formed. Both the origin and the orientation of the fundamental (reference) planes (or axes) are defined. A reference system also incorporates a specification of the fundamental models needed to construct the system; that is, the basis for the algorithms used to transform between observable quantities and reference data in the system. A *reference frame*, on the other hand, consists of a set of identifiable fiducial points on the sky along with their coordinates, which serves as the practical realization of a reference system. In this book, we will, however, not use this precise terminology, and we simply use the general phrase 'reference frame' to denote a coordinate system; only in Section 11.2 the difference between a reference frame and a reference system will be addressed again.

The classical concepts and systems applied to define reference frames and time were developed since the eighteenth century on basis of a consistent system of definitions. The last decades, major efforts were done to fine-tune the underlying definitions and concepts, and to operationalize the adopted reference frames and time systems. This has led to the introduction of quite complicated new concepts and definitions¹. These are, however, beyond the scope of this book and we will therefore limit ourselves to those 'classical' definitions and concepts that are fundamental to understand the basic aspects of reference frames, Earth's rotation and time systems applied in astrodynamics.

11.1. Position on the Earth's surface

To define a position relative to the Earth's surface, one usually takes a reference frame that is based on the Earth's axis of rotation. The points where this axis crosses the Earth's surface are called the *north pole* and the *south pole*. For simplicity, we will first assume that the shape of the Earth is exactly spherical. Then, the great circle on the Earth's surface halfway between the poles is the Earth's *equator*. The great circles passing through the poles are called *meridians*; they intersect the equator at right angles. By convention, we take a meridian to be bisected by the Earth's equatorial plane into an upper and lower branch; the upper branch containing an observer on the northern hemisphere. Henceforth, with a meridian we will always mean its upper branch. Through any point P (Figure 11.1) on the Earth's surface, the poles excepted, we can imagine just one meridian. The *geographic longitude*, Λ , of that point is defined as the arc length measured along the equator from a reference meridian, called the *prime meridian*, to the meridian passing through point P . Historically, various prime meridians have been used. In October 1884, the meridian passing through the Royal Observatory at Greenwich, England, indicated in Figure 11.1

¹ G.H. Kaplan, *The IAU Resolutions on Astronomical Reference Systems, Time Scales and Earth Rotation Models*, U.S. Naval Observatory, Circular No. 179, Washington, D.C., October 2005.

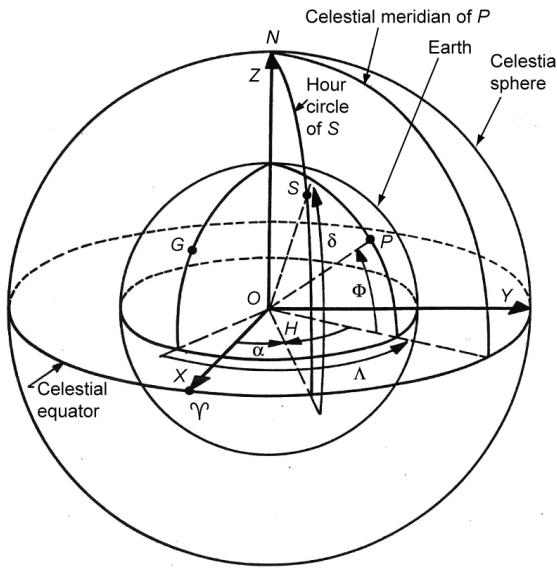


Figure 11.1: Various concepts used to describe positions on Earth and on the celestial sphere.

by G , was selected at a conference as the official prime meridian due to its popularity; this meridian was established by G.B. Airy (1801-1892) in 1851. However, French maps continued to use the Paris prime meridian for several decades. Presently, the *International Earth Rotation and Reference Systems Service* (IERS) maintains the *International Reference Meridian*, that is often (also in this book) still called the *Greenwich meridian*, but that actually passes Airy's original transit circle about 100 m east. This International Reference Meridian is the one applied for all present-day analyses and operational navigation purposes. Longitudes are measured either to the east or west of the Greenwich meridian from 0° to 180° , indicated by E or W. It is more convenient, however, when longitude is measured positively to the east of Greenwich and negatively to the west; in that case the indication W or E can be dropped. The *geocentric latitude*, ϕ , of P is given by the geocentric angle measured along its meridian, from the equator to P . Latitudes are measured either to the north or to the south of the equator from 0° to 90° ; positively to the north and negatively to the south of the equator.

Although we have assumed that the Earth is a perfect sphere, we know that this is not the case and that a better approximation of the Earth's shape is an ellipsoid of revolution (oblate spheroid); i.e. an ellipse rotated about its minor axis, where this minor axis coincides with the Earth's rotation axis. The best approximation is obtained when the lengths of the minor and major axes of this ellipsoid are selected such that deviations from the actual Earth's surface are minimized. We then arrive at the *standard ellipsoid* with a semi-major axis, a_e , (equatorial radius) of 6378.137 km and a semi-minor axis, a_p , (polar radius) of 6356.752 km. The parameter a_e defines the Earth's mean equatorial radius, which acts as a scaling factor in many geodetic and geophysical models. An important parameter related to the shape of this ellipsoid of revolution is the *flattening* (also called *oblateness*), denoted by f and defined as $f = (a_e - a_p)/a_e$. This parameter, which is usually given by its reciprocal value, is found to be approximately equal to 1/298.256. It is emphasized that the *geoid*, which is an equipotential surface accounting for the Earth's gravity field and for the Earth's rotational motion about its axis, in first approximation also can be described by an ellipsoid of revolution with its minor axis along the Earth's rotation axis. This correlation between the shape of the Earth and the shape of the geoid is to a large extent due to the fact that early in its history the Earth was more elastic than today.

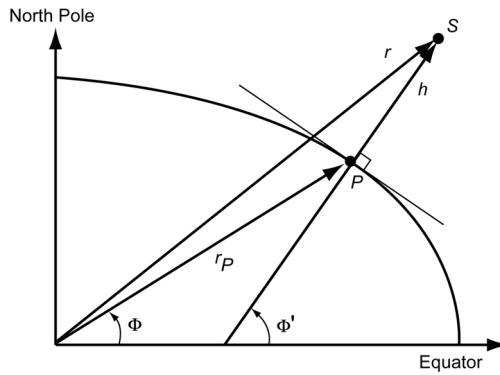


Figure 11.2: Geocentric and geodetic latitude, and altitude.

The deviation of the standard ellipsoid from a pure sphere has a severe consequence for the determination of the latitude of a place on Earth and requires the introduction of the concept of *geodetic latitude*. In addition, we also know *astronomical latitude*. The geocentric latitude, ϕ , of a point P on the surface of the standard ellipsoid is defined as the angle measured perpendicular to the equatorial plane between the equator and a line connecting the geometric center of the coordinate system with that point P . The geodetic latitude, ϕ' , of P (Figure 11.2) is defined as the angle measured perpendicular to the equatorial plane, between a line normal to the tangent plane touching the reference ellipsoid at point P and the equatorial plane. It may differ up to an amount of about $12'$ from the geocentric latitude at that place (Section 11.6). The astronomical latitude, ϕ_a , finally, is defined as the angle measured perpendicular to the equatorial plane, between the intersection of the local gravity vector (plumb line) at point P and the equatorial plane. Since this latitude is a function of the local gravity field, it is affected by the presence of any gravity anomaly due to mountains, seas, etc. Differences between geodetic and astronomical latitude are termed *station error*; differences of $5''$ are not uncommon, and they occasionally rise to $40''$. The flattening of the Earth also affects the measurement of the altitude, h (Figure 11.2). Since altitude is measured perpendicular to the surface, the altitude above a point on the ellipsoid is, in general, not measured along the geocentric radius vector to that point, as it is for a spherical Earth.

An even better approximation of the shape of the Earth's surface may be obtained by superimposing small corrections on the ellipsoid such that a *pear shape* in north-south direction and a slightly elliptical equator results. Then, the north polar radius is about 38 m longer than the south polar radius. The major axis of the equatorial ellipse intersects the Earth's surface at longitudes of about 15° W and 165° E, and is about 138 m longer than the minor axis of the equatorial ellipse.

11.2. Astronomical concepts

In denoting angular positions of celestial objects, it is convenient to use the concept of a fictitious *celestial sphere*. This is a sphere with an infinitely large radius, centered at an observer on Earth or at the mass center of the Earth. The remote stars appear to be set on the inner surface of this sphere. It will be clear that only for nearby objects, like spacecraft, different angular positions will be measured on the observer-centered celestial sphere and on the Earth-centered celestial sphere. For astrodynamical purposes we will mostly use the latter type; this situation is depicted in Figure 11.1. The *celestial poles* are defined as the points where the Earth's rotation axis intersects the celestial sphere; the north celestial pole corresponds to the Earth's north pole. This

celestial north pole presently lies close to a moderately bright star, α Ursae Minoris, called the *pole star* or *Polaris*. The intersection of the Earth's equatorial plane with the celestial sphere is the *celestial equator*. Great circles on the celestial sphere passing through the celestial poles are called *hour circles* (Figure 11.1). Again, we consider an hour circle to be limited to its upper branch.

When the meridian of an observer on Earth is projected onto the celestial sphere, we speak of the observer's *celestial meridian*. This celestial meridian, of course, passes through the celestial poles and through a point directly above the observer: his *zenith*. Owing to the Earth's rotation, the observer's celestial meridian continuously sweeps around the celestial sphere. To put it another way, the observer watches the celestial sphere rotating about the Earth's spin axis in a westward direction; celestial objects thereby passing the observer's stationary celestial meridian. This is the so-called *diurnal motion*.

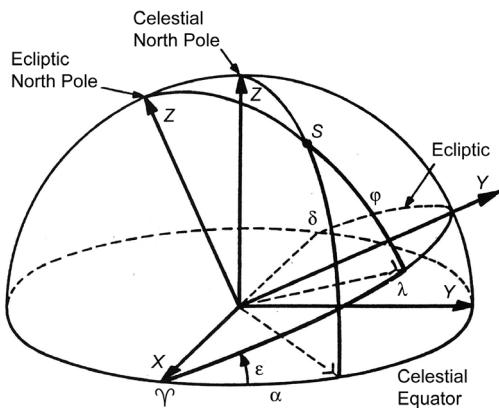


Figure 11.3: Geocentric equatorial and ecliptic systems of coordinates.

If the Sun is observed from the Earth, it is found to possess a second motion in addition to its diurnal motion. The Sun moves eastward among the stars at a rate of about 1° per day, returning to its original position on the celestial sphere in one year. The path of the Sun over the celestial sphere is called the *ecliptic*. The band of stellar constellations around this great circle on the celestial sphere is called the *zodiac*. It is important to realize that seen from the Sun, the ecliptic is nothing but the intersection of the plane of the Earth's orbit about the Sun with the celestial sphere². The ecliptic plane is inclined to the equatorial plane at an angle ϵ , referred to as the *obliquity of the ecliptic* (Figure 11.3). At the present time $\epsilon \approx 23^\circ 27'$. The *axis of the ecliptic*, being the line through the center of the celestial sphere perpendicular to the ecliptic, intersects the celestial sphere in the *ecliptic poles*. The angular distance between the celestial north pole and the ecliptic north pole equals the angle ϵ . Great circles on the celestial sphere through the ecliptic poles are called *circles of celestial longitude*. As we did with meridians and hour circles, we will only consider the upper branch of these great circles. The intersecting line of the equatorial plane and the ecliptic plane (Figure 11.3) plays a fundamental role in the definition of reference frames. This line is called the *equinox line* because when the Sun crosses this line, the Earth's axis of rotation is at right angles to the Sun-Earth line and, consequently, day and night have equal length, everywhere on Earth. This is the case around March 20/21 and September 22/23 each year. These crossing points are therefore called the *vernal equinox* and the *autumnal equinox*, respectively. They are also referred to as *first point of Aries*, denoted by Υ , and *Libra*, denoted

² A more precise definition is that the ecliptic is the mean plane of the orbit of the Earth-Moon barycenter around the barycenter of the solar system.

by Ω , respectively. From astronomical observations and the laws of celestial mechanics, we can determine at any moment of time the position of these points relative to the stars. The times of maximum ($\approx 23.4^\circ$) and minimum ($\approx -23.4^\circ$) solar declination are called *solstices*. This happens around June 20/21 and December 21/22 each year, and therefore the points are called the *summer solstice* and *winter solstice*, respectively. As an example, in 2013 the equinoxes and solstices happened on March 20, 11^h02^m UTC; June 21, 05^h04^m UTC; September 22, 20^h44^m UTC; and December 21, 17^h11^m UTC³. Note that the designations vernal and autumnal equinox, and summer and winter solstice, are somewhat misleading, since for an observer in the southern hemisphere autumn starts when the Sun crosses the vernal equinox and it is mid-winter when the Sun is at its summer solstice!

The point Υ is actually not fixed between the stars on the celestial sphere, but moves slowly across the sky because the celestial equator and the ecliptic are continuously in motion. Below, some major aspects of this motion will be described briefly.

The motion of the equator (or the celestial pole), is primarily due to the gravitational attraction by the Sun and the Moon on the spinning, oblate and elastic Earth. It consists of two components: *luni-solar precession* and *luni-solar nutation*. Precession is the smooth long-period motion of the mean celestial north pole around the ecliptic north pole; the mean celestial pole traces out a circle, about 23.4° in radius, centered on the ecliptic pole, taking about 25,800 years to complete one circuit. Precession does not change the obliquity of the ecliptic and causes the point Υ to move along the ecliptic at a rate of $-50''$ per year; so, in a westward direction. The discovery of precession is usually attributed to Hipparchos (~190-120 B.C.). Nutation is a relatively short-period motion of the true celestial pole around the mean celestial pole. The main nutation component is due to the rotation of the Moon's orbital plane about an axis perpendicular to the ecliptic (Section 17.1); this component traces out an ellipse with major and minor axes of about $18''$ and $14''$, respectively, about the mean celestial pole every 18.60 years. As a consequence, the obliquity of the ecliptic is varying with an amplitude of $9''$ and the point Υ is undergoing an oscillation with an amplitude of roughly $17''$ along the ecliptic. Many additional components are present, such as a semi-annual periodic component with an amplitude of $1.3''$ and a semi-monthly component with an amplitude of $0.2''$.

The motion of the ecliptic relative to inertial space is due to the gravitational attraction of the planets on the Earth, which results in a slow rotation of the ecliptic about a slowly-moving diameter; this motion is known as *planetary precession* and causes a precession of the point Υ of about $12''$ per century along the celestial equator and a decrease of the obliquity of the ecliptic of about $47''$ per century.

Luni-solar precession is about 420 times larger than planetary precession. Since, in addition to the Sun and the Moon, the gravitational attraction by the planets also cause a small movement of the Earth's axis of rotation in inertial space, making the contrast in the terms luni-solar versus planetary somewhat misleading. Therefore, the International Astronomical Union (IAU) recommended in 2006 that the dominant component be renamed *precession of the equator* and the minor component be renamed *precession of the ecliptic*; their combination is named *general precession*.

Finally, the stars are only at a finite distance, showing some of their own motion on the celestial sphere. The First Point of Aries, as its name suggests, was some 2000 years ago located in the constellation Aries (Ram). Presently, it is located in Pisces (Fish) and it is moving into the

³ The concept of UTC (Universal Time Coordinated) will be defined in Section 11.4.

direction of the constellation Aquarius (Water Carrier). That is called in poems and songs the ‘dawn of the age of Aquarius’. The exact moment when it will enter the constellation Aquarius depends on where we draw the boundaries of the star constellations.

For the definition of a reference frame, the equatorial plane, ecliptic plane and vernal equinox play a fundamental role. Because of all motions described above, we have to specify which orientation of the equatorial plane and of the ecliptic plane, and so which location of the point Υ , are taken to define the reference frame. Usually, the *mean* vernal equinox and equator (or ecliptic) of a reference epoch are selected, whereby ‘mean’ refers to the fact that relatively short-period motions have been filtered out. In astrodynamics, two choices for the reference epoch are commonly used: the beginning of the year 2000 or the date for which the computations are performed. We then speak of the *mean equinox of 2000 reference frame* or the *mean equinox of date reference frame*. However, we have to realize that the very definitions of these planes and their intersection are problematic for high-precision; there is no way magically to ‘paint’ them on the celestial sphere at any particular time! Therefore, in practice, a set of fiducial objects with assigned coordinates is used as the representation of an astronomical reference system. The scheme is completely analogous to how terrestrial reference systems are established using geodetic reference points on the Earth’s surface.

Nowadays, the fundamental celestial reference system for astronomical applications is the *International Celestial Reference System* (ICRS), which is a space-time coordinate system, within the framework of general relativity theory, with its origin at the center of mass of the solar system (*barycenter*) and axis directions that are fixed with respect to distant objects (quasars and active galactic nuclei) in the universe, which are observable at radio wavelengths. These radio sources are assumed to have no observable intrinsic angular motions. Thus, the ICRS is a ‘space-fixed’ system; i.e. a kinematically non-rotating system. The IAU has specified a set of more than 210 extragalactic radio sources, whose adopted coordinates, relative to the ICRS origin at the solar system barycenter and measured by Very Long Baseline Interferometry (VLBI), effectively define the directions of the ICRS axes. This ‘realization’ of the ICRS is called the *International Celestial Reference Frame* (ICRF). The ICRS is meant to represent the most appropriate coordinate system currently available for expressing reference data on the positions and motions of celestial objects, including spacecraft outside the immediate environment of the Earth. The ICRS axes (as implemented by the ICRF) closely approximate (to within about $0.02''$) the axes that would be defined by the dynamical mean equator and equinox of 2000 reference frame, if the latter is considered to be a barycentric system. The IAU has adopted a rotation matrix B containing three very small angular offsets to convert data in the ICRS system to corresponding data in the mean equinox of 2000 reference frame. Because the ICRS axes are meant to be space fixed, there is no date associated with the ICRS. Furthermore, since the defining radio sources are assumed to be so distant that their angular motions, seen from Earth, are negligible, there is no epoch associated with the ICRF. It will be clear that the same orientation of the ICRS axes can be used to define a geocentric kinematically non-rotating reference frame, which is generally called the *Geocentric Celestial Reference Frame* (GCRF). This is a ‘local’ reference frame that forms the basis for the definition of an appropriate non-rotating reference system in which positions and velocities of bodies in the near-Earth environment, e.g. satellites, should be expressed.

To describe the rotational motion of the Earth relative to the GCRF we have to apply models for the precession and nutation of the Earth. Over the years, the IAU has adopted models of ever increasing accuracy. The IAU 2000A precession-nutation model¹ consists of two parts: a preces-

sion algorithm and a nutation algorithm. The precession algorithm consists of a short polynomial series for the values of certain angles. The sines and cosines of these angles, in combination, then define the elements of a precession matrix, P . The nutation algorithm consists of a rather long series expansion in Fourier terms for the angular offsets of the actual celestial pole from the precession-only pole (true pole minus mean pole). The sines and cosines of these offsets, in combination, then define the elements of a nutation matrix, N . Both matrices are applied to the coordinates of celestial objects to transform them from the equator and equinox of one epoch to the equator and equinox of another epoch.

It is noted that the theories for precession and nutation hold for a dynamical mean equator and equinox of date reference frame. Therefore, when we start from positions relative to ICRS and want to compute coordinates with respect to the mean equinox of date reference frame, we first have to apply the transformation matrix B to convert ICRS data to the dynamical mean equator and equinox of 2000 reference frame, and then have to apply a transformation from that frame to the mean equinox of date reference frame, using the existing precession and nutation models.

Precession and nutation are motions of the Earth's rotation axis relative to inertial space. These motions affect the declination and right ascension of a celestial body and therefore have to be taken into account in the computation of the positions of celestial bodies. For an observer on Earth, the pole of the rotation axis moves relative to the Earth's surface as a result of the non-collinearity of the Earth's polar principal axis of inertia and the Earth's rotation axis. This phenomenon is called *polar motion*. Since the position of points on Earth is specified relative to an Earth-fixed rotating reference frame with its Z-axis pointing to the average position of the north pole over some time interval, polar motion affects the 'effective' latitude and longitude of points on the Earth's surface. It should certainly be accounted for when observations from tracking stations are used for the orbit computation of spacecraft. For completeness, the major aspects of polar motion will be described briefly.

In essence, polar motion is a free nutation of the Earth. The (instantaneous) rotation axis then describes a cone about the polar principle axis of inertia. It was first predicted by L. Euler (1707-1783) in 1765 using a rigid model of the Earth. He found that the true north pole performs a counterclockwise circular motion about the mean north pole with a period of about 305 days. However, the Earth is deformable, which complicates the picture. The resulting free nutation, called the *Chandler wobble*, is a periodic motion with a mean amplitude of $0.18''$ (5.5 m on the Earth's surface) and a mean period of 433 days. Polar motion is continuously excited by mass redistribution in the atmosphere, oceans, mantle and core. As a result, the amplitude and period of the Chandler wobble vary with time, and other periodic variations are introduced. These include an annual term with a mean amplitude of $0.09''$, a periodic term with an amplitude of $0.02-0.03''$ and a period of about 30 years (*Markowitz wobble*), and a semi-annual term with an amplitude of $0.01''$. It is emphasized, however, that the amplitudes of these periodic variations may change considerably over the years. In addition, polar motion includes a secular (non-periodic) term. Since observations began in 1899, the secular term has moved the north pole in the direction of Newfoundland (76.1° W) at a mean rate of $0.0033''$ (10 cm) per year. But in 2005, this drift changed abruptly; the pole began moving eastwards and continues to do so.

Polar motion can hardly be modeled and predicted with a reasonable accuracy and therefore one has to rely on observations. Information about the actual Earth's orientation variations, as determined from observations, is provided by the IERS Earth Orientation Centre at the Observatoire de Paris, Paris, and the IERS Rapid Service/Prediction Center at the U.S. Naval Observatory (USNO), Washington, DC.

In the following, we will assume that the orientation of the equatorial plane and ecliptic, the position of the vernal equinox and the coordinates of points on Earth are known with sufficient accuracy at all epochs of interest.

11.3. Topocentric, geocentric and heliocentric reference frames

A measurement on the position and velocity of a satellite (or interplanetary spacecraft) is usually taken by an observer on the surface of the rotating Earth. This means that such an observation is taken relative to a rotating *topocentric reference frame*, where the observer is at the origin of the coordinate system and the fundamental plane is the local horizontal plane, i.e. the plane tangent to the standard ellipsoid at the position of the observer (Figure 11.4). The $+X$ -axis is taken as pointing due north, the $+Z$ -axis is pointing away from the observer's zenith (so 'down'), and the east-pointing $+Y$ -axis completes the right-handed reference frame. The two angles needed to define the location of an object relative to the observer are the *elevation*, H , and the *azimuth*, ψ (Figure 11.4). The elevation is the angular position of an object relative to the local horizontal plane, and is measured from 0° to 90° above the horizontal plane and from 0° to -90° below the horizontal plane. The azimuth is the angle from the local north to the object's meridian, measured in clockwise direction from 0° to 360° in the local horizontal plane. Distance from the observer to the object is defined as the *slant range*, ρ .

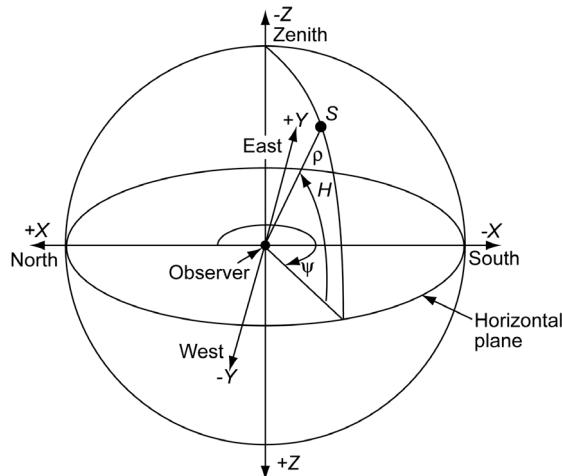


Figure 11.4: Position relative to the topocentric reference frame.

For describing the motion of a rocket or satellite with respect to the Earth's surface, usually a *geocentric rotating reference frame* is used. In this frame (Figure 11.1), the $+Z$ -axis is directed along the Earth's rotation axis towards the north pole and the $+X$ -axis is in the Earth's equatorial plane, crossing the Greenwich meridian. The $+Y$ -axis lies in the equatorial plane, oriented such as to make the reference frame right-handed. In spherical coordinates, position can be expressed by the length of the geocentric radius vector, r , the geocentric latitude, ϕ , and the geographic longitude, Λ .

For the description of the motion of satellites about the Earth, of spacecraft in a trajectory to the Moon, or of spacecraft or planets in their orbits about the Sun, normally a *non-rotating reference frame* is chosen that is suitable for the problem considered. For example, when the motion of a satellite about the Earth or the motion of an interplanetary spacecraft about the Sun is considered, the origin of the reference frame can be chosen at the center of the Earth or the Sun, respectively. One should realize that even when the motion of the spacecraft is described

as a pure two-body problem, these reference frames are non-inertial reference frames. After all, the origin of the reference frame describes a Keplerian orbit about the mass center of the system. That was precisely the reason why the ICRF (Section 11.2) was introduced.

Usually, the orientation of a non-rotating reference frame is defined by starting with a reference plane that serves as the XY -plane, in which a fixed reference direction that is determined with respect to the stationary celestial sphere, serves as the $+X$ -axis. Two obvious choices for the reference plane can be made. For describing the motion of satellites about the Earth, the XY -plane of the non-rotating reference frame with its origin at the mass center of the Earth is the equatorial plane. For the motion of interplanetary spacecraft, the ecliptic plane is used as the reference plane of the non-rotating reference frame with the Sun at its origin. These reference frames are called the *non-rotating geocentric equatorial reference frame* and the *non-rotating heliocentric ecliptic reference frame*, respectively, or shortly the *geocentric* and the *heliocentric* reference frame.

For the direction of the $+Z$ -axis, which is perpendicular to the reference plane, two options are still open. For the geocentric reference frame, the $+Z$ -axis is chosen from the center of the Earth in the direction of the north pole of the Earth. The $+Z$ -axis in the heliocentric reference frame is chosen such that the angle between this axis and the $+Z$ -axis in the geocentric reference frame is equal to the obliquity of the ecliptic ($\varepsilon = 23^\circ 27'$). In both reference frames, the $+Z$ -axis now defines the *north pole* (Figure 11.3). One of the obvious choices for the reference direction, the $+X$ -axis, in the geocentric reference frame is the direction towards the First Point of Aries, Υ (Figure 11.1). The $+Y$ -axis is chosen in the reference plane, the equatorial plane, in such a way that the reference frame is right-handed. The point Υ is also used as reference direction in the heliocentric reference frame. This is possible because the radius of the celestial sphere is infinitely large and the lines Earth- Υ and Sun- Υ are therefore parallel. Moreover, the line Sun- Υ lies in the ecliptic plane, which is a requirement for the X -axis in the heliocentric reference frame.

In the geocentric non-rotating equatorial reference frame the geocentric angle measured along the hour circle through a celestial object (star, planet, etc., spacecraft), indicated by S in Figure 11.1, from the celestial equator to that object is called the *declination*, δ , of the object. It is measured from 0° to 90° ; positive north of the equator, negative south of it. Declination in this reference frame is analogous to geocentric latitude on the surface of the Earth. The geocentric angle measured along the celestial equator, from the point Υ to the foot of the hour circle through S is called the *right ascension*, α , of the object. Right ascension is measured from 0° to 360° from point Υ eastward. Declination, right ascension and distance from the Earth's center describe the position of a vehicle relative to the geocentric equatorial reference frame. In the geocentric non-rotating ecliptic reference frame (Figure 11.3), the angular position of the celestial object S is specified by two quantities, *celestial longitude*, λ , and *celestial latitude*, φ . The celestial latitude is defined as the geocentric angular distance measured along the circle of celestial longitude passing through the object, from the ecliptic to that object. It is taken positive for objects north and negative for objects south of the ecliptic, and measured from 0° to 90° . The celestial longitude is defined as the angular distance measured along the ecliptic from point Υ to the foot of the circle of celestial longitude through that object. This angle is measured from 0° to 360° eastward along the ecliptic. In the heliocentric non-rotating ecliptic frame a position can be described by its *heliocentric radius*, r , and by its *heliocentric longitude* and *latitude*, defined analogously to celestial longitude and latitude in the geocentric ecliptic reference frame.

In all reference frames, the velocity vector of a rocket, satellite or spacecraft can be described by the magnitude of the velocity and by two angles determining the direction of the velocity vector. To this end, we introduce a *spacecraft local horizontal coordinate system*, where the local horizontal plane is defined as the plane through the spacecraft perpendicular to the radius vector, \bar{r}

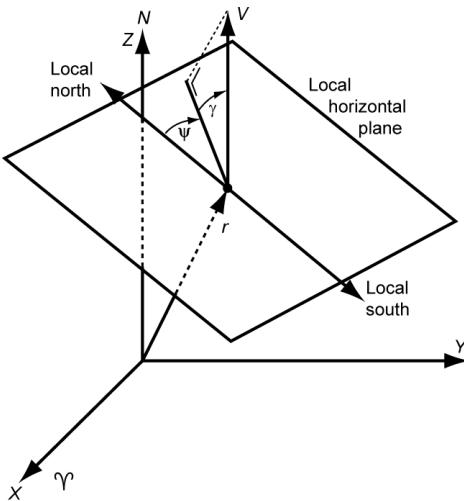


Figure 11.5: Spacecraft horizontal plane, flight path angle, γ , and flight path azimuth, ψ .

(Figure 11.5). The local north-south direction is defined by the intersection of the +Z-axis of the geocentric, heliocentric or planetocentric reference frame with this local horizontal plane (Figure 11.5). The +X-axis of the spacecraft local horizontal coordinate system points due north, the +Z-axis points in the opposite direction of \bar{r} ('down') and the east-pointing +Y-axis completes the right-handed reference frame. The *flight path angle*, γ , is the angle between the velocity vector and the horizontal plane, and was already introduced in Section 5.6 in the analysis of the two-body problem. This angle varies between -90° and $+90^\circ$; $+90^\circ$ corresponding with a radially outward directed velocity. The angle between the local north direction and the projection of the velocity vector onto the horizontal plane is called the *flight path azimuth*, ψ . This azimuth is measured from local north in a clockwise direction from 0° to 360° . Note the similarity between this concept of flight path azimuth and the azimuth introduced earlier for the description of a position in the topocentric reference frame.

11.4. Definition and measurement of time

Man has long held a belief that there is a 'uniform time'. This concept of uniformity, however, turns out to lack meaning. We have no master standard of which we are sure that it runs uniformly. In fact, we even do not need it; all we really need is that our chronometric variable, if inserted into our physical laws, furnishes reproducible predictions that are in agreement with our observations. In practice, we take a suitable repetitive physical phenomenon and define successive periods of repetition to be equal in length. Nowadays, it is felt that the time provided by an atomic clock should correlate to the highest accuracy with our physical laws, and we take by definition that this time progresses uniformly. We call this *Atomic Time* (AT). Though in this way we can define a unit of time, for the fitting of time in a conventional scale we are committed to astronomical observations; historically these form the basis for time reckoning systems.

In astronomy, the measurement of time intervals is based on the rotation of the Earth. The observer's time is defined as the angular distance covered by a reference object on the celestial sphere after its last crossing of the (upper branch of the) observer's celestial meridian. The time interval between two successive crossings is called a *day*. Though the actual length of a day will, in general, depend on the reference object chosen, each type of day is divided into 24 equal parts, called *hours*. These hours again are subdivided into 60 *minutes* and each minute into 60 *seconds*. The *hour angle*, H , of an object (Figure 11.1) is defined as the angular distance measured along

the celestial equator from the observer's meridian to the foot of the object's hour circle. Because for an observer on Earth, the apparent motion of celestial objects on the celestial sphere is in a westward direction, the hour angle is measured positively to the west so that the hour angle increases with time. As an increment in the hour angle of 360° corresponds to a time increment of exactly one day, time can be defined as the hour angle of the reference object. As reference object a particular star can be selected, e.g. Antares. When Antares crosses the observer's meridian, we may speak of $0^{\text{h}}0^{\text{m}}0^{\text{s}}$ Antares time. Instead of selecting a particular star, it turns out to be more convenient to use the vernal equinox or the Sun as reference object, resulting in *sidereal time* or *solar time*, respectively. As the hour angle of a given point on the celestial sphere differs at a certain epoch for observers with a different geographic longitude, both types of time are local and we therefore speak of *local sidereal time* and *local solar time*. From Figure 11.1 we can make the following observation. If the celestial object S transits the observer's meridian, the hour angle of S is zero and the hour angle of the point Υ is equal to the right ascension of the object, α . This leads to the important conclusion that a celestial object always transits the observer's meridian at a local sidereal time equal to the right ascension of that object.

The sidereal time at Greenwich is called the *Greenwich sidereal time*, θ_G . From our definitions, we conclude that the observer's sidereal time, θ , is obtained by adding the longitude of the observer, Λ , to the Greenwich sidereal time:

$$\theta = \theta_G + \Lambda \quad ; \quad 0 \leq \theta < 2\pi \quad (11.1)$$

Sidereal time appears in two forms, *mean* and *apparent*, depending on whether the mean or true (apparent) vernal equinox is the reference point. The difference between apparent and mean sidereal time is expressed by the *equation of the equinoxes*, which is a complex periodic function with a maximum amplitude of about 1.2 s. Apparent sidereal time is more relevant to actual observations, since it includes the effect of nutation. For astrodynamics, however, we usually select the mean vernal equinox of date as a reference, leading to *Greenwich mean sidereal time*, θ_{GM} , and the mean vernal equinox at a reference epoch, e.g. January 1, 2000, to define the *mean equinox of 2000 reference frame*. If we apply the concept of right ascension to describe the position of an observer (or any other point on Earth) relative to the geocentric non-rotating reference frame, we may write (Figure 11.1)

$$\alpha = \theta_{GM} + \Lambda \quad ; \quad 0 \leq \alpha < 2\pi \quad (11.2)$$

since both sidereal time and right ascension use the vernal equinox as reference. The value of θ_{GM} can be obtained from astronomical observations (see below). Once it is known, (11.2) provides a very important basic linkage between the rotating and non-rotating geocentric reference frames. We will use (11.2) in Section 17.4 and Section 18.2 to determine the right ascension of a launch site.

The *sidereal day* is defined as the interval between two successive passages of the mean vernal equinox across the observer's meridian. As the point Υ moves slowly at a non-uniform rate on the celestial sphere, the sidereal day will not have a constant length when compared to an atomic clock, but will vary slightly. The sidereal day, however, is to within 0.01 s equal to the true period of rotation of the Earth with respect to a fixed point on the celestial sphere. When periodic fluctuations in the length of a sidereal day are filtered out, we speak of a *mean sidereal day*. Sidereal time is very useful to astronomy; our lives and working days, however, are to a large extent regulated by the Sun. Therefore, it is often desirable to use solar time. A *solar day* is defined as the period of the Earth's rotation with respect to the Sun. As the Sun moves, for an observer on Earth, at about 0.986° per day along the ecliptic in an eastward direction, it will be clear from inspection of Figure 11.6 that a solar day is somewhat longer than a sidereal day. After

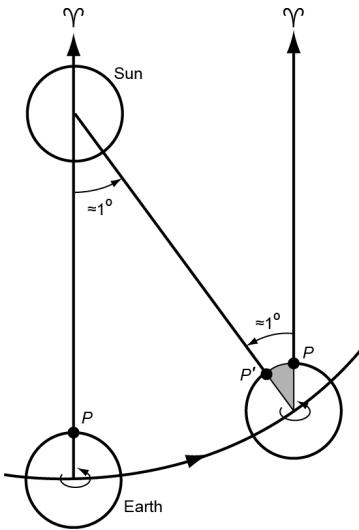


Figure 11.6: Difference between a sidereal day and a solar day.

one revolution of the Earth relative to point Υ , an observer P has still to cover the angle PP' before his solar time has increased by 24 hours. As it takes the Earth $3^{\text{m}}56^{\text{s}}$ to rotate through an angle of 0.986° , the solar day is about 4 min longer than the sidereal day. *Apparent solar time* is related to the hour angle of the Sun. At midday apparent solar time, the Sun passes the upper branch of the observer's meridian. It is convenient not to start a solar day at the Sun's crossing of the upper branch of the observer's meridian (noon), but at the crossing of the lower branch (midnight). Consequently, the elapsed apparent solar time since the beginning of a solar day is the hour angle of the Sun plus 12 hr. On September 22/23, the Sun passes through the autumnal equinox. The vernal equinox then passes the meridian of an observer for whom it is midnight. At that moment, local apparent solar time and local sidereal time are equal for the observer mentioned. With each succeeding day from then, sidereal time gains about 4 min on solar time, until just one year later the difference between them is 24 hr.

Owing to the ellipticity of the Earth's orbit about the Sun, and the fact that the equatorial plane and the ecliptic plane do not coincide, the apparent motion of the Sun on the celestial sphere certainly is not 'uniform'. For that reason, the solar day does not have a constant length. Therefore, the concept of a fictitious *mean Sun* was introduced by S. Newcomb (1835-1909) in 1895; this mean Sun is assumed to move at a uniform rate eastward along the celestial equator (not the ecliptic!), with a sidereal motion equal to the mean sidereal motion of the true Sun at January 0.5 (i.e. December 31, 12^h), 1900. Mathematical relations are available to compute the position of the mean Sun relative to the true Sun at any time. A *mean solar day*, or *civil day*, is the interval between two successive crossings of the mean Sun through the observer's meridian, and *mean solar time* is defined as the hour angle of the mean Sun plus 12 hr. During the year, the difference between both kinds of solar time varies between -14 min to +16 min. By definition, the length of a *mean solar second* is found by dividing a mean solar day by 86400 (= $24 \times 60 \times 60$) equal parts. In astrodynamics, we often use a standardized mean solar time: the *Greenwich mean solar time*, which is usually called *Universal Time (UT)*.

The rotation rate of the Earth is not constant, but shows various types of changes. Solid-Earth and ocean tides cause the rotation rate, $\dot{\Theta}$, to decrease secularly at a present rate of about $-5.5 \times 10^{-22} \text{ rad/s}^2$, or $\ddot{\Theta}_s \approx -2.4 \times 10^{-10} \dot{\Theta}$ per year. In addition, the rotation rate shows a number of variations with different periods. Because of these irregularities in the Earth's rotation rate, even the mean solar day varies slightly over the years as compared to atomic time. Over the last four

centuries the mean solar day has been increasing at a mean value of approximately 1.4 ms per century. Superimposed on this secular trend, short-term fluctuations of between -3 ms and + 4 ms have occurred on a time scale of decades. For very-precise time determination therefore a new kind of time was introduced in 1950: *Ephemeris Time* (ET). This is a time variable that is consistent with the laws of dynamics, and causes the position of celestial bodies, as predicted from the dynamical equations of motion, to agree with observation. The fundamental unit of ET is the *ephemeris second*, which was chosen such that it equals the length of a mean solar second at Greenwich mean noon, January 0, 1900. In 1975, ET and UT differed by about 45 s owing to changes in the Earth's rotation rate since January 0, 1900; in 1995 this difference had increased to about 62 s. The use of ET continued until 1984; in 1979 and 1982, *Terrestrial Dynamical Time* (TDT) and *Barycentric Dynamical Time* (*Temps Dynamique Barycentrique*, TDB) were introduced to replace ET. These are time variables that are consistent with the laws of dynamics, including relativistic effects, when the motion of planets is computed relative to a solar system barycentric reference frame. TDT refers to the Earth's center, while TDB refers to the solar system's barycenter. In 1991, *Terrestrial Time* (TT) was introduced as the new name for TDT. It is noted that all dynamical time scales have no relation to the rotation of the Earth!

We have found that, formally, two definitions of uniform time are in use. In astronomy, ET, TDT, TDB and TT were devised from the laws of dynamics; in physics, AT was introduced. The big advantage of atomic time is that it can be measured in a laboratory environment. The epoch at which atomic time starts is usually set at 1958, January 1, 0^h0^m0^s UT. *International Atomic Time* (*Temps Atomique International*, TAI) is a product of the Bureau International des Poids et Mesures (BIPM) in Sèvres, France, and is based on the analysis of about 200 frequency standards maintained by several countries to keep a unit of time as close to the ideal Système International (SI) second as possible. One SI second on the Earth's surface at sea level (specifically, the rotating geoid) equals the duration of 9,192,631,770 cycles of the wavelength associated with the radiation emitted by the electron transition between two hyperfine levels of the ground state of cesium-133. This numerical value makes the TAI second to be equal to the ET second as determined from astronomical observations over the period 1956 to 1965. In January 1972, the TAI second was adopted for all timing, replacing the ET second. On 1958, January 1, ET and TAI differed by 32.184 s and assuming that TAI and ET run at an equal rate, presently the difference is still 32.184 s. By definition, the difference TT - TAI is 32.184 s, which is the best available estimate of the difference between TDT and TAI on January 1, 1977, 0^h0^m0^s, and therefore makes TT to preserve continuity with the now obsolete time scales ET and TDT.

The universal time mentioned above (UT) is nowadays known as UT0. Because this time is based on astronomical observations, the effect of the secular decrease of the rotational velocity of the Earth has implicitly been taken into account. From this time, several other time indications have been deducted, which are corrected for known deviations in the orientation of the rotational axis and in the rotational velocity of the Earth. Polar motion results in variations in the moment of time at which the mean Sun passes the meridian of Greenwich, and thus in UT. The version of universal time that is corrected for this effect is called UT1, which makes this time independent of station location. For an observer at mid-latitudes, differences between UT1 and UT0 are at most about 0.030 s. When also season-bound variations in the rotational velocity of the Earth are corrected for, then the universal time is indicated by UT2. The correction from UT1 to UT2 is at most about 0.035 s. However, the concept of UT2 is nowadays considered obsolete.

Universal Time Coordinated (UTC) is a hybrid standard of time, the progression of which is determined by TAI and the reference to the Earth's rotation by UT1. This is the time that is broadcasted by radio signals of standard radio stations. Due to the fact that UTC is related to

atomic time, it shows a highly ‘uniform course’. It should be noticed, however, that when UTC was officially initiated on January 1, 1961, the UTC second was not identical to the TAI second. This UTC was intended to permit a very close approximation of UT2, within around 0.1 s. To achieve that goal, synchronization of UTC with UT2 and adjustment of the length of the UTC second had to be applied quite frequently. This led to practical problems and therefore the concept of UTC was adapted. Since January 1, 1972, the UTC second is identical to the TAI second, UTC tracks UT1 rather than UT2, and *leap seconds* are used to keep the difference between UT1 and UTC to within ± 0.9 s. Theoretically, a negative or positive leap second can be inserted at any time; a positive leap second corresponds to a decrease of the Earth’s rotation rate and a negative leap second to an increase of the rotation rate. Up to now, only positive leap seconds have been inserted on June 30 or December 31 of a particular year. UTC was set exactly 10 s behind TAI on January 1, 1972, 0^h0^m0^s, which was the approximate accumulated difference between TAI and UT1 since the inception of TAI in 1958. Consequently, $\Delta\text{AT} = \text{TAI} - \text{UTC}$ is always an integer number of seconds. The first leap second was inserted on June 30, 1972. Usually, leap seconds were inserted once a year; a major exception is the period January 1, 1999, to December 31, 2005, when the difference between UTC and UT1 increased slower than usual and no leap seconds had to be inserted. ΔAT has changed in steps of 1 s from 0 s on January 1, 1958, via 10 s on January 1, 1972, 25 s on January 1, 1990, 32 s from January 1, 1999, through December 31, 2005, 33 s from January 1, 2006, through December 31, 2008, 34 s from January 1, 2009, through June 30, 2012, and 35 since then (until at least June 2015).

The satellites of the Global Positioning System (GPS) transmit time signals that are related to UTC. *GPS Time* (GPST) is implemented by the atomic clocks in the GPS ground control stations and the GPS satellites. The difference between UTC and GPST is that GPST is not corrected to match the rotation of the Earth, so it does not contain leap seconds. GPST was synchronized with UTC on January 6, 1980, 0^h0^m0^s UTC, when ΔAT was 19 s, and has been steered in close synchrony with TAI since then. Therefore, the difference TAI - GPST has always been 19 s to within 50 ns. Consequently, GPST was ahead of UTC by 13 s during the period January 1, 1999, to December 31, 2005, by 14 s from January 1, 2006, through December 31, 2008, by 15 s from January 1, 2009, through June 30, 2012, and by 16 s since then; this time difference is also broadcasted by the GPS satellites. Receivers subtract this offset from GPST to calculate UTC.

For the user, then, UTC, which is widely available from GPS, radio broadcast services, and the Internet, is the practical starting point for computing any of the other time scales described above. We simply add the current value of ΔAT to UTC to obtain TAI; TT is then just 32.184 seconds ahead of TAI. The theoretical time scales TDB and ET can be obtained from TT using the appropriate mathematical formulas. For the time scales based on the rotation of the Earth, we again start with UTC and add the current value of ΔUT1 to obtain UT1. The various kinds of sidereal time can then be computed from UT1 using standard formulas.

Around 2008 one began to realize that the occasional insertion of leap seconds into UTC may create serious difficulties for operational navigation and communication systems. Therefore, the *International Telecommunication Union* (ITU) started a study to determine future requirements for continuous globally-accepted time scales for use both in navigation/telecommunication systems, and for civil time keeping. The study results should be available by 2015.

Year and month

We have seen that the length of the day is based on the rotation of the Earth. In addition to the day, we also use the year and the month to describe time intervals. The year is based on the motion of the Earth about the Sun, while the month is based on the motion of the Moon about the

Earth. There are various definitions of the year and the month; in this Section we will only mention the most important ones.

The *sidereal year* is the time it takes the Earth to complete one revolution about the Sun, when measured relative to an inertial reference frame; so, it is the orbital period of the Earth relative to the stars. The average length of the sidereal year is $365^{\text{d}}6^{\text{h}}9^{\text{m}}10^{\text{s}}$ (365.256363 mean solar days). The *anomalistic year*, with an average length of $365^{\text{d}}6^{\text{h}}13^{\text{m}}52^{\text{s}}$ (365.259636 mean solar days), is the period between successive perihelion passages of the Earth. The *tropical year* is the time between two successive passages of the Earth through the vernal equinox, and is the year that is synchronized with the seasons on Earth. The average length of the tropical year is $365^{\text{d}}5^{\text{h}}48^{\text{m}}45^{\text{s}}$ (365.242190 mean solar days). The differences in length of the anomalistic and tropical years relative to the sidereal year are a result of the fact that the ecliptic, the equatorial plane and the orientation of the (elliptical) Earth's orbit in the ecliptic vary with time due to the gravitational attraction by Sun, Moon and planets. A *Besselian year*, named after F.W. Bessel (1784-1846), is a kind of tropical year and refers to one complete revolution in right ascension of the (fictitious) mean Sun. It starts at the moment when the ecliptic longitude of the mean Sun is 280° ; this value was chosen because it is currently near the beginning of the Gregorian year (see below). The Besselian year has (about) the same length as the tropical year. Its use has become obsolete since 1984.

The *sidereal month* is the time it takes the Moon to complete one revolution about the Earth, when measured relative to an inertial reference frame; so, it is the orbital period of the Moon relative to the stars. The average length of the sidereal month is $27^{\text{d}}7^{\text{h}}43^{\text{m}}12^{\text{s}}$ (27.321661 mean solar days). The *anomalistic month* is the interval between two successive perigee passages of the Moon; its average length is $27^{\text{d}}13^{\text{h}}18^{\text{m}}33^{\text{s}}$ (27.554550 mean solar days). The *nodal or draconic (draconitic) month* is the interval between two successive passages of the Moon through its ascending node (ecliptic), its average length is $27^{\text{d}}5^{\text{h}}5^{\text{m}}36^{\text{s}}$ (27.212220 mean solar days). The *tropical month* is the time it takes the Moon to return to the same ecliptic longitude; its average length is $27^{\text{d}}7^{\text{h}}43^{\text{m}}5^{\text{s}}$ (27.321582 mean solar days). The *synodic month* is the interval between two successive New Moon's (a lunation); so, it is the orbital period of the Moon relative to the Sun and is on average $29^{\text{d}}12^{\text{h}}44^{\text{m}}3^{\text{s}}$ (29.530589 mean solar days). Again, the gravitational attraction by Sun, Earth and planets are the cause for the differences in length of the anomalistic, nodal and tropical months relative to the sidereal month. The difference of about 2.2 days between a synodic month and a sidereal month is a result of the motion of the Earth-Moon system about the Sun.

An interesting fact is that 223 synodic months is approximately equal to 239 anomalistic months, which is approximately equal to 242 draconic months, which is approximately equal to 6,585.322 days. This period of 6,585.322 days (18 years plus 10.322 days or 11.322 days, depending on the number of leap years, which is called today the *Saros cycle*, has been known since at least the Babylonians around the fifth century B.C. and was later known to Hipparchos, Gaius Plinius Secundus (Pliny; 23-79) and Claudius Ptolemy (latin: Ptolemaeus; ~85-165) under different names. The name 'Saros' was first given to this cycle by E. Halley (1656-1742) in 1691. It can be used to predict solar and lunar eclipses. These eclipses can only occur when the Moon is between the Sun and the Earth (New Moon) or when the Earth is between the Sun and the Moon (Full Moon), respectively. Suppose that at a certain moment of time an eclipse occurs. After one Saros cycle, the Earth-Sun-Moon geometry will be nearly identical and a nearly identical eclipse will occur. A complication is that the period of the Saros cycle is not an integer number of days, but contains a residual of about 1/3 of a day. Thus, as a result of the Earth's rotation, for each successive Saros cycle, an eclipse will occur about 8 hr later in the day. In the case of an eclipse of the Sun, this means that the region of visibility will shift westward by about

120°, and the two eclipses will thus not be visible from the same place on Earth. In the case of an eclipse of the Moon, the next eclipse might still be visible from the same location as long as the Moon is above the horizon. After three Saros cycles, the local time of day of an eclipse will be nearly the same. This period of three Saros cycles (19,755.97 days), is known as a *Triple Saros* and was already known to the ancient Greek astronomers.

Calendar

The calendar of ancient cultures was based on the synodic month, probably because changes in the Moon's phase are so readily apparent. At the simplest level, all lunar calendars are based on the approximation that 2 lunations last about 59 days. These calendars therefore used a 30 day full month followed by a 29 day hollow month. However, this system is only marginally accurate and quickly needs correction by using larger cycles, or the equivalent of leap days. To keep the lunar calendar (on the long run) synchronized with the seasons, one tried to use a system with an integer number of months in an integer number of years. The most common solution to this problem is the *Metonic cycle* (Meton of Athens; ~460-400 B.C.), which takes advantage of the fact that 235 lunations are approximately equal to 19 tropical years (6939.6016 days); the difference is only about 2.083 hours. The still more-accurate *Callippic cycle* (Callippus of Cyzicus; ~370-310 B.C.) consists of four Metonic cycles, less one day. A *Metonic calendar* (such as the *Hebrew calendar*) will drift against the seasons by about 1 day every 220 years. The problems of creating reliable lunar calendars may explain why solar calendars were introduced. These solar calendars, having months which no longer relate to the phase of the Moon, are based on the motion of the Sun against the sky; they have generally replaced lunar calendars for civil use.

Since the tropical year, which is synchronized with the seasons, corresponds to 365.242190 mean solar days and one used 29 or 30 days in a lunar calendar, it was quite logical to adopt 12 months in a solar calendar plus some 'additional' days. The ancient Egyptians had a calendar that consisted of 12 months of 30 days each, with 5 days extra set aside at the end of the year for celebrations. In that way the year had a length of 365 days. The calendar of ancient Rome, the *Roman calendar*, originally consisted of 10 months of 30 or 31 days; the calendar year lasted 304 days and there were about 61 days of winter which were not assigned to any month. The *Julian calendar*, instituted by the Roman emperor Julius Caesar in 46 BC, had 365 days divided, as we do today, in 7 months of 31 days, 4 months of 30 days and 1 month of 28 days; every fourth year there was an additional day at February 24. So, the average length of a Julian year is 365.25 mean solar days, which is about 11.25 min longer than the tropical year, leading to a discrepancy of 0.781 day per century. By the time that the Julian calendar was adopted on January 1, 45 BC, it had to be made of 445 days long in order to bring the seasons back on schedule. At the Council of Christian bishops convened at Nicaea in the year 325 the Julian calendar was 'calibrated' such that the Earth's passage through the vernal equinox occurred on March 21. By 1582, the difference between the tropical year and the Julian year since the Council of Nicaea had resulted in an accumulated difference of about 10 days. So, the date of the Earth's vernal equinox passage had retreated from March 21 to March 11. Accordingly, Pope Gregory XIII called upon the astronomer Clavius to revise the Julian calendar. His solution was to drop 10 days from the calendar, so that the day following Thursday, October 4, 1582, thereby became Friday, October 15, 1582. To avoid future discrepancies, he introduced the system that century years divisible by 400 do not have 366 days but 365 days. Consequently, the average length of the *Gregorian year* is 365.2425 mean solar days. As this values is very close to the length of a tropical year, this calendar is synchronized with the seasons to about 1 day in 3300 years. The *Gregorian calendar*, which is our present *civil calendar*, was readily adopted by the Catholic countries; Protestants

and Lutherans followed slowly in adopting this calendar. When Great Britain and the American colonies changed to the Gregorian calendar in 1752, September 2 had to be followed by September 14, in order to compensate for the discrepancy that had grown to 12 days since 1582. Early in the twentieth century the last countries in Europe and Asia finally adopted the Gregorian calendar or one very close to it.

Julian Date

The *Julian Date* (JD) is based on a system of Julian day numbers that involves a continuous count of mean solar days, and so provides a simpler way of determining the interval of time between two dates than by using the conventional calendar. This concept was introduced by J.J. Scaliger (1540-1609) in 1583. He combined the solar cycle (28 years), the Metonic cycle (19 years), and the Roman Indication (15 years)⁴. Scaliger defined the *Julian period* by multiplying these three chronological cycles, leading to a period of 7980 *Julian years* of 365.25 days. Because the three cycles were already established, he chose the last common point for these cycles as epoch at which the Julian Date was exactly zero: 4713 BC, January 1, noon at Alexandria, on the Julian calendar. This epoch pre-dated all historical dates known to Scaliger and is also close to the one that in medieval times was believed to have marked the creation of the Earth. Modern astronomy adopted the concept of Julian Date in the late nineteenth century, but used the meridian of Greenwich instead of Alexandria, after the former was made the Prime Meridian by an international conference in 1884. As an example, the Julian Date for 1995, August 1, 18^h UTC is 2,449,931.25; for the same time on August 1, 2008, the Julian Date is 2454680.25. In order to avoid those large numbers for indicating dates in our times, it is convenient to express epochs in *Modified Julian Date* (MJD), which is related to Julian Date by MJD = JD - 2,400,00.5. It is emphasized that while a Julian Date starts at Greenwich mean noon, a Modified Julian Date starts at Greenwich mean midnight, which is more convenient.

Computation of UT1, Julian Date and mean sidereal time

We have seen that the Greenwich mean sidereal time, θ_{GM} , is a crucial parameter to link the geocentric rotating and non-rotating reference frames. Here, a procedure will be given to compute θ_{GM} at a specific date and time, where that time is expressed in UTC, TAI, TT or GPST. We then first convert the time variable to UT1 by the following relations:

$$\text{UT1} = \text{UTC} + \Delta\text{UT1} ; \quad \text{TAI} = \text{UTC} + \Delta\text{AT} ; \quad \text{TT} = \text{TAI} + 32.184 ; \quad \text{TAI} = \text{GPST} + 19$$

where all times are in SI seconds. Tables of the differences ΔUT1 and ΔAT for various dates are published in the IERS Bulletins A and B mentioned in Section 11.2; both past observations and predictions are available. When we know the time in UT1, we continue by determining the Julian Date at the given date and time. A convenient formula that holds for the period March 1, 1900, to February 28, 2100, is given by D.A. Vallado⁵. If yr indicates the year (four digits), mo the number of the month and d the number of the day, then the Julian Date at 0^h0^m0^s UT1 can be found from

⁴ The *solar cycle* is a period of time under a calendar system during which a leap year day takes place on each of the seven days of the week. Under the Julian calendar it is a period of 28 years, because leap years occur every four years and there are seven possible days to start a leap year. The *Roman Indication* refers to a cycle of agricultural or land tax reassessment.

⁵ D.A. Vallado, *Fundamentals of Astrodynamics and Applications*, Microcosm Press and Kluwer Academic Publishers, 2001.

$$JD_0 = 1,721,013.5 + 367 \text{yr} - \text{INT} \left[\frac{7}{4} \left\{ \text{yr} + \text{INT} \left(\frac{mo+9}{12} \right) \right\} \right] + \text{INT} \left(\frac{275 mo}{9} \right) + d$$

where INT denotes ‘real’ truncation and indicates the integer (floor) value. If h indicates the hour of the day, min the minute and s the second, then the Julian Date at any other time (UT1) of that day can be found from

$$JD = JD_0 + \left(\frac{(s/60 + min)/60 + h}{24} \right)$$

The IAU has adopted a convention⁶ to determine Greenwich mean sidereal time at $0^{\text{h}}0^{\text{m}}0^{\text{s}}$ UT1 as a function of JD_0 :

$$\begin{aligned} T &= (JD_0 - 2,451,545.0)/36,525 \\ \theta_{GM_0} &= 100.4606184 + 36,000.77005361 T + 0.00038793 T^2 - 2.6 \cdot 10^{-8} T^3 \end{aligned} \quad (11.3-1)$$

where the number 2,451,545.0 is the Julian Date on January 1, 2000, $0^{\text{h}}0^{\text{m}}0^{\text{s}}$ UT1, and the value of θ_{GM_0} is expressed in degree and refers to the mean equinox of 2000 reference frame. The Greenwich mean sidereal time at any other time (UT1) of that day can be computed from

$$\begin{aligned} \theta_{GM} &= \theta_{GM_0} + \dot{\theta} \Delta t \\ \dot{\theta} &= \frac{1}{240} (1.002737909350795 + 5.9006 \cdot 10^{-11} T - 5.9 \cdot 10^{-15} T^2) \end{aligned} \quad (11.3-2)$$

where $\dot{\theta}$ is the Earth’s mean rotation rate expressed in degree per second, and Δt is the time in seconds from midnight UT1. The local mean sidereal time can then be found from (11.1).

11.5. Orbital elements

During the discussion of the two-body problem (Chapter 5) it was found that the orbit of a body (point mass), with respect to a non-rotating reference frame that is fixed to the center of the other body, is a conic section, of which the shape, size and orientation with respect to this reference frame do not change. In that Chapter, the differential equations for the motion in the orbital plane have been solved, and three integration constants were defined: a , e , and ω . The first two constants (semi-major axis and eccentricity) determine the shape and size of the orbit; the orientation of the conic section in the orbital plane is determined by the argument of pericenter, ω , which is the angle between a fixed reference direction in the orbital plane (Section 5.2) and the radius vector from origin to pericenter. A fourth integration constant, the time of pericenter passage, τ , was introduced in Chapters 6, 7 and 8, and is needed to link time and position in the orbit. Consequently, the orbit and the position of a body in that orbit can be described by the orbital plane, the above-mentioned four integration constants and time.

The orientation of the orbital plane relative to the reference frame can be determined by two angles. For this purpose, first the line of intersection of the orbital plane with the reference plane of the reference frame (equatorial or ecliptic plane) is considered (Figure 11.7). The points where this line of intersection crosses the orbit of the body are called *nodes*. Depending on whether the

⁶ P.K. Seidelmann, B. Guinot, L.E. Doggett, *Explanatory Supplement to the Astronomical Almanac*, University Science Books, California, 1992.

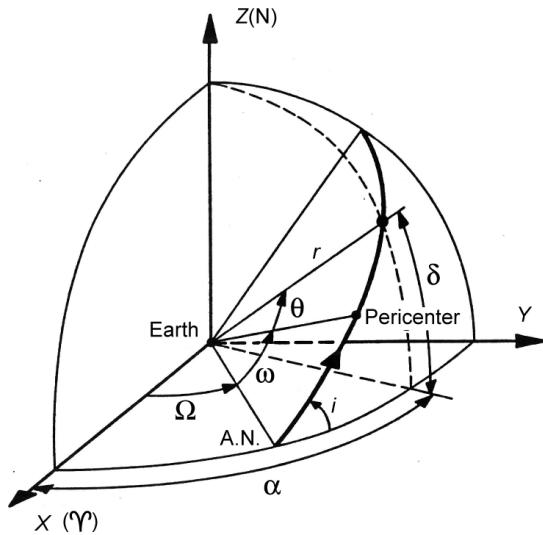


Figure 11.7: Definition of the orbital elements i , Ω , ω , and the position of a satellite in spherical coordinates.

body passes these nodes from south-to-north or from north-to-south, the node is referred to as the *ascending node* (A.N.) or *descending node*. The line that connects these two points, i.e. the line of intersection of the orbital plane and the reference plane, is called the *nodal line*. The first angle used for describing the orientation of the orbital plane is the angle from the reference direction towards the point Υ to the ascending node, measured in the reference plane. This angle is indicated by Ω (Figure 11.7) and is measured eastward from 0° to 360° . Depending on whether the reference plane is the ecliptic or the equatorial plane, this angle is called the *longitude of the ascending node* or the *right ascension of the ascending node*, respectively. The second angle is the angle between the orbital plane and the reference plane: the *inclination*, i (Figure 11.7). This angle is measured from the reference plane. To define this angle unambiguously, the convention is that when one looks from the ascending node to the origin of the reference frame, the angle i is measured counter-clockwise from the reference plane. An equivalent definition is that the inclination is the angle between the +Z-axis and the angular momentum vector of the motion of the body, \bar{H} . From these definitions, we conclude that the value of the inclination may range from 0° to 180° . Orbits may be categorized as *direct* or *retrograde*, based on their inclination. Direct means easterly; this is the direction in which the Sun, the Earth, and most of the planets and their moons rotate about their axes, and the direction in which all of the planets revolve around the Sun. A satellite with an orbital inclination $i < 90^\circ$ moves in a direct orbit and in the same direction as the Earth rotates about its axis. Satellite orbit inclinations of $90^\circ < i < 180^\circ$ imply retrograde orbits, in which the satellite moves in a direction opposite to the Earth's rotation. As the fixed reference direction in the orbital plane, which is needed for the definition of the *argument of pericenter*, ω , the direction of the vector from the origin to the ascending node is chosen. So, ω is measured from the nodal line in the direction of motion of the body in its orbital plane, from 0° to 360° . It are these angles i , Ω and ω , which were already used in classical astronomy, that were adopted by Euler for problems in mechanics and later became known as the (original) *Euler angles*. The six constants a , e , i , Ω , ω , τ are called the *classical orbital elements*.

When the equations that describe the motion of body i relative to body k would have been derived using the *Hamilton-Jacobi theory* that is commonly applied in physics, we would have obtained a solution that, of course, represents again a Keplerian orbit. This solution contains six integration constants, which are, for a certain solution method, related to the classical orbital

elements by $\alpha_1 = -\mu/2a$, $\alpha_2 = \sqrt{\mu a(1-e^2)}$, $\alpha_3 = \sqrt{\mu a(1-e^2)} \cos i$, $\beta_1 = -\tau$, $\beta_2 = \omega$, $\beta_3 = \Omega$. The α -constants represent the total energy, the total angular momentum and the component of the angular momentum about the Z-axis, all per unit of mass, respectively. It is this Hamilton-Jacobi theory and this set of orbital elements, or another set derived from it (Section 22.2), that form the basis for modern analytical theories for the analysis of perturbed orbits. However, for the discussions presented in this book, the Hamilton-Jacobi theory is not needed and the classical treatment even leads to a better insight into the practical applications of astrodynamics. Therefore, in the following we will not use this type of elements. Only in Section 22.2 we will briefly discuss them again.

In summary, the orbit and the instantaneous position of body i in its orbit can be described by six independent constant parameters: the orbital elements a , e , i , Ω , ω , τ , and time. That there are just six orbital elements is evident when we realize that position and velocity of the body are determined by the integration of three second-order differential equations. An analytical solution of this set of differential equations will contain six integration constants. The six classical orbital elements therefore have to be regarded as a set of cleverly chosen, physically interpretable integration constants. The angles i , ω and Ω provide a geometric insight into the orientation of the orbital plane and the orientation of the orbit in the orbital plane. The value of a gives, through (6.14), information about the total orbital energy and the type of orbit; a and e together determine the shape and size of the orbit. However, these orbital elements are nothing else than integration constants in the solution of the equations of motion, and therefore one can compose from these orbital elements any other set of orbital elements, as long as there exist unambiguous one-to-one relations between the elements of both sets. An example is the use of the mean anomaly at epoch, M_0 , introduced in Section 6.5, instead of the element τ . Sometimes, the mean anomaly itself is considered as an orbital element. Strictly, this is not correct since the mean anomaly varies linearly with time because of the motion of body i , and therefore should be regarded as an orbital parameter, just as the true anomaly, θ , and the eccentric anomaly, E , rather than as an orbital element. However, in this book we will not use this strict definition.

An alternative set of orbital elements has to be used when the classical orbital elements are not suitable for certain orbits, and then may lead to serious problems, e.g. in the mathematical treatment of perturbed satellite orbits (Section 22.4). When the orbital inclination is zero, both elements ω and Ω are undefined. For (near-)equatorial orbits we therefore use the orbital element $\tilde{\omega} = \Omega + \omega$, where $\tilde{\omega}$ is called the *longitude of perigee*, to describe the location of perigee. This angle is measured in the equatorial plane from the X-axis to the ascending node and subsequently in the orbital plane to perigee. For these (near-)equatorial orbits we use the *mean longitude* $\lambda_m = \tilde{\omega} + M$, *true longitude* $\lambda_t = \tilde{\omega} + \theta$, or *eccentric longitude* $\lambda_e = \tilde{\omega} + E$ to describe the in-orbit position of the satellite. When we deal with circular orbits, then the orbital elements ω and τ , and the orbital parameters θ and M , are not defined. For (near-)circular orbits we therefore select the line of nodes as a reference and use the *argument of latitude* $u = \omega + \theta$ to describe the in-orbit position. For equatorial orbits, of course, u is not defined and we therefore use for (near-)circular, (near-)equatorial orbits the mean, true or eccentric longitude to describe the in-orbit position of the satellite. For a parabolic orbit $a = \infty$ and $e = 1$, which means that both a and e does not allow us to distinguish one parabolic orbit from another. In that case, we use the orbital element $q = a(1-e)$, which indicates the perigee distance. It is emphasized that in the literature many other types of orbital elements are introduced, which have advantages for specific cases.

When we consider only the classical orbital elements, we may conclude that position and velocity of body i can be expressed in at least three ways; i.e. by the following sets of coordinates:

- rectangular coordinates : $x, y, z, \dot{x}, \dot{y}, \dot{z}$.
- spherical coordinates : $r, \alpha, \delta, V, \gamma, \psi$.
- orbital elements : $a, e, i, \Omega, \omega, \tau$.

Obviously, there will exist unambiguous relations between these three sets of parameters. The third set is fundamentally different from the first two. In a Keplerian orbit, both $x, y, z, \dot{x}, \dot{y}, \dot{z}$ and $r, \alpha, \delta, V, \gamma, \psi$ change continuously; consequently they are functions of time. The orbital elements $a, e, i, \Omega, \omega, \tau$ are constants and retain for every position in the Keplerian orbit the same value. When one knows at a certain moment of time the position and velocity of body i in, for example, rectangular coordinates and one wants to compute the position and velocity at another moment of time, then the standard calculation method is as follows. From the known values of $x, y, z, \dot{x}, \dot{y}, \dot{z}$ the orbital elements at that moment of time are computed. For the other moment of time these elements have the same values and subsequently the values of $x, y, z, \dot{x}, \dot{y}, \dot{z}$ at that moment of time can be determined through a reversed transformation. However, while there is no inherent difficulty in calculating position and velocity from known orbital elements with e and i close to zero, the reverse task may cause practical and numerical problems. These problems are due to singularities arising from the definition of some of the classical orbital elements.

In Sections 11.7 through 11.10, a number of important relations between rectangular coordinates, geodetic coordinates, spherical coordinates and orbital elements will be derived. When position and velocity are described by using orbital elements, we will use as independent variable the true anomaly, θ , instead of time. For elliptical orbits, this anomaly is directly related to time through Kepler's equation (6.36); for parabolic or hyperbolic through Barker's equation (7.11) or (8.23). We will restrict ourselves to elliptical orbits about the Earth; for parabolic and hyperbolic orbits similar relations can be derived.

11.6. Relation between geocentric and geodetic latitude

Figure 11.8 shows a cross-section of the standard ellipsoid that approximates the true shape of the Earth, and a circumscribing circle. The flattening of this ellipsoid is defined in Section 11.1 as

$$f = \frac{a_e - a_p}{a_e} \quad (11.4)$$

In Section 6.1, we have found the following relation between the semi-major and semi-minor

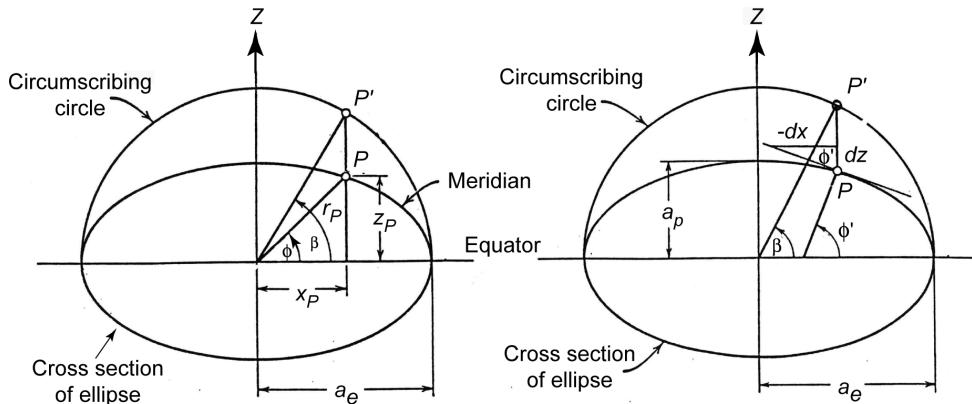


Figure 11.8: Cross section of the standard ellipsoid, with geodetic and geocentric latitude and auxiliary variables.

axes of an ellipse:

$$a_p = a_e \sqrt{1 - e^2} \quad (11.5)$$

where e is the eccentricity of the ellipse. Combination of (11.4) and (11.5) yields a relation between the flattening and the eccentricity of an ellipsoid:

$$e^2 = 2f - f^2 ; \quad 1 - e^2 = (1 - f)^2 \quad (11.6)$$

According to Appendix B: $f = 1/298.25642$, $e = 0.081818$. Below, we will use the parameter e as a measure of the shape of the ellipsoid.

In Figure 11.8 (left), P denotes the location of an observer on the standard ellipsoid. The projection of P on the circumscribing circle is indicated by P' . For this geometry we find

$$x_P = a_e \cos \beta$$

$$z_P = a_e \sqrt{1 - e^2} \sin \beta \quad (11.7)$$

$$r_P = \sqrt{x_P^2 + z_P^2} = a_e \sqrt{1 - e^2 \sin^2 \beta}$$

The geocentric latitude, ϕ , can be obtained in terms of β from

$$\begin{aligned} \sin \phi &= \frac{z_P}{r_P} = \frac{\sqrt{1 - e^2} \sin \beta}{\sqrt{1 - e^2 \sin^2 \beta}} \\ \cos \phi &= \frac{x_P}{r_P} = \frac{\cos \beta}{\sqrt{1 - e^2 \sin^2 \beta}} \end{aligned} \quad (11.8)$$

Multiplication of (11.8-2) by $\sqrt{1 - e^2}$, and squaring both the resulting relation and (11.8-1) yields

$$\sin^2 \phi = \frac{(1 - e^2) \sin^2 \beta}{1 - e^2 \sin^2 \beta}$$

$$(1 - e^2) \cos^2 \phi = \frac{(1 - e^2) \cos^2 \beta}{1 - e^2 \sin^2 \beta}$$

which, upon addition and after taking the square root on both sides, leads to

$$\sqrt{1 - e^2 \cos^2 \phi} = \frac{\sqrt{1 - e^2}}{\sqrt{1 - e^2 \sin^2 \beta}}$$

Substitution of this relation into (11.8-1) and (11.8-2) results in

$$\sin \beta = \frac{\sin \phi}{\sqrt{1 - e^2 \cos^2 \phi}} ; \quad \cos \beta = \frac{\sqrt{1 - e^2} \cos \phi}{\sqrt{1 - e^2 \cos^2 \phi}}$$

Substitution of these relations into (11.7) yields

$$x_P = a_e \frac{\sqrt{1 - e^2} \cos \phi}{\sqrt{1 - e^2 \cos^2 \phi}} \quad (11.9)$$

$$\begin{aligned} z_P &= a_e \frac{\sqrt{1 - e^2} \sin \phi}{\sqrt{1 - e^2 \cos^2 \phi}} \\ r_P &= a_e \sqrt{\frac{1 - e^2}{1 - e^2 \cos^2 \phi}} \end{aligned} \quad (11.9)$$

These relations give the rectangular position components of an observer P on the surface of the cross section of the adopted Earth's standard ellipsoid as a function of geocentric latitude, ϕ .

We continue with a similar analysis for the geodetic latitude, ϕ' , and now use the geometry indicated in Figure 11.8 (right). Differentiation of 11.7 gives

$$dx_P = -a_e \sin \beta d\beta \quad ; \quad dz_P = a_e \sqrt{1 - e^2} \cos \beta d\beta$$

Therefore,

$$ds \equiv \sqrt{(dx_P)^2 + (dz_P)^2} = a_e \sqrt{1 - e^2 \cos^2 \beta} d\beta$$

and

$$\begin{aligned} \sin \phi' &= -\frac{dx_P}{ds} = \frac{\sin \beta}{\sqrt{1 - e^2 \cos^2 \beta}} \\ \cos \phi' &= \frac{dz_P}{ds} = \frac{\sqrt{1 - e^2} \cos \beta}{\sqrt{1 - e^2 \cos^2 \beta}} \end{aligned} \quad (11.10)$$

Multiplication of (11.10-1) by $\sqrt{1 - e^2}$, and squaring both the resulting relation and (11.10-2), gives

$$(1 - e^2) \sin^2 \phi' = \frac{(1 - e^2) \sin^2 \beta}{1 - e^2 \cos^2 \beta}$$

$$\cos^2 \phi' = \frac{(1 - e^2) \cos^2 \beta}{1 - e^2 \cos^2 \beta}$$

Adding these two relations and subsequently taking the square root on both sides leads to

$$\sqrt{1 - e^2 \sin^2 \phi'} = \frac{\sqrt{1 - e^2}}{\sqrt{1 - e^2 \cos^2 \beta}}$$

Substitution of this relation into (11.10) gives

$$\sin \beta = \frac{\sqrt{1 - e^2}}{\sqrt{1 - e^2 \sin^2 \phi'}} \sin \phi' \quad ; \quad \cos \beta = \frac{1}{\sqrt{1 - e^2 \sin^2 \phi'}} \cos \phi'$$

Substitution of these relation into (11.7) leads to

$$\begin{aligned}x_P &= a_e \frac{\cos \phi'}{\sqrt{1 - e^2 \sin^2 \phi'}} \\z_P &= a_e \frac{(1 - e^2) \sin \phi'}{\sqrt{1 - e^2 \sin^2 \phi'}} \\r_P &= a_e \sqrt{\frac{1 - (2e^2 + e^4) \sin^2 \phi'}{1 - e^2 \sin^2 \phi'}}\end{aligned}\tag{11.11}$$

These expressions for the rectangular coordinates of an observer P on the surface of the cross-section of the standard ellipsoid are similar to the expressions (11.9). However, the variable is now ϕ' instead of ϕ in (11.9).

Combination of (11.9) and (11.11) results in

$$\frac{z_P}{x_P} = \tan \phi = (1 - e^2) \tan \phi'\tag{11.12-1}$$

or

$$\phi' = \arctan \left[\frac{1}{1 - e^2} \tan \phi \right]\tag{11.12-2}$$

Because $e \ll 1$; $\phi' \approx \phi$. A relation for the difference between ϕ' and ϕ can be found as follows. If Δ is defined as

$$\Delta = \phi' - \phi\tag{11.13}$$

we can write according to (11.12-1):

$$\tan \phi = (1 - e^2) \tan(\phi + \Delta)$$

or

$$\tan \phi = (1 - e^2) \frac{\tan \phi + \tan \Delta}{1 - \tan \phi \tan \Delta}$$

This relation can be written as

$$\tan \Delta = \frac{e^2 \sin \phi \cos \phi}{1 - e^2 \cos^2 \phi}\tag{11.14}$$

Because $\Delta \ll 1$ and $e \ll 1$, we may approximate (11.14) by the series expansion:

$$\Delta \approx \frac{1}{2} e^2 \sin 2\phi [1 + e^2 \cos^2 \phi] + O(e^6)\tag{11.15}$$

To determine the values of ϕ for which $|\Delta|$ is a maximum, we write (11.14) as

$$\Delta = \arctan \left[\frac{e^2 \sin \phi \cos \phi}{1 - e^2 \cos^2 \phi} \right]\tag{11.16}$$

Differentiation of this equation leads, after some algebraic manipulation, to

$$\frac{\partial \Delta}{\partial \phi} = e^2 \frac{(1 - e^2) \cos^2 \phi - \sin^2 \phi}{1 - e^2 (2 - e^2) \cos^2 \phi}$$

Because $e \ll 1$, the denominator in this expression is always positive and setting the numerator equal to zero yields for the value of ϕ for which $|\Delta|$ takes an extreme value:

$$\tan^2 \phi_{ext} = 1 - e^2 \quad (11.17)$$

From (11.16) and (11.17), and $e = 0.081818$, we find that Δ takes a maximum value of $11.55'$ at $\phi = 44.904^\circ$ and $\phi = 224.904^\circ$, and that Δ takes a minimum value of $-11.55'$ at $\phi = 135.096^\circ$ and $\phi = 315.096^\circ$. From (11.16) we find that $\Delta = 0$ at $\phi = 0^\circ, 90^\circ, 180^\circ, 270^\circ$.

11.7. Transformation from rectangular coordinates to geodetic coordinates

In this Section, it is assumed that at a specified time the position of a satellite is given in the rectangular coordinates x, y, z . Then, the distance of the satellite from the origin is given by

$$r = \sqrt{x^2 + y^2 + z^2} \quad (11.18)$$

The declination, δ , of the satellite is given by

$$\sin \delta = \frac{z}{r} \quad (11.19)$$

where $-90^\circ \leq \delta \leq 90^\circ$. For the satellite's right ascension, α , we find (Figure 11.7):

$$\sin \alpha = \frac{y}{r \cos \delta} ; \quad \cos \alpha = \frac{x}{r \cos \delta} \quad (11.20)$$

These relations uniquely determine α .

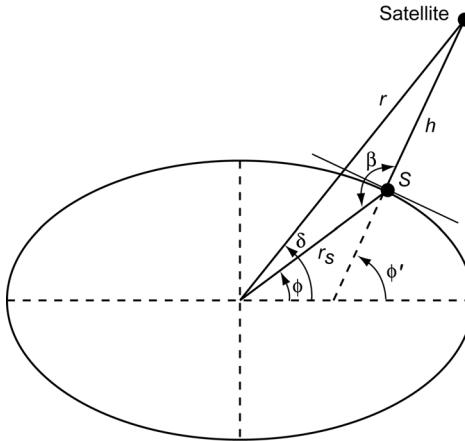


Figure 11.9: Position of a satellite in geodetic coordinates and some auxiliary variables.

In Figure 11.9, S is the sub-satellite point; i.e. the point on the standard ellipsoid that is nearest to the satellite. In this configuration, the satellite transits the meridian through S . With (11.2) and (11.3-2) we may write for the satellite's longitude

$$\Lambda = \alpha - [\theta_{GM_0} + \dot{\theta} \Delta t] \quad (11.21)$$

where θ_{GM_0} may be computed from (11.3-1), $\dot{\theta}$ is the Earth's mean rotation rate that may be computed from (11.3-2), and Δt is the time from midnight UT1.

The values of ϕ , ϕ' and h cannot be obtained in a direct way, but require an iterative solution. As a first approximation, we assume that the geocentric latitude of the sub-satellite point is equal to the declination of the satellite, i.e. $\phi = \delta$. Then, we compute r_s and ϕ' from the relations (11.9-3) and (11.12-2). These equations show that the error in the solution for ϕ' is of the same order as the error in ϕ , while the error in the solution of r_s is an order of magnitude smaller. We now write (Figure 11.9)

$$r^2 = h^2 + r_s^2 - 2hr_s \cos\beta \quad (11.22)$$

From the geometry depicted in Figure 11.9, we find

$$\beta = \pi - (\phi' - \phi) \quad (11.23)$$

So, we may write (11.22) as

$$r^2 = h^2 + r_s^2 + 2hr_s \cos(\phi' - \phi)$$

This relation can also be written as

$$h^2 + 2hr_s \cos(\phi' - \phi) + r_s^2 \cos^2(\phi' - \phi) = r^2 - r_s^2 + r_s^2 \cos^2(\phi' - \phi)$$

or

$$h = \sqrt{r^2 - r_s^2 \sin^2(\phi' - \phi)} - r_s \cos(\phi' - \phi) \quad (11.24)$$

Because h is expressed in terms of ϕ' - ϕ , the error in h is also one order of magnitude smaller than the error in ϕ . From the geometry depicted in Figure 11.9 and using (11.23), we find

$$\frac{r}{\sin(\phi' - \phi)} = \frac{h}{\sin(\delta - \phi)}$$

or

$$\phi = \delta - \arcsin \left\{ \frac{h}{r} \sin(\phi' - \phi) \right\} \quad (11.25)$$

Again, ϕ is expressed in terms of ϕ' - ϕ , leading to small errors. This value of ϕ is substituted in (11.12-2) and the iteration loop may be executed again. The iteration process should be continued until the difference between successive values of ϕ is below a specified limit. We then know the solutions of ϕ , ϕ' and h .

11.8. Transformation from orbital elements to rectangular coordinates

In this Section, relations will be derived that express the rectangular coordinates and velocity components at a certain time in terms of the orbital elements and the true anomaly at that time. We will apply two different methods; the first method is based on spherical geometry and the second method on rotation matrices. The reason for discussing these two methods is that it will introduce two mathematical techniques that are frequently applied in astrodynamics.

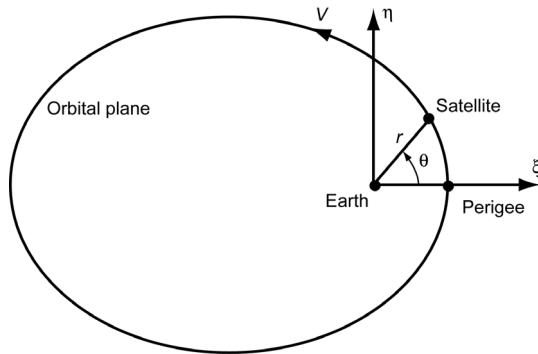


Figure 11.10: A satellite in an elliptical orbit about the Earth, and the geocentric non-rotating $\xi\eta\zeta$ reference frame.

To start with the method of spherical geometry, we first define a rectangular non-rotating reference frame $\xi\eta\zeta$ with its origin at the center of the Earth, the $\xi\eta$ -plane coinciding with the orbital plane and the ξ -axis pointing towards perigee (Figure 11.10). The direction of the η -axis is chosen equal to the direction of the velocity vector at perigee. Then,

$$\xi = r \cos\theta \quad ; \quad \eta = r \sin\theta \quad ; \quad \zeta = 0 \quad (11.26)$$

When the orbital elements are known, the solution of Kepler's equation (Section 6.5) for a certain time t_0 provides us with the value of the eccentric anomaly, E , at that time. With the relations presented in Section 6.5, we then may obtain the value of the true anomaly, θ , at that time. For the rectangular coordinates relative to the geocentric equatorial reference frame we find (Figure 11.7):

$$x = r \cos\delta \cos\alpha \quad ; \quad y = r \cos\delta \sin\alpha \quad ; \quad z = r \sin\delta \quad (11.27)$$

We now consider the angle α as the sum of two angles: $(\alpha - \Omega)$ and Ω , and rewrite (11.27) as

$$\begin{aligned} x &= r \cos\delta [\cos(\alpha - \Omega) \cos\Omega - \sin(\alpha - \Omega) \sin\Omega] \\ y &= r \cos\delta [\sin(\alpha - \Omega) \cos\Omega + \cos(\alpha - \Omega) \sin\Omega] \\ z &= r \sin\delta \end{aligned} \quad (11.28)$$

Equations (11.28) show that if one wants to express x, y, z in orbital elements, then the terms $\sin\delta, \cos\delta \cos(\alpha-\Omega)$ and $\cos\delta \sin(\alpha-\Omega)$ have to be expressed in orbital elements. This can be done by using classical relations from spherical trigonometry for a right-angle spherical triangle shown in Figure 11.11.

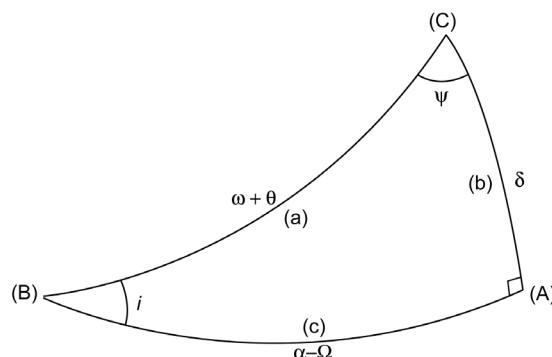


Figure 11.11: Right-angle spherical triangle in which the rules of spherical trigonometry are applied.

First cosine rule

$$\cos a = \cos b \cos c$$

Second cosine rule

$$\sin a \cos B = \cos b \sin c ; \quad \sin a \cos C = \cos c \sin b$$

Sine rule

$$\sin b = \sin B \sin a ; \quad \sin c = \sin C \sin a$$

When we substitute the corresponding orbital parameters, which are also indicated in Figure 11.11, for the quantities a, b, c , and B , we find

$$\begin{aligned} \cos(\omega + \theta) &= \cos \delta \cos(\alpha - \Omega) \\ \sin(\omega + \theta) \cos i &= \cos \delta \sin(\alpha - \Omega) \\ \sin \delta &= \sin i \sin(\omega + \theta) \end{aligned} \tag{11.29}$$

Substitution of (11.29) into (11.28) yields

$$\begin{aligned} x &= r [\cos(\omega + \theta) \cos \Omega - \sin(\omega + \theta) \cos i \sin \Omega] \\ y &= r [\sin(\omega + \theta) \cos i \cos \Omega + \cos(\omega + \theta) \sin \Omega] \\ z &= r \sin(\omega + \theta) \sin i \end{aligned} \tag{11.30}$$

After some trigonometric manipulation, we obtain

$$\begin{aligned} x &= r [\cos \omega \cos \theta \cos \Omega - \sin \omega \sin \theta \cos \Omega \\ &\quad - \sin \omega \cos \theta \cos i \sin \Omega - \cos \omega \sin \theta \cos i \sin \Omega] \\ y &= r [\sin \omega \cos \theta \cos i \cos \Omega + \cos \omega \sin \theta \cos i \cos \Omega \\ &\quad + \cos \omega \cos \theta \sin \Omega - \sin \omega \sin \theta \sin \Omega] \\ z &= r [\sin \omega \cos \theta \sin i + \cos \omega \sin \theta \sin i] \end{aligned}$$

These relations can be written as

$$\begin{aligned} x &= r [l_1 \cos \theta + l_2 \sin \theta] \\ y &= r [m_1 \cos \theta + m_2 \sin \theta] \\ z &= r [n_1 \cos \theta + n_2 \sin \theta] \end{aligned} \tag{11.31-1}$$

where

$$\begin{aligned} l_1 &= \cos \omega \cos \Omega - \sin \omega \sin \Omega \cos i \\ m_1 &= \cos \omega \sin \Omega + \sin \omega \cos \Omega \cos i \\ n_1 &= \sin \omega \sin i \\ l_2 &= -\sin \omega \cos \Omega - \cos \omega \sin \Omega \cos i \\ m_2 &= -\sin \omega \sin \Omega + \cos \omega \cos \Omega \cos i \\ n_2 &= \cos \omega \sin i \end{aligned} \tag{11.32}$$

and

$$r = \frac{a(1 - e^2)}{1 + e \cos \theta} \quad (11.33)$$

The parameters $l_1, l_2, m_1, m_2, n_1, n_2$ are combinations of constants (orbital elements) and are therefore constant themselves.

With (11.26) the relations (11.31-1) can also be written as

$$\begin{aligned} x &= l_1 \xi + l_2 \eta \\ y &= m_1 \xi + m_2 \eta \\ z &= n_1 \xi + n_2 \eta \end{aligned} \quad (11.31-2)$$

The velocity components $\dot{x}, \dot{y}, \dot{z}$ can be found by differentiating (11.31-2):

$$\begin{aligned} \dot{x} &= l_1 \dot{\xi} + l_2 \dot{\eta} \\ \dot{y} &= m_1 \dot{\xi} + m_2 \dot{\eta} \\ \dot{z} &= n_1 \dot{\xi} + n_2 \dot{\eta} \end{aligned} \quad (11.34)$$

The derivatives $\dot{\xi}$ and $\dot{\eta}$ can be obtained from (11.26):

$$\dot{\xi} = \dot{r} \cos \theta - r \dot{\theta} \sin \theta ; \quad \dot{\eta} = \dot{r} \sin \theta + r \dot{\theta} \cos \theta \quad (11.35)$$

In Chapter 5 we have derived expressions for \dot{r} and $r\dot{\theta}$ as a function of θ . Substituting (5.26) and (5.27) into (11.35), we obtain

$$\dot{\xi} = -\frac{\mu}{H} \sin \theta ; \quad \dot{\eta} = \frac{\mu}{H} (e + \cos \theta) \quad (11.36)$$

where the angular momentum, H , is given by

$$H = \sqrt{\mu a (1 - e^2)} \quad (11.37)$$

Substitution of (11.36) into (11.34) gives

$$\begin{aligned} \dot{x} &= \frac{\mu}{H} [-l_1 \sin \theta + l_2 (e + \cos \theta)] \\ \dot{y} &= \frac{\mu}{H} [-m_1 \sin \theta + m_2 (e + \cos \theta)] \\ \dot{z} &= \frac{\mu}{H} [-n_1 \sin \theta + n_2 (e + \cos \theta)] \end{aligned} \quad (11.38)$$

Relations (11.26), (11.31), (11.32) and (11.38) give the relationships between the orbital elements and the true anomaly on the one hand, and the rectangular coordinates and velocity components on the other hand. To find the true anomaly, the equation of Kepler has to be solved.

Obviously, the reverse problem can also be considered, i.e. the determination of the orbital elements as a function of $x, y, z, \dot{x}, \dot{y}, \dot{z}$. This will be done in Section 11.10.

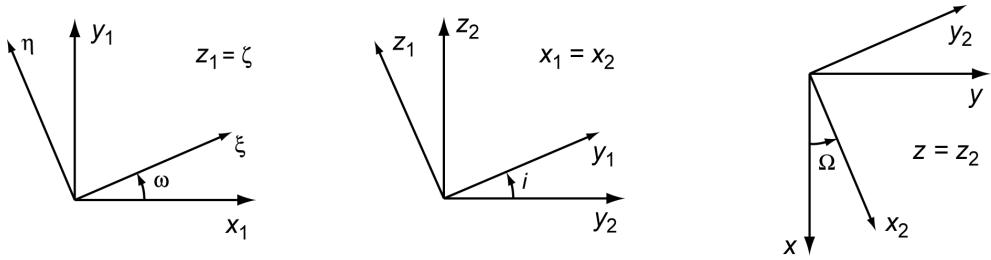


Figure 11.12: Geometry of reference frame rotations.

Now, we will consider the transformation of orbital elements to rectangular coordinates and velocity components by means of the application of rotation matrices. We start again from the geocentric non-rotating reference frame $\xi\eta\zeta$ (Figure 11.10), of which the $+\xi$ -axis points towards perigee, the $\xi\eta$ -plane is the orbital plane and the $+\zeta$ -axis is in the direction of the angular momentum vector. For the transformation from the $\xi\eta\zeta$ reference frame to the geocentric equatorial non-rotating XYZ reference frame (Figure 11.7), three rotations of the reference frame are required (Figure 11.12). First the $\xi\eta\zeta$ reference frame is rotated around the ζ -axis over an angle $-\omega$, until the ξ -axis coincides with the line of nodes. This is followed by a rotation over an angle $-i$ around the line of nodes and finally by a rotation over an angle $-\Omega$ around the Z -axis. We then obtain the following relations between the coordinates:

$$\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} = \begin{bmatrix} \cos\omega & -\sin\omega & 0 \\ \sin\omega & \cos\omega & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} \xi \\ \eta \\ \zeta \end{pmatrix} \quad (11.39-1)$$

$$\begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos i & -\sin i \\ 0 & \sin i & \cos i \end{bmatrix} \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} \quad (11.39-2)$$

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{bmatrix} \cos\Omega & -\sin\Omega & 0 \\ \sin\Omega & \cos\Omega & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} \quad (11.39-3)$$

Thus,

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{bmatrix} \cos\Omega & -\sin\Omega & 0 \\ \sin\Omega & \cos\Omega & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos i & -\sin i \\ 0 & \sin i & \cos i \end{bmatrix} \begin{bmatrix} \cos\omega & -\sin\omega & 0 \\ \sin\omega & \cos\omega & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} \xi \\ \eta \\ \zeta \end{pmatrix}$$

or

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{bmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \\ n_1 & n_2 & n_3 \end{bmatrix} \begin{pmatrix} \xi \\ \eta \\ \zeta \end{pmatrix} \quad (11.40)$$

where

$$\begin{aligned}
l_1 &= \cos\omega \cos\Omega - \sin\omega \sin\Omega \cos i \\
m_1 &= \cos\omega \sin\Omega + \sin\omega \cos\Omega \cos i \\
n_1 &= \sin\omega \sin i \\
l_2 &= -\sin\omega \cos\Omega - \cos\omega \sin\Omega \cos i \\
m_2 &= -\sin\omega \sin\Omega + \cos\omega \cos\Omega \cos i \\
n_2 &= \cos\omega \sin i \\
l_3 &= \sin\Omega \sin i \\
m_3 &= -\cos\Omega \sin i \\
n_3 &= \cos i
\end{aligned} \tag{11.41}$$

We see that the first six relations are identical to (11.32). The expressions for l_3 , m_3 and n_3 appear because we have considered a general transformation between reference frames, and have not yet used the condition that the satellite is always in the $\xi\eta$ -plane, i.e. $\zeta = 0$. In that case, we find from (11.40):

$$\begin{aligned}
x &= l_1 \xi + l_2 \eta \\
y &= m_1 \xi + m_2 \eta \\
z &= n_1 \xi + n_2 \eta
\end{aligned} \tag{11.42}$$

These equations are, of course, identical to (11.31-2). With (11.26) we then find for the position components the relations (11.31-1). For the rectangular velocity components the relations (11.38) hold.

From the characteristics of the matrix in (11.40) some relations can be derived that will be used in Section 11.10 and Section 22.1. Note that the matrix

$$A = \begin{bmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \\ n_1 & n_2 & n_3 \end{bmatrix} \tag{11.43}$$

is composed of three rotation matrices, which are each orthogonal. This means that also the matrix A is orthogonal and thus

$$A^T A = I$$

where A^T is the transpose of A , and I is the identity matrix (unit matrix). So, we may write

$$\begin{bmatrix} l_1 & m_1 & n_1 \\ l_2 & m_2 & n_2 \\ l_3 & m_3 & n_3 \end{bmatrix} \begin{bmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \\ n_1 & n_2 & n_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

which yields

$$l_1^2 + m_1^2 + n_1^2 = l_2^2 + m_2^2 + n_2^2 = l_3^2 + m_3^2 + n_3^2 = 1 \tag{11.44}$$

$$l_1 l_2 + m_1 m_2 + n_1 n_2 = l_1 l_3 + m_1 m_3 + n_1 n_3 = l_2 l_3 + m_2 m_3 + n_2 n_3 = 0 \quad (11.44)$$

These relations will be used in Section 22.1.

Because A is orthogonal:

$$A^{-1} = A^T$$

from which follows that we may write for the reversed transformation from x, y, z to ξ, η, ζ :

$$\begin{pmatrix} \xi \\ \eta \\ \zeta \end{pmatrix} = \begin{bmatrix} l_1 & m_1 & n_1 \\ l_2 & m_2 & n_2 \\ l_3 & m_3 & n_3 \end{bmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad (11.45)$$

With (11.41) we thus obtain the following transformation relations:

$$\begin{aligned} \xi &= (\cos\omega \cos\Omega - \sin\omega \sin\Omega \cos i) x \\ &\quad + (\cos\omega \sin\Omega + \sin\omega \cos\Omega \cos i) y + (\sin\omega \sin i) z \\ \eta &= (-\sin\omega \cos\Omega - \cos\omega \sin\Omega \cos i) x \\ &\quad + (-\sin\omega \sin\Omega + \cos\omega \cos\Omega \cos i) y + (\cos\omega \sin i) z \\ 0 &= (\sin\Omega \sin i) x - (\cos\Omega \sin i) y + (\cos i) z \end{aligned} \quad (11.46)$$

where we have used the fact that $\zeta = 0$. The third relation indicates that the satellite moves in a plane through the origin of the reference frame. The first two relations will be used in Section 11.10.

11.9. Transformation from spherical coordinates to orbital elements

Another important type of transformation concerns the calculation of orbital elements from the spherical position and velocity components $r, \alpha, \delta, V, \gamma, \psi$. For this type of transformation, we can use a number of equations derived in Chapter 5 and Chapter 6. For completeness, these equations are repeated here. For the semi-major axis and the eccentricity, the following relations were derived in Section 5.7 and Section 6.1:

$$a = \frac{r}{2 - r V^2 / \mu} \quad (11.47)$$

$$e^2 = 1 - \frac{r V^2}{\mu} \left(2 - \frac{r V^2}{\mu} \right) \cos^2 \gamma \quad (11.48)$$

In Section 6.5, the following relations were derived:

$$r = a (1 - e \cos E) \quad (11.49)$$

$$\dot{E} (1 - e \cos E) = \sqrt{\frac{\mu}{a^3}} \quad (11.50)$$

From (11.49) we obtain

$$e \cos E = 1 - \frac{r}{a} \quad (11.51-1)$$

Differentiation of (11.49) to time results in

$$\dot{r} = V \sin \gamma = a e \dot{E} \sin E \quad (11.52)$$

Substitution of (11.49) and (11.50) into (11.52) yields

$$e \sin E = \frac{r V \sin \gamma}{\sqrt{\mu a}} \quad (11.51-2)$$

From (11.51-1) and (11.51-2) E can be determined:

$$E = \arctan \left[\sqrt{\frac{a}{\mu}} \frac{r V \sin \gamma}{a - r} \right] \quad (11.53)$$

Because, by definition, $e \geq 0$, we can use (11.51-2) to determine in which quadrant E is located, which means that E can be determined unambiguously.

The true anomaly, θ , follows from

$$\tan \frac{\theta}{2} = \sqrt{\frac{1+e}{1-e}} \tan \frac{E}{2} \quad (11.54)$$

where $\theta/2$ and $E/2$ are located in the same quadrant (Section 6.5). The time of perigee passage, τ , follows from

$$\tau = t - \sqrt{\frac{a^3}{\mu}} (E - e \sin E) \quad (11.55)$$

where t is the moment of time for which the transformation is executed.

For the determination of the orbital elements i, ω, Ω , we start from the classical relations from spherical trigonometry, applied in a right-angle spherical triangle, as listed in Section 11.8. Using these relations for the spherical triangle shown in Figure 11.11, we can derive the following expressions:

$$\begin{aligned} \cos(\omega + \theta) &= \cos \delta \cos(\alpha - \Omega) \\ \sin(\omega + \theta) \cos i &= \cos \delta \sin(\alpha - \Omega) \\ \sin(\omega + \theta) \cos \psi &= \cos(\alpha - \Omega) \sin \delta \\ \sin \delta &= \sin i \sin(\omega + \theta) \\ \sin(\alpha - \Omega) &= \sin \psi \sin(\omega + \theta) \end{aligned} \quad (11.56)$$

This set is an extension of the set (11.29). From (11.56-2) and (11.56-5) it follows that

$$\cos i = \cos \delta \sin \psi \quad (11.57)$$

Because, by definition, $0^\circ \leq i < 180^\circ$, from this relation i can be solved unambiguously. From (11.56-3) and (11.56-4) follows

$$\cos(\alpha - \Omega) = \frac{\cos \psi}{\sin i} \quad (11.58-1)$$

From (11.56-2) and (11.56-4) follows

$$\sin(\alpha - \Omega) = \tan \delta \cot i \quad (11.58-2)$$

From (11.56-4) follows

$$\sin(\omega + \theta) = \frac{\sin \delta}{\sin i} \quad (11.59-1)$$

From (11.56-1) and (11.58-1) follows

$$\cos(\omega + \theta) = \cos \delta \frac{\cos \psi}{\sin i} \quad (11.59-2)$$

The orbital elements Ω and ω can be determined unambiguously from the relations (11.58) and (11.59), respectively.

11.10. Transformation from rectangular coordinates to orbital elements

In this Section, relations will be derived that express the orbital elements in terms of the rectangular coordinates and velocity components $x, y, z, \dot{x}, \dot{y}, \dot{z}$ at a certain time. First, the local distance and velocity of the satellite are computed from

$$r = \sqrt{x^2 + y^2 + z^2} \quad (11.60)$$

$$V = \sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2} \quad (11.61)$$

The semi-major axis can now be computed from (11.47). For the flight path angle, γ , we have

$$\bar{r} \cdot \bar{V} = r V \sin \gamma$$

or

$$\sin \gamma = \frac{x \dot{x} + y \dot{y} + z \dot{z}}{r V}$$

Substitution of this relation into (11.51-2) yields

$$e \sin E = \sqrt{\frac{1}{\mu a}} (x \dot{x} + y \dot{y} + z \dot{z}) \quad (11.62)$$

From (11.51-1) and (11.62) the parameters e and E can be computed unambiguously. Then, using (11.54) and (11.55), it is possible to compute the parameters θ and τ .

The angular momentum can be computed from (11.37). For the components of the angular momentum along the X , Y and Z axes we can write according to Figure 11.13:

$$\begin{aligned} H_z &= H \cos i = x \dot{y} - y \dot{x} \\ H_x &= H \sin \Omega = H \sin i \sin \Omega = y \dot{z} - z \dot{y} \end{aligned} \quad (11.63)$$

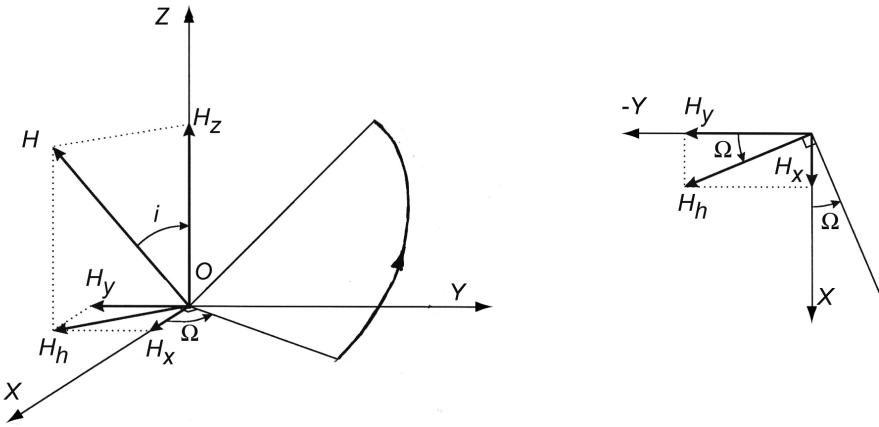


Figure 11.13: The angular momentum vector and its components.

$$H_y = H_h \cos \Omega = H \sin i \cos \Omega = x \dot{z} - z \dot{x} \quad (11.63)$$

From (11.63-1) we obtain

$$i = \arccos\left(\frac{x \dot{y} - y \dot{x}}{H}\right) \quad (11.64)$$

which gives an unambiguous solution for i , since $0^\circ \leq i < 180^\circ$. Subsequently, we write (11.63-2) and (11.63-3) as

$$\sin \Omega = \frac{y \dot{z} - z \dot{y}}{H \sin i} ; \quad \cos \Omega = \frac{x \dot{z} - z \dot{x}}{H \sin i} \quad (11.65)$$

These two relations yield an unambiguous solution for Ω . From (11.30-3) follows that

$$\sin(\omega + \theta) = \frac{z}{r \sin i} \quad (11.66-1)$$

In order to find an expression for $\cos(\omega + \theta)$, we start from (11.46). Multiplying (11.46-1) by $\cos \omega$ and (11.46-2) by $-\sin \omega$, and subsequently adding the resulting expressions leads to

$$\xi \cos \omega - \eta \sin \omega = x \cos \Omega + y \sin \Omega$$

Substitution of (11.26) into this relation yields

$$\cos(\omega + \theta) = \frac{x}{r} \cos \Omega + \frac{y}{r} \sin \Omega \quad (11.66-2)$$

With the expressions (11.66) an unambiguous solution for $(\omega + \theta)$ can be obtained, after which, because θ is known, the value of ω can be determined unambiguously.

In a slightly different way, the values of the angles i , Ω , ω , θ and u can be obtained from the following vector relations:

$$\cos i = \frac{\bar{e}_z \cdot \bar{H}}{H} \quad (11.67)$$

$$\bar{n}_{AN} = \bar{e}_z \times \bar{H} ; \quad n_{AN} = H \sin i$$

$$\begin{aligned}
 \cos\Omega &= \frac{\bar{e}_x \cdot \bar{n}_{AN}}{n_{AN}} && \text{if } (n_{AN})_y < 0 \text{ then } \Omega = 360^\circ - \Omega \\
 \cos\omega &= \frac{\bar{n}_{AN} \cdot \bar{e}}{n_{AN} e} && \text{if } (e)_z < 0 \text{ then } \omega = 360^\circ - \omega \\
 \cos\theta &= \frac{\bar{e} \cdot \bar{r}}{e r} && \text{if } \bar{r} \cdot \bar{V} < 0 \text{ then } \theta = 360^\circ - \theta \\
 \cos u &= \frac{\bar{n}_{AN} \cdot \bar{r}}{n_{AN} r} && \text{if } (r)_z < 0 \text{ then } u = 360^\circ - u
 \end{aligned} \tag{11-67}$$

where \bar{e}_x and \bar{e}_z are unit vectors in the X - and Z -directions, \bar{n}_{AN} is the so-called *ascending node vector*, i.e. a vector pointing towards the ascending node, $(\bar{n}_{AN})_y$ is the component of the ascending node vector in the Y -direction, \bar{e} is the eccentricity vector introduced in Section 5.7, $(\bar{e})_z$ is the component of the eccentricity vector in the Z -direction, and e is the eccentricity (magnitude of the eccentricity vector). Don't confuse the notation e for unit vectors and for the orbital eccentricity! When applying these relations it is assumed that the arccosine-function returns an angle between 0° and 180° . The expressions show that if i approaches 0° or 180° , the ascending node vector, \bar{n}_{AN} , approaches zero. So, if $i = 0^\circ$ or 180° , then Ω , ω and u are undefined. If $e = 0$, then ω and θ are undefined. These conclusions were already mentioned in Section 11.5.

11.11. *f*- and *g*-series

In the previous Sections it has been shown that when position and velocity of the spacecraft at a certain moment of time are known in either rectangular or spherical coordinates, the orbital elements can be computed. For any other time, position and velocity of the spacecraft can be computed from the orbital elements and the true anomaly by executing the reverse transformation. In this computation scheme, one implicitly uses the fact that the orbital elements are constants and thus have the same value at any moment of time. However, in this scheme we have to know the type of orbit, because $a > 0$, $e < 1$ for an ellipse, $a = \infty$, $e = 1$ for a parabola, $a < 0$, $e > 1$ for a hyperbola, and because the expression for the relation between θ and t is different for an ellipse, parabola or hyperbola. Various attempts have been made to provide sets of *universal variables* and formulas that can be used with all kinds of two-body orbital motion, even for the case of rectilinear motion. Pioneers in this field were K.F. Sundman (1873-1949), K. Stumpff (1895-1970), W.H. Goodyear (-) and S. Herrick (1911-1974), and some elegant mathematical formulations have been achieved. However, the esthetic attractions of the universal approach have to be weighed against a number of computational disadvantages, and in many occasions it is preferable to stick to the classical expressions for the elliptical, parabolic and hyperbolic orbits. In those cases where we do not know whether the orbit is a highly eccentric ellipse ($e > 0.995$) or a low-eccentricity hyperbola ($e < 1.005$), we then consider the orbit as a parabola and apply the relations derived for parabolic orbits. Therefore, in this book we will not present and discuss the universal parameters and formulations.

If the moment of time for which we want to compute position and velocity of the spacecraft is reasonably close to the moment of time on which position and velocity are known, then we can apply an alternative direct method that does not require knowledge about the type of orbit and

does not require the computation of orbital elements except for the semi-major axis. This is the method of the so-called *f- and g-series*, which will be treated in this Section.

Assume that at t_0 the position and velocity vectors \bar{r}_0 and $\dot{\bar{r}}_0$ are known. For the position vector \bar{r} at another moment of time, t , we may write with a Taylor series expansion

$$\bar{r} = \bar{r}_0 + \dot{\bar{r}}_0 \Delta t + \frac{1}{2} \ddot{\bar{r}}_0 (\Delta t)^2 + \frac{1}{6} \bar{r}_0^{(3)} (\Delta t)^3 + \frac{1}{24} \bar{r}_0^{(4)} (\Delta t)^4 + \frac{1}{120} \bar{r}_0^{(5)} + \dots \quad (11.68)$$

in which $\Delta t = t - t_0$, and the indices ⁽³⁾, ⁽⁴⁾ and ⁽⁵⁾ indicate third-, fourth- and fifth-order derivatives, respectively. From Newton's laws, we know that

$$\ddot{\bar{r}} = -\frac{\mu}{r^3} \bar{r} \quad (11.69-1)$$

By repeated differentiation of this equation, the higher-order derivatives in (11.68) can be computed:

$$\bar{r}^{(3)} = \frac{3\mu \dot{r}}{r^4} \bar{r} - \frac{\mu}{r^3} \dot{\bar{r}} \quad (11.69-2)$$

$$\bar{r}^{(4)} = -\frac{12\mu \dot{r}^2}{r^5} \bar{r} + \frac{3\mu \ddot{r}}{r^4} \bar{r} + \frac{6\mu \dot{r}}{r^4} \dot{\bar{r}} - \frac{\mu}{r^3} \ddot{\bar{r}} \quad (11.69-3)$$

$$\bar{r}^{(5)} = \frac{60\mu \dot{r}^3}{r^6} \bar{r} - \frac{36\mu \dot{r} \ddot{r}}{r^5} \bar{r} - \frac{36\mu \dot{r}^2}{r^5} \dot{\bar{r}} + \frac{3\mu \bar{r}^{(3)}}{r^4} \bar{r} + \frac{9\mu \ddot{r}}{r^4} \dot{\bar{r}} + \frac{9\mu \dot{r}}{r^4} \ddot{\bar{r}} - \frac{\mu}{r^3} \bar{r}^{(3)} \quad (11.69-4)$$

The position and velocity vectors, \bar{r} and $\dot{\bar{r}}$ are known at t_0 . In order to rewrite (11.68) in terms of \bar{r}_0 , $\dot{\bar{r}}_0$ and Δt , we first combine two relations found in Section 5.2 and Section 6.3:

$$\ddot{\bar{r}} - r \dot{\theta}^2 = -\frac{\mu}{r^2} \quad ; \quad V^2 = \dot{r}^2 + r^2 \dot{\theta}^2 = \mu \left(\frac{2}{r} - \frac{1}{a} \right)$$

and write

$$\ddot{\bar{r}} = \frac{\mu}{r^2} - \frac{\mu}{r a} - \frac{\dot{r}^2}{r} \quad (11.70-1)$$

Differentiation yields

$$\bar{r}^{(3)} = -\frac{4\mu \dot{r}}{r^3} + \frac{3\mu \dot{r}}{r^2 a} + \frac{3\dot{r}^3}{r^2} \quad (11.70-2)$$

Substitution of (11.69-1), (11.69-2) and (11.70) into (11.69-3) and (11.69-4) leads to

$$\bar{r}^{(4)} = \left[-\frac{15\mu \dot{r}^2}{r^5} + \frac{4\mu^2}{r^6} - \frac{3\mu^2}{r^5 a} \right] \bar{r} + \frac{6\mu \dot{r}}{r^4} \dot{\bar{r}} \quad (11.71-1)$$

$$\bar{r}^{(5)} = \left[\frac{105\mu\dot{r}^3}{r^6} - \frac{60\mu^2\dot{r}}{r^7} + \frac{45\mu^2\dot{r}}{r^6a} \right] \bar{r} + \left[-\frac{45\mu\dot{r}^2}{r^5} + \frac{10\mu^2}{r^6} - \frac{9\mu^2}{r^5a} \right] \dot{\bar{r}} \quad (11.71-2)$$

Substitution of (11.69-1), (11.69-2), (11.71-1) and (11.71-2) into (11.68), and evaluation of the resulting expression at t_0 , gives

$$\begin{aligned} \bar{r} = & \left[1 - \frac{1}{2} \frac{\mu}{r_0^3} (\Delta t)^2 + \frac{1}{2} \frac{\mu\dot{r}_0}{r_0^4} (\Delta t)^3 + \frac{1}{2} \frac{\mu^2}{r_0^6} \left(\frac{1}{3} - \frac{1}{4} \frac{r_0}{a} - \frac{5}{4} \frac{r_0\dot{r}_0^2}{\mu} \right) (\Delta t)^4 + \right. \\ & \left. - \frac{1}{2} \frac{\mu^2\dot{r}_0}{r_0^7} \left(1 - \frac{3}{4} \frac{r_0}{a} - \frac{7}{4} \frac{r_0\dot{r}_0^2}{\mu} \right) (\Delta t)^5 + O(\Delta t)^6 \right] \bar{r}_0 + \\ & \left[\Delta t - \frac{1}{6} \frac{\mu}{r_0^3} (\Delta t)^3 + \frac{1}{4} \frac{\mu\dot{r}_0}{r_0^4} (\Delta t)^4 + \right. \\ & \left. + \frac{1}{4} \frac{\mu^2}{r_0^6} \left(\frac{1}{3} - \frac{3}{10} \frac{r_0}{a} - \frac{3}{2} \frac{r_0\dot{r}_0^2}{\mu} \right) (\Delta t)^5 + O(\Delta t)^6 \right] \dot{\bar{r}}_0 \end{aligned} \quad (11.72)$$

Because a is solely a function of r and V , and is constant for a Keplerian orbit, a can easily be computed from (11.47) and the known values r_0 and V_0 . Expression (11.72) can be written as

$$\bar{r} = f\bar{r}_0 + g\dot{\bar{r}}_0 \quad (11.73)$$

where the series f and g can be written as

$$f = \sum_{n=0}^{\infty} f_n (\Delta t)^n \quad ; \quad g = \sum_{n=0}^{\infty} g_n (\Delta t)^n \quad (11.74)$$

where the coefficients f_n, g_n are given by

$$\begin{aligned} f_0 = g_1 = 1 & \quad ; \quad f_1 = g_0 = g_2 = 0 \quad ; \quad f_2 = -\frac{1}{2} \frac{\mu}{r_0^3} \\ f_3 = \frac{1}{2} \frac{\mu\dot{r}_0}{r_0^4} & \quad ; \quad f_4 = \frac{1}{2} \frac{\mu^2}{r_0^6} \left(\frac{1}{3} - \frac{1}{4} \frac{r_0}{a} - \frac{5}{4} \frac{r_0\dot{r}_0^2}{\mu} \right) \\ f_5 = -\frac{1}{2} \frac{\mu^2\dot{r}_0}{r_0^7} \left(1 - \frac{3}{4} \frac{r_0}{a} - \frac{7}{4} \frac{r_0\dot{r}_0^2}{\mu} \right) & \\ g_3 = -\frac{1}{6} \frac{\mu}{r_0^3} = \frac{1}{3} f_2 & \quad ; \quad g_4 = \frac{1}{4} \frac{\mu\dot{r}_0}{r_0^4} = \frac{1}{2} f_3 \end{aligned} \quad (11.75)$$

$$g_5 = \frac{1}{4} \frac{\mu^2}{r_0^6} \left(\frac{1}{3} - \frac{3}{10} \frac{r_0}{a} - \frac{3}{2} \frac{r_0 \dot{r}_0^2}{\mu} \right) = \frac{3}{5} \left(f_4 - \frac{1}{9} f_2^2 \right) \quad (11.75)$$

etc. Equation (11.73) was first used by J.L. Lagrange (1736-1813) in 1869, who gave explicit expressions for the coefficients f_n , g_n up to $n = 5$. Since then, a number of investigators have given explicit expressions for the coefficients up to order 27 and higher. We conclude from the relations (11.75) that g_3 , g_4 and g_5 can be expressed as functions of f_2 , f_3 and f_4 . This is a general property of the f_n - and g_n -terms: g_n can always be expressed as a series of f_m -terms, with $m = 2, 3, \dots, n-1$. The f - and g -terms can also be computed by using a recursive algorithm:

$$f_{n+1} = \frac{1}{n+1} \left(\frac{df_n}{dt} - \frac{\mu}{r^3} g_n \right) ; \quad g_{n+1} = \frac{1}{n+1} \left(\frac{dg_n}{dt} + f_n \right) \quad (11.76)$$

where the starting values are given by $f_0 = 1$, $g_0 = 0$. During the differentiation of f_n and g_n to time, formally every term has to be considered as a variable. After differentiation the index 0 has to be added again. The relations (11.76) are known as the *recursive formulas of Cipolletti* (D. Cipolletti, 1840-1874) and were discovered by him around 1872. Note that the Cipolletti formulation is recursive in only a limited sense, because the coefficients of a given order depend upon the time-derivatives of the coefficients of the previous order and there is no recursive formulation for these derivatives in Cipolletti's method.

Analogous to (11.73), the velocity vector can be written as

$$\dot{\vec{r}} = \dot{\vec{f}} \vec{r}_0 + \dot{\vec{g}} \vec{r}_0 \quad (11.77)$$

where for the series $\dot{\vec{f}}$ and $\dot{\vec{g}}$ holds

$$\dot{\vec{f}} = \sum_{n=1}^{\infty} n f_n (\Delta t)^{n-1} ; \quad \dot{\vec{g}} = \sum_{n=1}^{\infty} n g_n (\Delta t)^{n-1} \quad (11.78)$$

F.R. Moulton (1872-1952) has studied around 1903 the convergence of these series expansions. The practical usefulness of the series ends, however, when terms higher than f_5 and g_5 have to be taken into account, because these higher-order terms are fairly complicated. Roughly, it can be stated that the f - and g -series are only applicable if

$$\frac{\mu}{r_0^3} (\Delta t)^2 < 10^{-2} \quad (11.79)$$

An important advantage of these f - and g -series is that they are universal; i.e. they can be applied regardless whether the orbit is an ellipse, parabola or hyperbola. When the condition (11.79) is satisfied, the f - and g -series provide a quick, efficient method for the computation of position and velocity of the spacecraft, and they are extensively applied in the determination and prediction of satellite orbits, but also in computing rendez-vous orbits.

An interesting property of the f - and g -series can be found. Because in a Keplerian orbit the angular momentum of the motion is constant, we have

$$\bar{H} = \vec{r}_0 \times \dot{\vec{r}}_0 = \vec{r} \times \dot{\vec{r}}$$

Substitution of (11.73) and (11.77) into this expression yields

$$f\dot{g} - \dot{f}g = 1 \quad (11.80)$$

This relation can be used to verify the expressions derived for higher-order f_n - and g_n -terms.

It is also possible to derive closed-form expressions for the f - and g -series. We then start by writing the vectors \bar{r} and $\dot{\bar{r}}$ relative to the non-rotating reference frame $\xi\eta\zeta$ defined in Section 11.8 (Figure 11.10), where the ζ -axis is along the angular momentum vector perpendicular to the orbital plane and the ξ -axis points towards perigee. Because the satellite is always in the orbital plane: $\zeta = \dot{\zeta} = 0$, and

$$\bar{r} = \xi \bar{e}_\xi + \eta \bar{e}_\eta \quad ; \quad \dot{\bar{r}} = \dot{\xi} \bar{e}_\xi + \dot{\eta} \bar{e}_\eta \quad (11.81)$$

where \bar{e}_ξ and \bar{e}_η are unit vectors in the ξ and η directions, respectively. Taking the vector product of (11.73) and $\dot{\bar{r}}_0$ yields

$$\bar{r} \times \dot{\bar{r}}_0 = f \bar{r}_0 \times \dot{\bar{r}}_0 \quad (11.82)$$

Substituting (11.81) for \bar{r} and $\dot{\bar{r}}_0$ on the left-hand side of (11.82), and using the well-known relation for the orbital angular momentum

$$\bar{H} = \bar{r}_0 \times \dot{\bar{r}}_0 = \sqrt{\mu p} \bar{e}_\zeta$$

where \bar{e}_ζ is the unit vector in the ζ -direction, we find from (11.82)

$$f = \frac{\xi \dot{\eta}_0 - \eta \dot{\xi}_0}{\sqrt{\mu p}} \quad (11.83)$$

Taking the vector product of (11.73) and \bar{r}_0 yields

$$\bar{r} \times \bar{r}_0 = g \dot{\bar{r}}_0 \times \bar{r}_0 \quad (11.84)$$

Again substituting (11.81) and the relation for the angular momentum, we find from (11.84):

$$g = \frac{\eta \xi_0 - \xi \eta_0}{\sqrt{\mu p}} \quad (11.85)$$

Combination of (6.31), (6.32) and (11.26) results for an elliptical orbit in

$$\xi = r \cos \theta = a (\cos E - e) \quad ; \quad \eta = r \sin \theta = a \sqrt{1 - e^2} \sin E \quad (11.86)$$

Differentiation of (11.86) leads to

$$\dot{\xi} = -a \dot{E} \sin E \quad ; \quad \dot{\eta} = a \sqrt{1 - e^2} \dot{E} \cos E \quad (11.87)$$

Combination of (11.49) and (11.50) yields

$$\dot{E} = \frac{1}{r} \sqrt{\frac{\mu}{a}} \quad (11.88)$$

Substitution of (11.88) into (11.87) gives

$$\dot{\xi} = -\frac{1}{r} \sqrt{\mu a} \sin E \quad ; \quad \dot{\eta} = \frac{1}{r} \sqrt{\mu a (1 - e^2)} \cos E \quad (11.89)$$

Substitution of (11.86) and (11.89) into (11.83), and elaboration of the resulting equation leads to

$$f = \frac{a}{r_0} [(1 - e \cos E_0) - \{1 - \cos(E - E_0)\}]$$

or, with (6.33),

$$f = 1 - \frac{a}{r_0} [1 - \cos(E - E_0)] \quad (11.90)$$

Substitution of (11.86) into (11.85), and elaboration of the resulting equation leads to

$$g = \sqrt{\frac{a^3}{\mu}} [\sin(E - E_0) - e(\sin E - \sin E_0)] \quad (11.91)$$

According to (6.36):

$$M - M_0 = \sqrt{\frac{\mu}{a^3}} (t - t_0) = (E - E_0) - e(\sin E - \sin E_0) \quad (11.92)$$

Substitution of (11.92) into (11.91), finally, leads to

$$g = (t - t_0) - \sqrt{\frac{a^3}{\mu}} [(E - E_0) - \sin(E - E_0)] \quad (11.93)$$

Differentiation of (11.90) and (11.93) with respect to time yields

$$\dot{f} = -\frac{a}{r_0} \dot{E} \sin(E - E_0) \quad ; \quad \dot{g} = 1 - \sqrt{\frac{a^3}{\mu}} \dot{E} [1 - \cos(E - E_0)]$$

or, with (11.88),

$$\dot{f} = -\frac{\sqrt{\mu a}}{r r_0} \sin(E - E_0) \quad (11.94)$$

$$\dot{g} = 1 - \frac{a}{r} [1 - \cos(E - E_0)] \quad (11.95)$$

If \bar{r}_0 and $\dot{\bar{r}}_0$ are known, then it is possible to compute the orbital elements a and e , which are a function of r_0 , V_0 and γ_0 only (Section 11.9). Then, we solve (6.36) for $(E - E_0)$, substitute that value into (11.90), (11.93) to (11.95) and obtain values of f , g , \dot{f} and \dot{g} . These values may be used in (11.73) and (11.77) to obtain the position and velocity vectors at another time. These closed-form expressions find extensive use in preliminary orbit determination procedures. It is emphasized that the application of the closed-form expressions for f , g , \dot{f} and \dot{g} instead of the series expansions derived earlier, results in a loss of generality; they hold only for an elliptical orbit while the series expansions do not require any knowledge about the type of Keplerian orbit. By a similar process, closed-form expressions can also be derived for parabolic and hyperbolic

orbits.

12. TRANSFER BETWEEN TWO COPLANAR ORBITS

An important topic in spaceflight is the application of *transfer orbits*, where satellites are transferred from a given initial orbit around the Earth to a specified final orbit. The initial and final orbits can be either circular or elliptical. Obviously, many different transfer orbits can be flown. Usually, one will be interested in optimum transfer orbits, i.e. transfer orbits that require a minimum amount of propellant. In this Chapter, several examples of transfer orbits will be analyzed, where it is assumed that the initial orbit, the final orbit and the transfer orbit are all in the same plane (coplanar) and that the required maneuvers are executed by impulsive shots (Section 1.7). So, the analysis holds for maneuvers with traditional chemical rocket propulsion systems and definitely not for low-thrust electric propulsion systems (Chapter 19). We will apply an analysis method very similar to the method developed by M. Vertregt (1897-1973) around 1957 to analyze interplanetary trajectories.

A basic assumption in this Chapter, and in some subsequent Chapters, is that the sequence of maneuvers is accomplished by a single rocket stage using the same rocket engine with a constant (effective) exhaust velocity. Using *Tsiolkovski's law* ((1.49)) for the sequence of impulsive shots, we then find

$$\Delta V_{tot} = \sum_{i=1}^n (\Delta V)_i = V_j \sum_{i=1}^n \ln \frac{M_{0_i}}{M_{e_i}}$$

where $M_{0,i}$ is the rocket stage mass before the i^{th} impulsive shot, $M_{e,i}$ is the mass after the i^{th} impulsive shot, V_j is the (effective) exhaust velocity, and $(\Delta V)_i$ is the velocity increase produced by the i^{th} impulsive shot. When it is also assumed that in between successive impulsive shots no mass is rejected or expelled, we find

$$\Delta V_{tot} = V_j \left[\ln \frac{M_{0_1}}{M_{e_1}} + \ln \frac{M_{e_1}}{M_{e_2}} + \ln \frac{M_{e_2}}{M_{e_3}} + \dots \right] = V_j \ln \frac{M_0}{M_e} \quad (12.1)$$

where M_0 is the stage mass just before the first impulsive shot and M_e is the mass after the final impulsive shot. In Section 1.7 it was shown that, for a given initial mass, M_0 , and construction mass, M_c , a maximum payload mass, M_l , corresponds to a minimum value of M_0/M_e . From (12.1) then follows that the payload mass is maximum, and so the propellant consumption is minimum, when

$$\Delta V_{tot} = \text{minimum} \quad (12.2)$$

We will use this criterion to determine the optimum transfer orbit.

When the total transfer maneuver is performed by more than one rocket stage with rocket engines having different exhaust velocities, and when in between the impulsive shots mass is rejected (staging) or expelled (residual propellant dump), then criterion (12.2) obviously does not hold. In that case, an optimization analysis, in which the different exhaust velocities and mass distributions of the various rocket stages are accounted for, has to be performed in order to find the maximum payload mass, M_l , for a given initial mass, M_0 . Such an optimization analysis is, however, beyond the scope of this book.

12.1. Optimum transfer between two circular orbits

Application of the cosine rule in the velocity vector diagrams indicated in Figure 12.1 results in

$$\begin{aligned}\Delta V_1^2 &= V_{c_1}^2 + V_1^2 - 2V_1 V_{c_1} \cos \gamma_1 \\ \Delta V_2^2 &= V_{c_2}^2 + V_2^2 - 2V_2 V_{c_2} \cos \gamma_2\end{aligned}\quad (12.3)$$

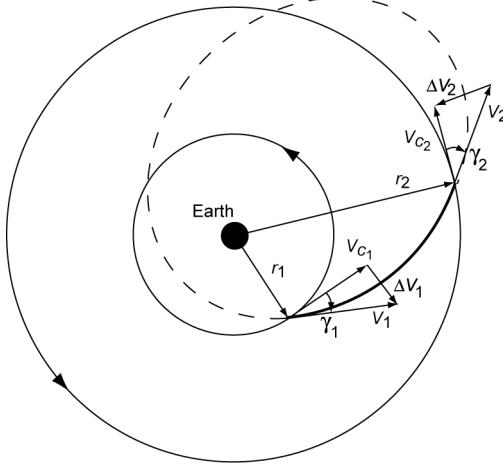


Figure 12.1: Geometry of a transfer orbit between two coplanar circular orbits, and the notation used.

In Chapters 5 and 6 we have derived expressions for V_1 , V_{c_1} , V_2 , V_{c_2} , γ_1 and γ_2 . For instance, in Chapter 6 we have found

$$p = a(1 - e^2) \quad ; \quad V^2 = \mu \left(\frac{2}{r} - \frac{1}{a} \right) \quad ; \quad V_c^2 = \frac{\mu}{r} \quad (12.4)$$

From these expression, we obtain

$$V_1^2 = V_{c_1}^2 \left(2 - \frac{r_1}{p} (1 - e^2) \right) \quad ; \quad V_2^2 = V_{c_2}^2 \left(2 - \frac{r_2}{p} (1 - e^2) \right)$$

With the notation

$$n = \frac{r_2}{r_1} \quad ; \quad z = \frac{p}{r_1} \quad (12.5)$$

and the relation

$$(V_{c_2})^2 = \frac{1}{n} (V_{c_1})^2$$

we find

$$V_1^2 = V_{c_1}^2 \left(2 - \frac{1 - e^2}{z} \right) \quad ; \quad V_2^2 = V_{c_1}^2 \left(\frac{2}{n} - \frac{1 - e^2}{z} \right) \quad (12.6)$$

Combination of (5.21), (5.25), (6.2) and (6.21) yields

$$\cos^2 \gamma = \frac{H^2}{r^2 V^2} = \frac{\mu p}{r^2 V^2} = \frac{p^2/r^2}{2p/r - p/a} = \frac{p^2/r^2}{2p/r - (1 - e^2)}$$

from which, after substitution of (12.5), we find

$$\cos \gamma_1 = \frac{z}{\sqrt{2z - (1 - e^2)}} ; \quad \cos \gamma_2 = \frac{z/n}{\sqrt{2z/n - (1 - e^2)}} \quad (12.7)$$

Substitution of (12.6) and (12.7) into (12.3) yields

$$\begin{aligned} \Delta V_1^2 &= V_{c_1}^2 \left[3 - 2\sqrt{z} - \frac{1 - e^2}{z} \right] \\ \Delta V_2^2 &= V_{c_1}^2 \left[\frac{3 - 2\sqrt{z/n}}{n} - \frac{1 - e^2}{z} \right] \end{aligned} \quad (12.8)$$

Consequently, for the total impulse required for this transfer flight we find

$$\frac{\Delta V_{tot}}{V_{c_1}} = \frac{\Delta V_1 + \Delta V_2}{V_{c_1}} = \sqrt{3 - 2\sqrt{z} - \frac{1 - e^2}{z}} + \sqrt{\frac{3 - 2\sqrt{z/n}}{n} - \frac{1 - e^2}{z}} \quad (12.9)$$

For an optimum transfer orbit, under the conditions mentioned before, one is obviously interested in the case that

$$\frac{\Delta V_{tot}}{V_{c_1}} = \text{minimum}$$

For any transfer orbit the following conditions must hold:

$$a(1 - e) \leq r_1 ; \quad a(1 + e) \geq r_2 \quad (12.10)$$

Substitution of (12.4) and (12.5) into (12.10) results in

$$z \leq 1 + e ; \quad z \geq n(1 - e) \quad (12.11)$$

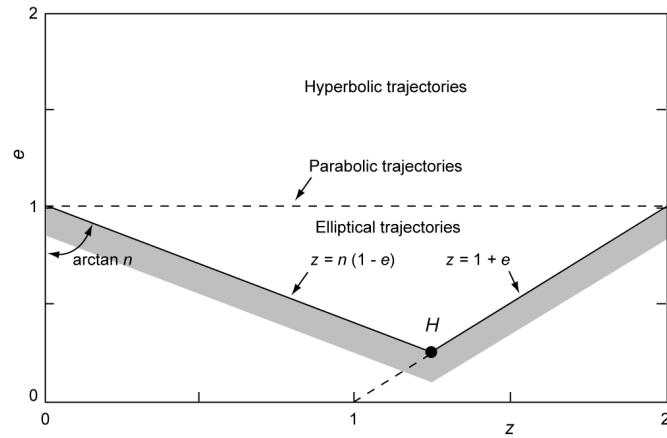


Figure 12.2: Region of possible transfer orbits between two coplanar circular orbits in the ez -plane.

Figure 12.2 shows the region in an ez -diagram where transfer orbits are feasible. We conclude that elliptical ($e < 1$) as well as parabolic ($e = 1$) and hyperbolic ($e > 1$) transfer orbits are feasible.

From the structure of (12.9) we note that for given values of z and n , the quantity $\Delta V_{tot}/V_{c_1}$

decreases when e decreases. The point in the ez -diagram that, for given values of z and n , represents a local minimum-energy transfer orbit therefore is located on the lines $z = n(1-e)$ or $z = 1+e$. We now examine the conditions on the boundary line $z = n(1-e)$. On this line:

$$e = 1 - \frac{z}{n} \quad (12.12)$$

Substitution of this relation into (12.9) leads to

$$\frac{\Delta V_{tot}}{V_{c_1}} = \sqrt{3 - 2\sqrt{z} - \frac{2}{n} + \frac{z}{n^2}} + \sqrt{\frac{1 - 2\sqrt{z/n}}{n} + \frac{z}{n^2}}$$

Differentiation of this relation to z gives

$$\frac{d}{dz} \left(\frac{\Delta V_{tot}}{V_{c_1}} \right) = \frac{1}{2} \frac{\frac{1}{n^2} - \frac{1}{\sqrt{z}}}{\sqrt{3 - 2\sqrt{z} - \frac{2}{n} + \frac{z}{n^2}}} + \frac{1}{2} \frac{\frac{1}{n^2} - \frac{1}{n\sqrt{nz}}}{\sqrt{\frac{1 - 2\sqrt{z/n}}{n} + \frac{z}{n^2}}} \quad (12.13)$$

Since $e \geq 0$, (12.12) shows that on the line $z = n(1-e)$: $z \leq n$. Because the radius of the final orbit is larger than the radius of the initial orbit: $n > 1$. Then, (12.13) shows that

$$\frac{d}{dz} \left(\frac{\Delta V_{tot}}{V_{c_1}} \right) < 0 \quad \text{on} \quad z = n(1-e) \quad (12.14)$$

So, on the line $z = n(1-e)$ the quantity $\Delta V_{tot}/V_{c_1}$ decreases when z increases.

For the other boundary line, we have

$$e = z - 1 \quad (12.15)$$

Substitution into (12.9) yields

$$\frac{\Delta V_{tot}}{V_{c_1}} = \sqrt{3 - 2\sqrt{z} - (2-z)} + \sqrt{\frac{3 - 2\sqrt{z/n}}{n} - (2-z)}$$

and after differentiation to z we obtain

$$\frac{d}{dz} \left(\frac{\Delta V_{tot}}{V_{c_1}} \right) = \frac{1}{2} \frac{1 - \frac{1}{\sqrt{z}}}{\sqrt{3 - 2\sqrt{z} - (2-z)}} + \frac{1}{2} \frac{1 - \frac{1}{n\sqrt{nz}}}{\sqrt{\frac{3 - 2\sqrt{z/n}}{n} - (2-z)}} \quad (12.16)$$

Since $e \geq 0$, on the line $z = 1+e$: $z \geq 1$. Because also $n > 1$, we find from (12.16):

$$\frac{d}{dz} \left(\frac{\Delta V_{tot}}{V_{c_1}} \right) > 0 \quad \text{on} \quad z = 1 + e \quad (12.17)$$

So, on the line $z = 1+e$ the quantity $\Delta V_{tot}/V_{c_1}$ increases when z increases.

From (12.14) and (12.17) we conclude that point H in the ez -diagram indicates the transfer orbit with minimum propellant consumption. Obviously, this orbit is an ellipse with the lowest

eccentricity of all possible transfer orbits and for which

$$z = n(1 - e) = 1 + e \quad (12.18)$$

or

$$z = \frac{2n}{n+1} ; \quad e = \frac{n-1}{n+1} \quad (12.19)$$

Combination of (12.4) and (12.18) yields

$$a = \frac{r_2}{1+e} = \frac{r_1}{1-e}$$

This condition is, according to (6.4), met when r_2 is the apogee distance and r_1 is the perigee distance of the elliptical transfer orbit. We thus may conclude that the transfer orbit for which the propellant consumption is minimum, is an elliptical orbit that touches the initial orbit at perigee and the final orbit at apogee. For this orbit, the velocity vectors ΔV_1 and ΔV_2 are directed tangentially to the initial and the final orbits. This minimum-energy transfer orbit is called a *Hohmann transfer orbit*, after W. Hohmann (1880-1945) who has proposed it for interplanetary spaceflight (Chapter 18) in 1925. It is noted that Hohmann only conjectured that this orbit is the minimum-energy impulsive transfer orbit; the rigorous demonstration came some 30 years later! For a Hohmann transfer orbit, in Figure 12.2 indicated by point H , the values of z and n are given by (12.19). Observing the structure of (12.9), (12.13) and (12.16) we conclude that when $\Delta V_2 = 0$, and thus the satellite is not injected into the final orbit, a Hohmann transfer orbit is also the optimum (minimum propellant consumption) trajectory to reach a distance r_2 .

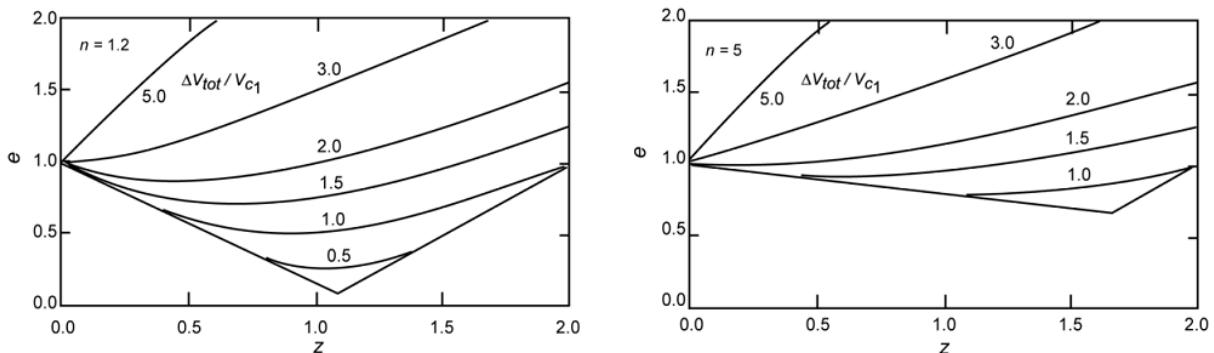


Figure 12.3: Curves of constant dimensionless total impulse $\Delta V_{tot}/V_{c1}$ for a transfer between two coplanar circular orbits, and for $n = 1.2$ and $n = 5$.

In Figure 12.3 several lines of constant $\Delta V_{tot}/V_{c1}$ are plotted in the ez -diagram for $n = 1.2$ (left) and $n = 5$ (right). When it is assumed that the altitude of the initial orbit is 250 km, then these values of n correspond to final circular orbital altitudes of about 1575 km and 26,750 km, respectively. We observe that, e.g. for $n = 1.2$ and $\Delta V_{tot}/V_{c1} = 1.5$, elliptical, parabolic and hyperbolic transfer orbits are feasible, which all require the same total impulse. The boundary line $z = 1 + e$ represents all orbits that touch the initial orbit ($\gamma_1 = 0^\circ$); the boundary line $z = n(1 - e)$ represents all orbits that touch the final orbit ($\gamma_2 = 0^\circ$). Therefore, for one particular value of $\Delta V_{tot}/V_{c1}$ larger than the value for the Hohmann transfer orbit, a transfer orbit can be flown either with $\gamma_1 = 0^\circ$, $\gamma_2 \neq 0^\circ$ or with $\gamma_1 \neq 0^\circ$, $\gamma_2 \neq 0^\circ$ or with $\gamma_1 \neq 0^\circ$, $\gamma_2 = 0^\circ$.

For a Hohmann transfer, the required impulsive shots can be computed easily. We then use a

standard computation scheme that will be used for most analyses of transfer orbits presented in this and following Chapters. It is based on the application of a number of relations derived in Sections 6.1 and 6.3 for elliptical orbits:

$$V_p = V_{c_p} \sqrt{1+e} \quad ; \quad V_a = V_{c_a} \sqrt{1-e} \quad (12.20)$$

$$e = \frac{r_a - r_p}{r_a + r_p} \quad (12.21)$$

where V_p and V_a are the velocities at perigee and at apogee, and V_{cp} and V_{ca} are the circular velocities at perigee and at apogee. For the two impulsive shots we find with (12.20)

$$\begin{aligned} \Delta V_1 &= V_{c_1} \sqrt{1+e} - V_{c_1} = V_{c_1} [\sqrt{1+e} - 1] \\ \Delta V_2 &= V_{c_2} - V_{c_2} \sqrt{1-e} = V_{c_1} \sqrt{\frac{1}{n} [1 - \sqrt{1-e}]} \end{aligned}$$

Substitution of (12.5) and (12.21) into these equations yields

$$\frac{\Delta V_1}{V_{c_1}} = \sqrt{\frac{2n}{n+1}} - 1 \quad (12.22)$$

$$\frac{\Delta V_2}{V_{c_1}} = \sqrt{\frac{1}{n} \left(1 - \sqrt{\frac{2}{n+1}} \right)} \quad (12.23)$$

$$\frac{\Delta V_{tot}}{V_{c_1}} = (n-1) \sqrt{\frac{2}{n(n+1)}} + \sqrt{\frac{1}{n}} - 1 \quad (12.24)$$

In Figure 12.4 the impulses $\Delta V_1/V_{c_1}$, $\Delta V_2/V_{c_1}$ and $\Delta V_{tot}/V_{c_1}$ are plotted as a function of n . An analysis of this Figure and the relations (12.22) through (12.24) shows that:

- For large values of n , $\Delta V_1/V_{c_1}$ asymptotically approaches the value 0.414. This corresponds with the impulsive shot that is needed to reach the escape velocity: $\Delta V_1/V_{c_1} = (\sqrt{2} - 1)$.
- The value of $\Delta V_2/V_{c_1}$ reaches a maximum of 0.19 for $n = 5.9$. The impulse that is needed to circularize the orbit at the final altitude therefore is maximum at this altitude.
- The dimensionless total impulse $\Delta V_{tot}/V_{c_1}$ reaches a maximum of 0.54 for $n = 15.6$.
- For values of $n > 3.3$, the dimensionless total impulse is larger than $(\sqrt{2} - 1)$. We therefore conclude that, for an initial orbital altitude of 250 km, it requires more propellant to reach a circular orbit at an altitude of $h > 15,488$ km than to accelerate from the initial orbit to escape velocity.

For the flight time in a Hohmann transfer orbit, we find from (6.25)

$$t_{f_H} = \pi \sqrt{\frac{a^3}{\mu}} = \pi \sqrt{\frac{(r_1 + r_2)^3}{8\mu}}$$

or, with (12.5),

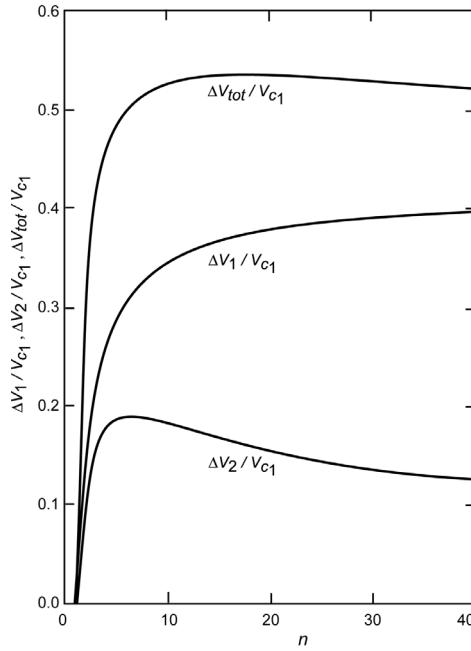


Figure 12.4: Required impulses $\Delta V_1/V_{c1}$, $\Delta V_2/V_{c1}$, $\Delta V_{tot}/V_{c1}$ as a function of n , for a Hohmann transfer between two coplanar circular orbits.

$$t_{f_H} = 2\pi \sqrt{\frac{r_1^3}{\mu}} \frac{1}{4} \sqrt{\frac{(n+1)^3}{2}} = \frac{1}{4} T_{c1} \sqrt{\frac{(n+1)^3}{2}} \quad (12.25)$$

where T_{c1} is the period of the original circular orbit. This equation shows that for large values of n the flight time in the Hohmann transfer orbit can become very long. For various missions this long flight time can be unacceptable and we then have to use faster transfer orbits, where it is accepted that these orbits demand more energy than the Hohmann transfer orbit.

12.2. Faster transfer between two circular orbits

In principle, an infinite number of transfer orbits between two given circular orbits can be flown. In this Section, only the case of a coplanar elliptical transfer orbit that touches the original circular orbit (Figure 12.5) and intersects the final circular orbit will be considered. This type of transfer is a realistic example of faster transfer orbits that do not require too much energy.

For this transfer, the following conditions hold:

- The perigee distance of the transfer orbit is equal to the radius of the initial orbit:

$$a(1 - e) = r_1 \quad (12.26)$$

where a and e are the orbital elements of the transfer orbit.

- At the intersection of the transfer orbit and the final circular orbit:

$$\frac{a(1 - e^2)}{1 + e \cos \theta_t} = r_2$$

where θ_t denotes the true anomaly at the point of intersection.

Analytical manipulation of these relations leads to

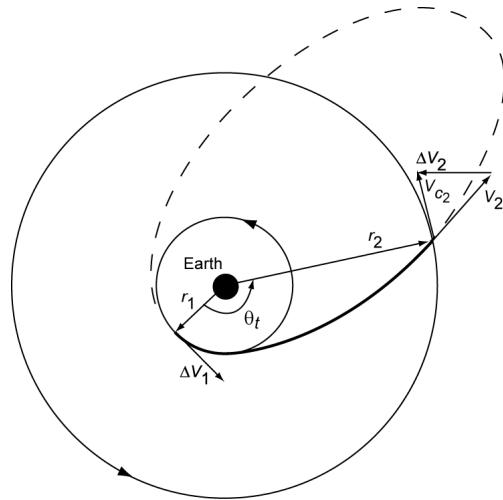


Figure 12.5: Geometry of a class of faster elliptical transfer orbits that are tangent to the original circular orbit.

$$a(1 - e^2) = p = r_1(1 + e) \quad ; \quad a(1 - e^2) = p = r_2(1 + e \cos \theta_t)$$

and, with (12.5),

$$z = 1 + e \quad ; \quad z = n(1 + e \cos \theta_t)$$

from which we find

$$e = \frac{n - 1}{1 - n \cos \theta_t} \quad (12.27)$$

$$z = \frac{n(1 - \cos \theta_t)}{1 - n \cos \theta_t} \quad (12.28)$$

From (12.26) and (12.27) follows

$$\frac{a}{r_1} = \frac{1 - n \cos \theta_t}{2 - n(1 + \cos \theta_t)} \quad (12.29)$$

Substitution of (12.27) and (12.28) into (12.9) yields

$$\begin{aligned} \frac{\Delta V_{tot}}{V_{c1}} &= \sqrt{3 - 2 \sqrt{\frac{n(1 - \cos \theta_t)}{1 - n \cos \theta_t} - \frac{2 - n(1 + \cos \theta_t)}{1 - n \cos \theta_t}}} \\ &\quad + \sqrt{\frac{3}{n} - \frac{2}{n} \sqrt{\frac{1 - \cos \theta_t}{1 - n \cos \theta_t} - \frac{2 - n(1 + \cos \theta_t)}{1 - n \cos \theta_t}}} \end{aligned} \quad (12.30)$$

For a Hohmann transfer between two specified circular orbits, (12.24) yields the value of $(\Delta V_{tot}/V_{c1})_{H}$. For the percentage increase of the total impulse (loss) required for the faster transfer orbit, we may write

$$L = \frac{\Delta V_{tot} - (\Delta V_{tot})_H}{(\Delta V_{tot})_H} * 100\% \quad (12.31)$$

For the flight time, t_f , in the transfer orbit, we may use (6.36-1) and find

$$t_f = \sqrt{\frac{a^3}{\mu}} (E_t - e \sin E_t) \quad (12.32)$$

where E_t is the eccentric anomaly corresponding to θ_t , which is given according to (6.35) by

$$\tan \frac{E_t}{2} = \sqrt{\frac{1-e}{1+e}} \tan \frac{\theta_t}{2}$$

Substitution of (12.27) results in

$$\tan \frac{E_t}{2} = \sqrt{\frac{2-n(1+\cos\theta_t)}{n(1-\cos\theta_t)}} \tan \frac{\theta_t}{2} \quad (12.33)$$

Equation (12.32) can also be written as

$$\frac{t_f}{T_{c_1}} = \frac{1}{2\pi} \sqrt{\left(\frac{a}{r_1}\right)^3} (E_t - e \sin E_t) \quad (12.34)$$

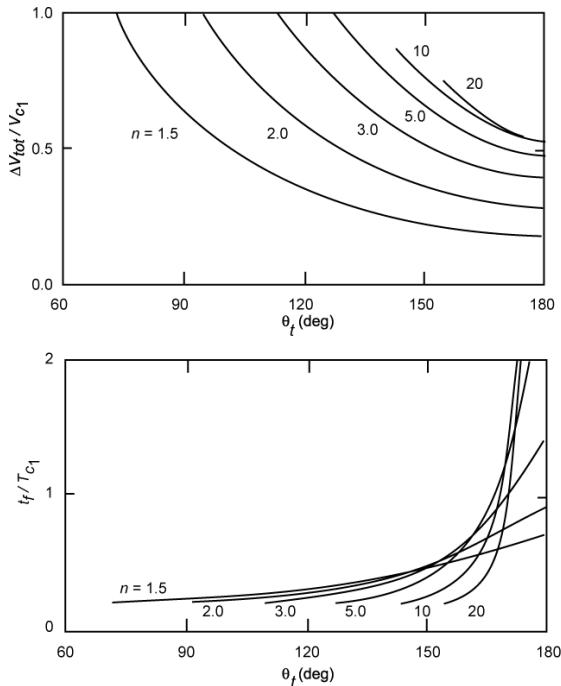


Figure 12.6: Dimensionless total impulse (top) and flight time (bottom) for a class of faster elliptical transfer orbits between two coplanar circular orbits, as a function of the transfer angle, θ_t , and n .

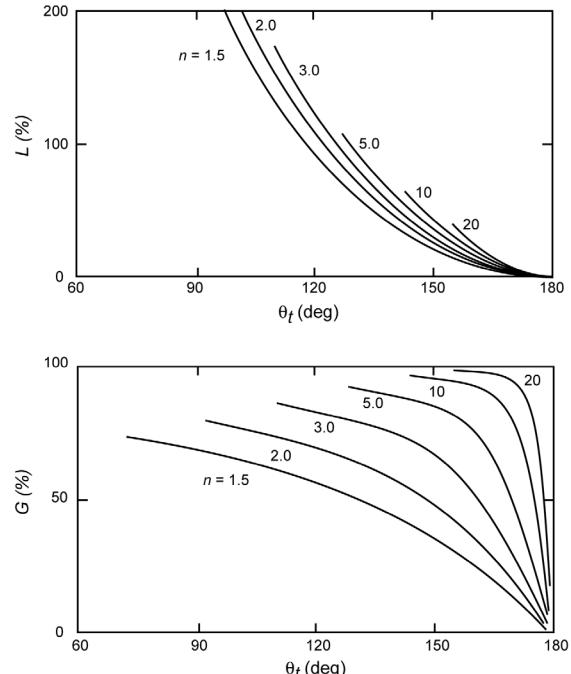


Figure 12.7: Percentage increase of the total impulse, L , (top) and the percentage reduction of the flight time, G , (bottom) for a class of faster elliptical transfer orbits between two coplanar circular orbits, as a function of the transfer angle, θ_t , and n .

Substitution of (12.27) and (12.29) into (12.34) yields

$$\frac{t_f}{T_{c_1}} = \frac{1}{2\pi} \sqrt{\left(\frac{1 - n \cos \theta_t}{2 - n(1 + \cos \theta_t)} \right)^3} \left[E_t - \frac{n-1}{1-n \cos \theta_t} \sin E_t \right] \quad (12.35)$$

For a Hohmann transfer orbit the flight time follows from (12.25). The percentage shortening (gain) of the flight time by using a faster transfer orbit may be written as

$$G = - \frac{t_f - t_{f_H}}{t_{f_H}} * 100\% \quad (12.36)$$

In Figure 12.6 the values of $\Delta V_{tot}/V_{c_1}$ and t_f/T_{c_1} are plotted as a function of θ_t , for different values of n . For all orbits: $e < 1$; the values at $\theta_t = 180^\circ$ hold for the Hohmann transfer. We observe that for small values of n ($n < 2.5$) the total impulse only slightly increases when θ_t decreases from 180° to 150° . However, the flight time then already decreases significantly. For large values of n ($n > 5$) the flight time strongly decreases when θ_t decreases from 180° to 150° . In Figure 12.7, the percentage increase of the required total impulse, L , and the percentage shortening of the flight time, G , are plotted as a function of θ_t , for different values of n ; for all orbits: $e < 1$. We see that for values of $\theta_t < 120^\circ$ the percentage increase in the required total impulse, with respect to the one for a Hohmann transfer orbit, may amount to more than 100%. The percentage shortening of the flight time is at $\theta_t = 150^\circ$ already 40-95%, depending on the value of n .

12.3. Multiple Hohmann transfer between two circular orbits

The question may be raised whether, for given initial and final circular orbits, less propellant is required when instead of one Hohmann transfer orbit multiple Hohmann transfer orbits are flown. In this Section, we will analyze the case of applying two successive Hohmann transfer orbits between the circular initial and the circular final orbit. For such a transfer flight three impulsive shots are required. First, the case will be considered where the apogee distance of the first transfer orbit is smaller than the radius of the final orbit. Secondly, the case where the apogee distance of the first transfer orbit is larger than the radius of the final orbit is treated.

Transfer orbit with a low apogee

In principle, there are two options for this case: 1) two successive impulsive shots at perigee of the two successive Hohmann transfer orbits, followed by a third impulsive shot to enter the final circular orbit (Figure 12.8), or 2) the second impulse at apogee of the first Hohmann transfer orbit. Here, we will analyze only the first case; it can be shown easily that the second option always costs more propellant. When we use the standard computation scheme introduced in Section 12.1, we can directly write (Figure 12.8):

$$\Delta V_1 = V_{c_1} \sqrt{1 + e_t} - V_{c_1} ; \quad e_t = \frac{r_a - r_1}{r_a + r_1} \quad (12.37)$$

$$\Delta V_2 = V_{c_1} \sqrt{1 + e} - V_{p_t} ; \quad e = \frac{r_2 - r_1}{r_2 + r_1} \quad (12.38)$$

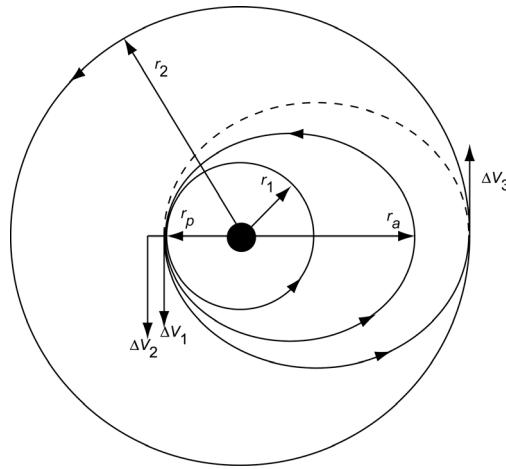


Figure 12.8: Geometry of a three-impulses Hohmann transfer trajectory with the apogee distance of the first Hohmann transfer orbit smaller than the radius of the final circular orbit.

$$\Delta V_3 = V_{c_2} - V_{c_2} \sqrt{1 - e} \quad ; \quad e = \frac{r_2 - r_1}{r_2 + r_1} \quad (12.39)$$

where the index t refers to the first transfer orbit and r_a denotes the apogee distance of this transfer orbit. We now use the relations derived in Chapter 6:

$$V_{p_t} = \sqrt{\frac{\mu}{a_t} \frac{1 + e_t}{1 - e_t}} \quad ; \quad a_t = \frac{r_1 + r_a}{2} \quad ; \quad V_{c_2} = \sqrt{\frac{r_1}{r_2}} V_{c_1}$$

from which follows

$$V_{p_t} = V_{c_1} \sqrt{\frac{2r_a}{r_a + r_1}}$$

Substitution of these relations into (12.37) to (12.39) leads to

$$\Delta V_1 = V_{c_1} \left[\sqrt{\frac{2r_a}{r_a + r_1}} - 1 \right] \quad (12.40)$$

$$\Delta V_2 = V_{c_1} \left[\sqrt{\frac{2r_2}{r_2 + r_1}} - \sqrt{\frac{2r_a}{r_a + r_1}} \right] \quad (12.41)$$

$$\Delta V_3 = V_{c_1} \sqrt{\frac{r_1}{r_2}} \left| 1 - \sqrt{\frac{2r_1}{r_2 + r_1}} \right| \quad (12.42)$$

or

$$\frac{\Delta V_{tot}}{V_{c_1}} = \frac{\Delta V_1 + \Delta V_2 + \Delta V_3}{V_{c_1}} = \sqrt{\frac{2r_2}{r_2 + r_1}} + \sqrt{\frac{r_1}{r_2}} - \sqrt{\frac{r_1}{r_2}} \sqrt{\frac{2r_1}{r_2 + r_1}} - 1$$

With (12.5) we may write this equation as

$$\frac{\Delta V_{tot}}{V_{c_1}} = (n - 1) \sqrt{\frac{2}{n(n + 1)}} + \sqrt{\frac{1}{n}} - 1 \quad (12.43)$$

Note that ΔV_{tot} is independent of the apogee altitude of the first transfer orbit and that (12.43) is identical to (12.24), which expresses the total impulse required for a direct Hohmann transfer. When we generalize this result, we may conclude that it does not matter how many intermediate orbits are selected; the total impulse is always the same, as long as $r_a < r_2$. This property is often used in the execution of space missions. Examples are the use of phasing orbits in order to reach a specific position in the final orbit (Chapter 14) and the application of multiple thrusting periods when low-thrust chemical rocket engines are used. With these propulsion systems a relatively long thrusting period is required in order to realize a desired orbit transfer. The thrust program that has to be applied in order to minimize the propellant consumption will mostly demand that the thrust vector is directed more or less tangentially to the orbit (Section 1.7 and Section 19.2). For long thrusting periods this implies that a component of the thrust will be directed oppositely to the gravitational attraction vector. This leads to a so-called *gravity loss* (Section 1.7). That loss can be limited by applying shorter thrusting periods during successive passages of perigee and reaching the final orbit through multiple intermediate orbits (Figure 12.9). As a result, the thrust will be directed ‘more perpendicular’ to the radius vector during the subsequent thrusting periods, causing the gravity losses to be smaller. A similar technique is applied in interplanetary spaceflight (Chapter 18), but then as a braking maneuver in order to reach lower orbits about a planet.

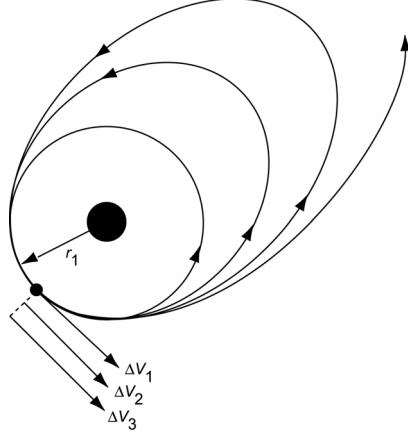


Figure 12.9: Application of multiple Hohmann transfer orbits with intermediate apogee altitudes and successive perigee thrusting periods.

Transfer orbit with a high apogee

Figure 12.10 depicts the situation where the apogee distance of the first transfer orbit (r_a) is larger than the radius of the final circular orbit (r_2). In this case, we introduce the new variable

$$m = \frac{r_a}{r_1} \quad (12.44)$$

Using the standard computation scheme introduced in Section 12.1, we write

$$\Delta V_1 = V_{c_1} \sqrt{1 + e} - V_{c_1} \quad ; \quad e = \frac{r_a - r_1}{r_a + r_1} \quad ; \quad V_a = V_{c_a} \sqrt{1 - e} \quad (12.45)$$

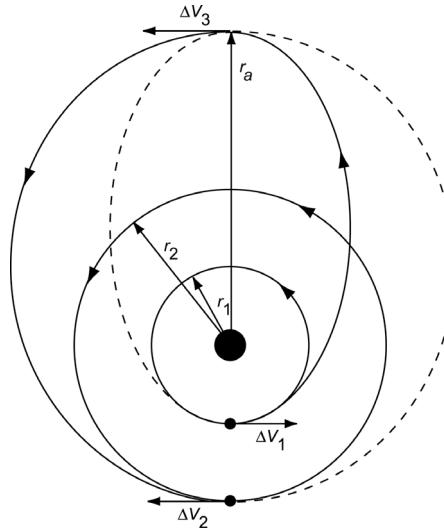


Figure 12.10: Geometry of a three-impulses Hohmann transfer with the apogee distance of the first Hohmann transfer orbit larger than the radius of the final circular orbit.

where V_a is the velocity at apogee of the first transfer orbit. For the apogee velocity required to reach the second transfer orbit, we have

$$V_a^* = V_{c_a} \sqrt{1 - e} \quad ; \quad e = \frac{r_a - r_2}{r_a + r_2} \quad (12.46)$$

So, the magnitude of the second impulsive shot is given by

$$\Delta V_3 = V_a^* - V_a \quad (12.47)$$

For the third impulsive shot, we write

$$\Delta V_2 = V_{c_2} \sqrt{1 + e} - V_{c_1} \quad ; \quad e = \frac{r_a - r_2}{r_a + r_2} \quad (12.48)$$

Substitution of the eccentricities of the corresponding transfer orbits in the expressions for the impulsive shots and using (12.5) and (12.44) yields

$$\Delta V_1 = V_{c_1} \left[\sqrt{\frac{2m}{m+1}} - 1 \right] \quad (12.49)$$

$$\Delta V_3 = V_{c_1} \sqrt{\frac{1}{m} \left[\sqrt{\frac{2n}{m+n}} - \sqrt{\frac{2}{m+1}} \right]} \quad (12.50)$$

$$\Delta V_2 = V_{c_1} \sqrt{\frac{1}{n} \left[\sqrt{\frac{2m}{m+n}} - 1 \right]} \quad (12.51)$$

or, after some algebraic manipulation,

$$\frac{\Delta V_{tot}}{V_{c_1}} = (m-1) \sqrt{\frac{2}{m(m+1)}} + \sqrt{\frac{2(n+m)}{mn}} - \sqrt{\frac{1}{n}} - 1 \quad (12.52)$$

The transfer with a high apogee distance demands less propellant than the classical Hohmann transfer with two impulsive shots when

$$\Delta \equiv \left(\frac{\Delta V_{tot}}{V_{c_1}} \right)_3 - \left(\frac{\Delta V_{tot}}{V_{c_1}} \right)_2 < 0$$

where the indices 2 and 3 indicate a two-impulses and three-impulses Hohmann transfer, respectively. Substitution of (12.24) and (12.52) results in

$$\Delta = (m-1) \sqrt{\frac{2}{m(m+1)}} - (n-1) \sqrt{\frac{2}{n(n+1)}} + \sqrt{\frac{2(m+n)}{mn}} - 2\sqrt{\frac{1}{n}} \quad (12.53)$$

In order to determine the values of m and n for which $\Delta < 0$, we analyze (12.53) numerically. But first, we determine two characteristics of this function. For $m = n$: $\Delta = 0$. This result is easy to understand since for $m = n$ the three-impulses transfer actually is a two-impulses transfer. For $n = 1$, $m \uparrow \infty$, we find: $\Delta \uparrow 2(\sqrt{2} - 1)$. This corresponds to twice the impulse that is required in the initial orbit to reach the escape velocity, which is also easy to understand. In Figure 12.11, the curve $\Delta(m,n) = 0$ is plotted along with the line $m = n$. Since always $m > n$, the line $m = n$ marks a boundary in $m-n$ space where three-impulses transfer orbits (with a high apogee) are possible. The curve $\Delta = 0$ marks the separation of the regions in $m-n$ space where either a two-impulses Hohmann transfer or a three-impulses Hohmann transfer is optimum. A numerical analysis of (12.53) shows that:

- For $n < 11.939$, a two-impulses Hohmann transfer is always optimum.
- For $11.939 < n < 15.582$, a three-impulses Hohmann transfer is optimum, if m is chosen sufficiently large.

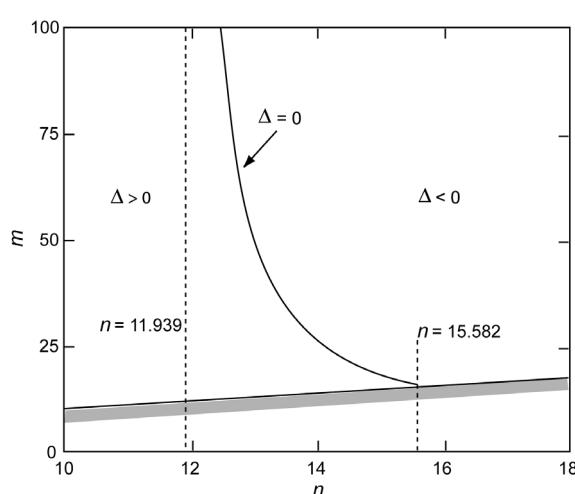


Figure 12.11: Regions in $n-m$ space where two-impulses or three-impulses Hohmann transfer trajectories between two coplanar circular orbits are optimum.

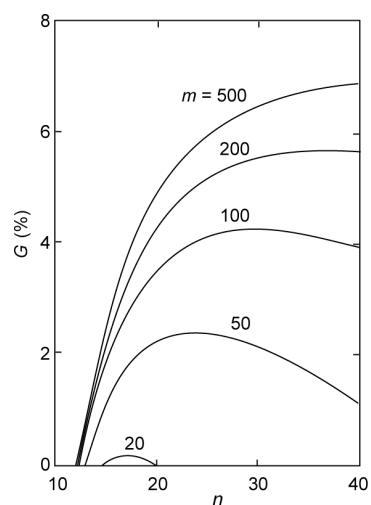


Figure 12.12: Percentage reduction of the total impulse, G , when a three-impulses Hohmann transfer trajectory between two coplanar circular orbits is flown instead of a two-impulses Hohmann trajectory, as a function of n and m .

– For $n > 15.582$, a three-impulses Hohmann transfer is optimum for any value of $m > n$. These results indicate that for certain values of n less propellant is required if the satellite first climbs from an initial circular orbit through a Hohmann transfer orbit to a high altitude, and then enters a second Hohmann transfer orbit to descend to the altitude of the final circular orbit. The advantage of the three-impulses transfer only exists if $n > 11.939$. This means a final circular orbit at a very high altitude ($h > 72,000$ km). In Figure 12.12 the percentage reduction of the total impulse achieved by using a three-impulses Hohmann transfer is plotted as a function of n for several values of m . This percentage reduction (gain) is defined as

$$G = - \frac{(\Delta V_{tot})_3 - (\Delta V_{tot})_2}{(\Delta V_{tot})_2} * 100\% \quad (12.54)$$

Note that for $m \leq 50$ a total impulse reduction of only 2% or less can be achieved by applying a three-impulses transfer; for $m > 50$ this reduction can be 5% or more. When the satellite is carrying cryogenic propellants, part of this advantage is lost because of the boiling off of propellant during the extended flight. However, if the satellite is designed for a long operational lifetime in the final orbit, the longer flight time in the transfer trajectory is not relevant and the three-impulses transfer is attractive.

12.4. Hohmann transfer from a circular orbit to an elliptical orbit

In this Section it is assumed that a satellite is in a circular orbit and that it has to go into a coplanar elliptical orbit with perigee distance larger than the radius of the initial orbit (Figure 12.13). Based on the results from Section 12.1, one may expect that a Hohmann type of transfer will be optimum. This still leaves two options, which are indicated in Figure 12.13 by I and II. In the following, the required total impulse for both types of transfer will be compared. In the analysis, the notation a and e will be used for the elements of the final elliptical orbit, and a_i and e_i for the elements of the transfer orbit.

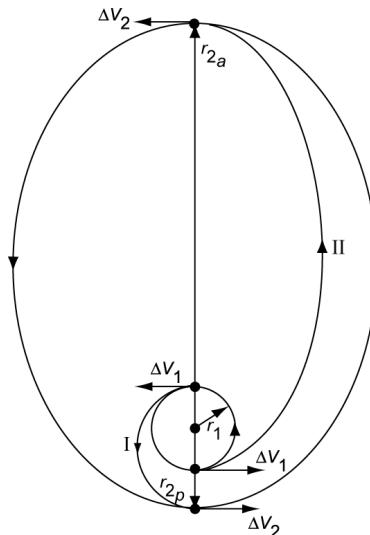


Figure 12.13: Two feasible Hohmann transfer orbits from an initial circular orbit to a coplanar elliptical final orbit.

Case I

From Figure 12.13 we conclude that we may write for the first and second impulsive shot

$$\Delta V_1 = V_{c_1} \sqrt{1 + e_t} - V_{c_1} ; \quad e_t = \frac{r_{2_p} - r_1}{r_{2_p} + r_1} \quad (12.55)$$

$$\Delta V_2 = V_{2_p} - V_{c_{2_p}} \sqrt{1 - e_t} ; \quad e_t = \frac{r_{2_p} - r_1}{r_{2_p} + r_1} \quad (12.56)$$

With the expressions for the eccentricities and the relations

$$r_{2_p} = a(1 - e)$$

$$V_{c_{2_p}} = \sqrt{\frac{\mu}{r_{2_p}}} = V_{c_1} \sqrt{\frac{r_1}{a(1 - e)}} ; \quad V_{2_p} = \sqrt{\frac{\mu}{a} \left(\frac{1 + e}{1 - e} \right)} = V_{c_1} \sqrt{\frac{r_1}{a} \left(\frac{1 + e}{1 - e} \right)}$$

we find, after some algebraic manipulation, for both impulsive shots

$$\Delta V_1 = V_{c_1} \left[\sqrt{\frac{2a(1 - e)}{a(1 - e) + r_1}} - 1 \right] \quad (12.57)$$

$$\Delta V_2 = V_{c_1} \left[-\frac{r_1}{a} \sqrt{\frac{2}{(1 - e)(1 - e + r_1/a)}} + \sqrt{\frac{r_1}{a} \frac{1 + e}{1 - e}} \right] \quad (12.58)$$

For the required total impulse then follows

$$\left(\frac{\Delta V_{tot}}{V_{c_t}} \right)_I = \sqrt{\frac{2}{(1 - e)(1 - e + r_1/a)}} \left(1 - e - \frac{r_1}{a} \right) + \sqrt{\frac{r_1}{a} \frac{(1 + e)}{(1 - e)}} - 1 \quad (12.59)$$

Case II

From Figure 12.13 we conclude that we may write for the first and second impulsive shot

$$\Delta V_1 = V_{c_1} \sqrt{1 + e_t} - V_{c_1} ; \quad e_t = \frac{r_{2_a} - r_1}{r_{2_a} + r_1} \quad (12.60)$$

$$\Delta V_2 = V_{2_a} - V_{c_{2_a}} \sqrt{1 - e_t} ; \quad e_t = \frac{r_{2_a} - r_1}{r_{2_a} + r_1} \quad (12.61)$$

With the expressions for the eccentricities and the relations

$$r_{2_a} = a(1 + e)$$

$$V_{c_{2_a}} = \sqrt{\frac{\mu}{r_{2_a}}} = V_{c_1} \sqrt{\frac{r_1}{a(1 + e)}} ; \quad V_{2_a} = \sqrt{\frac{\mu}{a} \frac{1 - e}{1 + e}} = V_{c_1} \sqrt{\frac{r_1}{a} \frac{1 - e}{1 + e}}$$

we find for both impulsive shots:

$$\Delta V_1 = V_{c_1} \left[\sqrt{1 + \frac{a(1+e) - r_1}{a(1+e) + r_1}} - 1 \right] \quad (12.62)$$

$$\Delta V_2 = V_{c_1} \left[-\sqrt{\frac{r_1}{a(1+e)}} \sqrt{1 - \frac{a(1+e) - r_1}{a(1+e) + r_1}} + \sqrt{\frac{r_1}{a} \left(\frac{1-e}{1+e} \right)} \right] \quad (12.63)$$

For the required total impulse then follows

$$\left(\frac{\Delta V_{tot}}{V_{c_1}} \right)_{II} = \sqrt{\frac{2}{(1+e)(1+e+r_1/a)}} \left(1 + e - \frac{r_1}{a} \right) + \sqrt{\frac{r_1}{a} \left(\frac{1-e}{1+e} \right)} - 1 \quad (12.64)$$

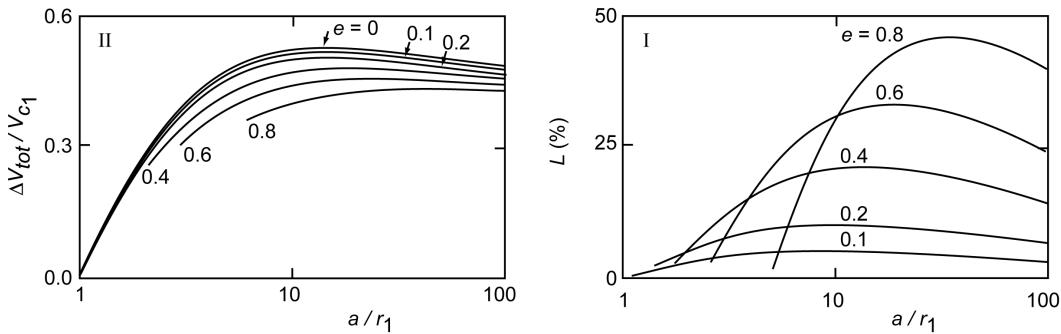


Figure 12.14: Dimensionless total impulse for transfer orbit II from an initial circular orbit to a coplanar elliptical final orbit (left) and the percentage increase of the total impulse for transfer orbit I (right), as a function of the parameters a/r_1 and e of the final orbit.

For given values of a and e we can compute the total impulse for both types of transfer orbits from (12.59) and (12.64). It then turns out that type II transfers are always preferable. So, injection at apogee of the final elliptical orbit demands less propellant than injection at perigee of the final elliptical orbit. This result can be verified with Figure 12.14. In the left part of this Figure, the dimensionless total impulse for transfer orbit II is plotted as a function of a/r_1 and e . In the right part of this Figure, the percentage increase of the dimensionless total impulse that is required for a type I transfer orbit is plotted. This percentage increase (loss) is defined as

$$L = \frac{(\Delta V_{tot})_I - (\Delta V_{tot})_{II}}{(\Delta V_{tot})_{II}} * 100\% \quad (12.65)$$

The Figure shows that for values of $e = 0 - 0.6$, all curves for the total impulse required for type II transfer orbits reach a maximum for $a/r_1 = 16-27$. For $e < 0.4$, up to 20% more impulse is needed when, instead of type II transfer orbits, type I transfer orbits are flown; for $0.5 < e < 0.8$ the percentage increase of the required impulse amounts up to 25 - 45%.

12.5. Hohmann transfer from an elliptical orbit to a circular orbit

In this Section, it is assumed that the satellite is in an elliptical orbit with orbital elements a and e and that it has to go into a coplanar circular orbit with a radius larger than the apogee distance of the initial orbit (Figure 12.15). Extrapolating the results of Section 12.1, one may expect that

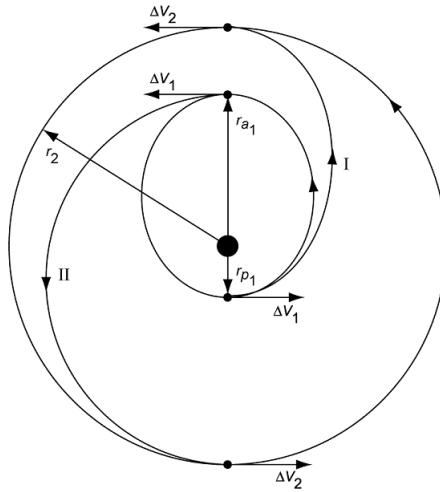


Figure 12.15: Two feasible Hohmann transfer orbits from an initial elliptical orbit to a coplanar circular final orbit.

a Hohmann type of transfer is optimum. This still leaves two possibilities, which are indicated in Figure 12.15 by I and II. In the following, the required total impulse for both types of transfer orbits will be compared, where e_t indicates the eccentricity of the transfer orbit.

Case I

From Figure 12.15, we conclude that we can write for the first impulsive shot

$$\Delta V_1 = V_{c_{p_1}} \sqrt{1 + e_t} - V_{c_{p_1}} \sqrt{1 + e} \quad ; \quad e_t = \frac{r_2 - r_{p_1}}{r_2 + r_{p_1}} \quad (12.66)$$

and for the second impulsive shot

$$\Delta V_2 = V_{c_2} - V_{c_2} \sqrt{1 - e_t} \quad ; \quad e_t = \frac{r_2 - r_{p_1}}{r_2 + r_{p_1}} \quad (12.67)$$

With the expressions for the eccentricities and the relations

$$r_{p_1} = a(1 - e) \quad ; \quad V_{c_{p_1}} = \sqrt{\frac{\mu}{r_{p_1}}} = V_{c_2} \sqrt{\frac{r_2}{a(1 - e)}}$$

we find, after some algebraic manipulation, for both impulsive shots:

$$\frac{\Delta V_1}{V_{c_2}} = \sqrt{\frac{r_2/a}{1 - e}} \left[\sqrt{\frac{2}{1 + a/r_2(1 - e)}} - \sqrt{1 + e} \right] \quad (12.68)$$

$$\frac{\Delta V_2}{V_{c_2}} = 1 - \sqrt{\frac{2a/r_2(1 - e)}{1 + a/r_2(1 - e)}} \quad (12.69)$$

and for the required total impulse:

$$\left(\frac{\Delta V_{tot}}{V_{c_2}} \right)_I = \sqrt{\frac{2r_2/a}{(1-e)[1+a/r_2(1-e)]}} \left(1 - \frac{a}{r_2}(1-e) \right) - \sqrt{\frac{r_2/a(1+e)}{1-e} + 1} \quad (12.70)$$

Case II

Now, we can write for the first impulsive shot

$$\Delta V_1 = V_{c_{a_1}} \sqrt{1+e_t} - V_{c_{a_1}} \sqrt{1-e} ; \quad e_t = \frac{r_2 - r_{a_1}}{r_2 + r_{a_1}} \quad (12.71)$$

and for the second impulsive shot

$$\Delta V_2 = V_{c_2} - V_{c_2} \sqrt{1-e_t} ; \quad e_t = \frac{r_2 - r_{a_1}}{r_2 + r_{a_1}} \quad (12.72)$$

With the expression for the eccentricity e_t and the relations

$$r_{a_1} = a(1+e) ; \quad V_{c_{a_1}} = \sqrt{\frac{\mu}{r_{a_1}}} = V_{c_2} \sqrt{\frac{r_2}{a(1+e)}}$$

we find, after some algebraic manipulation, for both impulsive shots:

$$\frac{\Delta V_1}{V_{c_2}} = \sqrt{\frac{r_2/a}{1+e}} \left[\sqrt{\frac{2}{1+a/r_2(1+e)}} - \sqrt{1-e} \right] \quad (12.73)$$

$$\frac{\Delta V_2}{V_{c_2}} = 1 - \sqrt{\frac{2a/r_2(1+e)}{1+a/r_2(1+e)}} \quad (12.74)$$

and for the required total impulse:

$$\left(\frac{\Delta V_{tot}}{V_{c_2}} \right)_{II} = \sqrt{\frac{2r_2/a}{(1+e)[1+a/r_2(1+e)]}} \left(1 - \frac{a}{r_2}(1+e) \right) - \sqrt{\frac{r_2}{a} \left(\frac{1-e}{1+e} \right)} + 1 \quad (12.75)$$

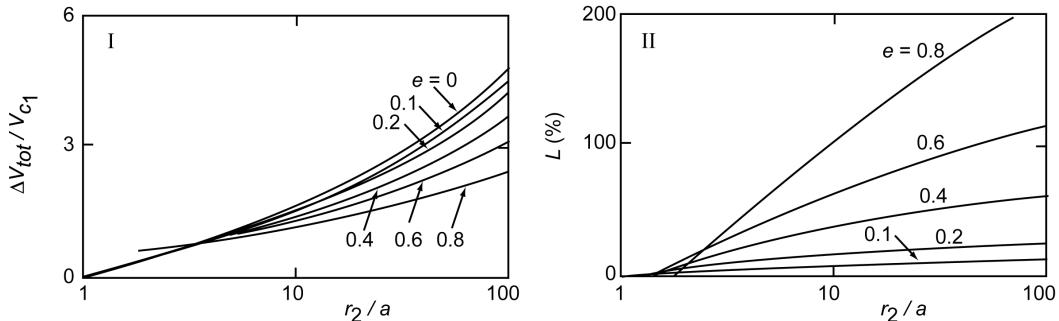


Figure 12.16: Dimensionless total impulse for transfer orbit I from an initial elliptical orbit to a coplanar circular final orbit (left) and the percentage increase of the total impulse for transfer orbit II (right), as a function of the parameters r_2/a and e of the initial orbit.

An analysis of (12.70) and (12.75) shows that type I transfer orbits, i.e. departure at perigee of the original elliptical orbit, demand, for any value of r_2/a and e , always less total impulse than type II transfer orbits. In Figure 12.16, the dimensionless total impulse required for a type I transfer orbit is plotted as a function of r_2/a and e . Also plotted is the percentage increase of the required total impulse when a type II orbit is flown. This percentage increase (loss) is defined as

$$L = \frac{(\Delta V_{tot})_{II} - (\Delta V_{tot})_I}{(\Delta V_{tot})_I} * 100\% \quad (12.76)$$

Note that for a given value of a , the total impulse for a type I transfer orbit decreases when e increases. For large values of r_2/a and e , the percentage increase in the required total impulse for type II transfer orbits is large, and may amount up to 100-200%.

12.6. Transfer from a circular orbit to an intersecting elliptical orbit

In this Section, we consider the case that the initial circular orbit and the final elliptical orbit intersect each other (Figure 12.17). Of the many feasible transfer trajectories, three types will be analyzed:

- Type I: single-impulse maneuver at the point where the initial and final orbits intersect.
- Type II: two-impulses maneuver with ΔV_2 at apogee of the final elliptical orbit.
- Type III: two-impulses maneuver with ΔV_2 at perigee of the final elliptical orbit.

In the following, the total impulse for these three types of transfer orbits will be determined.

Case I

From the general relations for the impulses required for transfer trajectories ((12.8)), we obtain for this case

$$\frac{\Delta V_1}{V_{c_1}} = \sqrt{3 - 2\sqrt{z} - \frac{1 - e^2}{z}} \quad (12.77)$$

where z and e refer to the final elliptical orbit. For the value of z , we have in this case

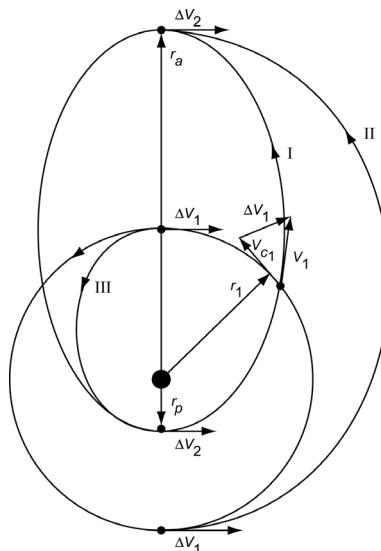


Figure 12.17: Transfer from an initial circular orbit to a coplanar crossing elliptical orbit. Shown are a direct injection maneuver (I) and two types of Hohmann transfer orbits (II and III).

$$z = \frac{p}{r_1} = \frac{a(1-e^2)}{r_1}$$

where a is the semi-major axis of the final elliptical orbit. We then find for the single impulsive shot

$$\left(\frac{\Delta V_1}{V_{c_1}} \right)_I = \sqrt{3 - \frac{r_1}{a} - 2\sqrt{\frac{1-e^2}{r_1/a}}} \quad (12.78)$$

Case II

From Figure 12.17 we conclude that we can write for the first impulsive shot

$$\Delta V_1 = V_{c_1} \sqrt{1+e_t} - V_{c_1} ; \quad e_t = \frac{r_a - r_1}{r_a + r_1} \quad (12.79)$$

and for the second impulsive shot

$$\Delta V_2 = V_{c_{r_a}} \sqrt{1-e_t} - V_{c_{r_a}} \sqrt{1-e} ; \quad e_t = \frac{r_a - r_1}{r_a + r_1} \quad (12.80)$$

where e_t denotes the eccentricity of the transfer orbit. With the expressions for the eccentricities and the relations

$$V_{c_{r_a}} = \sqrt{\frac{\mu}{r_a}} = V_{c_1} \sqrt{\frac{r_1}{a(1+e)}} ; \quad e_t = \frac{a(1+e) - r_1}{a(1+e) + r_1}$$

we find for the required total impulse:

$$\left(\frac{\Delta V_{tot}}{V_{c_1}} \right)_{II} = \sqrt{2 \frac{(1+e+r_1/a)}{1+e}} - \sqrt{\frac{r_1}{a} \left(\frac{1-e}{1+e} \right)} - 1 \quad (12.81)$$

Case III

From Figure 12.17 we conclude that we can write for the first impulsive shot

$$\Delta V_1 = V_{c_1} - V_{c_1} \sqrt{1-e_t} ; \quad e_t = \frac{r_1 - r_p}{r_1 + r_p} \quad (12.82)$$

and for the second impulsive shot

$$\Delta V_2 = V_{c_{r_p}} \sqrt{1+e} - V_{c_{r_p}} \sqrt{1+e_t} ; \quad e_t = \frac{r_1 - r_p}{r_1 + r_p} \quad (12.83)$$

With the expressions for the eccentricities and the relations

$$V_{c_{r_p}} = \sqrt{\frac{\mu}{r_p}} = V_{c_1} \sqrt{\frac{r_1}{a(1-e)}} ; \quad e_t = \frac{r_1 - a(1-e)}{r_1 + a(1-e)}$$

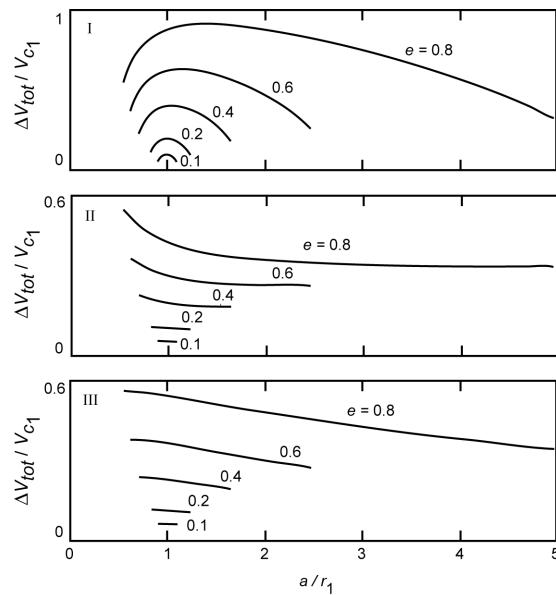


Figure 12.18: Dimensionless total impulse for the three types of transfer from an initial circular orbit to a coplanar intersecting elliptical final orbit, as a function of the parameters a/r_1 and e of the final orbit.

we find for the required total impulse

$$\left(\frac{\Delta V_{tot}}{V_{c1}} \right)_{\text{III}} = 1 - \sqrt{\frac{2(1 - e + r_1/a)}{1 - e}} + \sqrt{\frac{r_1}{a} \left(\frac{1 + e}{1 - e} \right)} \quad (12.84)$$

In Figure 12.18 the required dimensionless total impulse for the three types of transfer is plotted as a function of a and e . An analysis of the relations (12.78), (12.81) and (12.84), and an examination of Figure 12.18, shows that the single-impulse maneuver demands by far the most energy, and that the type II transfer always requires the least amount of propellant. However, for small values of e , a type III transfer demands only slightly more impulse ($< 5\%$) than a type II transfer, while the flight time in a type III transfer is significantly shorter than in a type II transfer. For a final elliptical orbit with $e < 0.2$, therefore, one will mostly select a type III transfer orbit.

12.7. Gravity losses

In the previous Sections, it has been assumed explicitly that the maneuvers required to enter and leave a transfer orbit are executed by impulsive shots. However, in reality the thrusting time of rocket engines is not infinitely short, and it has been shown in Section 1.7 that then a *gravity loss* may occur. For first-order analyses a classical approach is to assume that this gravity loss has to be compensated for by an additional ΔV . That ΔV has to be added to the ΔV computed for the impulsive shot and results in a gross ΔV , which, when substituted into Tsiolkovski's law (Section 1.7), yields the actually required amount of propellant for this maneuver. For a given mission, generally, the thrust vector program will be selected in such a way that the gravity loss is minimized. The computation of such optimum thrust programs is quite complicated and will not be discussed in this book. In the following, only a few examples of maneuvers and associated gravity losses will be given, from which a general picture of the actual level of gravity losses can

be obtained. These examples are taken from document ESA CR(P) 1910¹ that contains many plots from which the gravity losses can be estimated for many different initial and final orbits, and various thrusting arc characteristics.

For the numerical examples given in this Section, it is assumed that a maneuver is executed to bring a satellite from a circular orbit into an elliptical, parabolic or hyperbolic orbit with perigee distance equal to the radius of the initial circular orbit. If this maneuver would be accomplished by an impulsive shot, then the thrust vector should be oriented perpendicular to the position vector, which results, according to the discussion in Section 1.7, in a zero gravity loss. For the real maneuver it is assumed that the magnitude of the thrust and the effective exhaust velocity of the rocket engine are constant over the thrusting arc and that an optimum thrust angle program is applied; i.e. that the thrust angle, δ , varies such that the gravity loss during the thrusting arc is minimum. The thrust angle is defined as the angle between the thrust vector and the plane normal to the position vector (Section 1.7).

In the previously mentioned document, three parameters are defined which characterize the original orbit and the final orbit:

$$\mathcal{E}_1^* = \frac{\mathcal{E}_1}{V_{c_1}^2} \quad ; \quad \mathcal{E}_2^* = \frac{\mathcal{E}_2}{V_{c_1}^2} \quad ; \quad \Delta\mathcal{E}^* = \mathcal{E}_2^* - \mathcal{E}_1^* \quad (12.85)$$

where \mathcal{E} is the orbital energy, \mathcal{E}^* is the normalized orbital energy, both per unit of mass, V_{c_1} is the circular velocity in the initial orbit, and the indices 1 and 2 indicate the initial orbit and the final orbit, respectively. For the assumptions made in this Section, we find

$$\begin{aligned} \mathcal{E}_1^* &= \frac{-\mu/r_1 + \frac{1}{2}V_{c_1}^2}{V_{c_1}^2} = \frac{-V_{c_1}^2 + \frac{1}{2}V_{c_1}^2}{V_{c_1}^2} = -\frac{1}{2} \\ \mathcal{E}_2^* &= \frac{-\mu/r_1 + \frac{1}{2}V_{p_2}^2}{V_{c_1}^2} = -1 + \frac{1}{2}\frac{V_{p_2}^2}{V_{c_1}^2} \end{aligned} \quad (12.86)$$

where V_{p_2} is the perigee velocity in the final orbit. For an elliptical, parabolic or hyperbolic final orbit we may write

$$V_{p_2}^2 = V_{c_1}^2(1+e) \quad (12.87)$$

where e is the eccentricity of the final orbit. Substitution of (12.87) into (12.86-2) yields

$$\mathcal{E}_2^* = -\frac{1}{2}(1-e) \quad ; \quad \Delta\mathcal{E}^* = \frac{1}{2}e \quad (12.88)$$

Note that $\Delta\mathcal{E}^* < 0.5$ refers to an elliptical final orbit, $\Delta\mathcal{E}^* = 0.5$ to a parabolic final orbit, and $\Delta\mathcal{E}^* > 0.5$ to a hyperbolic final orbit. For an elliptical final orbit the eccentricity can be expressed by $e = (r_a - r_1)/(r_a + r_1)$, where r_a is the apogee distance. Three other dimensionless parameters introduced in document ESA CR(P) 1910 are

¹ J. Weiss, B. Metzger, M. Gallmeister, *Orbit maneuvers with finite thrust*, MBB/ERNO, ESA CR(P) 1910, 3 volumes, 1983.

$$k_2 = \frac{F}{g_1 M_0} = \frac{r_1^2}{\mu} \frac{F}{M_0} ; \quad k_3 = \frac{V_j}{V_{c1}} ; \quad t^* = \frac{t}{\sqrt{r_1^3/\mu}} = \frac{2\pi t}{T_1} \quad (12.89)$$

where F is the thrust, g_1 is the acceleration due to gravity at the injection point, M_0 is the initial mass of the spacecraft, μ is the gravitational parameter of the Earth, V_j is the (effective) exhaust velocity of the rocket engine, and T_1 is the period of the original circular orbit. The percentage gravity loss is defined as

$$\Delta\Delta V = \left(\frac{\Delta V - (\Delta V)_i}{(\Delta V)_i} \right) * 100\% \quad (12.90)$$

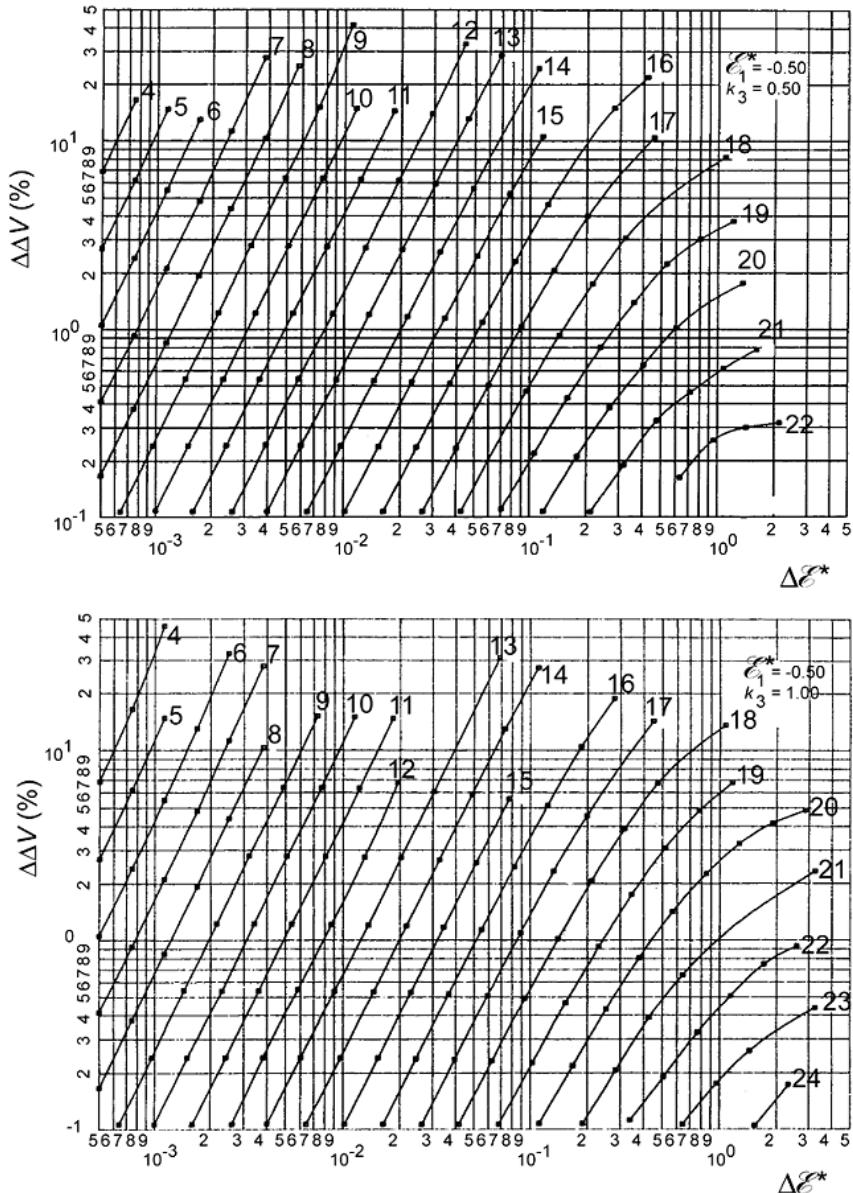


Figure 12.19: Percentage gravity loss for an optimum transfer from a circular orbit to a coplanar elliptical, parabolic or hyperbolic orbit that is tangent to the initial orbit, as a function of ΔE^* and k_2 . The curves hold for $V_j = 0.5 V_{cl}$ (top) and $V_j = V_{cl}$ (bottom).

Table 12.1: The values of k_2 that correspond to the indices indicated in Figure 12.19.

index	k_2	index	k_2	index	k_2	index	k_2	index	k_2
4	0.00040	8	0.0025	12	0.016	16	0.10	20	0.63
5	0.00063	9	0.0040	13	0.025	17	0.16	21	1.0
6	0.0010	10	0.0063	14	0.040	18	0.25	22	1.6
7	0.0016	11	0.010	15	0.063	19	0.40	23	2.5

where ΔV is the velocity increase that corresponds to the finite thrusting period and $(\Delta V)_i$ is the velocity increase in case the maneuver would have been executed by an impulsive shot:

$$(\Delta V)_i = V_{c1} \left(\sqrt{1+e} - 1 \right) \quad (12.91)$$

In the ESA report, many plots are presented for the value of $\Delta\Delta V$ as a function of $\Delta\mathcal{E}^*$, for specified values of \mathcal{E}_1^* , k_2 and k_3 . An example is given in Figure 12.19 for the case that $\mathcal{E}_1^* = -0.5$ (circular initial orbit), and $V_j = 0.5 V_{c1}$ or $V_j = V_{c1}$. The number indicated at each curve refers to the applied value of k_2 ; the values of k_2 corresponding to these numbers are listed in Table 12.1. The plots show that for values of $\Delta\mathcal{E}^* < 0.2$ ($e < 0.4$), the effect of different exhaust velocities (k_3) on the value of the gravity loss is very small. So, the effect of the exhaust velocity only becomes important for higher values of the eccentricity. For given values of k_3 and $\Delta\mathcal{E}^*$, the gravity loss decreases as k_2 increases. This is a trivial result, because for a given mission a larger value of k_2 results in a shorter thrusting period. For an escape flight: $\Delta\mathcal{E}^* = 0.5$. When we adopt $k_3 = 0.5$, we find from Figure 12.19 (top): $\Delta\Delta V < 5\%$ if $k_2 > 0.25$; for $k_2 = 0.10$ we find $\Delta\Delta V = 25\%$.

As an example, Figure 12.20 (left) shows, for an optimum maneuver to leave an initial circular orbit and to enter an elliptical final orbit with $r_a/r_1 = 2.0$ ($\mathcal{E}_2^* = -0.333$, $\Delta\mathcal{E}^* = 0.167$, $e = 0.334$) that touches the original orbit, the variation of the angle $\theta - \delta$, where θ is the true anomaly and δ is the thrust angle. For this example, the following values hold: $F/M_0 = 0.05 g_1$

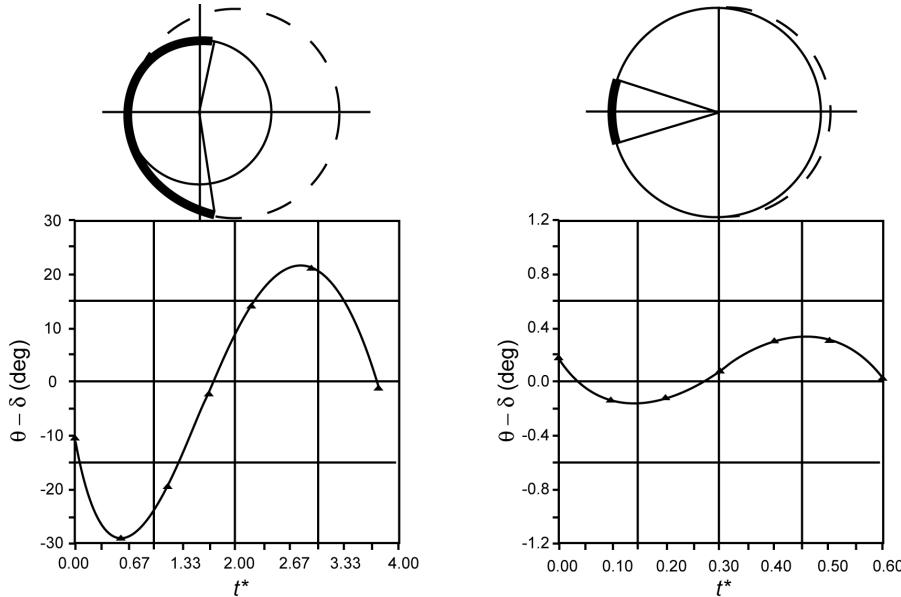


Figure 12.20: Variation of the angle $\theta - \delta$ for an optimum transfer from a circular orbit to a coplanar elliptical orbit that is tangent to the initial orbit. The plot on the left holds for $r_a/r_1 = 2.0$, $e = 0.334$, $F/M_0 = 0.05 g_1$, $V_j = V_{cl}$; the plot on the right for $r_a/r_1 = 1.028$, $e = 0.014$, $F/M_0 = 0.0123 g_1$, $V_j = 0.388 V_{cl}$.

($k_2 = 0.05$) and $V_j = V_{c1}$ ($k_3 = 1$). During this flight a prolonged thrusting phase is applied: $t \approx 0.6 T_1$ ($t^* \approx 3.77$) and $\theta - \delta$ varies between -29° and 21° . Inspection of Figure 12.19 (bottom) yields for $\Delta\mathcal{E}^* = 0.167$: $\Delta\Delta V \approx 34\%$. Obviously, this high value is a direct result of the long thrusting period, during which the thrust angle deviates significantly from zero. In this case, the gravity loss could have been reduced by applying during the first perigee passage ($\theta \approx 0^\circ$) a much shorter thrusting phase, and by applying at each succeeding perigee passage additional short thrusting phases (Section 12.3). Figure 12.20 (right) shows similar information, but now for an optimum maneuver with a fairly short thrusting phase of $t \approx 0.095 T_1$ ($t^* = 0.6$). In this case the satellite is transferred from an initial circular orbit to an elliptical final orbit with $r_a/r_1 = 1.028$ ($\mathcal{E}_2^* = -0.493$, $\Delta\mathcal{E}^* = 0.007$, $e = 0.014$). The applied thrust and the exhaust velocity are given by $F/M_0 = 0.0123$ g_1 ($k_2 = 0.0123$) and $V_j = 0.388 V_{c1}$ ($k_3 = 0.388$). The angle $\theta - \delta$ deviates during the entire propelled flight less than 0.4° from zero and the gravity loss amounts to only about 1.5%.

When we adopt the following values as characteristic for realistic maneuvers using chemical propulsion: $e < 0.4$ ($\Delta\mathcal{E}^* < 0.2$), $k_2 > 0.1$, $k_3 = 0.5$, then it follows from Figure 12.19 (top) that, for the kind of maneuvers considered in this Section, the gravity loss is less than 10%. Using electric propulsion, where very small values of k_2 occur, the gravity loss, in particular for large values of $\Delta\mathcal{E}^*$, will become very large. This topic is more thoroughly discussed in Chapter 19.

When the percentage gravity loss, $\Delta\Delta V$, is known, then the associated additional amount of propellant can be computed as follows. From *Tsiolkovski's law* (Section 1.7), we find for the propellant consumption of an impulsive shot (index i) and for a maneuver with finite thrusting time

$$\left(\frac{M_p}{M_0} \right)_i = 1 - \exp \left[-\frac{(\Delta V)_i}{V_j} \right] ; \quad \left(\frac{M_p}{M_0} \right) = 1 - \exp \left[-\frac{(\Delta V)_i}{V_j} \left(1 + \frac{\Delta\Delta V}{100} \right) \right] \quad (12.92)$$

where M_p is the required propellant mass and M_0 is the initial mass, and $\Delta\Delta V$ is expressed as a percentage. When the percentage increase of the required amount of propellant for a maneuver with finite thrusting time is defined as

$$\Delta M_p = \frac{M_p - (M_p)_i}{(M_p)_i} * 100\% \quad (12.93)$$

we find with (12.92)

$$\Delta M_p = 100 \left[\frac{1 - \exp \left\{ - \left((\Delta V)_i / V_j \right) (\Delta\Delta V / 100) \right\}}{\exp \left((\Delta V)_i / V_j \right) - 1} \right] \quad (12.94)$$

For a specified total mass, M_0 , any propellant increase, of course, implies an equivalent decrease of the payload capability. When we substitute (12.91) into (12.94), we obtain

$$\Delta M_p = 100 \left[\frac{1 - \exp \left\{ - (\sqrt{1+e} - 1) (V_{c1}/V_j) (\Delta\Delta V / 100) \right\}}{\exp \left\{ (\sqrt{1+e} - 1) (V_{c1}/V_j) \right\} - 1} \right] \quad (12.95)$$

It is emphasized that $\Delta\Delta V$ is a function of the altitude of the initial circular parking orbit, r_1 , the eccentricity of the final orbit, e , the initial thrust acceleration, F/M_0 , and the exhaust velocity, V_j . For specified values of r_1 , the quantities V_{c1} and g_1 can be computed. Then, for a specified value of e , the values of $(\Delta V)_i$ and $\Delta\mathcal{E}^*$ can be computed from (12.88) and (12.91). Subsequently, for specified values of $k_2 = F/g_1 M_0$ and $k_3 = V_j/V_{c1}$, the value of $\Delta\Delta V$ can be read from diagrams like

the ones presented in Figure 12.19. Finally, the value of ΔM_p can be computed from (12.95). As an example, Table 12.2 shows the results for an initial circular parking orbit about the Earth at an altitude of 300 km, $e = 0.01, 0.04, 0.2$, or 1.0 , $F/g_1 M_0 = 0.01, 0.04, 0.1$, or 0.4 , and $V_j/V_{c1} = 0.5$. The series of eccentricities correspond to $(\Delta V)_i = 0.038 \text{ km/s}, 0.153 \text{ km/s}, 0.737 \text{ km/s}$, and 3.200 km/s , respectively. The particular value of V_j was selected because then Figure 12.19 (top) can be used to read the values of $\Delta\Delta V$. From this Table, we conclude that for small maneuvers, leading to elliptical orbits with $e < 0.01$, the percentage gravity loss and percentage propellant increase are negligible for all realistic values of the thrust acceleration, $F/g_1 M_0$. For large maneuvers, the gravity loss and propellant increase may become significant, in particular for smaller thrust accelerations, i.e. longer engine burning times. The percentage propellant increase is always less than the percentage gravity loss; the largest differences occur at the smaller thrust accelerations. Of course, for larger or smaller values of V_j , and the same values of r_1 , e and $F/g_1 M_0$, the percentage gravity loss and percentage propellant increase will be smaller or larger, respectively.

Table 12.2: Values of $\Delta\Delta V$ and ΔM_p , both in percentages, for a transfer from a circular parking orbit at an altitude of 300 km to an elliptical orbit with perigee altitude of 300 km, for $V_j = 0.5V_{c1}$ and specified values of e and $F/g_1 M_0$.

$e = 0.01$		$e = 0.04$		$e = 0.2$		$e = 1.0$		
$F/g_1 M_0$	$\Delta\Delta V$	ΔM_p						
0.01	1.0	1.0	19.0	18.6	> 50	> 50	> 50	> 50
0.04	< 0.1	< 0.1	1.0	1.0	20.0	17.8	> 50	> 50
0.1	< 0.1	< 0.1	0.1	0.1	3.2	2.9	25.0	14.5
0.4	< 0.1	< 0.1	< 0.1	< 0.1	0.2	0.2	2.1	1.3

13. TRANSFER BETWEEN TWO ORBITS IN DIFFERENT ORBITAL PLANES

In Chapter 12, transfers between coplanar circular and elliptical orbits about the Earth have been discussed. However, in reality the initial and final orbits will mostly not lie in the same plane, and the impulsive shots have to be applied in such a way that the inclination, i , and/or the right ascension of the ascending node, Ω , (also) change. To achieve such a non-coplanar orbit change, a component of the impulsive shot must be directed perpendicular to the initial orbital plane; after all, components of the impulsive shot in the orbital plane only change the in-plane orbital parameters. When the out-of-plane impulsive shot is applied at an arbitrary point along the orbit both the inclination and the right ascension of the ascending node, and some other orbital elements, will change. In this Chapter, first a general overview of the geometry of orbital plane changes will be given, after which a few basic types of three-dimensional maneuvers will be discussed in some detail. In all cases, it will be assumed that the maneuvers may be considered as impulsive shots. The additional ΔV required to compensate for the gravity loss that occurs due to a non-zero rocket engine burning time can be estimated with the method described in Section 12.7.

13.1. Geometry of orbital plane changes

Figure 13.1 depicts the general case of a maneuver for changing the orbital plane of a satellite about the Earth. It is assumed that, originally, the satellite is in an elliptical orbit with the elements $a_1, e_1, i_1, \Omega_1, \omega_1, \tau_1$. At point P in this orbit, which is determined by the argument of latitude u_1 , an impulsive shot is applied such that an orbit with the elements $a_2, e_2, i_2, \Omega_2, \omega_2, \tau_2$ results. Consequently, point P is a point where both orbits intersect each other. For simplicity, it is assumed that the impulsive shot is executed over the northern hemisphere, which implies $u_1 < 180^\circ$, $u_2 < 180^\circ$, and that $\Omega_2 > \Omega_1$, $i_2 > i_1$. The velocity prior to the impulsive shot is V_1 . The impulsive shot results in the change in velocity ΔV , which has a component perpendicular to the original orbital plane. The velocity just after the impulsive shot is V_2 . Now, consider the oblique spherical triangle shown in Figure 13.1, which is determined by the angles $i_1, 180^\circ - i_2$ and A , and the corresponding (spherical) sides u_2, u_1 and $\Omega_2 - \Omega_1$. Note that the curves $\Omega_1 - P$ and $\Omega_2 - P$ are the intersections of the orbital planes with a sphere. The angle α is given by

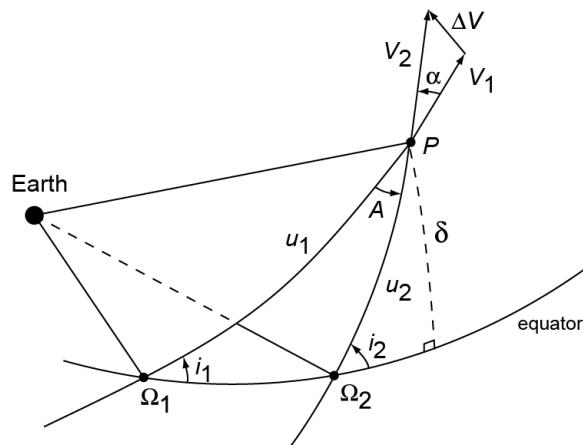


Figure 13.1: Geometry of a general orbital plane change maneuver.

$$\cos \alpha = \frac{\bar{V}_1 \cdot \bar{V}_2}{V_1 V_2}$$

For specified initial and final elliptical orbits, the vectors \bar{V}_1 and \bar{V}_2 are known and thus the values of ΔV and α can be computed. It is emphasized that, in general, $\alpha \neq A$. The reason is that at point P the velocity in both elliptical orbits will, in general, have a radial component, while the angle A is measured on a sphere. Only when point P is perigee or apogee of both orbits, or when the orbits are circular: $\alpha = A$; for low-eccentricity orbits: $\alpha \approx A$. Of course, $A \neq 0$, since we deal with two crossing orbital planes. To simplify our analysis, in this Chapter we will compute the magnitude of an impulsive shot only for cases where we may substitute α for A . From spherical trigonometry a whole suite of formulas is known that can be applied for the oblique spherical triangle indicated in Figure 13.1. For our analysis we only need a subset of these formulas.

Sine rule

$$\frac{\sin A}{\sin(\Omega_2 - \Omega_1)} = \frac{\sin(\pi - i_2)}{\sin u_1} = \frac{\sin i_1}{\sin u_2}$$

With the definition $\Delta\Omega = \Omega_2 - \Omega_1$, this expression can be written as

$$\frac{\sin A}{\sin \Delta\Omega} = \frac{\sin i_2}{\sin u_1} = \frac{\sin i_1}{\sin u_2} \quad (13.1)$$

First cosine rule

$$\cos u_2 = \cos u_1 \cos \Delta\Omega + \sin u_1 \sin \Delta\Omega \cos i_1 \quad (13.2)$$

Second cosine rule

$$\cos A = -\cos i_1 \cos(\pi - i_2) + \sin i_1 \sin(\pi - i_2) \cos \Delta\Omega$$

$$\cos i_1 = -\cos(\pi - i_2) \cos A + \sin(\pi - i_2) \sin A \cos u_2$$

$$\cos(\pi - i_2) = -\cos A \cos i_1 + \sin A \sin i_1 \cos u_1$$

or

$$\cos A = \cos i_1 \cos i_2 + \sin i_1 \sin i_2 \cos \Delta\Omega \quad (13.3)$$

$$\cos i_1 = \cos i_2 \cos A + \sin i_2 \sin A \cos u_2 \quad (13.4)$$

$$\cos i_2 = \cos i_1 \cos A - \sin i_1 \sin A \cos u_1 \quad (13.5)$$

Formula of Napier

One of the formulas derived by J. Napier (or Neper (1550-1617)) reads

$$\frac{\tan \frac{1}{2}(i_1 + \pi - i_2)}{\cot \frac{1}{2}A} = \frac{\cos \frac{1}{2}(u_2 - u_1)}{\cos \frac{1}{2}(u_2 + u_1)}$$

With the definition $\Delta i = i_2 - i_1$, this expression can be written as

$$\tan \frac{1}{2} \Delta i = \tan \frac{1}{2} A \frac{\cos \frac{1}{2}(u_2 + u_1)}{\cos \frac{1}{2}(u_2 - u_1)} \quad (13.6)$$

Figure 13.1 also shows a right-angle spherical triangle with the declination, δ , of point P as one of the right-angle sides. When we apply the sine rule in this spherical triangle, we find

$$\sin \delta = \sin i_1 \sin u_1 \quad (13.7)$$

Note that we have assumed that $i_1 \neq 0$ and $i_2 \neq 0$, since otherwise $\Delta\Omega$ is not defined.

In Sections 13.2 and 13.3 it will be shown that large changes in i and Ω cost a large amount of propellant. Therefore, in all practical situations $|\Delta i| < 90^\circ$, $|\Delta\Omega| < 90^\circ$, $A < 90^\circ$. Now, we assume that the elements of the initial orbit are known and that we want to compute the values of A , u_1 and u_2 that correspond to specified values of Δi and $\Delta\Omega$. These values of A , u_1 and u_2 can be computed unambiguously from (13.1) to (13.5). However, since we have assumed that the impulsive shot is applied over the northern hemisphere: $u_1 < 180^\circ$ and $u_2 < 180^\circ$, and therefore the angles can be computed unambiguously from the reduced set of equations (13.3) to (13.5):

$$\cos A = \cos i_1 \cos i_2 + \sin i_1 \sin i_2 \cos \Delta\Omega$$

$$\cos u_1 = \frac{\cos i_1 \cos A - \cos i_2}{\sin i_1 \sin A}$$

$$\cos u_2 = \frac{\cos i_1 - \cos i_2 \cos A}{\sin i_2 \sin A}$$

For a numerical example, we assume $i_1 = 20^\circ$, $i_2 = 40^\circ$, $\Delta\Omega = 20^\circ$. From the relations given above, we then obtain $A = 22.1^\circ$, $u_1 = 35.7^\circ$, $u_2 = 18.1^\circ$, $\delta = 11.5^\circ$. The magnitude and direction of the required impulsive shot cannot be determined, because we have not specified the other parameters of both orbits. It is emphasized that the impulsive shot required to achieve Δi and $\Delta\Omega$ will, in general, also affect other orbital elements. So, generally, multiple impulsive shots are required to realize a transfer between two specified non-coplanar elliptical orbits. In the following, three relatively simple three-dimensional transfer maneuvers will be discussed.

13.2. Changing Ω without changing i

In this Section, we will analyze the case that an impulsive shot is applied to change the right ascension of the ascending node from Ω_1 to Ω_2 , while the inclination is held constant. This implies that the orbital plane rotates about the polar axis. We then find from (13.1) and (13.6)

$$\sin u_1 = \sin u_2 \quad ; \quad u_1 + u_2 = \pi \quad (13.8)$$

From these conditions, we obtain the solution

$$u_1 = \frac{1}{2}\pi + \epsilon \quad ; \quad u_2 = \frac{1}{2}\pi - \epsilon \quad (13.9)$$

where ϵ is a function of i_1 and $\Delta\Omega$. So, the values of u_1 and u_2 are equally spaced on either side of $u = 90^\circ$. From (13.4), (13.5) and (13.7) we find for $i_1 = 90^\circ$: $\epsilon = 0^\circ$, $u_1 = u_2 = 90^\circ$, $\delta = 90^\circ$. When $\Delta\Omega$ is small, we may, using (13.8), approximate (13.2) by

$$\cos(\pi - u_1) \approx \cos u_1 + \sin u_1 \Delta\Omega \cos i_1$$

which leads with (13.9) to

$$\epsilon \approx \arctan\left(\frac{1}{2} \cos i_1 \Delta\Omega\right) \quad (13.10)$$

From (13.10) we conclude that for $\Delta\Omega < 20^\circ$, which will be the case for most practical situations, $\epsilon < 9.9^\circ$. For a numerical example, we assume $i_1 = 45^\circ$, $\Delta\Omega = 20^\circ$. From the relations given above, we then obtain $A = 14.1^\circ$, $u_1 = 97.1^\circ$, $u_2 = 82.9^\circ$, $\epsilon = 7.1^\circ$, $\delta = 44.6^\circ$, where ϵ is computed from (13.9). When we compute ϵ from the approximate relation (13.10) we find $\epsilon = 7.0^\circ$.

We now assume that the initial and final orbits are circular. Since they both go through P , the radii of both orbits are equal, and consequently $V_2 = V_1$. Since in this case $\alpha = A$, we find from (13.3)

$$\cos \alpha = \cos^2 i_1 + \sin^2 i_1 \cos \Delta\Omega$$

or

$$1 - \cos \alpha = \sin^2 i_1 (1 - \cos \Delta\Omega) = 2 \sin^2 i_1 \sin^2 \frac{1}{2} \Delta\Omega \quad (13.11)$$

When $\Delta\Omega$ is small then α is small, which implies that the angle between \bar{V}_1 and $\Delta\bar{V}$ will be close to 90° . According to the geometry depicted in Figure 13.1 we can write for the magnitude of the required impulse

$$(\Delta V)^2 = V_1^2 + V_2^2 - 2 V_1 V_2 \cos \alpha \quad (13.12)$$

This expression leads with $V_2 = V_1$ and (13.11) to

$$(\Delta V)^2 = 2 V_1^2 (1 - \cos \alpha) = 4 V_1^2 \sin^2 i_1 \sin^2 \frac{1}{2} \Delta\Omega$$

or

$$\frac{\Delta V}{V_1} = 2 \sin i_1 \sin \frac{1}{2} \Delta\Omega \quad (13.13)$$

This relation shows that, for prescribed values of V_1 and $\Delta\Omega$, the required impulse depends on the inclination of the initial orbit; the impulsive shot is maximum for polar orbits. For $\Delta\Omega = 60^\circ$ and $i_1 = 90^\circ$: $\Delta V = V_1$, which means that the velocity change required for the orbital plane rotation is equal to the velocity of the satellite just prior to the maneuver. This demonstrates that it costs very much propellant to produce large plane changes. As (13.13) shows, a change in Ω costs less propellant when V_1 is smaller. In Figure 13.2, the value of $y = \Delta V / \sin i_1$ is plotted as a function of $x = \Delta\Omega$ for circular orbits with specified orbital altitudes. The Figure shows that the higher the circular orbit (smaller V_1) the lower the impulse required to accomplish a specified change of the right ascension of the ascending node. For $h = 24,300$ km and given values of i_1 and $\Delta\Omega$, the required impulse is only about half the impulse required at an altitude of $h = 300$ km for the same values of i_1 and $\Delta\Omega$. For small values of $\Delta\Omega$: $\Delta V = V_1 \sin i_1 \Delta\Omega$, which fully agrees with the results found in Section 22.7.

When sufficient time is available, (part of) the required change in Ω may be realized by using the orbit perturbation due to the flattening of the Earth (Section 21.2, Section 23.2). That perturbation results, without demanding any rocket propellant, in a slow continuous precession

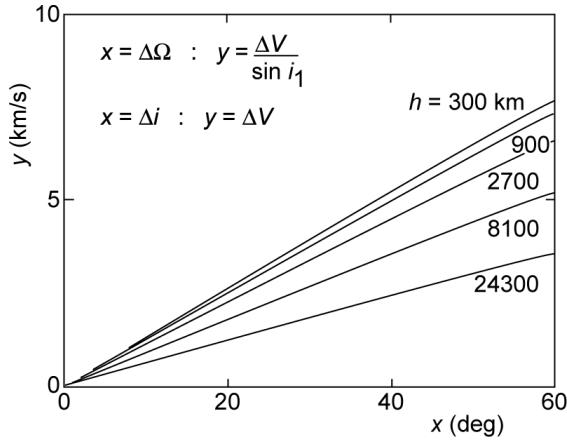


Figure 13.2: Required impulsive shot for a specified change of the right ascension of the ascending node or the inclination, as a function of the altitude of the initial circular orbit.

of the ascending node. However, this approach can only be applied if $\Delta\Omega < 0$ when $i_1 < 90^\circ$, or if $\Delta\Omega > 0$ when $i_1 > 90^\circ$ (Section 21.2).

13.3. Changing i without changing Ω

In this Section, we will analyze the case that an impulsive shot is applied to change the inclination of the orbit from i_1 to i_2 , while the right ascension of the ascending node is held constant. This implies that the orbital plane rotates about the line of nodes. We then find from (13.1) and (13.2)

$$\sin u_1 = 0 \quad ; \quad \sin u_2 = 0 \quad ; \quad \cos u_1 = \cos u_2 \quad (13.14)$$

From these conditions, we find the solution

$$u_1 = u_2 = k\pi \quad , \quad k = 0, 1 \quad (13.15)$$

So, the maneuver has to be executed at the ascending or descending node ($\delta = 0^\circ$). From (13.3) we then obtain

$$A = \Delta i \quad (13.16)$$

We now assume that the initial and final orbits are circular. Since then $\alpha = A$, $V_2 = V_1$, we find from (13.12) and (13.16)

$$\frac{\Delta V}{V_1} = 2 \sin \frac{1}{2} \Delta i \quad (13.17)$$

This equation shows that for $\Delta i = 60^\circ$: $\Delta V = V_1$. This again demonstrates that it costs very much propellant to produce large plane changes. Equation (13.17) shows that a change in inclination costs less propellant when V_1 is smaller. Figure 13.2 shows a plot of $y = \Delta V$ versus $x = \Delta i$. For circular orbits, the higher the orbit the lower the impulse required to accomplish a specified change in inclination. Also in this case, for a given value of Δi the impulsive shot required at an altitude of 24,300 km is only about half the impulse required at an altitude of 300 km. For small values of Δi : $\Delta V = V_1 \Delta i$, which fully agrees with the results found in Section 22.7.

In Figure 13.3 the geometry of an inclination change maneuver at the ascending node ($u_1 = 0^\circ$) of two circular orbits with the same orbital radius is depicted. Since $V_2 = V_1$, we can write for the

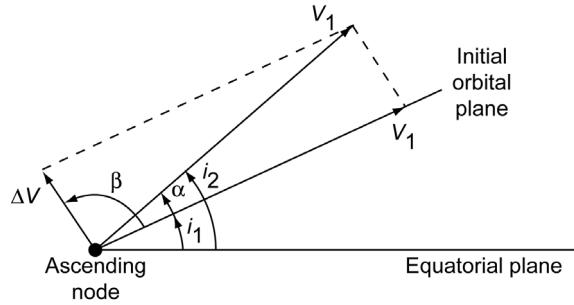


Figure 13.3: Geometry of an inclination change maneuver between two circular orbits with the same radius.

angle β , i.e. the direction of the impulsive shot with respect to the initial orbital plane

$$2(\pi - \beta) + \Delta i = \pi$$

or

$$\beta = \frac{1}{2}(\pi + \Delta i) \quad (13.18)$$

This relation shows that, generally, the required impulsive shot is not directed perpendicular to the initial orbital plane, but has a component opposite to the direction of the velocity in the initial orbital plane. This is to guarantee that the requirement $V_2 = V_1$ is satisfied. However, for small values of Δi we find that $\beta \approx 90^\circ$.

13.4. Transfer between two circular orbits with different inclinations

In the previous Sections, we have analyzed maneuvers to change the inclination and/or the right ascension of the ascending node of a satellite orbit. In reality, orbit change maneuvers will (mostly) be executed in such a way that several orbital parameters change simultaneously. In this Section, we will analyze the case of a transfer from an initial low-altitude circular orbit about the Earth with radius r_1 and inclination i_1 to a higher circular orbit with radius r_2 and inclination i_2 (Figure 13.4). Such maneuvers are frequently applied; for instance, when launching a geostationary satellite from a launch site that is not located on the equator. In Section 16.2, it will be shown that the latitude of the launch site determines the minimum inclination of the orbital plane of a satellite launched from that site. Therefore, the launcher usually, first injects the satellite into a low-altitude circular parking orbit with an inclination (about) equal to the latitude

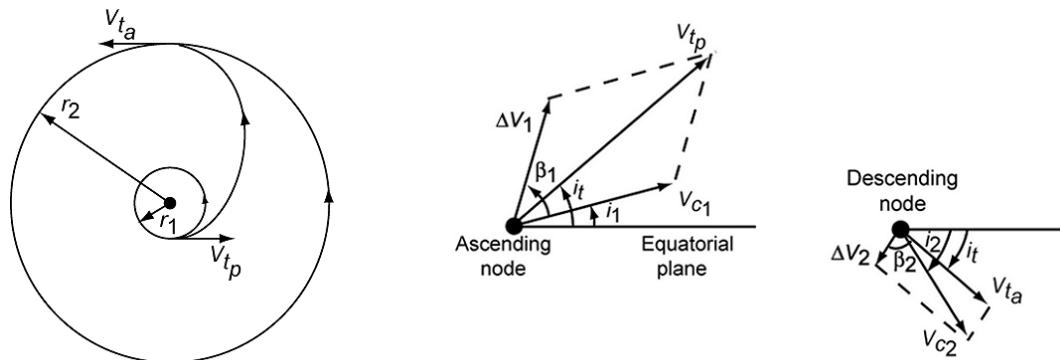


Figure 13.4: Geometry of a Hohmann transfer maneuver between two circular orbits with different inclinations.

of the launch site. At an ascending or descending node of this orbit, the satellite is injected into a Hohmann-type elliptical transfer orbit to the geostationary orbit ($i_2 = 0^\circ$). Just as in Chapter 12, it will be assumed that the total maneuver is performed by the same rocket engine and that in between rocket engine burns no mass is discarded or expelled.

It will be clear that the first impulsive shot, ΔV_1 , at a node of the initial circular orbit has to be applied such that this point becomes the perigee of the transfer orbit; this means that the impulsive shot should be directed perpendicular to the radius vector. The transfer orbit has an inclination i_t and an apogee distance equal to the radius of the final orbit, r_2 . The second impulsive shot, ΔV_2 , is executed at apogee of this transfer orbit and is directed perpendicular to the radius vector. This point lies in the equatorial plane and is therefore a node of the final orbit (if $i_2 \neq 0^\circ$). In the following analysis, it is assumed that the orbit transfer starts at the ascending node of the initial orbit.

According to the vector diagrams in Figure 13.4, we may write

$$\begin{aligned}\Delta V_1^2 &= V_{c_1}^2 + V_{t_p}^2 - 2V_{c_1}V_{t_p} \cos(i_t - i_1) \\ \Delta V_2^2 &= V_{c_2}^2 + V_{t_a}^2 - 2V_{c_2}V_{t_a} \cos(i_2 - i_t)\end{aligned}\quad (13.19)$$

With the standard relations for elliptical orbits, we write for the velocities at perigee and apogee

$$V_{t_p} = V_{c_1} \sqrt{1+e} \quad ; \quad V_{t_a} = V_{c_2} \sqrt{1-e} \quad ; \quad e = \frac{r_2 - r_1}{r_2 + r_1}$$

where e is the eccentricity of the transfer orbit, and for the circular velocity at the distance r_2

$$V_{c_2} = V_{c_1} \sqrt{\frac{r_1}{r_2}}$$

Substitution of these relations into (13.19) and using the parameter n introduced in Chapter 12 ($n = r_2/r_1$), we obtain after some algebraic manipulation

$$\begin{aligned}\frac{\Delta V_1}{V_{c_1}} &= \sqrt{\frac{3n+1}{n+1} - 2\sqrt{\frac{2n}{n+1}} \cos(\Delta i^*)} \\ \frac{\Delta V_2}{V_{c_1}} &= \sqrt{\frac{1}{n} \left\{ \frac{n+3}{n+1} - 2\sqrt{\frac{2}{n+1}} \cos(\Delta i - \Delta i^*) \right\}}\end{aligned}\quad (13.20)$$

where

$$\Delta i^* = i_t - i_1 \quad (13.21)$$

Equations (13.20) and (13.21) show that for given values of r_1 , r_2 and Δi , the total impulse $\Delta V_{tot} = \Delta V_1 + \Delta V_2$ is a function of Δi^* . Generally, there will be an optimum value of Δi^* for which ΔV_{tot} is minimum. This optimum value, and the corresponding values of ΔV_1 and ΔV_2 , can be determined by means of a numerical analysis of (13.20). In Figure 13.5 the optimum value of Δi^* is plotted as a function of n , for several values of the total inclination change, Δi . The Figure shows that the value of Δi^* is always considerably smaller than the value of Δi , and that the

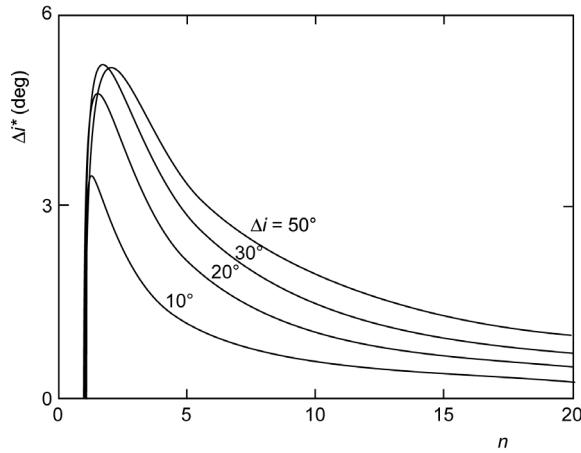


Figure 13.5: Optimum relative inclination of the transfer orbit, Δi^* , as a function of n and the total inclination change, Δi .

maximum value of Δi^* occurs at $n < 3$. For instance, for $\Delta i = 30^\circ$ the maximum value $\Delta i^* = 5.225^\circ$ occurs at $n = 1.75$. For a transfer from a circular orbit at an altitude of 300 km and with $i_1 = 28.5^\circ$ (launch from Cape Canaveral) to the geostationary orbit, we have $n = 6.32$, $\Delta i = -28.5^\circ$. Then, it follows from Figure 13.5 that $\Delta i^* = -2.19^\circ$. Because Δi^* is significantly smaller than Δi , in first-order analyses one usually assumes that the transfer orbit has the same inclination as the original circular orbit; i.e. $\Delta i^* = 0$.

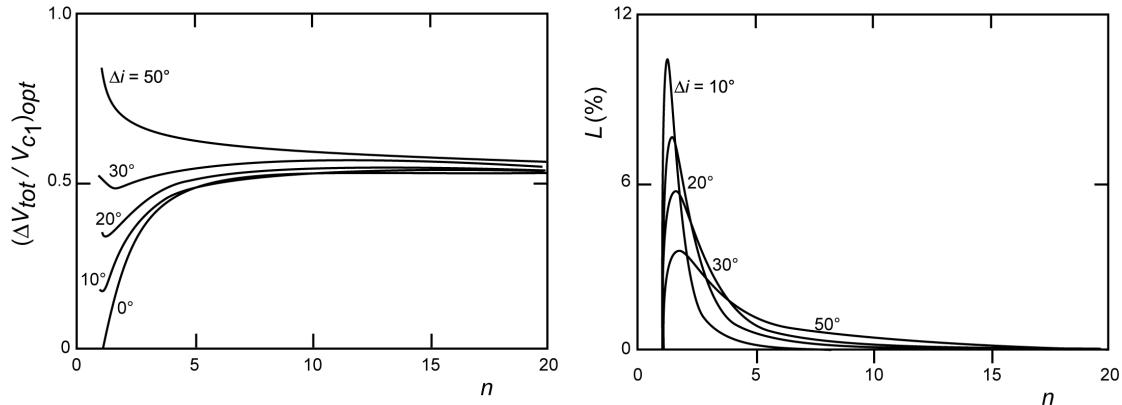


Figure 13.6: Dimensionless total impulse for an optimum Hohmann transfer between two inclined circular orbits (left) and the percentage increase in total impulse if $\Delta i^* = 0^\circ$ (right), both as a function of n and the total inclination change, Δi .

In Figure 13.6, the dimensionless total impulse for a transfer from a low-altitude circular orbit to a higher circular orbit is plotted as a function of n for some values of Δi ; the results hold for the case that the optimum transfer orbital inclination is applied. This Figure shows that for a total inclination change of about 30° : $\Delta V_{tot} \approx 0.5 V_{c1}$, independent of the value of n ; for $n > 5$: $\Delta V_{tot} \approx 0.5 - 0.6 V_{c1}$ for values of $|\Delta i|$ up to 50° . Also indicated is the percentage increase of the required impulse for the case that the transfer orbital inclination does not have its optimum value, but $i = i_1$. This percentage increase (loss) is defined as

$$L = \frac{(\Delta V_{tot})_{II} - (\Delta V_{tot})_I}{(\Delta V_{tot})_I} * 100\%$$

where the index I refers to the total impulse required when the optimum transfer orbital inclination is applied, and the index II to the total impulse required when the transfer orbit has the same inclination as the original circular orbit. As Figure 13.6 shows, the percentage increase is largest for small changes in inclination and then may amount up to 10%. For a transfer from a circular orbit at an altitude of 300 km and an inclination of 28.5° (launch from Cape Canaveral) to a geostationary orbit, the percentage increase amounts to about 0.6%.

For the thrust angles β_1 and β_2 , which determine the direction of the thrust vector in a plane perpendicular to the local radius vector and relative to the orbital plane just prior to the impulsive shots ΔV_1 and ΔV_2 , we can write according to Figure 13.4

$$V_{t_p}^2 = V_{c_1}^2 + \Delta V_1^2 - 2V_{c_1}\Delta V_1 \cos(\pi - \beta_1)$$

$$V_{c_2}^2 = V_{t_a}^2 + \Delta V_2^2 - 2V_{t_a}\Delta V_2 \cos(\pi - \beta_2)$$

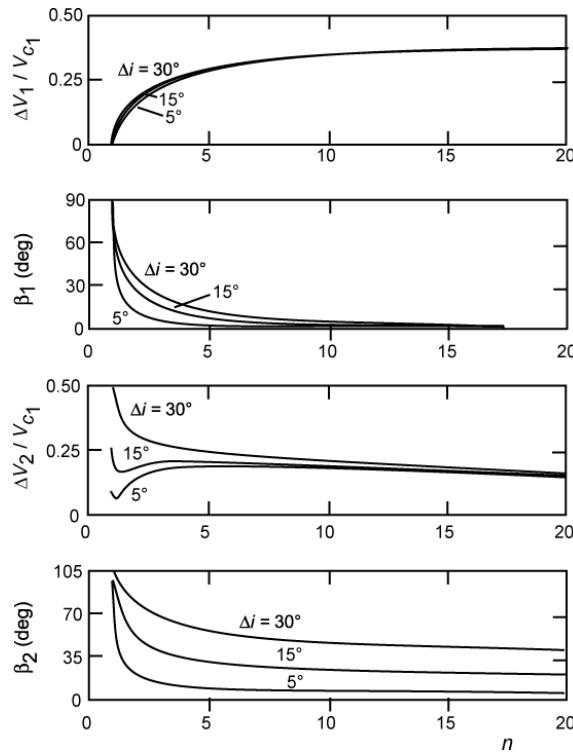


Figure 13.7: Dimensionless impulses and thrust angles for an optimum Hohmann transfer between two inclined circular orbits, as a function of n and the total inclination change, Δi .

With the standard relations for elliptical orbits ((12.19) to (12.21)), we can write these equations as

$$V_{c_1}^2 \left(\frac{n-1}{n+1} \right) - \Delta V_1^2 = 2V_{c_1}\Delta V_1 \cos\beta_1$$

$$V_{c_1}^2 \left(\frac{1}{n} \right) \left(\frac{n-1}{n+1} \right) - \Delta V_2^2 = 2V_{c_1} \sqrt{\frac{2}{n(n+1)}} \Delta V_2 \cos\beta_2$$

or

$$\begin{aligned}\cos\beta_1 &= \frac{1}{2} \left[\left(\frac{n-1}{n+1} \right) \frac{V_{c_1}}{\Delta V_1} - \frac{\Delta V_1}{V_{c_1}} \right] \\ \cos\beta_2 &= \frac{1}{2} \sqrt{\frac{n(n+1)}{2}} \left[\left(\frac{n-1}{n(n+1)} \right) \frac{V_{c_1}}{\Delta V_2} - \frac{\Delta V_2}{V_{c_1}} \right]\end{aligned}\quad (13.22)$$

From (13.20) and (13.21) the optimum inclination of the transfer orbit, and the corresponding values of ΔV_1 and ΔV_2 , can be computed for given values of n and Δi . Subsequently, the values of the thrust angles β_1 and β_2 , which hold for the optimum transfer orbit, can be computed with (13.22). In Figure 13.7 the values of $\Delta V_1/V_{c_1}$, $\Delta V_2/V_{c_1}$, β_1 and β_2 are plotted as a function of n and Δi for transfer orbits with the optimum inclination i_r . This Figure shows that for $n \downarrow 1$, the thrust is directed nearly perpendicular to the orbital plane at both locations where the impulsive shot is executed. For increasing values of n , the values of β_1 and β_2 decrease; for a given value of n , the values of β_1 and β_2 increase when Δi increases. For a transfer from a circular orbit at an altitude of 300 km and an inclination of 28.5° to the geostationary orbit, we find $\beta_1 = 9^\circ$ and $\beta_2 = 50^\circ$.

13.5. Inclination change at high altitude

In the previous Sections it has been shown that an orbital plane change is very costly in terms of ΔV . A comparison with results presented in Chapter 12 shows that an inclination change of 30° for a circular orbit at an altitude of 300 km costs as much ΔV as a coplanar Hohmann transfer to raise the orbital altitude from 300 km to 44,693 km. If the (effective) exhaust velocity of the rocket engine is 3 km/s, this ΔV requires, according to *Tsiolkovski's law* (Section 1.7), a propellant mass that is equal to about 74% of the initial vehicle mass. In Section 13.3 we have learned that the ΔV required for a plane change maneuver decreases if the maneuver is executed at higher altitudes. From Figure 12.4 we learn that a transfer from an initial 300 km altitude circular orbit to a coplanar elliptical orbit with perigee and apogee at 300 km altitude and 44,693 km altitude, respectively, costs about 64% of the total ΔV required to reach a circular orbit at an altitude of 44,693 km. These facts suggest that a transfer from a low circular orbit to a higher inclined circular orbit might be carried out most efficiently if the vehicle is first injected into a coplanar Hohmann orbit with high apogee, after which the full inclination change is performed at apogee and the vehicle descents to the final circular orbit along a Hohmann orbit that is coplanar with the final orbit. In this Section, we will analyze this high-altitude inclination change maneuver in some detail.

The geometry of the flight profile is depicted in Figure 13.8. For the velocities at perigee and apogee of the first transfer orbit we write with the standard relations for elliptical orbits

$$V_1 = V_{c_1} \sqrt{1 + e_{t_1}} \quad ; \quad V_{1_a} = V_{c_a} \sqrt{1 - e_{t_1}} \quad ; \quad e_{t_1} = \frac{r_a - r_1}{r_a + r_1}$$

where V_{c_a} is the circular velocity at apogee of the first transfer orbit and e_{t_1} is the eccentricity of this orbit. For the velocities at perigee and apogee of the second transfer orbit, we write

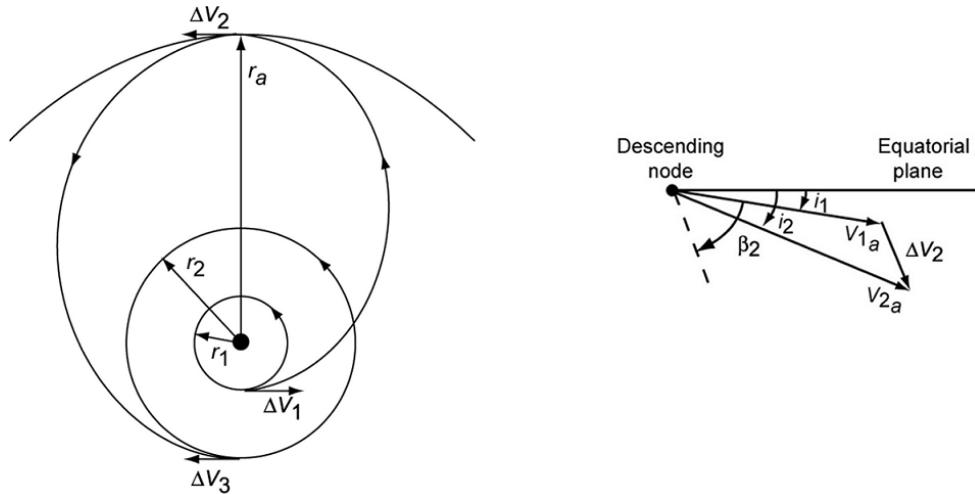


Figure 13.8: Geometry of a three-impulses Hohmann transfer maneuver between two inclined circular orbits, where the inclination change is executed at high altitude.

$$V_2 = V_{c_2} \sqrt{1 + e_{t_2}} \quad ; \quad V_{2a} = V_{c_a} \sqrt{1 - e_{t_2}} \quad ; \quad e_{t_2} = \frac{r_a - r_2}{r_a + r_2}$$

where e_{t_2} is the eccentricity of the second transfer orbit. When it is assumed that the total inclination change $\Delta i = i_2 - i_1$ is realized at apogee of the first transfer orbit by an impulsive shot ΔV_2 , we may write for the three impulsive shots (Figure 13.8)

$$\Delta V_1 = V_{c_1} \left[\sqrt{1 + e_{t_1}} - 1 \right] \quad (13.23)$$

$$\Delta V_2^2 = V_{c_a}^2 \left[(1 - e_{t_1}) + (1 - e_{t_2}) - 2\sqrt{(1 - e_{t_1})(1 - e_{t_2})} \cos \Delta i \right] \quad (13.24)$$

$$\Delta V_3 = V_{c_2} \left[\sqrt{1 + e_{t_2}} - 1 \right] \quad (13.25)$$

Substitution of the expressions for e_{t_1} and e_{t_2} , of the parameters $n = r_2/r_1$ and $m = r_a/r_1 > n$, and of the relations

$$V_{c_2} = \sqrt{\frac{1}{n}} V_{c_1} \quad ; \quad V_{c_a} = \sqrt{\frac{1}{m}} V_{c_1}$$

into (13.23) to (13.25) yields

$$\begin{aligned} \frac{\Delta V_1}{V_{c_1}} &= \sqrt{\frac{2m}{m+1}} - 1 \\ \frac{\Delta V_2}{V_{c_1}} &= \sqrt{\frac{2}{m} \left[\frac{1}{m+1} + \frac{n}{m+n} - 2\sqrt{\left(\frac{1}{m+1}\right)\left(\frac{n}{m+n}\right)} \cos \Delta i \right]} \end{aligned} \quad (13.26)$$

$$\frac{\Delta V_3}{V_{c1}} = \sqrt{\frac{1}{n} \left[\sqrt{\frac{2m}{m+n}} - 1 \right]} \quad (13.26)$$

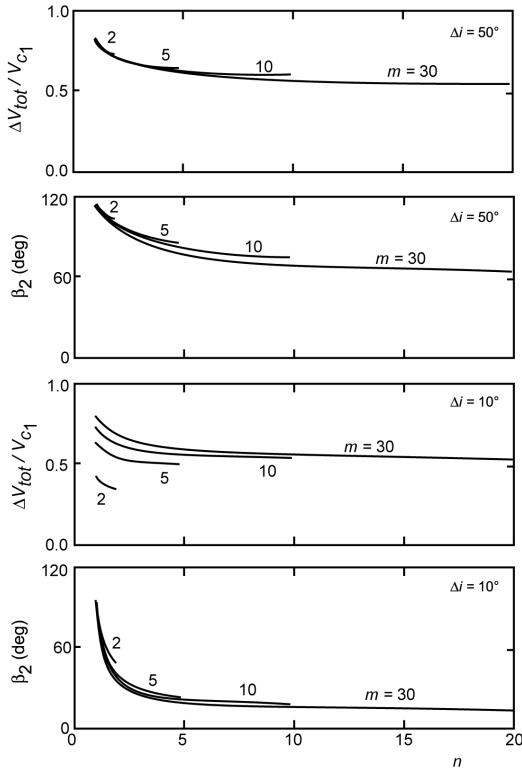


Figure 13.9: Total dimensionless impulse required for a three-impulses Hohmann transfer between two inclined circular orbits, where the total inclination change is executed at high altitude, and the thrust angle β_2 ; both as a function of the total inclination change, Δi , and the altitude ratios n and m .

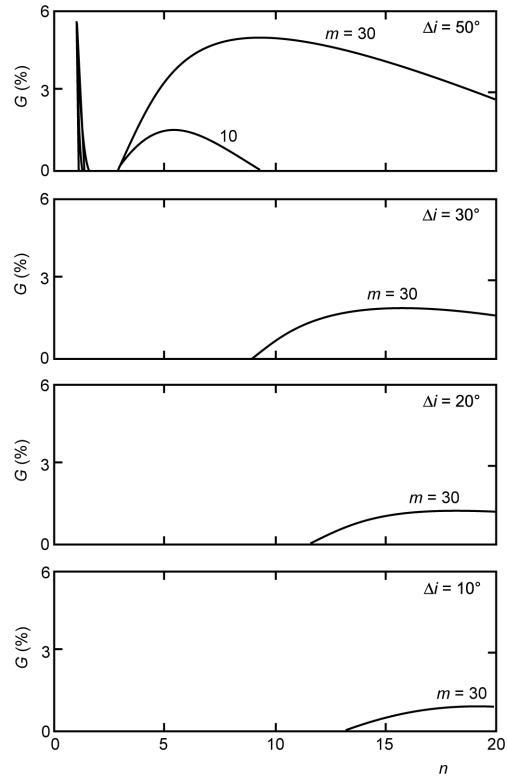


Figure 13.10: Percentage decrease of the total impulse required for a three-impulses Hohmann transfer between two inclined circular orbits, where the total inclination change is executed at high altitude, relative to a direct Hohmann transfer with an optimum inclination change distribution over the two impulses, as a function of the total inclination change, Δi , and the altitude ratios n and m .

and for the total impulse

$$\begin{aligned} \frac{\Delta V_{tot}}{V_{c1}} = & \sqrt{\frac{2m}{m+1}} + \sqrt{\frac{1}{n} \left[\sqrt{\frac{2m}{m+n}} - 1 \right]} - 1 \\ & + \sqrt{\frac{2}{m} \left[\frac{1}{m+1} + \frac{n}{m+n} - 2 \sqrt{\left(\frac{1}{m+1} \right) \left(\frac{n}{m+n} \right) \cos \Delta i} \right]} \end{aligned} \quad (13.27)$$

A numerical analysis of (13.27) shows that, for specified values of n and Δi , $\Delta V_{tot}/V_{c1}$ first increases, decreases or fluctuates at increasing values of m ; beyond a certain value of $m > n$, $\Delta V_{tot}/V_{c1}$ approaches asymptotically its value at $m \rightarrow \infty$. For $m \rightarrow \infty$, (13.27) reduces to

$$\left(\frac{\Delta V_{tot}}{V_{c_1}} \right)_{m \rightarrow \infty} = (\sqrt{2} - 1) \left(1 + \sqrt{\frac{1}{n}} \right)$$

This relation shows that for $m \rightarrow \infty$ the inclination change maneuver costs no energy and ΔV_{tot} consists of a component that represents the ΔV required to leave the initial circular orbit and to enter a parabolic orbit, and a component that represents the ΔV required to return from a parabolic orbit and to enter the final circular orbit. In Figure 13.9 the value of $\Delta V_{tot}/V_{c_1}$ is plotted as a function of n and m for two values of Δi . This Figure shows that for both values of Δi and a specified value of m , the value of $\Delta V_{tot}/V_{c_1}$ monotonously decreases for increasing values of n .

For the flight profile that is analyzed in this Section, the required orientation of the impulsive shot ΔV_2 can be found from the vector diagram in Figure 13.8:

$$V_{2_a}^2 = V_{1_a}^2 + \Delta V_2^2 - 2 V_{1_a} \Delta V_2 \cos(\pi - \beta_2)$$

After substitution of a number of relevant relations for elliptical orbits, we find

$$\cos \beta_2 = \sqrt{\frac{m+1}{2}} \left[\sqrt{\frac{1}{m}} \left\{ \frac{n}{m+n} - \frac{1}{m+1} \right\} \frac{V_{c_1}}{\Delta V_2} - \frac{1}{2} \sqrt{m} \frac{\Delta V_2}{V_{c_1}} \right] \quad (13.28)$$

From (13.26-2) we can compute the value of $\Delta V_2/V_{c_1}$ as a function of Δi , n and m . Then, we can compute from (13.28) the corresponding value of β_2 . In Figure 13.9, the value of β_2 is plotted as a function of n and m for two values of Δi . The Figure shows that for $n \approx 1$ the thrust vector is inclined at an angle $\beta_2 = 95^\circ - 115^\circ$ to the original orbital plane; for increasing values of n the value of β_2 decreases monotonously.

The equations (13.20) and (13.21) can be used to compute the minimum value of $\Delta V_{tot}/V_{c_1}$ when a direct transfer orbit is used and the total inclination change is divided in an optimum way over the two impulsive shots. Equation (13.27) yields the value of $\Delta V_{tot}/V_{c_1}$ when the flight profile described in this Section is applied. When the percentage decrease (gain) in the required ΔV_{tot} for this three-impulses profile is indicated by G , which is defined as

$$G = - \frac{(\Delta V_{tot})_3 - (\Delta V_{tot})_2}{(\Delta V_{tot})_2} * 100\%$$

where the indices 2 and 3 refer to the optimum two-impulses flight profile and the three-impulses flight profile, respectively, then G can be computed as a function of Δi , n and m . In Figure 13.10 the value of G is plotted as a function of n and m for four values of Δi , and for $G > 0$. The Figure shows that for small values of Δi only a minor percentage decrease in ΔV_{tot} may occur and only for large values of m . For $\Delta i = 50^\circ$, a ΔV_{tot} decrease of the order of 5% may be achieved for small values of n and a value of $m = 2 - 10$; for $n > 5$, a ΔV_{tot} decrease of 3 - 5% may be achieved if $m > 25$. The total flight time for such a three-impulses flight profile is, of course, significantly longer than for a two-impulses profile. For satellites that have to operate for extensive periods of time in their final orbits, this additional flight time is irrelevant and an inclination change maneuver at high altitude certainly is an attractive option.

13.6. Orbital plane changes in practice

In this and previous Chapters it has been assumed that all maneuvers are executed by the same rocket engine and that no mass is expelled or discarded between successive impulsive shots. So, it is implicitly assumed that no staging is performed. In practice, however, the various maneuvers are often executed by two or more rocket stages and parts of the initial vehicle are discarded between the maneuvers. An example is a geostationary transfer flight where the perigee burn is executed by the last rocket stage, after which the satellite is separated from that stage, travels along its geostationary transfer trajectory and, after arriving at geostationary altitude, fires its own apogee rocket engine. The advantage of such a maneuver is clear: by separating the satellite from the rocket stage useless mass is discarded and the propulsion system that performs the apogee burn can be used more efficiently. This is a direct consequence of *Tsiolkovski's law* (Section 1.7). Another example of a real-world strategy for a geostationary mission is the appropriate selection of the inclination of the initial parking orbit. It will be shown in Chapter 16 that the payload capability of a rocket is maximum when the rocket injects its payload into an orbit with inclination equal to the latitude of the launch site. However, depending on the launcher characteristics, the situation may occur that by selecting a parking orbital inclination that is not equal to the latitude of the launch site, a larger payload can be placed in geostationary orbit. In that case, the launcher payload capability loss is less than the gain in payload due to the smaller orbital plane change to be provided by the satellite's rocket engine, requiring less propellant on board of the satellite. So, in reality, the determination of an optimal strategy to launch the maximum mass into the final orbit by a specific launcher leaving from a specific launch site is a complicated problem. Important parameters in this optimization procedure are, among others, the launcher's payload capability as a function of the inclination of the ascent trajectory, the ratio between the mass of the satellite and the mass of the last rocket stage, the ratio between the exhaust velocity of the rocket engine of the satellite and that of the last rocket stage, etc. As an example, for the modern Atlas V/Centaur family of launchers and for today's satellite liquid-propellant apogee rocket engines, it was found that a 27° inclination of the geostationary transfer orbit is optimum for maximizing the satellite beginning-of-life mass in geostationary orbit, given an optimally-sized satellite propulsion system and a launch from Cape Canaveral (latitude 28.5°). The analysis of such optimum launch procedures falls outside the scope of this book. In order to provide at least some information about real-world maneuvers, two geostationary transfer schemes and a few specific orbit transfer maneuvers are briefly described below.

With satellites weighing less than the geostationary transfer orbit capability of the launch vehicle, excess launcher performance can be used to further reduce inclination, raise perigee or both. In addition, alternative transfer flight schemes can enhance the lifetime of geostationary satellites that use common sources of liquid propellant for orbit insertion and on-orbit stationkeeping. These are: *supersynchronous transfer* and *subsynchronous transfer*. The supersynchronous trajectory design is based on an intermediate orbit with apogee altitude significantly higher than the geostationary altitude. If the apogee altitude capability exceeds the satellite maximum allowable altitude, excess launch vehicle performance can be used to lower the transfer orbital inclination. At supersynchronous altitudes, the decreased velocity allows the satellite to make orbital plane changes more efficiently (Section 13.5), and so offers an opportunity to increase the amount of satellite beginning-of-life propellants enhancing the satellite's operational lifetime. The satellite makes the plane change and raises perigee to geostationary altitude in one or more apogee burns. It then coasts to perigee and circularizes into the final geostationary orbit. These supersynchronous transfer trajectories have been flown on many missions, starting with the

launch of the Eutelsat II F3 satellite by an Atlas II vehicle on December, 7, 1991. In a subsynchronous transfer scheme, the launch vehicle delivers the satellite with a mass larger than the launcher's geostationary transfer capability to an intermediate transfer orbit with apogee altitude less than the geostationary altitude and with an inclination of about 27° for a launch from Cape Canaveral. The separated satellite executes a series of engine burns when it passes perigee of successive revolutions in its transfer orbit, thereby gradually raising the apogee to geostationary altitude. This series of maneuvers is quite similar to the scheme presented in Section 12.3. After apogee has reached the geostationary altitude, the satellite fires its rocket engine during one or more apogee passages to lower the orbital inclination and to circularize the orbit. Also this mission concept was applied in many missions. By using these alternative transfer concepts, the operational lifetime of communications satellites with a designed lifetime of 10 years may be extended by up to 6 months.

Many modern geostationary communications satellites are equipped with ion engines for on-orbit stationkeeping (Section 19.1). These engines have very high exhaust velocities, which make them very attractive for orbit maneuvers. The satellites use their liquid-propellant orbit maneuvering engines and ion engines in combination to execute the orbit change maneuvers required for the supersynchronous and subsynchronous transfer concepts. By using the ion engines, as much as 450 kg of chemical propellants may be saved for a communications satellite with an initial mass of 5200 kg. This mass saving may be applied to increase the revenue-generating payload mass, to prolong operational service life, or to change to a less-expensive launch vehicle.

As a first example of specific orbit transfer maneuvers, we consider the launch of the Galaxy VII communications satellite by an Ariane 42P rocket from the ESA launch site at Kourou, French Guiana, which is located at 5.2° northern latitude, on October 27, 1992. For this flight the subsynchronous transfer scheme was used. The Ariane 42P launcher has a payload capability of 2750 kg for a normal geostationary transfer orbit, which has a perigee altitude of 200 km and an apogee altitude of 36,000 km. However, in this flight the satellite was filled with a greater propellant load, which resulted in a geostationary transfer orbit with a lower than normal apogee altitude: 27,673 km. During a perigee pass of the satellite, the satellite's own propulsion system was fired to raise the apogee to 36,000 km. After this, the satellite propulsion system was used as in normal missions to progressively decrease the inclination and increase the perigee altitude, thereby circularizing the orbit.

On July 25, 1996, an Atlas II launched from Cape Canaveral the US Navy communications satellite UHF F7 into a geostationary transfer orbit. The Centaur second stage was fired for 6 min and 49 s to place the spacecraft in a preliminary parking orbit. This orbit was maintained for only 11 min, before the twin Centaur RL10A engines were ignited again for 91 s. This second burn was initiated as the vehicle reached the equator off West Africa, was continued until Centaur propellant depletion and placed the spacecraft in a subsynchronous intermediate transfer orbit of 285 km by 27,320 km and inclined at 27° . Starting July 27, 1996, a series of six maneuvers were executed with the satellite's liquid-fueled rocket engine to boost the spacecraft to geostationary altitude and to bring the inclination down to zero. These maneuvers resulted by early August 1996 in the placement of the satellite near its initial checking position in the geostationary orbit over the Pacific south of Hawaii.

On July 27, 1997, the Japanese Superbird-C communications satellite was launched by an Atlas IIAS rocket from Cape Canaveral. The twin RL10A-4-1 engines of the Centaur second stage were commanded through a two-burn sequence to place the satellite into a supersynchronous transfer orbit with a 90,000 km apogee altitude, 168 km perigee altitude and 27.55° inclination. The Centaur was fired to propellant depletion to obtain the high apogee.

Before separation of the spacecraft, the Centaur's attitude control system was used to spin up the payload to 5 rpm for initial attitude control stability. Once separation was achieved, the satellite's own thrusters increased the spin rate to 10 rpm and a series of burns with the satellite's 500 N thrust liquid-propellant apogee rocket engine was executed to reduce the inclination and to circularize the orbit at geosynchronous altitude. The reduction of inclination was, of course, performed at the high apogee.

On April 18, 2000, an Ariane 42L rocket launched the Galaxy 4R communications satellite from Kourou. The launcher injected the satellite into a subsynchronous transfer orbit with perigee altitude of 199.6 km, apogee altitude of 32,231 km, and inclination of 6.998° . The spacecraft's propellant tanks were filled to the brim, and the launcher lifted the satellite to the highest possible transfer orbit by continuing thrusting until propellant depletion. The satellite then used its own onboard propulsion system to reach its final geostationary orbit. On this flight, the total payload mass was 3,737 kg, while the Ariane 42L version offers nominal capacity of only 3,560 kg with a 99% probability of reaching the targeted geostationary altitude before propellant depletion. By thrusting until propellant depletion the launcher could aim at an apogee of 33,448 km with over 50% probability of reaching it, while maintaining a 99% probability of reaching an apogee higher than 29,604 km. The actual apogee altitude of 32,231 km was 2,627 km higher than the minimum guaranteed altitude, which directly translated into extra propellant on the satellite and therefore a longer service life.

On the first flight of the Atlas V launch vehicle from Cape Canaveral on August 21, 2002, Eutelsat's Hot Bird 6 communications satellite was injected into a supersynchronous transfer orbit. About 4 min after liftoff, the propellants of the first stage were used up and the engine shut down. The spent stage was then jettisoned from the Centaur upper stage, and the Centaur's RL10A main rocket engine was ignited to continue the push to orbit. This first Centaur burn lasted for about 11 min, injecting the stage and attached satellite into a temporary parking orbit about the Earth with perigee at 199 km altitude, apogee at 1482 km altitude and inclination of 27.4° . The duo coasted in that orbit for 9 min until the Centaur fired again for 4 min, accelerating the satellite into the planned deployment orbit with 315 km perigee altitude, 45,850 km apogee altitude and inclination of 17.57° . Unlike Centaur missions that see the stage fire until it depletes its propellants, this launch was a 'guidance commanded shutdown' flight, in which the onboard computer cut off the engine when the desired orbit was reached. The Centaur stage then performed some maneuvers to orient itself into the proper position to release the satellite at 31.5 min after liftoff. Once flying on its own, Hot Bird 6 performed a series of onboard engine firings to achieve the geostationary orbit at 13° eastern longitude.

On September 25, 2005, the US Air Force launched a Delta II 7925 rocket carrying the GPS IIR-14 (M) navigation satellite from Cape Canaveral at a flight path azimuth of 110° . At 4.4 min after liftoff the first-stage propellants were depleted and the engine cut off. Then, the second stage separated and the rocket engine of the second stage ignited and burned for about 6.8 min. This second stage first burn placed the vehicle in an orbit with perigee and apogee altitudes of 174 km and 1113 km, respectively, and an inclination of 37.77° . The second stage restarted at approximately 20 min after liftoff and burned again for about 7.6 s. After this maneuver the vehicle was in an orbit with perigee at 190 km altitude, apogee at 1294 km altitude, and inclination of 37.84° . Then, the third stage was spun up and separated, after which the third stage rocket engine was ignited for a thrusting period of about 1.4 min, injecting the satellite into a transfer orbit with perigee at 281 km altitude, apogee at 20,368 km altitude and inclination of 39.5° . Separation of the satellite from the third stage occurred at 24.3 min after liftoff. The satellite traveled in this transfer orbit and reached apogee at 3 hr, 19.5 min after liftoff. Then, the onboard solid-propellant rocket kick motor was ignited to insert the satellite into its operational

circular orbit at an altitude of 20,200 km and with an inclination of 55° .

In this Chapter, orbital plane change maneuvers were analyzed and discussed, where it was assumed that the satellite flies an elliptical orbit in between successive maneuvers. However, in practice this assumption is not always valid for transfer orbits with a very high apogee altitude. For these orbits, the gravitational attraction by the Sun and the Moon may lead to significant orbit changes from one perigee passage to the next, and the analysis approach presented in this Chapter breaks down. It can be shown that for such transfer orbits the gravitational perturbations by the Sun and the Moon can be used to achieve plane changes at reduced ΔV costs over the classical approaches. This type of third-body plane-change maneuver is somewhat comparable to the classical bi-elliptical plane-change maneuver (Section 13.5) with the apogee plane change maneuver replaced by the cost-free action of third-body perturbations. The analysis of these transfer trajectories is beyond the scope of this book.

14. PHASING ORBITS

In Chapters 12 and 13 transfers between an initial and a final orbit have been analyzed. Often, a transfer orbit has to be selected such that not only the satellite will arrive in the chosen final orbit, but also that it will arrive at a specific position in that orbit. This is, for example, the case for *rendez-vous missions* (Chapter 15). In that case, the satellite has to arrive in the orbit of another satellite and in the direct vicinity of that satellite. Another example is a geostationary satellite, which not only has to reach the geostationary orbit but also has to be positioned at a selected geographic longitude in that orbit. In this case, the target is not another satellite but a specified position in the geostationary orbit.

In most cases, a direct transfer will not bring the satellite to its desired final position and we have to use *phasing orbits*. The simplest phasing technique is to wait in the original orbit until a point in that orbit is reached where injection into the (Hohmann) transfer orbit has to take place in order to arrive at the desired position in the final orbit. Often, this phasing technique is inadequate, because of the long waiting time or because of other constraints, and the satellite is brought into a special interim phasing orbit. For different types of missions many phasing techniques can be applied. In this Chapter, only some simple scenarios will be analyzed. First, we will analyze the transfer between two coplanar circular orbits. As a second example, we will analyze the transfer between two circular orbits with different inclinations. In both cases, it will be assumed that the radius of the final orbit is larger than that of the initial orbit and that an elliptical transfer orbit is used. The satellite that is originally in the lower orbit will be indicated by ‘satellite 1’; the satellite in the higher orbit by ‘satellite 2’. For the phasing orbit, we will assume that $r_a \leq r_2$; $r_p \geq r_1$, where r_1 and r_2 are the radii of the initial and final orbit, respectively, and r_a and r_p are the apogee and perigee distances of the phasing orbit, respectively. For the case that the initial and final orbits are inclined relative to each other, we will assume that the two ascending nodes coincide. This is a realistic assumption, because we have learned in Chapter 13 that a change of the right ascension of the ascending node costs a lot of propellant. Therefore, usually the ascending nodes of both orbital planes are made to coincide by an appropriate choice of the launch time of satellite 1.

In this Chapter, we will not discuss the energy requirement of different phasing strategies. After all, the results of Chapter 12 have shown that in case the initial and final orbit lie in the same plane, the energy requirement is independent of the number of intermediate orbits, as long as $r_a \leq r_2$. From the results of Chapter 13, it can be concluded that, when the initial and final orbit have different inclinations, the energy requirement is only weakly dependent on the inclination of the selected intermediate orbit and that always the optimum inclination of that orbit can be determined. Finally, it is noted that, in practice, the orbits of both satellites are always known with a limited accuracy. Therefore, in practice, one will always select the transfer and phasing orbits such that satellite 1 will not arrive at the position of satellite 2 but in a slightly lower orbit at some distance ‘behind’ satellite 2. As a result of the different angular velocities of both satellites they will then approach each other gradually (*drift orbit*). For satellites that have to reach a specified position in the geostationary orbit, *drift techniques* are generally applied that make use of the east-west deviations of the Earth’s gravity field that generate along-track accelerations (Section 21.7). For all numerical examples in this Chapter, we will assume that the altitude of the initial circular orbit is 300 km, which corresponds to $r_1 = 6671$ km.

14.1. Geometric aspects of direct transfer orbits

In this Section, we will analyze some geometric aspects of direct transfer orbits. First, we will consider a transfer between two coplanar circular orbits and subsequently we will analyze a transfer between two inclined circular orbits.

Coplanar circular orbits

Assume that at t_0 both satellites are located at the points that are indicated in Figure 14.1 by $s_{1,0}$ and $s_{2,0}$. In this notation, s_1 and s_2 indicate the positions of satellite 1 and satellite 2,

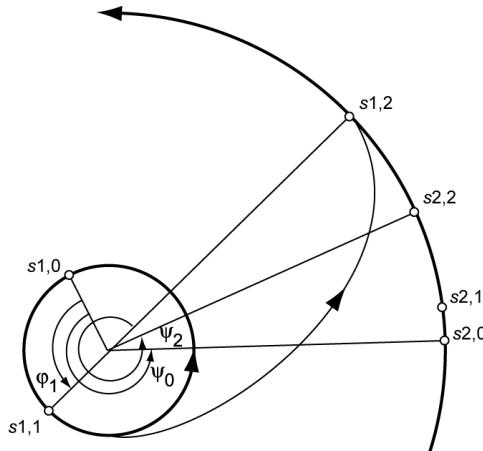


Figure 14.1: Geometry of a transfer flight between two coplanar circular orbits, with the notation used.

respectively, and the second index indicates the time when the satellite occupies that location. So, the notation $s_{1,1}$ indicates the position of satellite 1 at t_1 . The relative angular position of both satellites at t_0 is given by Ψ_0 ; their relative position at t_2 by Ψ_2 . Assume that satellite 1 first travels over an angular distance φ_1 in its orbit, before it is injected at t_1 into a Hohmann transfer orbit to reach the orbit of satellite 2. At t_2 , satellite 1 arrives in that orbit. When the flight time in the transfer orbit is indicated by t_f , then we find for the total flight time of satellite 1, t_f ,

$$t_f = t_2 - t_0 = \varphi_1/n_1 + t_f \quad (14.1)$$

where n_1 is the angular velocity (mean angular motion) in the low orbit. During this time interval, satellite 2 travels over an angular distance $n_2 t_f$, where n_2 is the angular velocity in the high orbit. From Figure 14.1 we obtain

$$\Psi_2 - \Psi_0 - (\pi - \varphi_1) = n_2 t_f$$

which, after combination with (14.1), leads to

$$\Psi_2 = \Psi_0 - \left(1 - \frac{n_2}{n_1} \right) \varphi_1 + n_2 t_f + \pi \quad (14.2)$$

From the theory of elliptical orbits (Chapter 6) we know that the following relations hold:

$$n_1 = \sqrt{\frac{\mu}{r_1^3}} \quad ; \quad n_2 = \sqrt{\frac{\mu}{r_2^3}} \quad ; \quad t_f = \pi \sqrt{\frac{(r_1 + r_2)^3}{8\mu}} \quad (14.3)$$

where μ is gravitational parameter of the Earth. Substitution of (14.3) into (14.2) yields

$$\Psi_2 = \Psi_0 - \left(1 - \sqrt{\left(\frac{r_1}{r_2} \right)^3} \right) \varphi_1 + \pi \left[\sqrt{\left\{ \frac{1}{2} \left(1 + \frac{r_1}{r_2} \right) \right\}^3} + 1 \right] \quad (14.4)$$

This expression shows that for given initial and final orbits (r_1 and r_2) and a given relative position of the satellites at t_0 (Ψ_0), the angle Ψ_2 is a function of φ_1 only. For any value of Ψ_2 the required value of φ_1 follows from (14.4). Therefore, any point in the final orbit can be reached by an appropriate selection of φ_1 . For *rendez-vous missions* or *interception missions* we must require $\Psi_2 = 2\pi$. In that case, we obtain from (14.4)

$$\varphi_1 = \frac{\pi \left[\sqrt{\left\{ \frac{1}{2} \left(1 + \frac{r_1}{r_2} \right) \right\}^3} + 1 \right] - (2\pi - \Psi_0)}{1 - \sqrt{\left(\frac{r_1}{r_2} \right)^3}} \quad (14.5)$$

When (14.5) yields a negative value of φ_1 , then injection has to take place before the relative position angle Ψ_0 is reached. A positive value of φ_1 indicates that satellite 1 has to stay in the low orbit for some time after the relative position angle reaches the value Ψ_0 , before injection into the transfer orbit has to take place.

If we consider the case where $\Delta r = r_2 - r_1 \ll r_1$, then we may linearize (14.5) and find

$$\varphi_1 \approx \frac{2}{3} \frac{r_1}{\Delta r} \Psi_0 - \frac{1}{2} \pi \quad (14.6)$$

This relation shows that if $\Psi_0 = 0^\circ$: $\varphi_1 \approx -\pi/2$ for all $\Delta r \neq 0$. So, in this case injection has to take place when satellite 1 is 90° behind the position for which the relative position angle Ψ_0 would be zero. If $\Psi_0 \neq 0^\circ$, then the value of φ_1 is proportional to Ψ_0 and inversely proportional to Δr . For very small values of Δr , satellite 1 will, in general, have to complete many revolutions in its initial orbit before injection into the transfer orbit should take place. When $r_2 \gg r_1$, we find from (14.5) $\Psi_0 - \varphi_1 \approx 2.031$ rad $\approx 116^\circ$. This means that injection into the transfer orbit has to take place at the moment that the geocentric angle between both satellites is about 116° , while satellite 2 is in the direction of motion ahead of satellite 1. When $r_2 \gg r_1$ and $\Psi_0 = 0^\circ$, we find $\varphi_1 = -116^\circ$. So, we may conclude that for all values of $\Delta r \neq 0$ and $\Psi_0 = 0^\circ$, φ_1 will vary between -90° and -116° . For the time it takes satellite 1 to reach the final orbit, we find from (14.1) and (14.3)

$$t_f = \sqrt{\frac{r_1^3}{\mu}} \left[\varphi_1 + \pi \sqrt{\left\{ \frac{1}{2} \left(1 + \frac{r_2}{r_1} \right) \right\}^3} \right]$$

or

$$\frac{t_f}{T_1} = \frac{\varphi_1}{2\pi} + \frac{1}{2} \sqrt{\left\{ \frac{1}{2} \left(1 + \frac{r_2}{r_1} \right) \right\}^3} \quad (14.7)$$

where T_1 is the period of the low-altitude orbit.

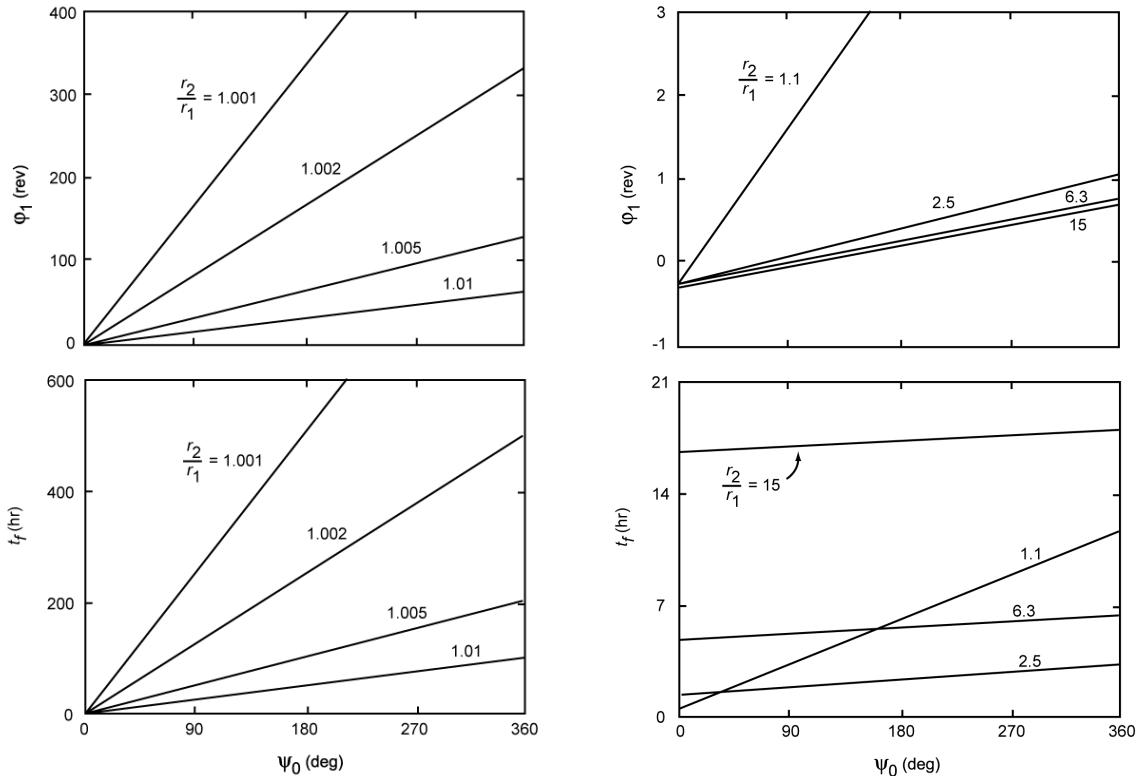


Figure 14.2: The values of φ_1 and the flight time for a transfer between two coplanar circular orbits as a function of ψ_0 and r_2/r_1 ; the flight time holds for $r_1 = 6671$ km.

In Figure 14.2, φ_1 and t_f are plotted as a function of r_2/r_1 and ψ_0 . Note that for small values of Δr and $\psi_0 > 90^\circ$, satellite 1 has to complete tens to hundreds of revolutions in its low orbit before it is injected into its transfer orbit, and that the flight time becomes very long. For large values of Δr and the same value of ψ_0 , the values of φ_1 and t_f are much smaller. Figure 14.2 (right, top) demonstrates what we already have found from (14.5), namely that for $\psi_0 = 0^\circ$: $\varphi_1 \approx -90^\circ$ to -116° , which corresponds to about a quarter of an orbital revolution. As a result of the scale of Figure 14.2 (top, left) this is not observable for values of $r_2/r_1 < 1.01$.

Now, it may happen that φ_1 cannot be chosen freely. For coplanar initial and final orbits, this may be the case if injection into the transfer orbit has to take place when the satellite is within the field of view of a ground station. Another example is when the orbital planes of the initial and the final orbits do not coincide and injection has to take place at a specified point of the initial orbit (Chapter 13). In the following, we will consider this case of mutually inclined orbits. Then, injection into the transfer orbit and into the final orbit have to take place when the satellite is at the intersecting line of the original orbital plane and the final orbital plane. The results that will be found for this case, of course, also apply for the case that the transfer takes place between two coplanar orbits but when the value of φ_1 is prescribed.

Inclined circular orbits

In Chapter 13 it was found that for a transfer between two circular orbits with different inclinations but equal right ascension of the ascending node, the impulsive shots have to be applied at the nodes of the orbits. These nodes lie in both orbital planes as well as in the equatorial plane. In Figure 14.3, the geometry of this problem and the notation used are indicated. Because the initial and the final orbit are in different planes, the angles are now defined with

respect to the intersecting line of the planes (line of nodes). In Figure 14.3 this is the line that connects the point where satellite 1 is injected into the transfer orbit ($s_{1,1}$) with the point where this satellite reaches the final orbit ($s_{1,2}$). The angles ψ_0 and ψ_2 are measured in the orbital plane of satellite 2. Of course, injection can also take place at the point in the low orbit that is located diametrically opposite to point $s_{1,1}$. In that case, the transfer orbit will be an exact mirror image. For simplicity, it will be assumed in this Chapter that injection always takes place at point $s_{1,1}$. Injection of satellite 1 into the transfer orbit is therefore only possible at each successive passage of point $s_{1,1}$.

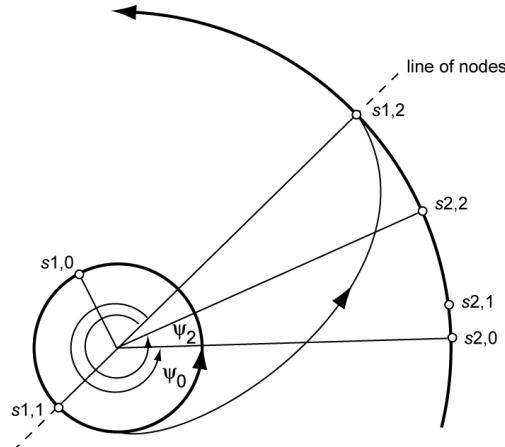


Figure 14.3: Geometry of a transfer flight between two inclined circular orbits, with the notation used.

Now, assume that injection occurs at the moment that satellite 1 passes point $s_{1,1}$ for the first time. The relative geocentric angle upon arrival in the orbit of satellite 2 is indicated by ψ_2 . When satellite 1 is injected k revolutions later, then the extra flight time of satellite 1 is

$$(\Delta t_f)_k = 2\pi \frac{k}{n_1}$$

In this period, satellite 2 covers an angular distance of

$$\Delta\psi_2 = 2\pi k \frac{n_2}{n_1}$$

The relative geocentric angle upon arrival then is

$$(\psi_2)_k = \psi_2 + 2\pi k \frac{n_2}{n_1} \quad (14.8)$$

When the difference between the angle ψ_2 after completing k revolutions in the low orbit and the angle ψ_2 at the first passage of the injection point is indicated by $(\Delta\psi_2)_k$, then we find from (14.3) and (14.8):

$$(\Delta\psi_2)_k = 2\pi k \sqrt{\left(\frac{r_1}{r_2}\right)^3}$$

So, $(\psi_2)_k$ increases by a discrete amount when satellite 1 completes an additional revolution in its low orbit before it is injected into the transfer orbit. The interval between two successive points of arrival in the high final orbit is called the *basic interval*. This interval is given in angular

units by

$$\alpha = 2\pi \sqrt{\left(\frac{r_1}{r_2}\right)^3} \quad (14.9)$$

When $2\pi/\alpha$ takes an integer value, satellite 1 can only reach the same discrete points in the orbit of satellite 2 after injection at any passage of point s1,1. The angular distance between successive points is then equal to α . When $2\pi/\alpha$ is not an integer value, satellite 1 can reach any point in the orbit of satellite 2, if it completes an appropriate number of revolutions in the low orbit. In general, to reach a given point in the final orbit, many revolutions in the low orbit may be needed and therefore usually a *phasing orbit* is introduced. This is an intermediate orbit with such a period that after a small number of revolutions in the low orbit and a small number of revolutions in this phasing orbit, an encounter between the two satellites in the high orbit occurs. In the following Sections, two characteristic types of phasing orbits will be analyzed. These are *low phasing orbits*, where $r_p = r_1$, $r_1 < r_a < r_2$, and *high phasing orbits*, where $r_a = r_2$, $r_1 < r_p < r_2$. Here, r_p and r_a are the perigee and apogee distances of the phasing orbit, respectively.

14.2. Geometric aspects of low phasing orbits

Figure 14.4 shows a diagram that indicates how the angle $(\psi_2)_k$ varies as a function of the number of revolutions that satellite 1 completes in the low orbit prior to injection into the transfer orbit. The position upon arrival in the orbit of satellite 2 is indicated relative to this satellite, causing satellite 2 to have a fixed position in the diagram.

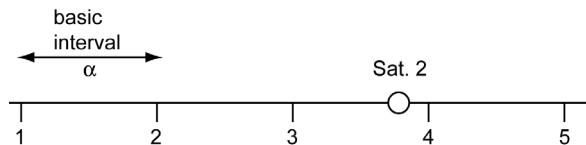


Figure 14.4: Series of possible relative positions of satellite 1 when it arrives in the orbit of satellite 2. Point 1 corresponds to injection into the transfer orbit when satellite 1 passes for the first time point s1,1 in Figure 14.3, etc.

It is assumed that when satellite 1 is injected into the transfer orbit during its first passage of point s1,1, it reaches the orbit of satellite 2 at point 1 of Figure 14.4. When injection takes place during its second passage of the injection point, the arrival in the high orbit takes place at point 2, etc. Successive points of arrival are separated by an angular distance α (basic interval) and are shifted oppositely to the direction of motion of satellite 2.

It will be obvious that for the case shown in Figure 14.4, an attractive strategy is to inject satellite 1 during its second passage of point s1,1 into a phasing orbit with $r_p = r_1$, $r_1 < r_a < r_2$, and with such a period that after one revolution in this orbit a second impulsive shot at s1,1 is applied to inject that satellite into a transfer orbit with $r_p = r_1$, $r_a = r_2$, where upon arrival in the high orbit it encounters satellite 2. Obviously, it is also possible to complete multiple revolutions in the phasing orbit. In that case, satellite 1 has to be injected into the phasing orbit one or more revolutions in its initial orbit earlier.

When we limit ourselves to the case that only one revolution in the phasing orbit is completed and we assume that an arbitrary number of revolutions in the low orbit are completed, then it follows that any point in the high orbit can be reached when the period of the phasing orbit

satisfies the condition

$$T_1 < T_{ph} < 2T_1 \quad (14.10)$$

where T_1 and T_{ph} indicate the period of the low orbit and the phasing orbit, respectively. For these periods we can write

$$T_1 = 2\pi \sqrt{\frac{r_1^3}{\mu}} ; \quad T_{ph} = 2\pi \sqrt{\frac{(r_1 + r_a)^3}{8\mu}} \quad (14.11)$$

Combining (14.10) and (14.11), we find

$$2\pi \sqrt{\frac{r_1^3}{\mu}} < 2\pi \sqrt{\frac{(r_1 + r_a)^3}{8\mu}} < 4\pi \sqrt{\frac{r_1^3}{\mu}}$$

or

$$1 < \frac{r_a}{r_1} < 32^{1/3} - 1 \approx 2.175 \quad (14.12)$$

Because always $r_a < r_2$, this strategy is only applicable if $r_2/r_1 > 2.175$. When this condition is not satisfied, then parts of the high orbit will not be accessible to satellite 1, unless it completes more revolutions in the phasing orbit.

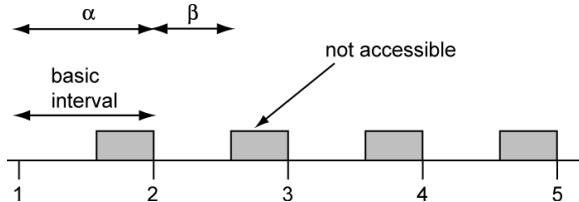


Figure 14.5: Illustration of the regions in the orbit of satellite 2 that are, for a specified value of $r_2/r_1 < 2.175$, not accessible (shaded area) for satellite 1, if satellite 1 completes only one revolution in the low phasing orbit.

When only one revolution in the phasing orbit is completed, and when for a given value of $r_2/r_1 < 2.175$ the size of the accessible part per basic interval α is indicated by β (Figure 14.5), then it follows that

$$\frac{\alpha + \beta}{\alpha} = \frac{T_{ph}^*}{T_1}$$

where T_{ph}^* is the period of a phasing orbit with $r_a = r_2 < 2.175 r_1$. From this relation and (14.11) we obtain

$$\frac{\beta}{\alpha} = \frac{T_{ph}^*}{T_1} - 1 = \sqrt{\left\{ \frac{1}{2} \left(\frac{r_2}{r_1} + 1 \right) \right\}^3} - 1 \quad (14.13)$$

For $1 < r_2/r_1 < 2.175$: $0 < \beta/\alpha < 1$. Combination of (14.9) and (14.13) yields for β , expressed in angular units,

$$\beta = 2\pi \left[\sqrt{\left\{ \frac{1}{2} \left(1 + \frac{r_1}{r_2} \right) \right\}^3} - \sqrt{\left(\frac{r_1}{r_2} \right)^3} \right] \quad (14.14)$$

from which we find for $1 < r_2/r_1 < 2.175$: $0 < \beta < 1.959$, or in degrees: $0^\circ < \beta < 112^\circ$. For $r_2/r_1 = 2.175$: $\alpha = \beta = 112^\circ$. In case $r_2/r_1 < 2.175$, the only possibility for still reaching any point in the high orbit is to inject the satellite into the phasing orbit one or more revolutions in the low orbit earlier and to complete this phasing orbit more than once.

Next, the case is considered that $r_2/r_1 > 2.175$ and that during the first passage of point s1,1 injection into a phasing orbit takes place such that

$$2.175r_1 < r_a < r_2 \quad (14.15)$$

In this case, a phasing orbit is applied with an apogee distance larger than $2.175 r_1$. In Figure 14.6, point 1 indicates the location where satellite 1 reaches the high orbit, if it is injected into the transfer orbit during its first passage of the injection point.

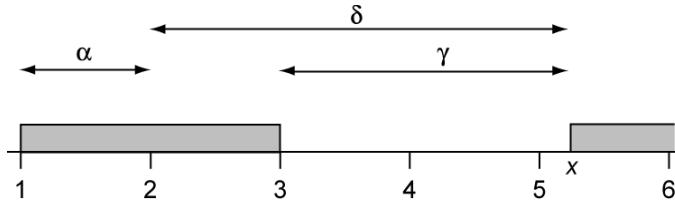


Figure 14.6: Illustration of the region in the orbit of satellite 2 that is, for specified values of $r_2/r_1 > r_a/r_1 > 2.175$, accessible (non-shaded region) for satellite 1, if satellite 1 is injected into the low phasing orbit when it passes point s1,1 for the first time and completes one revolution in the phasing orbit before it is injected into the transfer orbit.

When, before injection into the transfer orbit, the satellite would have completed one revolution in a phasing orbit with $r_a/r_1 = 2.175$, then satellite 1 would have arrived at point 3; when, before injection into the transfer orbit, the satellite would have completed one revolution in a phasing orbit with $r_a = r_2$, then satellite 1 would have arrived at point x. From Figure 14.6 we find

$$\frac{\gamma + 2\alpha}{\alpha} = \frac{T_{ph}^*}{T_1}$$

where T_{ph}^* is the period of a phasing orbit with $r_a = r_2$. Using (14.13) we may write this relation as

$$\frac{\gamma}{\alpha} = \sqrt{\left\{ \frac{1}{2} \left(\frac{r_2}{r_1} + 1 \right) \right\}^3} - 2 \quad (14.16)$$

For $2.175 < r_2/r_1 < \infty$: $0 < \gamma < \infty$. Combination of (14.9) and (14.16) yields for γ , expressed in angular units,

$$\gamma = 2\pi \left[\sqrt{\left\{ \frac{1}{2} \left(1 + \frac{r_1}{r_2} \right) \right\}^3} - 2 \sqrt{\left(\frac{r_1}{r_2} \right)^3} \right] \quad (14.17)$$

So, for $2.175 < r_2/r_1 < \infty$: $0 < \gamma < 2.221$, or in degrees: $0^\circ < \gamma < 127^\circ$.

From the results obtained so far, it may be concluded that when for $r_2/r_1 > 2.175$ satellite 1

is injected into a phasing orbit with $1 < r_a/r_1 < 2.175$ at the first passage of the injection point $s1,1$, any point in the high final orbit in between the points 2 and 3 can be reached. When during the first passage of the injection point the satellite is injected into a phasing orbit with $2.175 < r_a/r_1 < r_2/r_1$, then any point in the region between the points 3 and x can be reached. The total accessible interval is therefore $\alpha+\gamma$, measured from point 2. When this interval is indicated by δ (Figure 14.6), then we find from (14.9) and (14.17)

$$\delta = 2\pi \left[\sqrt{\left\{ \frac{1}{2} \left(1 + \frac{r_1}{r_2} \right) \right\}^3} - \sqrt{\left(\frac{r_1}{r_2} \right)^3} \right] \quad (14.18)$$

This expression is equal to the expression for β ((14.14)) that gives the accessible region per basic interval α when $r_2/r_1 < 2.175$. As a numerical example, we consider a transfer to geostationary orbit, for which $r_2/r_1 = 6.32051$. From (14.18) we then find $\delta = 217^\circ$.

14.3. Selection of a low phasing orbit

In this Section, we consider the general case where satellite 1, after passing for the first time point $s1,1$, completes k revolutions in the low orbit before it is injected into a phasing orbit with $r_p = r_1$, $r_1 < r_a < r_2$. In this phasing orbit the satellite completes l revolutions, after which it is injected at point $s1,1$ into the transfer orbit to reach the high orbit of satellite 2. The geometry of this case and the notation used are indicated in Figure 14.7. It is emphasized that the angles ψ_0 and ψ_e are measured in the orbital plane of satellite 2, while angle ϕ_0 is measured in the orbital plane of satellite 1.

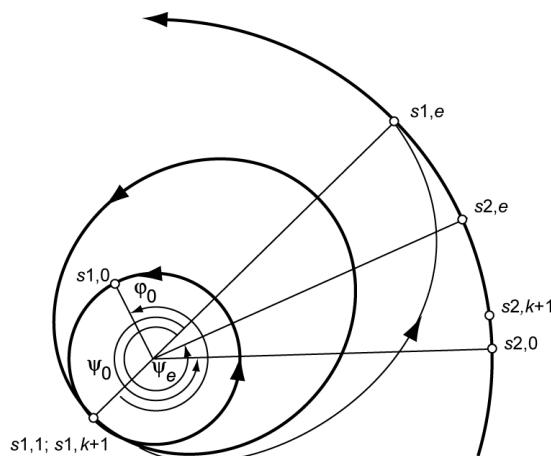


Figure 14.7: Geometry of a transfer flight between two inclined circular orbits, using a low phasing orbit, and the notation used.

For the total flight time of satellite 1, we can write

$$t_f = \frac{2\pi - \phi_0}{n_1} + k T_1 + l T_{ph} + t_t \quad (14.19)$$

where T_1 and T_{ph} are the periods of the low orbit and the phasing orbit, respectively, and t_t is the flight time in the transfer orbit. In this time interval, satellite 2 covers in its high orbit a geocentric angle of

$$\Psi_e = \Psi_0 + n_2 t_f \quad (14.20)$$

Both satellites meet each other when

$$\Psi_e = 2m\pi \quad (14.21)$$

The values of the integers k, l, m cover the range

$$k = 0, 1, 2, \dots ; \quad l = 0, 1, 2, \dots ; \quad m = 1, 2, 3, \dots$$

Substitution of (14.19) and (14.21) into (14.20) yields

$$2m\pi = \Psi_0 + n_2 \left[\frac{2\pi - \Phi_0}{n_1} + kT_1 + lT_{ph} + t_t \right]$$

or

$$lT_{ph} = \frac{2m\pi}{n_2} - \frac{\Psi_0}{n_2} - \frac{2\pi - \Phi_0}{n_1} - kT_1 - t_t \quad (14.22)$$

Substitution of (14.3) and (14.11) into (14.22) results, after some algebraic manipulation, in

$$l \sqrt{\left\{ \frac{1}{2} \left(\frac{r_a}{r_1} + 1 \right)^3 \right\}^3} = \frac{1}{2\pi} \left[\Phi_0 - \sqrt{\left(\frac{r_2}{r_1} \right)^3 \Psi_0} \right] - \sqrt{\frac{1}{32} \left(\frac{r_2}{r_1} + 1 \right)^3} + m \sqrt{\left(\frac{r_2}{r_1} \right)^3} - k - 1 \quad (14.23)$$

This relation shows that, for given values of Φ_0 , Ψ_0 and r_2/r_1 , in general, a number of combinations of k, l, m exist that yield a value of r_a with $r_1 < r_a < r_2$. All these combinations correspond to feasible phasing orbits; each with a different apogee distance.

A special case is a direct transfer without a phasing orbit, for which $k = l = 0 ; m = 1$. For this case, (14.23) simplifies to

$$\frac{1}{2\pi} \left[\Psi_0 + \sqrt{\left(\frac{r_1}{r_2} \right)^3 (2\pi - \Phi_0)} \right] = 1 - \sqrt{\frac{1}{32} \left(1 + \frac{r_1}{r_2} \right)^3}$$

This equation shows that for given values of Φ_0 and r_2/r_1 , such a transfer is only possible for one value of the relative position of satellite 2 at injection into the transfer orbit:

$$\frac{\Psi_0}{2\pi} = 1 - \sqrt{\frac{1}{32} \left(1 + \frac{r_1}{r_2} \right)^3} - \sqrt{\left(\frac{r_1}{r_2} \right)^3 \left(1 - \frac{\Phi_0}{2\pi} \right)} \quad (14.24)$$

For this case, the values of Ψ_0 and the total flight time are plotted in Figure 14.8 as a function of Φ_0 and r_2/r_1 . Note that for small values of r_2/r_1 the required value of Ψ_0 strongly depends on the value of Φ_0 , but that for large values of r_2/r_1 the value of Ψ_0 is for all values of Φ_0 within the interval $250^\circ - 300^\circ$.

In the following, we will consider again the case $l > 0$. In Section 14.2 it was found that for $r_2/r_1 > 2.175$ any point in the high orbit can be reached by selecting an appropriate phasing orbit and by completing that orbit only once ($l = 1$). In that case, the boundary conditions for the

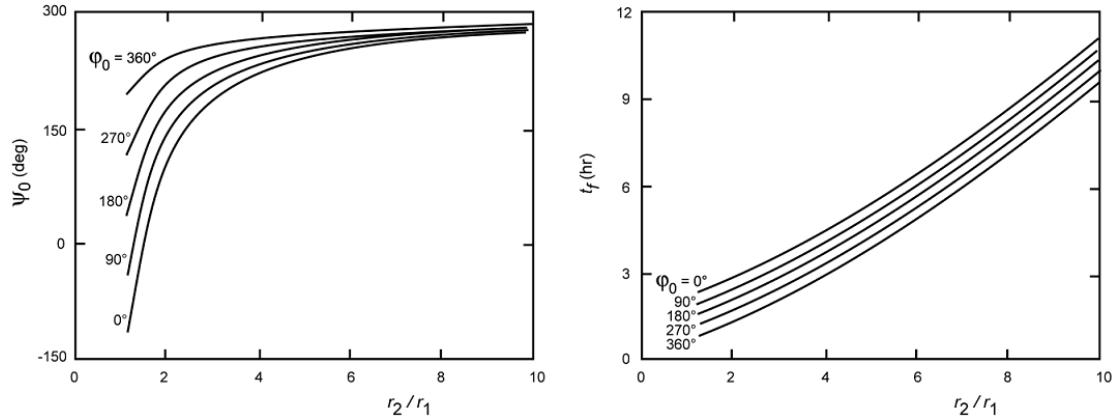


Figure 14.8: Values of ψ_0 and the flight time for a direct transfer ($k=l=0$, $m=1$) between two inclined circular orbits as a function of φ_0 and r_2/r_1 ; the flight time holds for $r_1 = 6671$ km.

apogee distance of the phasing orbit are given by (14.12). That relation can also be written as

$$1 < \sqrt{\left\{ \frac{1}{2} \left(\frac{r_a}{r_1} + 1 \right) \right\}^3} < 2$$

Substitution of (14.23) into this relation yields for $l = 1$:

$$1 < \frac{1}{2\pi} \left[\varphi_0 - \sqrt{\left(\frac{r_2}{r_1} \right)^3} \Psi_0 \right] - \sqrt{\frac{1}{32} \left(\frac{r_2}{r_1} + 1 \right)^3} + m \sqrt{\left(\frac{r_2}{r_1} \right)^3} - k - 1 < 2 \quad (14.25)$$

When a variable q is introduced, which is defined as

$$q = \frac{1}{2\pi} \left[\varphi_0 - \sqrt{\left(\frac{r_2}{r_1} \right)^3} \Psi_0 \right] - \sqrt{\frac{1}{32} \left(\frac{r_2}{r_1} + 1 \right)^3} - 3 \quad (14.26)$$

then we find from (14.25)

$$q + m \sqrt{\left(\frac{r_2}{r_1} \right)^3} < k < q + m \sqrt{\left(\frac{r_2}{r_1} \right)^3} + 1 \quad (14.27)$$

Equation (14.26) shows that for $l = 1$ the value of q is completely determined by the parameters φ_0 , Ψ_0 and r_2/r_1 , and is independent of k and m . Equation (14.27) shows that for any value of m and $l = 1$ there is not more than one value of k that satisfies this inequality. For the values of k and m , we have $k \geq 0$, $m \geq 1$. The computation scheme for possible phasing orbits is as follows. For given values of r_2/r_1 , φ_0 and Ψ_0 , the value of q is computed from (14.26). Next, for a series of selected values of m the corresponding values of k are determined with (14.27). For a set values of k and m , and $l = 1$, the required value of r_a/r_1 follows from (14.23).

In Section 14.2, it was also found that when $r_2/r_1 < 2.175$, in general, multiple revolutions in the phasing orbit are required in order to reach a given point in the high orbit. Obviously, also for $r_2/r_1 > 2.175$ multiple revolutions in the phasing orbit may be completed and therefore the following methodology is generally applicable, and the method described above may be regarded

as a special case of that methodology for $l = 1$, $r_2/r_1 > 2.175$. Because in the general case: $r_1 < r_a < r_2$, combination of (14.23) and (14.26) yields

$$l < q + 2 + m \sqrt{\left(\frac{r_2}{r_1}\right)^3} - k < l \sqrt{\left\{\frac{1}{2}\left(\frac{r_2}{r_1} + 1\right)\right\}^3}$$

or

$$q + 2 - l \sqrt{\left\{\frac{1}{2}\left(\frac{r_2}{r_1} + 1\right)\right\}^3} + m \sqrt{\left(\frac{r_2}{r_1}\right)^3} < k < q + 2 - l + m \sqrt{\left(\frac{r_2}{r_1}\right)^3} \quad (14.28)$$

In this case, the computation scheme for determining the possible phasing orbits is as follows. For given values of φ_0 , Ψ_0 and r_2/r_1 , the value of q is determined from (14.26). Now, a series of positive integers values of l and m is selected, starting with $l = 1$, $m = 1$. Subsequently, from (14.28) all possible integer values of $k \geq 0$ are determined. Then, for every set of values for k , l , m , (14.23) yields the corresponding value of r_a/r_1 .

Equation (14.28) shows that, in general, for given values of φ_0 , Ψ_0 and r_2/r_1 and selected values of l and m , multiple values of k are possible. Only one value of k is possible if

$$\left(q + 2 - l + m \sqrt{\left(\frac{r_2}{r_1}\right)^3}\right) - \left(q + 2 - l \sqrt{\left\{\frac{1}{2}\left(\frac{r_2}{r_1} + 1\right)\right\}^3} + m \sqrt{\left(\frac{r_2}{r_1}\right)^3}\right) < 1$$

or

$$\frac{r_2}{r_1} < 2\left(1 + \frac{1}{l}\right)^{2/3} - 1 \quad (14.29)$$

So, for $l = 1$, only one value of k is possible when $r_2/r_1 < 2.175$, which confirms the results obtained before. For $l = 2, 3$ and 5 , a transfer is possible for only one value of k when r_2/r_1 is smaller than 1.621, 1.423 and 1.258, respectively.

For the total flight time of satellite 1, we can write according to (14.20) and (14.21)

$$t_f = \frac{2m\pi - \Psi_0}{n_2} = 2\pi \sqrt{\frac{r_2^3}{\mu} \left(m - \frac{\Psi_0}{2\pi}\right)} = T_2 \left(m - \frac{\Psi_0}{2\pi}\right) \quad (14.30)$$

where T_2 is the period of the orbit of satellite 2. This equation shows that, for given values of Ψ_0 and r_2 , the total flight time of satellite 1 is determined by the value of m only, and not by the values of k and l . Therefore, generally, a phasing orbit will be selected for which the value of m is minimum. For this value of m , generally, multiple k,l -combinations are possible and the selection of k and l is done on basis of mission considerations.

In Table 14.1, possible combinations of k , l , m and the corresponding values of r_a/r_1 and t_f are listed for a transfer to geostationary orbit ($r_2/r_1 = 6.3205$), with $r_1 = 6671$ km, $\varphi_0 = \Psi_0 = 90^\circ$. The search process has covered the range $k = 0-8$, $l = 0-5$, and $m = 1-3$. The Table shows that for a given value of m indeed a number of combinations of k and l are possible and that the apogee distance of the phasing orbit ranges, for these k,l,m -combinations, from $1.174 r_1$ to $6.241 r_1$. In Figure 14.9, possible values of r_a/r_1 and the corresponding flight time are plotted as a function

Table 14.1: Characteristics of feasible transfer flights between two inclined circular orbits with $r_1 = 6671$ km and $r_2/r_1 = 6.3205$ (geostationary orbit), and using a low phasing orbit, for $\varphi_0 = \psi_0 = 90^\circ$. The search process has covered the range $k = 0-8$, $l = 0-5$, $m = 1-3$.

k	l	m	r_a / r_1	t_f (hr)	k	l	m	r_a / r_1	t_f (hr)
1	1	1	6.084	17.951	5	3	2	5.739	41.885
2	1	1	5.357	17.951	6	3	2	5.495	41.885
3	1	1	4.585	17.951	7	3	2	5.246	41.885
4	1	1	3.755	17.951	8	3	2	4.992	41.885
5	1	1	2.846	17.951	0	4	2	5.522	41.885
6	1	1	1.811	17.951	1	4	2	5.336	41.885
0	2	1	3.898	17.951	2	4	2	5.148	41.885
1	2	1	3.463	17.951	3	4	2	4.956	41.885
2	2	1	3.004	17.951	4	4	2	4.761	41.885
3	2	1	2.518	17.951	5	4	2	4.563	41.885
4	2	1	1.996	17.951	6	4	2	4.361	41.885
5	2	1	1.423	17.951	7	4	2	4.156	41.885
0	3	1	2.738	17.951	8	4	2	3.946	41.885
1	3	1	2.406	17.951	0	5	2	4.621	41.885
2	3	1	2.056	17.951	1	5	2	4.460	41.885
3	3	1	1.685	17.951	2	5	2	4.298	41.885
4	3	1	1.286	17.951	3	5	2	4.133	41.885
0	4	1	2.086	17.951	4	5	2	3.965	41.885
1	4	1	1.811	17.951	5	5	2	3.794	41.885
2	4	1	1.523	17.951	6	5	2	3.620	41.885
3	4	1	1.216	17.951	7	5	2	3.443	41.885
0	5	1	1.659	17.951	8	5	2	3.262	41.885
1	5	1	1.423	17.951	5	5	3	6.241	65.820
2	5	1	1.174	17.951	6	5	3	6.100	65.820
3	3	2	6.215	41.885	7	5	3	5.958	65.820
4	3	2	5.979	41.885	8	5	3	5.814	65.820

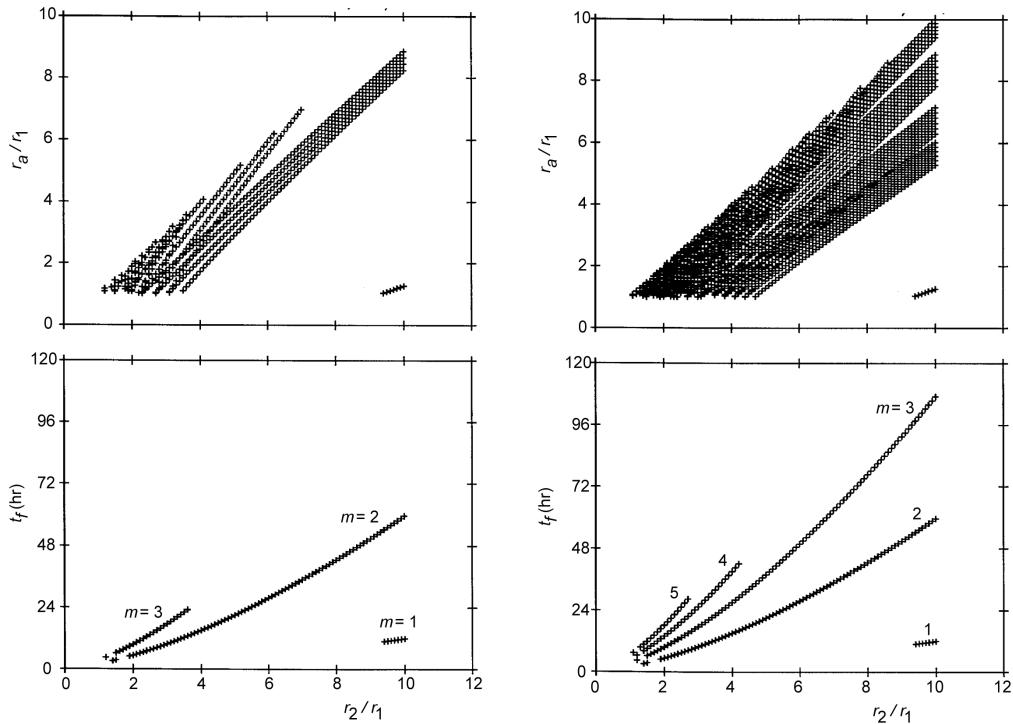


Figure 14.9: Values of r_a / r_1 and t_f for transfers between two inclined circular orbits, using a low phasing orbit and $r_1 = 6671$ km, $\varphi_0 = \psi_0 = 270^\circ$, as a function of r_2 / r_1 . The search process has covered the range $k = 0-3$, $l = 0-3$, $m = 1-3$ (left) or the range $k = 0-5$, $l = 0-5$, $m = 1-5$ (right).

of r_2/r_1 and for $r_1 = 6671$ km, $\varphi_0 = \psi_0 = 270^\circ$. In the plots on the left, results are shown for all orbits where the parameters k and l vary between 0 and 3, and m between 1 and 3; in the plots on the right, results are shown for the case that these parameters vary between 0 and 5, and 1 and 5, respectively. The Figure shows that for a given value of r_2/r_1 , many combinations of k , l and m are possible. In Figure 14.10, possible values of r_a/r_1 and the flight time are shown as a function of ψ_0 , for given values of r_2/r_1 and φ_0 . Here, all values of k and l between 0 and 3, and of m between 1 and 3, have been examined. Note that for the larger values of r_2/r_1 , the number of possible values of ψ_0 and the number of possible combinations of k , l , m are larger than for the smaller values of r_2/r_1 .

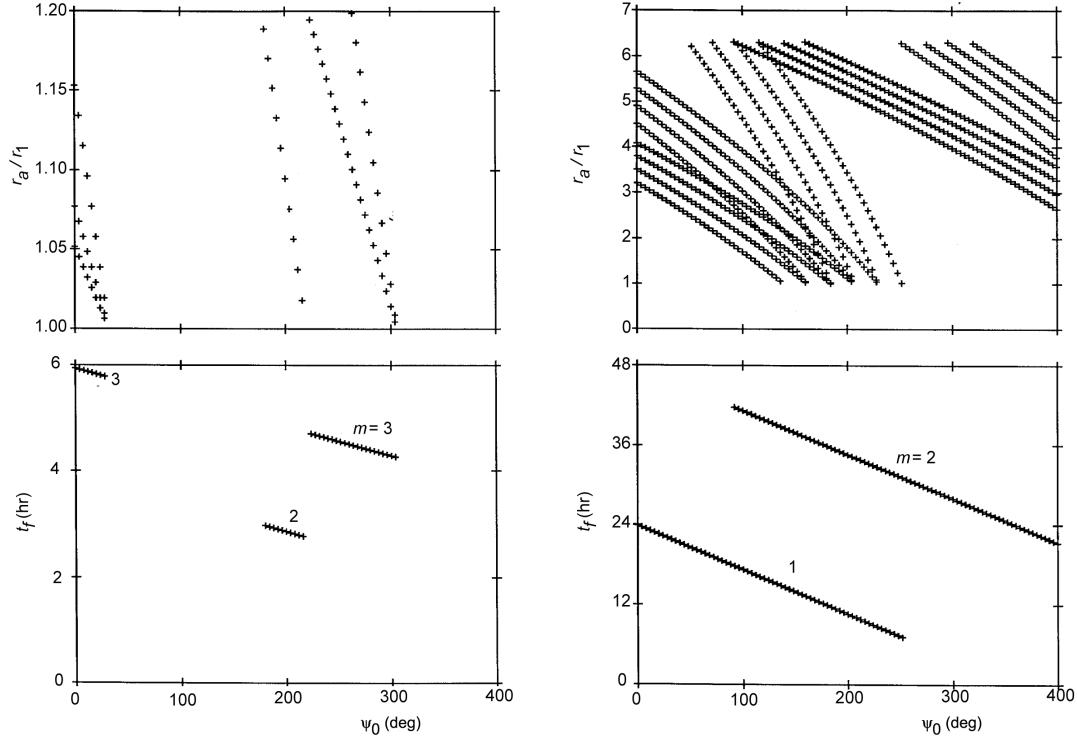


Figure 14.10: Values of r_a/r_1 and t_f for transfers between two inclined circular orbits, using a low phasing orbit, as a function of ψ_0 . The results hold for $r_2/r_1 = 1.2$ (left) and $r_2/r_1 = 6.3205$ (geostationary orbit, right) and $r_1 = 6671$ km, $\varphi_0 = 270^\circ$. The search process has covered the range $k = 0 - 3$, $l = 0 - 3$, $m = 1 - 3$.

14.4. Geometric aspects of high phasing orbits

In this Section, the case is considered where satellite 1, after it has completed k revolutions in the low orbit, is injected into a Hohmann transfer orbit to the high final orbit (Figure 14.11) at the moment it passes the nodal line of the two orbital planes (point s1,1). When the satellite reaches the high orbit, it is accelerated to a velocity that is somewhat lower than the local circular velocity. As a result, the satellite enters a phasing orbit with $r_a = r_2$, $r_1 < r_p < r_2$. The perigee distance of this orbit is selected such that satellite 1 reaches the desired point in the high orbit after one revolution in the phasing orbit. At that moment, satellite 1 is accelerated to the required local circular velocity. Because the period of the phasing orbit is shorter than the period of the high final orbit, it will be clear that at the moment that satellite 1 arrives for the first time in the high orbit, it will be located ‘behind’ satellite 2. After one revolution in the phasing orbit, satellite 2 has covered slightly less than one revolution in its orbit and both satellites approach each other. When the situation depicted in Figure 14.4 still holds, where the numbers indicate the position of arrival in the high orbit when satellite 1 is injected into the Hohmann transfer orbit during the

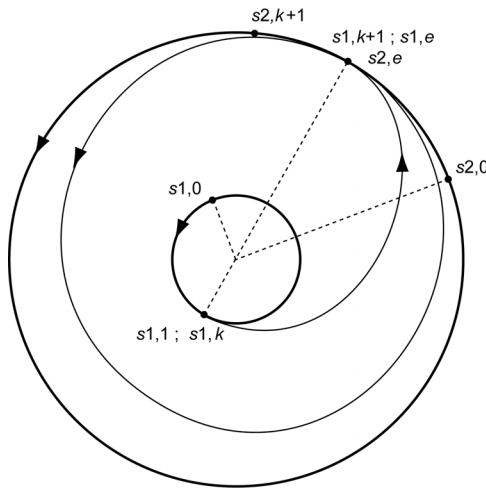


Figure 14.11: Geometry of a transfer flight between two inclined circular orbits, using a high phasing orbit, and the notation used.

k^{th} time that it passes point $s1,1$ ($k = 1, 2, \dots$), we conclude that now the strategy is to inject satellite 1 into a Hohmann transfer orbit when it passes for the fourth time the injection point. Then, the perigee distance of the phasing orbit has to be selected such that after one revolution in the phasing orbit satellite 1 approaches satellite 2. Obviously, it is also possible to inject satellite 1 during the fifth or a later passage of point $s1,1$ into the transfer orbit and to complete multiple revolutions in the phasing orbit.

When the case is considered that an arbitrary number of revolutions in the low orbit is completed, then it follows that any point in the high orbit can be reached, if (Figure 14.5)

$$0 < T_2 - T_{ph} < T_1 \quad (14.31)$$

where T_1 , T_2 and T_{ph} are the periods of the low orbit, the high orbit and the phasing orbit, respectively. From this inequality, we find

$$T_2 - T_1 < T_{ph} < T_2$$

or, with (14.3) and (14.11),

$$2\pi \left[\sqrt{\frac{r_2^3}{\mu}} - \sqrt{\frac{r_1^3}{\mu}} \right] < 2\pi \sqrt{\frac{(r_2 + r_p)^3}{8\mu}} < 2\pi \sqrt{\frac{r_2^3}{\mu}}$$

or

$$2 \left[1 - \sqrt{\left(\frac{r_1}{r_2} \right)^3} \right]^{2/3} - 1 < \frac{r_p}{r_2} < 1 \quad (14.32)$$

Since $r_p > r_1$, the limiting value of r_2/r_1 for which any point in the high orbit can be reached follows from

$$2 \left[1 - \sqrt{\left(\frac{r_1}{r_2} \right)^3} \right]^{2/3} = \frac{r_1}{r_2} + 1$$

A numerical solution of this equation yields $r_2/r_1 = 2.007$. So, for

$$\frac{r_2}{r_1} > 2.007 \quad (14.33)$$

any point in the basic interval between the points 3 and 4 in Figure 14.5 is accessible by completing one revolution in the high phasing orbit, with $r_1 < r_p < r_2$. By completing multiple revolutions in the low orbit, every basic interval is accessible. When the condition (14.33) is not satisfied, then parts of the high orbit will not be accessible, unless multiple revolutions in the phasing orbit are completed.

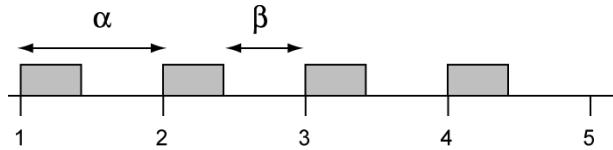


Figure 14.12: Illustration of the regions in the orbit of satellite 2 that are, for a specified value of $r_2/r_1 < 2.007$, not accessible (shaded area) for satellite 1, if satellite 1 completes only one revolution in the high phasing orbit.

When, for a given value of $r_2/r_1 < 2.007$, the size of the accessible part per basic interval α is indicated by β (Figure 14.12), then it follows that

$$\frac{\beta}{\alpha} = \frac{T_2 - T_{ph}^*}{T_1}$$

where T_{ph}^* is the period of a phasing orbit with $r_p = r_1$. We may write this relation as

$$\frac{\beta}{\alpha} = \sqrt{\left(\frac{r_2}{r_1} \right)^3} \left[1 - \sqrt{\left\{ \frac{1}{2} \left(1 + \frac{r_1}{r_2} \right) \right\}^3} \right] \quad (14.34)$$

For $1 < r_2/r_1 < 2.007$: $0 < \beta/\alpha < 1$. Combination of (14.9) and (14.34) yields for β , expressed in angular units,

$$\beta = 2\pi \left[1 - \sqrt{\left\{ \frac{1}{2} \left(1 + \frac{r_1}{r_2} \right) \right\}^3} \right] \quad (14.35)$$

from which we find that for $1 < r_2/r_1 < 2.007$: $0 < \beta < 2.209$, where the value of $\beta = 2.209$ is equal to the value of α for $r_2/r_1 = 2.007$. In degrees, the limiting value is $\beta = 127^\circ$. The only way to reach any point in the high orbit for $r_2/r_1 < 2.007$ is by completing multiple revolutions in an appropriately selected phasing orbit.

14.5. Selection of a high phasing orbit

In this Section, we consider the general case where satellite 1, after passing for the first time point $s_{1,1}$, completes k revolutions in the low orbit before it is injected into a Hohmann transfer orbit. Upon arrival in the high orbit, the satellite is injected into a phasing orbit with $r_1 < r_p < r_2$, $r_a = r_2$, which is subsequently flown for l revolutions, after which injection into the high orbit takes place. The geometry of this case and the notation used are indicated in Figure 14.13. At t_0 , the position of both satellites is indicated by φ_0 and ψ_0 . At t_1 , satellite 1 passes for the first time the

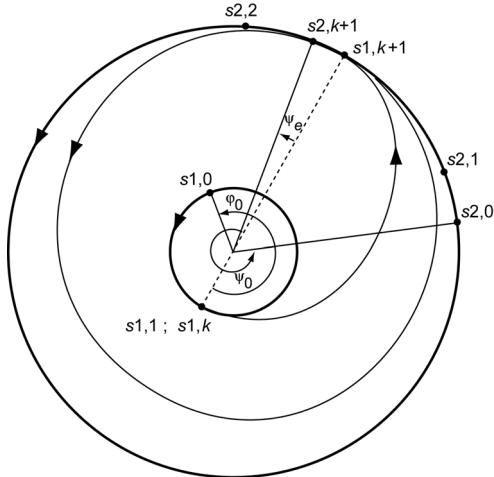


Figure 14.13: Geometry of a transfer flight between two inclined circular orbits, using a high phasing orbit, and the notation used.

injection point. When satellite 1 reaches the high orbit for the first time, satellite 2 has to be located ‘ahead’ of satellite 1. Now, satellite 1 is injected into the phasing orbit. After having completed one revolution in this orbit, satellite 2 has completed less than one revolution in its high orbit. After each revolution in the phasing orbit the geocentric angle between both satellites becomes smaller, while satellite 2 is at all times ahead of satellite 1. After satellite 1 has completed l revolutions in the phasing orbit, satellite 2 and satellite 1 approach each other and satellite 1 is injected into the final orbit.

For the total flight time of satellite 1 until injection into the high circular orbit, we can write

$$t_f = \frac{2\pi - \varphi_0}{n_1} + k T_1 + t_t + l T_{ph} \quad (14.36)$$

In this time interval, satellite 2 covers in its high orbit a geocentric angle of

$$\Psi_e + 2m\pi = \Psi_0 + n_2 t_f \quad (14.37)$$

where m is the number of revolutions that satellite 2 has completed in its high orbit. For the values of the integers k, l, m we have again

$$k = 0, 1, 2, \dots ; \quad l = 0, 1, 2, \dots ; \quad m = 1, 2, 3, \dots$$

Both satellites encounter each other when

$$\Psi_e = 0 \quad (14.38)$$

Substitution of (14.36) and (14.38) into (14.37) yields

$$2m\pi = \Psi_0 + n_2 \left[\frac{2\pi - \Phi_0}{n_1} + kT_1 + t_t + lT_{ph} \right]$$

or

$$lT_{ph} = \frac{2m\pi}{n_2} - \frac{\Psi_0}{n_2} - \frac{2\pi - \Phi_0}{n_1} - kT_1 - t_t \quad (14.39)$$

These equations are, of course, identical to the ones derived in Section 14.3. Substitution of (14.3) and (14.11) into (14.39) results, after some algebraic manipulation, in

$$l \sqrt{\left\{ \frac{1}{2} \left(\frac{r_2}{r_1} + \frac{r_p}{r_1} \right) \right\}^3} = \frac{1}{2\pi} \left[\Phi_0 - \sqrt{\left(\frac{r_2}{r_1} \right)^3} \Psi_0 \right] - \sqrt{\frac{1}{32} \left(\frac{r_2}{r_1} + 1 \right)^3} + m \sqrt{\left(\frac{r_2}{r_1} \right)^3} - k - 1 \quad (14.40)$$

This relation, which is similar to (14.23), shows that, for given values of Φ_0 , Ψ_0 and r_2/r_1 , in general, a number of combinations of k , l , m yield a value $r_1 < r_p < r_2$.

A direct transfer without a phasing orbit occurs when $k = l = 0$, $m = 1$. For this case, (14.40) simplifies to

$$\frac{1}{2\pi} \left[\Psi_0 + \sqrt{\left(\frac{r_1}{r_2} \right)^3} (2\pi - \Phi_0) \right] = 1 - \sqrt{\frac{1}{32} \left(1 + \frac{r_1}{r_2} \right)^3}$$

This equation is, of course, identical to (14.24) derived in Section 14.3.

In Section 14.4, it was found that when $r_2/r_1 > 2.007$ any point in the high orbit can be reached by selecting an appropriate phasing orbit and by completing that orbit only once ($l = 1$). In general, then $k > 0$. In that case, the boundary conditions for the perigee distance of a phasing orbit are given by (14.32). That relation may also be written as

$$\sqrt{\left(\frac{r_2}{r_1} \right)^3} - 1 < \sqrt{\left\{ \frac{1}{2} \left(\frac{r_2}{r_1} + \frac{r_p}{r_1} \right) \right\}^3} < \sqrt{\left(\frac{r_2}{r_1} \right)^3}$$

Substitution of (14.40) into this relation yields for $l = 1$

$$\sqrt{\left(\frac{r_2}{r_1} \right)^3} - 1 < \frac{1}{2\pi} \left[\Phi_0 - \sqrt{\left(\frac{r_2}{r_1} \right)^3} \Psi_0 \right] - \sqrt{\frac{1}{32} \left(\frac{r_1}{r_2} + 1 \right)^3} + m \sqrt{\left(\frac{r_2}{r_1} \right)^3} - k - 1 < \sqrt{\left(\frac{r_2}{r_1} \right)^3} \quad (14.41)$$

When a variable q is introduced, which is defined as

$$q = \frac{1}{2\pi} \left[\Phi_0 - \sqrt{\left(\frac{r_2}{r_1} \right)^3} \Psi_0 \right] - \sqrt{\frac{1}{32} \left(\frac{r_1}{r_2} + 1 \right)^3} - 1 \quad (14.42)$$

then we find from (14.41)

$$q + (m - 1) \sqrt{\left(\frac{r_2}{r_1}\right)^3} < k < q + (m - 1) \sqrt{\left(\frac{r_2}{r_1}\right)^3 + 1} \quad (14.43)$$

It is emphasized that the definition of q according to (14.42) strongly resembles the definition of q according to (14.26) for the low phasing orbit. Again, q is completely determined by φ_0, ψ_0 and r_2/r_1 . Equation (14.43) shows that, if $l = 1$ and $r_2/r_1 > 2.007$, for any value of m there cannot exist more than one value of k that satisfies the inequality. For the values of k and m we obviously have $k \geq 0, m \geq 1$. The computation scheme of possible phasing orbits therefore is as follows. For given values of φ_0, ψ_0 and r_2/r_1 , the value of q is computed from (14.42). Subsequently, for a series of selected values of m the corresponding values of k are determined with (14.43). For any set of values of k and m , and $l = 1$, (14.40) yields the required value of r_p/r_2 .

In Section 14.3, it was also found that when $r_2/r_1 < 2.007$, in general, multiple revolutions in the phasing orbit are required in order to reach a given point in the high orbit. Obviously, also for $r_2/r_1 > 2.007$ multiple revolutions in the phasing orbit may be completed. Therefore, the following methodology is generally applicable and the method described above may be considered as a special case of that methodology for $l = 1, r_2/r_1 > 2.007$. Because in the general case $r_1 < r_p < r_2$, combination of (14.40) and (14.42) yields

$$q + (m - l) \sqrt{\left(\frac{r_2}{r_1}\right)^3} < k < q + m \sqrt{\left(\frac{r_2}{r_1}\right)^3 - l \sqrt{\left\{\frac{1}{2} \left(\frac{r_2}{r_1} + 1\right)\right\}^3}} \quad (14.44)$$

Then, the computation scheme for determining the possible phasing orbits is as follows. For given values of φ_0, ψ_0 and r_2/r_1 , the value of q is determined from (14.42). Now, a series of positive integer values of l and m is selected, starting with $l = 1, m = 1$. Then, all possible positive integer values of k are determined from (14.44). For every k, l, m -combination, (14.40) yields the corresponding value of r_p/r_1 .

Equation (14.44) shows that, in general, multiple values of k are possible for given values of φ_0, ψ_0 and r_2/r_1 , and selected values of l and m . However, there exists only one value for k if

$$\left(q + m \sqrt{\left(\frac{r_2}{r_1}\right)^3} - l \sqrt{\left\{\frac{1}{2} \left(\frac{r_2}{r_1} + 1\right)\right\}^3} \right) - \left(q + (m - l) \sqrt{\left(\frac{r_2}{r_1}\right)^3} \right) < 1$$

or

$$l \left[\sqrt{\left(\frac{r_2}{r_1}\right)^3} - \sqrt{\left\{\frac{1}{2} \left(\frac{r_2}{r_1} + 1\right)\right\}^3} \right] < 1 \quad (14.45)$$

From (14.45) we find that for $l = 2, 3$ and 5 , the corresponding values of r_2/r_1 for which a transfer is possible for only one value of k are $1.560, 1.391$ and 1.245 , respectively.

Just as for the case of a low phasing orbit, the total flight time of satellite 1 is solely determined by the value of m , and not by the values of k and l . From (14.37) and (14.38) we find

$$t_f = \frac{2m\pi - \psi_0}{n_2} = T_2 \left(m - \frac{\psi_0}{2\pi} \right) \quad (14.46)$$

Table 14.2: Characteristics of feasible transfer flights between two inclined circular orbits with $r_1 = 6671$ km and $r_2/r_1 = 6.3205$ (geostationary orbit), and using a high phasing orbit, for $\varphi_0 = \psi_0 = 90^\circ$. The search process has covered the range $k = 0-8$, $l = 0-5$, $m = 1-3$.

k	l	m	r_p / r_1	t_f (hr)	k	l	m	r_p / r_1	t_f (hr)
0	1	1	1.455	17.951	4	3	3	4.055	65.820
8	1	2	6.143	41.885	5	3	3	3.859	65.820
0	2	2	4.033	41.885	6	3	3	3.661	65.820
1	2	2	3.738	41.885	7	3	3	3.461	65.820
2	2	2	3.438	41.885	8	3	3	3.259	65.820
3	2	2	3.134	41.885	0	4	3	2.877	65.820
4	2	2	2.825	41.885	1	4	3	2.721	65.820
5	2	2	2.510	41.885	2	4	3	2.563	65.820
6	2	2	2.190	41.885	3	4	3	2.404	65.820
7	2	2	1.864	41.885	4	4	3	2.244	65.820
8	2	2	1.531	41.885	5	4	3	2.082	65.820
0	3	2	1.581	41.885	6	4	3	1.919	65.820
1	3	2	1.355	41.885	7	4	3	1.754	65.820
2	3	2	1.127	41.885	8	4	3	1.587	65.820
8	2	3	6.232	65.820	0	5	3	1.605	65.820
0	3	3	4.821	65.820	1	5	3	1.471	65.820
1	3	3	4.632	65.820	2	5	3	1.335	65.820
2	3	3	4.441	65.820	3	5	3	1.198	65.820
3	3	3	4.249	65.820	4	5	3	1.060	65.820

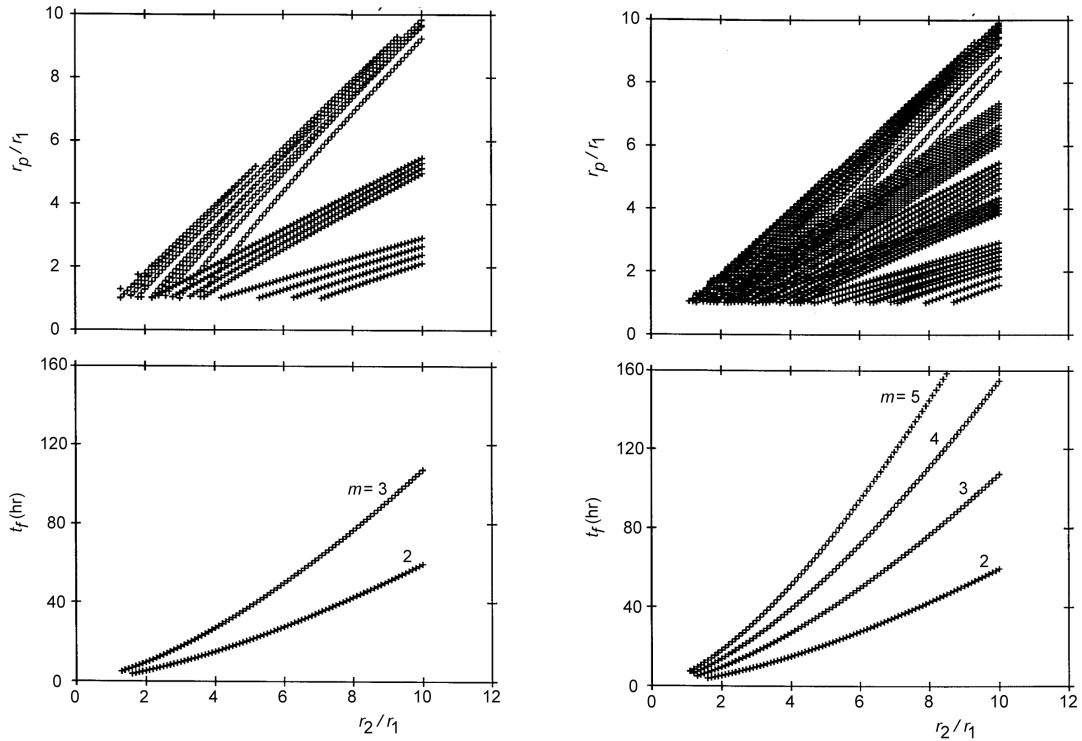


Figure 14.14: Values of r_p/r_1 and t_f for transfers between two inclined circular orbits, using a high phasing orbit and $r_1 = 6671$ km, $\varphi_0 = \psi_0 = 270^\circ$, as a function of r_2/r_1 . The search process has covered the range $k = 0-3$, $l = 0-3$, $m = 1-3$ (left) or the range $k = 0-5$, $l = 0-5$, $m = 1-5$ (right).

This relation is identical to (14.30) that holds for low phasing orbits and proves the statement that the flight time is independent of the values of k and l . Again, preferably a phasing orbit is selected for which the value of m is as small as possible. For this value of m , generally, multiple

combinations of k and l are possible and the selection of k and l is done on basis of mission considerations. In Table 14.2, possible k,l,m -combinations and the corresponding values of r_p/r_1 and t_f are listed for a transfer to geostationary orbit ($r_2/r_1 = 6.3205$), with $r_1 = 6671$ km, $\varphi_0 = \psi_0 = 90^\circ$. The search process has covered the range $k = 0\text{-}8$, $l = 0\text{-}5$, $m = 1\text{-}3$. In Figure 14.14 possible values of r_p/r_1 and the corresponding flight time are shown as a function of r_2/r_1 and $r_1 = 6671$ km, $\varphi_0 = \psi_0 = 270^\circ$. In this Figure, results are shown for all orbits for which the parameters k and l vary between 0 and 3, and the parameter m between 1 and 3 (left), or for which the parameters k and l vary between 0 and 5, and the parameter m between 1 and 5 (right). The Figure shows that for a given value of r_2/r_1 , many combinations of k , l and m are possible. In Figure 14.15 possible values of r_p/r_1 and the flight time are shown as a function of ψ_0 , for given values of r_2/r_1 and φ_0 . Here, all values of k and l between 0 and 3, and of m between 1 and 3, have been examined. This Figure shows that for the larger values of r_2/r_1 , the number of possible values of ψ_0 and the number of possible combinations of k , l , m are larger than for the smaller values of r_2/r_1 .

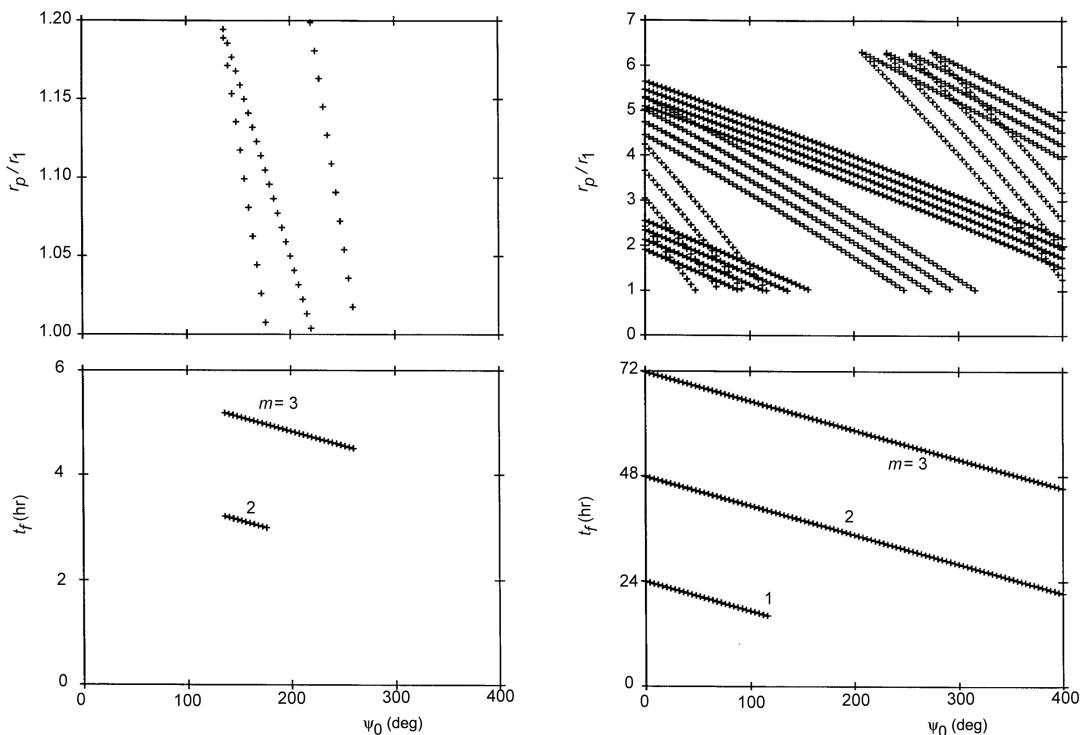


Figure 14.15: Values of r_p/r_1 and t_f for transfers between two inclined circular orbits, using a high phasing orbit, as a function of ψ_0 . The results hold for $r_2/r_1 = 1.2$ (left) and $r_2/r_1 = 6.320$ (geostationary orbit, right) and $r_1 = 6671$ km, $\varphi_0 = 270^\circ$. The search process has covered the range $k = 0\text{-}3$, $l = 0\text{-}3$, $m = 1\text{-}3$.

15. RENDEZ-VOUS FLIGHTS

A *rendez-vous flight* is a mission where a spacecraft approaches another spacecraft. In most cases, such a mission is completed by a *docking procedure* that results in a firm connection between the two spacecraft. In this Chapter, we will limit ourselves to rendez-vous missions involving satellites that move in orbits about the Earth. The satellite that is performing the approach flight is called the *chaser*; the satellite that takes a passive role and is approached by the chaser is called the *target*. The rendez-vous mission generally starts with a (Hohmann) transfer (Chapters 12 and 13) that brings the chaser relatively close to the target and in an orbit somewhat lower than the orbit of the target. Then the real rendez-vous phase starts. When the chaser is not too far from the target, the motion of the chaser relative to the target may be described by the Clohessy-Wiltshire equations that have been derived in Chapter 9. These equations hold for the case that the orbit of the target is circular, which is usually a good approximation, and when the motion of the chaser is described relative to a target-fixed rotating reference frame (Figure 15.1). The *X*-axis of this reference frame is directed radially away from the Earth, the *Y*-axis is tangential to the orbit of the target and is directed in the direction of motion of the target, and the *Z*-axis completes the right-handed reference frame. So, in this reference frame the *x* and *y* coordinates of the chaser are positive if the chaser is located ‘above’ and ‘ahead of’ the target. Before we start with an analysis of rendez-vous flights, Section 15.1 will first give an overview of the historical development of rendez-vous techniques and of some general aspects of rendez-vous.

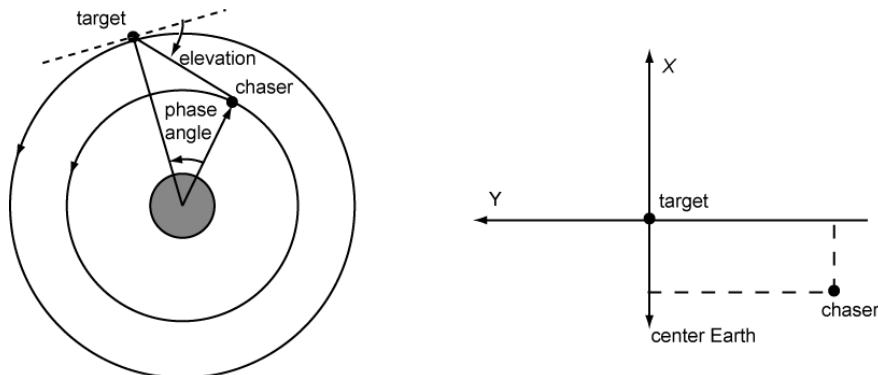


Figure 15.1: Two-dimensional geometry of the motion of the target and the chaser (left), and the reference frame used for describing the relative motion of the chaser (right).

15.1. General rendez-vous aspects

In the past, the US and the former USSR have gained much experience in rendez-vous and docking operations during the manned Gemini, Apollo and Skylab programs, and the manned or unmanned Cosmos, Soyuz and Progress missions.

The Gemini program served as a test bed for the Apollo manned lunar landing program, and various techniques that were required for the Apollo missions were tested in Earth orbit by Gemini spacecraft. On the Gemini IV mission (June 3-7, 1965) the first attempt was made to execute rendez-vous and *station keeping* maneuvers; in this case with the spent second stage of the Titan II launch vehicle. A phantom rendez-vous was performed during the Gemini V mission (August 21-28, 1965), which involved a maneuver sequence equal to part of the maneuver sequence designed for the Gemini VI-A mission. The first US real rendez-vous mission, without docking, involved the Gemini VI-A and Gemini VII spacecraft closing to each other to about 0.3

m in December 1965. The first rendez-vous mission with docking took place in March 1966, involving the Gemini VIII and the so-called Gemini Agena Target Vehicle (GATV). This GATV was capable to change the orbit of Gemini VIII. Gemini IX performed a rendez-vous with a vehicle launched to replace a GATV that was lost. Further and more-complex rendez-vous and docking operations were performed during the Gemini X, XI and XII missions.

As an example, the rendez-vous maneuvers executed during the Gemini VI-A, VIII and X missions will be described briefly. The chaser was first injected at perigee of its initial orbit into an elliptical orbit that was about coplanar with the target's orbit. The first orbital revolution was left free of maneuvers to check the chaser's functions. After one orbital revolution, an apogee height adjust maneuver was performed at perigee to correct for in-plane insertions errors. At the second apogee passage, a phase adjust maneuver raised the next perigee to provide the catch-up rate required for the proper phasing of the terminal phase initiation. That terminal phase occurred in a darkness period ('night') in order to have a good sight of the target; when the maneuver would have been performed in sunlight the Sun might have blinded the crew. An out-of-plane correction maneuver was applied at the nodal crossing after the second apogee to correct for the out-of-plane insertion errors. At the third apogee, a co-elliptical maneuver was performed to produce an almost constant altitude difference of 15 n.mi. between both spacecraft. The terminal phase initiation occurred when the chaser was about 40 n.mi. 'behind' the target. That phase was initiated with a velocity change along the line of sight (LOS) towards the target. The necessary velocity change was calculated by the onboard guidance system. Gravitational effects move the chaser away from this initial LOS direction (Section 15.4). Therefore, small corrections were performed by rocket thrust pulses to put the chaser on an interception course. Braking and LOS rate control were executed by manual operation during the final approach based upon radar and visual data.

In the Apollo program, rendez-vous and docking operations were carried out in low Earth orbit, translunar orbit and lunar orbit. In the follow-on Skylab program, the Apollo Command and Service Module (CSM) served as the chaser vehicle and Skylab as the target. One important feature of all US rendez-vous and docking missions was that none of these missions took place fully autonomously; in all missions a human operator in space interfered with the rendez-vous and docking procedure.

In October 1967, the Cosmos 186 and Cosmos 188 vehicles were the first Soviet spacecraft that performed rendez-vous and docking. A striking characteristic of this mission was that both spacecraft were unmanned and that the final phases of this rendez-vous and docking flight took place fully autonomously outside the field of view of the Soviet ground control center. In subsequent years, the USSR carried out many rendez-vous and docking operations with and without a human operator in space interfering. In the Apollo Soyuz Test Project (ASTP) the US and USSR have collaborated; a manned Apollo and a manned Soyuz spacecraft flew for some time closely together and docked to each other on July 17, 1975.

With the space station Mir the Soviets have significantly enlarged their experience in rendezvous and docking operations; over the years many USSR spacecraft, but also the US Space Shuttle, have docked with this station. And since December 1998, US (Space Shuttle) and USSR/Russian spacecraft rendez-vous and dock routinely with the International Space Station (ISS), bringing crew and materials to this space station and returning crew and materials back to Earth. In March 2008 the first rendez-vous and docking mission to the ISS was performed by an unmanned European ATV (Automated Transfer Vehicle) spacecraft (Section 15.6); in September 2009 by a Japanese HTV (H-II Transfer Vehicle) spacecraft; and in October 2012 by a commercial Dragon spacecraft developed and launched by Space Exploration Technologies (SpaceX).

In the future, many more rendez-vous and docking missions will be flown. Examples are missions in geostationary orbit for building up a communications platform, in geostationary transfer orbit between a propulsion module carrying an apogee boost motor and a communication payload module, and in low Earth orbit between a retrievable carrier and a dedicated test vehicle.

In general, a number of phases may be distinguished during a rendez-vous mission; these will be briefly described below.

Launch phase

For this phase basically two options exist. The first option is that when the launch site crosses the orbital plane of the target the chaser is launched directly into an elliptical or circular orbit that is (about) coplanar with and somewhat lower than the target's orbit. This enables the chaser to gradually catch up with the target and to perform a rendez-vous. The disadvantage of this option is that, depending on the position of the target at the time of launch of the chaser, the rendez-vous between both satellites can take quite a time as the difference in the orbital angular velocities of both spacecraft is quite small. An alternative option is that, when the launch site crosses the orbital plane of the target, the chaser is launched into a low-altitude parking orbit that is (about) coplanar with the target's orbit. In this case, the difference in the angular velocities of both spacecraft is larger, leading to a shorter time interval before rendez-vous can be accomplished. At the right time, the chaser is injected into a transfer orbit and subsequently into an elliptical or circular orbit that is (about) coplanar with and somewhat lower than the target's orbit from which it accomplishes the rendez-vous.

Mid-course phase

In this phase, the chaser is maneuvered sufficiently close to the target to enable the chaser's onboard guidance sensors to acquire the target, i.e. several hundreds of kilometers 'behind' and some tens of kilometers 'below' the target. Several maneuvers may be distinguished during this phase. First, we have the *height adjust maneuver*, which has to correct for in-plane insertion errors. This maneuver adjust the apogee altitude of the elliptical trajectory (or the radius of the circular orbit) to a desired value and can best be applied at perigee (in case of an elliptical orbit) of the chaser's orbit in order to minimize the fuel consumption (Section 12.5). A second type of maneuver is the *phase adjust maneuver*, which has to adjust the catch-up rate, i.e. the relative velocity of the chaser towards the target, to a value required for the next phase, i.e. the *terminal phase*. The *phase angle*, ψ , (Figure 15.1) is defined as the angle between the radius vector of the chaser and that of the target. This angle must have a certain value to continue with the terminal phase. These maneuvers are applied at apogee of the chaser's orbit to minimize the fuel consumption and raise the perigee altitude. Thus, a height adjust maneuver adjusts the apogee altitude and a phase adjust maneuver adjusts the perigee altitude of the chaser's orbit. A third type of maneuver is the *out-of-plane correction maneuver*. As stated above, the chaser is launched into an orbit essentially coplanar with the target. However, due to injection errors the chaser's orbital plane may differ, both in inclination and in right ascension of the ascending node, slightly from the target's orbital plane. This means that the chaser will have a relative velocity component perpendicular to the orbital plane of the target: the *out-of-plane velocity*. The purpose of out-of-plane corrections is to nullify the relative out-of-plane errors. This maneuver is performed at a properly selected point along the orbit, dependent on the actual errors in inclination and right ascension of the ascending node (Chapter 13). The last type of maneuver is the *co-elliptical maneuver*, which positions the chaser in an orbit with a constant altitude difference relative to the orbit of the target. After this maneuver, the catch-up rate has a constant value and the chaser drifts with a constant relative velocity towards the target. This characteristic

may be used for the calibration of guidance sensors.

Terminal phase

During this phase, the chaser is brought into an interception course with the target. The terminal phase starts at the *terminal phase initiation point*, after a drift period that follows the co-elliptical maneuver. Visibility aspects, for the crew or for the sensors, determine the position of this point, i.e. the target must be visible to the crew and/or to the sensors of the chaser. During this phase, the relative velocity and range of the chaser to the target must be nullified simultaneously in a controlled way to avoid collision. The main type of guidance technique is *forced rectilinear motion*, or rendez-vous along the LOS (Section 15.4). In principle, also a two-impulses elliptical rendez-vous trajectory may be applied. However, guidance techniques for such impulsive trajectories are difficult to implement for an autonomous terminal phase. The reason is that for impulsive trajectories the relative velocity has to be nullified in a short time, whereas with continuous steering this velocity can be nullified gradually. In addition, because of small injection errors at the first impulsive shot of the two-impulses trajectory the final approach may occur from another direction than the desired one, which in case of large space structures may easily lead to a collision. Another reason for active continuous LOS steering is to obtain a constant chaser-target configuration, which is attractive to enable the sensors to keep the target in its field of view.

During the ‘straight line’ approach, a LOS control technique is applied through which the time-derivative of the angle between the LOS direction and the local horizontal plane is nullified or kept between certain limits by operating the chaser’s thrusting system perpendicular to the LOS direction. Braking is applied until the relative velocity has become lower than a specified value. Often, docking mechanisms only operate satisfactorily if the two rendez-vous spacecraft still have a small but finite relative velocity difference. This difference then activates the latching system.

Station keeping and hold points

During the rendez-vous and docking flight usually several *hold points* are introduced. A hold point is a time-flexible element of the chaser’s flight to gain time without the chaser markedly changing its position/range relative to the target. When the target is in a circular orbit, a hold point should preferably be located in the orbit of the target, since then the angular velocities of both spacecraft about the Earth are equal. A hold point can be applied nominally or in contingency cases. Nominally, hold points may serve a number of purposes: verification on the ground of the performed chaser’s operations, checking the conditions for the next operational phase, calibration of the onboard guidance sensors, delaying the next operational step to improve environmental conditions (e.g. illumination), or modifying/updating input parameters for operational software. In contingency cases, the following purposes can be distinguished: saving the rendez-vous and docking mission, avoiding collision, gaining time for trouble shooting, etc.

Adopted flight profile

In order to provide more insight in rendez-vous flights, in the next Sections analytical relations will be derived for three important phases of a rendez-vous flight:

- *Phase 1*: Flight from an initial distance of more than 1000 km from the target to a distance of 1 - 10 km from the target. It is assumed that two impulsive shots are applied and that at the end of this phase the chaser is located at the last hold point on the Y-axis, 5 km ‘behind’ the target.
- *Phase 2*: Flight from this hold point to approximately 200 m away from the target. It is assumed that two impulsive shots are applied.

– *Phase 3*: The chaser approaches the target along the LOS using thrust pulses.

To simplify the analysis, it is assumed that the orbit of the target is circular and that at the end of phases 1 and 2 both the relative position and the relative velocity of the chaser are equal to zero. Regarding the velocity, this certainly is the case at the end of phase 1, but in reality needs not necessarily be the case at the end of phase 2. Regarding the position, the relative position is in reality certainly not zero at the end of phases 1 and 2. However, because for both phase 1 and phase 2 the distance between the chaser and the target is at the end of that phase much smaller than at the start of that phase, this approximation may be applied. In addition, it is assumed that after phase 1 has finished both spacecraft are moving in the same orbital plane.

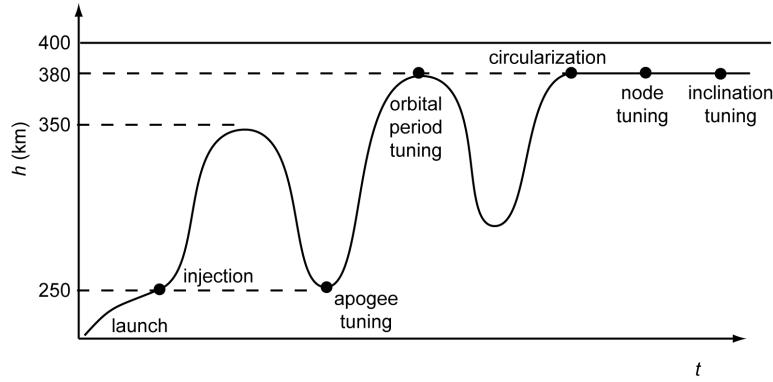


Figure 15.2: Typical rendez-vous mission profile.

We may imagine the following mission profile for this rendez-vous flight (Figure 15.2). The target is moving in a circular orbit at an altitude of 400 km and with a period of $T_t = 92.41$ min. The chaser is launched into an orbit with perigee altitude of 250 km and apogee altitude of 350 km. After one orbital revolution of the chaser, the rocket engine is ignited at perigee to raise the apogee altitude to 380 km. At apogee of this orbit, the engine is ignited again to change the perigee altitude such that when the chaser passes the apogee of its orbit again it approaches the target at an appropriate distance. It is assumed that at that moment the chaser is located ‘behind’ the target. Now, the orbit of the chaser is circularized and subsequent maneuvers are performed to nullify any difference in the right ascension of the ascending node and the inclination of the two orbital planes. The period of this orbit is $T_{ch} = 92.00$ min. Since the altitudes of both orbits differ by 20 km, the phase angle, ψ , decreases continuously at a rate of approximately $1^\circ/\text{hr}$. As soon as this angle is 20° , the first phase of the rendez-vous part of the flight starts. At that moment the relative position and velocity components of the chaser are $x_0 = -20$ km, $y_0 = -2358$ km, $z_0 = 0$, $\dot{x}_0 = \dot{z}_0 = 0$, $\dot{y}_0 = 11.4$ m/s. Note that these initial conditions are selected rather arbitrarily; they certainly do not correspond to a set of optimum initial conditions. However, the analysis of this non-optimal rendez-vous flight will show the major characteristics of rendez-vous trajectories.

15.2. Flight from a large distance to the last hold point

Suppose that the relative position and velocity of the chaser at t_0 are given by $x_0, y_0, z_0, \dot{x}_0, \dot{y}_0, \dot{z}_0$ relative to the target-fixed rotating reference frame, and that the chaser should reach the target at t_2 . It is assumed that at t_0 an impulsive shot is applied, as a result of which the chaser is injected into a collision trajectory with the target, and that at t_2 a second impulsive shot is given to nullify the relative velocity of the chaser with respect to the target. The flight time of this intercept trajectory is $t_f = t_2 - t_0$. In Section 9.3, the following relations were derived for the motion of the

chaser relative to the target:

$$\begin{aligned} x &= \left(4x_0 + \frac{2\dot{y}_0}{n} \right) + \frac{\dot{x}_0}{n} \sin nt - \left(3x_0 + \frac{2\dot{y}_0}{n} \right) \cos nt \\ y &= \left(y_0 - \frac{2\dot{x}_0}{n} \right) - 3 \left(2x_0 + \frac{\dot{y}_0}{n} \right) nt + 2 \left(3x_0 + \frac{2\dot{y}_0}{n} \right) \sin nt + \frac{2\dot{x}_0}{n} \cos nt \\ z &= z_0 \cos nt + \frac{\dot{z}_0}{n} \sin nt \end{aligned} \quad (15.1)$$

where n is the mean angular motion of the target in its orbit about the Earth. These equations describe the motion of the chaser in between the two impulsive shots of phase 1, where the time is counted from $t_0 = 0$; so, for the flight time we may write $t_f = t_2$. Note that the value of $y_0 = -2358$ km used for the adopted flight profile is certainly outside the range given in Section 9.2 for which (15.1) yields accurate results. The numerical results presented in this Section therefore only provide a qualitative picture of the relative motion. The components of the velocity immediately after the first impulsive shot (index 1) are indicated by $\dot{x}_1, \dot{y}_1, \dot{z}_1$. So, the state vector at the start of phase 1 has the components $x_0, y_0, z_0, \dot{x}_1, \dot{y}_1, \dot{z}_1$. At t_2 the position components should be zero and we then find from (15.1)

$$\begin{aligned} 0 &= \left(4x_0 + \frac{2\dot{y}_1}{n} \right) + \frac{\dot{x}_1}{n} \sin nt_2 - \left(3x_0 + \frac{2\dot{y}_1}{n} \right) \cos nt_2 \\ 0 &= \left(y_0 - \frac{2\dot{x}_1}{n} \right) - 3 \left(2x_0 + \frac{\dot{y}_1}{n} \right) nt_2 + 2 \left(3x_0 + \frac{2\dot{y}_1}{n} \right) \sin nt_2 + \frac{2\dot{x}_1}{n} \cos nt_2 \\ 0 &= z_0 \cos nt_2 + \frac{\dot{z}_1}{n} \sin nt_2 \end{aligned}$$

These equations can be rewritten as

$$\begin{aligned} \dot{x}_1 \sin nt_2 + 2\dot{y}_1 (1 - \cos nt_2) &= -x_0 n (4 - 3 \cos nt_2) \\ 2\dot{x}_1 (1 - \cos nt_2) + \dot{y}_1 (3nt_2 - 4 \sin nt_2) &= y_0 n - 6x_0 n (nt_2 - \sin nt_2) \\ \dot{z}_1 \sin nt_2 &= -z_0 n \cos nt_2 \end{aligned}$$

Multiplying the first expression by $2(1 - \cos nt_2)$ and the second expression by $-\sin nt_2$, and subsequent summation of the resulting expressions gives

$$\dot{y}_1 = \frac{n [x_0 \{6nt_2 \sin nt_2 - 14(1 - \cos nt_2)\} - y_0 \sin nt_2]}{8(1 - \cos nt_2) - 3nt_2 \sin nt_2} \quad (15.2-1)$$

Multiplying the first expression by $(3nt_2 - 4 \sin nt_2)$ and the second expression by $-2(1 - \cos nt_2)$, and subsequent summation of the resulting expressions gives

$$\dot{x}_1 = \frac{n [2y_0 (1 - \cos nt_2) - x_0 (4 \sin nt_2 - 3nt_2 \cos nt_2)]}{8(1 - \cos nt_2) - 3nt_2 \sin nt_2} \quad (15.2-2)$$

The third expression immediately gives

$$\dot{z}_1 = \frac{-z_0 n}{\tan nt_2} \quad (15.2-3)$$

Note that when the denominator in (15.2-1) and (15.2-2) becomes zero, the values of \dot{x}_1 and \dot{y}_1 may become infinitely large. Equations (15.2) show that the values of \dot{x}_1 and \dot{y}_1 are functions of the relative coordinates x_0 and y_0 and the flight time, t_f . So, the singularities are due to unfavorable values of these parameters. An analysis of the expression in the denominator shows that within the range $nt_2 = 0$ to $nt_2 = 5\pi$ it has five roots: $nt_2 = 0, 2\pi, 2.813\pi, 4\pi, 4.891\pi$. When we use the relation $t_2 = t_f$ and the relation for the orbital period of the target: $T_t = 2\pi/n$, we find that these roots correspond to flight times of $t_f = 0, T_t, 1.407T_t, 2T_t, 2.445T_t$. For values of nt_2 around these roots we may expect large values of both impulsive shots required to execute the transfer. When we consider the motion of both spacecraft relative to a geocentric non-rotating reference frame, both trajectories are conic sections about the Earth, which intersect at t_2 . It will be clear that there are combinations of x_0 and y_0 which are favorable to execute a rendez-vous flight with a relatively short flight time for a small ΔV_0 , but that other combinations lead to an unfavorable geometry and will require a very high, or sometimes even an infinitely large, ΔV_0 to execute the mission in a relatively short flight time. A trivial example is the singularity at $t_2 = t_f = 0$. This requires for non-zero values of x_0 and y_0 infinitely large values of \dot{x}_1, \dot{y}_1 and ΔV_0 .

Now that the velocity components just after the first impulsive shot are known, we can compute the rectangular components of the required velocity change at t_0 :

$$\Delta\dot{x}_0 = \dot{x}_1 - \dot{x}_0 \quad ; \quad \Delta\dot{y}_0 = \dot{y}_1 - \dot{y}_0 \quad ; \quad \Delta\dot{z}_0 = \dot{z}_1 - \dot{z}_0 \quad (15.3)$$

and the magnitude of the velocity change due to the impulsive shot at t_0 is

$$\Delta V_0 = ((\Delta\dot{x}_0)^2 + (\Delta\dot{y}_0)^2 + (\Delta\dot{z}_0)^2)^{1/2} \quad (15.4)$$

The distance between the chaser and the target at t_0 is

$$r_0 = (x_0^2 + y_0^2 + z_0^2)^{1/2} \quad (15.5)$$

When \bar{r}_0 is the position vector of the chaser in the target-centered rotating reference frame at t_0 , then $\bar{R}_0 = -\bar{r}_0$ is the position vector from the chaser to the target at t_0 . For the angle α_0 between \bar{R}_0 and $\Delta\bar{V}_0$ we can write

$$\cos \alpha_0 = \frac{\bar{R}_0 \cdot \Delta\bar{V}_0}{R_0 \Delta V_0} \quad (15.6)$$

where

$$\bar{R}_0 = -x_0 \bar{e}_x - y_0 \bar{e}_y - z_0 \bar{e}_z \quad ; \quad \Delta\bar{V}_0 = \Delta\dot{x}_0 \bar{e}_x + \Delta\dot{y}_0 \bar{e}_y + \Delta\dot{z}_0 \bar{e}_z$$

Substitution of these relations into (15.6) gives

$$\cos \alpha_0 = -\frac{x_0 \Delta\dot{x}_0 + y_0 \Delta\dot{y}_0 + z_0 \Delta\dot{z}_0}{r_0 \Delta V_0} \quad (15.7)$$

with $0^\circ \leq \alpha_0 < 180^\circ$.

When we differentiate (15.1) with respect to time, substitute for the initial conditions the

quantities $x_0, y_0, z_0, \dot{x}_1, \dot{y}_1, \dot{z}_1$ and apply the equations for t_2 , then we arrive at the following relations:

$$\begin{aligned}\dot{x}_2 &= \dot{x}_1 \cos nt_2 + (3nx_0 + 2\dot{y}_1) \sin nt_2 \\ \dot{y}_2 &= -3(2nx_0 + \dot{y}_1) + 2(3nx_0 + 2\dot{y}_1) \cos nt_2 - 2\dot{x}_1 \sin nt_2 \\ \dot{z}_2 &= -nz_0 \sin nt_2 + \dot{z}_1 \cos nt_2\end{aligned}$$

Substitution of (15.2) into these equations yields

$$\begin{aligned}\dot{x}_2 &= -n \left[\frac{2y_0(1 - \cos nt_2) - x_0(3nt_2 - 4\sin nt_2)}{8(1 - \cos nt_2) - 3nt_2 \sin nt_2} \right] \\ \dot{y}_2 &= -n \left[\frac{y_0 \sin nt_2 - 2x_0(1 - \cos nt_2)}{8(1 - \cos nt_2) - 3nt_2 \sin nt_2} \right] \\ \dot{z}_2 &= -n \frac{z_0}{\sin nt_2}\end{aligned}\tag{15.8}$$

These equations express the values of the rectangular components of the velocity just before the second impulsive shot. As the relative velocity has to be nullified by this impulsive shot, its magnitude is given by

$$\Delta V_2 = (\dot{x}_2^2 + \dot{y}_2^2 + \dot{z}_2^2)^{1/2}\tag{15.9}$$

For the total required impulse for this rendez-vous flight, we thus obtain

$$\Delta V_{tot} = \Delta V_0 + \Delta V_2\tag{15.10}$$

The expressions given above show that for specified initial conditions $x_0, y_0, z_0, \dot{x}_0, \dot{y}_0, \dot{z}_0$, the total required rocket impulse is a function of the flight time in phase 1, i.e. $t_f = t_2$. In general, one will choose an optimal rendez-vous flight, which means that a value of t_f will be selected for which ΔV_{tot} is a minimum. An analytical treatment of these optimum rendez-vous flights is quite complicated, and is beyond the scope of this book.

The amount of rocket propellant needed for this phase of the rendez-vous mission can be computed as follows. For an impulsive shot, we may use *Tsiolkovski's law* (Section 1.7):

$$\Delta V = V_j \ln \frac{M_0}{M_e}\tag{15.11}$$

where V_j is the (effective) exhaust velocity, M_0 is the initial mass of the chaser and M_e is the mass of the chaser after an impulsive shot. If it is assumed that ΔV_0 and ΔV_2 are executed by the same rocket engine and that no mass is discarded in between both impulsive shots, then we obtain from (15.11)

$$\frac{M_p}{M_0} = 1 - \exp \left(-\frac{\Delta V_{tot}}{V_j} \right)\tag{15.12}$$

where M_p is the total amount of propellant used. In Figure 15.3, the values of ΔV_0 , α_0 , ΔV_2 and M_p/M_0 (in percentage) are plotted as a function of the flight time to the target, for the initial

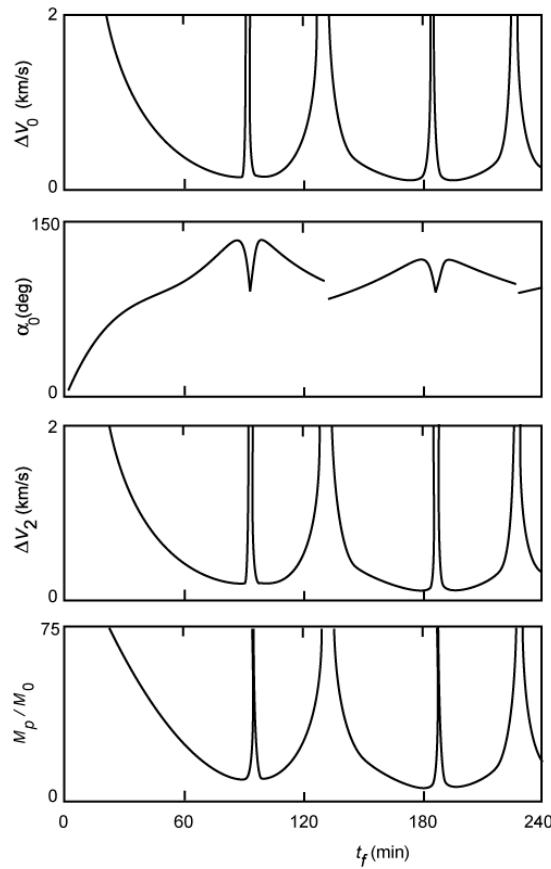


Figure 15.3: Variation of ΔV_0 , α_0 , ΔV_2 and M_p/M_0 as a function of the flight time for a flight from the initial relative position $x_0 = -20$ km, $y_0 = -2358$ km, $z_0 = 0$ km to the final hold point, and for $V_j = 3$ km/s. The target moves in a 400 km altitude circular orbit; the chaser moves initially in a 380 km altitude circular orbit.

conditions listed in Section 15.1 and $V_j = 3$ km/s. This Figure shows that for very short flight times $\alpha_0 \approx 0$, and ΔV_0 and ΔV_2 are very large, which was to be expected. A (local) minimum value of M_p/M_0 occurs for $t_f = 86.22$ min. For that flight time: $\alpha_0 = 134.2^\circ$, $\Delta V_0 = 152$ m/s, $\dot{x}_1 = -110$ m/s, $\dot{y}_1 = -94$ m/s, $\dot{x}_2 = 4.0$ m/s, $\dot{y}_2 = -139$ m/s, $\Delta V_2 = 139$ m/s, $M_p/M_0 \approx 9.3\%$. Note the large value of α_0 . When we consider the trajectory of the chaser again as an orbit about the Earth relative to a geocentric non-rotating reference frame, then it is interesting to compare that (local) optimum trajectory with a Hohmann transfer trajectory from the initial orbit of the chaser to the orbit of the target. For that trajectory we find: $\Delta V_0 = 5.68$ m/s, $\Delta V_2 = 5.68$ m/s, and $t_f = 46.10$ min. In the Hohmann trajectory the chaser meets the target after half a revolution and this requires the target to be only slightly ahead of the chaser at t_0 : $y_0 \approx -49$ km. In this Section we have assumed that the chaser is 2358 km behind the target at t_0 and the chaser therefore cannot fly a Hohmann-type of transfer trajectory, but has to fly an orbit about the Earth with a period much shorter of that of the target's orbit, and thus with a perigee distance smaller than the radius of the initial orbit of the chaser, in order to meet the target after about a full revolution at a point in the target's orbit about 336° ‘ahead’ of the target’s position at t_0 . A further analysis shows that this trajectory leads to an unrealistically low perigee altitude. However, in this Section we do not consider the various constraints that exist for feasible rendez-vous trajectories. The value of ΔV_{tot} is very dependent on the value of y_0 ; e.g. when $y_0 = -500$ km we find $\Delta V_{tot} = 76$ m/s and $t_f = 78.80$ min. When the chaser is allowed to complete more than one revolution in its rendez-vous trajectory,

the relative spacecraft geometry becomes somewhat better, leading to a smaller minimum value of ΔV_{tot} and M_p/M_0 , but the flight time, of course, increases. For example, when the chaser completes two revolutions about the Earth, the flight time increases from 86.22 min to 176.24 min, but ΔV_{tot} decreases from 291 m/s to 154 m/s. For a flight time of 92.41 min (orbital period of the target), 130.00 min, 184.83 min, 226.00 min, for which the values of nt_2 are 2π , 2.813π , 4π , 4.891π , respectively, a very unfavorable relative spacecraft geometry occurs, which results in very large values of ΔV_{tot} and thus of M_p/M_0 .

15.3. Flight from the last hold point

For the numerical examples presented in this Section, it is assumed that the last hold point is located on the Y -axis ‘behind’ the target, which moves in a circular 400 km altitude orbit, and that the initial conditions are $x_0 = z_0 = \dot{x}_0 = \dot{y}_0 = \dot{z}_0 = 0$, $y_0 = -5$ km. Two possible cases will be distinguished: a flight after a rocket pulse in the X -direction (radial) or a flight after a rocket pulse in the Y -direction (tangential).

Rocket pulse in the radial direction

Immediately after the first impulsive shot ΔV_0 the following conditions hold:

$$t = t_0 : \quad x_0 = z_0 = \dot{y}_1 = \dot{z}_1 = 0 \quad ; \quad y_0 = -5 \text{ km} \quad ; \quad \dot{x}_1 = \Delta V_0$$

With these conditions, we find from (15.1) for the motion of the chaser after the impulsive shot

$$\begin{aligned} x &= \frac{\Delta V_0}{n} \sin nt \\ y &= y_0 - 2 \frac{\Delta V_0}{n} (1 - \cos nt) \end{aligned} \tag{15.13}$$

By differentiating these equations, we find for the velocity components

$$\begin{aligned} \dot{x} &= \Delta V_0 \cos nt \\ \dot{y} &= -2\Delta V_0 \sin nt \end{aligned} \tag{15.14}$$

If we require that at t_2 : $x_2 = y_2 = 0$, then we obtain from (15.13)

$$\begin{aligned} \Delta V_0 \sin nt_2 &= 0 \\ 2\Delta V_0 (1 - \cos nt_2) &= ny_0 \end{aligned} \tag{15.15}$$

From (15.15-1) we find $\sin nt_2 = 0$, which leads to $nt_2 = 0, \pi, 2\pi, \dots$. We then find from (15.15-2)

$$\Delta V_0 = \infty \quad \text{for} \quad nt_2 = 0, 2\pi, 4\pi, \dots ; \quad y_0 \neq 0$$

$$\Delta V_0 = \frac{1}{4}ny_0 \quad \text{for} \quad nt_2 = \pi, 3\pi, 5\pi, \dots$$

So, the ‘real’ solution is

$$\Delta V_0 = \frac{1}{4}ny_0 \quad \text{for} \quad nt_2 = \pi, 3\pi, 5\pi, \dots \tag{15.16}$$

and we conclude that the chaser arrives at the target after half a revolution about the Earth. Note

that if $y_0 < 0$, which means that the chaser is originally ‘behind’ the target, $\Delta\bar{V}_0$ should be directed towards the center of the Earth (negative X -axis); if $y_0 > 0$, then $\Delta\bar{V}_0$ should be directed radially outward. If no braking maneuver is applied, both vehicles will approach each other again after each subsequent full revolution about the Earth. Substitution of $nt_2 = \pi, 3\pi, 5\pi, \dots$ into (15.14) gives

$$\dot{x}_2 = -\Delta V_0 \quad ; \quad \dot{y}_2 = 0$$

So, for the second impulsive shot we obtain

$$\Delta V_2 = |\dot{x}_2| = \frac{1}{4}n y_0 \quad (15.17)$$

Note that $\Delta\bar{V}_2$ has the same direction as $\Delta\bar{V}_0$. From (15.16) and (15.17) we find for the total velocity change required for this maneuver

$$\Delta V_{tot} = \frac{1}{2}n y_0 \quad (15.18)$$

An illustration of the relative trajectory for $y_0 = -5$ km is presented in Figure 15.4 as case I. The (relative) trajectory is an ellipse with eccentricity $e = \frac{1}{2}\sqrt{3}$, as was shown in Section 9.3.

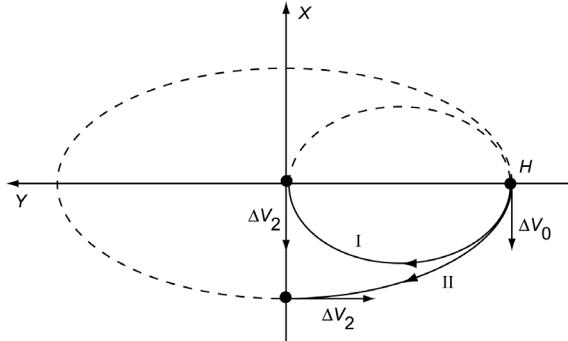


Figure 15.4: Rendez-vous trajectories after a radial impulsive shot in the final hold point at $y_0 = -5$ km. Case I corresponds to an impulsive shot $\Delta V_0 = 0.25 ny_0$; case II to $\Delta V_0 = 0.5 ny_0$.

For some missions it may be important that the chaser first arrives on the X -axis some distance away from the target, and that the final phase starts from this position (Figure 15.4, case II). In that case, we require that at t_2 : $y_2 = \dot{x}_2 = 0$ and (15.14-1) leads to

$$\Delta V_0 \cos nt_2 = 0$$

So, a necessary condition is $\cos nt_2 = 0$, which leads to $nt_2 = \pi/2, 3\pi/2, 5\pi/2, \dots$. This means that the chaser arrives on the X -axis after one quarter, three quarters, five quarters, etc. of an orbital revolution about the Earth. For these values of nt_2 , (15.13) and (15.14) give

$$x_2 = \pm \frac{1}{2}y_0 \quad ; \quad \Delta V_0 = \frac{1}{2}n y_0 \quad ; \quad \dot{y}_2 = \pm 2\Delta V_0$$

and we find for the second impulsive shot that is needed to brake on the X -axis

$$\Delta V_2 = |\dot{y}_2| = 2\Delta V_0 = ny_0 \quad (15.19)$$

So, the total required velocity change for this maneuver is

$$\Delta V_{tot} = \frac{3}{2}ny_0 \quad (15.20)$$

This total impulse is three times the impulse that is needed for the first case. The resulting relative trajectory for $y_0 = -5 \text{ km}$ and $nt_2 = \pi/2$ is shown in Figure 15.4 as case II.

Rocket pulse in the along-track direction

In this case, the following conditions hold immediately after the impulsive shot ΔV_0 :

$$t = t_0 : \quad x_0 = z_0 = \dot{x}_1 = \dot{z}_1 = 0 \quad ; \quad y_0 = -5 \text{ km} \quad ; \quad \dot{y}_1 = \Delta V_0$$

and (15.1) yields for the motion of the chaser after the impulsive shot:

$$\begin{aligned} x &= 2 \frac{\Delta V_0}{n} (1 - \cos nt) \\ y &= y_0 - \frac{\Delta V_0}{n} (3nt - 4 \sin nt) \end{aligned} \tag{15.21}$$

Differentiation of these equations gives

$$\begin{aligned} \dot{x} &= 2 \Delta V_0 \sin nt \\ \dot{y} &= -\Delta V_0 (3 - 4 \cos nt) \end{aligned} \tag{15.22}$$

If we require that upon arrival at the target $x_2 = y_2 = 0$, then we find from (15.21)

$$\begin{aligned} \Delta V_0 (1 - \cos nt_2) &= 0 \\ \Delta V_0 (3nt_2 - 4 \sin nt_2) &= ny_0 \end{aligned} \tag{15.23}$$

From (15.23-1) we find $\cos nt_2 = 1$, which leads to $nt_2 = 0, 2\pi, 4\pi, \dots$. We then find from (15.23-2)

$$\Delta V_0 = \frac{ny_0}{6N\pi} \quad \text{for } N = 0, 1, 2, 3, \dots$$

where N is the number of revolutions that both spacecraft complete about the Earth before a braking maneuver is executed. Note that ΔV_0 becomes smaller when N is larger. When we assume that encounter takes place just after one orbital revolution, we obtain for the first impulsive shot

$$\Delta V_0 = \frac{ny_0}{6\pi} \tag{15.24}$$

Note that if $y_0 < 0$, which means that the chaser is originally ‘behind’ the target, $\Delta \bar{V}_0$ should be directed in the negative Y -direction; if $y_0 > 0$, then $\Delta \bar{V}_0$ should be directed in the positive Y -direction. So, in both cases the impulsive shot is *directed away* from the target. This is contrary to what one would expect intuitively! The astronauts of the Gemini IV spacecraft, who experimented with rendez-vous maneuvers (Section 15.1) were not aware of this phenomenon. When they fired their thrusters in the direction of the spent rocket stage, they found themselves moving away and downward. After several tries to get closer and with half their thruster propellant spent, they gave up.

The physical explanation of this phenomenon is quite simple. Suppose that satellite 2 is initially ‘behind’ satellite 1 and that the impulsive shot ΔV_0 is given in the $-Y$ -direction. Then the

point in the orbit of satellite 1 where this impulsive shot is applied becomes apogee of the orbit of satellite 2 about the Earth. The perigee distance of this orbit is less than the radius of the orbit of satellite 1, and therefore the period of the orbit is shorter than the period of the orbit of satellite 1. So, after satellite 2 has completed one orbital revolution, and has arrived again in the orbit of satellite 1, satellite 1 has not yet completed an orbital revolution and, consequently, satellite 2 has moved in the $+Y$ -direction relative to satellite 1. By selecting an appropriate value of ΔV_0 , and thus of the period of the orbit of satellite 2, satellite 1 arrives just at apogee of the orbit of satellite 2 when that satellite has completed one orbital revolution. For the case that satellite 1 is initially ‘ahead’ of satellite 1 and ΔV_0 is given in the $+Y$ -direction, a similar reasoning shows that then satellite 2 moves in an elliptical orbit with a period longer than the period of the orbit of satellite 1, and therefore has moved in the $-Y$ -direction relative to satellite 1 when it arrives in the orbit of satellite 1. Note that this maneuver, in essence, is a high phasing orbit maneuver, which was discussed in Section 14.4.

For the velocity at t_2 , we find from (15.22) for $nt_2 = 0, 2\pi, 4\pi, \dots$:

$$\dot{x}_2 = 0 \quad ; \quad \dot{y}_2 = \Delta V_0$$

This gives for the second impulsive shot

$$\Delta V_2 = |\dot{y}_2| = \frac{n y_0}{6N\pi} \quad \text{for } N = 0, 1, 2, 3, \dots \quad (15.25)$$

This impulsive shot is also directed away from the target. Combination of (15.24) and (15.25) yields for the total velocity change required for this maneuver

$$\Delta V_{tot} = \frac{n y_0}{3N\pi} \quad \text{for } N = 0, 1, 2, 3, \dots \quad (15.26)$$

The resulting relative trajectory for $N=1$ is shown in Figure 15.5; the dashed part of the curve indicates the relative trajectory if the second impulsive shot would not be applied.

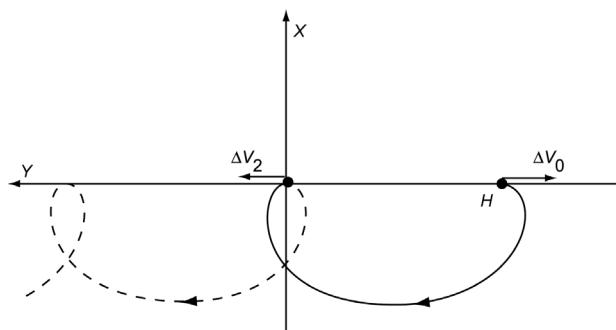


Figure 15.5: Rendez-vous trajectory after an along-track impulsive shot $\Delta V_0 = ny_0/6\pi$ in the final hold point at $y_0 = -5$ km.

Propellant consumption

For the computation of the amount of propellant, M_p , required for these maneuvers, we start from (15.12) and substitute the values of the total velocity change for the maneuvers discussed above. When we assume that both impulsive shots are executed by the same rocket engine, we find for the two cases considered for pulses in the radial direction:

$$\text{case I: } \frac{M_p}{M_0} = 1 - \exp\left(-\frac{ny_0}{2V_j}\right) \quad ; \quad \text{case II: } \frac{M_p}{M_0} = 1 - \exp\left(-\frac{3ny_0}{2V_j}\right) \quad (15.27)$$

For the case of rocket pulses in the along-track direction, we find

$$\frac{M_p}{M_0} = 1 - \exp\left(-\frac{ny_0}{3N\pi V_j}\right) \quad (15.28)$$

As $3N\pi > 2$ for $N \geq 1$, a comparison of (15.27) and (15.28) shows that, for the same values of V_j and y_0 , a maneuver with an impulsive shot in the along-track direction is always more efficient in terms of propellant consumption.

15.4. Last phase of the rendez-vous flight

In the last phase of the rendez-vous flight, the chaser is within a distance of 300 m from the target. Then, the chaser can ‘observe’ the target and, usually, it approaches the target along a straight line: the LOS. During the LOS approach, rocket engine thrust pulses are periodically applied along the LOS in order to reduce the velocity of the chaser relative to the target and to eventually nullify this relative velocity. As mentioned in Section 15.1, for the last phase this approach strategy is to be preferred over a two-impulses approach maneuver along a curved trajectory. In Section 9.3 it has been shown that the true trajectory of the chaser relative to the target is an ellipse, if no force other than that expressed by a Newton gravity field is acting on both satellites. Consequently, in general, the chaser will gradually move away from the LOS and therefore also rocket thrust pulses perpendicular to the LOS are required periodically in order to keep the chaser near the LOS.

In principle, the chaser may approach the target from all directions, but in practice two approaches are frequently used: the *R-bar approach* and the *V-bar approach*. In the *R-bar* approach, the chaser approaches the target along the *X*-axis, i.e. along the target’s orbital radius vector (\bar{R}). In the *V-bar* approach, the chaser approaches the target along the *Y*-axis, i.e. along the axis that is directed along the target’s velocity vector (\bar{V}). Below, these two characteristic types of LOS approach maneuvers will be analyzed. It should be mentioned here that in our definition *R-bar* is directed radially outwards, but that in most US literature *R-bar* is directed towards the center of the Earth; in both cases *V-bar* is directed along the velocity vector.

R-bar approach

For this approach flight, the initial conditions are $y_0 = z_0 = \dot{y}_0 = \dot{z}_0 = 0$. If there is no thrust acting on the chaser, the equations of motion (15.1) in the *XY*-plane simplify to

$$\begin{aligned} x &= (4 - 3 \cos nt) x_0 + \frac{\dot{x}_0}{n} \sin nt \\ y &= -6(nt - \sin nt) x_0 - 2 \frac{\dot{x}_0}{n} (1 - \cos nt) \end{aligned} \quad (15.29)$$

Equation (15.29-2) shows that for $t > 0$: $y \neq 0$; consequently, the spacecraft will drift away from the *X*-axis. So, rocket pulses in the *Y*-direction are required to keep the chaser near the *X*-axis. In between successive rocket pulses, the radial velocity and acceleration of the spacecraft can be found by differentiation of (15.29-1), which yields

$$\begin{aligned}\dot{x} &= 3x_0 n \sin nt + \dot{x}_0 \cos nt \\ \ddot{x} &= 3x_0 n^2 \cos nt - \dot{x}_0 n \sin nt\end{aligned}\tag{15.30}$$

where x_0 and \dot{x}_0 refer to the conditions just after the previous impulsive shot and t is the time elapsed since that impulse. The interval between successive impulses is relatively short, and therefore $nt \ll \pi/2$. It is clear that both for the case that the chaser approaches the target from ‘above’ ($x_0 > 0, \dot{x}_0 < 0$) and for the case that the approach is from ‘below’ ($x_0 < 0, \dot{x}_0 > 0$), x_0 and \dot{x}_0 will have opposite signs. We conclude from (15.30-2) that for both cases the radial acceleration is directed away from the target. This means that the chaser experiences a ‘natural braking effect’; during the approach this effect has to be counteracted by rocket thrust pulses along the LOS in the direction to the target. If these thrusters would fail, the ‘natural braking effect’ will cause the chaser to move away from the target, at least when the approach velocity is not too large. So, it is a ‘build-in safety mechanism’. The physical explanation for this braking phenomenon is as follows. If the chaser is (in between successive pulses) moving along the X -axis and approaches the target from ‘above’, then its velocity in the Y -direction is larger than the local circular velocity. So, the chaser experiences a relative acceleration in the $+X$ -direction, which is the ‘braking effect’ mentioned above. If the chaser is moving along the X -axis but approaches the target from ‘below’, then its velocity is smaller than the local circular velocity and it will experience a relative acceleration in the $-X$ -direction, again away from the target.

Because $nt \ll \pi/2$, we may approximate (15.30) by applying series expansions for the sine- and cosine-terms and retaining only the leading terms of the resulting expressions. We then find for the motion in between successive pulses

$$x \approx x_0 + \left(\frac{\dot{x}_0}{n} + \frac{3}{2} x_0 n t \right) n t ; \quad y \approx - \left(\frac{\dot{x}_0}{n} + x_0 n t \right) (n t)^2$$

The second term of the expression in brackets in the first equation expresses the ‘natural braking effect’. These relations shows that to first-order approximation: 1) the variation of x and y is linearly dependent on the value of \dot{x}_0 ; 2) the variation of x is linearly dependent on time, while the variation of y is quadratically dependent on time; 3) for an approach from ‘above’ the chaser starts to lead the target in the Y -direction, for an approach from ‘below’ the chaser starts to lag ‘behind’ the target in the Y -direction.

As an example, Table 15.1 lists for a realistic R -bar approach from ‘above’ to a target moving in a circular 400 km altitude orbit about the Earth, the relative position, radial velocity and radial acceleration at $t = 30$ s. It was found that the first term on the right-hand side of (15.30-2) contributes about 99% of the total radial acceleration. The ‘natural braking effect’ was found to

Table 15.1: Position, velocity and acceleration of the chaser at 30 s after the previous thruster impulsive shot during an R -bar or a V -bar approach of a target moving in a circular orbit at an altitude of 400 km above the Earth.

R-bar approach		V-bar approach	
$x_0 = 200$ m; $\dot{x}_0 = -0.2$ m/s		$y_0 = -200$ m; $\dot{y}_0 = 0.2$ m/s	
x (m)	194	y (m)	-194
\dot{x} (m/s)	-0.177	\dot{y} (m/s)	0.200
\ddot{x} (m/s ²)	7.76×10^{-4}	\ddot{y} (m/s ²)	-3.08×10^{-5}
y (m)	0.196	x (m)	0.204

be $7.8 \times 10^{-4} \text{ m/s}^2$.

V-bar approach

In this case, the initial conditions are $x_0 = z_0 = \dot{x}_0 = \dot{z}_0 = 0$. If there is no thrust acting on the chaser, then the equations of motion (15.1) in the XY-plane simplify to

$$\begin{aligned} x &= 2 \frac{\dot{y}_0}{n} (1 - \cos nt) \\ y &= y_0 - \frac{\dot{y}_0}{n} (3nt - 4\sin nt) \end{aligned} \quad (15.31)$$

Equation (15.31-1) shows that for $t > 0$: $x \neq 0$; consequently, the chaser has the tendency to drift away from the Y-axis. So, rocket thrust pulses perpendicular to the Y-axis are needed in order to keep the chaser near the Y-axis. In between successive rocket pulses, the along-track velocity and acceleration of the spacecraft can be found by differentiation of (15.31-2), which yields

$$\begin{aligned} \dot{y} &= -\dot{y}_0 (3 - 4\cos nt) \\ \ddot{y} &= -4\dot{y}_0 n \sin nt \end{aligned} \quad (15.32)$$

where \dot{y}_0 denotes the along-track velocity just after the previous impulsive shot and t is the time elapsed since that impulse. Note that both \dot{y} and \ddot{y} are a function of \dot{y}_0 , but not of y_0 . Again, we may assume that $nt \ll \pi/2$. Equation (15.32-2) shows that both for the case that the chaser approaches the target from ‘behind’ ($y_0 < 0, \dot{y}_0 > 0$) and for the case that the chaser approaches the target from ‘ahead’ ($y_0 > 0, \dot{y}_0 < 0$), the acceleration is directed away from the target. So, also in this case the chaser experiences a ‘natural braking effect’, which may be considered as a ‘build-in safety mechanism’, at least when the approach velocity is not too large. During the approach this effect has to be counteracted by rocket thrust pulses along the LOS. In this case, the physical explanation for the braking phenomenon is as follows. If the chaser is moving along the Y-axis and approaches the target from ‘behind’, its velocity in the Y-direction is slightly larger than the local circular velocity. So, the chaser will start to enter a higher orbit and consequently will move in the +X-direction. This leads to a lower velocity, which corresponds to an acceleration away from the target. If the chaser is moving along the Y-axis but approaches the target from ‘ahead’, then its velocity is slightly smaller than the local circular velocity and it will enter a lower orbit. This leads to a higher velocity, which corresponds again to an acceleration away from the target. It is emphasized that, in contrast to the situation during an R-bar approach, in this case the braking effect is an ‘indirect’ effect, because it is produced by a departure of the chaser from the Y-axis.

Because $nt \ll \pi/2$, we may approximate (15.31) by applying series expansions for the sine- and cosine-terms and retaining only the leading terms of the resulting expressions. We then find for the motion in between successive pulses

$$x \approx \dot{y}_0 n t^2 \quad ; \quad y \approx y_0 + \dot{y}_0 t$$

This relation shows that to first-order approximation: 1) the variation of x and y is linearly dependent on the value of \dot{y}_0 ; 2) the variation of x is quadratically dependent on time, while the variation of y is linearly dependent on time; 3) for an approach from ‘behind’, the chaser starts to move in the +X-direction; for an approach from ‘ahead’, the chaser starts to move in the -X-direction.

As an example, Table 15.1 lists for a realistic *V*-bar approach from ‘behind’ to a target moving in a circular 400 km altitude orbit about the Earth, the relative position, velocity and acceleration in the *Y*-direction at $t = 30$ s. In this case, the ‘natural braking effect’ was found to be about 3.1×10^{-5} m/s², or only about 4% of the braking effect during the *R*-bar approach. In fact, this deceleration is so small that, without additional braking along the *Y*-axis, in practice the chaser’s velocity along the *Y*-axis will not be reduced to zero before the chaser reaches the target. So, we conclude that the braking effect along a *V*-bar approach is of little practical value.

15.5. Typical Space Shuttle rendez-vous flight

To provide an overview of a real rendez-vous mission, in this Section a description will be given of a typical rendez-vous flight of the former Space Shuttle with a target vehicle that is already in orbit at the time of the Shuttle launch.

The Space Shuttle was a reusable launch system and orbital spacecraft operated by NASA for human spaceflight missions. The total Space Shuttle system was officially named the Space Transportation System (STS) and Shuttle flights were indicated by the abbreviation STS followed by a number. Six airworthy Orbiters were built; the first, Enterprise, was not built for orbital space flight, and was used only for testing gliding and landing. Enterprise was followed by Columbia, Challenger, Discovery, Atlantis, and Endeavour. The first flight to space (STS-1) was performed by the Shuttle Columbia, launched on April 12, 1981; the last flight (STS-135) by the Shuttle Atlantis, launched on July 8, 2011. In total, the Shuttles were launched 135 times; on 9 missions a docking with the Mir and on 36 missions a docking with the International Space Station (ISS) was established. Challenger disintegrated 73 s after launch on January 28, 1986, and Endeavour was built as a replacement for Challenger. Columbia broke apart during reentry on February 1, 2003. The earliest Shuttle flights had the minimum crew of two; many later missions a crew of five or seven (commander, pilot, several mission specialists, and rarely a flight engineer). On two occasions, eight astronauts have flown. Eleven people could be accommodated in an emergency mission. Major missions have included launching numerous satellites and interplanetary spacecraft, conducting space science experiments, and servicing and construction of the Mir and the ISS.

Launch phase

For rendez-vous missions, the Shuttle is launched near the time of day that the Earth’s rotation brings the target’s orbital plane over the launch site. The launch window around the optimal launch time mainly depends on the inclination of the target’s orbit; targets with an inclination of 28.5° have rendez-vous launch windows up to one hour long, while targets in higher inclination orbits can have launch windows less than 5 min in duration. After launch, the Shuttle is steered to the target’s orbital plane with the main engines using yaw steering. At the time of launch, the target can be anywhere in its orbital plane. Due to performance limitations and constrained crew activity timelines, only certain phase angles at launch are permissible for rendez-vous. This phase window can be anywhere from 10 min long (40° of on-orbit phase angle make-up capability) to 90 min long (360° of phase angle make-up), depending on the target altitude, crew timeline, and crew day that the actual rendez-vous occurs. The relationship of the phase window to the launch window changes each day during the launch period and depends on the target’s orbital period and inclination. More information about the launch phase is presented in Section 16.3.

Orbit adjust phase

The Shuttle enters this flight trajectory regime after the ascent phase. Using the Shuttle *Standard*

Insertion (SI) ascent procedure, the *Orbital Maneuvering System* (OMS) engines are fired approximately 2 min after Shuttle *Main Engine Cut-Off* (MECO), which occurs about 8.5 min after launch, to raise the apogee altitude above 185 km. This maneuver is called the *OMS-1 burn*. Approximately 30 min later at apogee, the engines are fired again (OMS-2) to raise perigee to a safe altitude. For rendez-vous missions, the OMS-2 burn is targeted as a posigrade phasing burn and is the first on-orbit maneuver to begin the process of rendez-vous. Another available ascent technique is called *Direct Insertion* (DI), in which the Shuttle main engine performance is used to establish a sufficiently high apogee at the time of MECO so that the OMS-1 burn is not necessary. The first maneuver on orbit is still called OMS-2 and is targeted in the same manner as for standard insertions. With the completion of OMS-2, the Shuttle is phasing towards the target. In order to provide maximum flexibility, this mission-unique phase of the rendez-vous is controlled from the ground.

The Shuttle state vector and attitude information is supplied by the onboard *Inertial Measurement Units* (IMUs). To compensate for the drift of these systems, the state vector information provided by the IMUs is corrected by a ground uplink of orbital parameters. To this end, the Shuttle is tracked using the *Tracking and Data Relay Satellite System* (TDRSS) S-band communication links, and S-band and C-band ground tracking systems. To correct the attitude information, the Shuttle takes star sightings to keep the error in the attitude below 0.25° . The target is tracked with S-band ground tracking systems (if the target carries the appropriate transponder) and C-band systems are used to skin track the target. These C-band tracking systems have been used on most Shuttle rendez-vous missions when the targets were passive disabled satellites. Determination of the semi-major axis of each vehicle is the primary concern during rendez-vous, since errors in the semi-major axis cause secular along-track dispersions (Section 10.1). Generally, the Shuttle's semi-major axis is known to within 500 m (3σ) and the target's semi-major axis is known to better than 30 m (3σ) after two to three orbits of tracking data. Shuttle semi-major axis knowledge is gradually degraded due to translation effects caused by attitude maneuvers, attitude dead-banding, Shuttle venting, and translational maneuvering trim limits.

In the mid 1990s, a five-channel Global Positioning System (GPS) receiver was integrated into the Shuttle avionics system due to the anticipated start of Tactical Air Control and Navigation (TACAN) phase-out in the year 2000. The Shuttle used TACAN data during reentry and the flight to the landing strip. The use of GPS data from a single receiver in parallel with TACAN during entry was successfully demonstrated during four Shuttle missions in 2006 and 2007. A slip in TACAN phase-out initiation to 2010 allowed the program to continue using TACAN for operational purposes and to further experiment with GPS. Flight STS-118 on August 8, 2007, was the first Shuttle mission flown with three GPS receivers in place of the three legacy TACAN units. During later flights the GPS information was also available to the crew during the orbital phasing and rendez-vous phases. It was used for limited orbital operations but not for rendez-vous; neither for absolute positioning nor for positioning relative to a target. The reasons were: 1) the standard navigation concept was accurate enough and fully flight proven; 2) the Shuttle program had a requirement to rendez-vous with passive targets; 3) the obscuration of GPS signals by the ISS and signal multi-path by reflections of various ISS structure elements could pose serious problems for relative GPS in close proximity to the station.

Upon completion of navigation processing, the *Rendez-vous Targeting Processor* (RTP) in the *Mission Control Center* (MCC) in Houston is used to compute the remaining rendez-vous maneuvers. This processor is a generalized targeting program, which allows flexible scheduling of maneuvers. Maneuver types include horizontal phasing and altitude adjustments, co-elliptic maneuvers, plane change maneuvers, and two-impulse Lambert type (Section 6.7) maneuvers.

Scheduling options include geometric constraints (e.g. node crossing, apogee, elevation angle) and lighting constraints. Upon completion of the rendez-vous targeting, the results are transferred into an ephemeris¹ maintained in the MCC and the upcoming maneuver is uplinked to the crew. Onboard support is limited to crew execution of the ground targeted maneuvers using *External Delta-V Guidance* (EXDV). The uplinked maneuver is loaded by the crew into the *Maneuver Execute Software* (MES). Within 15 s of planned *Time of Ignition* (TIG), the crew enters an execute command to allow for automatic execution of the required ΔV . EXDV guidance is then used during the burn, which means that accelerometers sense the thrusting every second, decrement the commanded inertial ΔV to go, and compute the time to thrust termination. The engines shut down automatically and the crew then manually trims the maneuver residuals usually to better than 0.1 m/s.

After each orbit maneuver, ground tracking is used again to update the estimated states of both vehicles and the navigation-target-burn sequence is recycled. Pre-mission guidance, navigation and control errors are analyzed in Monte Carlo type of simulations to evaluate trajectory and propellant dispersions and to determine the frequency of maneuvering. Ground controllers design the orbit adjust phase to terminate with the Shuttle less than 100 km directly ‘behind’ the target at orbital noon on the morning of the crew day on which the retrieval will be completed. At this point, onboard sensors can optically detect the target and dedicated crew rendez-vous operations begin.

Relative navigation phase

When the Shuttle closes to within 100 km of the target, the relative navigation phase starts. Maneuvers are scheduled closer together, usually less than one revolution apart. The objective of this phase is to approach to within 2 km of the target, using only passive skin tracking of the target from the Shuttle by its Ku-band radar system. The Ku-band system is first given the general location of the space hardware from the Orbiter computer; then the antenna makes a spiral scan of the area to pinpoint the target. Star trackers onboard the Shuttle can track objects brighter than the third magnitude. The target attitude profile and reflectance must be such that the Shuttle can track the target from orbital noon through orbital sunset as the approach continues. Star tracker angle measurements are made 30 times per second, averaged, and output to the onboard guidance, navigation and control *General Purpose Computers* (GPCs).

Onboard, relative navigation software selects data once every 8 s for processing and a Kalman filter is used to estimate the relative state. The relative state information is combined with the onboard propagated target inertial state vector to compute the estimated Shuttle inertial state vector. The output of the rendez-vous navigation software is the current target and Shuttle estimated inertial states. Typically, the relative state vector converges after the first star tracker pass to within 2.3 km (3- σ) and 0.8 m/s (3- σ). The crew then uses the onboard navigation results to compute two-impulses targeting maneuvers, using the onboard *Orbit Targeting Specialist Function* (OTSF). Pre-mission offset points and transfer times are stored in the flight software data initialization load and are used in the targeting. The maneuver time relative to a reference time, the transfer time, and the aim point are the components of a target set. The aim point is a point relative to the target in the target-centered coordinates frame, where the Shuttle is to be at the end of the transfer. Up to 40 different target sets are defined pre-mission in the data initialization loads. Additionally, the crew can manually overwrite any of the target sets should

¹ *Ephemeris*, plural *ephemerides*, is a table or computer file of values that give the positions of astronomical objects in the sky at a given time or times.

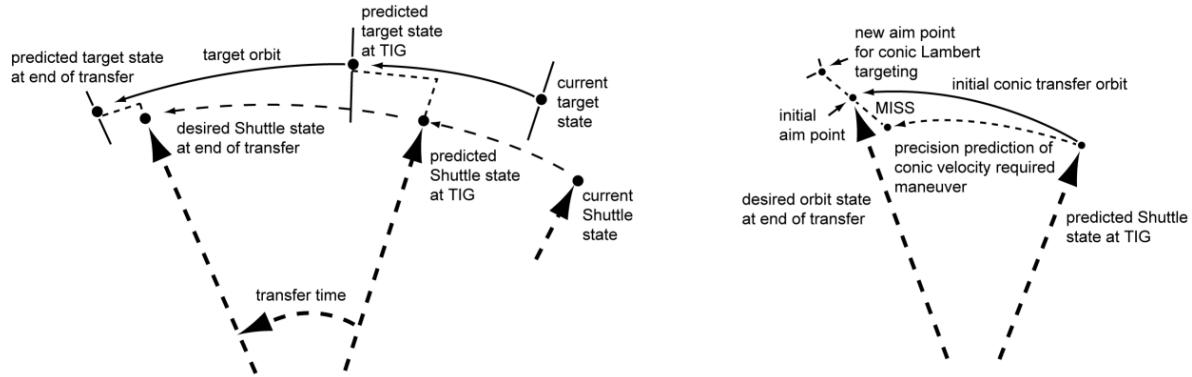


Figure 15.6: Lambert conic targeting technique (left) and a correction method (right).

the need arise.

Two different two-impulses targeting algorithms exist in the OTSF. One algorithm uses a Lambert targeting technique (Section 6.7), in which the current Shuttle and target inertial state vectors are predicted forward to the maneuver time and the relative state vector is computed and displayed to the crew (Figure 15.6). The target inertial vector is then predicted forward through the transfer time, is combined with the aim point and a desired Shuttle inertial position vector is then constructed. A conic Lambert targeting routine then computes the conic transfer orbit. Onboard software is limited to less than one revolution transfer times. The resulting transfer orbit is then predicted forward through the transfer. This results in the predicted Shuttle state at the end of the transfer, missing the aim point slightly due to the non-conic effects. The aim point is then automatically adjusted slightly and the cycle is repeated until convergence is reached. Generally, the convergence criterion for this Lambert targeting technique was 80 m. The other targeting scheme available is based on the Clohessy-Wiltshire closed-form relative-motion equations (Section 9.2). Current onboard inertial state vectors are predicted forward to the ignition time and the relative state is computed. This initial relative state and the desired relative state at the end of the transfer are used directly to compute the velocity change required at TIG. Again, a precision prediction scheme is wrapped around the targeting scheme and the aim point miss is evaluated. If it exceeds the tolerance (one percent of the relative range at the end of the transfer), the aim point is modified and the process is repeated. The resulting maneuver is again automatically transferred to the MES and is executed by the crew using EXDV guidance.

A standard trajectory profile has been developed for the relative navigation phase. Ground controllers place the Shuttle on a relative trajectory that will touch a point 75 km directly ‘behind’ the target at orbital noon. At that point, a ground computed phasing burn is performed to approach to 15 km ‘behind’ the target in either one or two orbital revolutions (Figure 15.7). A star tracker pass occurs immediately after the phasing burn and a second star tracker pass one revolution later, if the mission design uses a two-revolution transfer. After the second star tracker pass, a Lambert maneuver is targeted onboard to correct for in-plane and out-of-plane trajectory dispersion observed with the star tracker. Guidance, navigation and control Monte Carlo simulations should indicate that the Shuttle will miss the 15 km point by less than 1 km (3σ). After the mid-course correction burn, the Shuttle Ku-band system is used in the rendez-vous radar mode to track the target. Radar data is incorporated every 8 s and the relative state navigation uncertainty is reduced to 275 m and 0.5 m/s (3σ) by the time the Shuttle reaches the 15 km range. The maneuver at the 15 km range is called the *Transition Initiation* (TI) maneuver. It is the maneuver where the Shuttle transitions to actually targeting directly for intercept with the

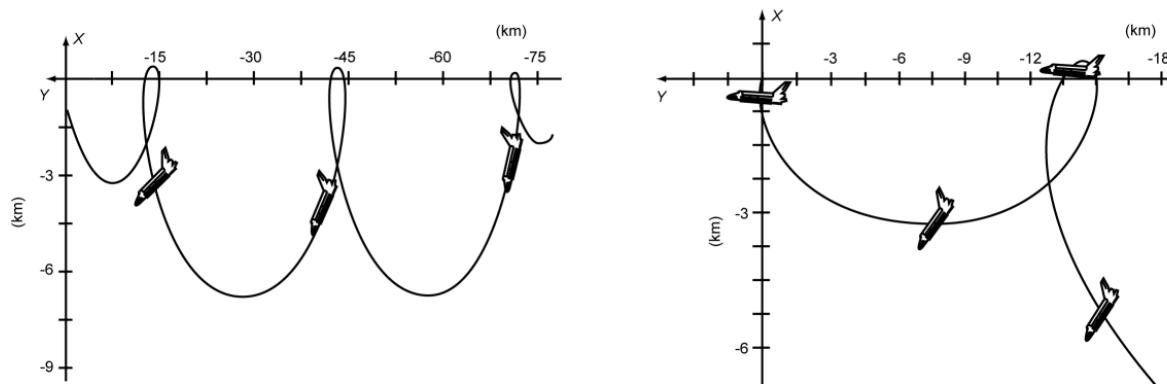


Figure 15.7: Standard Space Shuttle rendez-vous profile.

target. Just prior to TI, rendez-vous guidance, navigation and control data are evaluated to determine if the crew is ‘go for TI’. If there is some concern about continuing the rendez-vous, a new maneuver may be executed at TI to interrupt the sequence and allow for station-keeping (hold point) to resolve the situation. Upon completion of TI, the Shuttle again uses rendez-vous radar information to update its knowledge of the relative state and five mid-course maneuvers are scheduled to correct the trajectory during the next hour. If the radar has failed, the orbital lighting is such that optical tracking with the star tracker is used immediately following TI.

Proximity operations phase

By the completion of the last mid-course maneuver, orbital sunrise has occurred, the range has decreased to about 2 km, and the relative rates are under 2 m/s. Additionally, the relative trajectory has been designed such that the inertial LOS rates to the target are nearly zero at manual takeover and the Shuttle is ‘below’ the target altitude so that the background view is dark. The target appears as a bright star frozen against a stellar background. Because of single point failures possible with the rendez-vous radar, this final rendez-vous phase has been designed to allow for manual piloting since no relative navigation tracking support can be assured. Two minutes after the last mid-course maneuver, the crew changes the Orbiter’s automated pointing mode from target-track to inertial hold. The trajectory control requirements now are to control range-rate as a function of the decreasing range while maneuvering perpendicular to LOS to keep the target centered in the *Crew Optical Alignment Sight* (COAS) mounted on the upper windows. The *V*-bar approach is the preferred approach technique. A standard range-rate versus range-braking schedule has been developed, which minimizes plume impingement on the target and yet allows the range to decrease to 120 m in 20 min as the Shuttle reaches the *V*-bar down-track position. As the Shuttle nears the *V*-bar, the automated attitude control mode is switched from inertial to a Shuttle tail-to-Earth pointing mode with its bottom into the flight direction (Figure 15.8, left); the relative rates have now decreased to under 0.2 m/s. When crossing the *V*-bar, normal to the LOS rates are nulled to keep the Shuttle on the *V*-bar and the closing rate is adjusted through a new braking schedule. As the Shuttle continues the approach, orbital dynamics require the crew to periodically fire the aft thrusters to maintain a LOS approach along the *V*-bar (Section 15.4). Since these engines are canted, incremental braking occurs each time these engines are fired. The LOS approach to the target continues as orbital sunset occurs. It is necessary to be within 60 m for optical tracking during the night pass and it is desirable to be within 20 m at orbital sunset, so that the payload bay floodlights can illuminate the payload for grappling operations. The rendez-vous radar drops out at a range of 25 m due to transmit/receive switch-over limitations.

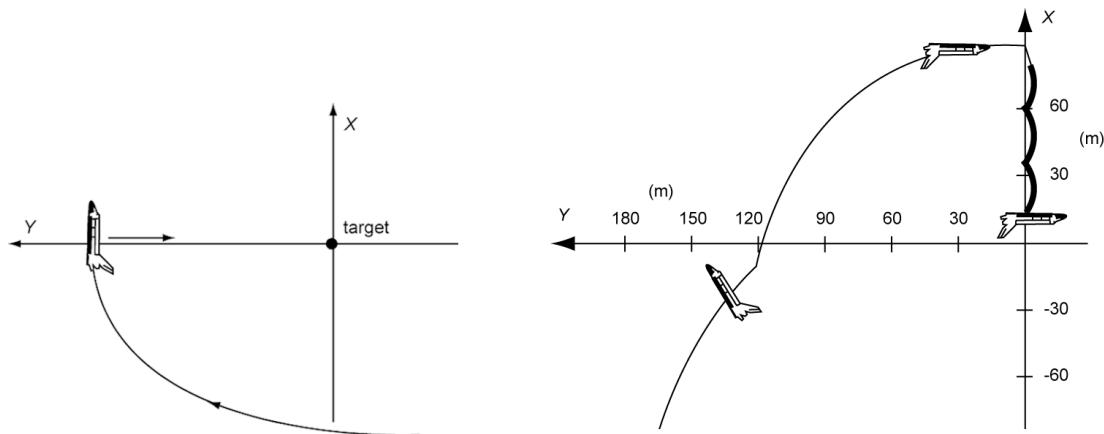


Figure 15.8: Examples of a Space Shuttle V-bar (left) and R-bar (right) approach.

From an operational point of view, the *V*-bar approach is desirable, because: 1) it is relatively fuel efficient; 2) the constant Earth horizon orientation is a good piloting reference; 3) closing rates can be easily and immediately nulled with subsequent efficient stationkeeping should some Shuttle or payload system anomaly occur. However, other techniques for final approach have also been used. For the *Long Duration Exposure Facility* (LDEF), which has been retrieved in January 1990, a mission was designed in which the inertial approach through the *V*-bar crossing was continued for 23 more minutes and subsequently an approach radially down on top of LDEF (Figure 15.8, right) was performed. The Shuttle yawed as necessary during this approach to set up the proper geometry for grapple. In November 1994, for the first time maneuvers to rendezvous along the *R*-bar from ‘below’ were tested with the Russian space station Mir, without actually making contact with Mir. This rendez-vous from ‘below’ with docking was applied for the first time for the docking of an Orbiter to Mir in November 1995. It has the advantage that the target is clearly visible against a dark stellar background, while during a rendez-vous from ‘above’ the Earth’s albedo may blind the crew or the onboard sensors. The big advantage of rendez-vous along the *R*-bar is that the thrusters facing in the opposite direction of the target have to be fired to control the closure rate, while orbital mechanics provides a ‘natural braking effect’ (Section 15.4), which seriously decreases the risk of a collision if the thrusters fail. Approaches along any LOS direction have been proposed to allow even more design flexibility. Retrieval of targets without the standard Shuttle grapple fixture is also possible through the use of the *Manned Maneuvering Unit* (MMU) as demonstrated in a number of satellite retrievals. Unique crew techniques for arresting target rotation rates and attaching grappling hardware have also been demonstrated with the US Navy Syncron IV-3 salvage mission on STS-51-I already in early September 1985. Since then, many more sophisticated rendez-vous, docking and retrieval maneuvers have been executed.

15.6. Space Shuttle rendez-vous with ISS

The Space Shuttle has performed many rendez-vous and docking missions with/to the ISS, transferring astronauts and cargo to the ISS and returning astronauts and cargo to Earth. The ISS is a large research laboratory in the space environment, where crews conduct scientific experiments each day. Scientists on Earth have a swift access to the crew’s data and can modify experiments or launch new ones. The presence of a permanent crew affords the ability to monitor, replenish, repair and replace experiments and components of the spacecraft itself. The assembly of the ISS began on November 20, 1998, with the launch of the Russian Zarya module by a

Proton rocket and the launch of the US Unity module on December 4, 1998, by the Space Shuttle. The launch and docking of the Russian Zvezda service module on July 12, 2000, enabled permanent habitation of the station. The first resident crew arrived on November 2, 2000, on a Russian Soyuz spacecraft. Since then, the ISS has been gradually enlarged with new modules, large solar panels, etc. From November 2, 2000, to August 1, 2014, there have been 55 Progress (Russia), 39 Soyuz (Russia), 32 Shuttle (USA), 5 ATV (Europe), 4 Dragon (USA), 4 HTV (Japan), 3 Cygnus (USA) successful flights to the station. The choice of the ISS orbital altitude is a compromise between a small decay rate due to atmospheric drag, which requires a high orbit, and a large payload capability of the vehicles that dock to the ISS, which requires a low orbit. The ISS is maintained in a nearly circular orbit with a minimum altitude of 280 km and a maximum of 450 km, and an inclination of 51.6° ; it needs to be boosted to a higher altitude several times each year. This boost can be performed by the station's two main engines on the Zvezda module, or by a docked Shuttle, Progress or ATV. In April 2011 the ISS orbital altitude was 347 km; in March 2012 it was 390 km.

On February 1, 2003, the reentry of the Space Shuttle Columbia ended in a disaster. Columbia had sustained damage to its heat shield due to insulating foam breaking off the external tank and hitting the shield at liftoff. The damage was too great for the heat shield to protect the Shuttle from the heat and structural strain of atmospheric reentry, causing it to break apart. For this reason, the integrity of the heat shield has been a critical concern of NASA ever since and a special maneuver, called the *backflip maneuver*, was designed that allowed inspection of the heat shield from the ISS. For this maneuver the Shuttle approaches the station from 'below' along the *R*-bar and at a small distance from the ISS, usually around 200 m, the Shuttle performs a slow 360° pitch maneuver, during which it exposes the heat shield on the upper surface and underside to the ISS. The crew inside the ISS visually inspects and photographs the heat shield to determine whether or not it has been damaged during liftoff and ascent. This maneuver requires skilled piloting, as the Shuttle commander must fly very close to the ISS without the station always in full view. After the maneuver has completed, the Shuttle flies from the *R*-bar to the *V*-bar, and then continues along the *V*-bar to close in on the ISS and eventually complete the docking. This maneuver was first performed on Shuttle mission STS-114, launched on July 26, 2005. Below, the rendez-vous and docking of the Shuttle with/to the ISS on flight STS-120 will be summarized.

On October 23, 2007, the Shuttle was launched on a flight to the ISS. The main mission goal was to deliver the Harmony module and to reconfigure a portion of the station in preparation for future assembly missions. To this end the cargo included four DC-to-DC converter unit racks and three zero-g storage racks, a power and data grapple fixture for the station's robot arm, and a Shuttle power distribution unit. During a two-days chase of the station with periodic rocket engine firings the Shuttle was gradually brought to about 15 km behind (-Y-axis) the station—the starting point for the final approach. About 2.5 hr before docking, the Shuttle's OMS engines were fired to begin the final phase of the rendez-vous. As the Shuttle moved closer to the ISS, the Shuttle's Ku-band rendez-vous radar system tracked the ISS and provided range and closing rate data to the crew. During the final approach, the Shuttle executed several small correction burns that placed it about 300 m directly below (-X-axis) the station. From that moment on the Shuttle commander manually controlled the Shuttle for the remainder of the approach and docking. The approach was stopped about 180 m below the station to ensure proper lighting for imagery prior to initiating the 360° backflip maneuver. After some time, the maneuver was executed and the station crew took as many as 300 digital pictures of the Shuttle's upper surface and underside. These pictures had a resolution of up to 3 cm. The photos were taken out of

windows in the Zvezda service module and were downlinked to Earth for analysis. When the Shuttle had completed its rotation, its payload bay was facing the station. The commander then moved the Shuttle to a position about 120 m directly in front of (+Y-axis) the station in preparation for the final approach to docking to the station's Destiny docking port. The crew members operated laptop computers processing the navigational data, the laser ranging system and the docking mechanism. Using a view from a camera mounted in the center of the Shuttle's docking system, the commander precisely matched up the docking ports of the two spacecraft. The Shuttle homed in at a relative speed of about 3 cm/s, and the docking mechanisms were kept aligned to within a tolerance of 8 cm. When the Shuttle made contact with the station, preliminary latches automatically attached the two spacecraft. Immediately thereafter, the Shuttle's steering engines were deactivated to reduce the forces acting at the docking interface. Shock absorber springs in the docking mechanism damped any relative motion between the Shuttle and the station. Once the motion between the spacecraft had been stopped, the docking ring was retracted to close a final set of latches between the two vehicles.

Figure 15.9 shows the last part of the trajectory of a very similar rendez-vous flight of the Space Shuttle Endeavour with ISS on May 18, 2011. This was the 25th and last flight of the Space Shuttle Endeavour, the 36th Shuttle flight to ISS, and the last but one flight of the Space Shuttle.

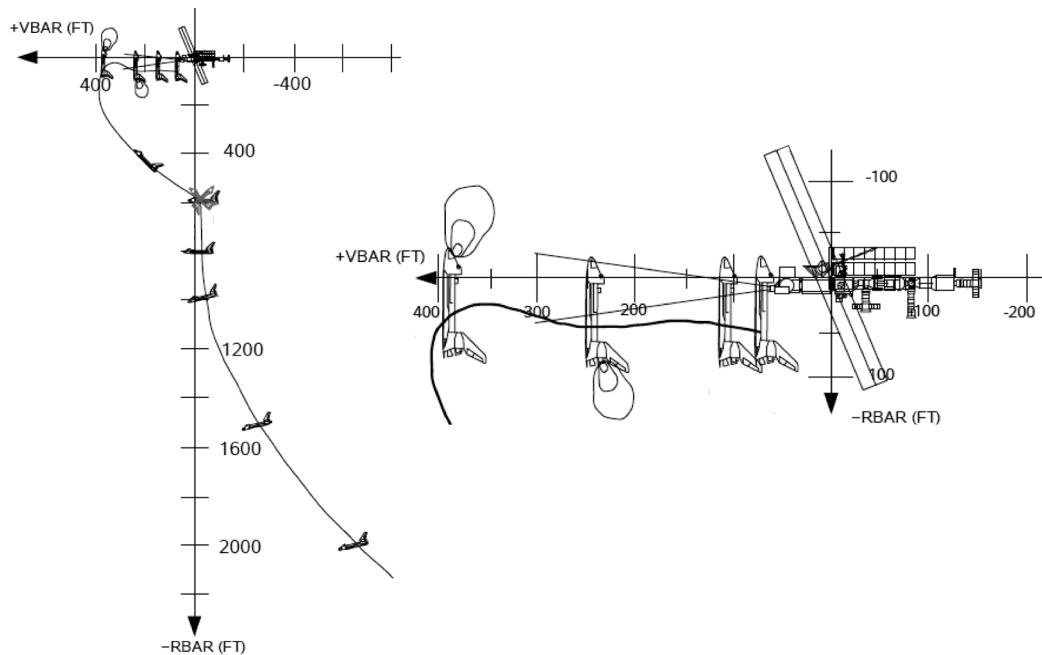


Figure 15.9: The final phase of the Space Shuttle's rendez-vous flight with the International Space Station (ISS) during the STS-134 mission in May 2011.

15.7. ATV rendez-vous with ISS

The Automated Transfer Vehicle (ATV) is an expendable, unmanned resupply spacecraft developed by ESA and designed to supply the ISS with propellant, water, air, payload and experiments. As soon as cargo is removed from the ATV, solid and liquid waste from the station is loaded in its place, meaning that the ATV will slowly fill up with waste over the course of some months. In addition, ATVs can boost the station into a higher orbit. The ATV is designed to complement the Russian Progress spacecraft, having three times its capacity. It is launched by a two-stage Ariane 5 ES rocket from the ESA Kourou Space Center in French Guiana (Section

16.2). After completing its mission, the ATV reenters the Earth's atmosphere and burns up. The ATV uses the GPS system for navigation and orbit control, and its docking system includes two videometers and two telegoniometers. A videometer is based on imaging sensor technology, where laser pulses are sent to and reflected by retroreflectors on the Zvezda module. The resulting image is then processed to provide range, LOS angles and attitude relative to the Zvezda module. A telegoniometer emits laser pulses towards retroreflectors on the Zvezda module. The travel time of the reflected pulses gives the distance between the two spacecraft. The direction from the ATV to the Zvezda module is given by the orientation of the telegoniometer's two built-in mirrors, which rotate to aim the laser towards the retroreflectors. Visual imagery is provided by a camera on the Zvezda module. The first ATV was successfully launched on March 9, 2008; the fifth and last ATV on July 30, 2014. Below, the rendez-vous and docking flight of the second ATV (ATV-2) will be summarized.

ATV-2 was launched atop an Ariane 5 ES rocket from Kourou on February 16, 2011. It had a total mass of 20,050 kg and carried 7,085 kg cargo, consisting of 4,535 kg propellant for ISS reboost and attitude control, 851 kg ISS refuel propellant, 100 kg oxygen gas, and 1,599 kg dry cargo (food, clothes, equipment). It did not carry any water, as enough water already was on board of the ISS. ATV-2 separated from the Ariane second stage when it reached a 260 km circular orbit at the same inclination as the ISS orbit. At that time the ISS moved in a 351 km altitude orbit. During the next part of the flight, an eight-days phasing period, ATV-2 used the GPS system for navigation, and the ATV-2 mission was controlled from the ATV Control Center in Toulouse, France. Data communication between the control center and ATV-2 followed one of two different paths; either via the NASA TDRSS system or via the ESA Artemis satellite. ATV-2 controllers commanded spacecraft's thrusters into a programmed firing sequence to move the spacecraft to around 35 km behind (-Y-axis) and 4 km below (-X-axis) the ISS, at which point direct space-to-space communication coverage began and the ISS crew could monitor ATV-2's systems. From there, the rendez-vous phase was initiated and the ATV software performed the guidance, navigation and control functions autonomously. The control center performed health checks at pre-defined hold points and authorized the go-ahead for the subsequent automatic approach sequence. During the rendez-vous phase the Zvezda service module transmitted GPS receiver data via a radio-frequency link to ATV-2. The ATV's guidance, navigation and control software calculated the vehicle's position relative to the ISS, based on at least four GPS satellites seen by both the ISS and ATV-2. At a distance of 4.5 km, this relative GPS navigation took over from absolute GPS navigation that was used at a larger distance. The Russian Kurs radio-frequency system installed on the Zvezda module provided independent range and range-rate data. A series of maneuvers brought ATV-2 to the first hold point at 3.5 km behind and 100 m above the ISS. Once authorized to proceed, ATV-2 moved closer to another hold point, at 250 m behind the ISS. At this point, ATV-2 switched over from relative GPS navigation to videometer and telegoniometer navigation. These systems were used up to the docking. From the 250 m hold point, ATV-2 moved in a straight line towards the ISS with another two hold points at 19 m and 11 m. The ISS crew began taking distance measurements on ATV-2 once the vehicle passes inside the 20 m mark. The relative approach velocity was then about 15 cm/s. At each hold point, the vehicle stopped, allowing time for ground controllers and the ISS crew to assess all ATV-2 systems and ATV's behavior before authorizing the next step. For ISS safety reasons, the crew had the capability and authority to intervene in case of problems. At 1 m distance, the ISS crew went 'hands off' and let ATV-2 perform its final docking sequence. During the rendez-vous and docking sequence, the station's solar arrays were placed into a special configuration to protect for GPS multipathing. In total, the rendez-vous took about 3.5 hr and ATV-2 docked to

the Zvezda module on February 24. The station has no window facing out in this direction, but the crew could observe the ATV-2 approaching the ISS via a camera mounted at the end of the Zvezda module. This camera was aligned with a target on ATV-2, providing the crew with necessary visual information. The docking phase started as soon as the ATV's nose probe touched the inside of the docking cone on the Zvezda module. The docking system automated sequence controlled this phase, which included the retraction of the probe to bring the two vehicles close enough to close the hooks, and the connection of fluid, electrical and data interfaces. The docking occurred at a relative speed of about 7 cm/s. ATV-2 met the docking port center point with a precision of 10 cm. As soon as all interfaces were established, the hatch was opened and the crew could enter the pressurized part of ATV-2.

One of the main mission tasks of ATV-2 was to raise the ISS orbit by about 40 km in order to reduce the atmospheric drag experienced by the station. This stepwise maneuver was performed using two of the four 490 N main rocket engines, which consumed nearly 4,500 kg of propellants for the orbit raising. On March 18, 2011, ATV-2's thrusters were used to increase the orbital altitude of the ISS by 3.9 km; the engines burnt for 14 min and 42 s. On April 2, ISS mission controllers used ATV-2's thrusters to boost the ISS into a 0.8 km higher orbit so as to avoid any possibility of intercepting a catalogued piece of space debris. Further orbit altitude boosts were applied in the period May-June 2011. ATV-2 undocked from the station on June 20, 2011. Some hours later, the spacecraft fired its rocket engines for about 10 min to enter an elliptical orbit. Three hours later the engines were fired for a period of about 14 min. Just before hitting the outer layers of the atmosphere, Johannes Kepler was commanded to begin a tumbling to ensure that it would disintegrate in the atmosphere, and would burn up, with its ISS waste, in the atmosphere. Surviving pieces struck the ocean surface in the south Pacific on June 21, 2011, at around 21 hr UTC; about 2500 km east of New Zealand, 6000 km west of Chili and 2500 km south of French Polynesia.

16. LAUNCHING OF A SATELLITE

Satellites are launched from the Earth's surface by means of multi-stage rockets. After ignition of the first-stage rocket engine(s), the launch vehicle ascends vertically from the launch platform. During this portion of the flight relatively large gravity and drag losses (Section 1.7) occur. After some seconds to some tens of seconds, depending on the vehicle, a *pitch-over maneuver* is executed by gimbaling the rocket engine(s) slightly. This will create a torque on the vehicle, turning it so that it enters a curved trajectory. Because the pitch over is carried out while the vertical velocity is still quite small, large aerodynamic loads on the vehicle during the maneuver are avoided. The rocket's angle of attack is adjusted to zero, which reduces lateral aerodynamic loads and produces negligible lift forces during the ascent. This small steering maneuver is the only time during an ideal so-called *gravity turn ascent trajectory*, or its more-modern analogue: a *controlled transition trajectory*, that thrust must be used for purposes of steering. After the pitch over is complete, the engines are reset to point straight down along the axis of the rocket again. The 'direction' of the pitch over determines the launch azimuth and so the flight path azimuth. From a propellant consumption perspective, the optimum ascent trajectory lies in a single plane that is determined by the center of the Earth, the launch site and the flight path azimuth at the start of the rocket's curved trajectory. Ideally, that ascent trajectory plane coincides with the orbital plane selected for the satellite. Then, no plane change maneuvers, which consume a significant amount of propellant, have to be executed. If the rocket is a multi-stage system where stages fire sequentially, the rocket's ascent burn will not be continuous. Obviously, some time must be allowed for stage separation and engine ignition between each successive stage, and many times free-flight periods between successive powered phases, so-called *coasting periods*, are required. When the vehicle has reached an altitude where the aerodynamic forces are negligible, a closed-loop inertial guidance system takes over the control of the vehicle to correct for trajectory deviations and to attain the desired orbit. The final stage rocket engine(s) cut off when the vehicle has a predetermined velocity and position.

The determination of the optimum ascent trajectory of a particular launch vehicle for a prescribed satellite orbit is very complex and beyond the scope of this book. In this Chapter, we will discuss some general aspects of ascent trajectories. Launch vehicle performance data for orbital missions will be presented in graphical form. A more-detailed discussion will be given on the relations between the position and velocity of a satellite at the moment that it is injected into its orbit and the orbital parameters, and on the orbit deviations due to injection errors.

16.1. Launch vehicle ascent trajectories

Two basic types of satellite launch vehicle ascent trajectories can be distinguished (Figure 16.1): *Direct Ascent* (DA) and *Hohmann Transfer Ascent* (HTA). The DA trajectory is selected such that its summit point coincides with the required satellite orbit. When the satellite, which in most cases is still attached to the final rocket stage, approaches this summit altitude, the final-stage engine is ignited and the satellite is accelerated to the required velocity. In some cases, the rocket engine required for this acceleration is not the engine of the last stage, but is an integral part of the satellite itself (*apogee engine*). In the HTA, the satellite and the last stage first attain a low-altitude circular parking orbit, just outside the densest part of the atmosphere. Usually, the altitude of this parking orbit is about 200 km. For preliminary mission and performance analyses, mostly a parking orbit altitude of 185 km (100 n.mi.) is assumed. Either shortly after parking orbit injection, or after a coasting period in this parking orbit, the vehicle is injected into a Hohmann elliptical transfer orbit. At perigee, this transfer ellipse is (about) tangential to the

parking orbit and at apogee (about) tangential to the required satellite orbit (Section 12.1). When the satellite is at apogee of the transfer orbit, the last-stage rocket engine, or the satellite's apogee engine, is fired to accelerate the satellite into its final orbit.

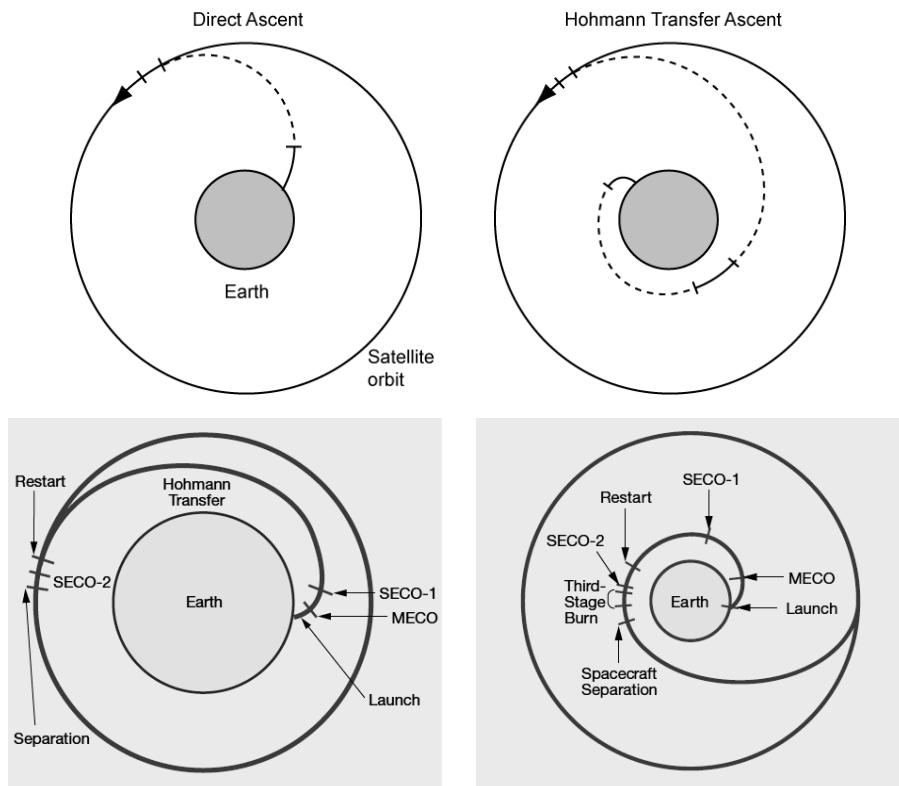


Figure 16.1: Sketch of a Direct Ascent (DA) and a Hohmann Transfer Ascent (HTA) (top), and of a typical HTA two-stage (bottom left) and three-stage (bottom right) mission profile. MECO = Main Engine Cut Off; SECO = Second stage Engine Cut Off.

Comparing these two types of ascent trajectories, it can be shown that, generally, an HTA requires less propellant for ascent to a specified orbit. The main reason is that the ascent trajectory for a DA is steeper than for an HTA, yielding larger gravity losses (Section 1.7). The HTA would even be the absolute minimum-energy trajectory, if the altitude of the parking orbit would be very low, and if the rocket would be launched horizontally. However, as a result of the presence of the atmosphere and launch site constraints, this trajectory, of course, cannot be flown and a low-altitude parking orbit has to be used. Only for satellite orbits below 300 km altitude does a DA sometimes require less propellant. However, it should be realized that apart from propellant consumption, many operational aspects should be taken into account in the selection of an ascent trajectory. Obviously, the shorter coasting periods in a DA may be advantageous for stages with cryogenic propellants or for military missions. Problems of attitude control and other aspects of vehicle engineering are mostly reduced too when using a DA. In addition, due to the comparatively steeper DA, the launch vehicle can be tracked longer from the launch site. For the HTA, a world-wide network of tracking stations is necessary. From the viewpoint of booster recovery, the shorter horizontal distance traveled by the booster during a DA is advantageous. Finally, it will be clear, that, in general, the HTA requires the application of restartable rocket engines. For launch vehicles using only solid propellants, this restart capability is difficult to provide.

As examples of ascent trajectories, Figure 16.2 shows typical Delta II 7320/7420 (top) and

Atlas V 552 (bottom) ascent profiles for injecting a satellite into a low Earth orbit, and the history of some relevant parameters during the ascent flights.

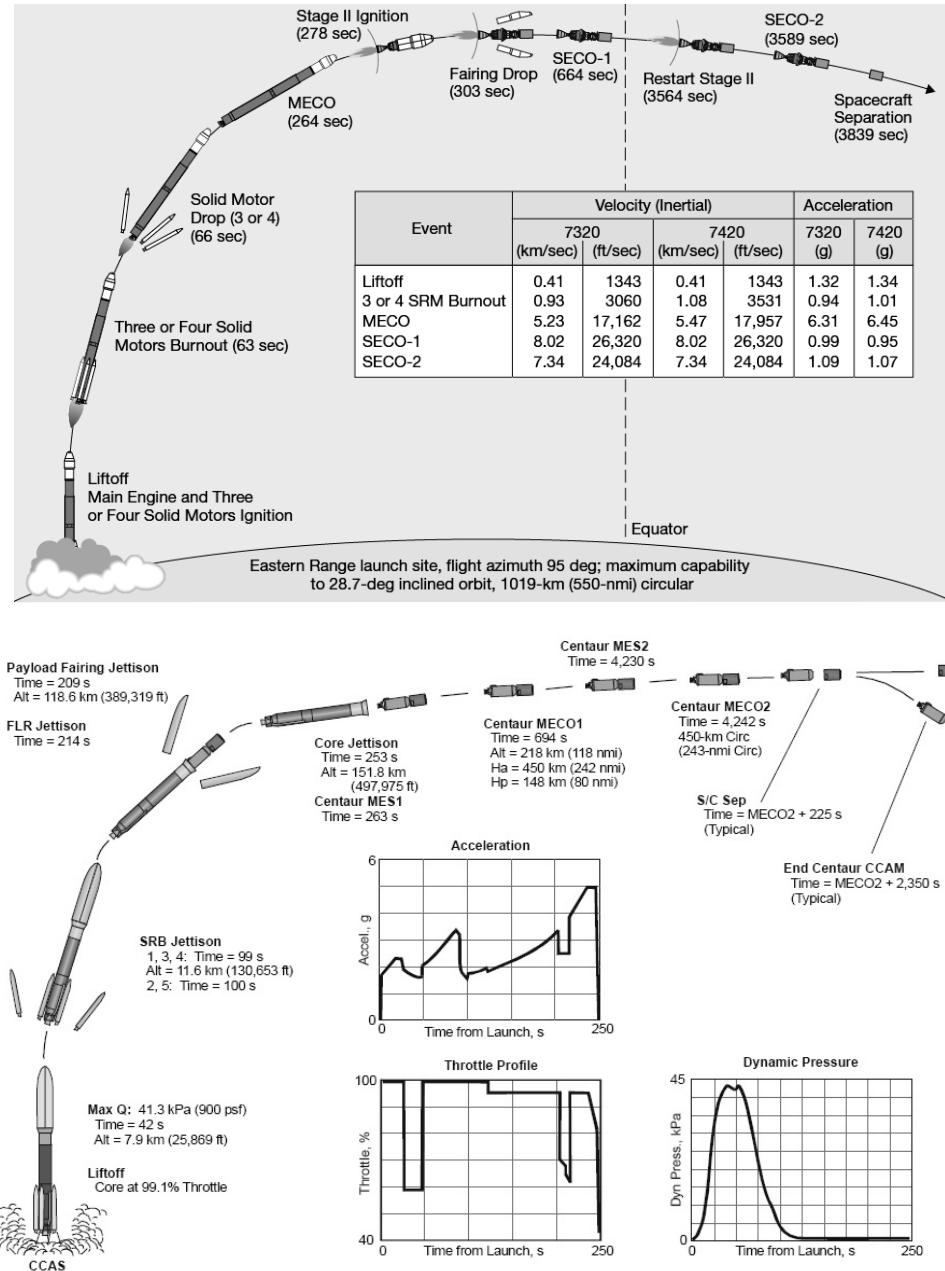


Figure 16.2: Typical Delta II 7320/7420 (top) and Atlas V 552 (bottom) ascent profiles for injecting a satellite into a low Earth orbit.

16.2. Orbit injection and injection constraints

Whatever ascent trajectory is flown, the satellite, and often also the last stage, will ultimately reach the position and velocity required to enter a prescribed orbit about the Earth. Usually, then the satellite is separated from the last stage, e.g. by releasing a compressed-spring mechanism, and small rocket thrusters on the last stage are fired to move that stage away from the satellite. For liquid-propellant stages, usually the propellant residuals are also exhausted, without combustion, through the main engine(s), in order to increase the relative velocity between rocket

stage and satellite. The time of separation is called the *injection epoch*; the position and velocity of the satellite at that epoch constitute the *injection parameters*. Injection epoch and injection parameters together completely determine the orbit in which the satellite will move about the Earth.

We now assume that at injection epoch, t_0 , the injection parameters distance, r_0 , declination, δ_0 , right ascension, α_0 , velocity, V_0 , flight path angle, γ_0 , and flight path azimuth, ψ_0 (Chapter 11) are known. The orbit that the satellite will follow can then be computed by using the relations between spherical coordinates and orbital elements, which were derived in Section 11.9 for elliptical orbits,

$$\alpha = \frac{r_0}{2 - \frac{r_0 V_0^2}{\mu}} \quad (16.1)$$

$$e^2 = 1 - \frac{r_0 V_0^2}{\mu} \left(2 - \frac{r_0 V_0^2}{\mu} \right) \cos^2 \gamma_0 \quad (16.2)$$

$$E_0 = \arctan \left[\sqrt{\frac{\alpha}{\mu}} \frac{r_0 V_0 \sin \gamma_0}{\alpha - r_0} \right] \quad (16.3-1)$$

$$e \sin E_0 = \frac{r_0 V_0 \sin \gamma_0}{\sqrt{\mu \alpha}} \quad (16.3-2)$$

where the two expressions (16.3) allow an unambiguous determination of E_0 , and

$$\tan \frac{\theta_0}{2} = \sqrt{\frac{1+e}{1-e}} \tan \frac{E_0}{2} \quad (16.4)$$

where $\theta_0/2$ and $E_0/2$ are in the same quadrant, and

$$\cos i = \cos \delta_0 \sin \psi_0 \quad ; \quad 0 \leq i < \pi \quad (16.5)$$

$$\sin(\omega + \theta_0) = \sin \delta_0 / \sin i \quad ; \quad \cos(\omega + \theta_0) = \cos \delta_0 \cos \psi_0 / \sin i \quad (16.6)$$

$$\sin(\alpha_0 - \Omega) = \tan \delta_0 \cot i \quad ; \quad \cos(\alpha_0 - \Omega) = \cos \psi_0 / \sin i \quad (16.7)$$

where the two equations (16.6) and the two equations (16.7) allow an unambiguous determination of ω and Ω , and

$$\tau = t_0 - \sqrt{\frac{\alpha^3}{\mu}} (E_0 - e \sin E_0) \quad (16.8)$$

We may write (16.1) to (16.8) in symbolic form as

$$\begin{aligned} \alpha &= \mathcal{F}_1(r_0, V_0) \quad ; \quad e = \mathcal{F}_2(r_0, V_0, \gamma_0) \quad ; \quad i = \mathcal{F}_3(\delta_0, \psi_0) \\ \omega &= \mathcal{F}_4(r_0, V_0, \gamma_0, \delta_0, \psi_0) \quad ; \quad \Omega = \mathcal{F}_5(\alpha_0, \delta_0, \psi_0) \quad ; \quad \tau = \mathcal{F}_6(t_0, r_0, V_0, \gamma_0) \end{aligned} \quad (16.9)$$

where \mathcal{F}_j is a known function for each j . For the argument of latitude, u_0 , true anomaly, θ_0 , and eccentric anomaly, E_0 , we have

$$u_0 = \omega + \theta_0 = \mathcal{F}_7(\delta_0, \psi_0) ; \quad \theta_0 = \mathcal{F}_8(r_0, V_0, \gamma_0) ; \quad E_0 = \mathcal{F}_9(r_0, V_0, \gamma_0) \quad (16.10)$$

These relations clearly show that the orbital elements and orbital parameters can be separated into three groups:

- The orbital elements a , e and τ , and the orbital parameters θ_0 and E_0 , which are a function of the *in-plane injection parameters* r_0 , V_0 , γ_0 only.
- The orbital elements i and Ω , and the orbital parameter u_0 , which are a function of the *out-of-plane injection parameters* α_0 , δ_0 , ψ_0 only.
- The orbital element ω , which is a function of the in-plane as well as of the out-of-plane injection parameters.

From (16.5) a number of interesting conclusions can be drawn:

- For a given declination of the injection point, we obtain

$$-\cos\delta_0 \leq \cos i \leq \cos\delta_0 \quad \text{or, since } 0 \leq i < \pi : \quad |\delta_0| \leq i \leq \pi - |\delta_0|$$

For example, if $\delta_0 = 20^\circ$ then $20^\circ \leq i \leq 160^\circ$. So, we conclude that, for a specified launch site, the inclination of the orbit cannot be chosen arbitrarily.

- For a given injection point, the minimum inclination ($i = |\delta_0|$) is reached when $\psi_0 = 90^\circ$; i.e. an injection due east. The maximum inclination ($i = \pi - |\delta_0|$) is reached for an injection due west: $\psi_0 = 270^\circ$.

– If a satellite has to enter an equatorial orbit ($i = 0^\circ$), we must require $\delta_0 = 0^\circ$, $\psi_0 = 90^\circ$. If the required inclination of the orbit is within one of the ranges $0 < i < |\delta_0|$ or $\pi - |\delta_0| < i < \pi$, the launch site will never cross the orbital plane and, consequently, the plane of the ascent trajectory cannot coincide with the orbital plane. In these cases, a plane change has to be executed during the ascent trajectory, at orbit injection or at a certain position along the orbit. In Section 13.3 it was shown that for a satellite that is already in orbit an inclination change maneuver should be executed at the satellite's ascending or descending node. These maneuvers are performed at relatively high velocities and are very expensive in terms of propellant consumption. For launching satellites into low-inclination orbits, therefore an equatorial launch site is very attractive. Two near-equatorial sites exist: the Italian San Marco platform in the Indian Ocean off the coast of Kenya, and the French/European site Kourou in French Guiana (Table 16.1). In 1988 the last rocket was launched from the San Marco platform. Presently, it is only used for satellite communications. Ariane rockets are launched on a routine basis from Kourou. The first Ariane 1 rocket was launched in December 1979; in February 2011 the 200th flight of an Ariane family vehicle took place. In October 2011 the first Soyuz-2 rocket was launched from Kourou; in February 2012 the first Vega rocket was launched from this site. The three US launch sites, located north of the equator, are Western Test Range (Vandenberg Air Force Base) in California, Eastern Test Range (Cape Canaveral) in Florida and Wallops Island in Virginia. The western and eastern test ranges are used on a routine basis for a large variety of satellite and planetary missions. Wallops Island is in operation since 1945 and from this range smaller orbital-class rockets are launched. Another US launch site is the Sea Launch platform, a self-propelled converted oil-drilling platform that is loaded with a three-stage Zenit launcher and a satellite in a port at Long Beach, California. The platform and its accompanying Assembly and Command Ship then sail to the launch location at 154° W on the equator, south of the Hawaiian Islands. From this platform launches to all inclinations can be executed. In the period 1999 to 2009 thirty rockets have been launched. In 2009, the provider of the launch services ran

into financial problems and a Russian corporation is presently the majority owner of the reorganized entity. The launching of satellites from this platform was resumed in September 2011. The three Russian launch sites listed in Table 16.1, Plesetsk, Kapustin Yar and Baikonur, are located at fairly high latitudes of over 45° N; the Indian launch site Sriharikota at a fairly low latitude of about 14° N; the Chinese launch sites Shuang and Xichang at latitudes of about 40° N and 28° N; the Japanese launch site Kagoshima (also called Uchinoura) at about 31° N.

Table 16.1: Permitted range of launch azimuth and corresponding orbital inclination for various launch sites. [coordinates and launch azimuth constraints were taken from D.A. Vallado, Fundamentals of Astrodynamics and Applications, Kluwer Academic Publishers, 2001.]

Site	Location (°)	Azimuth (°)	Inclination (°)
Cape Canaveral (ETR)	28.5 N, 80.6 W	37 - 112	28 - 58
Vandenberg (WTR)	34.6 N, 120.6 W	147 - 201	63 - 107
Wallop Island	37.8 N, 75.5 W	30 - 125	38 - 67
San Marco Platform	2.9 S, 40.2 E	50 - 150	3 - 60
Kourou	5.2 N, 52.8 W	-20 - 100	5 - 110
Plesetsk	62.8 N, 40.6 E	-30 - 90	63 - 103
Kapustin Yar	48.4 N, 45.8 E	-10 - 90	48 - 97
Baikonur	45.6 N, 63.4 E	-20 - 90	46 - 104
Sriharikota	13.7 N, 80.2 E	100 - 290	17 - 166
Shuang	40.4 N, 99.8 E	-10 - 120	40 - 98
Xichang	28.2 N, 102.2 E	94 - 105	28 - 32
Kagoshima	31.2 N, 131.1 E	20 - 150	31 - 73

For all launch sites, the allowed launch azimuth is restricted by safety aspects and the availability of tracking stations. For example, for a launch from the Eastern Test Range (ETR) the launch azimuth, generally, must be within the range 37° - 112° ; for a launch from the Western Test Range (WTR) a launch azimuth between 147° and 201° is generally acceptable. Therefore, the Eastern Test Range is used primarily for launches into a more or less eastward direction, which employ the rotation of the Earth to increase effectively the inertial velocity of the vehicle, thereby improving the launch vehicle's payload capability. In Table 16.1 the permitted range of launch azimuths and the range of orbital inclinations are listed for the various launch sites. The Table shows that polar orbits can be reached when launching from Vandenberg, Kourou, Plesetsk, Kapustin Yar, Baikonur, Sriharikota, and Shuang. For Vandenberg and Sriharikota, the launch must then be in southern direction; for Kourou, Plesetsk, Kapustin Yar, Baikonur and Shuang, in northern direction.

Sometimes, the required orbital inclination is, for a given launch site, outside the range of inclinations that can be achieved for a launch azimuth within the launch safety constraints. In that case, a so-called *dogleg maneuver* is executed, where the launch vehicle first flies a trajectory within the permitted range of flight azimuths. During its climb to orbital altitude an out-of-plane maneuver is executed with the rocket's main propulsion system, and sometimes with additional vernier rocket engines, to change the plane of the ascent trajectory such that the required inclination can be reached. In some cases two dogleg maneuvers are executed; one when the first stage is operating and another when the second stage is operating. Usually, a dogleg maneuver lasts for 30 - 60 s. Of course, a dogleg maneuver can only be executed if the range of permitted flight azimuths during that part of the flight is wider than at launch. It is emphasized that such

a maneuver costs a relative large amount of propellant and can therefore only be applied for relatively small inclination changes. On the other hand, since it is executed at a time when the velocity is significantly lower than when the vehicle has reached orbital altitude, this maneuver may be preferable over an inclination change maneuver when the spacecraft enters orbit or is already in orbit. This last type of plane change maneuver, which is often also called a dogleg maneuver, is discussed in Chapter 13.

If $|\delta_0| < i < \pi - |\delta_0|$, then the launch site will cross the selected orbital plane twice a day. Because of the constraints on launch azimuth, generally, only one of these crossings can be used to launch a satellite into its orbit along a coplanar ascent trajectory. If small plane change maneuvers are accepted, the satellite can be launched within a short period around the orbital plane crossing time, constituting the so-called *daily launch window*. We now meet another advantage of using a parking orbit. For a given launch time and a DA trajectory that is prescribed by the launch vehicle capabilities, the point where the satellite enters its orbit is completely determined. Often, such as, for instance, in rendez-vous missions or for applications satellites, the satellite has to enter its orbit at a specified location. Then, a coplanar launch within the launch window is mostly impossible. The best alternative to long periods of inaccessibility or expensive non-planar transfers is using a parking orbit in combination with an HTA. By carefully selecting the point of departure from this parking orbit, it is possible to reach any point in the final satellite orbit. If the satellite has to reach a specific location in the final orbit at a specified time, then, generally, an intermediate phasing orbit is required (Chapter 14).

16.3. Launching of the Space Shuttle

To give an impression of the complexity of launching a spacecraft and of the processes that take place during launch, in this Section a description will be given of the mission plan for the first launch of the Space Shuttle Columbia on April 12, 1981, from Cape Canaveral (Space Transportation System flight STS-1). During this flight the Shuttle was injected into a 40° inclination orbit at an altitude of 240 km.

At 3.5 s before launch time, which is indicated in this Section by T , the three main engines of the Orbiter are ignited within 120 ms of each other. Since the resulting thrust vector does not pass through the points on which the Shuttle rests on the launch platform, after ignition of the main engines a rotation about the center of mass occurs. As a result, the cockpit that houses the astronauts moves about 50 cm. At $T+0.24$ s the engines should be at 90% of their nominal thrust level. If this is the case, at $T+3$ s the two *Solid Rocket Booster* (SRB) engines are ignited. The five rocket engines together produce a thrust of 31 MN. With a total mass of the Shuttle of about 2040 ton, this results in an initial upward acceleration of the Shuttle of about 5 m/s^2 , or about half the gravitational acceleration on Earth. Immediately after lift-off, a maneuver is performed to make the thrust vector to pass through the center of mass of the vehicle. This maneuver and the entire control during launch are performed by pivoting the nozzles of the three main engines as well as of the two solid rockets. The acoustic load during launch is considerable. Without precautions, the sound level during launch would be more than 168 dB in the cargo bay of the Shuttle. This sound level would lead to an acoustic load on the payload that is too high. Therefore, special precautions are taken. The cooling system that protects the bottom of the launch tower against excessive heating is especially designed to also reduce the sound level. For this purpose, more than 3.4 million liters of water per minute are poured out over the base of the launch platform during the first phase of the launch; this reduces the sound level by 8 dB.

At $T+8$ s the Shuttle has gained an altitude where the nozzles of the boosters are about 10 m above the top of the launch tower. At that moment, a maneuver is performed (Figure 16.3) that

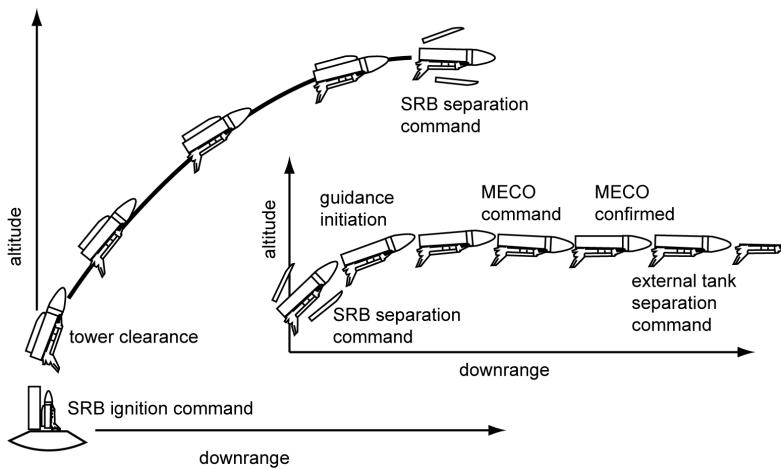


Figure 16.3: Characteristic Space Shuttle ascent trajectory.

is aimed to change the vertical trajectory of the vehicle into a curved trajectory in a plane that is determined by the desired inclination of the final orbit around the Earth. During the first flight the launch azimuth is 60.2° , which results in an inclination of 40.3° . At $T+20$ s this maneuver is completed; the vehicle has performed a rotation about its longitudinal axis and the longitudinal axis is at an angle of 78° with the horizon. The Orbiter now flies upside down. This orientation guarantees, together with the selected ascent trajectory, that the aerodynamic loads on the entire launch vehicle do not become excessive. The aerodynamic control surfaces located on the Orbiter's wings and horizontal stabilizer are constantly kept in such a position that the loads on the connecting elements between the Orbiter, the boosters and the external tank do not become too large and that these elements are kept under tensile loading. At $T+32$ s the Shuttle reaches an altitude of 2.4 km and a velocity of 740 km/hr. The thrust of the three main engines is then reduced by 10% per second until 65% of their initial value to limit the aerodynamic forces that act on the launch vehicle when it breaks the sound barrier (Mach number (M) = 1) and when the maximum so-called dynamic pressure is reached. At $T+52$ s, at an altitude of 7.3 km, the local speed of sound is exceeded. The thrust program of the solid rocket boosters is designed in such a way that about 55 s after launch the thrust of each engine decreases from 12.9 MN to 9.5 MN, to limit the loads on the launch vehicle when the maximum dynamic pressure is reached. At $T+69$ s the period of maximum dynamic pressure has passed and the thrust of the main engines of the Orbiter is returned to its initial level. At an altitude of 36.6 km a velocity of $M = 3$ is reached. From that moment on the crew can no longer use the ejection seats in case of an emergency. At $T+132$ s, when the Shuttle has reached an altitude of 50 km and a velocity of $M = 4.5$, the empty boosters are separated from the launch vehicle. Small auxiliary thrusters are used to move the boosters away from the Orbiter and the boosters fall in a ballistic trajectory back to Earth. At an altitude of about 5 km parachutes are deployed, after which the boosters fall into the Atlantic Ocean at about 275 km from the launch site. There they are picked up by two specially designed ships and taken back to Cape Canaveral for refurbishment.

Until booster separation, the flight is performed using *open-loop control*. This means that the roll, pitch and yaw commands are programmed as a function of velocity, and that the control system only executes these programmed commands. In the computer the control programs for both a nominal flight and the flight in case of an engine failure are stored. If this happens, the computer automatically switches over to the appropriate control program. Four seconds after booster separation, control is switched to *closed-loop control*. That means that the computer calculates the trajectory that the vehicle must follow, starting from the known conditions when

closed-loop control is initialized, to reach the required condition at the time of *Main Engine Cut-Off* (MECO) with minimum propellant consumption. These MECO conditions are stored in the computer memory before the flight in the form of four parameters: velocity, flight path angle, distance from the center of the Earth, and inclination. These MECO conditions are chosen in such a way that an appropriate point of impact of the external tank can be guaranteed. In the first phase of the flight under closed-loop control, a rather steep trajectory is flown to let the vehicle gain altitude fast. This is necessary to reach a situation that, in case of an engine failure, an emergency landing can be performed after one orbit about the Earth. At an altitude of 116 km and a velocity of $M = 6.4$, approximately 4 min after launch, this situation is achieved. Then, the second phase of the closed-loop control is started. Under computer control the vehicle now switches to a trajectory that requires a minimum amount of propellant for three working engines. For this purpose, the attitude of the Orbiter is changed such that the nose, which first pointed 19° above the horizon, now points 4° below the horizon. In this phase, the altitude first still increases to 130 km, after which the Orbiter enters a slow dive in which the altitude decreases to 117 km, while the velocity increases more and more. At $T+390$ s, the Orbiter reaches an altitude of 130 km and a velocity of $M = 15$. From that moment on, even if two main engines would fail, a successful reentry into the atmosphere and a landing on a specially reserved runway would be possible. After $T+420$ s it is even possible to reach the desired orbit if two main engines fail. At this stage of the flight, the acceleration of the Orbiter and the external tank increases to 3 g , i.e. three times the acceleration due to gravity on Earth. To prevent higher accelerations, the third phase of the closed-loop control is started. During this phase the thrust is continuously decreased such that the acceleration does not exceed 3 g . If the Orbiter is carrying much cargo, or if the three main engines do not reach their nominal performance, it may be possible that this third phase is never entered. If during this period an engine fails, the computer immediately switches back to the second control phase.

At 7 s before MECO, which occurs 8.5 min after launch, the thrust of the three main engines is reduced to 65% of their initial thrust. About 6.5 s after that, when the Orbiter has reached a velocity of $M = 25$, the three main engines are stopped. The Orbiter and the external tank now have a velocity of 7.82 km/s, a radial velocity of 67 m/s and have entered an Earth orbit with perigee at 24 km and apogee at 150 km altitude. Approximately 20 s later, the external tank separates. The thrusters of the *Reaction Control System* (RCS) fire for 24 s to move the Orbiter away from the tank with a relative velocity of 1.2 m/s. The tank now falls in a ballistic trajectory back to Earth. At an altitude of about 80 km it will break up into pieces and partially burn up, after which the debris will fall into the Indian Ocean.

Two minutes after MECO, the two engines of the *Orbital Maneuvering System* (OMS) ignite; this OMS is located at the back of the Orbiter above the three main engines. These engines fire for 87 s, as a result of which the Orbiter is injected into an orbit with perigee at 105 km and apogee at 240 km altitude. If one of the OMS engines would fail, the engines of the RCS can also be used to perform this maneuver. During this powered phase, also the 2.5 ton of liquid hydrogen and oxygen that is still in the pipelines to the main engines, is dumped. The liquid oxygen flows away through a nozzle, the liquid hydrogen is removed to space via special valves. When the Orbiter passes the apogee of its orbit about half an hour later, the two OMS engines are ignited for the second time, now for 75 s, to circularize the orbit of the Orbiter at an altitude of 240 km. This completes the launch 45 minutes after leaving Cape Canaveral.

16.4. Performance of launch vehicles

To select a nominal orbit for a specific satellite mission, the performances of available launch

vehicles have to be known. After all, one has to know which rocket can launch the satellite into the desired orbit. For first-order analyses, a method based on the concept of the *mission characteristic velocity*, V_{char}^* , is generally used. This quantity is defined as the sum of all velocity increments, which are needed to inject a payload from a 100 n.mi. (185 km) altitude, 28.5° inclination parking orbit into the desired orbit; the velocity of the satellite in its parking orbit ($V_c = 7.80$ km/s) has to be added. The choice of the reference inclination of 28.5° refers to a launch due east from Cape Canaveral (Eastern Test Range). Figure 16.4 illustrates the concept

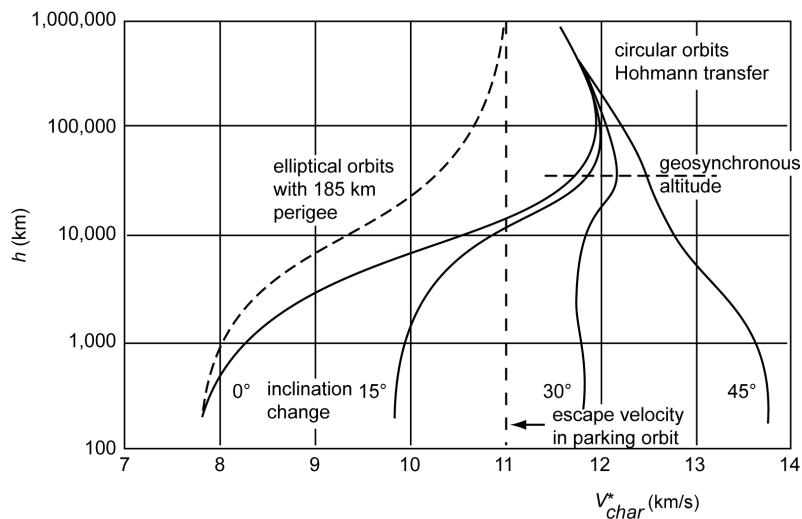


Figure 16.4: Mission characteristic velocity for various final orbits about the Earth.

of mission characteristic velocity. The dashed curve indicates the sum of the initial velocity of the satellite in its parking orbit and the velocity increase ΔV_1 required in the parking orbit to enter a coplanar Hohmann transfer orbit with apogee altitude equal to a desired final orbit altitude (Chapter 12). To circularize the orbit at apogee, a velocity increase ΔV_2 has to be applied. The sum of ΔV_1 and ΔV_2 is also indicated in Figure 16.4 (solid curve for 0° inclination change). The particular shape of this curve is discussed in Section 12.1. The mission characteristic velocity for a three-dimensional Hohmann transfer (Chapter 13) to a final circular orbit at a particular altitude is also indicated for three values of the inclination change. For this three-dimensional transfer, it is assumed that when the vehicle crosses the equatorial plane it is injected into a coplanar Hohmann transfer trajectory. At apogee of this trajectory, which also coincides with an equatorial crossing, an impulsive shot is applied such that the required plane change is effectuated and the orbit is circularized simultaneously. This mission profile is discussed in Section 13.4. From Figure 16.4 we conclude that for most missions the mission characteristic velocity, V_{char}^* , falls in the range of 9–13 km/s.

The performance of a launch vehicle can also be expressed in a type of characteristic velocity. This *launch vehicle characteristic velocity*, V_{char} , is defined as the total velocity to which a payload can be accelerated by a particular launch vehicle from an altitude of 100 n.mi. after a launch due east from Cape Canaveral, plus the velocity of the vehicle in its parking orbit ($V_c = 7.80$ km/s). Of course, for a given mission always a launch vehicle has to be selected that can provide a launch vehicle characteristic velocity larger than the mission characteristic velocity. As an example, Figure 16.5 shows the payload capability, M_l , as a function of the launch vehicle characteristic velocity for a series of launchers. Although many of the launchers included are historical and not available anymore, and the Saturn 5/Centaur has only existed on the drawing table, this Figure is valuable for understanding the character of the $M_l - V_{char}$ curves. The curves

for launchers that are optimized for launching payloads into low-Earth orbits fall off at relatively low values of V_{char} , while the curves for launchers with *kick stages*, which are optimized for high-energy missions (e.g. planetary missions), show significant payload capabilities at higher values of V_{char} . When the launch azimuth is not equal to 90° , and thus the initial parking orbit has an inclination larger than 28.5° , or when the spacecraft is launched from another launch site, a V_{char} correction has to be applied. Figure 16.6 shows this ΔV_{char} penalty for a launch from Cape Canaveral (Eastern Test Range) but using a parking orbit with an inclination larger than 28.5° (left), and for a launch from Vandenberg (Western Test Range) and using a parking orbit with

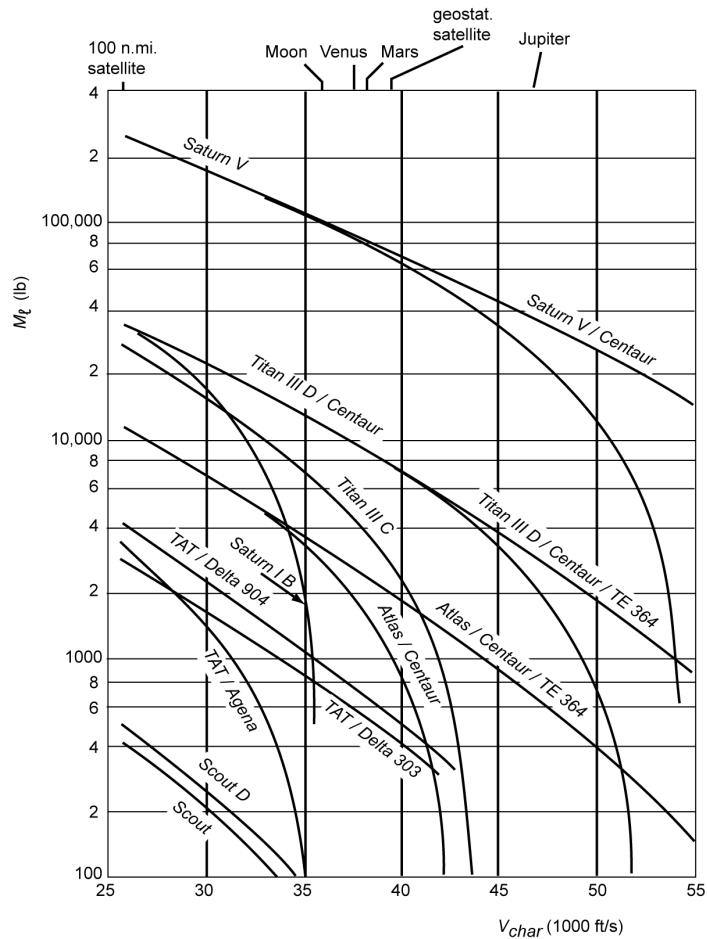


Figure 16.5: Launch vehicle payload capability as a function of the launch vehicle characteristic velocity.

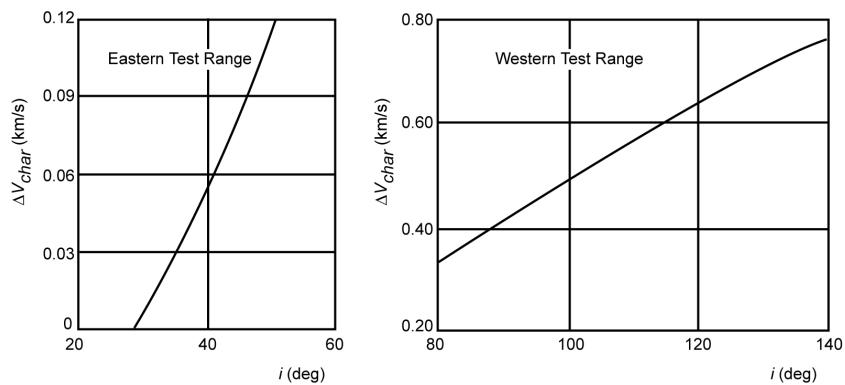


Figure 16.6: Launch site and launch azimuth characteristic velocity penalties.

an inclination between 80° and 140° (right). This corrected value of V_{char} should be used when the payload capability of a specific launch vehicle is read from Figure 16.5. In this Figure typical values of the required mission characteristic velocity for six missions are indicated. These are a satellite in a circular orbit at an altitude of 185 km: $V_{char}^* = 7.8$ km/s; a geostationary satellite: $V_{char}^* = 11.7$ km/s; and a flyby mission to the Moon, Venus, Mars and Jupiter (Chapter 18). For a favorable position of the planets and for a minimum-energy direct trajectory (Chapters 17 and 18) to the Moon, Venus, Mars and Jupiter, the corresponding V_{char}^* -values are about 10.9 km/s, 11.3 km/s, 11.4 km/s and 14.1 km/s, respectively. From this Figure we conclude that a Saturn 1B rocket can deliver a payload of approximately 15 ton into a low satellite orbit and that the payload capacity of the Atlas/Centaur and Scout D rockets is for the same orbit approximately 5200 kg and 225 kg, respectively. A geostationary mission can neither be executed by the Scout nor by the large Saturn 1B rocket; the Atlas/Centaur can deliver about 500 kg of payload into the

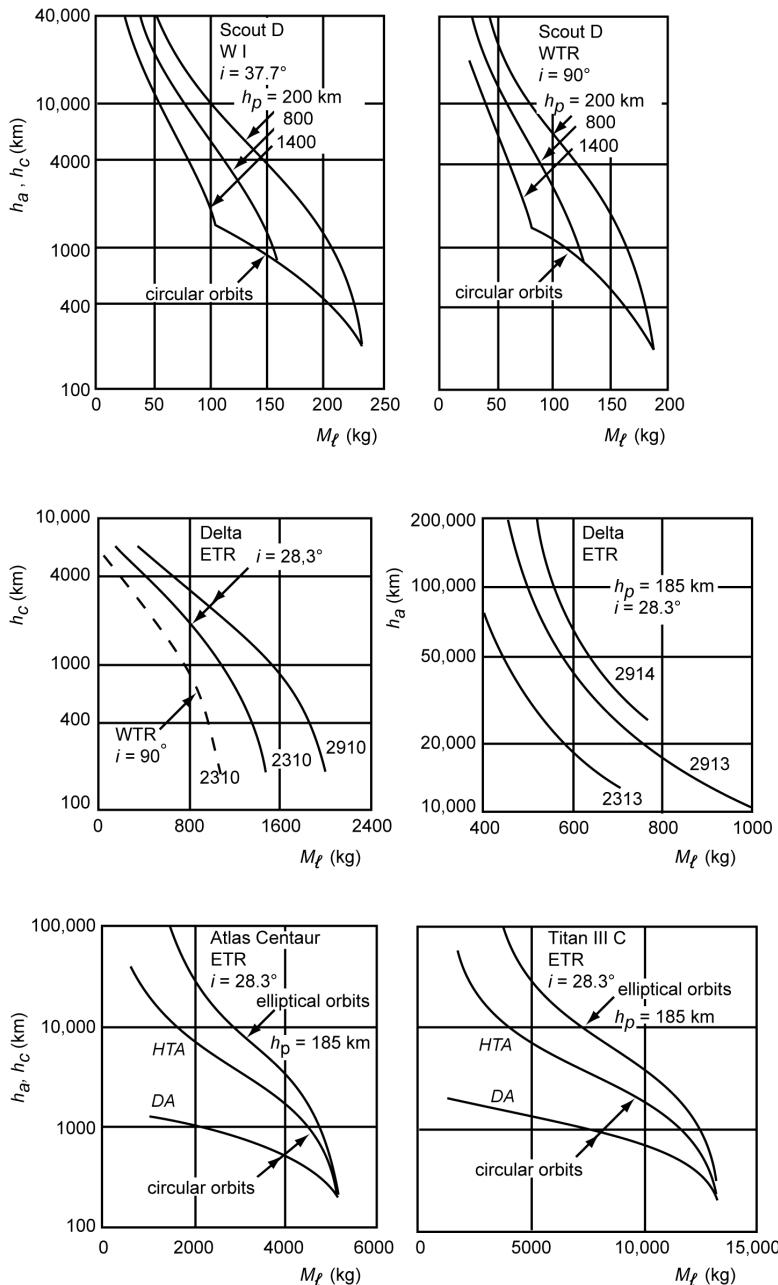


Figure 16.7: Earth-orbital performance of some satellite launch vehicles.

geostationary orbit. A mission to Jupiter can only be flown by the Saturn 5, Titan 3D/Centaur and Atlas/Centaur/TE364 launchers.

Most launch vehicles are constrained with regard to the number of times a rocket engine can be restarted, the permitted time between two thrusting phases (coasting period), etc. As a result, it may not be possible to accelerate the payload at the desired times. For a more precise mission planning, therefore a different type of performance diagram has to be used. Such a diagram gives the launch vehicle payload capability, M_p , as a function of perigee and apogee altitudes for a given launch site and specified orbital inclination. Figure 16.7 presents examples of such diagrams for Scout D, Delta, Atlas/Centaur and Titan rockets. A comparison of the performance of the Scout for a launch from Wallops Island (WI) or from the Western Test Range (WTR) shows the benefit of launching in an easterly direction. In that case, the Earth's rotational velocity at the Earth's surface reduces the ΔV that the launch vehicle has to deliver for a certain mission. This results in a significantly larger launch vehicle payload capability for that mission. Since the rotational velocity at the Earth's surface is largest at the equator (≈ 0.46 km/s), it is clear that a launch site on or near the equator, from which satellites can be launched in an easterly direction, is very attractive.

Figure 16.8 presents the payload capability of a few more-modern launch vehicles as a

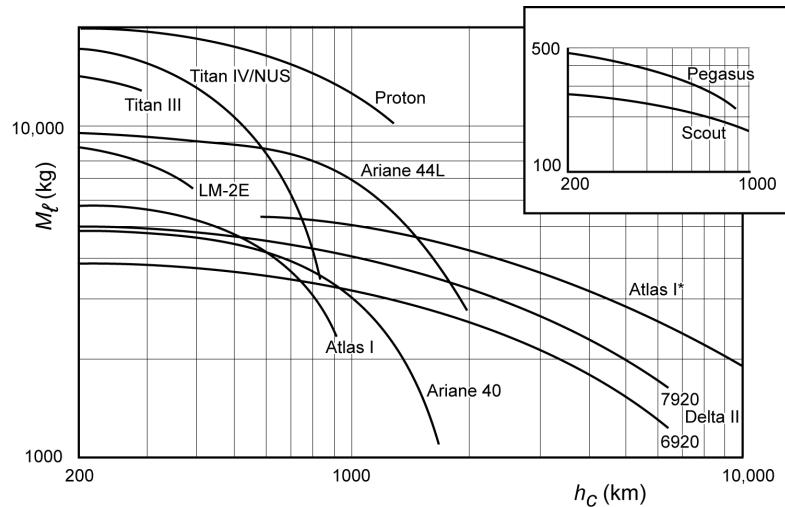


Figure 16.8: Launch vehicle performance for launches due east into a circular orbit. [copied from W.J. Larson and J.R. Wertz (eds.), *Space Mission Analysis and Design*, second edition, Kluwer Academic Publishers, 1997.]

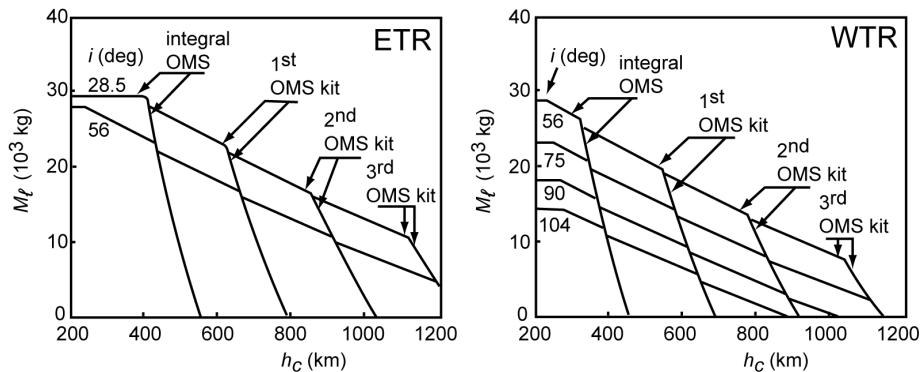


Figure 16.9: Space Shuttle payload capability for delivery missions from ETR and WTR.

function of the altitude of a circular orbit in which the payload is injected after a launch due east from Cape Canaveral (US launchers), Kourou (Ariane), Baikonur (Proton) and Xichang (Long March (LM)). The modern heavy-lift launchers Delta 4050H-19 and Atlas V 552 (not included in Figure 16.8) can inject payloads of about 22,500 kg and about 18,100 kg, respectively, into a 800 km altitude circular orbit after a launch due east from Cape Canaveral; for an orbital altitude of 2000 km, the corresponding values are about 19,300 kg and 14,600 kg, respectively. These launchers can also inject payloads of about 10,700 kg and 8,980 kg, respectively, into a standard geostationary transfer orbit. Nowadays, the launchers Ariane 5, Soyuz-2 and Vega (also not included in Figure 16.8) are operational at Kourou. From this launch site, Ariane 5 ECA can inject payloads of 17,000 kg into a Sun-synchronous (Section 23.7) 800 km altitude orbit, and a payload of 10,500 kg into a standard geosynchronous transfer orbit; for Soyuz 2-1b the corresponding values are 4750 kg and 3060 kg, respectively. Vega can inject a payload of 1350 kg into a Sun-synchronous 800 km altitude orbit. Figure 16.9 gives an impression of the payload capability of the (former) Space Shuttle for circular orbits. Note that this vehicle could deliver about 29,000 kg of payload into a low orbit with an inclination of 28.5° when launched from Cape Canaveral (ETR). When launched from Vandenberg (WTR), this vehicle could deliver only

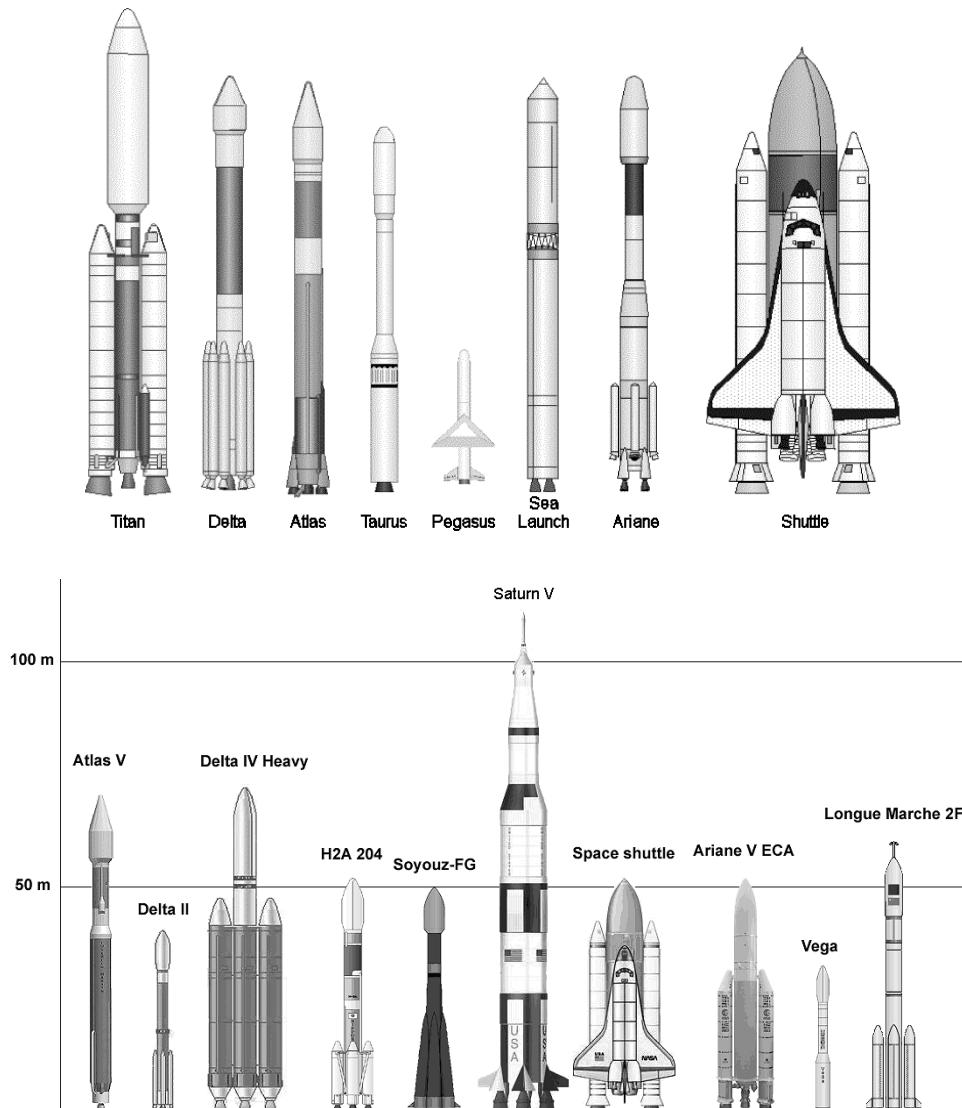


Figure 16.10: Sketch of some launch vehicles.

about 18,000 kg of payload into a low polar orbit. The Figure clearly shows that the Space Shuttle was designed for missions to relatively low Earth orbits; the payload capacity becomes zero for an orbital altitude of approximately 1200 km. When performance diagrams such as the ones presented in Figures 16.7 to 16.9 are used, it should be realized that operational constraints imposed by the launch vehicle are already taken into account. This means that the performance curves should never be extrapolated outside the areas indicated in the diagrams.

Figure 16.10 presents sketches of some launch vehicles, showing the relative size of each vehicle. For reference: a Solid Rocket Booster (SRB) of the Space Shuttle is 45.5 m long.

16.5. Effects of variations in the in-plane injection parameters

The effects of variations in the injection parameters on the satellite's orbit can be investigated with (16.1) to (16.8). In practice, perigee and apogee distances, true anomaly of the injection point, and the difference between time of injection and time of perigee passage, $t_0 - \tau$, are important parameters. Perigee and apogee distances are functions of a and e : $r_p = a(1-e)$, $r_a = a(1+e)$. In Section 16.2 it was found that the orbital elements a and e , the true anomaly of the injection point, θ_0 , and the time difference $t_0 - \tau$ are functions of the *in-plane* injection parameters r_0 , V_0 , γ_0 only. Therefore, in this Section we will analyze the variations of the parameters a , e , r_p , r_a , θ_0 and $t_0 - \tau$ as a function of the injection parameters r_0 , V_0 , γ_0 .

We first derive expressions for θ_0 as a function of r_0 , V_0 and γ_0 . In Section 5.6 the following relations were found:

$$\dot{r}_0 = V_0 \sin \gamma_0 = \frac{\mu e \sin \theta_0}{H} \quad ; \quad r_0 \dot{\theta}_0 = V_0 \cos \gamma_0 = \frac{\mu}{H} (1 + e \cos \theta_0)$$

$$H = r_0^2 \dot{\theta}_0 = r_0 V_0 \cos \gamma_0$$

from which we obtain

$$e \sin \theta_0 = \frac{r_0 V_0^2 \sin \gamma_0 \cos \gamma_0}{\mu} \quad ; \quad e \cos \theta_0 = \frac{r_0 V_0^2 \cos^2 \gamma_0}{\mu} - 1 \quad (16.11)$$

For the circular velocity at the injection point, we have

$$V_{c_0}^2 = \frac{\mu}{r_0}$$

with which the term $r_0 V_0^2 / \mu$ in (16.11) can be written as

$$\frac{r_0 V_0^2}{\mu} = \left(\frac{V_0}{V_{c_0}} \right)^2 = k_0 \quad (16.12)$$

Note that $k_0 > 1$ means that the injection velocity is higher than the local circular velocity, and that $k_0 < 1$ means that the injection velocity is lower than the local circular velocity. When we substitute k_0 for the term $r_0 V_0^2 / \mu$ in (16.1), (16.2) and (16.11), we find

$$\frac{a}{r_0} = \frac{1}{2 - k_0} \quad (16.13)$$

$$\begin{aligned} e^2 &= 1 - k_0(2 - k_0) \cos^2 \gamma_0 \\ e \sin \theta_0 &= k_0 \sin \gamma_0 \cos \gamma_0 \\ e \cos \theta_0 &= k_0 \cos^2 \gamma_0 - 1 \end{aligned} \quad (16.13)$$

From (16.13-3) and (16.13-4) the value of θ_0 can be determined unambiguously. As the injection point is known, the parameter θ_0 determines the position of the perigee of the orbit relative to the injection point.

Substitution of (16.12) and (16.13-1) into (16.3) leads to

$$\begin{aligned} e \sin E_0 &= \sqrt{k_0(2 - k_0)} \sin \gamma_0 \\ E_0 &= \arctan \left(\frac{\sqrt{k_0(2 - k_0)}}{k_0 - 1} \sin \gamma_0 \right) \end{aligned} \quad (16.14)$$

From these equations the value of E_0 can be determined unambiguously. For the time difference $t_0 - \tau$ we may write with (16.8) and (6.25)

$$\frac{t_0 - \tau}{T} = \frac{E_0 - e \sin E_0}{2\pi} \quad (16.15)$$

where T is the orbital period, which is a function of the orbital element a .

For the perigee or apogee distance we can write

$$\bar{r} = a(1 \pm e)$$

where \bar{r} denotes the perigee or apogee distance, and the plus-sign corresponds to apogee and the minus-sign to perigee. This expression can also be written as

$$\left(\frac{\bar{r}}{a} - 1 \right)^2 = e^2$$

Substitution of (16.13-1) and (16.13-2) into this relation leads to

$$\left(\frac{\bar{r}}{r_0} (2 - k_0) - 1 \right)^2 = 1 - k_0(2 - k_0) \cos^2 \gamma_0 \quad (16.16)$$

This equation has, for given values of γ_0 and k_0 (r_0 and V_0), in general two solutions for \bar{r} : the perigee and apogee distances.

From the expressions given above we conclude that the quantities a , e , θ_0 , $(t_0 - \tau)/T$ and \bar{r}/r_0 are fully determined by only two parameters: k_0 and γ_0 , where, of course, k_0 is a function of r_0 and V_0 . In Figure 16.11 the values of e , θ_0 , \bar{r}/r_0 and $(t_0 - \tau)/T$ are shown as a function of k_0 and γ_0 . From this Figure we conclude that:

- The curves for the variation of the eccentricity are symmetric about $\gamma_0 = 0^\circ$. Identical curves hold for $k_0 = 0.25$ and $k_0 = 1.75$, $k_0 = 0.50$ and $k_0 = 1.50$, etc.
- For a given value of k_0 , the eccentricity is minimum at $\gamma_0 = 0^\circ$.
- For near-circular orbits ($k_0 \approx 1$), small variations of γ_0 around $\gamma_0 = 0^\circ$ cause large variations of the eccentricity.
- For all values of $k_0 > 1$ and $\gamma_0 = 0^\circ$: $\theta_0 = 0^\circ$, indicating that the injection point coincides with

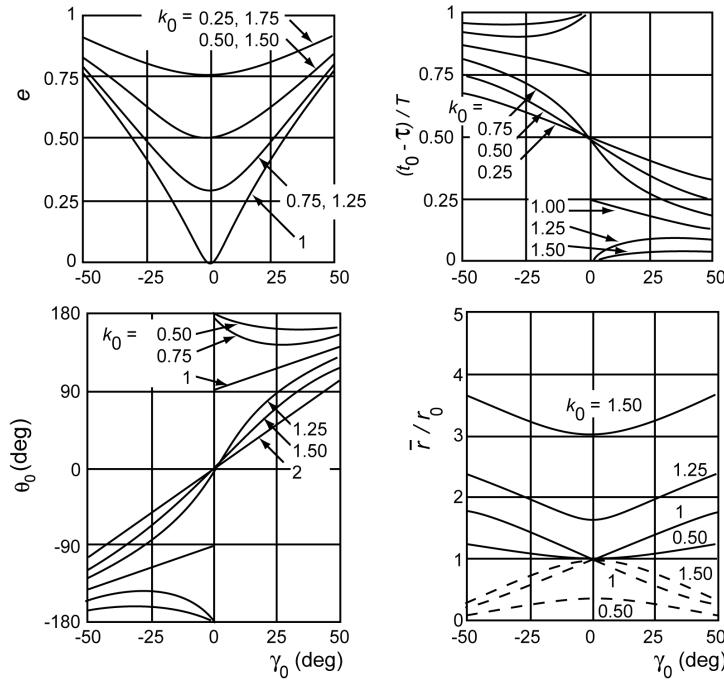


Figure 16.11: Values of some orbital parameters as a function of the injection parameters.

perigee. For all values of $k_0 < 1$ and $\gamma_0 = 0^\circ$: $\theta_0 = 180^\circ$, indicating that the injection point coincides with apogee. If k_0 is varied between 0 and 2, the position of perigee changes discontinuously by an amount of 180° at $k_0 = 1$.

- For a circular orbit ($k_0 = 1, \gamma_0 = 0^\circ$), the location of perigee is undetermined. For near-circular orbits ($k_0 \approx 1$), small variations of γ_0 about $\gamma_0 = 0^\circ$ cause large changes in the location of perigee.
- For given values of r_0 and k_0 , the perigee distance is maximum and the apogee distance is minimum at $\gamma_0 = 0^\circ$.
- For all values of $k_0 < 1$, the apogee distance at $\gamma_0 = 0^\circ$ is equal to r_0 ; for all values of $k_0 > 1$, the perigee distance at $\gamma_0 = 0^\circ$ is equal to r_0 .
- For $k_0 > 1$ and given values of γ_0 and r_0 , variations of k_0 lead to relatively small changes in perigee distance; in the apogee distance large changes occur. So, the apogee distance is sensitive to the injection velocity, in particular for highly eccentric orbits ($k_0 \approx 2$).
- The plot for the variation of $(t_0 - \tau)/T$ with variations in k_0 and γ_0 is a kind of mirror image of the plot for the variation of θ_0 . However, note the difference in the legend of ‘similar’ curves in both plots.

Since the density of the atmosphere decreases approximately exponentially with increasing altitude, it is in particular the perigee altitude that determines the orbital lifetime of a satellite. The injection point will always be chosen as low as possible to maximize the launcher payload capability. For the selected injection altitude the perigee altitude should be as high as possible to maximize the orbital lifetime. Furthermore, it is important that variations in perigee altitude due to injection errors will be as small as possible. Figure 16.11 shows that both requirements are satisfied if $\gamma_0 = 0^\circ$ at injection. To give an example of the required accuracy of the injection parameters, we consider an injection at $\gamma = 0^\circ$ into an orbit with nominal perigee and apogee altitudes of 500 km and 600 km, respectively. The magnitude of the nominal injection velocity is $V_0 = 7.642$ km/s. Figure 16.12 shows the range of variations in the values of r_0 , V_0 and γ_0 that lead to satellite orbits with perigee and apogee altitudes between 425 km and 675 km, or between

350 km and 750 km. This corresponds to significant orbit deviations of 75 km and 150 km, respectively. Note that for both cases the flight path angle, γ_0 , may deviate less 1.5° from its nominal value of $\gamma_0 = 0^\circ$; the magnitude of the injection velocity may deviate less than 70 m/s, the injection altitude less than 80 km. In reality, the orbit has to meet more-stringent requirements, which lead to much smaller permitted deviations in r_0 , V_0 , γ_0 .

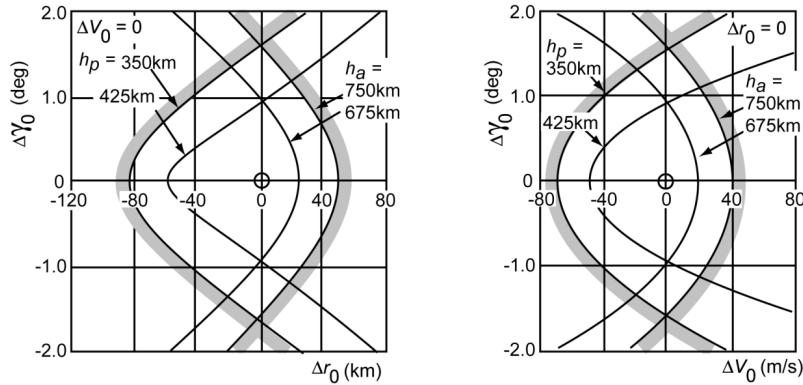


Figure 16.12: Range of acceptable in-plane injection parameter deviations for specified maximum departures in perigee and apogee altitude. Nominal orbit: $\gamma_0 = 0^\circ$, $h_p = 500\text{ km}$, $h_a = 600\text{ km}$.

16.6. Effects of injection errors

As a compromise between the desired satellite mission characteristics and the payload capability of available launch vehicles, a nominal orbit is selected. The true orbit of the satellite, however, will always deviate from the nominal orbit as a result of unavoidable launch trajectory errors and orbit injection errors. Both types of errors manifest themselves in deviations of the position of the injection point and of the velocity at that point: *injection errors*. It are these injection errors, which, although their numerical values may look small at first glance, may cause considerable deviations from the nominal orbit. Therefore, in the selection of a nominal orbit one always has to check whether the possible orbit deviations due to injection errors are acceptable. As an example, Figure 16.13 shows the $1-\sigma$ value (standard deviation) of a number of injection parameters for the (historical) Scout D launcher as a function of injection altitude. It was assumed that the errors in each injection parameter show a normal (Gaussian) distribution with zero mean. From statistics we know that the $1-\sigma$ value then corresponds to a probability of 68.3% that the absolute value of the error in a parameter is less than the $1-\sigma$ value of that parameter. So, Figure 16.13 should be read as follows: when a satellite is injected at an altitude of 400 n.mi. (740 km) by a Scout D into a low-eccentricity orbit, then there is a 68% probability that the absolute value of the errors in injection velocity, injection altitude, injection azimuth and injection flight path angle are smaller than 25.6 m/s, 8.1 km, 0.625° and 0.52° , respectively. To modern standards, Scout D had a poor guidance system. But even for this launcher the injections errors are relatively small; the $1-\sigma$ value of the injection velocity is only 25.6 m/s compared to an injection velocity of about 7.5 km/s; so, the error is about 0.3% of the nominal value.

Figure 16.14 (left) shows for nominal circular orbits the $1-\sigma$ values of the difference between apogee altitude and perigee altitude (Δh) and of the inclination due to injection errors of a Delta 3E launcher as a function of the nominal orbital altitude. Figure 16.14 (right) shows for the same launcher and for elliptical orbits with a perigee altitude of 100 n.mi. the $1-\sigma$ values of apogee altitude and inclination as a function of the nominal apogee altitude. Note that even for launch vehicles with a relatively good guidance system as the Delta 3E rocket, apogee altitude errors of several hundreds of kilometers could occur for highly eccentric orbits. For more-modern

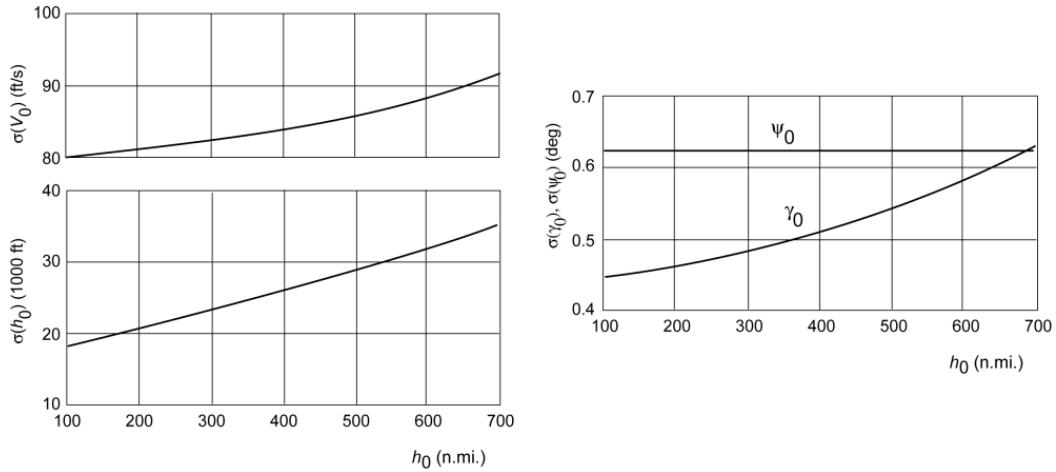


Figure 16.13: Standard deviation of some injection parameters for a Scout D launcher.

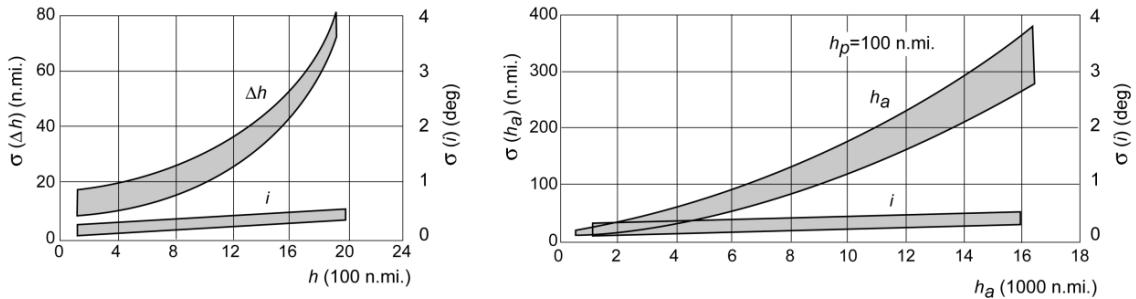


Figure 16.14: Standard deviation of the difference between apogee and perigee altitude for circular orbits (left) or of apogee altitude for elliptical orbits (right), and of the inclination, for an injection by a Delta 3E launcher.

launchers, like Delta II, Atlas/Centaur, Soyuz and Ariane 4, the injection errors are definitely smaller, leading to smaller orbit errors. For low Earth orbits, characteristic 3- σ values of the errors in a , e , i , Ω are then 10 km, 0.002, 0.1°, 0.1°, respectively; for a standard geosynchronous transfer orbit, characteristic 3- σ values of the errors in i , ω , Ω , h_p , h_a are 0.05°, 0.2°, 0.2°, 5 km, 100 km, respectively. For the modern Delta IV launcher, the 3- σ values of the injection errors in h_p , h_a , i are 7.4 km, 7.4 km, 0.04° when injecting the spacecraft into a circular 500 km altitude polar orbit, and 5.6 km, 93 km, 0.03° when injecting the spacecraft into a standard geosynchronous transfer orbit. For the current Ariane 5 and Soyuz-2 launchers and a standard geosynchronous transfer orbit from Kourou, the 1- σ values of the injection errors in i , ω , Ω , h_p , h_a are 0.02°, 0.2°, 0.2°, 1.3 km, 80 km, and 0.05°, 0.08°, 0.08°, 6.6 km, 40 km, respectively. For these launchers, the 1- σ values of the injection errors in a , e , i , Ω after a launch from Kourou into a Sun-synchronous 800 km altitude circular orbit are 2.5 km, 0.0004, 0.04°, 0.03°, and 3.3 km, 0.0007, 0.03°, 0.05°, respectively.

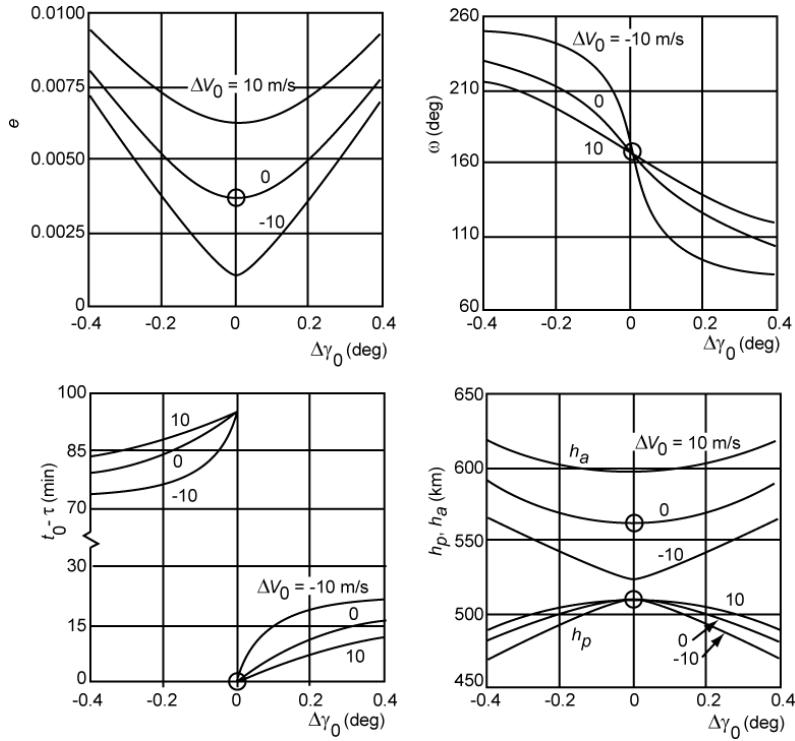
To illustrate the effects of injection errors in more detail, we consider the launch of the *Astronomical Netherlands Satellite* (ANS), which was launched on August 30, 1974, by a Scout D from the Western Test Range. Nominal perigee and apogee altitudes were 510 km and 560 km, respectively. The nominal values of the injection parameters and the corresponding orbital elements at injection are listed in Table 16.2. With (11.58) we can compute the right ascension of the injection point: $\alpha_0 = 63.56^\circ$. The (Keplerian) orbital period is 95.34 min. Now, we assume

Table 16.2: Nominal injection parameters and orbital elements of ANS.

$t_0 = \text{August 27, 1974; } 14^{\text{h}} 9^{\text{m}} 41^{\text{s}} \text{ UTC}$	
$r_0 = 6887.92 \text{ km}$	$a = 6913.078 \text{ km}$
$\delta_0 = 13.917^\circ$	$e = 0.003640$
$\Lambda_0 = -125.540^\circ$	$i = 97.801^\circ$
$V_0 = 7.6211 \text{ km/s}$	$\omega = 167.329^\circ$
$\gamma_0 = -0.005^\circ$	$\Omega = 241.618^\circ$
$\psi_0 = 188.038^\circ$	$t_i - \tau = 94.975 \text{ min}$

Actual launch took place on August 30, 1974; $14^{\text{h}} 7^{\text{m}} 40^{\text{s}}$ UTC.

that all injection parameters but V_0 and γ_0 have their nominal values. From Figure 16.13 we find the following $1-\sigma$ values for these parameters at the nominal injection altitude: $\Delta V_0 = 25 \text{ m/s}$, $\Delta \gamma_0 = 0.48^\circ$. For a numerical example, we assume that errors in V_0 and γ_0 well within these $1-\sigma$ limits occur: $\Delta V_0 = -10$ to 10 m/s , $\Delta \gamma_0 = -0.4^\circ$ to 0.4° . With the equations given in Section 16.2, we can compute the deviations of the orbital parameters due to errors in V_0 and γ_0 . For the orbital parameters e , ω , τ , h_p and h_a the results are shown in Figure 16.15. This Figure shows that the eccentricity may vary between 0.001 and 0.009 as a result of injection errors; the apogee altitude

Figure 16.15: Deviation of ANS orbital parameters due to injection errors within their $1-\sigma$ limits.

may deviate up to about $+60 \text{ km}$ or -35 km ; the perigee altitude up to about -50 km ; the argument of perigee up to about $\pm 90^\circ$ from its nominal value. It is emphasized that this Figure shows the effects of individual injection errors on the orbit of ANS separately. In reality, errors in all components of the injection state vector will occur simultaneously, while for each component a certain probability distribution holds. Through Monte Carlo simulations we can compute a large series of perigee and apogee altitude deviations when the values of the injection parameters are

varied according to their known probability distributions. Figure 16.6 shows contours which contain 75%, 95% and 99.5% of all computed combinations of perigee and apogee altitude deviations, for a satellite that is injected at an altitude of 300 n.mi. (560 km) by a Scout D rocket into its orbit. This Figure should be read as follows. If a satellite is injected into a circular orbit (left) at an altitude of 300 n.mi. then there is a 95% probability (2σ value) that combinations of deviations in perigee altitude and apogee altitude occur that falls within the 0.95-contour. This includes the following combinations: $\Delta h_p = -120$ n.mi., $\Delta h_a = 0$ n.mi. and $\Delta h_p = 0$ n.mi., $\Delta h_a = 120$ n.mi. Of course, the first combination would be disastrous for the orbital lifetime of the satellite. Figure 16.16 (right) shows that if the satellite is injected at the same altitude but with a higher velocity, i.e. into an orbit with a higher apogee altitude (550 n.mi.), the influence of injection errors on the perigee altitude is statistically considerably smaller, but the deviation in apogee altitude increases. However, this is generally less of a problem, because it is in particular the perigee altitude that determines the satellite's orbital lifetime. So, when lifetime is a critical factor, one will prefer an injection with a velocity higher than the (local) circular velocity to minimize possible negative perigee altitude deviations due to injection errors.

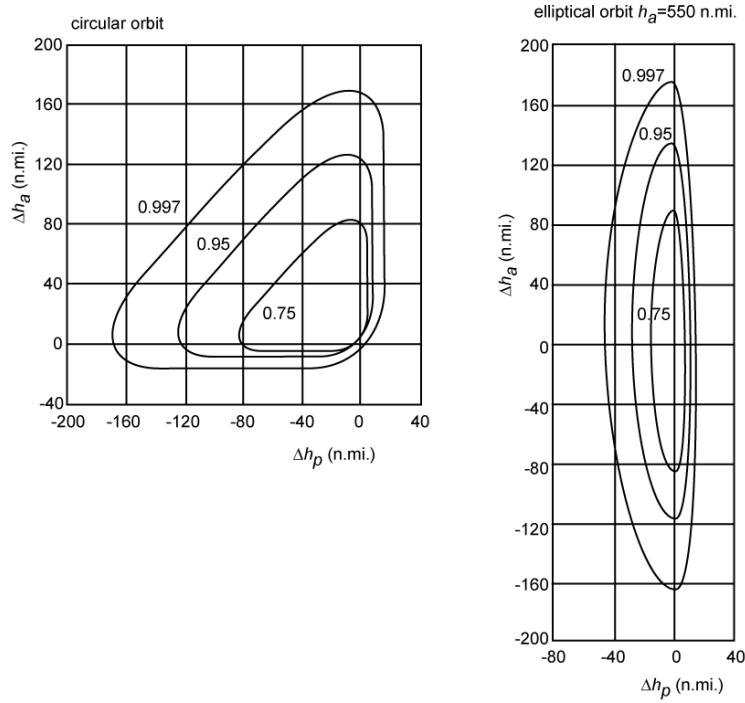


Figure 16.16: Probability contours of simultaneous deviations in perigee and apogee altitude due to injection errors of a Scout D launcher, for a nominal circular orbit (left) and a nominal elliptical orbit (right). In both cases injection takes place at an altitude of 300 n.mi.

When we limit ourselves to small injection errors, we can derive analytical expressions for the sensitivity of orbital parameters to injection errors. For this purpose, we consider the relations between the injection parameters r_0 , δ_0 , α_0 , V_0 , γ_0 , ψ_0 and the orbital elements a , e , i , ω , Ω and the orbital parameters r_a and r_p , as given in Section 16.2. The orbital element τ is omitted, because for this element we would get quite complicated expressions and, moreover, because in this Chapter we are more interested in the orbit than in the relation between time and position of the satellite in the orbit. If x denotes one of the orbital elements or orbital parameters, then we may approximate the deviation (Δx) of an orbital element or parameter from its nominal value (x_0) as a result of injection errors by a Taylor series expansion:

$$\Delta x = \left(\frac{\partial x}{\partial r} \right)_0 \Delta r + \left(\frac{\partial x}{\partial \delta} \right)_0 \Delta \delta + \left(\frac{\partial x}{\partial \alpha} \right)_0 \Delta \alpha + \left(\frac{\partial x}{\partial V} \right)_0 \Delta V + \left(\frac{\partial x}{\partial \gamma} \right)_0 \Delta \gamma + \left(\frac{\partial x}{\partial \psi} \right)_0 \Delta \psi + O(\Delta^2) \quad (16.17)$$

where the partial derivatives have to be evaluated for the nominal conditions. This implies that in the expressions for the partial derivatives the orbital elements have their nominal values. We now limit ourselves to small deviations in the injection parameters, so that second-order terms, indicated in (16.17) by $O(\Delta^2)$, may be neglected. The partial derivatives can be found by partial differentiation of the relevant expressions given in Section 16.2 and by using (16.12). As an example, we will derive an expression for the partial derivative of e to γ . From (16.2) we find

$$2e \left(\frac{\partial e}{\partial \gamma} \right)_0 = 2 \frac{r_0 V_0^2}{\mu} \left(2 - \frac{r_0 V_0^2}{\mu} \right) \cos \gamma_0 \sin \gamma_0$$

or, with (16.12),

$$\left(\frac{\partial e}{\partial \gamma} \right)_0 = \frac{k_0 (2 - k_0)}{2e} \sin 2\gamma_0$$

In a similar way, we can derive expressions for the other partial derivatives and finally obtain

$$\begin{aligned} \frac{\partial a}{\partial r} &= \frac{2}{(2 - k_0)^2} ; \quad \frac{\partial a}{\partial V} = \frac{2k_0}{(2 - k_0)^2} \frac{r_0}{V_0} ; \quad \frac{\partial a}{\partial \delta} = \frac{\partial a}{\partial \alpha} = \frac{\partial a}{\partial \gamma} = \frac{\partial a}{\partial \psi} = 0 \\ \frac{\partial e}{\partial r} &= \frac{k_0 (k_0 - 1)}{e r_0} \cos^2 \gamma_0 ; \quad \frac{\partial e}{\partial V} = \frac{2k_0 (k_0 - 1)}{e V_0} \cos^2 \gamma_0 \\ \frac{\partial e}{\partial \gamma} &= \frac{k_0 (2 - k_0)}{2e} \sin 2\gamma_0 ; \quad \frac{\partial e}{\partial \delta} = \frac{\partial e}{\partial \alpha} = \frac{\partial e}{\partial \psi} = 0 \\ \frac{\partial i}{\partial \delta} &= \frac{\sin \delta_0 \sin \psi_0}{\sin i} ; \quad \frac{\partial i}{\partial \psi} = -\frac{\cos \delta_0 \cos \psi_0}{\sin i} ; \quad \frac{\partial i}{\partial r} = \frac{\partial i}{\partial \alpha} = \frac{\partial i}{\partial V} = \frac{\partial i}{\partial \gamma} = 0 \\ \frac{\partial \omega}{\partial r} &= \frac{k_0}{e^2 r_0} \sin \gamma_0 \cos \gamma_0 ; \quad \frac{\partial \omega}{\partial \delta} = \frac{\cos \psi_0}{\sin^2 i} ; \quad \frac{\partial \omega}{\partial V} = \frac{2k_0}{e^2 V_0} \sin \gamma_0 \cos \gamma_0 \\ \frac{\partial \omega}{\partial \gamma} &= -1 - \frac{k_0 - 1}{e^2} ; \quad \frac{\partial \omega}{\partial \psi} = \frac{\sin \delta_0 \cos \delta_0 \sin \psi_0}{\sin^2 i} ; \quad \frac{\partial \omega}{\partial \alpha} = 0 \\ \frac{\partial \Omega}{\partial \delta} &= -\frac{\sin \psi_0 \cos \psi_0 \cos \delta_0}{\sin^2 i} ; \quad \frac{\partial \Omega}{\partial \alpha} = 1 ; \quad \frac{\partial \Omega}{\partial \psi} = -\frac{\sin \delta_0}{\sin^2 i} ; \quad \frac{\partial \Omega}{\partial r} = \frac{\partial \Omega}{\partial V} = \frac{\partial \Omega}{\partial \gamma} = 0 \\ \frac{\partial \bar{r}}{\partial r} &= \frac{1}{2 - k_0} \left[\frac{2(1 \pm e)}{2 - k_0} \pm \frac{k_0(k_0 - 1)}{e} \cos^2 \gamma_0 \right] ; \quad \frac{\partial \bar{r}}{\partial \gamma} = \pm \frac{r_0 k_0}{2e} \sin 2\gamma_0 \end{aligned} \quad (16.18)$$

$$\frac{\partial \bar{r}}{\partial V} = \frac{2k_0 r_0}{V_0(2-k_0)} \left[\frac{1 \pm e}{2-k_0} \pm \frac{k_0 - 1}{e} \cos^2 \gamma_0 \right] ; \quad \frac{\partial \bar{r}}{\partial \delta} = \frac{\partial \bar{r}}{\partial \alpha} = \frac{\partial \bar{r}}{\partial \psi} = 0 \quad (16.18)$$

where we have omitted for simplicity the index 0 at each partial derivative. In these expressions \bar{r} denotes perigee or apogee distance, and the plus-sign in the expressions for $\partial \bar{r}/\partial r$, $\partial \bar{r}/\partial \gamma$ and $\partial \bar{r}/\partial V$ corresponds to apogee distance and the minus-sign to perigee distance.

If we now consider the case that injection occurs at perigee of the nominal orbit, we have $\gamma_0 = 0^\circ$, $k_0 > 1$, and (16.13-2) simplifies to

$$e = k_0 - 1 \quad (16.19)$$

Substitution of (16.13-1) and (16.19) into (16.18) leads for the partial derivatives to

$$\begin{aligned} \frac{\partial a}{\partial r} &= \frac{2}{2-k_0} \frac{a}{r_0} ; \quad \frac{\partial a}{\partial V} = \frac{2k_0}{2-k_0} \frac{a}{V_0} ; \quad \frac{\partial a}{\partial \delta} = \frac{\partial a}{\partial \alpha} = \frac{\partial a}{\partial \gamma} = \frac{\partial a}{\partial \psi} = 0 \\ \frac{\partial e}{\partial r} &= \frac{k_0}{r_0} ; \quad \frac{\partial e}{\partial V} = 2 \frac{k_0}{V_0} ; \quad \frac{\partial e}{\partial \delta} = \frac{\partial e}{\partial \alpha} = \frac{\partial e}{\partial \gamma} = \frac{\partial e}{\partial \psi} = 0 \\ \frac{\partial i}{\partial \delta} &= \frac{\sin \delta_0 \sin \psi_0}{\sin i} ; \quad \frac{\partial i}{\partial \psi} = - \frac{\cos \delta_0 \cos \psi_0}{\sin i} ; \quad \frac{\partial i}{\partial r} = \frac{\partial i}{\partial \alpha} = \frac{\partial i}{\partial V} = \frac{\partial i}{\partial \gamma} = 0 \\ \frac{\partial \omega}{\partial \delta} &= \frac{\cos \psi_0}{\sin^2 i} ; \quad \frac{\partial \omega}{\partial \gamma} = - \frac{k_0}{k_0 - 1} ; \quad \frac{\partial \omega}{\partial \psi} = \frac{\sin 2 \delta_0 \sin \psi_0}{2 \sin^2 i} ; \quad \frac{\partial \omega}{\partial r} = \frac{\partial \omega}{\partial \alpha} = \frac{\partial \omega}{\partial V} = 0 \\ \frac{\partial \Omega}{\partial \delta} &= - \frac{\sin 2 \psi_0 \cos \delta_0}{2 \sin^2 i} ; \quad \frac{\partial \Omega}{\partial \alpha} = 1 ; \quad \frac{\partial \Omega}{\partial \psi} = - \frac{\sin \delta_0}{\sin^2 i} ; \quad \frac{\partial \Omega}{\partial r} = \frac{\partial \Omega}{\partial V} = \frac{\partial \Omega}{\partial \gamma} = 0 \\ \frac{\partial r_p}{\partial r} &= 1 ; \quad \frac{\partial r_p}{\partial \delta} = \frac{\partial r_p}{\partial \alpha} = \frac{\partial r_p}{\partial V} = \frac{\partial r_p}{\partial \gamma} = \frac{\partial r_p}{\partial \psi} = 0 \\ \frac{\partial r_a}{\partial r} &= \frac{4}{(2-k_0)^2} - 1 ; \quad \frac{\partial r_a}{\partial V} = \frac{4k_0}{(2-k_0)^2} \frac{r_0}{V_0} ; \quad \frac{\partial r_a}{\partial \delta} = \frac{\partial r_a}{\partial \alpha} = \frac{\partial r_a}{\partial \gamma} = \frac{\partial r_a}{\partial \psi} = 0 \end{aligned} \quad (16.20)$$

where, again, the index 0 has been omitted at each partial derivative. Note that the value of some partial derivatives approaches infinity, and consequently the linearized series expansion (16.17) becomes invalid, if $e \rightarrow 0$ ($k_0 \rightarrow 1$), $i \rightarrow 0$ or $V_0/V_{c0} \rightarrow \sqrt{2}$ ($k_0 \rightarrow 2$). In these cases the argument of perigee is undefined, or the right ascension of the ascending node and the argument of perigee are undefined, or the orbit of the spacecraft is no longer an ellipse but a parabola. These situations are a consequence of the chosen set of orbital elements; in such situations other sets of orbital parameters should be used (Sections 11.5 and 22.4).

From the expressions (16.20) we conclude that when a satellite is injected at perigee of its nominal orbit, in first approximation:

- The semi-major axis and the eccentricity are insensitive to errors in δ_0 , α_0 , γ_0 , ψ_0 .
- The error in the semi-major axis due to errors in r_0 and V_0 increases with increasing values of k_0 . So, for highly eccentric orbits large deviations in a may occur.

- The inclination is insensitive to errors in r_0 , α_0 , V_0 and γ_0 .
- The argument of perigee is only affected by errors in δ_0 , γ_0 and ψ_0 .
- For near-circular orbits large deviations in the argument of perigee due to errors in γ_0 may occur.
- The right ascension of the ascending node is only affected by errors in δ_0 , α_0 and ψ_0 .
- Perigee distance is only affected by errors in r_0 ; apogee distance is only affected by errors in r_0 and V_0 . The error in the perigee distance is just equal to the error in the injection distance.
- The error in the apogee distance increases with increasing values of k_0 . So, for highly eccentric orbits large deviations in r_a may occur.

When, as a numerical example, the ANS satellite is considered again and the nominal injection parameters are taken from Table 16.2, then the values of the partial derivatives at injection can be computed by substituting these injection parameters into the expressions (16.20). We then assume that injection takes place at perigee with $\gamma_0 = 0^\circ$, which is close to the nominal mission plan. The results are listed in Table 16.3, from which we conclude that in particular the influence of γ on ω is very large.

Table 16.3: Partial derivatives of the orbital parameters to the injection parameters of ANS.

$\partial a / \partial r = 2.015 \text{ km/km}$	$\partial a / \partial V = 1.828 \text{ km/m.s}^{-1}$	
$\partial e / \partial r = 1.457 * 10^{-4} \text{ /km}$	$\partial e / \partial V = 2.634 * 10^{-4} \text{ /m.s}^{-1}$	
$\partial i / \partial \delta = -0.0339 \text{ }^\circ/\text{°}$	$\partial i / \partial \psi = 0.970 \text{ }^\circ/\text{°}$	
$\partial \omega / \partial \delta = -1.0088 \text{ }^\circ/\text{°}$	$\partial \omega / \partial \gamma = -274 \text{ }^\circ/\text{°}$	$\partial \omega / \partial \psi = -0.0333 \text{ }^\circ/\text{°}$
$\partial \Omega / \partial \delta = -0.137 \text{ }^\circ/\text{°}$	$\partial \Omega / \partial \alpha = 1 \text{ }^\circ/\text{°}$	$\partial \Omega / \partial \psi = -0.245 \text{ }^\circ/\text{°}$
$\partial r_p / \partial r = 1 \text{ km/km}$		
$\partial r_a / \partial r = 3.029 \text{ km/km}$	$\partial r_a / \partial V = 3.655 \text{ km/m.s}^{-1}$	

All other partial derivatives are zero.

We have found that for injection at perigee, the apogee distance is only a function of r_0 and V_0 . We thus may write according to (16.17)

$$\Delta \mathbf{r}_a = \left(\frac{\partial \mathbf{r}_a}{\partial \mathbf{r}} \right)_0 \Delta \mathbf{r} + \left(\frac{\partial \mathbf{r}_a}{\partial V} \right)_0 \Delta V$$

Substitution of the relevant expressions for the partial derivatives leads to

$$\Delta \mathbf{r}_a = \frac{1}{(2 - k_0)^2} \left[k_0 (4 - k_0) \Delta \mathbf{r} + 4 \frac{k_0 r_0}{V_0} \Delta V \right]$$

Assuming a spherical Earth and using $h = r - R$, where h is altitude and R is the Earth radius, and (16.19), we find

$$\Delta h_a = \frac{1 + e}{(1 - e)^2} \left[(3 - e) \Delta h + 4 \frac{r_0}{V_0} \Delta V \right] \quad (16.21)$$

Since the injection point coincides with perigee, we may write $V_0 = V_{c0} \sqrt{1 + e}$ and thus (16.21) can be written as

$$\frac{\Delta h_a}{h_0} = \frac{(1+e)}{(1-e)^2} \left[(3-e) \frac{\Delta h}{h_0} + \frac{4}{\sqrt{1+e}} \left(1 + \frac{R}{h_0} \right) \frac{\Delta V}{V_{c_0}} \right] \quad (16.22)$$

For injection into a low-eccentricity orbit with $e < 0.02$, we may approximate this relation by

$$\frac{\Delta h_a}{h_0} = 3 \frac{\Delta h}{h_0} + 4 \left(1 + \frac{R}{h_0} \right) \frac{\Delta V}{V_{c_0}} \quad (16.23)$$

Usually, injection takes place at $h_0 = 200\text{-}400$ km. Then, the term $1 + R/h_0$ is always larger than 16. So, (16.22) shows that an error Δh leads to a deviation Δh_a that is more than three times as large, and that the apogee altitude is very sensitive to errors in ΔV . As an example, we consider the injection of a satellite at perigee of an orbit with nominal injection altitude of 300 km and nominal apogee altitude of 2000 km by a modern launcher that achieves 3- σ injection errors of $\Delta h \approx 2$ km and $\Delta V \approx 3$ m/s. We then find $e = 0.113$, $\Delta h/h_0 = 0.67\%$, and $\Delta V/V_{c_0} = 0.039\%$. From (16.23) we obtain for the contribution of Δh to the apogee altitude deviation: $\Delta h_a = 8.17$ km and $\Delta h_a/h_0 = 2.72\%$, and for the contribution of ΔV : $\Delta h_a = 13.91$ km and $\Delta h_a/h_0 = 4.64\%$. So, the adopted error of 2 km in injection altitude leads to about four times that error in apogee altitude, while the adopted error of 3 m/s in injection velocity yields an error of about 14 km in apogee altitude.

It is emphasized that the discussion in this Section refers to the ‘normal’ case, where the errors in the injection parameters show a Gaussian distribution with zero mean. However, in practice much larger injection errors, or even injection failures, may occur. As an example, we consider again the launch of ANS. During the powered flight an error occurred in the control of the rocket that led to large deviations of the injection parameters. As a result, the satellite entered an orbit with perigee altitude of 266 km and apogee altitude of 1175 km; these values differed considerably from the nominal values of 510 km and 560 km, respectively. The very different orbit severely affected the satellite’s observation program and its operational lifetime. The satellite reentered the atmosphere over the Indonesian archipelago on June 14, 1977, at about 2^h UTC.

17. LUNAR FLIGHTS

From ancient times down to the time of Newton's work in the seventeenth century, lunar theories were composed mainly with the help of geometric ideas, inspired more or less directly by long series of positional observations of the Moon. From the mid-seventeenth century, the Moon has played a significant role in assisting the progress of celestial mechanics. The first explanation of the irregularities in the motion of the Moon was given by I. Newton (1643-1727) in his *Principia*. He nevertheless regarded the lunar theory as very difficult and confined to E. Halley (1656-1742) that it "made his head ache and kept him awake so often that he would think of it no more". In the eighteenth to the early twentieth centuries, lunar theory was developed analytically by L. Euler (1707-1783), A.C. Clairaut (1713-1765), J.B. le Rond d'Alembert (1717-1783), J.L. Lagrange (1736-1813), P.S. Laplace (1749-1827), P.A. Hansen (1795-1874), C.E. Delaunay (1816-1872), S. Newcomb (1835-1909), G.W. Hill (1838-1914), E.W. Brown (1866-1938), and others. Laplace brought his special form of mathematical elegance to the solution of the Moon's motion. He used the true longitude as the independent variable in the equations of motion. He explained the secular acceleration of the true longitude; it depends on the eccentricity of the Earth's orbit, which changes over time. Hansen restructured much of Laplace's work and developed tables that were used from 1862 to about 1922. Delaunay's theory was accurate enough to predict the Moon's motion to one radius over a period of 20 years. Newcomb worked on improving ephemerides for navigation. He made some empirical corrections to Hansen's tables, and his modified results were used after 1883. A more-exact theory, using the theory of infinite matrices, was published by Hill in 1878. Brown published very accurate tables, based on a lunar theory consisting of more than 1500 terms, in 1908. They replaced Hansen's tables in about 1923 and were used for decades afterwards.

Since the 1960s, lunar theory has been further developed in a somewhat different way. This has been stimulated by the availability of high-speed computers and modern high-precision observations. W.J. Eckert (1902-1971) and colleagues used early computers to expand Brown's lunar theory by 6000 terms in 1964. Most significant has been the record of lunar laser ranging measurements, obtained using lasers on Earth and retro-reflectors placed on the surface of the Moon. The time-of-flight of a pulse of laser light to one of the reflectors and back gives a measure of the Moon's distance at that time. The first of four reflectors, which are still operational today, was taken to the Moon in the Apollo 11 spacecraft in July 1969. The lunar theory, as developed numerically to fine precision using these modern measures, is based on a larger range of considerations than the classical theories; it takes account not only of gravitational forces (with relativistic corrections) but also of many tidal and geophysical effects and a greatly extended theory of lunar librations. A leading position in these developments was taken by the NASA Jet Propulsion Laboratory (JPL), and names particularly associated with the transition from classical lunar theories towards the modern state of the art include those of J.D. Mulholland (1934-2008) and J.G. Williams (-). In parallel with these developments, a new class of analytical lunar theories has also been developed since the 1960s. Using computer-assisted algebra, the analytical developments have been taken further than previously could be done by the classical analysts working manually. A leading scientist from this period is A. Deprit (1926-2006), who reproduced Delaunay's lunar theory with computers and symbolic manipulation around 1970. He found several errors in Delaunay's theory and therefore improved its accuracy significantly. Also, some of these new analytical theories have been fitted to the numerical lunar ephemerides previously developed at JPL. The main aims of these advanced analytical theories, in contrast to the aims of the classical theories of past centuries, have not been to generate improved positional data for current dates; rather, their aims have included the study of further aspects of the motion,

such as long-term properties, which may not so easily be apparent from the modern numerical theories themselves.

Since the 1960s, the Moon is an important target for manned and unmanned space missions, and presently concrete plans exist to build a permanent manned lunar basis and to use the Moon as a starting point for a manned mission to Mars. In the next Section of this Chapter¹, we will briefly describe the Earth-Moon system, together with some of the irregularities of the Moon's motion. In subsequent Sections, we will analyze the launching of a spacecraft from the Earth to the Moon.

The first spacecraft to the Moon was launched by the former USSR in January 1959: Luna 1. This spacecraft passed the surface of the Moon at an altitude of about 6000 km. The first US spacecraft that passed the Moon was Pioneer 4, which was launched in March 1959. It passed the Moon at a distance of about 60,000 km, which was much too far to provide any useful measurements. Since then, both the USA and the USSR/Russia have launched many spacecraft towards the Moon. In January 1966, the USSR launched Luna 9, which became the first spacecraft to soft land on the Moon. Equipped with a camera, Luna 9 provided a panoramic view of the lunar surface surrounding its landing site. In April 1966, the USSR launched Luna 10, which became the first spacecraft to orbit another celestial body. Upon arrival, Luna 10 conducted multiple spectral, gravitational, solar plasma and meteorite detection studies.

As a prelude to the US manned Apollo program, the USA have sent a series of Ranger, Surveyor and Lunar Orbiter spacecraft to the Moon. In the period April 1962 to March 1965, five Ranger spacecraft have impacted on the Moon. During the last part of their approach, three spacecraft made a series of pictures of the surface of the Moon at ever decreasing altitudes and increasing velocities, and transmitted these immediately to the Earth. They gave a good impression of what astronauts could expect on the surface of the Moon. In the period May 1966 to January 1968, five Surveyor spacecraft made a soft landing on the Moon, and have sent series of pictures to the Earth. In addition, these spacecraft have measured a number of physical and mechanical parameters of the lunar soil. In the period August 1966 to August 1967, a total of five Lunar Orbiters were injected into an orbit about the Moon. They have taken a large number of high-quality photo's, from which an almost complete photographic map of the lunar surface could be produced and many potential Apollo landing sites were photographed in detail. The first three missions were flown in low-inclination low-altitude orbits; the last two missions were flown in high-altitude polar orbits, in which the altitude ranged from 45 km over the near side of the Moon to 6000 km over the far side of the Moon. The first manned flight towards the Moon (Section 3.11) took place in December 1968, and Apollo 8 entered on December 24 an orbit about the Moon at an altitude of about 113 km. After ten orbits about the Moon, the rocket engine fired again for the return to Earth. The first man setting foot on the Moon was Neil Armstrong, one of the three astronauts of the Apollo 11 mission, who landed, together with Edwin Aldrin, on the Moon on July 20, 1969. The astronauts collected and returned almost 22 kg of lunar rock and soil for scientists to study on Earth. After this flight, six more Apollo flights to the Moon have taken place; the last one was Apollo 17 that landed on the Moon on December 11, 1972. In total, twelve US astronauts have spent 12 days and 12 hrs on the Moon, and the astronauts have walked for 3 days and 8 hrs on the Moon. In September 1970, the USSR landed Luna 16 on the Moon, which included an ascent vehicle. After a sample of the surface of the Moon was retrieved, the ascent vehicle launched itself to Earth and a capsule with the sample landed safely

¹ This Chapter is partly based on Chapter 7 of: R.R. Bate, D.D. Mueller, J.E. White, *Fundamentals of Astrodynamics*, Dover Publications, New York, 1971.

on Earth. In November 1970, the USSR successfully sent a lander/rover to the Moon: Luna 17/Lunokhod 1. The rover was equipped with four television cameras and special extendable devices to impact the lunar soil for soil density and mechanical properties tests. On December 14, 2013, China became the third nation that has successfully soft-landed a spacecraft on the Moon: Chang'e 3, with a landing mass of 1200 kg. On the same day, a 140 kg rover, named Yutu, separated from the lander. Both spacecraft carried a suite of scientific instruments and cameras; the rover carried a ground-penetrating radar, allowing for the first direct measurement of the structure of the lunar soil down to a depth of 30 m. Below, some lunar missions will be described in more detail.

On January 24, 1990, the Japanese Muses-A spacecraft was launched; after its successful launch, the spacecraft was renamed Hiten. It was launched into a highly eccentric orbit about the Earth with an apogee altitude of 290,000 km. Using its propulsion system, the apogee was gradually raised and ten lunar swingby's (Section 18.11) were performed. The mission of this spacecraft is described in some detail in Section 17.5.

On January 25, 1994, the 227 kg Clementine spacecraft was launched from Vandenberg (Western Test Range) by a Titan 23G into a 260 km by 300 km altitude orbit with inclination of 67° . It was a joint project between the Strategic Defense Initiative Organization and NASA. Clementine was the first US spacecraft launched to the Moon since Explorer 49 in June 1973. The objective of the mission was to test advanced sensors and spacecraft components, and to make scientific observations of the Moon and the near-Earth asteroid 1620 Geographos. The observations included imaging in the ultraviolet, visible and infrared parts of the spectrum, laser altimetry, and charged particle measurements. The spacecraft remained in low-Earth orbit until February 3, 1994, when a solid-propellant rocket ignited to send the spacecraft into a very eccentric orbit with apogee distance somewhat less than lunar distance. After two successive Earth perigee passages on February 5 and February 15, the spacecraft's rocket engine fired and put Clementine successfully into a highly elliptical polar orbit about the Moon on February 21. This orbit had a period of 5 days and a perilune altitude of 400 km. A second engine burn, performed one day later, placed the spacecraft in its polar mapping orbit. This orbit had a period of $4^{\text{h}}58^{\text{m}}$, a perilune altitude of 429 km, and an apolune altitude of 2924 km. Initially, the mapping orbit had its argument of perilune 30° below the lunar equator; after one month a maneuver was executed to move the argument of perilune to 30° above the lunar equator. From its orbit, Clementine transmitted about 1.6 million digital images of the lunar surface. After completing its mission goals over 297 orbits around the Moon, Clementine's thrusters were fired on 5 May, 1994, to inject the spacecraft into a trajectory that would lead to a rendez-vous with the asteroid 1620 Geographos in August 1994. However, due to a computer problem on May 7, which caused a thruster to fire for 11 min and use up all propellant, the spacecraft was put into an uncontrollable tumble at about 80 rpm with no spin control. This made the planned flyby of the asteroid impossible. The spacecraft remained in geocentric orbit and continued testing the spacecraft components until the mission was terminated in June 1994, when falling power supply levels no longer allowed sufficient telemetry exchange.

Following launch on January 7, 1998, aboard a three-stage Athena II rocket from Cape Canaveral, the NASA 296 kg Lunar Prospector entered a 185 km altitude, 29.2° inclination, circular parking orbit about the Earth. About 42 min later, a STAR 37FM rocket engine was fired to deliver a ΔV of 3.142 km/s over the course of a 64 s burn. The spacecraft then entered a 105 hr transfer orbit to the Moon with semi-major axis of 200,240 km, eccentricity of 0.9672, and inclination of 29.18° . The spacecraft passed the Moon at a closest distance of 1819 km on January 11. Then, the spacecraft's rocket engine was fired for about 30 min to inject the spacecraft into an 11.63 hr period orbit about the Moon with inclination of 89.8° relative to the

Moon's equator. After 24 hours, Lunar Prospector was inserted into a 3.52 hr period intermediate orbit, followed 24 hours later, on January 13, by a transfer into a 92 km by 153 km altitude 2.00 hr period preliminary mapping orbit, and then on January 16 by insertion into the near-circular 100 km altitude nominal lunar polar mapping orbit with an inclination of 90.55° and a period of 117.9 min. The objective of the mission was to map the surface composition and possible polar ice deposits, to measure the magnetic and gravity fields, and to study lunar outgassing events. The data collection from this orbit was periodically interrupted for orbital maintenance burns to re-circularize the orbit whenever the perilune or apolune altitude was more than 20 km from the 100 km nominal orbit, about once a month. On December 19, 1998, a maneuver lowered the orbit to 40 km altitude to perform higher-resolution studies. The orbit was altered again on January 28, 1999, to a 15 km by 45 km altitude orbit. In this orbit, Lunar Prospector has passed as low as 8 km over lunar mountains in the southern polar and farside regions. This extended mission phase of six months has enabled the spacecraft to obtain science data up to three orders of magnitude better than at the primary mapping orbit. The mission ended on July 31, 1999, when Lunar Prospector deliberately impacted in a permanently shadowed area of the Shoemaker crater near the lunar south pole. It was hoped that the impact would liberate water vapor from the suspected ice deposits in the crater and that the plume would be detectable from Earth; however, no such plume was observed.

The ESA SMART-1 spacecraft was launched by an Ariane 5 rocket, together with a communications satellite, into a geostationary transfer orbit on September 27, 2003. The small thrust produced by its onboard ion propulsion system (Section 19.1) made the spacecraft to spiral out over a fourteen months period until being captured by the Moon on November 14, 2004. The mission of this spacecraft is described in some detail in Section 17.5

The 2885 kg Japanese Kaguya spacecraft (also known as SELENE) was launched on September 14, 2007, by an H-IIA rocket from Tanegashima Space Center. The primary objectives of this lunar mission were a global survey of the Moon, obtaining data on elemental abundance, mineralogical composition, topography, geology, gravity and magnetic fields, lunar and solar-terrestrial plasma environments, and the development of critical technologies for future lunar exploration. The spacecraft consisted of three satellites; an orbiter containing most of the scientific equipment, and two almost identical 50 kg subsatellites: Ouna and Okina. Both subsatellites had no propulsion systems and carried an *X*-band and three *S*-band radio transmitters. The spacecraft was injected into an orbit with perigee and apogee altitudes of 281 km and 232,805 km, respectively. An apogee maneuver raised perigee altitude to 924 km. During its second passage of this orbit's perigee the spacecraft's propulsion system was fired to inject the spacecraft into a lunar transfer trajectory. The lunar orbit injection maneuver was executed on October 3, and the spacecraft entered a lunar orbit with perilune and apolune altitudes of 101 km and 11,741 km, respectively. Over a two-week period a series of deceleration maneuvers around perilune were executed, which brought the orbiter in its polar science orbit, with perilune and apolune altitudes of 80 km and 128 km, respectively, on October 19. During the transition to this science orbit, the data-relay subsatellite Okina was released into a 100 km by 2400 km altitude polar orbit on October 9 and Ouna was released into a 100 by 800 km altitude orbit on October 12. Normal operations for the Kaguya spacecraft started in early November 2007. Ouna, in conjunction with Okina, has enabled differential Very-Long Baseline Interferometry (VLBI) observations from the ground to determine the satellites positions very accurately. Okina has relayed four-way doppler ranging signals between a ground station and the orbiter for the determination of the far side lunar gravity field. By the end of October 2008, Kaguya completed the planned operation, and entered an extended operational phase. It was planned to continue through March 2009, then lower the orbit to a 50 km altitude circular orbit, then to a 20 km by

100 km altitude elliptical orbit, and to let it impact on the Moon by August 2009. However, due to the degraded performance of a reaction wheel the plan was changed so that from February 1, 2009, the orbit was lowered to an altitude of 50 km and the lunar surface impact occurred on June 10, 2009, on the lunar near side at 80.4° E, 65.5° S. This impact has been successfully captured by at least one observatory in Australia. Okina ended its mission on February 12, 2009, when it impacted into the lunar far side.

Chandrayaan-1, India's first spacecraft to the Moon, was launched on October 22, 2008, by an PSLV-C11 rocket from India's Space Center at Sriharikota. The 1380 kg spacecraft carried eleven scientific instruments, and a bi-propellant integrated propulsion system. This system consisted of one 440 N engine and eight 22 N thrusters. Two tanks stored 390 liters of fuel and oxidizer. The mission was controlled from a spacecraft control center at Bangalore, using India's Deep Space Network at Byalalu. The launcher placed the spacecraft in an elliptical orbit with perigee altitude of 255 km, apogee altitude of 22,860 km, and inclination of 17.9° . Over a period of two weeks, the spacecraft's liquid-propellant rocket engine has been fired five times to increase the apogee altitude, first to 37,900 km, then to 74,715 km, later to 164,600 km, after that to 267,000 km and finally to 380,000 km. To execute these maneuvers, the 440 N rocket engine was fired for 18 min, 16 min, 9.5 min, 3 min and 2.5 min, respectively. On November 8, 2008, Chandrayaan-1's rocket engine was fired for 817 s when the spacecraft passed at a distance of about 500 km from the Moon to reduce its velocity such that it entered an orbit around the Moon. This was an elliptical orbit that passed over the polar regions of the Moon; the perilune altitude was 504 km and the apolune altitude 7502 km; the orbital period was about 11 hr. Subsequently, the altitude of the orbit was reduced in steps to achieve a final polar orbit at an altitude of about 100 km. At this time, the spacecraft's mass had decreased to about 600 kg. On November 14, 2008, the 34 kg Moon Impact Probe (MIP) was released from the spacecraft. A retrorocket was fired to decrease its velocity and this started a free fall to the lunar surface. As it fell, it kept sending information back to Chandrayaan-1, which, in turn, beamed the information back to Earth. When the MIP was close to the surface, rockets were fired to slow down its speed and to soften impact. It crash-landed near the crater Shackleton at the south pole. Following the successful deployment of MIP, the other scientific instruments were turned on one-by-one starting the next phase of the two-year mission. On May 19, 2009, the orbit was raised to 200 km altitude. This higher altitude enables further studies on orbit perturbations and lunar gravity field variations, and made it possible to image the lunar surface with a wider swath. After 312 days in orbit, contact with the spacecraft was lost on August 28, 2009.

The computation of precise lunar trajectories can only be done by numerical integration of the equations of motion, taking into account the gravitational attraction by the Earth, Sun and Moon, solar radiation pressure, etc. Because of the complex motion of the Moon (Section 17.1), actual mission planning places heavy reliance on an existing lunar ephemeris. The general procedure is to select the initial conditions, \bar{r}_1 and \bar{V}_1 , of the spacecraft at the injection point in Earth orbit and then use a numerical integration method to determine the subsequent trajectory. Depending on the selected initial conditions, the spacecraft may hit the Moon or miss it entirely. The idea is to adjust the injection conditions by some iterative method, until a suitable lunar trajectory is found. Even on modern high-speed computers this procedure can take many hours of computation time for a single launch date. If we have to explore a large number of launch dates and a variety of injection conditions, some approximate analytical method is needed to narrow down the choice of launch time and injection conditions. That analytical method should in any case retain the predominant features of the actual problem.

The first option for an approximative analytical method is to consider the lunar trajectory to

be shaped only by the gravity field of the Earth and to neglect the effects of the lunar gravitational attraction. It will be clear that this two-body approach (Chapter 5), which will be discussed in Section 17.2, can provide only some crude first-order information about the characteristics of lunar trajectories. As a much better approximation, we may assume that a lunar trajectory is shaped by the gravity fields of the Earth and the Moon, and that the gravity fields of the Sun and the planets produce only perturbations on this trajectory. The physical reason for this is the fact that the *sphere of influence* (Section 4.3) of the Moon is relatively small and falls completely within the sphere of influence of the Earth (Sections 4.3 and 17.3). This is in contrast to the situation for interplanetary trajectories (Chapter 18), where the sphere of influence of the target planet falls completely outside the sphere of influence of the Earth. Within the sphere of influence of the Earth or the Moon, the trajectory is considered as a conic section (two-body problem) about the respective celestial body. So, if we use the *patched-conic approach* (Section 17.3) to approximate a particular Earth-Moon trajectory, two two-body trajectories have to be patched, while an interplanetary trajectory requires the patching of at least three two-body trajectories.

17.1. Earth-Moon system

The radius of the Moon is about 1737 km; the radius of the Earth about 6371 km. Earth and Moon revolve about their common center of mass: the *barycenter*. The mean distance between the center of the Earth and the center of the Moon is 384,401 km. From this distance and the fact that the mass of the Earth is about 81.30 times the mass of the Moon, we find that the barycenter is located at 4671 km from the center of the Earth or about three-quarters of an Earth radius from the center of the Earth. It is common to draw the trajectory of the Earth from the point of view of the Sun, and the trajectory of the Moon from the point of view of the Earth. This could give the impression that the Moon circles around the Earth in such a way that sometimes it goes backwards when viewed from the Sun's perspective. However, since the orbital velocity of the Moon about the Earth is small (≈ 1 km/s) compared to the orbital velocity of the Earth about the Sun (≈ 30 km/s), this actually never occurs.

The barycenter revolves around the Sun once a year. In fact, this period is the definition of the tropical year (Section 11.4). Earth and Moon revolve about their barycenter (relative to inertial space) once in 27.32 days. As a result, the longitude of e.g. the Sun or a near-by planet exhibits fluctuations with a period of 27.32 days, arising from the fact that we observe it from the Earth and not from the center of mass of the Earth-Moon system. These periodic fluctuations in longitude were the most reliable source for determining the Moon's mass, until Ranger 5 flew within 725 km of the surface of the Moon in October 1962. The perturbations produced by the Moon's gravity field on the trajectory of Ranger 5 could be used to determine the mass of the Moon with an, for that time, unprecedented accuracy.

When viewed from the center of the Earth, the Moon's motion may be described by the six classical orbital elements. Primarily due to the perturbing effect of the Sun, the orbital elements are constantly changing with time; their value at any particular time can be obtained from a lunar ephemeris, such as published in the American Ephemeris and Nautical Almanac². In order to

² The first *Nautical Almanac and Astronomical Ephemeris* was published in the UK in 1766. The US Nautical Almanac Office was established in 1849 and the first *American Ephemeris and Nautical Almanac* was published in 1852. In 1960 the title of the UK almanac changed to *The Astronomical Ephemeris*, and the UK and US almanacs became identical in content. In 1981 both publications were combined under the title *The Astronomical Almanac*. In the 1970s, NASA JPL made use of high precision radar, laser and spacecraft observations to develop a complete new

illustrate the complexity, some of the characteristics and principal perturbations of the Moon's motion will be summarized.

The mean value of the semi-major axis is 384,401 km and the average time for the Moon to complete one revolution about the Earth relative to the stars (sidereal month) is 27.32166 days (Section 11.4). Solar gravitational perturbations yield periodic variations of the sidereal month by as much as 7 hr; its mean value, however, remains steady over many centuries to within one second. The mean eccentricity of the Moon's orbit is 0.054900489. Periodic changes in the orbital eccentricity occur at intervals of 31.81 days. This effect, known as the *evection*, was discovered more than 2000 years ago by Hipparchos (~190-120 B.C.) and is due to solar gravitational attraction.

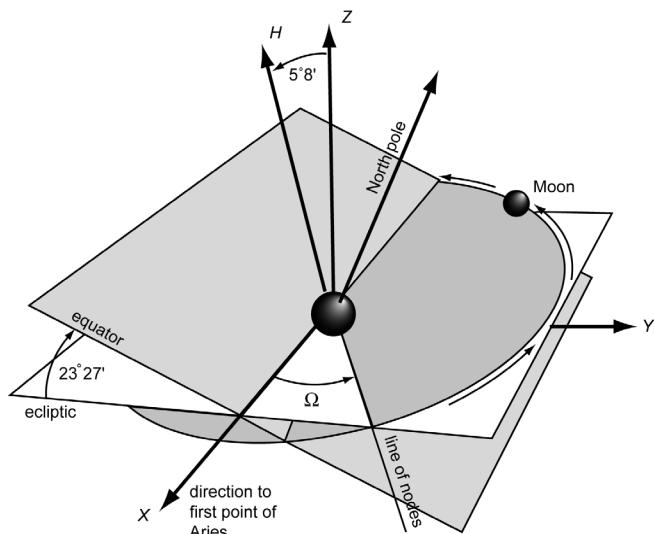


Figure 17.1: Geometry of the Moon's orbit, the equator and the ecliptic, and the precession of the Moon orbit's line of nodes.

Unlike most satellites of other planets, the Moon orbits near the ecliptic and not near the Earth's equatorial plane (Figure 17.1). The orientation of the orbital plane and of the major axis in the orbital plane are not fixed in space, but precess over time. The major axis rotates counterclockwise in the orbital plane (same direction as the Moon itself), and completes a full revolution in 8.8504 years (3,233 days). Newton tried to explain this effect in his *Principia*, but his prediction accounted only for about half of the observed rotation. In 1749, A.C. Clairaut was able to derive the correct result from theory, but more than a century later, in 1872, the correct calculations were also discovered among Newton's unpublished papers. This indicates that Newton had detected his own error, but had never bothered to correct it in print. When viewed from celestial north, the orbital plane itself rotates clockwise about an axis perpendicular to the ecliptic. So, the line of nodes (in the ecliptic, Figure 17.1) precesses with time, completing one revolution in 18.5996 years (6,793.5 days). This period is close to the period of the Saros cycle, which was

set of orbit calculations for the Moon and planets. These were introduced in The Astronomical Almanac for 1984. In the 1970s, the offices began exploring means for presenting astronomical data in computerized formats. The first effort was the *Almanac for Computers*; it was published annually for the years 1977 through 1991. The *Floppy Almanac*, an almanac program for early personal computers, was produced for the years 1986 through 1996. By the late 1980s, evolving computer technology created the demand for a new computer almanac. This has resulted in the *Multiyear Interactive Computer Almanac* (MICA) that was first released in 1993. The reference for the definitions of models and transformations is the *Explanatory Supplement to The Astronomical Almanac*, P. K. Seidelmann (ed.), University Science Books, 1992.

discussed in Section 11.4. The inclination of the Moon's orbit with respect to the ecliptic varies between 4.99° and 5.30° ; its mean value is 5.145396° . Both the slight variation in inclination relative to the ecliptic and the regression of the line of nodes were first observed by J. Flamsteed (1646-1719) in about 1670. The lunar axis of rotation maintains an inclination of 1.5424° with respect to the normal to the ecliptic plane, and the lunar equator is inclined to its orbital plane by a constant value of 6.6878° (*obliquity*). One might be tempted to think that as a result of the precession of the Moon's orbital plane, the angle between the lunar equator and the ecliptic would vary between the sum (11.8332°) and difference (1.5424°) of the inclination and the obliquity. However, as was discovered by J. Cassini (1677-1756) in 1721, the rotation axis of the Moon precesses with the same rate as the Moon's orbital plane, but is 180° out of phase. So, although the rotation axis of the Moon is not fixed with respect to the stars, the angle between the ecliptic and the lunar equator is always 1.5424° . So, during a full rotation of the orbital plane, the inclination with respect to the Earth's equator, which is itself inclined at 23.44° to the ecliptic, varies between $23.44^\circ + 5.14^\circ = 28.58^\circ$ and $23.44^\circ - 5.14^\circ = 18.30^\circ$. Simultaneously, the axis of lunar rotation sees its tilt with respect to the Moon's orbital plane vary between $5.14^\circ + 1.54^\circ = 6.68^\circ$ and $5.14^\circ - 1.54^\circ = 3.60^\circ$. Note that the Earth's rotation axis reacts to this process and itself varies by 0.00256° on either side of its mean value; this is the *nutation* discussed in Section 11.2. Once every 18.5996 years, the declination of the Moon reaches a maximum of 28.5847° . Some exact times for the occurrence of this maximum declination are: March 29, 1969, 02^h45^m UTC; November 8, 1987, 11^h45^m UTC; June 19, 2006, 20^h48^m UTC; January 29, 2025, 05^h56^m UTC.

The Moon's gravity field raises tides on Earth displacing the Earth's water outward by about 1 m in the open oceans and by as high as 18 m in some coastal waters, like the Bay of Fundy; it also raises tides in the Earth's solid surface by about 30 cm. The *tidal bulges* are dragged along with the Earth's surface and since the Earth spins faster about its axis than the Moon moves around the Earth, the bulges are not along the Earth-Moon line but are displaced somewhat 'in front of' the Moon. As a result, the tidal bulges generate a gravitational force on the Moon with a component along the lunar orbit in the Moon's 'forward' direction; the same force acts on the Earth and has a component in a direction opposite to the Earth's rotation. Because the force on the Moon acts in its direction of motion, the Moon moves to a higher orbit resulting in a decrease of its mean angular motion about the Earth. Measurements reveal that the Moon's distance to the Earth presently increases by 38.4 mm/yr, leading to a secular deceleration of the Moon of $25.858''/\text{century}^2$ in ecliptic longitude. The Earth's rotation rate is gradually decreasing; the present rate is about $-5.5 \times 10^{-22} \text{ rad/s}^2$ (Section 11.4), leading to a lengthening of the day by about 23 μs every year. From historical records over the past 2700 years an average value of 17 $\mu\text{s}/\text{yr}$ for the lengthening of the day is found. The difference in both values can be explained largely by changes in the Earth's moments of inertia due to the so-called *post-glacial rebound*. The present lengthening of the day would make each Earth day one minute (current length of minute) longer every three million years. Looking back and adopting the historical value of 17 $\mu\text{s}/\text{yr}$, the day was a mere 23.5 hr in length when the dinosaurs roamed the Earth 100 million years ago and the terrestrial day was approximately 20 hr about 850 million years ago. The tidal bulges on the Moon, generated by the Earth, produce a decrease of the Moon's rotation rate. This has lead to the present situation where the Moon presents only one face toward the Earth. Most significant moons in the solar system are *tidally locked* with their primaries, since they orbit quite closely their planets. Notable exceptions are the irregular outer moons of the giant outer planets, which orbit much further away than the large moons of these planets. Eventually, the Earth and the Moon would become caught up in what is called a *spin-orbit resonance*, in which the Moon would encircle the Earth in about 47 days, and both Moon and Earth would rotate around their

axes in the same time, always facing each other with the same side. However, long before this would happen the Sun will evolve into a red giant and will destroy both the Earth and the Moon. Pluto and Charon are already tidally locked in spin-orbit resonance.

If the Moon's orbit were circular and if its axis of rotation were perpendicular to its orbit, we would see exactly half of the Moon's surface. Actually, we see, at one time or another, about 59% of the lunar surface from the Earth, because of a phenomenon known as *geometric lunar libration*. This libration has three causes. The *geometric libration in latitude* occurs because the Moon's equator is inclined 6.7° to the plane of its orbit. At one time during the month the Moon's north pole is tipped toward the Earth and half a month later the south pole is tipped toward the Earth, allowing us to see slightly beyond each pole. The *geometric libration in longitude* is due to the eccentricity of the orbit. The rotation of the Moon about its axis is uniform, but the angular velocity around its orbit is not, since the Moon moves faster near perigee and slower near apogee. This permits us to see about 6.3° around each limb of the Moon. The third type of geometric libration has to do with the fact that an observer's location on Earth will change his perspective. An observer far north sees more of the Moon's northern part, while one observing the Moon at moonrise sees more of the leading limb. This effect alters the coordinates of the apparent center by up to about $\pm 1^\circ$. In addition to these types of geometric librations, there is an actual libration called *physical libration*. This phenomenon, with a maximum amplitude of about 0.03° , is caused by the attraction of the Earth on the non-spherical Moon.

17.2. Two-body lunar trajectories

As mentioned before, the simplest approach to analyze a lunar flight is to neglect the gravitational attraction by the Moon and to consider the trajectory as a conic section about the Earth that intersects the lunar orbit. This approach is only useful to gain some insight into the problem of selecting optimum launch dates and approximate injection conditions, and to estimate the flight

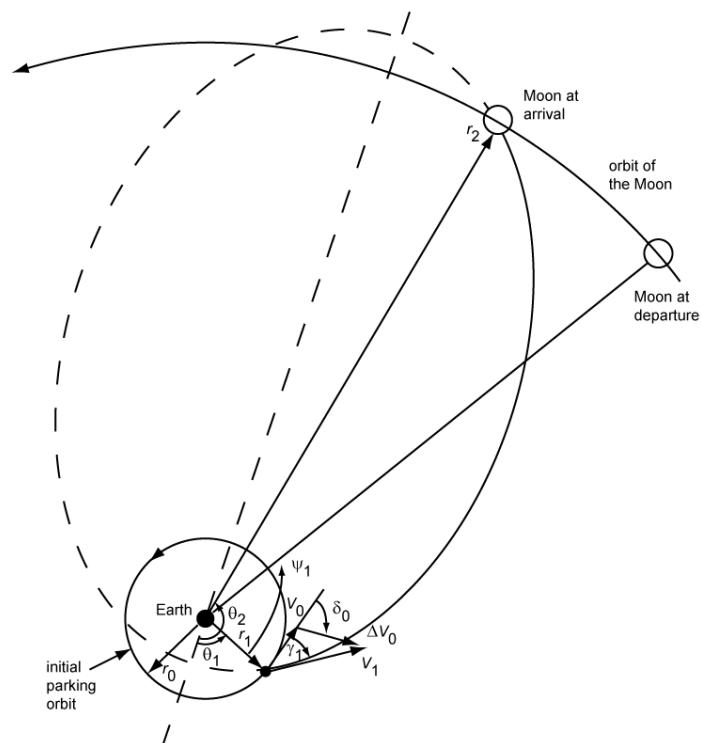


Figure 17.2: Geometry of a two-body lunar trajectory and a definition of various parameters.

times involved in lunar missions. For this simplified case, we assume that the Moon's orbit is circular with a radius of 384,401 km and that the lunar trajectory of the spacecraft is coplanar with the Moon's orbit. It is emphasized that in a precision trajectory computation the launch time is selected such that this coplanarity is approximately true in order to minimize the ΔV required for the mission, since plane changes are expensive in terms of propellant consumption (Chapter 13). A sketch of the conic trajectory is shown in Figure 17.2, where it has been assumed that the mission starts from a circular parking orbit with radius r_0 about the Earth. The Figure also defines some parameters that are used in the following analysis. In Section 17.3, we will extend the analysis by taking into account the terminal attraction by the Moon, although we will still consider the motion as two-dimensional.

In the following, we will use the index 1 to denote the conditions at the start of the lunar trajectory, just after injection into this trajectory. As the injection is considered as an impulsive shot: $r_1 = r_0$. To compute the orbit for adopted values of r_1 , V_1 , γ_1 , we apply a series of relations derived in Chapters 5 and 6. The energy and angular momentum of the spacecraft, both per unit of mass, follow from

$$\mathcal{E} = \frac{1}{2} V_1^2 - \frac{\mu}{r_1} \quad (17.1)$$

$$H = r_1 V_1 \cos \gamma_1 \quad (17.2)$$

where μ is the gravitational parameter of the Earth. The semi-latus rectum, semi-major axis and eccentricity of the trajectory may then be obtained from

$$p = \frac{H^2}{\mu} \quad (17.3)$$

$$a = -\frac{\mu}{2\mathcal{E}} \quad (17.4)$$

$$e = \sqrt{1 - \frac{p}{a}} \quad (17.5)$$

Solving the equation of a conic section for the true anomaly, we get

$$\cos \theta = \frac{p - r}{er} \quad (17.6)$$

When we substitute r_1 for r in this equation, we can solve for θ_1 ; when we substitute the radius of the Moon's orbit, r_M , for r , we can find the true anomaly upon arrival at the Moon's orbit, θ_2 . It is emphasized that, in general, (17.6) has two solutions for a given value of r , corresponding to different flight times along the same trajectory. Usually, for an actual mission $90^\circ < \theta_2 \leq 180^\circ$. We now have enough information to determine the time-of-flight, t_f , from Earth to Moon for any set of injection conditions using the equations derived in Section 6.5 for elliptical orbits, Section 7.3 for parabolic orbits or Section 8.3 for hyperbolic orbits. As an example, only the computation for the case of an elliptical orbit will be given below.

We first compute the eccentric anomalies at departure, E_1 , and arrival, E_2 , from

$$E = 2 \arctan \left[\sqrt{\frac{1-e}{1+e}} \tan \frac{\theta}{2} \right]$$

where we substitute θ_1 for θ to obtain E_1 , and θ_2 for θ to obtain E_2 . The values of E_1 and E_2 may be determined uniquely by requiring that $\theta/2$ and $E/2$ lie in the same quadrant (Section 6.5). The time-of-flight then is obtained from

$$t_f = \sqrt{\frac{a_3}{\mu}} [E_2 - E_1 - e(\sin E_2 - \sin E_1)] \quad (17.7)$$

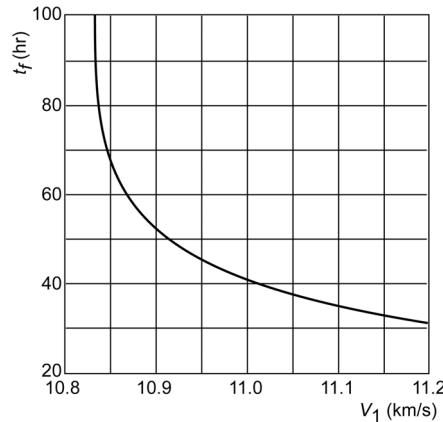


Figure 17.3: Flight time in a two-body lunar trajectory versus injection velocity, for $r_i = 6695$ km and $\gamma_i = 0^\circ$.

In Figure 17.3, t_f is plotted as a function of the injection velocity, V_1 , for an injection altitude of 320 km above the Earth and an initial flight path angle of $\gamma_1 = 0^\circ$. Actually, for small flight path angles the curve is nearly independent of the flight path angle at injection. Note that in this Figure the maximum flight time is 100 hr, and that a significant reduction in flight time is possible with only a modest increase in V_1 . For manned missions, the life-support requirements increase with mission duration, so the slightly higher injection velocities required to achieve a shorter flight time pays for itself up to a point. Requirements on the arrival conditions may also ask for faster transfer orbits. It is interesting to note that the flight time (between lunar trajectory injection and insertion into lunar orbit) chosen for the first Apollo lunar landing mission (Apollo 11) was 73.1 hr; for later flights trajectories with flight times between 71.9 hr (Apollo 16) and 83.0 hr (Apollo 17) were flown. Luna-1, the first mission to the Moon, flew a 34 hr transfer in January 1959; Luna 9, the first spacecraft to perform a soft landing on the Moon on February 3, 1966, used a 79 hr transfer; Lunar Prospector flew a 105 hr transfer in January 1998. In order to expand the launch window, several missions have flown added phasing loops (Chapter 14) before the final transfer from the Earth to the Moon was initiated. These phasing loops also allow for lower ΔV corrections, since many corrections can be applied at apogee. Clementine (January 1994) e.g. used 2.5 phasing loops, which could be adjusted in orbital period with perigee maneuvers to allow for more launch days per month while maintaining the same arrival day to meet specified lunar lighting conditions.

If we assume that injection into the lunar trajectory occurs at perigee, i.e. $\theta_1 = \gamma_1 = 0^\circ$, then the minimum-energy lunar trajectory is a *Hohmann trajectory* (Section 12.1), which is at apogee tangent to the Moon's orbit. This Hohmann transfer orbit requires the minimum injection V_1 and

yields the maximum time-of-flight. For an injection altitude of 320 km, this minimum injection velocity is 10.82 km/s; the maximum time-of-flight is 7172 min, or about 119 hr. The eccentricity of the minimum-energy trajectory is 0.966 and represents the minimum eccentricity for an elliptical orbit that reaches as far as the Moon, for the given injection altitude. Assuming no lunar gravity field, the velocity upon arrival at the Moon's orbit along the minimum-energy trajectory is 0.188 km/s. Since the Moon's orbital velocity is about 1.02 km/s, the Moon would literally run into our spacecraft from 'behind', resulting in an impact on the 'leading edge' (western limb) of the Moon. Spacecraft with a higher arrival speed would have $\gamma_2 > 0^\circ$ and would tend to impact somewhere on the side of the Moon facing us. The angle traversed by the vehicle in its lunar transfer trajectory, which is called the *geocentric transfer angle*, $\Delta\theta = \theta_2 - \theta_1$ (Figure 17.2), is a function of the injection velocity for a fixed injection altitude and flight path angle. An increase in $\Delta\theta$ (up to 180°) corresponds to a decrease in injection velocity. Therefore, if we are trying to minimize injection velocity, as is usually the case, we should try to select a trajectory that has a transfer angle of nearly 180° . We will use this general principle later in this Chapter.

In trying for a direct hit on the Moon, we should time the launch such that the spacecraft crosses the Moon's orbit just at the time the Moon is at the crossing point. Using our simplified model, we may obtain some idea of how much we will miss the center of the Moon if, due to guidance errors or other factors, the injection conditions are not exactly as specified. In such a case, both the geocentric transfer angle and the time-of-flight will differ from their nominal values and the trajectory of the spacecraft will cross the Moon's orbit at a different point and time than predicted. In the case of a direct (eastward) launch, the effects of such injection errors tend to cancel, as may be concluded from Figure 17.4. If, for example, the initial velocity is too high, the geocentric transfer angle will be smaller than predicted; i.e. the spacecraft will cross the Moon's orbit east of the predicted point by some small amount. But, the time-of-flight will be shorter, so the Moon will be east of the predicted interception point. Neglecting the effects of the lunar gravity field, the angular miss distance along the Moon's orbit is the difference of the two opposite effects. It may be demonstrated that for an injection velocity of about 11 km/s, which results in a transfer angle of about 160° , the effects of errors in injection altitude and velocity exactly cancel and, for all practical purposes, the miss distance at the Moon is a function only of errors in the flight path angle. For this condition, an error of 1° in flight path angle produces a miss distance of about 1300 km. It is interesting to note that the effect of lunar gravity is to reduce this miss distance.

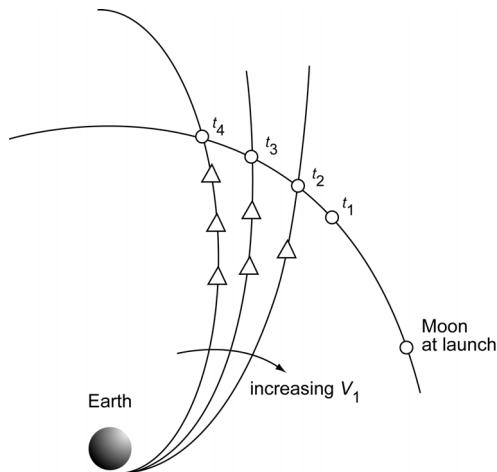


Figure 17.4: Effects of injection errors tend to cancel.

17.3. Patched-conic two-dimensional lunar trajectories

While it is acceptable to neglect lunar gravitational attraction when we only are interested in determining approximate injection conditions that result in particular conditions at lunar arrival, it is necessary to account for the terminal attraction of the Moon if we want to predict the lunar arrival conditions more precisely. We may take lunar gravitational attraction into account and still use two-body orbital mechanics by the simple expedient of considering the spacecraft to be under the gravitational attraction by the Earth alone until it enters the *sphere of influence* (Section 4.3) of the Moon and assuming that within this sphere it moves only under the gravitational attraction by the Moon. It should be clear that this process still is an approximation. The transition from geocentric motion to selenocentric motion actually is a gradual process, which takes place over a finite arc of the trajectory where both Earth and Moon affect the trajectory. However, the simple strategy of patching two conics together at the edge of the Moon's sphere of influence turns out to be a sufficiently good approximation for preliminary mission analyses. On the other hand, it should be kept in mind that this *patched-conic analysis* is certainly not good enough for the computation of the precise perilune altitude and the trajectory after leaving the Moon's sphere of influence again. For the analysis that follows, we will adopt the definition of the sphere of influence as suggested by P.S. Laplace (Section 4.3), which leads to a value of $R_{s,i} = 66,183$ km for the radius of the Moon's sphere of influence. This is about one-sixth of the distance from the Moon to the Earth. Given the fact that the radius of the Earth's sphere of influence is about 10^6 km (Section 4.3), we conclude that the Moon's sphere of influence falls completely within the sphere of influence of the Earth, as was already mentioned before.

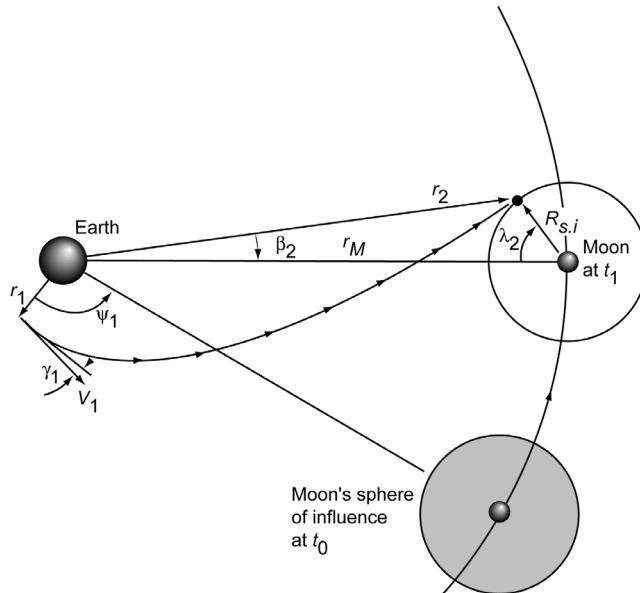


Figure 17.5: Geometry of a patched-conic lunar trajectory and the definition of various parameters.

Figure 17.5 shows the geometry of the geocentric departure trajectory. Just as in the previous Section, we assume that the lunar flight starts from a circular parking orbit about the Earth, with radius r_0 , and that the spacecraft's lunar transfer trajectory is coplanar with the Moon's orbit. As the injection is considered as an impulsive shot, the geocentric distance of the spacecraft just after the impulsive shot, r_1 , is equal to r_0 . The four quantities that completely specify the geocentric phase of the trajectory are r_1 , V_1 , γ_1 and ψ_1 , where ψ_1 is the lunar relative position angle at departure; this angle is generally called the *lunar phase angle at departure*. It defines the moment

of time when the impulsive shot is applied. The difficulty with these four quantities as independent variables is that the determination of the point where the geocentric trajectory crosses the sphere of influence of the Moon involves an iterative procedure in which the time-of-flight must be computed during each iteration. Therefore, it is more convenient to select three initial conditions and one arrival condition as the independent variables. A particularly convenient set is r_1 , ΔV_0 , δ_0 , λ_2 , where ΔV_0 is the magnitude of the impulsive shot in the parking orbit about the Earth, δ_0 is the thrust angle (angle between ΔV_0 and the normal to \bar{r}_0 (Figure 17.2)), and λ_2 specifies the point where the geocentric trajectory crosses the Moon's sphere of influence. Given these four quantities, we can compute the remaining arrival conditions on the Moon's sphere of influence: r_2 , V_2 , γ_2 and β_2 . We assume that, in case the geocentric trajectory is elliptical, lunar arrival occurs prior to apogee of the transfer trajectory; so, $\theta_2 < 180^\circ$.

The relations between the initial conditions in the transfer trajectory and the parameters defining the applied impulsive shot in the parking orbit, ΔV_0 and δ_0 , are (Figure 17.2)

$$\begin{aligned} (\Delta V_0)^2 &= V_0^2 + V_1^2 - 2V_0 V_1 \cos \gamma_1 \\ V_1^2 &= V_0^2 + (\Delta V_0)^2 + 2V_0 \Delta V_0 \cos \delta_0 \end{aligned} \quad (17.8)$$

where V_0 is the velocity in the circular parking orbit. For a selected set of initial conditions, (17.8) yields the values of V_1 and γ_1 . In determining the value of γ_1 , it should be realized that if $\delta_0 > 0^\circ$: $\gamma_1 > 0^\circ$, and if $\delta_0 < 0^\circ$: $\gamma_1 < 0^\circ$. The energy, \mathcal{E} , and angular momentum, H , of the trajectory, both per unit of mass, can then be determined from (17.1) and (17.2). From the law of cosines, the distance r_2 at arrival on the Moon's sphere of influence is given by (Figure 17.5)

$$r_2 = \sqrt{r_M^2 + R_{s,i}^2 - 2r_M R_{s,i} \cos \lambda_2} \quad (17.9)$$

where r_M is the (mean) distance between the center of the Earth and the center of the Moon (radius of the Moon's orbit). The velocity and flight path angle at arrival on the sphere of influence then follow from the conservation of energy and angular momentum:

$$V_2 = \sqrt{2\left(\mathcal{E} + \frac{\mu}{r_2}\right)} \quad ; \quad \cos \gamma_2 = \frac{H}{r_2 V_2} \quad (17.10)$$

where γ_2 is known to lie between 0° and 90° since we have assumed that arrival occurs prior to apogee passage. In that case, (17.10-2) yields a unique solution for γ_2 . According to (17.9), the geocentric distance at arrival, r_2 , is completely determined by λ_2 . It may happen that the trajectory is not sufficiently energetic to reach the specified point on the sphere of influence as determined by λ_2 . If so, the quantity under the square-root sign in (17.10) will be negative and the computation process fails. Finally, from geometry (Figure 17.5) we find

$$\sin \beta_2 = \frac{R_{s,i}}{r_2} \sin \lambda_2 \quad (17.11)$$

We now have the conditions at arrival on the Moon's sphere of influence for the chosen values of r_1 , ΔV_0 , δ_0 and λ_2 . These arrival conditions determine the trajectory within the sphere of influence, and we will proceed with the analysis of that part of the flight.

In Figure 17.6 the geometry at arrival on the Moon's sphere of influence is shown in detail. Note that \bar{V}_2 , generally, is not perpendicular to \bar{r}_3 , as the Figure seems to suggest. If we use the

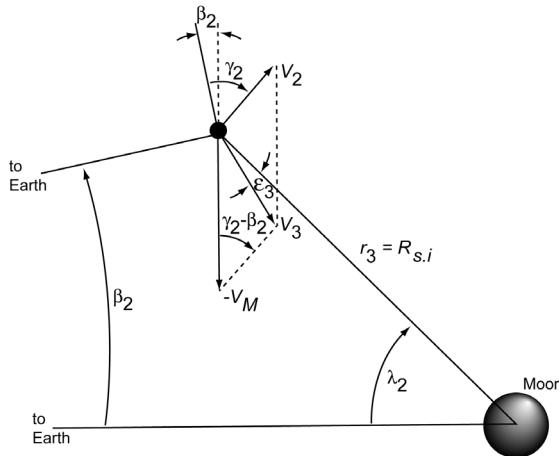


Figure 17.6: Geometry of the arrival on the Moon's sphere of influence and a definition of various parameters.

subscript 3 to indicate the initial conditions relative to a non-rotating reference frame centered at the Moon, then the selenocentric distance at the entrance of the sphere of influence is given by $r_3 = R_{s.i}$. The velocity of the spacecraft at that moment relative to the Moon-centered reference frame is given by

$$\bar{V}_3 = \bar{V}_2 - \bar{V}_M$$

where \bar{V}_M is the velocity of the Moon relative to a non-rotating Earth-centered reference frame. For our simplified Earth-Moon model, the orbital velocity of the Moon is $V_M = 1.018 \text{ km/s}$. The magnitude of the selenocentric arrival velocity, V_3 , may be obtained by applying the law of cosines to the vector triangle in Figure 17.6:

$$V_3 = \sqrt{V_2^2 + V_M^2 - 2V_2 V_M \cos(\gamma_2 - \beta_2)} \quad (17.12)$$

The angle ϵ_3 in Figure 17.6 defines the direction of \bar{V}_3 relative to the line connecting the spacecraft when entering the Moon's sphere of influence and the Moon's center. A detailed analysis of the geometry indicated in Figure 17.6 shows that we may write for the component of \bar{V}_3 perpendicular to \bar{r}_3

$$V_3 \sin \epsilon_3 = V_M \cos \lambda_2 - V_2 \cos(\lambda_2 + \beta_2 - \gamma_2)$$

from which we find

$$\epsilon_3 = \arcsin \left[\frac{V_M}{V_3} \cos \lambda_2 - \frac{V_2}{V_3} \cos(\lambda_2 + \beta_2 - \gamma_2) \right] \quad (17.13)$$

According to the definition of ϵ_3 in Figure 17.6, the spacecraft will move about the Moon in an anti-clockwise direction if $\epsilon_3 > 0^\circ$; if $\epsilon_3 < 0^\circ$ the spacecraft's motion about the Moon will be clockwise; for a dead-center hit on the Moon, ϵ_3 should be zero.

Now that the selenocentric initial conditions r_3 , V_3 and ϵ_3 are known, we can compute the conditions at any other point along the selenocentric trajectory, including the point of impact on the Moon or the point of closest approach: *perilune*. Three important classes of lunar missions can be discerned:

- *Lunar impact*, in which case the perilune distance, r_p , is less than the radius of the Moon, R_M , and we may wish to compute the spacecraft's velocity at impact.

- *Lunar orbit*, in which case we may wish to compute the ΔV needed to enter a circular lunar orbit at perilune altitude.
- *Circumlunar flight*, in which case we probably would want to compute both the perilune conditions and the conditions upon exit from the Moon's sphere of influence.

In general, the conditions at perilune will be certainly of interest, and we will continue with the computation of these parameters.

The energy and angular momentum relative to the Moon-centered reference frame, both per unit of mass, are given by

$$\mathcal{E} = \frac{1}{2} V_3^2 - \frac{\mu_M}{r_3} ; \quad H = r_3 V_3 \sin e_3 \quad (17.14)$$

where μ_M is the gravitational parameter of the Moon. The semi-latus rectum and the eccentricity of the selenocentric orbit can then be computed from (Sections 5.3 and 6.1)

$$p = \frac{H^2}{\mu_M} ; \quad e = \sqrt{1 + 2 \frac{H^2 \mathcal{E}}{\mu_M^2}} \quad (17.15)$$

The conditions at perilune are then obtained from

$$r_p = \frac{p}{1+e} ; \quad V_p = \sqrt{2 \left(\mathcal{E} + \frac{\mu_M}{r_p} \right)} \quad (17.16)$$

and the deceleration impulsive shot, ΔV_p , to enter a circular orbit about the Moon with radius equal to the perilune distance of the approach trajectory from

$$\Delta V_p = V_p - \sqrt{\frac{\mu_M}{r_p}} \quad (17.17)$$

If the adopted injection conditions do not result in a satisfactory lunar approach trajectory, we have to adjust the values of r_1 , ΔV_0 , δ_0 and λ_2 in a systematic search procedure.

If the arrival conditions are satisfactory, we proceed by computing the values of p , a and e of the geocentric trajectory from (17.3) to (17.5). Then, we determine the values of θ_1 and θ_2 from (17.6), where we assume $-90^\circ < \theta_1 < 90^\circ$, $\theta_2 < 180^\circ$. Note that for determining the value of θ_2 , the value of r_2 , i.e. the geocentric distance upon entering the Moon's sphere of influence, has to be substituted in (17.6). We then apply the classical relations between time and position in the trajectory, as given in Section 6.5 for elliptical orbits, in Section 7.3 for parabolic orbits, and in Section 8.3 for hyperbolic orbits, to determination the time-of-flight, $t_f = t_2 - t_1$, from injection to arrival on the sphere of influence.

The semi-major axis of the hyperbolic selenocentric trajectory can be computed from (17.5) by substituting the values of p and e obtained from (17.15). From (17.6) we find

$$\cos \theta_3 = \frac{p - r_3}{e r_3} ; \quad \theta_p = 0$$

where θ_3 is the true anomaly in the selenocentric trajectory at entrance of the sphere of influence and θ_p is the true anomaly at perilune. With (8.20), we compute the hyperbolic anomaly at entrance of the sphere of influence, F_3 , from θ_3 ; the hyperbolic anomaly at perilune is zero. Then, we compute the time-of-flight from entering the sphere of influence to passing perilune from

(8.23):

$$t_{f_M} = \sqrt{\frac{-\alpha^3}{\mu_M}} (e \sinh F_3 - F_3) \quad (17.18)$$

The total flight time, from departure in the parking orbit about the Earth to arrival at perilune, is given by

$$t_{f_{tot}} = t_f + t_{f_M} \quad (17.19)$$

The Moon moves through an angle $n_M t_f$ between injection and arrival on the Moon's sphere of influence, where n_M is the angular velocity of the Moon in its orbit about the Earth. The lunar phase angle at departure, Ψ_1 , can then be determined from (Figures 17.2 and 17.5)

$$\Psi_1 = \theta_2 - \theta_1 - \beta_2 - n_M t_f \quad (17.20)$$

To get an impression of the general characteristics of such two-dimensional lunar trajectories, a numerical experiment was performed, where the values of r_M , μ_M and n_M were taken from Appendix B or computed from relevant data in Appendix B. It was assumed that the lunar flight starts in a circular parking orbit about the Earth at an altitude of 200 km. The value of ΔV_0 was varied from 3 km/s to 5 km/s in steps of 10 m/s; the value of δ_0 was varied from -20° to 20° in steps of 0.1° ; the value of λ_2 was varied from 0° to 90° in steps of 0.1° . The reason for selecting only positive values of λ_2 is that geometric considerations show that $\lambda_2 < 0^\circ$ may only occur for very high values of ΔV_0 . Of all about 72.4 million trajectories thus computed, only those with a perilune altitude between 499.5 km and 500.5 km were stored. It was found that 1177 trajectories satisfied this constraint, and that all but 99 of the possible geocentric trajectories are hyperbolic; elliptical orbits only occur for $\Delta V_0 \leq 3.31$ km/s. Figure 17.7 shows for these 1177 trajectories the Moon's sphere of influence crossing angle, λ_2 , lunar transfer trajectory eccentricity, e , geocentric transfer angle, $\Delta\theta$, deceleration impulsive shot, ΔV_p , and total flight time, $t_{f,tot}$, as a function of the injection impulsive shot, ΔV_0 ; and the injection impulsive shot, ΔV_0 , sphere of influence crossing angle, λ_2 , geocentric transfer angle, $\Delta\theta$, and total flight time, $t_{f,tot}$, as a function of the lunar phase angle at departure, Ψ_1 . The plots for λ_2 versus ΔV_0 and λ_2 versus Ψ_1 clearly show two bands for $\Delta V_0 > 3.3$ km/s and $\Psi_1 < 130^\circ$, respectively. The upper band corresponds to orbits of spacecraft that move in a clockwise direction about the Moon ($\varepsilon_3 < 0^\circ$) and the lower band to orbits of spacecraft that move in an anti-clockwise direction about the Moon ($\varepsilon_3 > 0^\circ$). It was found that only solutions exist for $3.7^\circ < \lambda_2 < 79.1^\circ$. For all solutions, $\Delta\theta$ varies between 115° and 188° ; the total flight time, $t_{f,tot}$, from 15.3 hr to 125.1 hr. The flight time to the moment of entrance into the Moon's sphere of influence, t_f , was found to vary from 12.6 to 105.6 hr; the flight time within the sphere of influence, t_{f_M} , from 2.7 to 19.6 hr; the selenocentric velocity at entrance of the sphere of influence, V_3 , from 0.81 to 6.67 km/s; the angle ε_3 from ± 2.0 to $\pm 5.3^\circ$. The perilune velocity, V_p , was found to vary from 2.21 to 6.98 km/s; the deceleration impulsive shot, ΔV_p , from 0.73 to 5.50 km/s.

Many interesting conclusions can be drawn from these plots; here only a few are given:

- For $\Delta V_0 > 3.5$ km/s, the spacecraft enters the Moon's sphere of influence at $\lambda_2 < 18^\circ$; for smaller values of ΔV_0 the value of λ_2 increases up to $\lambda_2 = 79.1^\circ$.
- The eccentricity of the geocentric trajectory increases about linearly with increasing injection impulse, ΔV_0 .
- For $\Delta V_0 > 4.5$ km/s, the geocentric transfer angle, $\Delta\theta$, ranges from 113° to 145° ; for $\Delta V_0 < 3.5$ km/s, this angle ranges from 143° to 188° .

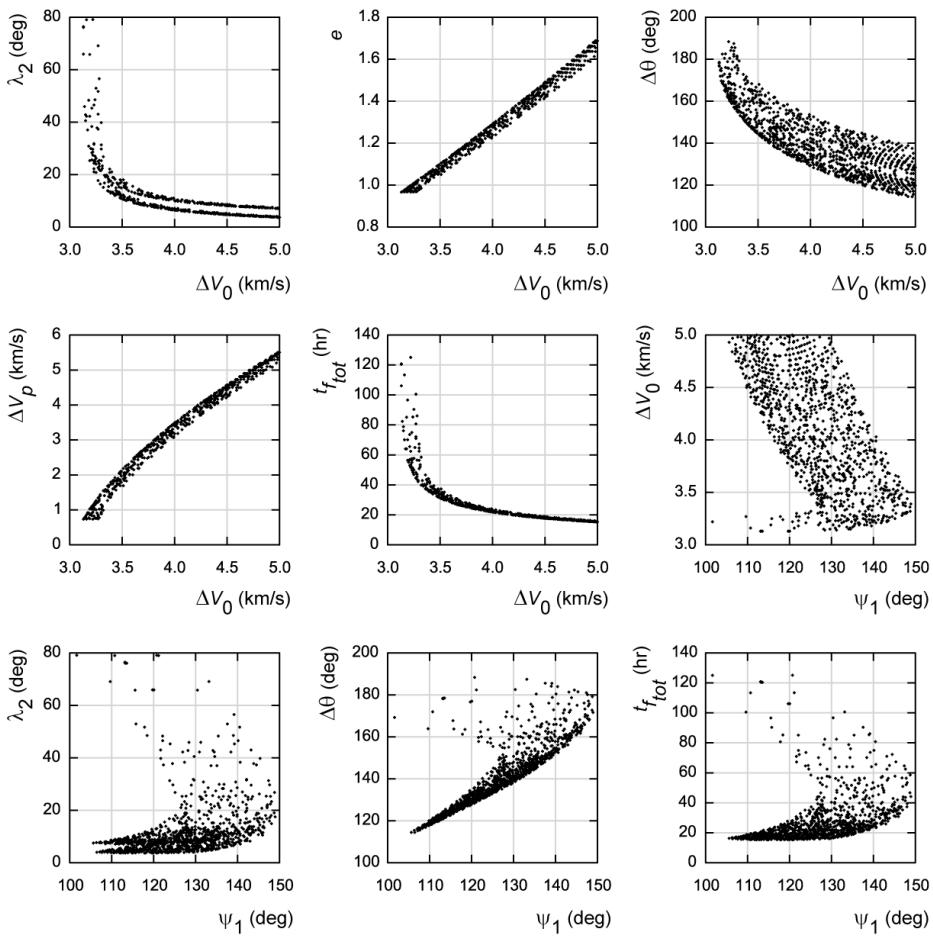


Figure 17.7: Dependence of various parameters of two-dimensional lunar trajectories that start from a 200 km altitude circular parking orbit about the Earth and approach the Moon at a perilune altitude between 499.5 km and 500.5 km, as a function of the injection impulsive shot, ΔV_0 , and the lunar phase angle at departure, ψ_1 .

- The partial derivative $\partial V_p / \partial V_0$ is always larger than one, which means that at increasing values of the departure impulsive shot, ΔV_0 , the deceleration impulsive shot, ΔV_p , increases faster.
- Launch should occur at a lunar phase angle at departure, ψ_1 , of $100\text{--}150^\circ$; for higher injection impulsive shots the value of ψ_1 , generally, is smaller.
- The plots for the variation of ΔV_0 , λ_2 , $\Delta\theta$, and $t_{f,tot}$ versus launch time, expressed through the angle ψ_1 , show an interesting ‘curved band’ structure, indicating that multiple solutions exist for a particular value of ψ_1 .

From this numerical experiment it was found that the injection conditions leading to the smallest value of ΔV_0 are $\Delta V_0 = 3.13 \text{ km/s}$, $\delta_0 = \pm 0.1\text{--}0.3^\circ$, $\psi_1 = 113\text{--}120^\circ$. To find the true minimum-energy transfer trajectory, a refined numerical searching procedure was performed by systematically varying the values of the independent variables ΔV_0 , δ_0 , λ_2 in steps of 1 m/s for ΔV_0 , and in steps of 0.05° for δ_0 and λ_2 . The minimum value of ΔV_0 was found to be 3.130 km/s. A total of six different trajectories for this minimum ΔV_0 were found to exist; they all cross the Moon’s sphere of influence at $\lambda_2 = 75.6\text{--}76.4^\circ$ and they all yield a total flight time of 119.7–120.8 hr. It is interesting to note that this total flight time is close to the flight time of the simplified direct Hohmann trajectory discussed in Section 17.2. The parameters of one of these minimum-energy trajectories are listed in Table 17.1.

Table 17.1: Parameters of a two-dimensional minimum-energy trajectory from departure in a 200 km altitude circular parking orbit about the Earth to injection into a 500 km altitude circular orbit about the Moon. Listed are the parameters of the transfer trajectory to interception of the Moon's sphere of influence, and the parameters of the trajectory within that sphere of influence.

Injection parameters: $\Delta V_0 = 3.130 \text{ km/s}$, $\delta_0 = 0.30^\circ$, $\psi_1 = 113.21^\circ$	
<i>Transfer trajectory</i>	<i>Trajectory about Moon</i>
$a = 192,582 \text{ km}$	$a = -9605 \text{ km}$
$e = 0.96584$	$e = 1.23299$
$V_1 = 10.914 \text{ km/s}$	$V_3 = 0.811 \text{ km/s}$
$\gamma_1 = 0.09^\circ$	$\varepsilon_3 = 5.28^\circ$
$\theta_1 = 0.18^\circ$	$\theta_3 = 138.58^\circ$
$\theta_2 = 178.32^\circ$	$h_p = 499.8 \text{ km}$
$\lambda_2 = 76.10^\circ$	$V_p = 2.212 \text{ km/s}$
$V_2 = 0.248 \text{ km/s}$	$\Delta V_p = 0.732 \text{ km/s}$
$\gamma_2 = 39.28^\circ$	$t_{fM} = 19.6 \text{ hr}$
$t_f = 100.7 \text{ hr}$	

If the spacecraft would not be decelerated when passing perilune, the spacecraft would continue its motion along its selenocentric hyperbolic trajectory. When it crosses the Moon's sphere of influence again, its velocity relative to the Moon has again the magnitude V_3 , but the direction of this velocity vector will, generally, differ from the direction of the velocity vector at entrance of the Moon's sphere of influence. Consequently, the velocity of the spacecraft relative to the Earth-centered reference frame at the point where it leaves the Moon's sphere of influence will differ from the velocity relative to the Earth-centered reference frame at the point where it enters the sphere of influence. In other words: passing the Moon at a relatively short distance will change the orbital energy of the spacecraft relative to the Earth-centered reference frame. This is an example of the so-called *gravity-assist* or *swingby* effect that will be discussed in Section 18.11.

17.4. Three-dimensional lunar trajectories

The preceding analysis was based on the assumption that the lunar trajectory lies in the plane of the Moon's orbit. As was already mentioned, the inclination of the Moon's orbit relative to the Earth's equator varies between about 18.3° and 28.6° over a period of 18.6 years. So, the orbital plane of a lunar transfer trajectory originating from Cape Canaveral (ETR) at a latitude of 28.5° , can, without executing dogleg maneuvers (Section 16.2), only be coplanar with the Moon's orbital plane if the inclination of the Moon's orbit is at its maximum value. Then, the required launch azimuth is 90° and the launcher's payload capability is maximum. This has occurred in March 1969, November 1987, and in June 2006; it will occur again in January 2025. In this Section, we will investigate the general characteristics of non-coplanar lunar trajectories and we will describe a method for selecting acceptable launch dates and injection conditions. In that analysis, we again consider the trajectory as a conic section about the Earth that intersects the lunar orbit. If we want to include the effect of the lunar gravity field on the trajectory, we have to extend the analysis by incorporating the methodology presented in Section 17.3. We first assume that the ascent trajectory lies in the plane of the lunar transfer trajectory, which means that no dogleg maneuvers have to be applied.

If there are no restrictions on the launch conditions or on the conditions at lunar approach, then there are no limitations on the time of the lunar month at which the spacecraft can approach the Moon. Practical considerations, such as launch site location, launch range safety, accuracy tolerances, and the limited range of attainable injection conditions, impose restrictions on the lunar intercept declination which can be accommodated. A typical design restriction for lunar missions is the specification of the lighting conditions on the surface of the Moon as determined by the phase of the Moon. For a particular year, the declination of the Moon at a given phase varies between maximum and minimum values, which correspond approximately to plus and minus the mean inclination of the Moon's orbit for that year. Another typical restriction concerns the permissible direction of launch from a particular site. In the following analysis, the launch site is assumed to be Cape Canaveral. The launch azimuth, ψ_l , must then lie between 37° and 112° , as specified by the Eastern Test Range safety requirements (Table 16.1); a launch azimuth of 90° will result in an orbital inclination of 28.5° .

We first analyze the case that the spacecraft intercepts the lunar orbit, without checking whether the Moon is at the interception point or not. An important parameter of the transfer orbit is the geocentric transfer angle, $\Delta\theta = \theta_2 - \theta_1$ from launch to lunar orbit intercept. The total transfer angle, $\Delta\theta_t$, consists of the free-flight transfer angle, $\Delta\theta_{ff}$, from injection to intercept, plus the transfer angle from launch to injection, $\Delta\theta_c$. Depending on the launch technique applied, $\Delta\theta_c$ may be simply the rocket engine burning arc for a direct-ascent launch or it may be the burning arc plus the angular distance traveled during a coasting period prior to injection. While the angle $\Delta\theta_c$ may be selected arbitrarily, the free-flight transfer angle, $\Delta\theta_{ff}$, is determined by the injection conditions r_1 , V_1 and γ_1 . The value of $\Delta\theta_{ff}$ at given injection conditions can be computed from the equations in Sections 17.2 and 17.3. In Figure 17.8, the free-flight transfer angle, $\Delta\theta_{ff}$, is plotted versus the initial velocity, V_1 , and for several values of γ_1 and a fixed injection altitude of 320 km. Lunar intercept is assumed to occur at a distance of 384,401 km.

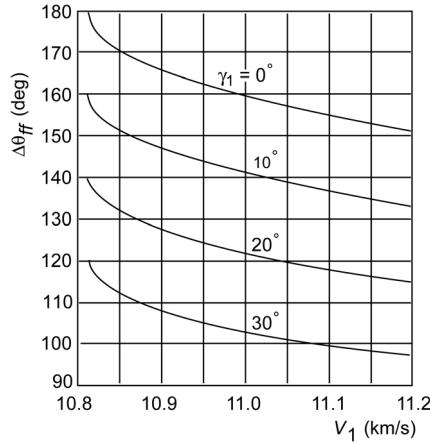


Figure 17.8: Free-flight transfer angle, $\Delta\theta_{ff}$, versus injection velocity, V_1 , and flight path angle, γ_1 , for an injection altitude of 320 km.

If we arbitrarily select $\Delta\theta_c$, we obtain the total transfer angle from

$$\Delta\theta_t = \Delta\theta_{ff} + \Delta\theta_c \quad (17.21)$$

Since the latitude of the launch site is known ($\phi_l = \delta_l = 28.5^\circ$), we may determine the declination of the spacecraft after it has traveled an arc $\Delta\theta_t$, δ_i , if we know the launch azimuth ψ_l , or we may determine ψ_l if δ_i is known. This is essentially a problem of spherical geometry and is illustrated in Figure 17.9. When we consider the spherical triangle containing the angles ψ_l and $\Delta\alpha$, and the

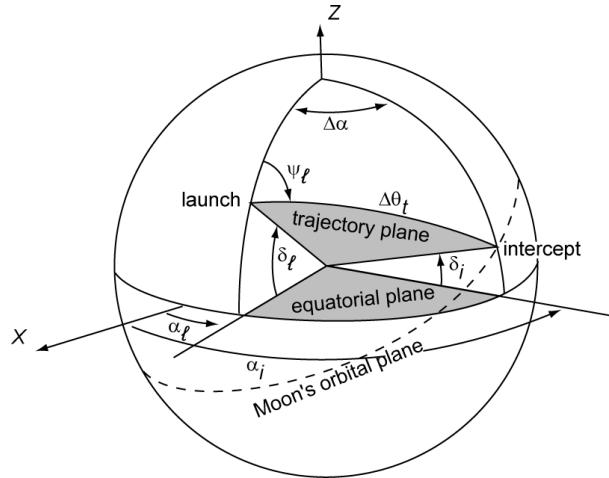


Figure 17.9: Lunar orbit interception angular relationships.

side $\Delta\theta_t$, we obtain from the first cosine rule for an oblique spherical triangle (Section 13.1)

$$\cos\left(\frac{1}{2}\pi - \delta_i\right) = \cos\left(\frac{1}{2}\pi - \delta_l\right) \cos\Delta\theta_t + \sin\left(\frac{1}{2}\pi - \delta_l\right) \sin\Delta\theta_t \cos\psi_l$$

or

$$\sin\delta_i = \sin\delta_l \cos\Delta\theta_t + \cos\delta_l \sin\Delta\theta_t \cos\psi_l \quad (17.22)$$

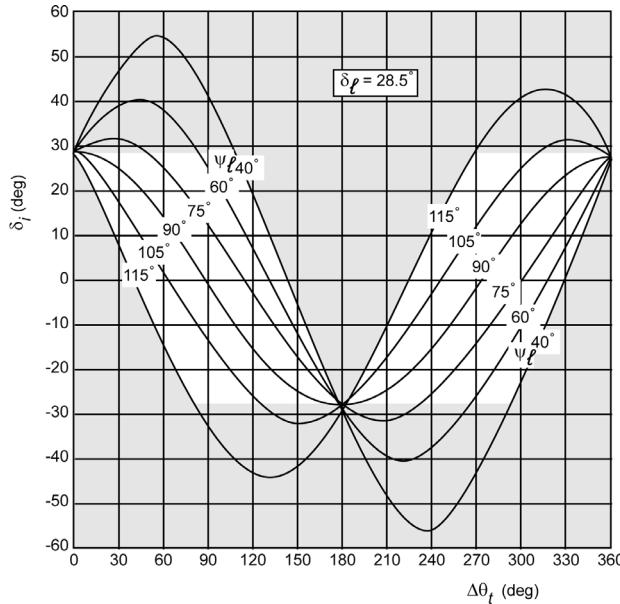


Figure 17.10: Interception declination, δ_i , versus the total transfer angle, $\Delta\theta_t$, for a launch from Cape Canaveral and a launch azimuth, ψ_l , between 40° and 115° . Note that these boundary values differ slightly from the launch azimuth boundary values of 37° and 112° listed in Table 16.1.

In Figure 17.10, δ_i , as computed from (17.22), is plotted versus total transfer angle, $\Delta\theta_t$, for launch azimuths between 40° and 115° , and for a launch from Cape Canaveral. The Figure shows white regions that hold for the early part of 1969, the last part of 1987, mid 2006 and the early part of 2025, when the Moon's orbital inclination relative to the Earth's equator is 28.6° and, consequently, the Moon's declination at intercept is limited between $+28.6^\circ$ and -28.6° . The shaded region of the graph indicates the impossible launch conditions for those years, because

of the lunar declination limits at intercept and the launch site safety restrictions. There are several interesting features of the Figure worth noting. Perhaps the most interesting is the fact that a transfer angle of 180° is possible only if we intercept the Moon at its maximum southern declination of -28.6° . For this condition, any launch azimuth (within the range of allowable azimuths) is feasible. Free-flight transfer angles of less than about 120° require high injection velocities, which is undesirable. As a result, if we are interested in minimizing injection velocity, we must intercept the Moon when it is near its maximum southern declination for direct ascent launches or launches where the coasting period is small enough to keep $\Delta\theta_t$ less than 180° . However, if we add a coasting arc, $\Delta\theta_c$, large enough to make $\Delta\theta_t$ greater than 180° , we can intercept the Moon at any point along its orbit. It is emphasized that these conclusions only hold for launches in the years indicated. For other years, the white regions in Figure 17.10 are different, which leads to other conclusions about the declination of feasible and optimum interception points.

If lunar declination at intercept is specified, then launch azimuth, ψ_l , can be computed from (17.22). The next step is to search through a lunar ephemeris to find a time when the Moon is at the correct declination. If lighting conditions are important, we must find a time when both declination and phase are simultaneously correct. In the past, the American Ephemeris and Nautical Almanac (Section 17.1) was frequently used for these purpose; nowadays we search through computer files. Now, suppose that we have selected a time t_i for lunar intercept that meets all constraints. The right ascension of the Moon at t_i is then extracted from the ephemeris computer file; this angle α_i is indicated in Figure 17.9. The difference in right ascension between launch and intercept, $\Delta\alpha = \alpha_i - \alpha_l$, is fixed by the geometry of Figure 17.9. Considering again the spherical triangle containing the angles ψ_l and $\Delta\alpha$, and the side $\Delta\theta_t$, we obtain from the first cosine rule for an oblique spherical triangle (Section 13.1)

$$\cos \Delta\theta_t = \cos(\frac{1}{2}\pi - \delta_l) \cos(\frac{1}{2}\pi - \delta_i) + \sin(\frac{1}{2}\pi - \delta_l) \sin(\frac{1}{2}\pi - \delta_i) \cos \Delta\alpha$$

or

$$\cos \Delta\alpha = \frac{\cos \Delta\theta_t - \sin \delta_l \sin \delta_i}{\cos \delta_l \cos \delta_i} \quad (17.23)$$

The next step is to establish the exact time of launch, t_l , and the right ascension of the launch site, α_l , at this launch time. To establish t_l , we need to compute the total flight time from launch to intercept. This consists of the free-flight time from injection to intercept, t_{ff} , which can be computed from the injection conditions, plus the time to traverse the engine-burning arc and coasting arc, t_c . Thus, the total flight time, $t_{f,tot}$, is

$$t_{f,tot} = t_{ff} + t_c \quad (17.24)$$

The launch time, t_l , may then be obtained from

$$t_l = t_i - t_{f,tot} \quad (17.25)$$

The right ascension of the launch site, α_l , may be obtained from (11.2)

$$\alpha_l = \Theta_{GM} + \Lambda_l \quad (17.26)$$

where Λ_l is the geographic longitude of the launch site and Θ_{GM} is the Greenwich mean sidereal time at launch (Section 11.4). It would be sheer coincidence if the difference $\alpha_i - \alpha_l$, where α_l is

now computed from (17.26), would have the same value as $\Delta\alpha$ computed from (17.23). Since the right ascension of the Moon changes slowly, whereas the right ascension of the launch site changes by nearly 360° in a day, it is possible to adjust t_i slightly and recompute t_i and α_i until $\alpha_i - \alpha_l$ and the required $\Delta\alpha$ from (17.23) agree. The lunar declination at intercept will change very slightly as t_i is adjusted and it may be necessary to go back to (17.22) and redetermine the launch azimuth ψ_l . We now know the injection conditions r_1 , V_1 , γ_1 , ψ_l , and the exact day and time of launch and of lunar intercept that will satisfy all of the constraints set forth earlier. These launch conditions should provide us with sufficiently accurate initial conditions to begin the computation of a precision lunar trajectory using extensive numerical methods.

If a specified lunar declination at intercept requires a launch azimuth that falls outside the launch site safety restrictions, then the ascent trajectory cannot be coplanar with the lunar transfer trajectory and dogleg maneuvers are required. However, as explained in Sections 13.2 and 13.3, large plane change maneuvers are very costly in terms of propellant consumption. Therefore, if the lunar flight occurs on a date when large plane changes are required, it is better to use another strategy. One example is to launch the spacecraft due east into an orbit with inclination equal to the latitude of the launch site and with an apogee distance of about 10^6 km. This adds only about 72 m/s to the required injection ΔV compared with a transfer orbit apogee distance equal to the Moon's distance. At apogee, the spacecraft's velocity is very small and a relatively small ΔV of about 0.3 km/s is enough to change the orbital plane such that the orbit's return leg meets the Moon at the specified lunar declination. This *bi-elliptical transfer* (Section 13.5) reduces the cost of the plane change considerably. However, the transfer flight time extends to 50 days. Arrival conditions at the Moon are comparable to those of a direct transfer, but the spacecraft now arrives around perigee of the transfer orbit, while in a 'classical' transfer it approaches the Moon around apogee of the lunar transfer orbit. Since its velocity is now higher as compared to the 'classical' transfer, the relative velocity with respect to the Moon is lower and therefore the ΔV needed for insertion into an orbit about the Moon is also lower. Additional burns may be required to position the perilune, for example, above the lunar south pole. These long transfers, though more complex, make lunar missions less dependent on appropriate launch windows.

The total ΔV required for a lunar orbiter or lander mission can be further reduced by injecting the spacecraft such that it flies a so-called non-Hohmann low-energy transfer trajectory. This type of flights will be discussed in Section 17.5.

Figure 17.11 shows an example of a 'classical' trajectory to the Moon and back to Earth. The plot on the left shows a projection of the trajectory, relative to a non-rotating geocentric reference frame, onto the Earth's equatorial plane, where the X -axis points towards the vernal equinox. The plot on the right shows a projection of the trajectory, relative to a rotating geocentric reference frame, onto the Moon's orbital plane, with the X -axis along the Earth-Moon line. Figures 17.12 and 17.13 show cartoons of the lunar and return flight in the manned Apollo program, and the lunar trajectory of the Chandrayaan-1 spacecraft, respectively, while Table 17.2 presents numerical details of the flight of Apollo spacecraft to the Moon.

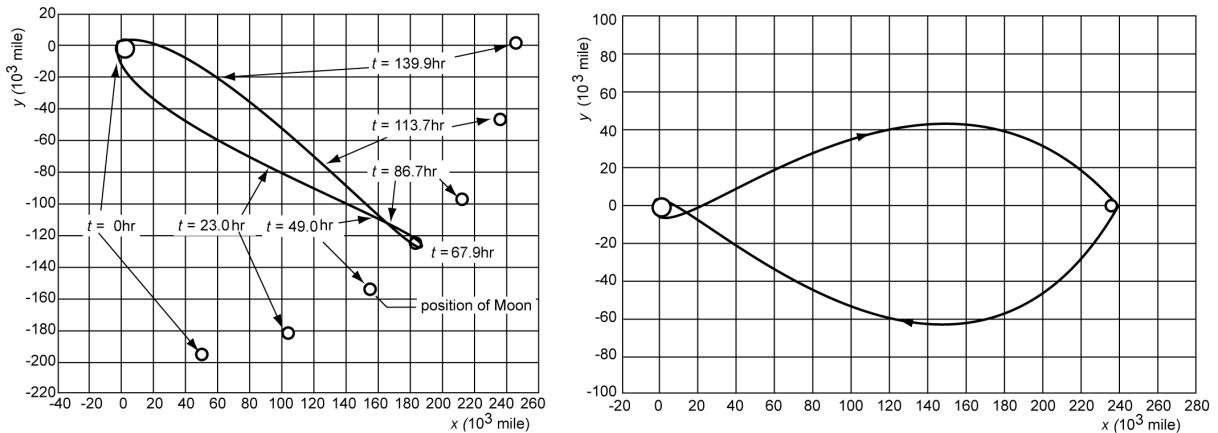


Figure 17.11: Example of a lunar trajectory. The plot on the left shows a projection of the trajectory relative to a non-rotating geocentric reference frame onto the Earth's equatorial plane; the plot on the right shows a projection of the trajectory relative to a rotating geocentric reference frame onto the Moon's orbital plane.

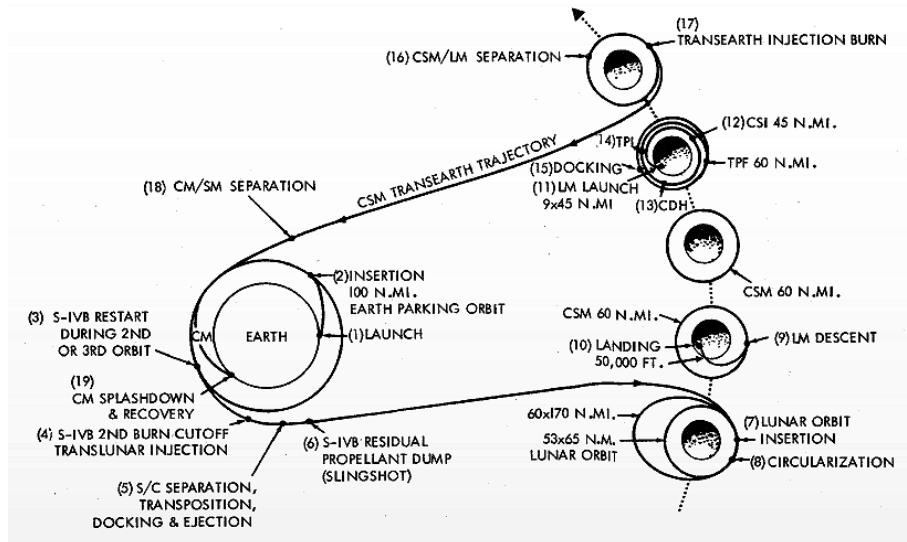


Figure 17.12: Cartoon of the lunar and return flight in the manned Apollo mission.

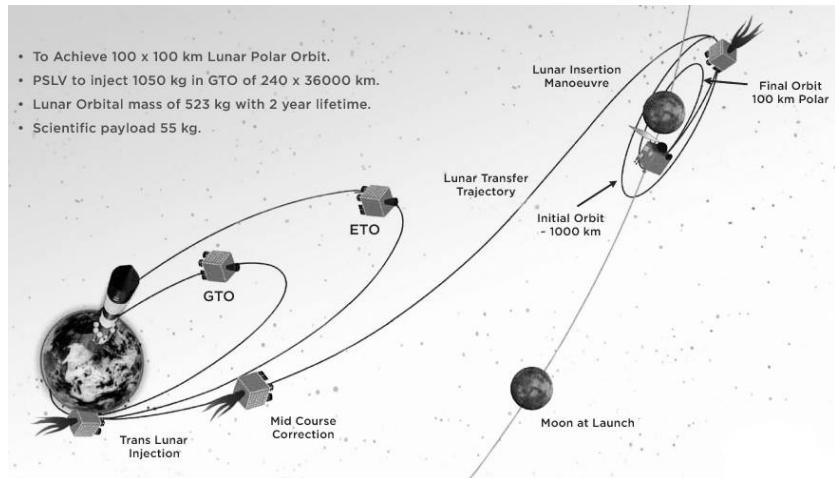


Figure 17.13: Cartoon of the lunar trajectory flown by the Chandrayaan-1 spacecraft.

Table 17.2: Apollo lunar trajectory data; compiled from NASA Apollo mission evaluation reports.

	Apollo 8	Apollo 10	Apollo 11	Apollo 12	Apollo 13	Apollo 14	Apollo 15	Apollo 16	Apollo 17
Launch									
Date (day/month/year)	21/12/1968	18/05/1969	16/07/1969	14/11/1969	11/04/1970	31/01/1971	26/07/1971	16/04/1972	07/12/1972
Time (hr:min:s UTC)	12:51:00	16:49:00	13:32:00	16:22:00	19:13:00	21:03:02	13:34:00	17:54:00	05:33:00
Earth orbit insertion									
Mission time (s)	695.0	713.8	709.3	703.9	759.8	710.6	704.7	716.2	712.6
Altitude (km)	191.36	191.37	191.08	190.91	191.63	190.92	172.63	172.93	170.49
Velocity (km/s)	7.7928	7.7931	7.7931	7.7925	7.7925	7.7925	7.8037	7.8044	7.8041
Latitude (°)	32.474	32.530	32.503	31.513	32.525	31.081	29.205	32.526	24.538
Longitude (°)	306.708	307.474	307.306	306.869	309.510	307.017	306.919	307.470	306.189
Flight path angle (°)	0.0006	-0.0049	0.012	-0.014	0.005	-0.003	0.015	0.001	0.003
Flight path azimuth (°)	88.532	89.933	88.848	88.580	90.148	91.656	95.531	88.932	105.021
Apogee (km)	185.28	185.89	186.04	185.48	185.86	185.48	169.55	169.18	167.33
Perigee (km)	184.50	184.76	183.26	181.22	184.00	183.26	166.03	166.77	166.77
Period (min)	88.19	88.20	88.18	88.16	88.19	88.18	87.84	87.85	87.83
Inclination (°)	32.509	32.546	32.521	32.540	32.547	31.120	29.679	32.542	28.526
R.A. descending node (°)	42.415	123.132	123.088	123.126	123.084	117.455	109.314	123.123	86.978
Translunar injection									
Mission time (hr:min:s)	2:50:37	2:39:21	2:44:16	2:53:14	2:41:47	2:34:33	2:56:04	2:33:37	3:12:38
Altitude (km)	346.92	333.39	334.62	368.79	338.07	332.69	321.83	317.31	313.90
Velocity (km/s)	10.8220	10.8396	10.8343	10.7868	10.8321	10.8239	10.8445	10.8406	10.8373
Latitude (°)	21.3460	-13.5435	9.9204	16.0791	-3.8635	-19.4388	24.8341	-11.9117	4.6824
Longitude (°)	216.0758	159.9201	195.1627	205.7202	167.2074	141.7312	217.8705	162.4820	306.8810
Flight path angle (°)	7.897	7.379	7.367	8.584	7.635	7.480	7.430	7.461	7.379
Flight path azimuth (°)	67.494	61.065	60.073	63.902	59.318	65.583	73.173	59.524	118.110
Eccentricity	0.97553	0.97834	0.97696	0.96966	0.9772	0.9722	0.9760	0.9741	0.9722
Inclination (°)	30.636	31.698	31.383	30.555	31.817	30.834	29.696	32.511	28.466
R.A. descending node (°)	38.983	123.515	121.847	120.388	122.997	117.394	108.439	122.463	86.042
Lunar orbit insertion									
Mission time (hr:min:s)	69:08:20	75:55:54	75:49:50	83:25:23	----	81:56:41	78:31:46	74:28:28	86:14:23

Mission time is zero at launch; R.A. = right ascension.

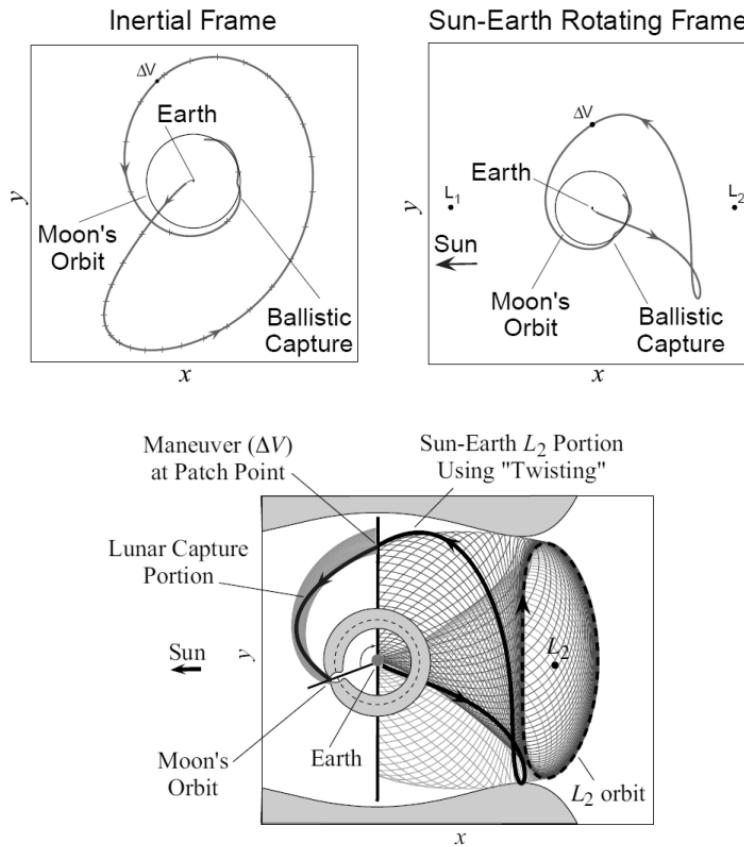
17.5. Non-Hohmann low-energy transfer trajectories

In Section 17.3 we have found that an optimum two-dimensional classical two-impulses transfer from a 200 km altitude circular parking orbit about the Earth to a 500 km altitude circular orbit about the Moon requires an injection impulsive shot of $\Delta V_0 = 3.130$ km/s and a braking impulsive shot of $\Delta V_p = 0.732$ km/s (Table 17.1). At entrance of the Moon's sphere of influence the velocity of the spacecraft relative to the Moon is $V_3 = 0.811$ km/s. The total travel time is 120.3 hr, or 5 days. Note that an escape flight would have required an injection impulsive shot of $\Delta V_0 = 3.224$ km/s; so only 0.1 km/s more than a flight to the Moon. These values may be considered as representative for the full three-dimensional case. Over the years, many studies have been performed to determine multi-impulses (three-dimensional) transfer trajectories which require less propellant. In these studies, the transfer was considered as a (circular) restricted three-body (Earth-Moon-spacecraft) or even four-body (Sun-Earth-Moon-spacecraft) problem. These studies have indicated that some gain in the required total ΔV can be realized. However, a real breakthrough came around 1987 when E.A. Belbruno (1951-) described a method to obtain transfer trajectories with only a very small hyperbolic excess velocity upon arrival at the Moon (Section 3.12). These trajectories require a slightly higher injection velocity to reach an apogee distance of about 1.5×10^6 km, but the total mission ΔV is about 350 m/s less than for an optimum classical two-impulses transfer, leading to a significantly larger spacecraft payload capability. The travel time is, however, more than ten times longer than for a Hohmann transfer. For the design of these special non-Hohmann low-energy lunar transfer trajectories extensive use is made of the concepts of *ballistic capture*, *weak stability boundary* and *invariant manifold*, which were introduced in Section 3.12.

For these low-energy transfers, the dynamics of the spacecraft are very complicated, because the combined simultaneous gravitational attractions by the Sun, the Earth, and the Moon have to be taken into account. This is in contrast to a relatively fast direct (Hohmann-type) transfer, for which the effect of the gravitational attraction by the Sun is rather small, and the attraction by the Moon becomes dominant only for the last part of the trajectory. Since the flight time of a non-Hohmann low-energy transfer is significantly longer, all these gravitational attractions are subtly acting for a long time, affecting the entire trajectory substantially. Therefore, for these low-energy trajectories conic sections are not good approximations anymore. Consequently, the elegant *patched-conic* method (Section 17.3) does not apply and a different approach must be developed for the design and optimization of these trajectories. *Dynamical systems theory* (Section 3.12) has provided many of the methodologies that allow mission designers to divide such low-energy missions into manageable pieces, to optimize those pieces, and then to combine each trajectory piece again. This theory may be applied in the Earth-Moon-spacecraft three-body system, where the Sun acts as a perturbing body. In more-advanced analysis methods, the Sun-Earth-Moon-spacecraft system is considered as a superposition of two three-body systems: Sun-Earth-spacecraft system and Earth-Moon-spacecraft system. The rotating reference frame applied for the analysis of the motion in the Earth-Moon-spacecraft system, of course, rotates relative to the rotating reference frame applied for the analysis of the motion in the Sun-Earth-spacecraft system (Figure 3.9), and this time-varying geometry has to be taken into account. Using our understanding of the invariant manifold tube structures (Section 3.12) associated with the Lagrange points L_1 and L_2 , the initial part of the transfer trajectory is designed such that the spacecraft crosses the region where the associated manifold structure of the Sun-Earth L points interacts with the manifold structure of the Earth-Moon L points. Then, the sensitivity of the ‘twisting’ of trajectories near the invariant manifold tubes in the Sun-Earth L points region is used to find a propellant-efficient transfer from the Sun-Earth system to the Earth-Moon system.

The manifold tubes of the Earth-Moon system provide the dynamical channels in phase space that enable ballistic capture of the spacecraft by the Moon.

Such a low-energy lunar mission may begin in any low-altitude Earth parking orbit. A specific transfer trajectory has a narrow launch window, but in the event of a delayed launch a nearby transfer may be easily found with similar properties to open a new launch window. Alternatively, a small maneuver ‘en route’ may be applied to correct the trajectory to reach the proper lunar transfer orbit. After the spacecraft leaves its low-altitude Earth parking orbit, it travels toward the region around e.g. the Earth-Sun L_2 point (Figure 17.14, top) at a distance of about 1.5×10^6 km from Earth. We will describe the spacecraft’s motion both from a two-body perspective and from a three-body perspective, since both perspectives help to paint the full picture. From a two-body perspective, the spacecraft begins by transferring from its low Earth orbit onto a highly eccentric orbit with apogee far beyond the Moon’s orbital radius. As the spacecraft approaches and traverses the apogee region of this orbit, it lingers long enough to give the Sun a large amount of time to perturb the orbit. For certain positions of the Sun relative to the transfer trajectory, the Sun’s gravitational attraction may change the spacecraft’s orbit such that, by the time the spacecraft has begun to return back to its perigee, the perigee distance has increased to about the radius of the Moon’s orbit about the Earth. During its flight from apogee



*Figure 17.14: (top) Sun-perturbed Earth-Moon low-energy transfer trajectory with ballistic capture involving weak stability boundaries. The plot on the left shows the trajectory relative to an inertial reference frame; the plot on the right the trajectory relative to a rotating reference frame with the X-axis coinciding with the Sun-Earth line. (bottom) The same trajectory relative to the rotating reference frame together with a Sun-Earth L_2 Lyapunov orbit, the associated invariant manifolds, the maneuver to transfer to an Earth-Moon manifold, and the capture by the Moon (schematic). [copied from W.S. Koon et al., *Dynamical Systems and Space Mission Design*, presentation at California Institute of Technology, Pasadena, June 2000.]*

to the Moon's orbit, the continuously acting solar gravitational attraction may also change the velocity of the spacecraft such that when the spacecraft approaches perigee its velocity is about equal to the velocity of the Moon about the Earth. From a three-body perspective, the spacecraft begins by transferring from its low-altitude Earth parking orbit onto a trajectory that shadows the stable manifold of the Sun-Earth L_2 point. The spacecraft approaches a periodic orbit, but does not enter it (Figure 17.14, bottom). The spacecraft then transfers to a trajectory that shadows the unstable manifold of the Sun-Earth L_2 point. This trajectory takes the spacecraft to the lunar encounter. As the spacecraft approaches the Moon, it arrives onto the stable manifold of the Earth-Moon L_1 or L_2 point. As it follows this stable manifold, it asymptotically approaches its final orbit about the Earth-Moon L_1 or L_2 points. The orbit is typically chosen to either meet certain mission requirements or to be used as a transitional orbit before transferring to a final lunar orbit or to the surface of the Moon itself.

As was mentioned above, the trajectory can be designed such that upon arrival at a point close to the Moon the spacecraft's velocity is about equal to the velocity of the Moon about the Earth. When the spacecraft crosses the weak stability boundary around the Moon for that velocity it can be ballistically captured (Section 3.12) by the Moon. This capture process is gradual and extends over a period of about two weeks prior to the spacecraft's closest approach to the Moon.

The first lunar spacecraft that has utilized the concept of weak stability boundaries transfer was the Japanese Hiten spacecraft (originally named Muses-A), which was designed to test and verify technologies for future lunar and planetary missions. Hiten was launched into a highly eccentric Earth orbit on a Mu-3SII-5 rocket from Kagoshima Space Center in Japan on January 24, 1990. The actual injection velocity was 50 m/s less than the nominal value, resulting in an apogee altitude of only 290,000 km compared to the planned 476,000 km. Using its onboard propulsion system, a number of trajectory correction maneuvers were performed and Hiten was put back in its nominal orbit. On March 18, 1990, as Hiten approached its first lunar flyby, the small Hagoromo satellite was released. This satellite was planned to go into an orbit about the Moon with the Hiten spacecraft remaining in its eccentric Earth orbit as a communications relay satellite. However, the S-band transmitter aboard Hagoromo had failed on February 21, 1990, and therefore its mission essentially was a failure. Hiten completed seven more lunar swingbys by March 4, 1991, and then started the aerobraking portion of its mission. On March 19, 1991, Hiten flew into the Earth's upper atmosphere at an altitude of 125.5 km over the Pacific at 11.0 km/s. Atmospheric drag reduced the velocity by 1.7 m/s and the apogee altitude by 8665 km. This was the first time aerobraking was used to modify a spacecraft orbit at close to escape velocity. Another aerobraking maneuver was performed on March 30, 1991, at 120 km altitude, reducing velocity by 2.8 m/s and apogee by 14,000 km. This concluded the primary mission and the extended mission phase started.

The goal of that follow-on phase was to bring Hiten in an orbit about the Moon. Because the spacecraft did not carry enough propellant to execute the required maneuver, a low-energy transfer using the weak stability boundary concept described in Section 3.12 was designed by Belbruno in 1990. Hiten left its eccentric orbit about the Earth on April 24, 1991, with a tiny maneuver. It went out to roughly 1.5×10^6 km from the Earth, taking about 45 days. A ninth lunar swingby was used to increase the apogee distance to 1.532×10^6 km, where it crossed the unstable region around the Sun-Earth L_2 point with a low apogee velocity. Then, Hiten executed a tiny maneuver to enter an orbit back to the Moon, and reached a weak stability boundary around the Moon, corresponding to its relatively high velocity at that time, in October 1991. It turned out that the sum of the maneuver to leave its eccentric orbit about the Earth and the second maneuver at its very high altitude was only 48 m/s. On October 2, a small maneuver was executed to

achieve temporarily lunar capture. Some time later, Hiten was put into a looping orbit that passed through the Earth-Moon L_4 and L_5 stable libration points to look for trapped dust particles. No increase of the particle density was found. On February 15, 1992, at a closest lunar approach of 422 km, most of Hiten's remaining fuel was used to put it into a highly eccentric orbit about the Moon with a period of about 3.5 days. After almost 13 months in this orbit, the very last fuel was used to have the spacecraft purposely crashed into the lunar surface on April 10, 1993, at 55.6° E, 34.3° S.

The ESA SMART-1 spacecraft also utilized a non-Hohmann low-energy trajectory to the Moon. However, the trajectory design was completely different and was based on interior low-energy transfer (Section 3.12). The purpose of this mission was the testing and proving of an ion engine and miniaturized instruments, along with investigations of lunar surface chemistry and a search for ice at the lunar south pole. The spacecraft was launched on September 27, 2003, from Kourou, French Guiana, as passenger on an Ariane 5G rocket that carried two large communications satellites as its primary payload. It was put into a geostationary transfer orbit with perigee and apogee altitudes of 655 km and 35,890 km, respectively, and inclined at 6.9° to the equator. The total mass of the spacecraft was 367 kg at launch, including 82.5 kg of xenon propellant for its primary Hall-effect ion propulsion system (Section 19.1). This system provided a thrust of about 70 mN at an exhaust velocity of about 16 km/s. The electrical power for the propulsion system was delivered by two solar arrays producing 1850 W at the start of the mission. Powered by its ion engine, SMART-1 gradually spiraled away from Earth. The particular trajectory initially required thrusting for about one-third to one-half of every orbital revolution, in the perigee part of the orbit. Up to mid September 2004 these thrusting arcs resulted in an increase of the apogee distance to about 300,000 km. Subsequently, thrusting arcs around the apogee region increased the perigee distance to about 170,000 km. These apogee thrusting arcs also increased the inclination of the orbit to 12° . From an apogee distance of 200,000 km onwards, the Moon's gravitational attraction started to significantly perturb the orbit and helped to widen the spiral orbit. The spacecraft utilized three lunar resonance phases on August 19, September 15 and October 12, while the spacecraft was at apogee. At these times the apogee distances were about 230,000 km, 290,000 km and 324,000 km, respectively, and the ratio of the spacecraft's orbital period and the Moon's orbital period was 5:1, 4:1, and 3:1, respectively. During such a resonance phase, the Moon's gravitational attraction produces a relatively large change of the spacecraft's orbit. Its final continuous thrusting maneuver took place over a period of 100 hr from October 10 to 14. Then, the ion engine was shut off and the spacecraft coasted to a point in space in the region of the Earth-Moon L_1 Lagrange point. Lunar orbit capture occurred on November 14, at a distance of about 60,000 km from the lunar surface. The capture itself costed no propellant! Up to that moment, the electric propulsion system had started 288 times, had accumulated 3652 hours of firing time, and had consumed about 59 kg of xenon propellant. The first orbit about the Moon had a perilune distance of 6704 km, apolune distance of 53,208 km, inclination of 81.1° , and period of 129.23 hr. On November 15, the electric propulsion system began firing again, now continuously, for a period of 4.5 days to enter a stable lunar orbit and to begin the lunar descent phase. When the propulsion system would not have been functioning at that time, the spacecraft would have escaped from the Moon around December 18. It took until February 28, 2005, using its ion engine, to reach the final operational orbit about the Moon. To achieve this, an additional 236 thrusting arcs were executed, adding another 953 hours to the cumulative thrusting time. The operational polar orbit had perilune (close to the south pole) and apolune altitudes of 465 km and 2865 km, respectively, and an inclination of about 90° ; the total flight time to reach this orbit was about 1.5 year.

From this operational orbit, SMART-1 has made many scientific observations. The orbital elements varied during the operational period under the effect of the Earth gravity field perturbations. The perilune altitude slowly decreased to about 300 km and the apolune altitude increased to about 3000 km. On August 2, 2005, the electric propulsion system was restarted to re-boost the spacecraft orbit. It was fired repeatedly until September 17, when virtually all xenon propellant was spent. The end of the re-boost phase therefore also marked the end of the electric propulsion system's active live. In total, the propulsion system had been fired for 4,958.3 hr. SMART-1 was left in a natural orbit determined by lunar gravity and perturbations caused by the gravity fields of the Earth and the Sun. That would have resulted in an impact of the spacecraft on the far side of the Moon on August 17, 2006. Using the hydrazine attitude control thrusters a series of orbit trim maneuvers was executed in the period June 19 to July 3, 2006, to increase the perilune altitude and by that to adjust the time and place of the impact on the Moon, such that the impact could be observed from Earth. The spacecraft impacted on the near side of the Moon at a speed of about 2 km/s on September 3, 2006, at 46.1° W, 34.2° S. The impact flash was captured by several telescopes on Earth.

On September 10, 2011, the two NASA GRAIL spacecraft were launched by a Delta II 7920H-10C two-stage rocket with nine strap-on solid rocket motors from Cape Canaveral on a mission to the Moon. The spacecraft were nearly identical; each with a total mass of 307 kg, including 105 kg of propellant. They were initially called GRAIL A and GRAIL B, but were renamed Ebb and Flow in March 2012. Each spacecraft carried two solar panels and a lithium ion battery; each solar panel was capable of producing 700 W at the end of the mission. Each battery had a capacity of 30 Ah, and was used to provide power when the spacecraft was in the Moon's shadow. Each spacecraft had a 22 N liquid hydrazine catalytic thruster for lunar orbit insertion and trajectory changes, and a hydrazine warm-gas system with eight 0.9 N thrusters for attitude control and other small maneuvers. Ebb was separated from the launcher about nine minutes after the upper stage engine cut off; Flow followed about eight minutes later. The two spacecraft carried a single science package, consisting of an ultra-stable oscillator, microwave assembly, time transfer assembly, and gravity recovery processor assembly. The ultra-stable oscillator provided a steady reference signal that was used by all the instrument subsystems. The microwave assembly converted the oscillator's reference signal to the Ka-band frequency, which was transmitted to the other orbiter. During the science phase of the mission, the radio-frequency system measured the rate of change of the distance between both spacecraft very precisely. The time transfer assembly provided a two-way time transfer link between the spacecraft to both synchronize and measure the offset between the clocks aboard the two spacecraft. The gravity recovery processor assembly combined all the inputs received to produce the radiometric data that was downlinked to the ground. These data were translated into a high-resolution map of the Moon's gravity field. In addition, each spacecraft carried a small camera, called MoonKAM (Moon Knowledge Acquired by Middle school students). Thousands of students have selected target areas on the lunar surface and have sent requests for study to the MoonKAM Mission Operations Center in San Diego.

The two GRAIL spacecraft were injected into nearly identical Earth-departure orbits from a low Earth parking orbit, much like conventional direct lunar transfers. However, they were injected at a somewhat higher velocity such that their initial orbits had apogees far beyond the Moon. Both spacecraft flew a low-energy trajectory, which allowed them to efficiently venture from Earth to the Moon via the unstable region around the Sun-Earth L_1 Lagrange point. As each spacecraft traversed its orbit's apogee, the Sun's gravity pulled on the spacecraft and raised its orbit's perigee altitude. By the time each spacecraft fell back towards the Earth, its perigee

altitude had risen high enough such that the spacecraft encountered the Moon and could immediately perform a lunar orbit insertion. On September 30, the rocket engine of Ebb (GRAIL A) burned for 127 s and imparted a velocity change of 14 m/s on the spacecraft, while expending 1.9 kg of propellant. On October 5, the rocket engine of Flow (GRAIL B) burned for 234 s and imparted a velocity change of 25 m/s, while expending 3.7 kg of propellant. These burns were designed to begin distancing Ebb's and Flow's arrival times at the Moon by approximately one day and to insert them into the desired lunar approach trajectories. Figure 17.15 shows an overview of the spacecraft trajectories relative to the Sun-Earth rotating reference frame. On December 28, Ebb was at a distance of 106,000 km from the Moon and closing at a speed of 333 m/s. Flow was then at a distance of 128,000 km from the Moon and closing at a speed of 341 m/s. During their final approaches to the Moon, both orbiters moved toward it from the south, flying nearly directly over the lunar south pole. The lunar insertion burn for Ebb occurred on December 31; it took approximately 40 min and changed the spacecraft's velocity by about 191 m/s; the spacecraft entered a near-polar orbit of 90 km by 8363 km altitude and with a period of about 11.5 hr. Flow's insertion burn occurred about 25 hr later, lasted about 39 min and changed the spacecraft's velocity by 192 m/s; it entered almost the same orbit. From there they performed a series of maneuvers during a period of five weeks to transition into their science orbit. The orbital period for Flow was targeted to be approximately three minutes longer than the period for Ebb. This difference in period meant that Ebb, flying below Flow, would lap Flow once every three days. A series of maneuvers was performed in this phase to achieve the desired initial separation distance and ensure that Flow was ahead of Ebb in the formation. At the start of the science phase on March 6, 2012, the two GRAIL spacecraft were in a near-polar, near-circular orbit at an average altitude of about 55 km. This orbit was selected such that no orbit maintenance maneuver was required during the science data collection phase. This phase, which ended on May 29, 2012, was divided into three 27.3-day mapping cycles; each mapping cycle corresponded to one rotation of the Moon underneath the GRAIL orbit. During the first mapping cycle the mean separation distance between the two spacecraft increased from approximately 100 km to 225 km. A small orbit trim maneuver executed near the end of this mapping cycle produced a change of the separation drift rate. After this maneuver, the mean separation distance decreased from 225 km to approximately 65 km at the end of the third mapping cycle. These changes in separation distance were required to meet the GRAIL science objectives. Because the spacecraft were still in an excellent condition, an extended mission phase started on August 30, 2012, while the two spacecraft were 30 km above the Moon. During the science phase of the extended mission, which

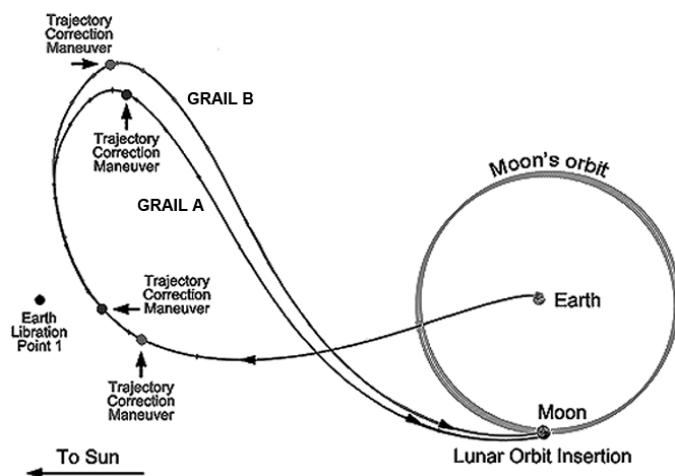


Figure 17.15: Trajectory of the GRAIL spacecraft relative to the rotating Sun-Earth reference frame.

ended on December 3, the average orbital altitude was decreased to 23 km, which placed the spacecraft within 8 km of some of the Moon's higher surface features. On December 14, 2012, the spacecraft each completed a thruster burn that brought them lower and lower over the Moon's surface and would result in an impact on December 17 on the southern face of an approximately 2.5 km tall mountain near the Goldschmidt crater, near the Moon's north pole. On December 17, the final maneuvers began when Ebb fired its main engines for 56 s, changing its orbital velocity by 4.6 m/s. Sixteen seconds later, Flow executing a 55 s burn with a resulting velocity change of 4.6 m/s. The satellites hit the lunar surface as planned at 22^h28^m UTC at a speed of 1.7 km/s. Fifty minutes prior to impact, the spacecraft fired their engines until the propellant was depleted; Ebb fired its engines for about 4 min and Flow fired its engine for about 5 min.

18. INTERPLANETARY FLIGHTS

For planetary missions a spacecraft is launched from the Earth and accelerated to a velocity higher than the local escape velocity. The spacecraft will then recede from Earth along a hyperbolic trajectory. While its distance from the Earth increases, the solar gravitational attraction will gradually become more important until, ultimately, it will become the dominating force and the spacecraft will enter a heliocentric orbit. The trajectory is chosen such that the spacecraft can perform a specified mission. When the spacecraft has to pass a target planet at a relatively small distance, we speak of a *flyby mission*; when the spacecraft has to be decelerated by a rocket engine and, if applicable, by atmospheric drag to enter an orbit about that planet, we call it an *orbiter mission*. For a *lander mission*, the final velocity of the spacecraft relative to the planet's surface has to be reduced to a very small value by means of atmospheric braking and/or rocket engine deceleration.

The first successful interplanetary mission was that of the US Mariner 2 spacecraft, which has flown by Venus at a minimum distance of 34,773 km on December 14, 1962. On July 14, 1965, the US Mariner 4 spacecraft became the first spacecraft to successfully fly by Mars at a minimum distance of 9,846 km. On October 18, 1967, the Venera 4 spacecraft, launched by the former Soviet Union (USSR), entered the Venusian atmosphere and transmitted back to Earth data about Venus' atmosphere. Contact was lost when the spacecraft reached an altitude of about 25 km. In December 1970, the Venera 7 spacecraft made a soft landing on Venus and transmitted signals for 23 min from Venus' surface. The first soft landing on Mars was performed by the Mars 3 spacecraft, launched by the USSR in December 1971. However, after 20 s, data return was terminated. In July and September 1976, the US Viking 1 and 2 spacecraft soft landed on Mars and performed extensive measurement programs. Since then, many interplanetary spacecraft have been launched by the USSR/Russia, USA, Europe (ESA) and Japan to perform flyby missions to all planets, flybys of comets, asteroids, and moons of planets, landings on Mars, Venus, Titan and asteroids, and an impact on a comet.

Until the early 1970s, the concept of interplanetary spaceflight was based on the application of *direct-transfer trajectories* achieved by a chemical rocket engine injection maneuver directly after launch of the spacecraft from Earth or after a period in an Earth parking orbit. In comparison with Earth satellite missions, these interplanetary trajectories require a much higher propulsive energy level and result in long flight times, which range from several months to some tens of years. For such types of missions, the trajectory that requires the least amount of propulsive energy is the *Hohmann transfer trajectory* (Section 18.4). But even for this minimum-energy trajectory, the relatively low energy density and exhaust velocity of chemical propellants only allow the launch of relatively small payloads, in particular to the more distant planets. Therefore, one had to accept that only a small fraction of the solar system could ever be reached and explored by instrumented spacecraft, or one had to develop more powerful propulsion systems, such as thermal-nuclear or nuclear-electric (Chapter 19) propulsion systems. Although the USA and the USSR have put a large effort over many years, beginning in the 1950s, in trying to solve the technical problems associated with the development of these advanced propulsion systems, the problems experienced were simply too difficult to surmount. When this conclusion became an undeniable fact in the mid-1960s, it appeared that much of the solar system would indeed remain unknown for a long period of time.

However, in the early 1960s a fundamental new concept of space travel was developed by M.A. Minovitch (1935-), then a graduate mathematics student, working at NASA's Jet Propulsion Laboratory (JPL): the *swingby* concept, also known as the *gravity-assist* or *slingshot* concept

(Section 18.11). The idea is to launch a spacecraft to an easy-to-reach nearby planet using a chemical rocket propulsion system, and to pass that planet on a precise trajectory such that the planet's gravity field would change the vehicle's orbital energy relative to the Sun and catapult it to another more distant planet, where its gravity field could then be used to catapult it to another planet, etc. In fact, this is an application of what was found in Section 2.1 for the n -body problem, namely that a close approach of two bodies leads to a high velocity of at least one of the bodies. Because the planet is much more massive than the spacecraft, it is, of course, the spacecraft that is accelerated to the high velocity. In this concept, the energy required for exploring the solar system is taken from the solar system itself! The 'propulsive force' automatically increases in direct proportion to the mass of the spacecraft. Hence, after the spacecraft is injected onto the first 'leg' of its interplanetary trajectory, the acceleration experienced by the spacecraft in a swingby maneuver is independent of its mass. No other propulsion system has this unique operating characteristic! By using these swingby maneuvers, existing launch vehicles can send relatively large spacecraft to distant planets.

The first spacecraft that has experienced a substantial gravity assist was NASA's Pioneer 10, which was launched on March 3, 1972, by an Atlas/Centaur/TE364-4 rocket. In December 1973, it approached Jupiter, traveling at 9.8 km/s. The spacecraft passed Jupiter at a minimum altitude of 130,350 km on December 3, 1973, and, as a result of this passage, it sped off into deep space at 22.4 km/s. On January 23, 2003, the last, very weak, signal from Pioneer 10 was received. On December 30, 2005, the spacecraft was 89.7 AU¹ away from the Sun, heading in the direction of the star Aldebaran in the constellation Taurus at roughly 2.6 AU per year. If Aldebaran had zero relative velocity, it would take Pioneer about 2 million years to reach it. The US Mariner 10 spacecraft, which was launched on November 3, 1973, towards Mercury and performed a flyby of this planet at a minimum altitude of 704 km on March 29, 1974, was the first spacecraft to use the swingby effect to reach another planet; it passed Venus on February 5, 1974, at a closest distance of 5768 km. In its orbit about the Sun the spacecraft made two additional encounters with Mercury: on September 21, 1974, it passed Mercury at an altitude of 48,069 km and on March 16, 1975, at an altitude of 327 km. Since then, this swingby technique has been applied for nearly all interplanetary missions and this was the reason that most of the solar system could already be explored, including regions close to the Sun, regions far above and below the ecliptic (Section 11.2) and deep space far beyond the orbit of Pluto².

The main practical limit to the use of the swingby concept is that planets are not always in the right places to help a voyage to a particular destination. For example, the Voyager missions, which started in the late 1970s, were made possible by the 'Grand Tour' alignment of Jupiter, Saturn, Uranus and Neptune. A similar alignment will not occur again until the middle of the twenty-second century. That is an extreme case, but even for less ambitious missions there are years when the planets are not in places that make swingbys useful. Another limitation is caused by the atmosphere of the available planet. The closer the spacecraft can get, the more boost it gets from the gravitational field. However, if a spacecraft gets too far into the atmosphere, the energy lost to friction can exceed that gained from the planet or the spacecraft can even burn up in the atmosphere.

Although the introduction of the swingby concept in interplanetary spaceflight was an enormous step forward, we should realize that it is not the ultimate solution to the realization of fast interplanetary space travel. The swingby technique requires that the spacecraft passes one or more other planet(s) on its travel to the target planet. So, in general, its overall trajectory is much longer than that of a direct-transfer mission. This means that, although the passing of the

¹ AU = Astronomical Unit. This is the mean distance between the Sun and the Earth; 1 AU \approx 149.6*10⁶ km.

² As already stated in Section 4.2, in this book Pluto is considered as a planet.

intermediate planet(s) will, in general, increase the spacecraft's velocity relative to the Sun considerably, in most cases the application of the swingby technique will lead to long total flight times. To execute real fast direct interplanetary transfer missions, high-energy thermal-nuclear or nuclear-electric propulsion systems have to be developed. Without succeeding in the development of such advanced propulsion systems, the possibilities for fast interplanetary travel, which is essential in particular for manned missions, will remain rather limited.

Since the 1980s a completely new type of non-Hohmann low-energy trajectories has been studied and applied for interplanetary missions. The design of these trajectories is based on the characteristics of the invariant manifolds of the Sun-planet systems (Section 3.12). Examples of missions based on this kind of low-energy trajectories are given in Section 18.12; the main part of this Chapter, however, is devoted to the analysis of 'classical' interplanetary trajectories.

18.1. General aspects of an interplanetary trajectory

It will be clear that for the precise computation of an interplanetary trajectory, the motion of the spacecraft must be considered as a many-body problem. Apart from the gravitational attractions by the Earth, the Sun and the target planet, the attractions by the Moon and a number of planets also have to be taken into account. In addition, solar radiation pressure may substantially affect the trajectory and should be accounted for. Because the spacecraft's mass is completely negligible relative to that of the Earth, Moon, Sun and planets, we may assume that the orbits of the Moon and the planets will not be affected by the gravitational forces between each of these bodies and the spacecraft. Consequently, for the orbits of the Moon and the planets we may use the ephemerides provided by e.g. NASA JPL or by the digital almanacs released by the US and UK Almanac Office (Section 17.1). These files contain precise ephemerides of the Sun, Moon, planets and their satellites, and sometimes minor planets, asteroids and comets, together with data relating to Earth orientation, coordinate systems and time scales.

For high-precision computations generally one of the planetary and lunar ephemerides released by NASA JPL under the name *Development Ephemeris* (DE) are used. Such an ephemeris contains Chebyshev polynomial coefficients from which the three-dimensional positions and velocities of the Sun, Moon and nine planets can be recovered as a function of Barycentric Dynamical Time (TDB), which is a relativistic coordinate time scale (Section 11.4). The ephemeris holds for a range of dates, usually both in the past and the future, and the recovered positions and velocities are relative to a rectangular barycentric non-rotating reference frame. There is a set of coefficients for each coordinate component of position and velocity for each object; each set covers a time span of several days. An ephemeris is the result of a numerical integration of an n -body system. The equations of motion used for generating the recent ephemerides include contributions from Newtonian gravitational point mass interactions among the Moon, planets and Sun; general relativity (isotropic, parameterized post-Newtonian metric); Newtonian gravitational perturbations from more than 300 selected asteroids; action upon the figure of the Earth by the Moon and Sun; action upon the figure of the Moon by the Earth and Sun; physical libration of the Moon; effect upon the Moon's motion caused by tides raised upon the Earth by the Moon and Sun. The orbit of the Sun is not integrated in the same way as the orbits of the planets. Instead, the position and velocity of the Sun are derived at each integration time step to keep the solar system barycenter at the origin of the coordinate system. The most advanced physical models for the various bodies are used in the numerical integration process. The initial conditions are adjusted in an iterative process such that the resulting orbits fit all available observations as good as possible. These observations include classical optical ground-based measurements of the Sun, Moon, planets and asteroids; ground-based radar ranging

measurements of the planets; ground-based lunar laser ranging measurements; spacecraft range, range-rate and VLBI measurements; among others.

Over the years, the reference frame to which the ephemeris refers has been tied to very distant celestial objects with ever increasing precision. The first widely-used ephemeris is DE118 that was released in 1981 and that refers to a non-rotating equatorial reference frame defined by the mean orientation of the Earth's equator and ecliptic at the beginning of the Besselian year 1950 (Section 11.4). Another widely used ephemeris is DE200, released in 1982, which basically is DE118 rotated into the mean equinox of 2000 reference frame (Section 11.2). DE200 has been the basis for the calculation of Astronomical Almanac planetary tables since 1984. Since then, a number of improved ephemerides have been developed, in particular for missions to the outer planets, based on more-accurate observations and improved dynamical models. From 2003 onward the Astronomical Almanac has been based on JPL ephemeris DE405 that was released in 1998. In this ephemeris the positions of the celestial bodies are relative to the International Celestial Reference Frame (ICRF, Section 11.2). One of the more recent ephemerides is DE421 that was released in 2009. With this ephemeris, where the solar system body positions also refer to the ICRF, the orbits of Venus, Earth and Mars can be recovered with an accuracy of better than a few hundreds of meters; the orbit of Mercury with an accuracy of a few kilometers; the orbits of Jupiter and Saturn with an accuracy of some tens of kilometers; the orbits of Uranus, Neptune and Pluto with an accuracy of some hundreds to thousands of kilometers.

For less accurate first-order analyses, usually the orbital elements of the planets (or other solar system bodies) are used in combination with the Keplerian formulas given in Section 11.8 to compute from the orbital elements the position and velocity of the planets relative to a non-rotating heliocentric reference frame. It must be noted that the orbital elements refer to an elliptical orbit that best fits the true orbit of the planet. Since the true orbit is perturbed by the gravitational attraction of the other planets and by other forces, any set of orbital elements is only valid for a given time or time span. In Table B5 of Appendix B the orbital elements of the planets, describing an elliptical orbit with the center of the Sun at a focal point (heliocentric reference frame), are listed for the epoch noon January 1, 2000; they hold with respect to the mean ecliptic and equinox of that epoch.

Even when the positions and velocities of the planets are known, the analyses of interplanetary trajectories is quite complex. For, one has to determine the spacecraft's injection conditions such that the spacecraft intercepts a target planet in its orbit about the Sun at the right time and with the right velocity relative to that planet. The search for the required interplanetary trajectory can only be made in an iterative way. The first step in that process is to assume that the interplanetary trajectory is part of a heliocentric conic section between the centers of the Earth and of the target planet. After a range of feasible trajectories has been determined, the computations are extended assuming that the interplanetary trajectory can be divided into three Keplerian trajectories. Within the *sphere of influence* (Section 4.3) of the Earth, the trajectory is considered as a geocentric hyperbola that is solely determined by the gravity field of the Earth; within the sphere of influence of the target planet the trajectory is considered as a planetocentric hyperbola that is solely determined by the gravity field of that planet; the interplanetary part of the trajectory is considered to be a heliocentric conic section that is solely determined by the Sun's gravity field. This method is known as the *patched-conic method*, which was introduced in Section 17.3 for the computation of trajectories towards the Moon. The main difference is that for lunar trajectories the sphere of influence of the Moon lies fully within the sphere of influence of the Earth, while for interplanetary trajectories the spheres of influence of the Earth and of the other planets do not overlap. Consequently, while for lunar trajectories the patched-conic approximation involves two Keplerian segments, in interplanetary missions at least three Keplerian

segments are involved. After a series of feasible trajectories has been obtained, the patched-conic approximation is replaced by a numerical integration analysis, which continuously accounts for the attraction by the Sun, the Moon and the planets on the spacecraft.

In the patched-conic approach, one implicitly uses a property of hyperbolic orbits that was discussed in Section 8.2, namely that the geocentric velocity of a spacecraft in a hyperbolic orbit about the Earth is at a distance of a few hundreds of thousands of kilometers about equal to the velocity that the spacecraft would have at an infinitely large distance from the Earth, $V_{\infty e}$. The velocity $V_{\infty e}$ is often called the *hyperbolic excess velocity*. This means that we may assume that the spacecraft has a velocity $V_{\infty e}$ with respect to the Earth when it crosses the Earth's sphere of influence. In Section 4.3, the radius of the Earth's sphere of influence relative to the Sun was found to be about 10^6 km. A similar assumption can be made for the velocity of a spacecraft in a hyperbolic orbit about another planet when it crosses that planet's sphere of influence. Although the radius of a planetary sphere of influence is quite large, both in an absolute sense and in relation to the radius of that planet, the radius of the sphere of influence is always small when measured on the scale of the solar system. Table 4.3 shows that for all planets this radius is less than 6.2% of the value of the semi-major axis of that planet's orbit about the Sun; for most planets it is less than 2% and for the Earth even only about 0.6%. This means that when measured on the scale of the solar system, the spacecraft reaches the velocity $V_{\infty e}$ very close to the Earth. So, at the moment that the interplanetary spacecraft leaves the sphere of influence of the Earth, its position is about that of the Earth when measured on the scale of the solar system, while it has a velocity $V_{\infty e}$ relative to the Earth.

The direction of the velocity vector $\bar{V}_{\infty e}$ with respect to the velocity vector of the Earth relative to the Sun, \bar{V}_e , depends on the time of launch from the Earth's surface, or, in case the spacecraft is injected into its interplanetary trajectory from a parking orbit about the Earth, on the position in the parking orbit where injection takes place (Figure 18.1). The velocity of the spacecraft relative to the Sun at the moment that it leaves the Earth's sphere of influence, \bar{V}_1 , is found by adding \bar{V}_e and $\bar{V}_{\infty e}$ vectorially. This velocity \bar{V}_1 and the position of the Earth relative to the Sun at that time determine the heliocentric trajectory of the spacecraft. When the spacecraft enters the sphere of influence of the target planet, the same computation scheme can be applied: the velocity of the spacecraft relative to the target planet, $\bar{V}_{\infty t}$, is found by subtracting the velocity of the target planet relative to the Sun, \bar{V}_t , vectorially from the velocity of the spacecraft relative to the Sun, \bar{V}_2 . The hyperbolic trajectory of the spacecraft within the sphere of influence of the target planet is completely determined by the point where the sphere of influence is entered and

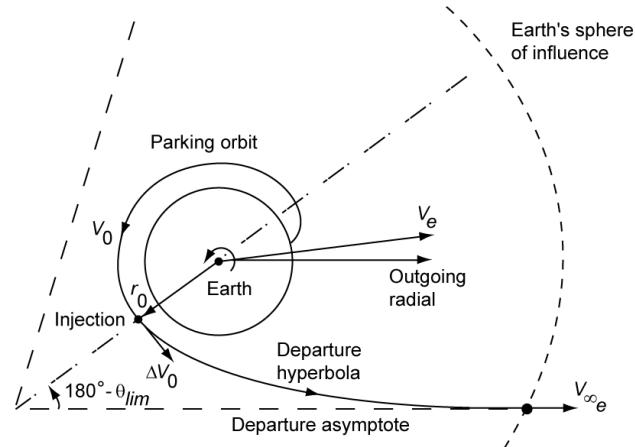


Figure 18.1: Matching of ascent trajectory, coasting phase and escape hyperbola for an interplanetary flight.

by the magnitude and direction of \bar{V}_{∞_t} .

18.2. Launching of interplanetary spacecraft

The launch date, time of the day, and selected interplanetary trajectory fully determine the required hyperbolic excess velocity, \bar{V}_{∞_e} , relative to a heliocentric reference frame. By applying the usual transformation relations, we can find \bar{V}_{∞_e} with respect to a non-rotating geocentric equatorial reference frame. Implicit in the method outlined in Section 18.1 is that this \bar{V}_{∞_e} may be considered as a vector passing through the center of the Earth. This vector determines the direction of the *outgoing radial* (Figure 18.1). Because the radius of the Earth's sphere of influence is very large as compared to the radius of the Earth, we may assume that the *outgoing asymptote* of an escape hyperbola is parallel to the outgoing radial and that both the outgoing radial and outgoing asymptote 'pierce' the sphere of influence at a point with the geocentric coordinates: declination, δ_{as} , and right ascension, α_{as} . The values of δ_{as} and α_{as} required for a particular mission strongly influence the selection of the ascent trajectory to be flown.

In order to prevent excessive dogleg maneuvers (Section 16.2) prior to injection, the launch vehicle ascent trajectory should lie preferably in the plane of the escape hyperbola. Then, the plane in which the motion within the sphere of influence of the Earth takes place is uniquely determined by the center of the Earth, the launch site and the outgoing asymptote. All American interplanetary missions start from Cape Canaveral (Eastern Test Range, ETR). For this launch site, a launch azimuth range from 90° to 115° ³ is most desirable for interplanetary missions to take full advantage of the Earth's eastward rotational velocity and of existing tracking facilities. According to (11.57) and (11.58), we may write for a coplanar launch

$$\begin{aligned} \cos i &= \cos \delta_l \sin \psi_l = \cos \delta_{as} \sin \psi_{as} \\ \cos(\alpha_l - \Omega) &= \frac{\cos \psi_l}{\sin i} ; \quad \sin(\alpha_l - \Omega) = \tan \delta_l \cot i \\ \cos(\alpha_{as} - \Omega) &= \frac{\cos \psi_{as}}{\sin i} ; \quad \sin(\alpha_{as} - \Omega) = \tan \delta_{as} \cot i \end{aligned} \quad (18.1)$$

where ψ denotes the flight path azimuth and the index l refers to the launch conditions. Because by definition $0^\circ \leq i < 180^\circ$, i can be solved unambiguously if $\cos i$ is known. From (18.1-1) follows

$$\sin^2 \psi_l \leq \frac{\cos^2 \delta_{as}}{\cos^2 \delta_l}$$

From Figure 18.2, where this inequality is plotted for a launch from Cape Canaveral ($\delta_l = 28.5^\circ$), we note that, for a launch from Cape Canaveral without applying a dogleg maneuver, values of δ_{as} within the range -35° to 35° can be reached. For a particular launch site, δ_l is known. When values of ψ_l and δ_{as} are selected, (18.1-1) yields the values of i and ψ_{as} , (18.1-2) the value of $\alpha_l - \Omega$ and (18.1-3) the value of $\alpha_{as} - \Omega$. So, for a specified launch site, we may compute from (18.1) the dependence of ψ_l and δ_{as} on the difference in right ascension $\alpha_{as} - \alpha_l$. These results are depicted in Figure 18.3 for a launch from Cape Canaveral. The shaded areas represent the range of

³ Just as in Section 17.4, here a boundary value of 115° is adopted that differs slightly from the boundary value of 112° listed in Table 16.1.

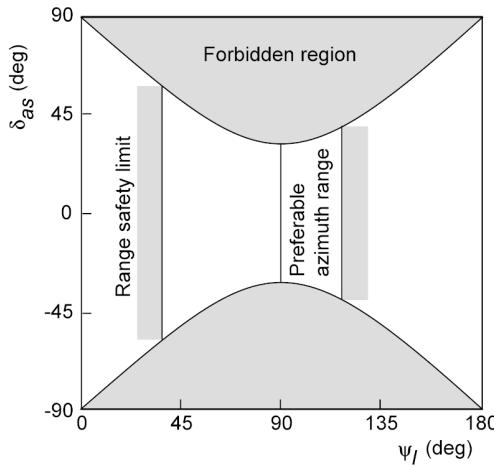


Figure 18.2: Permissible region of the declination of the outgoing asymptote for the launch of an interplanetary mission from Cape Canaveral ($\delta_i = 28.5^\circ$).

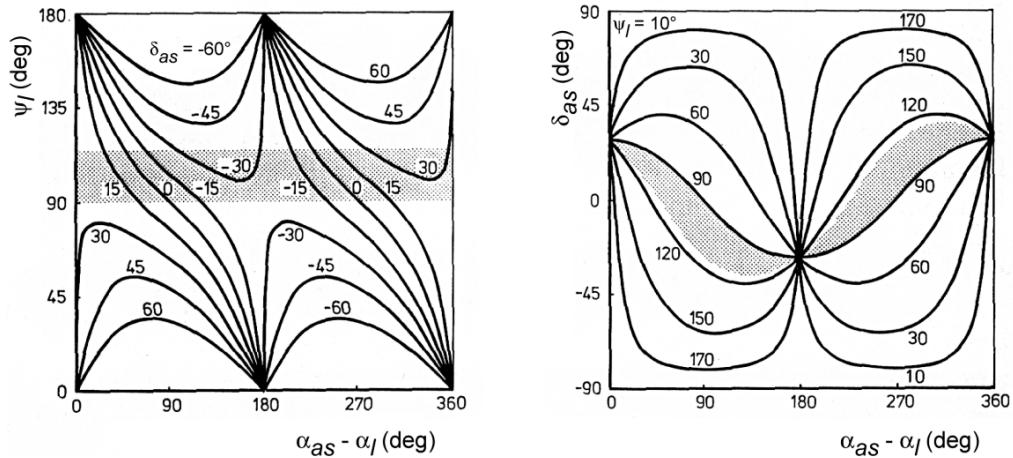


Figure 18.3: Launch azimuth and right ascension of the launch site versus declination and right ascension of the outgoing asymptote for a launch from Cape Canaveral.

preferable launch azimuths. During a given launch day and for a specified interplanetary mission, $V_{\infty e}$ is almost independent of the location of the launch site and of the time of launch. Owing to the rotation of the Earth, however, α_l varies linearly with launch time, and we may write with (11.2) and (11.3-2):

$$\alpha_l = \Theta_{GM0} + \Lambda_l + \dot{\Theta} t_l \quad (18.2)$$

where Θ_{GM0} is the Greenwich mean sidereal time (Section 11.4) at 0 hr UT1 of the launch day, Λ_l is the geographic longitude of the launch site, $\dot{\Theta}$ is the Earth's mean rotation rate, and t_l is the time of launch from midnight on the launch day (UT1). Therefore, the time of launch is determined solely by ψ_l , δ_{as} and α_{as} . Note from Figure 18.3 that, generally, for specified values of δ_{as} and α_{as} ($V_{\infty e}$), and of ψ_l , two launch times are possible each day.

Usually, a parking orbit *coasting period* is required to match the launch vehicle ascent trajectory and the required hyperbolic escape trajectory. The launch vehicle ascent trajectory is chosen such that, within the constraints of that vehicle, a maximum payload can be placed in the parking orbit. This ascent trajectory determines the altitude of the parking orbit and the position where the

spacecraft enters that parking orbit. Final injection takes place near perigee of the escape hyperbola. For a given $V_{\infty e}$, the location of this injection point in the parking orbit is fully determined. Thus, the spacecraft moves in its parking orbit until it reaches the point where the final burn has to be executed (Figure 18.1). We found that, generally, two launch times are possible each day. The required parking orbit coasting period and the location of the final injection point relative to the Earth's surface (tracking network) differ for these two cases. The launch time yielding the more favorable launch conditions will be selected.

From the viewpoint of launch vehicle reliability, tracking and telemetry coverage, a *Direct Ascent* (Section 16.1), without the application of a parking orbit, is very attractive. However, such an ascent flight can only be used if at the end of the powered phase the launch vehicle can inject the spacecraft near perigee of a hyperbola satisfying the prescribed direction of the outgoing asymptote. Generally, it is found that for Cape Canaveral a combination of a northeast launch and an outward radial declination of less than -15° result in feasible direct ascent trajectories for planetary missions.

For a specified interplanetary trajectory, there will be one instant of time for which the launch of a spacecraft requires a minimum amount of propellant. But because of the complexity of both launch vehicle and spacecraft, a launch exactly at that optimum moment is most unlikely, and an adequate firing window must be available around the optimum instant of time: the *daily launch window* (Section 16.2). Generally, a launch window of about one hour is available. As launch time varies within this launch window, both launch azimuth and parking orbit coasting time must be adapted continuously.

18.3. Assumptions and approximations

For the analyses presented in this Chapter, it will be assumed that the interplanetary flight starts from a circular parking orbit about the Earth and that a flight to an outer planet is considered. The same methods, however, can be applied for a direct ascent flight and for a flight to an inner planet. For the numerical examples presented in this Chapter, the physical and orbital parameters of the planets listed in Appendix B are used. To simplify the analysis, four approximations are introduced.

In the first part of this Chapter, it is assumed that the planetary orbits are circular and coincide with the plane in which the Earth moves about the Sun (ecliptic). In that case, the interplanetary trajectory and the planes of the circular parking orbit about the Earth and, if applicable, of the final orbit about the target planet should preferably also coincide with the ecliptic; this leads

Table 18.1: Basic data on the orbits of the planets.

Planet	a (AU)	e	i (deg)	r_a (AU)	r_p (AU)	V_a (km/s)	V_p (km/s)	T (sid.yr)
Mercury	0.38710	0.20564	7.005	0.4667	0.3075	38.858	58.977	0.241
Venus	0.72334	0.00678	3.395	0.7282	0.7184	34.784	35.259	0.615
Earth	1.00000	0.01671	0.000	1.0167	0.9833	29.291	30.287	1.000
Mars	1.52371	0.09339	1.850	1.6660	1.3814	21.972	26.498	1.881
Jupiter	5.20289	0.04839	1.304	5.4546	4.9511	12.441	13.706	11.87
Saturn	9.53668	0.05386	2.486	10.050	9.0230	9.1386	10.179	29.45
Uranus	19.1892	0.04726	0.773	20.096	18.282	6.4852	7.1286	84.06
Neptune	30.0699	0.00859	1.770	30.328	29.812	5.3851	5.4784	164.9
Pluto	39.4821	0.24883	17.14	49.306	29.658	3.6763	6.1119	248.1

to so-called *two-dimensional transfer trajectories*. Table 18.1 summarizes some characteristics of the orbits of the planets of our solar system; it is emphasized that all planets move in the same direction about the Sun. The values of the semi-major axis, eccentricity and inclination are copied from the data given in Appendix B. The other values are computed, using the standard relations for elliptical trajectories. Note that for all planets, except for Mercury and Pluto, the orbital eccentricity and inclination (relative to the ecliptic) are small, which justifies the approximations mentioned above.

As a second approximation, it is assumed that the thrusting time of a rocket engine is negligibly short compared to the time that the spacecraft moves within the sphere of influence of a planet. Then, the concept of *impulsive shot* (Section 1.7) can be applied. We then assume that a thrusting period results in a discontinuous velocity change ΔV of the spacecraft, while its position remains unchanged. The objective of the escape flight from the Earth is to increase the orbital energy of the spacecraft using a minimum amount of propellant. For, the rocket stage that has to produce the ΔV is launched together with the spacecraft in the parking orbit. If less propellant is required for the acceleration of the spacecraft, then the mass of the spacecraft can be larger or the applied launcher can be smaller and cheaper. If the spacecraft has to be decelerated at the target planet, also this maneuver has to be accomplished for the smallest amount of propellant. For an impulsive shot the minimum propellant requirement simplifies to the question: “How should the impulsive shot be applied such that the change of the spacecraft’s kinetic energy (per unit of mass) is maximum for a given magnitude of the total ΔV ?”

As a third approximation, it is assumed that the spacecraft is accelerated from an initial parking orbit about the Earth to the required velocity. For this maneuver, the question formulated above has been answered in Section 1.7. There, it was found that for a given magnitude of ΔV the increase of the spacecraft’s energy (per unit of mass) is maximum if the maneuver is executed at perigee of the parking orbit and if ΔV is directed tangentially to the orbit and in the direction of the orbital velocity at perigee. To inject with a given launcher the largest possible spacecraft mass into a parking orbit, one will always select a circular parking orbit at the lowest possible altitude. Then, there is no preference for the point of injection and ΔV should be directed perpendicular to the local radius vector. For the deceleration maneuver at the target planet, it was found in Section 1.7 that ΔV should be applied at the point in the flyby trajectory where the velocity is maximum, and in a direction opposite to the spacecraft’s velocity vector at that point; i.e. at pericenter of the flyby trajectory and perpendicular to the local radius vector.

As a fourth approximation, a hybrid model will be used to compute the interplanetary trajectory. The trajectory geometry is computed as if the trajectory is a heliocentric conic section between the centers of the Earth and of the target planet, while the ΔV ’s required to leave the parking orbit about the Earth and to enter an orbit about the target planet are computed by using the patched-conic model. The justification for using this hybrid model is the fact that the radius of a planetary sphere of influence is small when measured on the scale of the solar system.

A comparison with the results of detailed numerical computations shows that the results of this approximative method are reasonably accurate and that this method is well suited to first-order mission analyses.

Because $\Delta \bar{V}$ is oriented tangentially to the circular parking orbit about the Earth, the point where ΔV is applied (injection point) becomes the perigee of the hyperbolic escape trajectory. If the index 0 refers to the conditions at this injection point, the velocity of the spacecraft just after the impulsive shot, V_0 , is given by

$$V_0 = V_{c_0} + \Delta V_0 \quad (18.3)$$

where V_{c_0} is the circular velocity at the injection point and ΔV_0 is the magnitude of the impulsive shot applied at the injection point. For given values of r_0 and ΔV_0 , the shape of the hyperbolic escape trajectory is completely determined and so the angle between $\bar{V}_{\infty e}$ and the radius vector to the pericenter of the hyperbola (Figure 18.1) is fully determined. For this angle we find from (8.3) and (8.14)

$$\cos \theta_{lim} = -\frac{1}{e} = \frac{1}{1 - (V_0/V_{c_0})^2} \quad (18.4)$$

Consequently, the angle between $\bar{V}_{\infty e}$ and \bar{V}_e is determined by given values of r_0 and ΔV_0 , determined by the point in the (circular) parking orbit where ΔV_0 is applied. This shows an important aspect of the application of a parking orbit: within one revolution about the Earth (≈ 1.5 hr) any angle between $\bar{V}_{\infty e}$ and \bar{V}_e can be realized by applying the impulsive shot at the right time. For a launch directly from the surface of the Earth (Direct Ascent), a specified angle between $\bar{V}_{\infty e}$ and \bar{V}_e can be realized only once a day.

To compute the velocity to which the spacecraft has to be accelerated in order to reach a specified value of $V_{\infty e}$, (8.15) can be applied:

$$V_0^2 = V_{esc_0}^2 + V_{\infty e}^2 \quad (18.5)$$

where V_{esc_0} is the escape velocity at the injection point and $V_{\infty e}$ is given by (8.12):

$$V_{\infty e}^2 = -\frac{\mu}{a} \quad (18.6)$$

where a is the semi-major axis of the hyperbolic escape trajectory and μ is the gravitational parameter of the Earth. The total energy of the spacecraft (per unit of mass) is given by (8.8)

$$\mathcal{E} = -\frac{\mu}{2a} = \frac{1}{2} V_{\infty e}^2 \quad (18.7)$$

In the literature on interplanetary spaceflight often the *injection energy*, C_3 , is used, which is defined as

$$C_3 = 2\mathcal{E} = V_{\infty e}^2 \quad (18.8)$$

Because $\Delta \bar{V}_0$ is oriented tangentially to the circular parking orbit, the magnitude of ΔV_0 required to reach a specified value of $V_{\infty e}$ can be obtained from (18.5):

$$\Delta V_0 = \sqrt{V_{esc_0}^2 + V_{\infty e}^2} - V_{c_0} = \sqrt{\frac{2\mu}{r_0} + V_{\infty e}^2} - \sqrt{\frac{\mu}{r_0}} \quad (18.9)$$

For *flyby missions*, where the spacecraft is not decelerated when it arrives at the target planet, one will be interested in interplanetary trajectories for which $\Delta V_0 = \text{minimum}$.

When the spacecraft has to land on the target planet, or has to be injected into an orbit about that planet, then it has to be decelerated. Here, we assume that the spacecraft has to enter a parking orbit about the target planet and that the deceleration maneuver is executed at pericenter of the hyperbolic trajectory of the spacecraft about the planet. Suppose that the deceleration maneuver occurs at a distance r_3 from the center of the planet and that after deceleration the spacecraft has a velocity V_{par3} relative to the planet. Of course, $V_{par3} < \sqrt{2\mu/r_3}$, where μ is the gravitational parameter of that planet; for $V_{par3} = \sqrt{\mu/r_3}$ the spacecraft enters a circular orbit about that planet. For the magnitude of the deceleration impulse, we find

$$\Delta V_3 = \sqrt{2V_{c_3}^2 + V_{\infty_t}^2} - V_{par_3} = \sqrt{\frac{2\mu}{r_3} + V_{\infty_t}^2} - V_{par_3} \quad (18.10)$$

where V_{c_3} is the circular velocity at the point of deceleration. For this type of missions, one will be interested in that interplanetary trajectory for which, for a specified total initial mass, the final mass of the spacecraft will be maximum. It is emphasized that only for the case that a single rocket engine delivers both the impulsive shot ΔV_0 and the impulsive shot ΔV_3 , and no mass is expelled or rejected between both impulses, the optimization criterion is $\Delta V_0 + \Delta V_3 = \text{minimum}$ (Chapter 12). However, these conditions are not realistic. In interplanetary flights, the rocket stage that delivers the impulsive shot ΔV_0 is separated after depletion of the propellants, the rocket engine that delivers the impulsive shot ΔV_3 has different characteristics than the rocket engine that delivers the impulsive shot ΔV_0 , and sometimes the spacecraft that has to enter an orbit about the planet separates from the ‘mother’ spacecraft when it approaches the target planet. For these reasons, in Section 18.4 only the criterion $\Delta V_0 = \text{minimum}$ will be used. In fact, this means that the escape maneuver and the deceleration maneuver are optimized separately, which, generally, is a good approximation of the situation in a real planetary mission.

18.4. Optimum two-dimensional direct-transfer trajectories

When the simplifying approximations introduced in Section 18.3 are applied, then the analysis of determining the optimum interplanetary transfer trajectory shows a great similarity with the analysis of determining the optimum transfer trajectory between two circular coplanar orbits about the Earth, as discussed in Section 12.1. Therefore, in this Section a number of relevant equations are copied from Section 12.1, without repeating their derivation.

A necessary condition for a feasible interplanetary transfer trajectory is

$$r_p \leq r_e \quad ; \quad r_a \geq r_t$$

where r_e and r_t are the radius of the orbit of the Earth and of the target planet about the Sun, respectively, and r_a and r_p are the apohelion and perihelion distance of the interplanetary transfer trajectory, respectively. When the notation

$$n = \frac{r_t}{r_e} \quad ; \quad z = \frac{p}{r_e} \quad (18.11)$$

is introduced, we can, just as we did in Section 12.1, derive from these relations the necessary conditions for an interplanetary transfer trajectory:

$$z \leq 1 + e \quad ; \quad z \geq n(1 - e) \quad (18.12)$$

Figure 18.4 shows the range of feasible interplanetary trajectories to Mars ($n = 1.524$) in the ez -diagram, according to the conditions (18.12). In this Figure the region of elliptical trajectories ($e < 1$), the line representing parabolic trajectories ($e = 1$) and the region of hyperbolic trajectories ($e > 1$) are indicated. The trajectory with the lowest value of the eccentricity is indicated by H . For that trajectory, we find from (18.12)

$$z = \frac{2n}{n+1} \quad ; \quad e = \frac{n-1}{n+1} \quad (18.13)$$

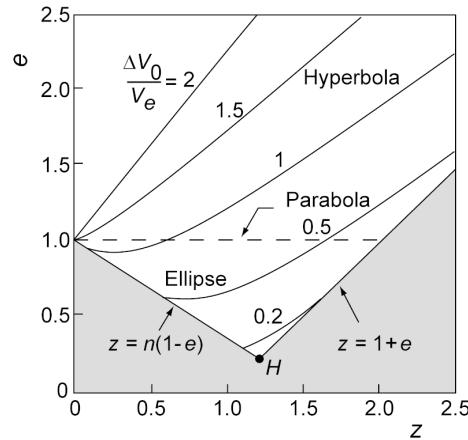


Figure 18.4: Constant ΔV_0 contours in the ez -diagram for two-dimensional trajectories to Mars. Assumptions: orbits of Earth and Mars are circular; 185 km altitude circular initial parking orbit.

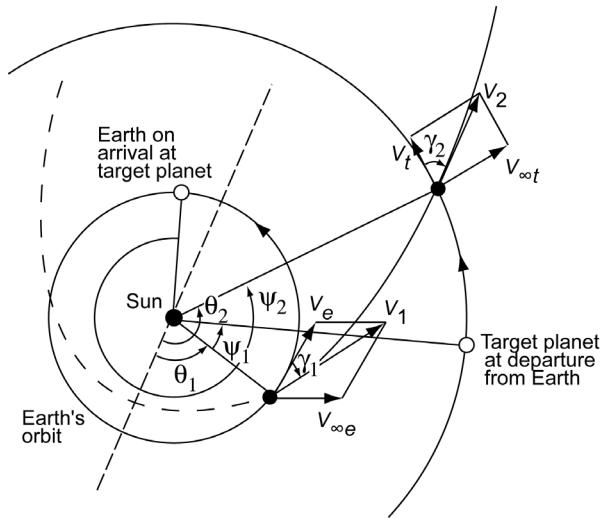


Figure 18.5: Geometry of two-dimensional interplanetary trajectories.

Figure 18.5 shows the geometry of a heliocentric interplanetary trajectory and the notations used for the analyses. Application of the cosine rule in the velocity diagrams indicated in this Figure results in

$$V_{\infty e}^2 = V_1^2 + V_e^2 - 2 V_1 V_e \cos \gamma_1 \quad ; \quad V_{\infty t}^2 = V_2^2 + V_t^2 - 2 V_2 V_t \cos \gamma_2 \quad (18.14)$$

Because the planets are assumed to move in circular orbits about the Sun, we have

$$\frac{V_t^2}{V_e^2} = \frac{r_e}{r_t} = \frac{1}{n} \quad (18.15)$$

Expressions for the parameters V_1 , V_2 , γ_1 and γ_2 of the heliocentric Keplerian transfer trajectory can be found in a way similar to the one applied in Section 12.1, and we find

$$V_1^2 = V_e^2 \left[2 - \frac{1 - e^2}{z} \right] \quad ; \quad V_2^2 = V_t^2 \left[2 - \frac{n}{z} (1 - e^2) \right] \quad (18.16)$$

$$\cos \gamma_1 = \frac{z}{\sqrt{2z - (1 - e^2)}} ; \quad \cos \gamma_2 = \frac{z/n}{\sqrt{2z/n - (1 - e^2)}} \quad (18.16)$$

Substitution of (18.15) and (18.16) into (18.14) yields, after some algebraic manipulation,

$$\frac{V_{\infty_e}}{V_e} = \sqrt{3 - 2\sqrt{z} - \frac{1 - e^2}{z}} ; \quad \frac{V_{\infty_t}}{V_e} = \sqrt{\frac{3 - 2\sqrt{z/n}}{n} - \frac{1 - e^2}{z}} \quad (18.17)$$

According to the criterion set in Section 18.3, the optimum transfer trajectory is characterized by the condition $\Delta V_0/V_e = \text{minimum}$. With (18.9) we may write

$$\frac{\Delta V_0}{V_e} = \sqrt{\frac{V_{esc_0}^2}{V_e^2} + \frac{V_{\infty_e}^2}{V_e^2} - \frac{V_{c_0}}{V_e}}$$

For a specified initial parking orbit about the Earth, the values of the parameters V_{c_0} en V_{esc_0} are known, and thus we may write with (18.17)

$$\frac{\Delta V_0}{V_e} = \sqrt{2c_1 + 3 - 2\sqrt{z} - \frac{1 - e^2}{z}} - \sqrt{c_1} \quad (18.18)$$

where $c_1 = (V_{c_0}/V_e)^2$. Figure 18.4 shows some curves of constant $\Delta V_0/V_e$ as a function of z en e , for a flight to Mars that starts from an initial circular parking orbit about the Earth at an altitude of 185 km ($c_1 = 6.85 \cdot 10^{-2}$).

Equation (18.18) is very similar to the first term on the right-hand side of (12.9), except for the appearance of c_1 in (18.18). Therefore, basically the same analysis as given in Section 12.1 can be applied to determine the optimum interplanetary transfer trajectory. We then arrive at the conclusion that point H in the ez -diagram indicates the optimum transfer trajectory. So, the heliocentric trajectory that requires a minimum ΔV_0 is an ellipse of which the values of e and z are given by (18.13); the apohelion distance of this ellipse is equal to r_t and the perihelion distance is equal to r_e (for a flight to an outer planet). This optimum interplanetary trajectory is tangent both to the orbit of the Earth and to the orbit of the target planet. This means that the velocity vectors \vec{V}_{∞_e} and \vec{V}_{∞_t} are also tangent to the orbits of both planets about the Sun, and we may write for a flight to an outer planet

$$V_1 = V_e + V_{\infty_e} ; \quad V_2 = V_t - V_{\infty_t} \quad (18.19-1)$$

and for a flight to an inner planet

$$V_1 = V_e - V_{\infty_e} ; \quad V_2 = V_t + V_{\infty_t} \quad (18.19-2)$$

This minimum-energy trajectory is called a *Hohmann trajectory* (Section 12.1), in honor of W. Hohmann (1880-1945) who has proposed this trajectory for interplanetary spaceflight in 1925.

Substitution of (18.13) into (18.17) and (18.18) yields for a Hohmann transfer trajectory

$$\left(\frac{V_{\infty_e}}{V_e} \right)_H = \sqrt{3 - 2\sqrt{\frac{2n}{n+1}} - \frac{2}{n+1}} ; \quad \left(\frac{V_{\infty_t}}{V_e} \right)_H = \sqrt{\frac{3 - 2\sqrt{\frac{2}{n+1}}}{n} - \frac{2}{n+1}} \quad (18.20)$$

$$\left(\frac{\Delta V_0}{V_e} \right)_H = \sqrt{2c_1 + 3 - 2\sqrt{\frac{2n}{n+1}} - \frac{2}{n+1}} - \sqrt{c_1} \quad (18.20)$$

For a given planetary mission (specified n) and a given parking orbit about the Earth (specified c_1), the values of $V_{\infty e}$, $V_{\infty t}$ and ΔV_0 corresponding to a Hohmann transfer trajectory can be computed from (18.20). If the parking orbit about the target planet is known, then the value of ΔV_3 can be computed from (18.10).

Some parameters of Hohmann trajectories to the planets are listed in Table 18.2. Because the planetary orbits are assumed to be circular (and coplanar), the value of n is set equal to the ratio between the semi-major axis of the orbit of the target planet and that of the orbit of the Earth. Note that for the outer planets the transfer trajectory has a high eccentricity. An important conclusion that can be drawn from this Table is that the flight time, t_f , to the outer planets is very long. For this reason, Hohmann trajectories to these planets are not very practical and faster direct-transfer trajectories are needed, although they require more propellant, or swingby trajectories (Section 18.11) have to be flown. The parameters T_{syn} , Ψ_1 , Ψ_2 and D will be discussed later. This Table also presents a survey of the characteristic velocities and impulsive shots associated with Hohmann transfer trajectories to the planets. For the computation of the impulsive shots it has been assumed that: 1) the spacecraft departs from a circular parking orbit at an altitude of 185 km above the Earth; 2) the spacecraft approaches the target planet along a

Table 18.2: Parameters of Hohmann transfer trajectories from the Earth to the planets. The orbits of the planets are assumed to be circular and coplanar. The velocity data refer to a departure from a 185 km altitude circular parking orbit about the Earth and a final circular parking orbit with a radius of 1.1 times the planet radius.

Target planet	n	a (AU)	e	t_f (yr)	T_{syn} (yr)	Ψ_1 (deg)	Ψ_2 (deg)	D (AU)
Mercury	0.3871	0.6935	0.4419	0.2888	0.317	108.3	76.0	0.981
Venus	0.7233	0.8617	0.1605	0.3999	1.599	306.0	36.0	0.594
Mars	1.5237	1.2618	0.2075	0.7087	2.135	44.3	284.8	1.594
Jupiter	5.2029	3.1014	0.6776	2.7310	1.092	97.2	276.8	5.180
Saturn	9.5367	5.2683	0.8102	6.0462	1.035	106.1	163.4	10.50
Uranus	19.189	10.095	0.9009	16.036	1.012	111.3	166.9	20.16
Neptune	30.070	15.535	0.9356	30.615	1.006	113.2	318.6	29.33
Pluto	39.482	20.241	0.9506	45.532	1.004	113.9	348.4	38.50

Target planet	V_e (km/s)	V_1 (km/s)	$V_{\infty e}$ (km/s)	V_t (km/s)	V_2 (km/s)	$V_{\infty t}$ (km/s)	ΔV_0 (km/s)	ΔV_3 (km/s)
Mercury	29.78	22.25	7.53	47.87	57.48	9.61	5.56	7.56
Venus	29.78	27.29	2.49	35.02	37.73	2.71	3.51	3.26
Mars	29.78	32.73	2.94	24.13	21.48	2.65	3.61	2.09
Jupiter	29.78	38.58	8.79	13.06	7.41	5.64	6.31	16.90
Saturn	29.78	40.07	10.29	9.64	4.20	5.44	7.28	10.34
Uranus	29.78	41.06	11.28	6.80	2.14	4.66	7.98	6.47
Neptune	29.78	41.44	11.65	5.43	1.38	4.05	8.25	6.92
Pluto	29.78	41.60	11.81	4.74	1.05	3.69	8.36	3.04

hyperbolic trajectory with a minimum distance of $r_3 = 1.1R$, where R is the (equatorial) radius of the planet⁴, and is then decelerated to enter a circular orbit with a radius of $1.1R$ about that planet. From this Table a number of conclusions can be drawn, e.g.:

- The required ΔV_0 is minimum for a mission to Venus; then follow flyby missions to Mars, Mercury, Jupiter and the outer planets.
- The required ΔV_3 to enter a circular orbit with a radius of $1.1R$ about a target planet is minimum for a mission to Mars. For the giant planets Jupiter and Saturn ΔV_3 is very large.
- If $\Delta V_0 = 8.74$ km/s, then $V_{\infty_e} = 12.34$ km/s and $V_1 = 42.12$ km/s, which is just the escape velocity from the Sun's gravity field at the (mean) distance of the Earth from the Sun. So, for this ΔV_0 the spacecraft will escape from the solar system. Note that this ΔV_0 is only 6% larger than the ΔV_0 required to reach the planet Neptune along a Hohmann trajectory.
- Except for the planets Venus and Mars, it costs less propellant to escape from the solar system than to fly along a Hohmann trajectory to a planet and to enter a circular orbit with radius $1.1R$ about that planet.

The modern heavy-lift launchers Delta 4050H-19 and Atlas 551 can inject a payload of about 2900 kg and 2250 kg, respectively, into an escape trajectory with $V_{\infty_e} = 7.5$ km/s after a launch due east from Cape Canaveral; this velocity corresponds to $V_0 = 13.3$ km/s and $\Delta V_0 = 5.5$ km/s when injection takes place at 185 km altitude. The Ariane 5 ECA launcher can inject a payload of about 4100 kg into an escape trajectory with $V_{\infty_e} = 3.5$ km/s after a launch due east from Kourou; this velocity corresponds to $V_0 = 11.6$ km/s and $\Delta V_0 = 3.8$ km/s when injection takes place at 185 km altitude. Comparing these values with the ones listed in Table 18.2, we conclude that even these heavy-lift launchers cannot launch a payload with a mass of more than 2 ton to Jupiter and beyond along a minimum-energy Hohmann trajectory. So, additional so-called *kick stages*, more-powerful launchers or swingby techniques (Section 18.11) are needed to launch appreciable payloads to the outer planets.

18.5. Insertion into an orbit about the target planet

A deceleration maneuver to enter an orbit about the target planet, generally, has to satisfy various requirements. In this context, it is noted that the initial parking orbit about the Earth always is (near-)circular and at a low altitude. Any higher circular orbit or elliptical orbit costs more propellant. This argument does not hold for the final orbit about a target planet. There, the requirement of low propellant consumption leads to a highly eccentric parking orbit. If the orbit about the planet is specified, then it is most efficient to apply the impulsive shot at pericenter of this orbit and perpendicular to the radius vector (Section 1.7). By applying small maneuvers during the interplanetary flight (*mid-course corrections*), the pericenter distance, r_3 , of the hyperbolic flyby trajectory can be adjusted such that it equals the pericenter distance of the desired orbit about the planet, while the value of V_{∞_t} remains about constant. After the deceleration maneuver the velocity should be smaller than the local escape velocity, and we find for the magnitude of the deceleration impulsive shot from (18.10)

$$\Delta V_3 > \sqrt{\frac{2\mu}{r_3}} + V_{\infty_t}^2 - \sqrt{\frac{2\mu}{r_3}} \quad (18.21)$$

where μ is the gravitational parameter of the target planet and V_{∞_t} follows from the interplanetary trajectory flown. For increasing values of ΔV_3 , the eccentricity of the parking orbit, of course,

⁴ For simplicity, in this Chapter we neglect the eventual existence of a planetary atmosphere or, where appropriate, assume that the radius R includes the appreciable part of the atmosphere.

decreases. The lowest eccentricity is reached if the velocity after deceleration is just equal to the local circular velocity, which requires a deceleration impulsive shot with magnitude

$$\Delta V_3 = \sqrt{\frac{2\mu}{r_3} + V_{\infty_t}^2} - \sqrt{\frac{\mu}{r_3}} \quad (18.22)$$

For the numerical examples presented in this Section, we assume that V_{∞_t} is 10% higher than the value of V_{∞_t} corresponding to a Hohmann transfer trajectory to a particular planet, which is listed in Table 18.2. We then find for a Mars orbiter mission with $r_3 = 3746$ km ($h_3 = 350$ km): $\Delta V_3 > 0.82$ km/s; for $\Delta V_3 = 2.22$ km/s the spacecraft will enter a circular orbit at an altitude of 350 km about Mars. These values provide some feeling for the sensitivity of the orbit on the magnitude of the deceleration impulsive shot.

In general, the orbit about the target planet has to satisfy certain mission criteria. Below, a few special cases will be analyzed in some detail and expressions for the required deceleration impulse will be given. In all cases, it is assumed that the deceleration maneuver is applied at pericenter of the final orbit about the planet.

Orbits with specified eccentricity

In Section 6.3 we have derived an expression ((6.23)) for the velocity at pericenter of an elliptical orbit, V_p . With the notation used in this Chapter, that expression reads for an orbit about the target planet:

$$V_{p_3} = V_{c_3} \sqrt{1+e} = \sqrt{\frac{\mu}{r_3} (1+e)}$$

Substitution of this expression into (18.10) yields

$$\Delta V_3 = \sqrt{\frac{2\mu}{r_3} + V_{\infty_t}^2} - \sqrt{\frac{\mu}{r_3} (1+e)} \quad (18.23)$$

Note that for any value of r_3 and V_{∞_t} , ΔV_3 decreases when e increases. So, ΔV_3 is, for specified values of r_3 and V_{∞_t} , minimum if the parking orbit is almost parabolic. To investigate whether there exists an optimum pericenter distance, r_3 , for which ΔV_3 is minimum for specified values of V_{∞_t} and e , we differentiate (18.23) partially with respect to r_3 and find

$$\frac{\partial \Delta V_3}{\partial r_3} = \frac{1}{2} \frac{\mu}{r_3^2} \left[\sqrt{\frac{1+e}{\frac{\mu}{r_3}}} - \frac{2}{\sqrt{\frac{2\mu}{r_3} + V_{\infty_t}^2}} \right]$$

A numerical analysis of this expression shows that, for any value of V_{∞_t} and $e < 1$, the expression on the right-hand side has one root and runs from negative to positive over the range $0 < r_3 < R_{s,i}$, where $R_{s,i}$ is the radius of the sphere of influence. This implies that ΔV_3 is minimum when $\partial \Delta V_3 / \partial r_3 = 0$, and that ΔV_3 keeps increasing at increasing values of r_3 larger than the optimum value of r_3 for which ΔV_3 is minimum. Setting the expression on the right-hand side equal to zero, we obtain for the optimum value of r_3

$$\left(\frac{r_3}{R} \right)_{opt} = \left(\frac{V_{esc,surf}}{V_{\infty_t}} \right)^2 \frac{1-e}{1+e} \quad (18.24)$$

where

$$V_{esc_{surf}} = \sqrt{\frac{2\mu}{R}}$$

is the escape velocity at the surface of the planet. Of course, this optimum only has a physical meaning if $(r_3/R)_{opt} \geq 1$. For the limiting case $(r_3/R)_{opt} = 1$, we find from (18.24)

$$e_{lim} = \frac{1 - (V_{\infty_t}/V_{esc_{surf}})^2}{1 + (V_{\infty_t}/V_{esc_{surf}})^2} < 1$$

So, two cases can be distinguished, depending on the value of e . For these cases, we find with (18.23)

$$\begin{aligned} e < e_{lim}: \left(\frac{r_3}{R} \right)_{opt} &= \left(\frac{V_{esc_{surf}}}{V_{\infty_t}} \right)^2 \frac{1-e}{1+e} ; \quad (\Delta V_3)_{min} = \frac{1}{2}\sqrt{2} V_{\infty_t} \sqrt{1-e} \\ e \geq e_{lim}: \left(\frac{r_3}{R} \right)_{opt} &= 1 ; \quad (\Delta V_3)_{min} = V_{esc_{surf}} \left[\sqrt{1 + \left(\frac{V_{\infty_t}}{V_{esc_{surf}}} \right)^2} - \sqrt{\frac{1+e}{2}} \right] \end{aligned} \quad (18.25)$$

Note that for $e < e_{lim}$ the optimum value of r_3/R decreases with increasing values of e and increasing values of V_{∞_t} ; for $e \geq e_{lim}$ the optimum value of r_3/R is always one. The value of $(\Delta V_3)_{min}$ decreases with increasing values of e and decreasing values of V_{∞_t} . With the assumption about the value of V_{∞_t} mentioned above, we find for the planets Venus, Mars and Jupiter: $e_{lim} = 0.847$, $e_{lim} = 0.496$ and $e_{lim} = 0.978$, respectively. For an orbit with e.g. $e = 0.7$, we find from (18.25) for Venus: $(r_3/R)_{opt} = 2.13$, $(\Delta V_3)_{min} = 1.15$ km/s; for Mars: $(r_3/R)_{opt} = 1$, $(\Delta V_3)_{min} = 1.18$ km/s; and for Jupiter: $(r_3/R)_{opt} = 16.25$, $(\Delta V_3)_{min} = 2.40$ km/s.

The absolute minimum value of ΔV_3 is obtained when $e \approx 1$, so for an almost parabolic parking orbit. For the limiting case $e = 1$ we find from (18.25-2)

$$\left(\frac{r_3}{R} \right)_{opt} = 1 ; \quad (\Delta V_3)_{min,abs} = V_{esc_{surf}} \left[\sqrt{1 + \left(\frac{V_{\infty_t}}{V_{esc_{surf}}} \right)^2} - 1 \right] \quad (18.26)$$

It is noted that $r_3/R = 1$ implies a zero pericenter altitude, which is impossible, and the minimum pericenter altitude is determined by the acceptable level of atmospheric drag acting on the spacecraft and by the acceptable level of charged particles radiation in the planet's radiation belts. However, the analysis shows that the absolute minimum value of ΔV_3 corresponds to an orbit about the planet that has a very low pericenter altitude and a very high eccentricity ($e < 1$). When we consider, as an example, missions to Venus, Mars and Jupiter, we find, with the assumption about V_{∞_t} mentioned above, from (18.26): $(\Delta V_3)_{min,abs} = 0.42$ km/s, 0.78 km/s, and 0.32 km/s, respectively.

For many planetary orbiters a circular orbit is very attractive because of the constant observation altitude. For this case $e = 0$, and (18.25-1) simplifies to

$$\left(\frac{r_3}{R} \right)_{opt} = \left(\frac{V_{esc_{surf}}}{V_{\infty_t}} \right)^2 ; \quad (\Delta V_3)_{min} = \frac{1}{2}\sqrt{2} V_{\infty_t} \quad (18.27)$$

When we substitute the respective values of $V_{\infty t}$ into (18.27), we find for Mercury and Pluto: $(r_3/R)_{opt} < 1$; for all other planets, except Mars, $(r_3/R)_{opt} > 12$, which means that for these planets optimum circular orbits are of little practical value. For Mars, we find: $(r_3/R)_{opt} = 2.97$, $h_3 = 6691$ km, $(\Delta V_3)_{min} = 2.06$ km/s. To give an impression of the variation of ΔV_3 in the region around the optimum altitude, we conclude from (18.27) that for the optimum value of r_3 : $\Delta V_3 = 0.71 V_{\infty t}$. When we substitute $e = 0$, $r_3/R = 2(r_3/R)_{opt}$ into (18.23) we find: $\Delta V_3 = 0.72 V_{\infty t}$; for $e = 0$, $r_3/R = 0.5(r_3/R)_{opt}$ we find: $\Delta V_3 = 0.73 V_{\infty t}$. So, we conclude that although there exists an optimum value of r_3 , the variation of ΔV_3 is quite ‘flat’ when r_3 varies over a significant range about its optimum value.

Orbits with specified semi-major axis

The semi-major axis determines the period of the orbit about the planet and therefore governs the observation program that can be executed by the spacecraft. It is emphasized that for a specified value of a , the apocenter distance decreases if the pericenter distance, r_3 , increases, while for a specified value of e , which case was discussed before, the apocenter distance increases if r_3 increases. Using the *vis-viva integral* (6.21), (18.10) may be rewritten as

$$\Delta V_3 = \sqrt{\frac{2\mu}{r_3} + V_{\infty t}^2} - \sqrt{\frac{2\mu}{r_3} - \frac{\mu}{a}} \quad (18.28)$$

To investigate whether there exists an optimum pericenter distance for which ΔV_3 is a minimum for specified values of $V_{\infty t}$ and a , we differentiate (18.28) partially with respect to r_3 and find

$$\frac{\partial \Delta V_3}{\partial r_3} = \frac{\mu}{r_3^2} \left[\sqrt{\frac{1}{\frac{2\mu}{r_3} - \frac{\mu}{a}}} - \sqrt{\frac{1}{\frac{2\mu}{r_3} + V_{\infty t}^2}} \right]$$

Since $a > 0$, this expression shows that $\partial \Delta V_3 / \partial r_3 > 0$ for all values of r_3 , which means that ΔV_3 will increase continuously for increasing values of r_3 . So, the minimum deceleration impulse ΔV_3 is required if the deceleration altitude (pericenter of orbit) is as low as possible. When, for simplicity, we again assume that the minimum value of r_3 is equal to R , we find from (18.28)

$$(\Delta V_3)_{min} = V_{esc_{surf}} \left[\sqrt{1 + \left(\frac{V_{\infty t}}{V_{esc_{surf}}} \right)^2} - \sqrt{1 - \frac{R}{2a}} \right] \quad (18.29)$$

Note that the value of $(\Delta V_3)_{min}$ increases if $V_{\infty t}$ increases or a decreases. An elementary analysis shows that the apocenter altitude and eccentricity of the resulting orbit are then given by $h_a = 2(a - R)$, $e = 1 - R/a$. The absolute minimum deceleration impulsive shot corresponds to the case that after deceleration the orbit is a parabola, for which $a = \infty$. In that case, (18.29) becomes identical to (18.26), as could be expected.

With the expression for the orbital period, T , of an elliptical orbit ((6.25)), (18.28) can be written as

$$\Delta V_3 = \sqrt{\frac{2\mu}{r_3} + V_{\infty t}^2} - \sqrt{\frac{2\mu}{r_3} - \left(\frac{2\pi\mu}{T} \right)^{2/3}} \quad (18.30)$$

When we consider orbits about Mars with a pericenter altitude of 350 km and an orbital period of 5 hr, 10 hr, 50 hr, and 300 hr, we find: $\Delta V_3 = 1.50$ km/s, 1.24 km/s, 0.96 km/s, and 0.86 km/s,

respectively. A special case is a *Mars surface synchronous orbit*, which is the Martian equivalent of a geosynchronous orbit. Since the rotation period of Mars is 24.623 hr, we find from (6.25): $a = 20,427$ km. From (18.30) we find for a pericenter altitude of 350 km: $\Delta V_3 = 1.04$ km/s. For a circular surface synchronous orbit ('Mars-stationary' orbit): $r_3 = 20,427$ km = $6.01 R$, $\Delta V_3 = 2.11$ km/s. Note that the ratio r_3/R for this orbit is reasonably close to the value of $r/R = 6.61$ for a circular geosynchronous orbit (geostationary orbit).

Aerobraking

If the target planet has an appreciable atmosphere, atmospheric drag generated by the spacecraft's motion through that atmosphere can be used to decelerate the spacecraft, thereby reducing the ΔV to be delivered by the spacecraft's thrusters, which saves mass.

The concept of *aerobraking* can be described as follows. By applying appropriate mid-course maneuvers, the spacecraft enters the planet's sphere of influence at the right position and with the right velocity, leading to a selected value of the pericenter distance of the flyby hyperbola. When the spacecraft passes pericenter, a rocket engine is ignited and a deceleration maneuver is applied that brings the spacecraft into a highly eccentric elliptical orbit about that planet. After check-out of the spacecraft's health, the *walk-in period*, which lasts about a week, starts. During this period, the spacecraft's thrusters are fired at successive apocenter passages to slightly decelerate the spacecraft and thereby gradually lower the pericenter from its original orbit insertion altitude to its aerobraking altitude. This phase is used as a calibration period to improve the models that describe the effects of the atmosphere on the spacecraft in the aerobraking region. Once the orbiter has reached its operational 'drag orbit' pericenter altitude, the *main phase* of aerobraking begins. During this phase, which lasts about half a year, the apocenter altitude continuously decreases due to the drag experienced at the pericenter region. It will be clear that aerobraking is an extremely sensitive technique. The pericenter altitude should be low enough to generate an appreciable drag, but not too low in order to limit the forces acting on the spacecraft, and in particular on its fragile solar panels, and the heat dissipated to the spacecraft. Therefore, small propulsive maneuvers at apocenter are occasionally required to raise or lower the pericenter altitude somewhat, depending on the actual atmospheric density at the pericenter region. When the apocenter altitude has decreased to a value close to its desired final value, the *walk-out phase* starts, which lasts about a week. Thrusters are fired at apocenter to slightly accelerate the spacecraft and thereby gradually increase the pericenter altitude, causing the orbit to shrink more slowly. When the apocenter has reached the desired altitude, thruster firings at apocenter raise the pericenter altitude out of the atmosphere to the desired altitude and the aerobraking maneuver is completed. A few additional small maneuvers are required to fine tune the final orbit.

The aerobraking technique was first applied in the Magellan mission (launch May 4, 1989) to Venus, and then in the Mars Global Surveyor mission (launch November 7, 1996), the Mars Odyssey mission (launch April 7, 2001), and the Mars Reconnaissance Orbiter mission (launch August 12, 2005). During these missions, the pericenter altitude in the main phase of aerobraking was as low as about 140 km for the mission to Venus and about 100 km for the missions to Mars. The Mars Climate Orbiter, launched on December 11, 1998, also aimed for the application of aerobraking. However, when the spacecraft approached Mars on September 23, 1999, a mix-up of English and metric units in the transmission of commands for the preparation of the orbit insertion maneuver actually brought the pericenter at an altitude of only 57 km instead of the planned 140 - 150 km altitude. At this low altitude, the atmospheric density was much too high, resulting in a disintegration and burn up of the spacecraft in the Martian atmosphere. As an example of a successful mission, the aerobraking maneuver of the Mars Reconnaissance Orbiter

will be described in some detail.

Mars Reconnaissance Orbiter arrived at Mars on March 10, 2006. Its interplanetary trajectory made it to pass under the southern hemisphere at an altitude of about 300 km, traveling at about 3 km/s. The desired final orbit was an elliptical one with pericenter over the south pole at an altitude of 255 km and apocenter over the north pole with an altitude of 320 km; the period of that orbit is 1 hr 53 min. When the spacecraft passed the pericenter of its hyperbolic trajectory, the six 170 N thrusters of the spacecraft's orbit maneuvering system were fired for about 25 min to slow down the spacecraft by about 1 km/s and to put it into a highly eccentric orbit with pericenter and apocenter altitudes of 300 km and 45,000 km, respectively; the period of this orbit is about 35 hr. Aerobraking started on April 5, 2006, by firing the thrusters at apocenter (in the direction opposite to the spacecraft's velocity vector) to initiate a 5-day walk-in phase, during which thruster burns near apocenter incrementally lowered the pericenter distance enough for the spacecraft to begin encountering atmospheric drag. Once the pericenter was lowered to the altitude where the optimum amount of atmospheric drag was found, the main phase of aerobraking began. In this phase, more than 500 carefully calculated aerobraking dips into the upper atmosphere occurred over the course of about six months; during this phase, the pericenter altitude ranged from 98 to 110 km. Experience gained from earlier US Mars orbiters, plus constant monitoring of the realized orbit changes, has aided calculations of how low to bring the pericenter from day to day. In its configuration for aerobraking, the orbiter presented a frontal area of 37.5 m^2 , including the back surfaces of the solar panels and high-gain antenna, to the incoming gas flow. The more compact and massive main structure of the spacecraft was out in front of these larger surfaces in the direction of motion. This configuration gives the spacecraft a self-righting stability. The latitude of the pericenter started near 65° S , moved southward past the winter pole and was almost back to the equator when the walk-out phase started in the last week of August. When the apocenter altitude had been reduced to 450 km, an exit maneuver on August 30 ended aerobraking. Three subsequent propulsive orbit adjustment maneuvers then established the spacecraft's final orbit. By using this aerobraking technique, the project saved carrying 600 kg of additional propellant for the spacecraft's thrusters system to Mars.

18.6. Faster two-dimensional direct-transfer trajectories

If the spacecraft is accelerated along its escape trajectory from the Earth to a velocity somewhat higher than required to enter a Hohmann transfer trajectory, then the spacecraft moves, in general, along a segment of a heliocentric ellipse that intersects both the orbit of the Earth and that of the target planet at two points (Figure 18.6). This implies that for a heliocentric elliptical trajectory with specified values of e and z , theoretically four different missions can be flown where different segments of the same ellipse are traversed (sketches a) to d)); these four missions require the same amount of propellant. Obviously, the flight indicated by sketch a) leads to the shortest flight time. It is emphasized that in all cases the spacecraft moves in the same direction

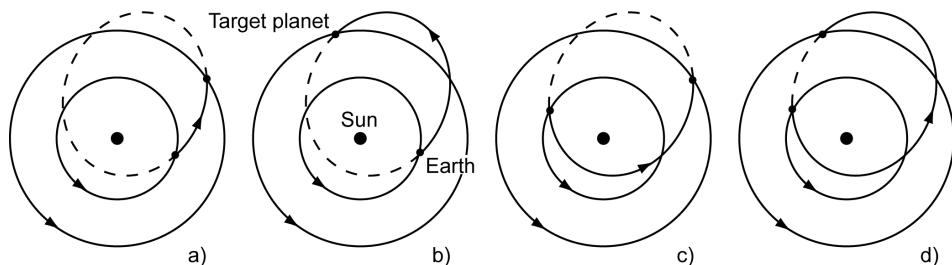


Figure 18.6: Four ways of reaching the target planet along the same elliptical trajectory.

as the planets about the Sun. When the spacecraft is accelerated in its parking orbit about the Earth to a velocity higher than the local escape velocity with respect to the solar gravity field ($V_1 \approx 42.1$ km/s), then the heliocentric trajectory will be a hyperbola. In that case, of course, only the flights indicated by sketches a) and c) are feasible.

When for these faster trajectories the impulse ΔV_0 is applied in such a way that $\bar{V}_{\infty e}$ coincides with \bar{V}_e , which requires $\gamma_1 = 0^\circ$, then the perihelion distance of the heliocentric transfer trajectory is r_e . In other words: the line $z = 1 + e$ in Figure 18.4 represents the collection of all heliocentric transfer trajectories that touch the orbit of the Earth. Likewise, the line $z = n(1 - e)$ in Figure 18.4 represents the collection of all heliocentric transfer trajectories that touch the orbit of the target planet ($\gamma_2 = 0^\circ$). So, it may be concluded from Figure 18.4 that for a specified value of ΔV_0 larger than the value required for the Hohmann trajectory, an interplanetary transfer trajectory can be flown with $\gamma_1 = 0^\circ, \gamma_2 \neq 0^\circ; \gamma_1 \neq 0^\circ, \gamma_2 = 0^\circ; \gamma_1 \neq 0^\circ, \gamma_2 \neq 0^\circ$. Another conclusion is that there exist hyperbolic transfer trajectories that touch the orbit of the Earth, which require less energy than elliptical transfer trajectories that touch the orbit of the target planet. For all trajectories with the same value of ΔV_0 , the arrival conditions at the target planet and the flight time to the target planet will, generally, be different. Which trajectory is chosen, depends on the requirements set for the flight time, distance between target planet and Earth upon arrival at the target planet, direction and magnitude of $\bar{V}_{\infty t}$, etc.

In order to prevent the flight times to become unrealistically long, the spacecraft should be injected with not too small a ΔV_0 . However, the possibilities are limited by the performance of available launchers. Figure 16.5 shows a diagram of the launch vehicle payload capability versus the launch vehicle characteristic velocity, V_{char} , (Section 16.4) for some launch vehicles. This Figure shows that, for a reasonable payload capability, a characteristic velocity of about 16 km/s may be considered as an upper limit, even for the most-powerful launchers. Only the Saturn V/Centaur four-stage launcher, which has only existed on the drawing table, would have achieved a higher V_{char} . A characteristic velocity of 16 km/s corresponds to a ΔV_0 of about 8 km/s. Table 18.2 shows that even a Hohmann transfer trajectory to the planets Uranus, Neptune or Pluto requires a ΔV_0 of 8.0 - 8.4 km/s. So, the limited capabilities of the present launchers with their chemical propulsion systems seriously constrain the execution of fast direct interplanetary missions. This is precisely the reason why swingby flights (Section 18.11) are chosen nowadays for most planetary missions.

We now consider the case of transfer trajectories that touch the orbit of the Earth ($\gamma_1 = 0^\circ, \theta_1 = 0^\circ$) and assume that the flight of the spacecraft corresponds to sketch a) of Figure 18.6; for this case: $0^\circ < \gamma_2 \leq 90^\circ, \theta_2 < 180^\circ$. These trajectories form a characteristic class of faster interplanetary trajectories. In Figure 18.4, these trajectories all lie on the line $z = 1 + e$. Substitution of this relation and (18.15) into (18.16) and (18.17) yields

$$\begin{aligned} \frac{V_1}{V_e} &= \sqrt{z} \quad ; \quad \frac{V_2}{V_e} = \sqrt{z - 2 + 2/n} \quad ; \quad \cos \gamma_2 = \frac{z/n}{\sqrt{z^2 - 2z(1 - 1/n)}} \\ \frac{V_{\infty e}}{V_e} &= \sqrt{z} - 1 \quad ; \quad \frac{V_{\infty t}}{V_e} = \sqrt{z + \frac{3 - 2\sqrt{z/n}}{n} - 2} \end{aligned} \tag{18.31}$$

The heliocentric transfer angle covered by the spacecraft during its motion about the Sun is found by substitution of $r = r_t, \theta = \theta_2, e = z - 1$, and (18.11) into the equation of a conic section (5.22). We then find

$$\cos \theta_2 = \frac{z/n - 1}{z - 1} \quad (18.32)$$

The flight time to the target planet can be found by applying the relations between Θ and time that hold for elliptical, parabolic or hyperbolic trajectories (Section 6.5, 7.3 or 8.3). For example, if the interplanetary transfer trajectory is an ellipse, then

$$t_f = \sqrt{\frac{a^3}{\mu_s}} (E_2 - e \sin E_2) \quad (18.33)$$

where μ_s is the gravitational parameter of the Sun and E_2 is the eccentric anomaly upon arrival at the target planet. Using the relations $e = z - 1$, (6.2) and (18.11), we find

$$\sqrt{\frac{a^3}{\mu_s}} = \frac{r_e}{V_e} \sqrt{\frac{1}{(2 - z)^3}} \quad (18.34)$$

For the angle E_2 we write according to (6.34) and (6.35):

$$\tan \frac{E_2}{2} = \sqrt{\frac{1 - e}{1 + e}} \tan \frac{\theta_2}{2} \quad (18.35)$$

where

$$\tan \frac{\theta_2}{2} = \sqrt{\frac{1 - \cos \theta_2}{1 + \cos \theta_2}} \quad (18.36)$$

Substitution of $e = z - 1$, (18.32) and (18.36) into (18.35) yields

$$\tan \frac{E_2}{2} = \sqrt{\frac{(2 - z)(1 - 1/n)}{z(1 + 1/n) - 2}} \quad (18.37)$$

The relations (18.33), (18.34) and (18.37) yield the flight time as a function of z and n . Some important conclusions concerning this class of transfer trajectories can be drawn from (18.11), (18.31) to (18.33). Because V_1 increases at increasing values of ΔV_0 , (18.31-1) shows that consequently also the value of z increases at increasing values of ΔV_0 . So, the equations show that for a specified target planet (n fixed) and increasing values of ΔV_0 (increasing values of z):

- The velocity V_2 and the flight path angle γ_2 increase.
- The velocity V_{∞} increases. Consequently, also the velocity V_3 increases for a fixed value of r_3 . This implies that faster interplanetary trajectories require larger deceleration impulses to enter a specified orbit about the target planet.
- The semi-major axis and eccentricity of the transfer trajectory increase.
- The heliocentric transfer angle, $\Delta\theta = \theta_2$, and the flight time, t_f , decrease.

Some parameters of two-dimensional trajectories with $\gamma_1 = 0^\circ$ to Mars and Jupiter are plotted as a function of ΔV_0 in Figure 18.7. It was assumed that the missions start from a 185 km altitude parking orbit about the Earth and that the (minimum) flyby distance is $2R$, where R is the planet's (equatorial) radius. From this Figure, it can be concluded that even a small increase of ΔV_0 above the value required for a Hohmann transfer trajectory leads to a significantly shorter flight time. For a Hohmann transfer to Mars the flight time is 259 days and the required ΔV_0 is 3.61 km/s, leading to a mission characteristic velocity, V_{char}^* , of 11.42 km/s ($\approx 37,600$ ft/s). Decreasing the

flight time by 50% requires a ΔV_0 increase of only 0.7 km/s ($\approx 19\%$), leading to a V_{char}^* of 12.12 km/s ($\approx 39,700$ ft/s). According to Figure 16.5, an Atlas/Centaur/TE 364 launcher can inject a payload of 1900 lb (860 kg) into this faster trajectory, compared to a payload of about 2600 lb (1180 kg) for the Hohmann flight. So, halving the flight time corresponds to a reduction of 27% in the payload capability. From Figure 18.7 we conclude that for flight times of less than 70 days to Mars, the transfer trajectory is a hyperbola and the heliocentric transfer angle, $\Delta\theta$, is less than 75° . These transfers would require an impulsive shot in the parking orbit of $\Delta V_0 > 8.6$ km/s or $V_{char}^* > 16.4$ km/s, which would lead even for the most powerful launchers to a payload mass of less than 400 kg (Figure 16.5). A characteristic phenomenon, in particular for flights to Venus and Mars, is the fast increase of the flyby velocity for a given value of r_3 at decreasing flight times. If the spacecraft has to enter a specified orbit about the target planet, one will aim at a relatively small value of V_3 , in order to limit the amount of propellant needed for the deceleration maneuver. Therefore, for planetary orbiter missions direct transfer trajectories that are relatively close to a Hohmann transfer trajectory are preferable.

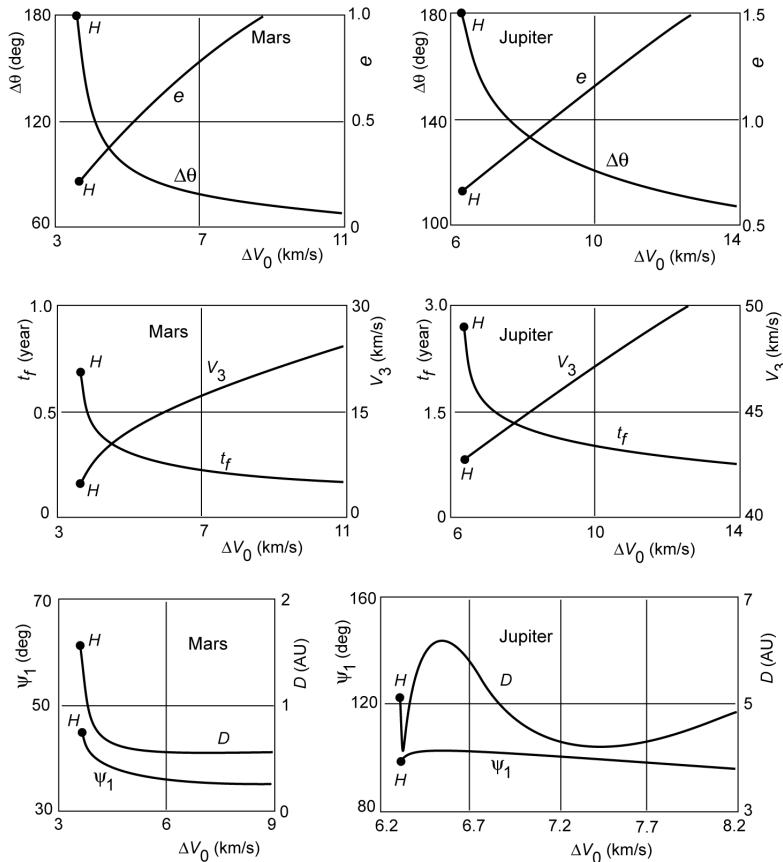


Figure 18.7: Parameters of two-dimensional trajectories with $\gamma_i = 0^\circ$ to Mars (left) and Jupiter (right) as a function of ΔV_0 . Assumptions: orbits of the planets are circular; 185 km altitude initial parking orbit; minimum flyby distance: $2R$.

18.7. Launch opportunities

A necessary condition for an interplanetary trajectory is that when the spacecraft crosses the orbit of the target planet about the Sun, the planet is at that position in its orbit. For a particular interplanetary transfer trajectory and flight time, this requirement defines the launch time. For a two-dimensional transfer, we conclude from Figure 18.5 that the geometric configuration of

Earth and target planet at the time of launch should satisfy the relation

$$\Psi_1 = (\theta_2 - \theta_1) - n_t t_f \quad (18.38)$$

where n_t is the (mean) angular (orbital) motion of the target planet in its orbit about the Sun and t_f is the flight time. For a specified interplanetary transfer trajectory, the parameters θ_1 , θ_2 and t_f are known. Then, the required configuration of Earth and target planet at launch, and thus the time of launch, can be determined from (18.38). According to Figure 18.5, the planetary configuration at the time of arrival at the target planet should satisfy the relation

$$\Psi_2 = (\theta_2 - \theta_1) - n_e t_f \quad (18.39)$$

where n_e is the (mean) angular (orbital) motion of the Earth in its orbit about the Sun. From this relation, the configuration of Earth and target planet upon arrival at the target planet can be computed. The parameters Ψ_1 and Ψ_2 are called the *planet phase angle at departure* and the *planet phase angle at arrival*, respectively. As a special case, we now consider a Hohmann transfer trajectory to an outer planet, for which $\theta_{1H} = 0^\circ$, $\theta_{2H} = 180^\circ$. We then obtain from (18.38) and (18.39)

$$\Psi_{1H} = \pi - n_t t_{fH} ; \quad \Psi_{2H} = \pi - n_e t_{fH} \quad (18.40)$$

The values of these angles are listed in Table 18.2 for all target planets.

The angle Ψ_2 is an important parameter in the selection of suitable interplanetary trajectories. When for a flight to an inner or outer planet $\Psi_2 \approx 180^\circ$, then the Sun is between the spacecraft and the Earth when the spacecraft arrives at the target planet. This will block or at least seriously hamper the receipt of signals transmitted by the spacecraft, because the ground antenna will then ‘look’ at the Sun and will also receive the solar emission at radio-frequency wavelengths. In addition, then the communication distance, D , between spacecraft and Earth is very large, which requires a high-power transmitter on the spacecraft. For a flight to an outer planet, $\Psi_2 \approx 0^\circ$ is attractive because of the short communication distance. However, then the spacecraft will ‘look’ at the Sun leading to the same type of communication problems. When for a flight to an inner planet $\Psi_2 \approx 0^\circ$, then the Sun is ‘behind’ the target planet when the spacecraft arrives at the target planet. Again, this will seriously hamper the receipt of signals transmitted by the spacecraft. In general, a Sun-Earth-spacecraft angle of larger than 3° is pursued so that the spacecraft stays outside the region where solar interference degrades spacecraft communication with Earth ground stations. The communication distance, D , can be determined from the relation (Figure 18.5)

$$D^2 = r_e^2 + r_t^2 - 2r_e r_t \cos(2\pi - \Psi_2)$$

or, with (18.11),

$$\frac{D^2}{r_e^2} = 1 + n^2 - 2n \cos \Psi_2 \quad (18.41)$$

For a Hohmann transfer trajectory, substitution of (18.40) into (18.41) gives

$$\left(\frac{D}{r_e} \right)_H = \sqrt{1 + n^2 + 2n \cos(n_e t_{fH})} \quad (18.42)$$

This value of $(D/r_e)_H$ is also listed in Table 18.2 for all target planets.

When, at a certain time t_0 , the relative configuration of Earth and target planet is indicated

by the angle $\Psi_{1,0}$, then at a time $t > t_0$ the angle Ψ_1 will differ from $\Psi_{1,0}$ due to the different angular motions of the Earth and the target planet in their orbits about the Sun. After a certain period, which is called the *synodic period*, T_{syn} , the relative configuration of Earth and target planet is $|\Psi_1 - \Psi_{1,0}| = 2\pi$. So, when at a certain time t_0 the configuration of Earth and target planet is ideal to fly a specified transfer trajectory to the target planet, for instance a Hohmann trajectory, then we have to wait for a time T_{syn} before the planet can be reached along a similar transfer trajectory. Thus, the planetary phase angle Ψ_1 defines the launch time for a selected transfer trajectory within each synodic period. From Figure 18.5 we may derive the following relation:

$$\Psi_1 = \Psi_{1,0} - n_e(t - t_0) + n_t(t - t_0) \quad (18.43)$$

For $t - t_0 = T_{syn}$: $|\Psi_1 - \Psi_{1,0}| = 2\pi$. Substitution of these relations into (18.43) gives

$$|(n_e - n_t)| T_{syn} = 2\pi$$

or, with $n_e = 2\pi/T_e$ and $n_t = 2\pi/T_t$, where T_e and T_t are the periods of the Earth and the target planet in their orbits about the Sun,

$$\frac{1}{T_{syn}} = \left| \frac{1}{T_e} - \frac{1}{T_t} \right| \quad (18.44)$$

For the outer planets, $T_t \gg T_e$. So, for a flight to the outer planets we find from (18.44): $T_{syn} \approx T_e = 1$ year, which means that every year a spacecraft can be launched along a minimum-energy transfer trajectory to an outer planet. In Table 18.2, the synodic periods are listed for all target planets. Note that Mercury can be reached three times a year along a Hohmann transfer trajectory. A flight along a Hohmann transfer trajectory to Mars can only be executed once every 2.14 years. This result is clearly observable in the launch dates of actual missions to Mars; e.g. Mars Observer (September 1992), Mars Global Surveyor (November 1996), Mars 96 (November 1996), Mars Pathfinder (December 1996), Mars Climate Orbiter (December 1998), Mars Polar Lander (January 1999), Mars Odyssey (April 2001), Mars Express (June 2003), Mars Exploration rover Spirit (June 2003), Mars Exploration rover Opportunity (July 2003), Mars Reconnaissance Orbiter (August 2005), Phoenix (August 2007), Phobos Grunt (November 2011), Mars Science Laboratory rover Curiosity (November 2011), Mangalyaan (November 2013), MAVEN (November 2013).

In Section 18.4, the *ez*-diagram (Figure 18.4) was used to determine the optimum interplanetary transfer trajectory. We may also use it for the selection of feasible faster transfer trajectories. With an analysis similar to the one presented in Section 18.6, but not using anymore the condition $e = z - 1$, we can derive expressions for $V_1, V_{\infty e}, V_2, V_{\infty t}, \gamma_1, \gamma_2, \theta_1, \theta_2, t_f, \Psi_1, \Psi_2$ and D as a function of e and z for arbitrary two-dimensional transfer trajectories to a specific planet. These expressions can be used to draw in an *ez*-diagram curves of constant values of these parameters. By superimposing the various diagrams, it is possible to perform all kinds of quick analyses. As an example, Figure 18.8 shows curves of constant t_f and Ψ_1 for a two-dimensional flight to Venus ($n = 0.723$). Consider a hypothetical flight to Venus that starts on a date when $\Psi_1 = -10^\circ$ and with a flight time of 30 days. From Figure 18.8 we then obtain for the values of e and z of this transfer trajectory: $e = 0.508$, $z = 0.864$. With these values of e and z , we can read from the *ez*-diagrams with the contours for the other parameters the values of those parameters for the selected trajectory. Of course, these can also be computed by substituting the values of e and z into the relevant relations derived in Sections 18.4 and 18.6, and using a number of classical relations for Keplerian trajectories. We find for this example of transfer trajectory and an initial parking orbit

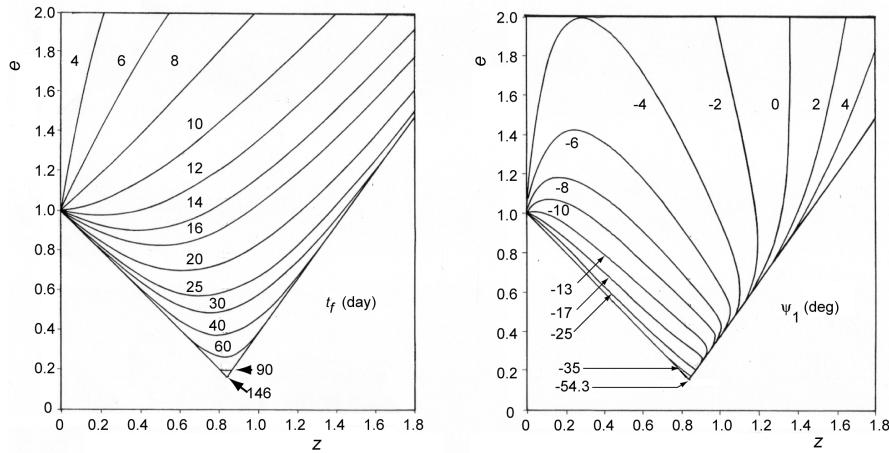


Figure 18.8: Curves of constant flight time, t_f , or constant planetary phase angle at departure, ψ_1 , in the ez -plane, for two-dimensional trajectories to Venus.

altitude of 200 km:

$$\begin{aligned} r_a &= 1.755 \text{ AU} ; \quad r_p = 0.573 \text{ AU} ; \quad D = 0.304 \text{ AU} \\ \theta_1 &= 254.5^\circ ; \quad \theta_2 = 292.5^\circ ; \quad \psi_2 = -8.5^\circ \\ V_1 &= 31.8 \text{ km/s} ; \quad \gamma_1 = -29.5^\circ ; \quad V_2 = 41.1 \text{ km/s} ; \quad \gamma_2 = -21.4^\circ \\ V_{\infty e} &= 15.8 \text{ km/s} ; \quad V_{\infty f} = 15.4 \text{ km/s} ; \quad \Delta V_0 = 11.5 \text{ km/s} \end{aligned}$$

With the present chemical propulsion systems this fast transfer trajectory is, however, beyond our capabilities.

Much more interesting, however, are diagrams in which contours of constant injection energy are plotted versus flight time, t_f , and planet phase angle at departure, ψ_1 . For this purpose, we write (18.18) as

$$\frac{\Delta V_0}{V_e} + \sqrt{c_1} = \sqrt{2c_1 + 3 - 2\sqrt{z} - \frac{1-e^2}{z}}$$

Squaring this equation leads, after some algebraic manipulation, to

$$e = \sqrt{1 - z \left[2c_1 + 3 - 2\sqrt{z} - \left(\frac{\Delta V_0}{V_e} + \sqrt{c_1} \right)^2 \right]} \quad (18.45)$$

For a specified initial parking orbit about the Earth and for a selected value of $\Delta V_0/V_e$, (18.45) can be used to compute the value of e for any value of z . As mentioned before, then the values of ψ_1 and t_f can be computed. In this way, it is possible to compute in a closed analytical way curves of constant $\Delta V_0/V_e$ and to draw them in a $\psi_1 t_f$ -diagram. Such a diagram is shown for a (hypothetical) two-dimensional flight to Mars in Figure 18.9 (top). This kind of contour plots are sometimes called ‘porkchop plots’. Also indicated in this Figure is a line for which $\Delta\theta = \theta_2 - \theta_1 = 180^\circ$. Trajectories with $\Delta\theta < 180^\circ$ are called *Type 1 transfer trajectories*. The dashed part of the curves, for which $\Delta\theta > 180^\circ$, represents *Type 2 transfer trajectories*. This kind of plots is very useful for determining the minimum ΔV_0 for a flight to a planet at a specified launch time (specified ψ_1), and for determining the *launch window*, as will be shown in Section 18.8.

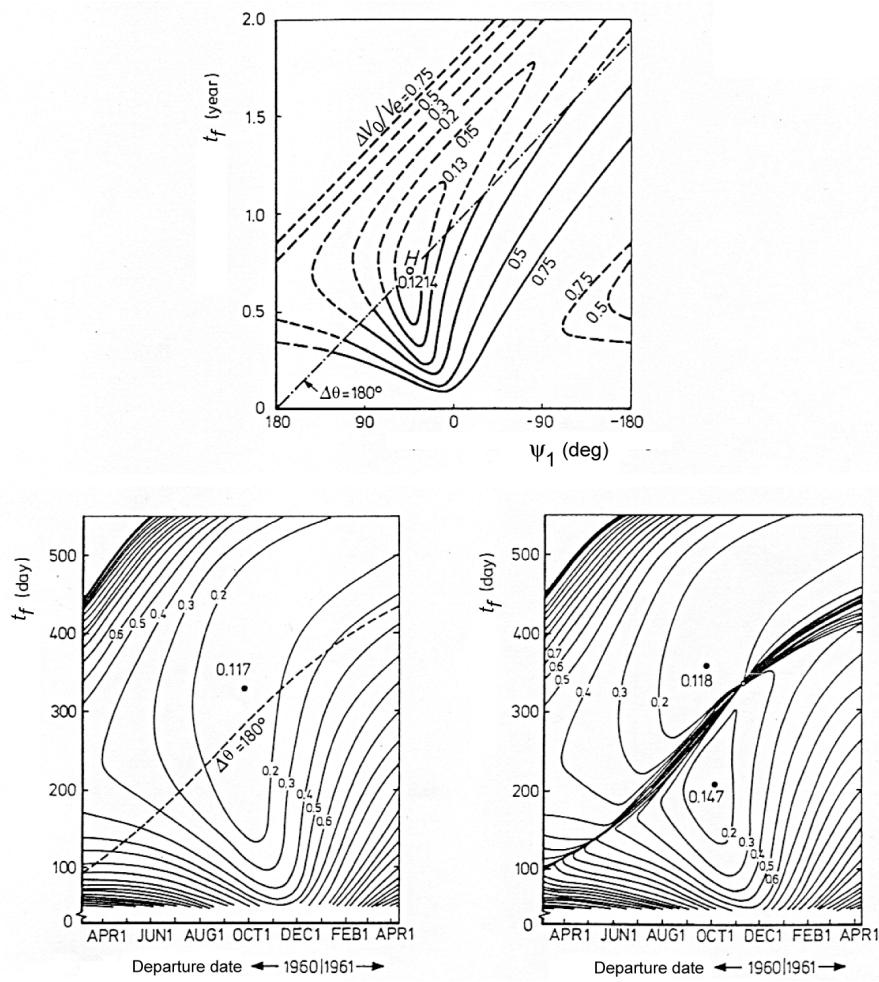


Figure 18.9: (top) Flight time, t_f , versus planetary phase angle at departure, ψ_1 , for two-dimensional constant ΔV_0 trajectories to Mars. Assumptions: orbits of the planets are circular; 185 km altitude initial parking orbit. (bottom) Contours of constant departure hyperbolic excess velocity, V_∞ , for missions to Mars; (bottom left) coplanar elliptical planetary orbits, (bottom right) mutually inclined elliptical planetary orbits; values at the curves refer to the ratio V_∞/V_e .

18.8. Three-dimensional direct-transfer trajectories

Until now, all analyses were based on the assumption that the planetary orbits are circular, that their orbital planes coincide with the plane of the Earth's orbit (ecliptic), and that the transfer trajectory lies in the ecliptic. However, in reality the planetary orbits are elliptical and mutually inclined, and transfer trajectories are three-dimensional. The computation of these three-dimensional transfer trajectories is quite complicated. An additional problem is that the relative configuration of the Earth and the target planet will not repeat every synodic period, due to the eccentricity and inclination of the planetary orbits. Consequently, although during each synodic period there exists a minimum-energy transfer trajectory, the injection energy required for that trajectory differs for each synodic period. In the following, a general outline will be given of the basic scheme to compute three-dimensional interplanetary trajectories.

We assume that the transfer trajectory is an elliptical heliocentric orbit from the center of the Earth to the center of the target planet, but a similar methodology may be applied for parabolic

and hyperbolic transfer trajectories. It is supposed that the basic variables in the computation process are the time of departure from the Earth, t_1 , and the flight time, t_f . Then, the time of arrival at the target planet, t_2 , is known. At t_1 , the coordinates of the Earth relative to a non-rotating heliocentric ecliptic reference frame can be computed approximately by using the Earth's mean orbital elements, or can be extracted from planetary ephemerides computer files. The heliocentric ecliptic coordinates of the target planet at t_2 can be found in the same way.

The heliocentric transfer trajectory of the spacecraft lies, excluding *mid-course plane changes*, in the plane through the Sun, the Earth at time of departure and the target planet at time of arrival. The heliocentric transfer angle, $\Delta\theta$, traversed by the spacecraft is given by

$$\cos\Delta\theta = \frac{\bar{r}_e \cdot \bar{r}_t}{r_e r_t} \quad (18.46)$$

where \bar{r}_e and \bar{r}_t are the heliocentric position vectors of the Earth at departure and of the target planet on arrival. The value of $\Delta\theta$ may be determined uniquely by realizing that $\sin\Delta\theta$ must have the same sign as $(\bar{r}_e \times \bar{r}_t) \cdot \bar{e}_z$, where \bar{e}_z is the unit vector along the Z-axis; i.e. pointing to the ecliptic north pole. The chord, c , joining Earth at departure and target planet on arrival is found from

$$c^2 = r_t^2 + r_e^2 - 2r_t r_e \cos\Delta\theta \quad (18.47)$$

Now that the radii r_e and r_t , and the chord c are known, *Lambert's equation* (Section 6.7) can be solved iteratively for the specified flight time to yield the semi-major axis, a , of the transfer trajectory. Of course, the four cases of Lambert's equation should be investigated separately. Then, the heliocentric velocities V_1 and V_2 may be computed from

$$V_1 = \sqrt{\mu_s \left(\frac{2}{r_e} - \frac{1}{a} \right)} \quad ; \quad V_2 = \sqrt{\mu_s \left(\frac{2}{r_t} - \frac{1}{a} \right)} \quad (18.48)$$

For the computation of the flight path angles γ_1 and γ_2 we need two well-known geometric characteristics of conic sections. In Figure 18.10 the geometry of an elliptical interplanetary trajectory is shown for one of the four cases of Lambert's equation. In addition to the usual triangle through the attraction center F_1 (Sun), a second triangle through the empty focal point F_2 is indicated. The first characteristic of an ellipse that will be used is that the sum of the distances of a point on the ellipse to the two focal points equals $2a$. In fact, this characteristic is one way of defining an ellipse. This relation allows the determination of the lengths of the sides in the second triangle. The second characteristic that will be used is that the normal at a point on

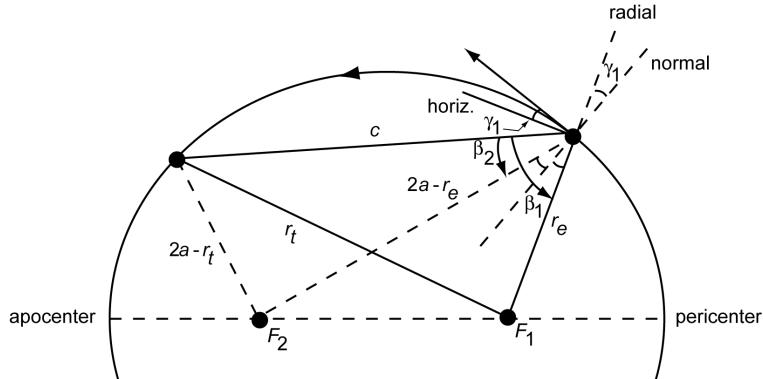


Figure 18.10: Geometry for the determination of the flight path angle γ_1 .

the ellipse bisects the angle between the lines to both focal points. So, for γ_1 we may write

$$\gamma_1 = \frac{1}{2}(\beta_1 - \beta_2) \quad (18.49)$$

Application of the first characteristic and the cosine rule gives

$$\cos\beta_1 = \frac{r_e^2 - r_t^2 + c^2}{2cr_e} ; \quad \cos\beta_2 = \frac{(2a - r_e)^2 - (2a - r_t)^2 + c^2}{2c(2a - r_e)} \quad (18.50)$$

As r_e , r_t , a and c are known, solution of (18.50) and subsequent substitution into (18.49) yields γ_1 . It is emphasized that for each of the four cases of Lambert's equation the quadrants of the angles β_1 and β_2 should be determined and that the functional relation (18.49) is slightly different for the other three cases. A similar analysis yields the value of γ_2 .

Once the flight path angles are known, we proceed by computing the semi-latus rectum of the interplanetary trajectory:

$$p = \frac{(r_e V_1 \cos\gamma_1)^2}{\mu_s} \quad (18.51)$$

The eccentricity can be obtained from

$$e = \sqrt{1 - \frac{p}{a}} \quad (18.52)$$

The heliocentric true anomaly at the time of departure from the Earth follows from

$$\theta_1 = \arccos \left\{ \frac{1}{e} \left(\frac{p}{r_e} - 1 \right) \right\} \quad (18.53)$$

The value of θ_1 is uniquely determined, as for $0^\circ \leq \theta_1 \leq 180^\circ$: $\gamma_1 \geq 0^\circ$, and for $180^\circ \leq \theta \leq 360^\circ$: $\gamma_1 \leq 0^\circ$. The heliocentric true anomaly on arrival is found from

$$\theta_2 = \theta_1 + \Delta\theta \quad (18.54)$$

where $\Delta\theta$ is given by (18.46). The inclination of the trajectory with respect to the ecliptic is found from

$$\cos i = \bar{e}_w \cdot \bar{e}_z \quad (18.55)$$

where \bar{e}_w is the unit vector normal to the orbital plane:

$$\bar{e}_w = \frac{\bar{r}_e \times \bar{r}_t}{r_e r_t \sin\Delta\theta} \quad (18.56)$$

Now that the inclination of the orbital plane is known, we may apply spherical geometry to determine the three-dimensional directions of \bar{V}_1 and \bar{V}_2 , of which the magnitudes are given by (18.48). The hyperbolic excess velocities $\bar{V}_{\infty e}$ and $\bar{V}_{\infty t}$ are given by

$$\bar{V}_{\infty e} = \bar{V}_1 - \bar{V}_e ; \quad \bar{V}_{\infty t} = \bar{V}_2 - \bar{V}_t \quad (18.57)$$

where \bar{V}_e is the Earth's heliocentric velocity at departure and \bar{V}_t the target planet's heliocentric velocity on arrival.

A very instructive discussion on the effects of the eccentricity and inclination of the planetary orbits is given by J.V. Breakwell (1917-1991) et al.⁵, from which Figure 18.9 (bottom) is taken. In that publication a flyby mission to Mars is considered and contours of constant $V_{\infty e}/V_e$ are plotted as a function of launch date in 1960/1961 and flight time. For the circular coplanar case, curves similar to Figure 18.9 (top) were obtained, where the optimum transfer is a Hohmann transfer. Introducing the planetary orbital eccentricities, the minimum departure excess velocity trajectory for that period now involves a transfer angle $\Delta\theta = 210^\circ$ and a longer flight time (Figure 18.9, bottom left). After introducing also the planetary orbital inclination, we see in Figure 18.9 (bottom right) that the original Hohmann trajectory has bifurcated into two local minimum-energy trajectories with two different flight times. The two groups of closed contours represent the Type 1 and Type 2 trajectories already mentioned in Section 18.7. Note that two trajectories usually exist for a given type, $V_{\infty e}$ and launch date. Trajectories corresponding to the lower parts of each contour are designated as *Class 1 trajectories* and yield shorter flight times and smaller transfer angles; trajectories corresponding to the upper part of the contours are designated as *Class 2 trajectories*. For a given launch date and $V_{\infty e}$, the Type 1 Class 1 trajectory has the shortest flight time and the Type 2 Class 2 trajectory has the longest flight time.

If the heliocentric transfer angle is about 180° , the plane determined by Sun, Earth and target planet will, as a result of the planetary inclination, generally, be highly inclined to the ecliptic. This implies excessively high values of ΔV_0 , and for orbiter missions also of ΔV_3 . The required high values of ΔV_0 are reflected by the very narrow spacing of the curves in Figure 18.9 (bottom right) near the bifurcation region. These problems may be circumvented by applying mid-course plane changes. One possibility is to depart in the plane of the Earth's motion and to apply a mid-course impulsive shot at the intersection of the orbital planes of the Earth and the target planet. This would rotate the transfer orbital plane through the small angle between the planes of the spacecraft trajectory and the target planet's orbit. Such a maneuver requires only a modest propulsive capability during the mid-course phase of the trip. However, it is generally even more advantageous to perform the mid-course plane change at a point somewhat removed from the line of intersection of the two orbital planes, and also to depart from the Earth and arrive at the target planet in planes which are slightly inclined to the planetary motions (Figure 18.11). An optimization procedure is needed to determine that position of the mid-course maneuver and those values of the inclination of both branches of the transfer orbit that yield the minimum overall mission

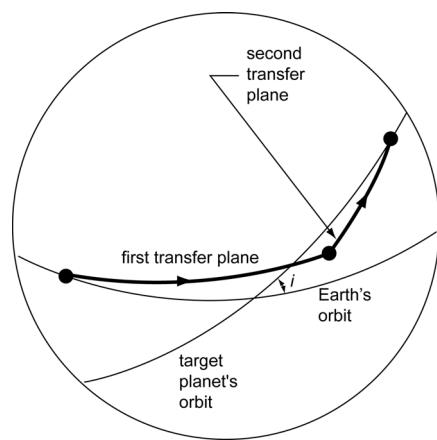


Figure 18.11: Mid-course maneuver leading to a broken-plane transfer trajectory.

⁵ J.V. Breakwell, R.W. Gillespie, S. Ross, *Researches in interplanetary transfer*, A.R.S. Journal, Vol. 31, p. 201-208, 1961.

propellant requirement. It is emphasized that in practice these near- 180° transfer angle trajectories are avoided for a number of operational aspects, and that mid-course maneuvers are mainly executed to eliminate trajectory errors and to guarantee the planetary encounter to be as accurate as possible.

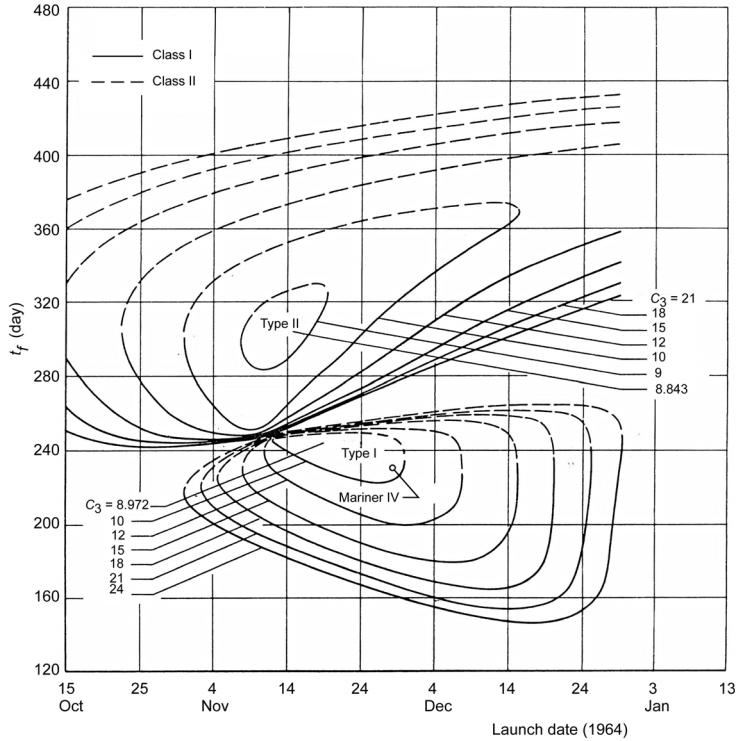


Figure 18.12: Contours of constant injection energy, C_3 , plotted versus launch date in 1964 and flight time, t_f ; numbers at the curves refer to the value of C_3 in km^2/s^2 .

Figure 18.12 shows, for the Mariner 4 flight to Mars in 1964, a diagram in which the flight time is plotted versus the time of launch from Earth and in which contours of constant injection energy $C_3 (= V_{\infty e}^2)$ are drawn. We recognize the two sets of contours discussed before; each set leading to a minimum-energy trajectory:

Minimum for Type 2 trajectories

Launch date: November 14, 1964, $C_3 = 8.843 \text{ km}^2/\text{s}^2$, $V_{\infty e} = 2.974 \text{ km/s}$, $t_f = 305 \text{ day}$

Minimum for Type 1 trajectories

Launch date: November 18, 1964, $C_3 = 8.972 \text{ km}^2/\text{s}^2$, $V_{\infty e} = 2.995 \text{ km/s}$, $t_f = 245 \text{ day}$

When these numbers are compared to those for a two-dimensional Hohmann transfer trajectory between coplanar circular planetary orbits: $V_{\infty e} = 2.94 \text{ km/s}$, $t_f = 259 \text{ days}$, we conclude that the required injection energy compares reasonably well and that also the flight time does not differ very much. Figure 18.12 shows that a Type 1 trajectory was selected for Mariner 4; the injection energy of this trajectory ($V_{\infty e} = 3.11 \text{ km/s}$) differs very little from the minimum injection energy for Type 1 trajectories in this period. For both types of trajectories the local minimum-energy trajectory at a specific launch date can be determined by adding vertical tangents to the injection energy contours in Figure 18.12. The minimum injection energy thus obtained can be plotted versus the launch date and compared to the maximum injection energy that can be provided by a particular launch vehicle to the spacecraft. In that way, the so-called *launch window*, during which injection of that spacecraft by that rocket into an interplanetary transfer trajectory can take place, can be determined. The selection of the nominal transfer trajectory for a particular mission

is a complex problem, in which a trade off has to be made between the required injection energy, i.e. the launcher payload capability, and operational constraints, such as communication distance upon arrival, geometric configuration of Earth and target planet upon arrival, direction and magnitude of planetary approach velocity, etc. The large majority of planetary spacecraft have flown Type 1 trajectories. One of the few exceptions was the Phoenix spacecraft that traveled along a Type II trajectory to Mars.

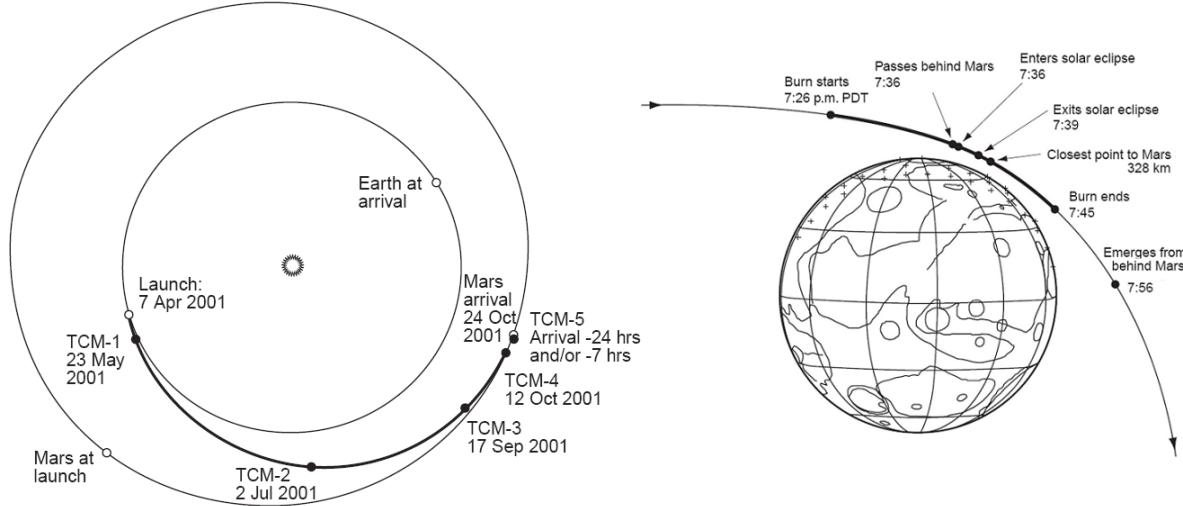


Figure 18.13: Interplanetary trajectory of the NASA Mars Odyssey spacecraft (left), and the spacecraft's approach of Mars and deceleration maneuver to enter an orbit about Mars (right).

In Figure 18.13 (left) the trajectory flown by the Mars Odyssey spacecraft is plotted relative to a non-rotating heliocentric reference frame. The spacecraft was launched on April 7, 2001, by a Delta II rocket, and arrived at Mars, after a flight of 200 days along a Type 1 trajectory, on October 24, 2001. During the flight, five trajectory correction maneuvers (TCM) were executed. Figure 18.13 (right) shows the spacecraft's approach trajectory to Mars and the deceleration maneuver to enter an orbit about Mars. Note that the deceleration burn lasted for 19 min and that the burn ended 11 min before the spacecraft emerged from behind Mars when viewed from Earth.

Figure 18.14 shows the mission characteristic velocity, V_{char}^* , for flights to Venus and Mars via minimum-energy transfer trajectories for the period 1975-1990. For each launch opportunity, the minimum characteristic velocity and the characteristic velocity for a 30-days launch window are indicated. For comparison, the mission characteristic velocity for a two-dimensional Hohmann

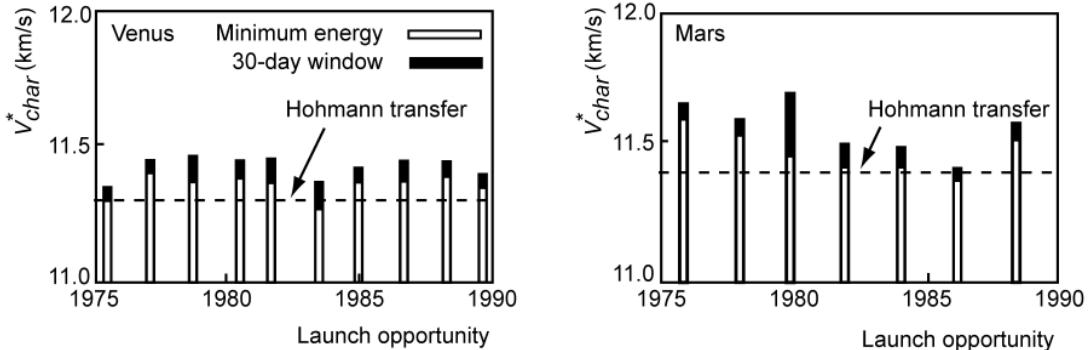


Figure 18.14: Yearly variation of the mission characteristic velocity for minimum-energy flights to Venus and Mars.

transfer trajectory is also indicated, for the case that Earth, Venus and Mars would move in circular coplanar orbits about the Sun. The Figure shows that for flyby missions to Mars a launch in the years 1982, 1984 and 1986 was very attractive. However, in these years no US spacecraft were sent to Mars.

18.9. Gravity losses

Until now, it has been assumed that thrusting periods during an interplanetary mission may be considered as impulsive shots (Section 1.7). In reality, the rocket engines will have a non-zero thrusting time and consequently gravity losses (Section 1.7) will occur, unless the thrust vector is continuously oriented perpendicular to the position vector. Of course, the largest gravity losses will occur during injection into the interplanetary trajectory and during the deceleration maneuver after arrival at the target planet. In Section 12.7 it was shown how the information provided by ESA CR(P) 1910 can be used to estimate the magnitude of the gravity loss for an optimum transfer from a circular orbit to an elliptical orbit with perigee distance equal to the radius of the initial circular orbit. There, it was assumed that the magnitude of the thrust and the (effective) exhaust velocity are constant, and that the optimum history of the thrust angle, δ , is applied. The methodology presented in that ESA report may also be applied to estimate the gravity loss for a transfer from a circular parking orbit about the Earth to a specified hyperbolic orbit (escape from Earth), and for a transfer from a hyperbolic orbit about a target planet to an elliptical or circular orbit about that target planet. In this Section, for both cases a numerical example will be presented.

Escape from a parking orbit about the Earth

In Section 12.7, a number of dimensionless parameters were introduced, which read in the notation applied in this Chapter:

$$\begin{aligned} \mathcal{E}_{0,1}^* &= \frac{\mathcal{E}_{0,1}}{V_{c_0}^2} \quad ; \quad \mathcal{E}_{0,2}^* = \frac{\mathcal{E}_{0,2}}{V_{c_0}^2} \quad ; \quad \Delta\mathcal{E}^* = \mathcal{E}_{0,2}^* - \mathcal{E}_{0,1}^* \\ k_2 &= \frac{F}{g_0 M_0} = \frac{r_1^2}{\mu} \frac{F}{M_0} \quad ; \quad k_3 = \frac{V_j}{V_{c_0}} \quad ; \quad t^* = \frac{t}{\sqrt{r_0^3/\mu}} = \frac{2\pi t}{T_0} \end{aligned} \quad (18.58)$$

where \mathcal{E} is the orbital energy, \mathcal{E}^* is the normalized orbital energy, both per unit of mass, r_0 is the radius of the initial circular parking orbit, V_{c_0} is the circular velocity in this parking orbit, F is the thrust, g_0 is the acceleration due to gravity in the parking orbit, M_0 is the initial mass of the spacecraft, μ is the gravitational parameter of the Earth, V_j is the (effective) exhaust velocity of the rocket engine, and T_0 is the period of the parking orbit. The notation 0,1 and 0,2 indicate the conditions just before and just after the thrusting period in the parking orbit, respectively. The percentage gravity loss is defined as

$$\Delta\Delta V = \left(\frac{\Delta V - (\Delta V)_i}{(\Delta V)_i} \right) * 100\% \quad (18.59)$$

where ΔV is the actual impulse transferred to the spacecraft during the finite thrusting period and $(\Delta V)_i$ is the impulse in case the maneuver would have been executed by an impulsive shot. In the ESA report, many plots are presented for the value of $\Delta\Delta V$ as a function of $\Delta\mathcal{E}^*$, for specified values of $\mathcal{E}_{0,1}^*$, k_2 and k_3 . Because the initial parking orbit is circular, the following relations hold

according to Section 12.7:

$$\mathcal{E}_{0,1}^* = -\frac{1}{2} \quad ; \quad \mathcal{E}_{0,2}^* = -\frac{1}{2}(1-e) \quad ; \quad \Delta\mathcal{E}^* = \frac{1}{2}e \quad (18.60)$$

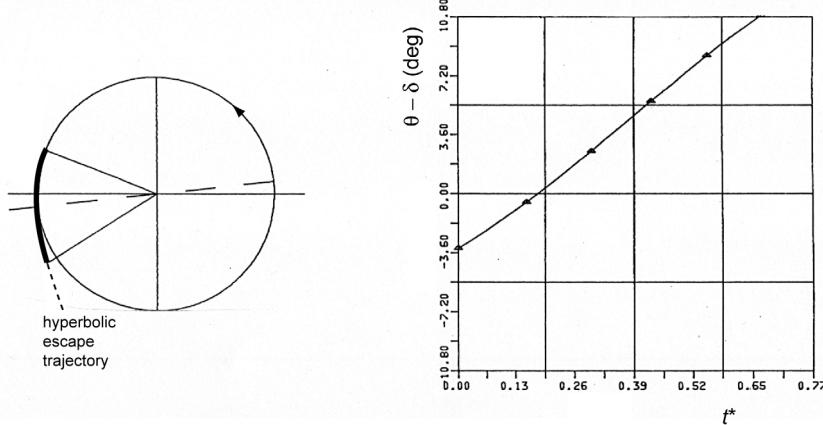


Figure 18.15: Powered part of a hyperbolic escape trajectory (left) for a mission to Mars and the optimum variation of the angle $\theta - \delta$ (right) during the powered flight for $\mathcal{E}_{0,1}^* = -0.5$, $\mathcal{E}_{0,2}^* = 0.5$, $k_2 = 0.5$, $k_3 = 0.5$.

Figure 18.15 shows, as an example, the history of the optimum variation of the angle $\theta - \delta$, where θ is the true anomaly and δ is the thrust angle, for a thrusting period of $t \approx 0.1034 T_0$ ($t^* = 0.65$) and $\mathcal{E}_{0,1}^* = -0.5$, $\mathcal{E}_{0,2}^* = 0.5$, $e = 2$ ($\Delta\mathcal{E}^* = 1.0$), $F/M_0 = 0.5g_0$ ($k_2 = 0.5$), $V_j = 0.5V_{c0}$ ($k_3 = 0.5$). So, this case concerns a transfer from an initial circular orbit to a hyperbolic orbit with $e = 2$. According to (18.4) this value of e corresponds to $V_0 = \sqrt{3}V_{c0}$, which leads for a hyperbolic trajectory perigee altitude of 300 km to $V_0 = 13.38$ km/s and, according to (18.5), to $V_{\infty e} = 7.73$ km/s. Inspection of Table 12.1 and Figure 12.19 (top) shows that for this escape flight: $\Delta\Delta V \approx 2\%$.

Insertion into an orbit about the target planet

As an example of an orbit insertion maneuver at a target planet, we consider a transfer from a hyperbolic trajectory with $\mathcal{E}_{3,1}^* = 0.567$ about Mars to an elliptical orbit with $\mathcal{E}_{3,2}^* = -0.179$ about Mars. Here, the index 3 refers to the conditions at pericenter of the elliptical orbit, and the indices 3,1 and 3,2 refer to the conditions just before and just after the thrusting period. For this case, we can derive the following expressions:

$$\mathcal{E}_{3,1}^* = \frac{1}{2}(e_1 - 1) \quad ; \quad \mathcal{E}_{3,2}^* = -\frac{1}{2}(1 - e_2) \quad ; \quad \Delta\mathcal{E}^* = \frac{1}{2}(e_2 - e_1) \quad (18.61)$$

where e_1 and e_2 are the eccentricities of the initial hyperbolic orbit and of the final elliptical orbit, respectively. From these relations, and the specified values of $\mathcal{E}_{3,1}^*$ and $\mathcal{E}_{3,2}^*$, we obtain $e_1 = 2.134$, $e_2 = 0.642$. The value of e_2 yields a value of 4.59 for the ratio between the apocenter and pericenter distances of the final elliptical orbit. If we assume a pericenter altitude of 350 km, then we obtain from (6.19), (18.5), (6.23) and (8.11): $T_3 = 1.934$ hr, $V_{\infty M} = 3.60$ km/s, $V_{3,1} = 5.98$ km/s, $V_{3,2} = 4.33$ km/s. So, the rocket engine thrusting period has resulted in a velocity decrease of $\Delta V_3 = 1.65$ km/s. Figure 18.16 shows the history of the optimum variation of the angle $\theta - \delta$ for two values of the thrust: $F/M_0 = 2.0373g_3$ ($k_2 = 2.0373$) and $F/M_0 = 0.20373g_3$ ($k_2 = 0.20373$); for both cases: $V_j = 0.9608V_{c3}$ ($k_3 = 0.9608$). The engine burning time is for the higher thrust level: $t \approx 0.03 T_3$ ($t^* = 0.188$), and for the lower thrust level: $t \approx 0.318 T_3$ ($t^* = 2.0$). When we use from the ESA report the appropriate plot similar to Figure 12.19, we find for the gravity loss at the

higher thrust level: $\Delta\Delta V \approx 0.15\%$ and at the lower thrust level: $\Delta\Delta V \approx 9.2\%$. We conclude that for the lower thrust level and corresponding extended burning time, the percentage gravity loss is quite significant.

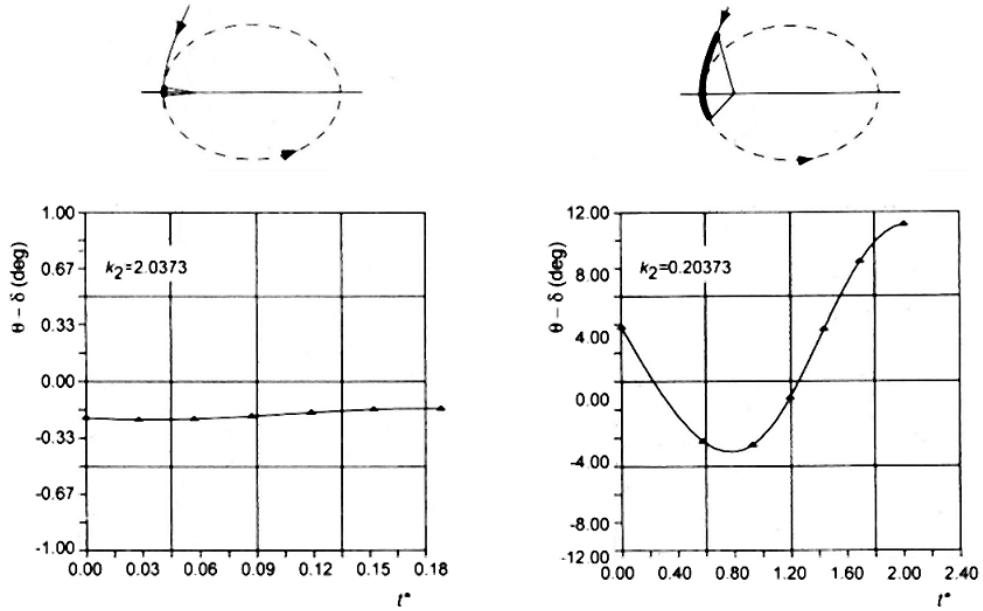


Figure 18.16: Transfer from a hyperbolic approach trajectory to an elliptical orbit about Mars and the optimum variation of the angle $\theta - \delta$ during the powered flight for $\mathcal{E}_{3,1}^* = 0.567$, $\mathcal{E}_{3,2}^* = -0.179$, $k_3 = 0.9608$, and $k_2 = 2.0373$ (left) or $k_2 = 0.20373$ (right).

18.10. Approach and flyby of a planet

During the approach of a planet the influence of the gravity field of that planet on the heliocentric trajectory of the spacecraft will gradually increase. Ultimately, the gravitational attraction of the planet will become the dominating force acting on the spacecraft. In the patched-conic approach, we assume that when the spacecraft enters the sphere of influence of the planet appropriate coordinate and velocity transformations are applied and from then on the hyperbolic trajectory about the planet is described relative to a planetocentric non-rotating reference frame. Because the radius of the sphere of influence is small relative to the distance of the planet from the Sun (Table 4.3), in first approximation it may be assumed that, if the planet would not influence the trajectory of the spacecraft, the heliocentric velocity of the spacecraft would hardly change in magnitude and direction over a distance corresponding to the diameter of the sphere of influence. Therefore, we may assume that the velocity and flight path angle at the moment of entrance into the sphere of influence are given by V_2 and γ_2 , independent of the location of the point of entrance. The location of the point of entrance into the sphere of influence can be influenced by adapting the launch time and the selected interplanetary trajectory, and by applying mid-course corrections. Figure 18.17 shows, for the case that the planes of the spacecraft's trajectory and the planet's orbit coincide, the effect of the location of the point of entrance on the spacecraft's trajectory about the planet. In this Figure, the intersection of the heliocentric trajectory with two instantaneous positions of the (moving) sphere of influence is sketched. Depending on the position of the spacecraft along its heliocentric trajectory, it will enter the sphere of influence at t_0 or t_1 and will fly for both cases entirely different trajectories about the planet. In the case that the spacecraft enters the sphere of influence at t_0 it will move in a counter-clockwise direction about the planet; in the case that it enters the sphere of influence at t_1 it will move in a clockwise direction.

direction about the planet. By applying small mid-course trajectory corrections, the time of entrance into the sphere of influence is tuned such that the hyperbolic trajectory about the planet precisely satisfies the mission requirements.

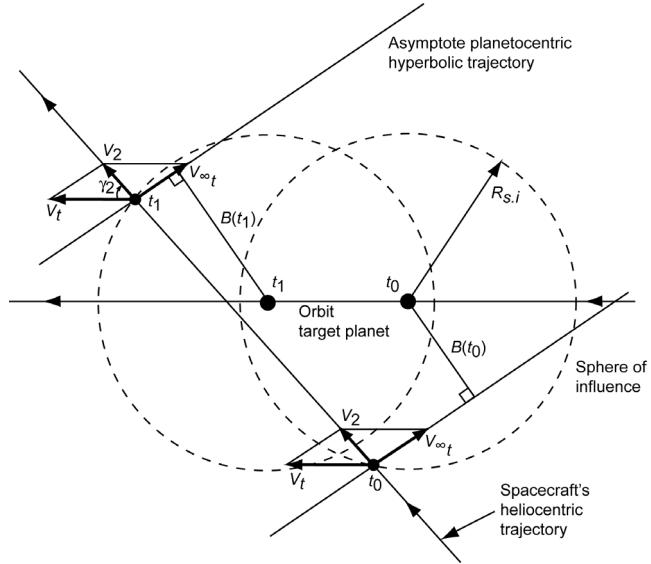


Figure 18.17: Sketch of the heliocentric trajectory of a spacecraft when it passes close to a planet, for two relative positions of that planet and for the case that the spacecraft and the planet move in the same plane.

For a specified time of launch, interplanetary trajectory and position of the spacecraft along its trajectory, the time when and location where the spacecraft enters the planet's sphere of influence and the spacecraft's hyperbolic excess velocity with respect to the target planet, $\bar{V}_{\infty,t}$, are fixed. These parameters determine the hyperbolic trajectory of the spacecraft within the sphere of influence. An important parameter of that trajectory is the *impact parameter*, \bar{B} . To define this impact parameter, we first introduce a non-rotating planetocentric coordinate frame *RST* (Figure 18.18). The *S*-axis is parallel to the incoming asymptote of the approach hyperbola; its positive direction is the direction of $\bar{V}_{\infty,t}$. The *T*-axis is parallel to the ecliptic plane; its positive direction is defined by

$$\bar{e}_T = \frac{\bar{e}_S \times \bar{e}_N}{|\bar{e}_S \times \bar{e}_N|}$$

where \bar{e}_S and \bar{e}_T are unit vectors along the *S*- and *T*-axis, and \bar{e}_N is the unit vector pointing towards the ecliptic north pole. The *R*-axis completes the right-handed orthogonal coordinate frame. So, the *RT*-plane is normal to the incoming asymptote. The vector \bar{B} , which specifies the *aiming point*, is a vector from the center of the planet to the point where the incoming asymptote pierces the *RT*-plane. Usually, the aiming point is defined by its two components $\bar{B} \cdot \bar{e}_T$ and $\bar{B} \cdot \bar{e}_R$. Spacecraft engineering requirements and scientific experiment considerations, or the swingby mission (Section 18.11) requirements, impose numerous constraints on the selection of the aiming point. Once the aiming point is selected, it is the goal of mid-course maneuvers to nullify any error in the trajectory such that the incoming asymptote pierces the *RT*-plane as close to the aiming point as possible. For this, an accurate tracking of the spacecraft during its interplanetary flight is indispensable.

The plane of the planetocentric hyperbola is determined by \bar{B} and $\bar{V}_{\infty,t}$. This plane is oriented perpendicular to the *RT*-plane. We now consider the motion of the spacecraft in its orbital plane

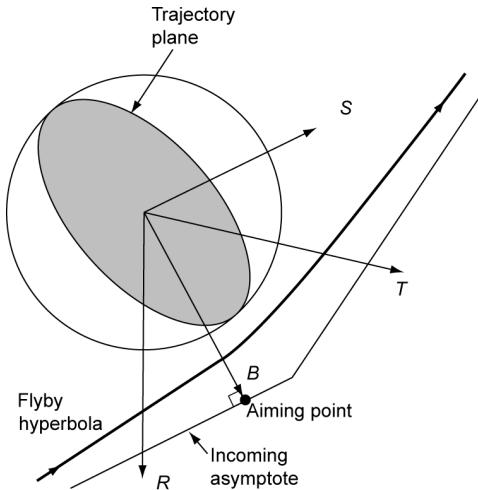


Figure 18.18: Encounter geometry and aiming point reference frame.

for the case that the spacecraft is passing ‘behind’ the swingby planet; this motion is illustrated in Figure 18.19. For a passage ‘ahead of’ the swingby planet a similar methodology can be developed. The velocity \vec{V}'_t indicated in this Figure is the projection of the target planet’s heliocentric velocity vector, \vec{V}_t , on this plane. It is important to realize that $|\vec{V}_{\infty t}^*| = |\vec{V}_{\infty t}|$. The state vector of the spacecraft when it enters the sphere of influence is determined by four scalar parameters: the position of the spacecraft by $R_{s,i}$ and δ , and the velocity of the spacecraft by $V_{\infty t}$ and φ_2 . For a specified interplanetary trajectory, δ , $V_{\infty t}$ and φ_2 are known; these parameters determine the hyperbolic trajectory flown by the spacecraft about the planet. In practice, the application of the three variables B , $V_{\infty t}$, β instead of the set δ , $V_{\infty t}$, φ_2 is preferable. Of course, we then first have to prove that there exist unique relations between both sets of variables.

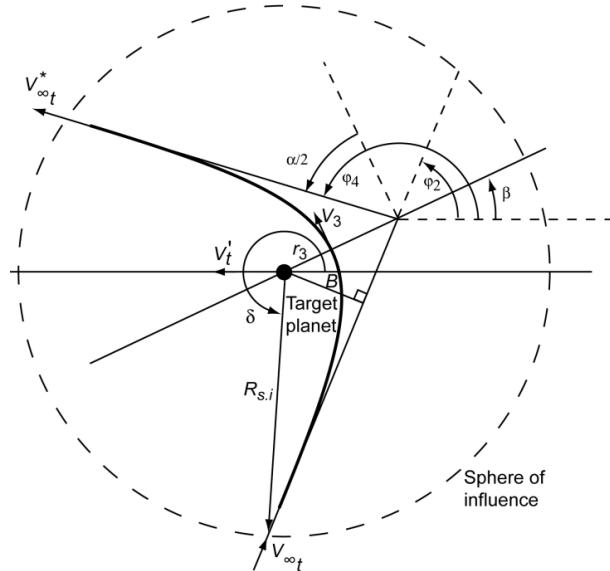


Figure 18.19: The in-plane geometry of hyperbolic encounter trajectories.

From Figure 18.19 we obtain

$$B = R_{s,i} \cos\left(2\pi - \delta - \left(\frac{1}{2}\pi - \varphi_2\right)\right) = R_{s,i} \sin(\varphi_2 - \delta) \quad (18.62)$$

Because the angular momentum of motion is constant for the entire trajectory about the planet, we can write

$$B V_{\infty_t} = r_3 V_3 \quad (18.63)$$

where the index 3 denotes the conditions at pericenter. With (18.15) we write

$$V_3^2 = V_{esc_3}^2 + V_{\infty_t}^2 = \frac{2\mu}{r_3} + V_{\infty_t}^2$$

where μ is the gravitational parameter of the planet. Substitution of this expression into (18.63) gives

$$B V_{\infty_t} = r_3 \sqrt{\frac{2\mu}{r_3} + V_{\infty_t}^2}$$

Squaring this relation yields for the pericenter distance of the hyperbolic trajectory about the planet

$$r_3 = \frac{\mu}{V_{\infty_t}^2} \left[\sqrt{1 + \frac{B^2 V_{\infty_t}^4}{\mu^2}} - 1 \right] \quad (18.64)$$

Substitution of (18.64) into (18.63) gives for the velocity at pericenter

$$V_3 = \frac{\mu}{B V_{\infty_t}} \left[\sqrt{1 + \frac{B^2 V_{\infty_t}^4}{\mu^2}} + 1 \right] \quad (18.65)$$

In Chapter 8 we have found the following relations for hyperbolic trajectories:

$$\cos \frac{1}{2}(\pi + \alpha) = -\frac{1}{e} \quad ; \quad r_3 = a(1 - e) \quad ; \quad V_{\infty_t}^2 = -\frac{\mu}{a} \quad (18.66)$$

From these relations we find for the *asymptotic deflection angle*, α (Figure 18.19),

$$\sin \frac{1}{2}\alpha = \frac{1}{1 + r_3 V_{\infty_t}^2 / \mu} \quad ; \quad 0 < \frac{1}{2}\alpha < \frac{1}{2}\pi \quad (18.67)$$

Substitution of (18.64) into this relation gives

$$\sin \frac{1}{2}\alpha = \frac{1}{\sqrt{1 + B^2 V_{\infty_t}^4 / \mu^2}} \quad (18.68)$$

For the angle β in Figure 18.19 we can write

$$\beta = \varphi_2 + \frac{1}{2}\alpha - \frac{1}{2}\pi$$

or, with (18.68),

$$\beta = \varphi_2 + \arcsin \left\{ \frac{1}{\sqrt{1 + B^2 V_{\infty_t}^4 / \mu^2}} \right\} - \frac{1}{2}\pi \quad (18.69)$$

Expressions (18.62) and (18.69) constitute unique relations between the set of variables δ , $V_{\infty t}$ and φ_2 and the set of variables B , $V_{\infty t}$, β . For the discussions below, the set of variables B , $V_{\infty t}$, β will be used; the values of these parameters follow, just as the values of the variables δ , $V_{\infty t}$ and φ_2 , from the interplanetary trajectory flown by the spacecraft. The parameters B and β can be adapted by applying mid-course maneuvers, which is crucial for the execution of swingby flights; the parameter $V_{\infty t}$ is much less sensitive to mid-course maneuvers. It follows from (18.64), (18.65) and (18.68) that, for specified values of B and $V_{\infty t}$, a larger planet (larger value of μ) yields a smaller value of r_3 , a larger value of V_3 , and a larger value of α . The values of r_3 , V_3 and α are independent of the value of β . It also follows from these expressions that at increasing values of B the value of r_3 increases, the value of V_3 decreases and the value of α decreases.

When the pericenter distance, r_3 , is smaller than the radius of the planet, R , then the spacecraft will impact somewhere on the planet. So, according to (18.64), the condition for impact is

$$\sqrt{1 + \frac{B^2 V_{\infty t}^4}{\mu^2}} < \frac{V_{\infty t}^2 R}{\mu} + 1$$

or

$$\left(\frac{B}{R}\right)^2 < 1 + \frac{2\mu}{R V_{\infty t}^2} = 1 + \frac{V_{esc, surf}^2}{V_{\infty t}^2} \quad (18.70)$$

As an example, we assume, just as we did in Section 18.5, that the spacecraft approaches Mars with a velocity $V_{\infty t}$ that is 10% higher than the value of $V_{\infty t}$ corresponding to a Hohmann transfer trajectory (Table 18.2). We then find that the spacecraft impacts on Mars if $B < 2.0 R$. So, for an impact on Mars it is not necessary that the incoming asymptote of the spacecraft's planetocentric trajectory intersects the surface of Mars; it is sufficient that this asymptote intersects a sphere about Mars with a radius of $2R$. Obviously, the Martian gravity field has a 'focusing effect' on the trajectory. This radius is sometimes called the *capture radius*. Of course, if one wants to execute a landing on the planet, then the spacecraft has to be decelerated by some means such that the velocity just above the surface is reduced to a very small value. In practice, one will require that the spacecraft lands in a certain region, which poses, for a given value of $V_{\infty t}$, narrow constraints on the admissible value of B . In addition, if we wish to take advantage of atmospheric braking, then the aiming point B should lie within a thin annulus about the planet and the atmosphere should be entered under a predetermined flight path angle (*reentry corridor*).

Figure 18.20 shows the minimum distance, r_3 , and the asymptotic deflection angle, α , as a function of B/R and a series of realistic values of $V_{\infty t}$ for a flyby of Jupiter. Note that, for the $V_{\infty t}$ -interval considered, the maximum deflection angle is $110\text{--}150^\circ$. The maximum deflection angle can be determined analytically by substituting $r_3 = R$ into (18.67):

$$\alpha_{max} = 2 \arcsin \left(\frac{1}{1 + V_{\infty t}^2 / V_{c, surf}^2} \right) \quad (18.71)$$

where $V_{c, surf}$ is the circular velocity at the surface of the planet. This maximum deflection angle is shown in Figure 18.21 (left) as a function of $V_{\infty t}$ for the various planets. Clearly visible is the increase of α_{max} at increasing values of μ and decreasing values of $V_{\infty t}$.

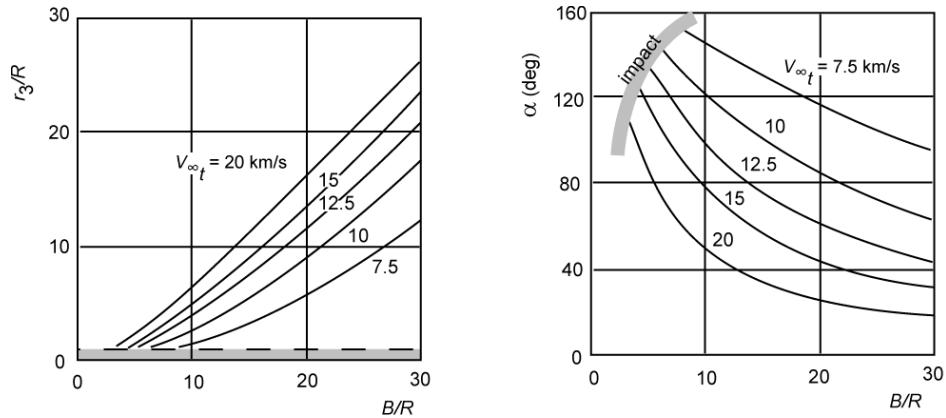


Figure 18.20: Distance at closest approach (left) and asymptotic deflection angle (right) of Jupiter flyby missions for several values of $V_{\infty t}$.

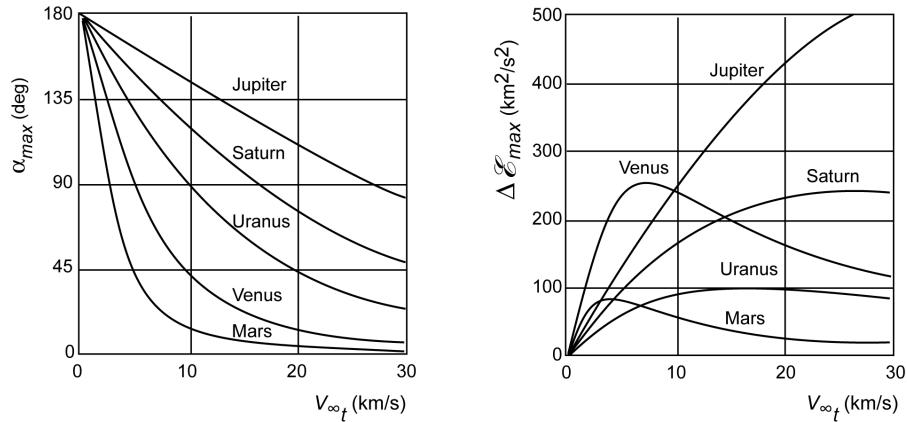


Figure 18.21: Maximum asymptotic deflection angle (left) and maximum heliocentric orbital energy increase (right) for swingby missions. For the energy increase, it has been assumed that $V'_t = V_t$.

18.11. Swingby flights

Figure 18.19 shows that the direction of $\bar{V}_{\infty t}^*$ is different from the direction of $\bar{V}_{\infty t}$. Consequently, the spacecraft's heliocentric velocity at the moment that it leaves again the planet's sphere of influence, \bar{V}_4 , differs from the heliocentric velocity at the moment that the spacecraft enters the sphere of influence, \bar{V}_2 . So, the result of passing a planet at a short distance is that the kinetic energy of the spacecraft in its heliocentric trajectory changes without using a rocket engine and expelling propellants. This, in essence, is the *swingby effect*. Because the dimension of the planet's sphere of influence is small on the scale of the solar system, the potential energy of the spacecraft in the gravity field of the Sun hardly changes over a distance corresponding to the diameter of the planet's sphere of influence. Consequently, we conclude that, when viewed on the scale of the solar system, the passage of a planet produces a discontinuous change of the total energy of the spacecraft's heliocentric trajectory. This is an illustration of the fact that in a many-body system the total energy of a particular body is not constant (Section 2.1). Of course, the change of the spacecraft's total orbital energy has to be compensated by a minute change in the orbit of the swingby planet. So, the spacecraft's orbital energy change comes from the planet's motion about the Sun. Generally, the swingby technique is used to increase a spacecraft's velocity and propel it outward in the solar system. However, the swingby effect can also be used to decrease a spacecraft's orbital energy or to change the inclination of the trajectory, which

would require a large amount of propellant if this would be performed by rocket thrust (Section 13.3). The swingby concept can also be used to increase the payload mass that can be carried by a launcher for a specific target planet. To illustrate this, the following example may serve. Consider a mission to Saturn, but assume that the spacecraft is so massive that the selected launcher can only inject it into a heliocentric trajectory with apohelion between the orbits of Jupiter and Saturn. Then, the spacecraft cannot reach the surroundings of Saturn. But if the trajectory is selected such that the spacecraft passes Jupiter at a close distance, then the swingby effect can produce an increase of the semi-major axis (total orbital energy) of the heliocentric trajectory that results in an apohelion beyond the orbit of Saturn, which allows the spacecraft to reach the surroundings of Saturn. The swingby effect can also be used for entering an (elliptical) orbit around another planet. In that case, which is called *gravity-assisted capture*, the gravity field of a moon of that planet is used to reduce, sometimes in combination with a rocket engine thrusting period, the energy of the hyperbolic orbit of the spacecraft about the planet so that it enters an orbit about that planet.

The application of the swingby effect is well known in the interplanetary space programs since the 1960s. Many swingby missions have already been flown, involving (a) single or multi-planet flyby(s). Examples are: Earth-Venus-Mercury-Mercury-Mercury (Mariner 10), Earth-Jupiter-interstellar (Pioneer 10), Earth-Jupiter-Saturn-interstellar (Pioneer 11), Earth-Jupiter-Saturn-interstellar (Voyager 1), Earth-Jupiter-Saturn-Uranus-Neptune-interstellar (Voyager 2), Earth-Jupiter-out of ecliptic (Ulysses), Earth-Venus-Earth-Jupiter (Galileo), Earth-Venus-Venus-Earth-Jupiter-Saturn (Cassini), Earth-Earth-Mars-Earth-Comet (Rosetta), Earth-Earth-Venus-Venus-Mercury-Mercury-Mercury (Messenger). Our own Moon has also been used for the same purpose, although it is a poor accelerator.

It should be remarked that the basic principles of the gravity-assist concept have been well known by astronomers for at least about 150 years in relation to the effects of Jupiter's gravitational attraction on the motion of comets. It is clear, for instance, that U.J.J. Le Verrier (1811-1877) understood the gravity-assist mechanism and that several astronomers have repeatedly derived the basic formulas, which are now finding such important applications in space missions. They, however, never proposed the application of this mechanism to make the solar system accessible to spacecraft launched from the Earth. As an illustration, we consider the motion of a comet that passes Jupiter at a close distance, but the same methodology can be applied for an asteroid and a flyby of any other giant outer planet.

When we consider the motion of the comet as a swingby flight, then we view the trajectory of the comet before it arrives in the surroundings of Jupiter as a Keplerian orbit about the Sun, and after it has passed Jupiter again as a Keplerian trajectory about the Sun. During the intermediate period the gravity field of Jupiter perturbs the orbit significantly. Many times, this leads to the question whether a comet that is detected when it moves close to the Sun is the same comet that was detected in deep space before it approached Jupiter. To analyze that problem, we consider the trajectory of the comet partly as a patched-conic problem (Section 18.1) and partly as a circular restricted three-body problem (Section 3.3), where the two main bodies are the Sun and Jupiter, and the body with negligible mass is the comet. Note that in the three-body model, there is no sphere of influence and the gravitational forces of Sun and Jupiter act continuously on the comet. In the discussion on the circular restricted three-body problem, we have derived equation (3.57), which expresses Jacobi's integral in terms of position and velocity relative to a barycentric inertial (non-rotating) reference frame. The various terms in that equation are expressed in the non-dimensional units introduced in Section 3.3. When we write that equation in the usual physical units and apply the resulting expression for the Sun-Jupiter system, we find

$$V^2 - 2G\left(\frac{m_S}{r_1} + \frac{m_J}{r_2}\right) - 2n_J(\xi \dot{\eta} - \eta \dot{\xi}) = -r_J^2 n_J^2 C \quad (18.72)$$

where V is the velocity of the comet relative to the non-rotating barycentric (inertial) reference frame defined in Figure 3.4, m_S and m_J are the masses of the Sun and Jupiter, respectively, r_1 and r_2 are the distances of the comet from the Sun and Jupiter, respectively, n_J is the mean (angular) motion of Jupiter in its orbit about the Sun, ξ and η are coordinates relative to the barycentric reference frame, r_J is the radius of Jupiter's orbit about the Sun, and C is the constant introduced in (3.53). Because $m_S \gg m_J$, the origin of the reference frame is close to the center of mass of the Sun. Therefore, we may use the following approximations: the barycentric distance r ($r \approx r_1$) and V indicate the position and velocity of the comet relative to a non-rotating heliocentric reference frame, and the term $(\xi \dot{\eta} - \eta \dot{\xi})$ describes the component of the comet's orbital angular momentum about the Sun and perpendicular to Jupiter's orbital plane, H_ζ . Because in the restricted circular three-body problem Jupiter moves in a circular orbit about the Sun:

$$n_J = \sqrt{\frac{\mu_S}{r_J^3}}$$

where μ_S is the gravitational parameter of the Sun ($\mu_S = Gm_S$). With this relation and the approximations mentioned above, (18.72) can be written as

$$V^2 - 2G\left(\frac{m_S}{r} + \frac{m_J}{r_2}\right) - 2\sqrt{\frac{\mu_S}{r_J^3}} H_\zeta = -\frac{\mu_S}{r_J} C \quad (18.73)$$

Before and after the passage of Jupiter, when the comet is far from Jupiter, we may approximate the comet's trajectory by a Keplerian orbit about the Sun. So, for both parts of the trajectory the classical relations for Keplerian orbits hold:

$$V^2 - \frac{2\mu_S}{r} = -\frac{\mu_S}{a} ; \quad H_\zeta = \sqrt{\mu_S a (1 - e^2)} \cos i \quad (18.74)$$

where a , e , and i are the semi-major axis, eccentricity, and inclination (relative to the orbital plane of Jupiter) of the comet's trajectory. Substitution of (18.74) into (18.73) yields

$$\frac{\sqrt{r_J^3}}{a} \left(1 + 2 \frac{m_J}{m_S} \frac{a}{r_2} \right) + 2 \sqrt{a(1 - e^2)} \cos i = \sqrt{r_J} C \quad (18.75)$$

Since $m_J/m_S \ll 1$, we may conclude that far from Jupiter the second part of the term in brackets is much smaller than one and therefore may be neglected. For both parts of the trajectory, generally, the values of a , e and i will be different. Substitution of (18.74) into (18.75) then yields

$$\frac{\sqrt{r_J^3}}{2a_4} + \sqrt{a_4(1 - e_4^2)} \cos i_4 = \frac{\sqrt{r_J^3}}{2a_2} + \sqrt{a_2(1 - e_2^2)} \cos i_2 \quad (18.76)$$

where the indices 2 and 4 indicate the situation before and after the swingby about Jupiter, respectively. It is remarkable that the gravitational parameter of the Sun does not appear in this equation. Expression (18.76) is known as the *criterion of Tisserand*; it was derived by F.F. Tisserand (1845-1896) in 1889. The value of the expression on either side of the equal-sign in

(18.76) is called the *Tisserand parameter*. As mentioned before, in classical astronomy this criterion is used to verify whether a comet of which the orbit has been determined before it made a flyby of an outer planet really is the same body as a comet of which the orbit has been determined when it moves relatively close to the Sun. If for both orbits substitution of the values of the semi-major axis, eccentricity and inclination into (18.76) leads to the same Tisserand parameter, then it is quite likely that it is the same body, and only then laborious calculations are started to follow the entire trajectory in detail. Nowadays, Tisserand's criterion is used to help clarify the physical processes involved in the origin of short-period comets. It has also found application in the computation of trajectories of interplanetary spacecraft; for example, to obtain some physical insight in the characteristics of swingby trajectories, or to check the accuracy of a numerical integration of the spacecraft's trajectory. It is emphasized that in the application of Tisserand's criterion to swingby trajectory design, the procedure used for the classical astronomical application is reversed. Transfer trajectories from the Earth to the swingby planet and from the swingby planet to the target planet are computed using Lambert's theorem (Section 6.7) or equivalent relations for parabolic and hyperbolic trajectories. These trajectories are then matched at the swingby planet based on Tisserand's criterion to identify viable launch and encounter opportunities.

We now return to the analysis of swingby trajectories of interplanetary spacecraft, and will investigate some characteristics of these types of flights.

The interplanetary trajectory before the approach of a swingby planet, and the position of the spacecraft along this trajectory, determines the point of entry into the swingby planet's sphere of influence and the planetocentric hyperbolic excess velocity, $\bar{V}_{\infty t}$, at that moment. These parameters determine the flyby trajectory about the swingby planet, the point of exit of the swingby planet's sphere of influence and the planetocentric hyperbolic excess velocity vector, $\bar{V}_{\infty t}^*$, at that moment. These parameters in turn determine the spacecraft's heliocentric trajectory after the swingby maneuver. If the swingby planet approach trajectory lies in the orbital plane of the swingby planet, and if the planet's gravity field is described by that of a point mass, then the plane of the swingby trajectory and that of the spacecraft's heliocentric trajectory after leaving the swingby planet's sphere of influence also coincide with the orbital plane of the swingby planet. However, in general, we deliberately select the three successive orbital planes of the spacecraft not to coincide. By selecting appropriate values of the 'latitude' and 'longitude' of the point of entrance into the swingby planet's sphere of influence, which requires a certain inclination of the approach trajectory relative to the planet's orbital plane, and the magnitude and direction of $\bar{V}_{\infty t}$, the flyby trajectory (including its inclination) and the heliocentric trajectory after the spacecraft has left the swingby planet's sphere of influence will satisfy the mission requirements. The computation of the transitions from one section of the entire trajectory to the next one, and of the corresponding inclinations of the three orbital planes, is in essence a

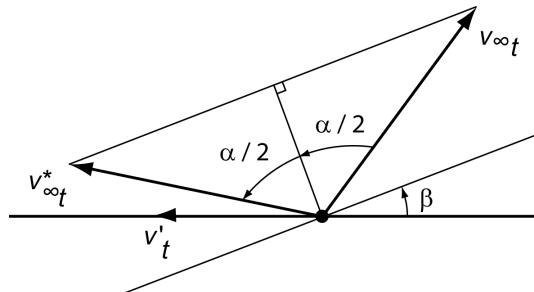


Figure 18.22: Diagram of the velocity vectors that characterize the swingby maneuver.

geometric problem, which will not be discussed in the Chapter. Here, we will only discuss some characteristics of the motion in the orbital plane of the swingby trajectory. For the heliocentric velocities of the spacecraft before and after the swingby flight we may write

$$\bar{V}_2 = \bar{V}_t + \bar{V}_{\infty_t} ; \quad \bar{V}_4 = \bar{V}_t + \bar{V}_{\infty_t}^* \quad (18.77)$$

Figure 18.22 shows the geometry of the vectors \bar{V}_{∞_t} and $\bar{V}_{\infty_t}^*$, together with the relevant angles α and β , and the projection of the swingby planet's heliocentric velocity on the spacecraft's orbital plane, \bar{V}' . These vectors constitute the plane of the swingby trajectory. As was already mentioned before: $|\bar{V}_{\infty_t}^*| = |\bar{V}_{\infty_t}|$. For the change in the velocity of the spacecraft due to the passage of the planet we can write with (18.77)

$$\Delta \bar{V} = \bar{V}_4 - \bar{V}_2 = \bar{V}_{\infty_t}^* - \bar{V}_{\infty_t}$$

From this relation, we obtain

$$(\Delta V)^2 = \Delta \bar{V} \cdot \Delta \bar{V} = (\bar{V}_{\infty_t}^* - \bar{V}_{\infty_t}) \cdot (\bar{V}_{\infty_t}^* - \bar{V}_{\infty_t})$$

Elaboration of the scalar product and using the geometry depicted in Figure 18.22, we obtain

$$(\Delta V)^2 = V_{\infty_t}^{*2} - 2 V_{\infty_t}^* V_{\infty_t} \cos \alpha + V_{\infty_t}^2 = 2 V_{\infty_t}^2 (1 - \cos \alpha)$$

or

$$\Delta V = 2 V_{\infty_t} \sin \frac{1}{2} \alpha \quad (18.78)$$

This velocity change can be decomposed into a radial component, ΔV_r , and a normal component, ΔV_n ; both components lie in the plane of the swingby trajectory. For these components we can write with (18.77)

$$\Delta V_r = \bar{V}_4 \cdot \bar{e}_r - \bar{V}_2 \cdot \bar{e}_r = (\bar{V}_{\infty_t}^* - \bar{V}_{\infty_t}) \cdot \bar{e}_r$$

$$\Delta V_n = \bar{V}_4 \cdot \bar{e}_n - \bar{V}_2 \cdot \bar{e}_n = (\bar{V}_{\infty_t}^* - \bar{V}_{\infty_t}) \cdot \bar{e}_n$$

where \bar{e}_r and \bar{e}_n are unit vectors along the Sun-planet vector and along the planet's velocity vector, respectively. Using the geometry depicted in Figure 18.22, we may write these expressions as

$$\Delta V_r = V_{\infty_t}^* \cos \left(\frac{1}{2} \pi - \left(\frac{1}{2} \pi - \frac{1}{2} \alpha - \beta \right) \right) - V_{\infty_t} \cos \left(\frac{1}{2} \pi - \left(\frac{1}{2} \pi - \frac{1}{2} \alpha + \beta \right) \right)$$

$$\Delta V_n = V_{\infty_t}^* \cos \left(\frac{1}{2} \pi - \frac{1}{2} \alpha - \beta \right) - V_{\infty_t} \cos \left(\frac{1}{2} \pi - \frac{1}{2} \alpha - \beta + \alpha \right)$$

or, after some trigonometric manipulation,

$$\begin{aligned} \Delta V_r &= -2 V_{\infty_t} \sin \frac{1}{2} \alpha \sin \beta = -\Delta V \sin \beta \\ \Delta V_n &= 2 V_{\infty_t} \sin \frac{1}{2} \alpha \cos \beta = \Delta V \cos \beta \end{aligned} \quad (18.79)$$

Because the radius of the sphere of influence is very small compared to the scale of the solar system, the change of the spacecraft's heliocentric potential energy (per unit of mass) over a

distance corresponding to the diameter of the sphere of influence may be neglected. Then, we can write for the change of the total energy of the spacecraft (per unit of mass) due to the passage of the planet

$$\Delta E = \frac{1}{2}(V_4^2 - V_2^2) = \frac{1}{2}(\bar{V}_4 + \bar{V}_2) \cdot (\bar{V}_4 - \bar{V}_2)$$

Substitution of the expressions (18.77) and elaboration of the scalar product leads to

$$\Delta E = \bar{V}_t' \cdot (\bar{V}_{\infty_t}^* - \bar{V}_{\infty_t})$$

where we have used the fact that the scalar products of the component of the planet's velocity vector perpendicular to the plane of the swingby trajectory and the hyperbolic excess velocities are zero. Using the angles indicated in Figure 18.22, we may write this expression as

$$\Delta E = V_t' V_{\infty_t}^* \cos\left(\frac{1}{2}\pi - \frac{1}{2}\alpha - \beta\right) - V_t' V_{\infty_t} \cos\left(\frac{1}{2}\pi - \frac{1}{2}\alpha - \beta + \alpha\right)$$

or, after some trigonometric manipulation,

$$\Delta E = 2V_t' V_{\infty_t} \sin\frac{1}{2}\alpha \cos\beta = V_t' \Delta V \cos\beta \quad (18.80)$$

We can compute the change of the semi-major axis of the spacecraft's interplanetary trajectory due to a swingby flight by starting from the classical relation

$$E = -\frac{\mu_s}{2a}$$

From this relation we obtain

$$\Delta\left(\frac{1}{a}\right) = -\frac{2}{\mu_s} \Delta E$$

Substitution of (18.80) into this relation gives

$$\Delta\left(\frac{1}{a}\right) = -\frac{4}{\mu_s} V_t' V_{\infty_t} \sin\frac{1}{2}\alpha \cos\beta \quad (18.81)$$

This equation is written in the parameter $1/a$, because a swingby may result in a transition from an elliptical trajectory to a hyperbolic trajectory. In that case, the semi-major axis shows a discontinuity and goes through infinity at an eccentricity equal to one, while the inverse semi-major axis then goes through zero and changes smoothly from positive to negative values.

With (18.68) the relations (18.78) to (18.81) may also be written in terms of the selected set of independent parameters B , V_{∞_t} , β :

$$\begin{aligned} \Delta V &= \frac{2V_{\infty_t}}{\sqrt{1 + B^2 V_{\infty_t}^4 / \mu^2}} \\ \Delta V_r &= -\frac{2V_{\infty_t} \sin\beta}{\sqrt{1 + B^2 V_{\infty_t}^4 / \mu^2}} \quad ; \quad \Delta V_n = \frac{2V_{\infty_t} \cos\beta}{\sqrt{1 + B^2 V_{\infty_t}^4 / \mu^2}} \end{aligned} \quad (18.82)$$

$$\Delta\mathcal{E} = \frac{2V'_t V_{\infty_t} \cos\beta}{\sqrt{1 + B^2 V_{\infty_t}^4/\mu^2}} ; \quad \Delta\left(\frac{1}{a}\right) = -\frac{4}{\mu_s} \frac{V'_t V_{\infty_t} \cos\beta}{\sqrt{1 + B^2 V_{\infty_t}^4/\mu^2}} \quad (18.82)$$

From these expressions a number of interesting conclusions can be drawn:

- ΔV is independent of β . For specified values of V_{∞_t} and β , the quantities ΔV , $|\Delta V_r|$, $|\Delta V_n|$, $|\Delta\mathcal{E}|$ and $|\Delta(1/a)|$ increase at decreasing values of B .
- When the spacecraft passes ‘ahead’ of the planet ($90^\circ < \beta < 270^\circ$; Figure 18.19) the trajectory’s total energy will decrease. This decrease is maximum at $\beta = 180^\circ$; then the major axis of the flyby hyperbola coincides with the velocity vector V'_t .
- In most cases, one aims for an increase of \mathcal{E} . This requires the spacecraft to pass ‘behind’ the planet ($-90^\circ < \beta < 90^\circ$). The energy increase is maximum at $\beta = 0^\circ$; then the major axis of the flyby hyperbola coincides with the velocity vector V'_t .
- For the change of the radial and normal velocity components the following relations hold. If $0^\circ < \beta < 180^\circ$, then $\Delta V_r < 0$; if $180^\circ < \beta < 360^\circ$, then $\Delta V_r > 0$. For $\beta = 0^\circ, 180^\circ$: $\Delta V_r = 0$. If $-90^\circ < \beta < 90^\circ$, then $\Delta V_n > 0$; if $90^\circ < \beta < 270^\circ$, then $\Delta V_n < 0$. For $\beta = 90^\circ, 270^\circ$: $\Delta V_n = 0$.
- For specified values of B , V_{∞_t} , β , and μ , $|\Delta\mathcal{E}|$ is maximum when V'_t is maximum; i.e. if the plane of the swingby trajectory coincides with the orbital plane of the swingby planet ($V'_t = V_t$). $\Delta\mathcal{E} = 0$ when the plane of the swingby trajectory is perpendicular to V_t ($V'_t = 0$).
- When the angle between the plane of the swingby trajectory and the orbital plane of the swingby planet is relatively small: $V'_t \approx V_t$. Then, two planetary parameters, V_t and μ , affect the change of \mathcal{E} ; $|\Delta\mathcal{E}|$ increases when V_t and μ increase. In practice, these are conflicting requirements: for the inner planets μ is small and V_t is large, while for the giant outer planets μ is large and V_t is (relatively) small.
- If $-90^\circ < \beta < 90^\circ$, then $\Delta(1/a) < 0$ and $\Delta a > 0$, both for elliptical trajectories ($a > 0$) and for hyperbolic trajectories ($a < 0$). If $90^\circ < \beta < 270^\circ$, then $\Delta(1/a) > 0$ and $\Delta a < 0$. If $\beta = 90^\circ, 270^\circ$, then $\Delta(1/a) = 0$ and $\Delta a = 0$.
- When the plane of the swingby trajectory is perpendicular to V_t , then $\Delta(1/a) = 0$ and $\Delta a = 0$.

Using (18.64) and (6.18), we can write (18.82-4) as

$$\Delta\mathcal{E} = \frac{2V'_t V_{\infty_t} \cos\beta}{1 + V_{\infty_t}^2/V_{c_3}^2} \quad (18.83)$$

where V_{c_3} is the circular velocity at the pericenter distance, r_3 , of the flyby trajectory. Note that the smaller r_3 , and consequently the larger V_{c_3} , the larger $|\Delta\mathcal{E}|$. As already mentioned in Section 18.10, the interplanetary trajectory flown before reaching the swingby planet determines the values of the parameters V_{∞_t} , β and B , and so of r_3 ; the values of r_3 and β can be adjusted by executing mid-course corrections. In order to achieve a maximum positive $\Delta\mathcal{E}$, we must require, according to (18.83) $\beta = 0^\circ$; $r_3 = R$, and we find

$$\Delta\mathcal{E}_{max} = \frac{2V'_t V_{\infty_t}}{1 + V_{\infty_t}^2/V_{c_{surf}}^2} \quad (18.84)$$

where $V_{c_{surf}}$ is the circular velocity at the surface of the planet. For a maximum negative $\Delta\mathcal{E}$, the requirements are $\beta = 180^\circ$; $r_3 = R$, and we find, except for the minus-sign, again (18.84). For the ΔV associated with a maximum value of $|\Delta\mathcal{E}|$, we find from (18.82-1)

$$(\Delta V)_{\Delta \mathcal{E}_{max}} = \frac{2 V_{\infty_t}}{1 + V_{\infty_t}^2 / V_{c_{surf}}^2} \quad (18.85)$$

The value of the maximum positive energy increase, $\Delta \mathcal{E}_{max}$, is plotted in Figure 18.21 (right) as a function of V_{∞_t} . Here, it was assumed that the plane of the swingby trajectory coincides with the plane in which the planet moves about the Sun ($V'_t = V_t$). Note that for small values of V_{∞_t} the planets Venus and Mars generate reasonably large values of $\Delta \mathcal{E}_{max}$. Therefore, for a mission to Mercury it is attractive to perform a swingby about Venus. For missions to Jupiter, Mars is a ‘reasonable’ swingby planet; for flights to the outer planets, Jupiter is an ideal swingby planet, in particular for larger values of V_{∞_t} , with Saturn in second place.

The value of V_{∞_t} for which the absolute maximum positive $\Delta \mathcal{E}_{max}$ is achieved and the corresponding value of $\Delta \mathcal{E}_{max}$ can be found analytically by partial differentiation of (18.84) to V_{∞_t} and by setting the result equal to zero. We then find for $V'_t = V_t$

$$V_{\infty_{t, opt}} = V_{c_{surf}} \quad ; \quad \Delta \mathcal{E}_{max, abs} = V_t V_{c_{surf}} \quad (18.86)$$

For such a flight, we find from Figure 18.19 and with (6.23), (18.64), (18.65), (18.68) to (18.70), (18.78), (18.79) and (18.81)

$$\alpha = 60^\circ \quad ; \quad \beta = 0^\circ \quad ; \quad \varphi_2 = 60^\circ \quad ; \quad \varphi_4 = 120^\circ$$

$$B = \sqrt{3}R \quad ; \quad e = 2 \quad ; \quad r_3 = R \quad ; \quad V_3 = \sqrt{3}V_{c_{surf}}$$

$$\Delta V = V_{c_{surf}} \quad ; \quad \Delta V_r = 0 \quad ; \quad \Delta V_n = V_{c_{surf}} \quad ; \quad \Delta \left(\frac{1}{a} \right) = -\frac{2}{\mu_S} V_t V_{c_{surf}}$$

So, for the case of an absolute maximum positive $\Delta \mathcal{E}$: $\Delta V = V_{\infty_{t, opt}}$. For a swingby about Jupiter we find from (18.86) that the absolute maximum value of $\Delta \mathcal{E} = 550 \text{ km}^2/\text{s}^2$ occurs at $V_{\infty_t} = 42.1 \text{ km/s}$. For a swingby about Mars, the corresponding values are $\Delta \mathcal{E} = 85.68 \text{ km}^2/\text{s}^2$ and $V_{\infty_t} = 3.55 \text{ km/s}$. Of course, these examples are unrealistic since for these flights $r_3 = R$, which is impossible because of the presence of planet topographic features and an atmosphere, and for Jupiter the presence of high-energy particles radiation belts. In addition, the optimum hyperbolic excess velocity for the approach of Jupiter requires an unrealistically high spacecraft heliocentric approach velocity. If the spacecraft is first injected into an orbit about the Sun in such a way that it passes the Earth after some time, a swingby about the Earth can also increase the spacecraft’s heliocentric orbital energy. From (18.86) we find that for an Earth swingby the absolute maximum value of $\Delta \mathcal{E} = 235 \text{ km}^2/\text{s}^2$ occurs at $V_{\infty_t} = 7.91 \text{ km/s}$. These values are quite close to the absolute maximum value of $\Delta \mathcal{E} = 257 \text{ km}^2/\text{s}^2$ and the optimum value of $V_{\infty_t} = 7.33 \text{ km/s}$ for a Venus swingby.

For specified values of V_{∞_t} , β and B , the direction and magnitude of the heliocentric velocity of the spacecraft after the swingby maneuver (V_4 , γ_4) are fixed. These values determine the flight time from the swingby planet to the next swingby planet or to the final target planet. Note that there are two conflicting effects: a large positive value of $\Delta \mathcal{E}$ results in a large value of V_4 , which tends to decrease the flight time to the next planet, but it also results in a large value of α , so a small value of V_r and thus of γ_4 , which tends to increase the flight time to the next planet. We have to rely on numerical search processes to determine, for a specified overall flight time to the target planet, the minimum injection impulsive shot, ΔV_0 , and the optimum injection conditions

and mid-course corrections to be applied, such that the spacecraft's arrival conditions at the swingby planet generate an appropriate swingby maneuver that sets the spacecraft on the right course to the next planet. When we consider, for example, Jupiter swingby missions with a total flight time of three to four years to Saturn, we generally find for minimum ΔV_0 trajectories the following characteristic values:

$$\text{Earth: } \Delta V_0 = 6.6 - 7.1 \text{ km/s} ; \gamma_1 = 0 - 1^\circ$$

$$\text{Jupiter: } r_3 = 10 - 11 R_J ; \gamma_4 - \gamma_2 = -23^\circ \text{ to } -29^\circ ; \Delta E = 155 - 165 \text{ km}^2/\text{s}^2$$

$$\text{Saturn: } V_\infty = 11.5 - 14.5 \text{ km/s}$$

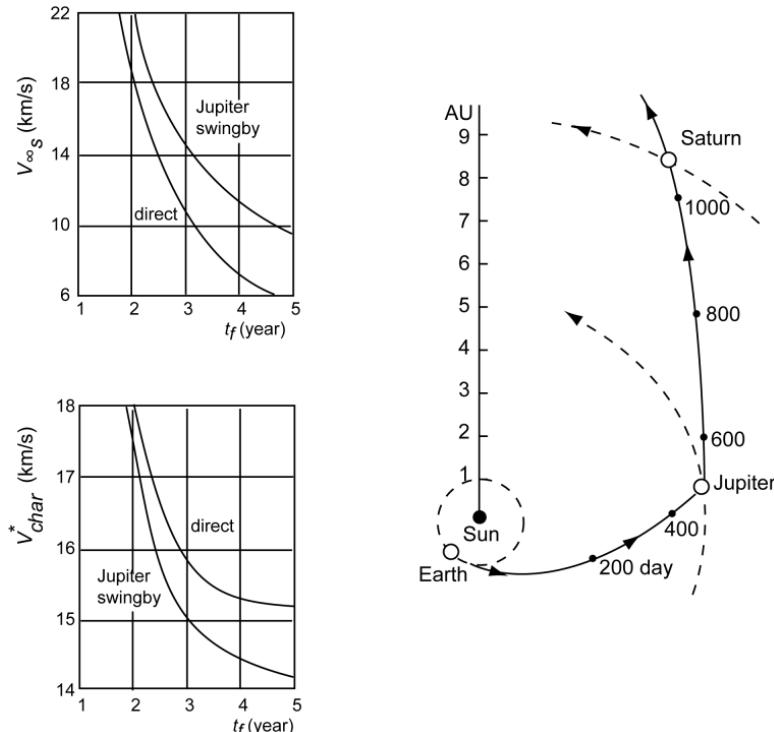


Figure 18.23: Mission characteristic velocity, V_{char}^* , and target planet hyperbolic excess velocity, $V_{\infty S}$, versus flight time, t_f , for a direct flight to Saturn and a flight that includes a Jupiter swingby (left), and an example of a swingby trajectory to Saturn (right).

Figure 18.23 (left) shows the decrease in mission characteristic velocity, V_{char}^* , for a flight to Saturn if a Jupiter swingby is applied, or stated differently: the reduction of the overall flight time to Saturn for the same mission characteristic velocity. This Figure also shows that the hyperbolic excess velocity relative to Saturn, $V_{\infty S}$, increases significantly due to the Jupiter swingby. This means that if the spacecraft has to enter an orbit about Saturn, a much larger deceleration impulse has to be delivered than without a Jupiter swingby. Figure 18.23 (right) depicts a fast trajectory to Saturn with a swingby at Jupiter and a total flight time of 2.9 year. The characteristic velocity of this mission is about 15.1 km/s; the minimum distance to Jupiter is $4R_J$, which results in an increase of the heliocentric velocity of the spacecraft to about 18.5 km/s. Figure 18.24 shows the mission characteristic velocity, V_{char}^* , of direct and swingby flights to Mercury and Uranus, where for a specified overall flight time, t_f , the optimum swingby maneuver about a swingby planet is executed. Note that for very short flight times to Mercury, a direct flight is more attractive than a Venus swingby flight. For missions to Uranus, a Jupiter swingby, or to a somewhat lesser extent a Saturn swingby, results in a significant saving of rocket engine propellants for a specified overall flight time, or in a significant reduction of the overall flight time for the same

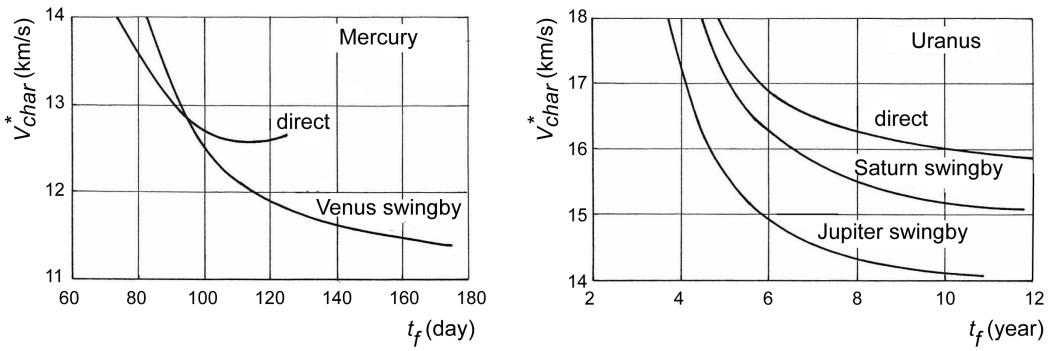


Figure 18.24: Mission characteristic velocity, V_{char}^* , versus flight time, t_f , for a direct and a Venus-swingby flight to Mercury (left), and for a direct, Jupiter- or Saturn-swingby flight to Uranus (right).

value of ΔV_0 .

A way to get even more energy from a swingby is to fire a rocket engine near the pericenter of a flyby trajectory to increase the spacecraft's velocity. For a given vehicle, a specified rocket engine impulsive shot always provides the same change in velocity (ΔV), but the change in kinetic energy is proportional to the vehicle's velocity at the time of the burn. Equation (1.55) shows that to get the most additional kinetic energy from the burn, it must occur at the vehicle's maximum velocity, i.e. at pericenter, and the thrust has to act along the velocity vector. In that case, (1.55) gives

$$\Delta \mathcal{E}_{\text{thr}} = \frac{1}{2} (\Delta V_{\text{thr}})^2 + V_3 \Delta V_{\text{thr}}$$

where the index *thr* is used to indicate that ΔV and $\Delta \mathcal{E}$ are produced by rocket thrust. This maneuver is known as the *Oberth maneuver* after H.J. Oberth (1894-1989), one of the founders of spaceflight, who first described it in 1929. The effect of thrusting can be large; for a Jupiter flyby at a pericenter distance of $6R$, $V_{\infty t} = 7.5$ km/s, and $\beta = 0^\circ$, we find from (18.82) a maximum value of $\Delta \mathcal{E} = 164$ km 2 /s 2 . With (18.5) we find $V_3 = 25.4$ km/s and thus a rocket engine thrusting period producing $\Delta V_{\text{thr}} = 2$ km/s results in an additional $\Delta \mathcal{E}_{\text{thr}} = 53$ km 2 /s 2 .

As already mentioned before, on a single interplanetary mission multiple swingby maneuvers can be performed. The flight then may be considered as a ‘cosmic billiard’, where the spacecraft is catapulted from one planet to the next planet and so on. In principle, many planetary combinations are possible. For instance, when the target planet is Saturn, it is possible that the spacecraft is launched from the Earth towards the inner planet Venus, performs a swingby maneuver at Venus, returns to the Earth, performs a swingby maneuver about the Earth, flies back to Venus for a second swingby about that planet, which accelerates the spacecraft towards Jupiter for the last swingby on its route to Saturn. The design of these multiple swingby trajectories involves detailed numerical search procedures that start from accurate three-dimensional planetary orbits. From these orbits, the instantaneous positions and velocities of the planets can be computed. For a series of launch times within a certain period and a selected sequence of swingby planets, it can be investigated whether feasible multi-swingby trajectories towards the target planet exist. To enter the sphere of influence of each swingby planet in an appropriate way, small trajectory corrections that require only small amounts of rocket propellant may be executed when the spacecraft has left the sphere of influence of a planet and is on its way to the next planet. Of all possible solutions found, then the trajectory is selected that best satisfies the mission constraints, e.g. low injection energy, suitable arrival conditions at the target planet, etc.

As examples of multiple swingby missions, some trajectory information for the Voyager-1 and -2, Galileo, Ulysses, Cassini and Rosetta missions is presented.

The two (almost) identical 722 kg NASA Voyager spacecraft provide a classical example of utilizing swingbys to reach the outer planets (Figure 18.25). They were launched aboard Titan-III/Centaur rockets from Cape Canaveral with destinations of Saturn and beyond. But their launch vehicles could provide only enough energy to get them to Jupiter and therefore they needed the swingby effect to increase their orbital energy. Voyager 1 was launched on September 5, 1977, and is the most distant human-made object in existence. Around August 2012, the spacecraft left the heliosphere and entered interstellar space. As of September 19, 2014, Voyager 1 was at a distance of 19.31×10^9 km (129.1 AU) from the Sun, and was traveling at 17.03 km/s (3.6 AU per year) relative to the Sun. It gained the energy to escape the Sun's gravity completely by performing swingby maneuvers around Jupiter and Saturn. It passed these planets on March 5, 1979, and November 12, 1980, at a minimum distance of 348,900 km and 184,300 km, respectively, from the center of the planet, and left the ecliptic plane at an angle of about 35° . In 40,000 years, it should pass within 1.6 light-year the star Gliese 445 in the Camelopardalis (Giraffe) constellation. The Voyager 2 mission started on August 20, 1977; so, slightly before Voyager 1. During this mission four outer planets (Jupiter (July 9, 1979), Saturn (August 26, 1981), Uranus (January 24, 1986), Neptune (August 25, 1989)) were passed at a minimum distance of 721,700 km, 161,000 km, 107,000 km, and 29,200 km, respectively, from the center of the planet. The passage of Jupiter resulted in a change of the spacecraft's velocity relative to the Sun from about 10 km/s to about 27 km/s. When the spacecraft arrived at Saturn, its heliocentric velocity had decreased to about 16 km/s. The passage of Saturn resulted in an increase of the heliocentric velocity to about 34 km/s. The spacecraft was able to undertake this 'Grand Tour' along the planets because of the particular way the outer planets were lined up in the late 1970s; this situation repeats every 176 years. After conclusion of the Neptune encounter, Voyager 2 left the ecliptic plane at an angle of about -45° . As of September 19, 2014, Voyager 2 was in the outermost layer of the heliosphere at a distance of about 15.85×10^9 km (105.9 AU) from the Sun, escaping the solar system at a speed of about 15.41 km/s (3.2 AU per year). In 40,000 years, it should pass within 1.7 light-years the star Ross 248 in the Andromeda constellation.

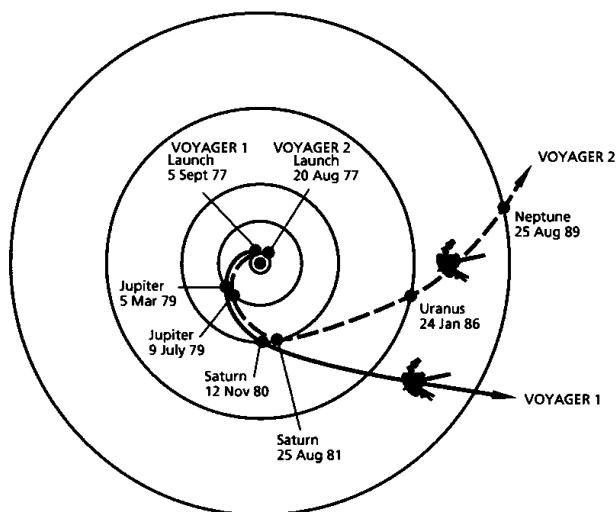


Figure 18.25: Interplanetary trajectories of the NASA Voyager-1 and Voyager-2 spacecraft.

The 2223 kg NASA Galileo spacecraft (including 925 kg of propellant) and its two-stage Inertial Upper Stage were launched on October 18, 1989, by the Space Shuttle from Cape Canaveral. The

two-stage upper stage accelerated the spacecraft out of Earth orbit towards the planet Venus (Figure 18.26, left) for the first of three planetary gravity assists designed to boost Galileo toward Jupiter. The spacecraft executed swingby maneuvers about Venus (February 10, 1990), Earth (December 8, 1990) and again Earth (December 8, 1992), with minimum flyby altitudes of 16,000 km, 960 km, and 303 km, respectively, before it arrived at Jupiter (December 7, 1995), where it used a gravity-assist flyby in front of Jupiter's largest moon, Io. In this way, it was possible for Galileo to decrease the amount of rocket engine propellant needed to insert into an orbit about Jupiter. On July 12, 1995, Galileo's descent probe, which has been carried aboard the spacecraft, was released and began a five-month journey to Jupiter. The probe had no thrusters, so its flight path was established by pointing of the spacecraft and spinning it up to 10.5 rpm before the spin-stabilized probe was released. When Galileo reached Jupiter on December 7, 1995, and while the probe was still approaching the planet, the spacecraft flew by two of Jupiter's major moons. It passed Europa at an altitude of about 33,000 km and Io at an altitude of about 900 km. About four hours after leaving Io, the spacecraft started receiving data from the descent probe (Figure 18.26, right), which slammed into the top of the Jovian atmosphere at a flight path angle of about -9° and at a speed of about 47.8 km/s. During entry it experienced temperatures of over 14,000 K and decelerations of over 300 g. Two minutes after entry a drogue parachute was released and the aft cover removed; somewhat later the main parachute deployed and the heat shield dropped off. The probe floated down under its parachute about 180 km through the clouds, transmitting information about the atmospheric conditions to the spacecraft. Fifty-eight minutes into its descent, high temperatures (130° C) and pressures (20 bar) silenced the probe's transmitters. An hour after receiving the last transmission from the probe, at an altitude of about 210,000 km above the planet, the Galileo spacecraft fired its 400 N thrust main engine for about 47 min to brake into orbit about Jupiter. During its primary mission orbital tour, Galileo performed four flybys of Jupiter's moon Ganymede, three of Callisto and three of Europa. The extended mission, called the 'Galileo Europa Mission', started in November 1997 and included eight consecutive close encounters of Europa. The flyby altitudes ranged from 201 km to 3580 km. This phase also provided the opportunity to make two additional flybys of Io and four addi-

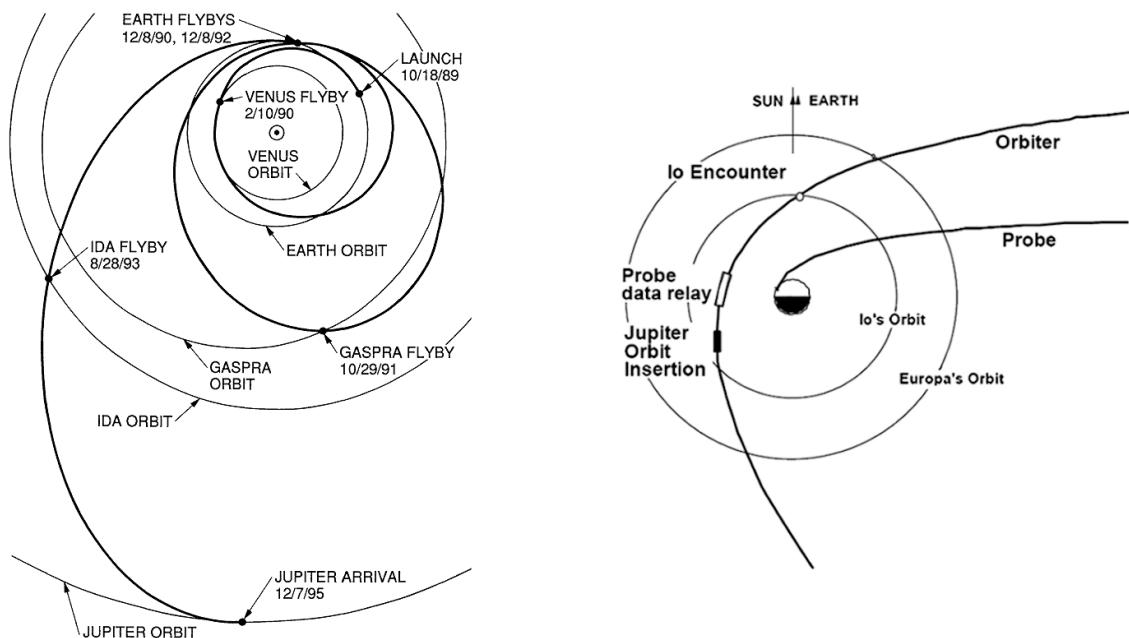


Figure 18.26: Interplanetary trajectory of the NASA Galileo spacecraft to Jupiter (left), and the trajectory of the orbiter and probe at Jupiter arrival (right).

tional flybys of Callisto. The follow-on mission extension, called the ‘Galileo Millenium Mission’, added another year of operation, including more flybys of Io and Ganymede. On Galileo’s first trip through the asteroid belt, the spacecraft made a flyby of the asteroid Gaspra at a minimum distance of 1601 km and a relative speed of about 8 km/s on October 29, 1991, and took detailed images of the asteroid; this was the first close approach to an asteroid. On a second pass through the asteroid belt, Galileo made a flyby of asteroid Ida on August 28, 1993, at a minimum distance of 2400 km, and discovered a miniature moon orbiting this asteroid. In July 1994, Galileo was perfectly positioned to watch the fragments of comet Shoemaker-Levy 9 crash into Jupiter within the period July 16 - 22. The impacts took place on the side of Jupiter hidden from Earth, and Galileo, then at a distance of 1.6 AU from the planet, could see the impacts as they occurred. Earth-based telescopes had to wait to see the impact sites as they rotated into view. The spacecraft’s fourteen years odyssey came to an end on September 21, 2003, when the spacecraft entered the giant planet’s atmosphere at about 0.2° south of Jupiter’s equator. Galileo crossed the altitude of the cloud tops at a flight path angle of about -22° and a relative velocity of 48.2 km/s, and disintegrated in Jupiter’s dense atmosphere somewhat later.

On October 6, 1990, the ESA/NASA 371 kg Ulysses spacecraft was launched by the Space Shuttle to study the polar regions of the Sun; places that are hidden from any observing location on Earth. To move to an orbit passing over the poles of the Sun, the spacecraft would have to eliminate the 30 km/s speed it inherited from the Earth’s orbital motion around the Sun and gain the speed needed to orbit the Sun in the pole-to-pole plane—tasks which are impossible with current spacecraft propulsion systems. Therefore, the spacecraft was sent towards Jupiter, aimed to arrive at a point in space just ‘in front of’ and ‘below’ the planet. The spacecraft then used a gravity assist to change the trajectory’s (ecliptic) inclination by about 80° . When separated from the Space Shuttle, the spacecraft consisted of Ulysses and a three-stage booster, consisting of a two-stage Boeing Inertial Upper Stage plus a McDonnell Douglas Payload Assist Module, to boost the spacecraft to the required high velocity. In February 1992, the spacecraft arrived at Jupiter; the closest approach to Jupiter occurred on February 8 at a distance of 6.3 Jovian radii or about 4.5×10^5 km from the center of the planet. This encounter changed the semi-major axis, eccentricity and inclination (relative to the ecliptic) of the spacecraft’s heliocentric orbit from 8.99 AU, 0.8892, 1.99° , respectively, to 3.37 AU, 0.6031, 79.13° , respectively. Before the encounter, the orbit’s perihelion and aphelion distances were about 1.0 AU and 17.0 AU; after the Jupiter encounter 1.34 AU and 5.41 AU. This post-encounter orbit has a period of 6.2 years, while the orbital period of Jupiter is about 11.9 years. So, about twelve years after the Jupiter encounter the spacecraft approached Jupiter again. This approach occurred on February 4, 2004, at a closest distance of about 1683 Jupiter radii (120×10^6 km). Because of this large distance, the orbit was now not significantly changed by Jupiter’s gravity field. After its first encounter with Jupiter, Ulysses reached a maximum southern heliocentric latitude of 80.2° on September 13, 1994. On March 13, 1995, it passed perpendicular through the ecliptic plane up to its northernmost heliocentric latitude of 80.2° on July 31, 1995. On May 1, 1996, the spacecraft unexpectedly crossed the ion tail of comet Hyakutake, revealing the tail to be at least 3.8 AU in length. Encounter with a comet tail happened again on February 5 - 9, 2007, when the spacecraft passed through the tail of comet McNaught at a distance of about 1.7 AU (254×10^6 km) from the nucleus. During these passes valuable information about comets was obtained. On January 14, 2008, the spacecraft reached maximum solar northern latitude on its third orbit about the Sun. Around that time the power supply of the X-band transmitter onboard the spacecraft could not be switched on again following what was planned to be a temporary switch off. This has left the mission with only S-band communication. This communication link can be used for playback of recorded data for very limited periods only, and since then the majority of the spacecraft data was

received in real-time during a tracking pass. The temporary switch off of the *X*-band transmitter was part of a plan to manage the naturally decreasing amount of power generated by the spacecraft's solar array after, at that time, more than seventeen years of operations. Due to that power generation degradation, the spacecraft no longer had enough power to run all of its communication, heating and scientific equipment simultaneously. As a result of the failure to switch on the *X*-band transmitter again, the heat input into the spacecraft has decreased considerably, which may result in hydrazine propellant freezing in the propellant pipework in the lower spacecraft and consequently to a loss of the spacecraft's attitude control. Analyses made shortly after the *X*-band transmitter switch on failure predicted the end of the spacecraft's operational life by July 2008. However, this prediction has proved to be too pessimistic. This is largely due to the introduction of the 'fuel bleeding' strategy, whereby two oppositely directed thrusters are fired simultaneously every two hours, causing the hydrazine to move through a short length of pipework where propellant is likely to freeze. This strategy has proven to be very successful; it kept the spacecraft operational for almost another year. On June 30, 2009, the mission was ended because the overall return of scientific data had decreased to a level where it was hard to justify the cost of keeping the spacecraft in operation. A command was sent to the spacecraft to switch the onboard communications system into the receive-only mode; no further contact with Ulysses is planned.

The NASA/ESA/ASI Cassini mission (Figure 18.27) aimed at studying the Saturnian system. The spacecraft was launched by a Titan IVB/Centaur rocket from Cape Canaveral on October 15, 1997. Its mass at launch was about 5700 kg (more than half was propellant) and no existing launch vehicle could have sent this massive spacecraft directly to Saturn. Therefore, four swingby maneuvers were applied: the spacecraft passed by Venus twice (April 26, 1998; June 24, 1999), then Earth (August 18, 1999), and finally Jupiter (December 30, 2000) on the way to Saturn at flyby distances of 284 km, 603 km, 1173 km, and 9.72×10^6 km, respectively. The 6.7 years transit was longer than the 6.05 years needed for a Hohmann transfer, but reduced the required injection velocity such that available launchers could launch the heavy Cassini probe on a journey to Saturn. During the first swingby about Venus the minimum flyby altitude was 287 km; this maneuver gave the spacecraft a boost in speed of about 7 km/s. Leaving Venus, Cassini was then moving at more than 39.2 km/s relative to the Sun. The second flyby took it within 600 km of the planet. Subsequently, the spacecraft successfully completed a highly accurate flyby of Earth on August 18, 1999. It flew past Earth at a minimum altitude of about 1171 km, above the eastern South Pacific at -23.5° latitude and 231.5° longitude. This flyby gave Cassini a 5.5 km/s boost

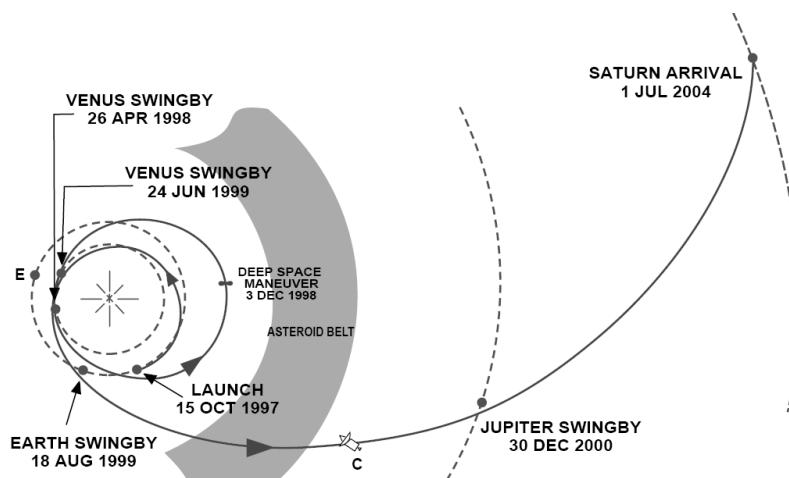


Figure 18.27: Interplanetary trajectory of the NASA Cassini spacecraft to Saturn.

in speed. The final gravity assist was from Jupiter; it flew by this planet at a minimum altitude of about 9.7×10^6 km on December 30, 2000, accelerated the spacecraft by about 2 km/s and gave Cassini the final thrust of energy it needed to reach its ultimate destination, Saturn. On arrival at Saturn, Cassini made a close approach to the planet, passing only 20,000 km above the cloud tops. Cassini fired one of its two redundant 445 N rocket engines on July 1, 2004, for 96 min to slow down by 622 m/s for Saturn orbit insertion; this maneuver has consumed 830 kg of propellant. The spacecraft entered an elliptical orbit about Saturn with pericenter and apocenter distances of 1.33 and 178 Saturn radii, respectively, with an orbital inclination of 16.8° relative to Saturn's equatorial plane and an orbital period of 148 days. Thirteen days after apocenter passage on the first orbit, a 50 min 335 m/s burn raised the pericenter distance to 8.2 Saturn radii in order to target the Orbiter for the first Titan encounter on October 26. A second Titan flyby at an altitude of 1,200 km on December 13 placed the Cassini-Huygens spacecraft on a trajectory that, if uncorrected, would lead to a subsequent flyby at an altitude of 4600 km. A probe targeting maneuver was executed on December 17 and placed the spacecraft on a direct impact trajectory with Titan during the next encounter. On December 25, the spring-loaded Huygens separation mechanism provided a velocity increment relative to the Cassini spacecraft of 35 cm/s. It also imparted to the lander an anti-clockwise spin of 7.5 rpm (when viewed from the orbiter). This provided inertial stability during the ballistic trajectory and atmospheric entry. After the separation of Huygens, Cassini performed a deflection maneuver, so that it would not crash into Titan, and an additional maneuver for final adjustment of its trajectory. These were on December 28, 2004, and January 3, 2005, respectively, and placed Cassini on the correct trajectory for receiving data from Huygens during its descent. Following its release from Cassini, Huygens reached Titan's outer atmosphere after 20 days. The probe entered the upper layers of Titan's atmosphere at an altitude of about 1000 km with a velocity of 6.1 km/s and flight path angle of -65.4° on January 14, 2005. The entry phase lasted about 3 min, during which the aeroshell surrounding the descent module reduced the probe's velocity to about 400 m/s, and a peak deceleration of about 20 g was experienced. Three parachutes were used during the probe's descent. At a velocity of Mach 1.5 near the end of the entry deceleration phase at an altitude of about 190 km, a 2 m diameter pilot parachute deployed, pulling off the aft cover. This was followed immediately by the deployment of a 8.3 m diameter main parachute. About 30 s after deployment of this parachute, the probe's velocity dropped from Mach 1.5 to Mach 0.6. The front heat shield was then released, and from an altitude of about 170 km the probe descended slowly below the main parachute for about 15 min. The main parachute then separated from the probe, at an altitude of about 140 km, and released a smaller 3 m diameter drogue parachute, which allowed the probe to descend faster, and ensured it would reach the surface before the batteries would expire. It arrived at the surface in 2.5 hr, with an impact velocity of about 6 m/s. The lander operated successfully for another 1 hr 10 min on the surface. Throughout the parachute descent to the surface and the subsequent active period on the surface Huygens has measured the chemical, meteorological, dynamical and physical properties of the Titan atmosphere and surface. Throughout this time, Cassini was oriented to receive the two incoming radio signals from the probe by continuously pointing its high gain antenna to the predicted Huygens landing point. After listening for the longest possible duration of the Huygens probe's visibility, the orbiter was commanded to re-point its high gain antenna to Earth for transmission of the stored Huygens telemetry data. At that time, Cassini was at a distance of 8.07 AU from the Earth. The orbiter kept examining Saturn's fascinating realm as it repeatedly looped around the planet, and performed a long series of swingby flights about many Saturnian moons and made observations of these moons. In its four-years nominal mission, Cassini made nine close targeted flybys of Enceladus (4), Phoebe, Hyperion, Dione, Rhea and Iapetus, with flyby altitudes ranging from 55 km to

125,000 km. Many distant flybys were made of other major moons. Originally, the mission was scheduled to end in July 2008. However, because of its great success, NASA and ESA decided in April 2008 to extend the mission for two more years. During this extended phase, Cassini performed several close and distant flybys of Saturn's moons. The encounters included 26 Titan flybys, 7 of Enceladus, 6 of Mimas, 8 of Tethys, and one flyby each of Dione, Rhea and Helene. A second mission extension was approved in February 2010. It will see Cassini orbiting Saturn 155 more times, conducting 54 additional flybys of Titan and 11 more of Enceladus. The chosen mission ending is a series of close Saturn passes, passing inside the rings, then a plunge into the Saturn atmosphere in 2017, to destroy the spacecraft.

ESA's Rosetta mission began on March 2, 2004, when an Ariane 5G launch vehicle lifted off from the ESA launch site at Kourou with on board the 3005 kg Rosetta spacecraft. The goal of the mission is to rendezvous with comet 67P/Churyumov-Gerasimenko in summer 2014. The spacecraft consists of an orbiter, which features twelve instruments, and the 108 kg Philae robotic lander with an additional nine instruments. The spacecraft carries a bi-propellant propulsion system capable of providing both attitude control and orbit control and maneuvering; the propellant load is 1670 kg. The system consists of 24 thrusters, each producing 10 N of thrust and running on monomethyl hydrazine as fuel and nitrogen tetroxide as oxidizer. On May 10, 2004, a deep-space maneuver was carried out to bring the spacecraft on a course in which it would perform a swingby about the Earth. Four of the spacecraft's thrusters were fired for more than 3.5 hr, to achieve a velocity change of 153 m/s. Two more small trajectory corrections were executed in November 2004 and February 2005. The swingby about the Earth occurred on March 4, 2005. At closest approach, Rosetta passed above the Pacific Ocean just west of Mexico at an altitude of 1954 km and a velocity relative to the Earth of 10.6 km/s. On July 4, 2005, the imaging instruments on board observed the collision between comet Temple 1 and the impactor of the NASA Deep Impact spacecraft. At that time Rosetta was about 0.53 AU (79×10^6 km) from the comet; the comet was then about 0.89 AU (133×10^6 km) from the Earth. On February 25, 2007, the spacecraft made a swingby maneuver about Mars, passing before the planet at a minimum altitude of 250 km and reducing the spacecraft's velocity by about 2.2 km/s, relative to the Sun, such that a second Earth swingby maneuver would occur. On November 13, 2007, Rosetta performed this flyby. It passed the Earth at a minimum altitude of 5301 km above a region of the Pacific Ocean, just off the coast of Chile; its velocity relative to Earth was about 12.5 km/s. Rosetta passed asteroid 2867/Steins at a minimum distance of 803 km and a relative velocity of about 8.6 km/s on September 5, 2008. A third and final swingby maneuver about the Earth has taken place on November 13, 2009, when Rosetta passed the Earth at a minimum altitude of 2481 km above the Indian Ocean just south of the island of Java. At that time, the spacecraft had a velocity of about 13.3 km/s relative to the Earth; the swingby maneuver provided a velocity increase of 3.6 km/s. On July 10, 2010, when the spacecraft was 454×10^6 km from Earth, it flew past the asteroid 21/Lutetia at a minimum distance of 3,168 km and a relative velocity of about 15 km/s. In the period January 18-25, 2011, Rosetta's four 10 N thrusters were ignited during five thrusting periods to produce a ΔV of 0.79 km/s. This major deep-space maneuver placed the spacecraft onto an interception trajectory with comet Churyumov-Gerasimenko. On June 8, 2011, when the spacecraft was 549×10^6 km from the Earth and about 670×10^6 km from the Sun, the spacecraft was commanded to put itself into a hibernation mode during which all instruments and almost all control systems were turned off; only the computer and several heaters remained active. On January 20, 2014, when Rosetta's orbit has brought the spacecraft back to within 673×10^6 km from the Sun, the solar radiation intensity was sufficient again to power the spacecraft fully and a timer has woken up the spacecraft. After warming up its key navigation instruments, coming out of a stabilizing spin, and aiming its main radio

antenna at Earth, Rosetta sent a signal to let mission operators know it had survived the most distant part of its journey. From May 7 to August 6, 2014, a series of ten maneuvers was executed in order to reduce the distance and relative velocity between Rosetta and the target comet. Three major maneuvers were executed on May 21, June 4 and June 18. On May 21, when the spacecraft was at a distance of about 10^6 km from the comet and had a velocity of 754 m/s relative to the comet, four 10 N thrusters were switched on for 7^h21^m to reduce the speed relative to the comet by 290 m/s; on June 4 the thrusters were fired for 6^h45^m to reduce the speed by 270 m/s; on June 18 the thrusters were fired for 2^h20^m to further reduce the speed by 91 m/s. During these maneuvers about 215 kg, 190 kg and 65 kg of propellant was used. The last of the series of ten maneuvers was executed on August 6, when the spacecraft was at a distance of about 100 km from the comet. After the thrusters were fired for 6^m26^s , the relative velocity of the spacecraft was reduced to less than 1 m/s. In the next weeks a number of small maneuvers were executed until the spacecraft entered an orbit about the comet. A detailed mapping of the comet's surface was conducted and a landing site was selected for Philae. The lander was dropped from an altitude of about 22 km on November 12, and touched the nucleus about 7 hr later at a speed of the order of 1 m/s. Because the anchoring harpoons did not fire, Philae bounced off the surface and flew up to an altitude of about 1 km. About 2 hr after first contact, it again touched down. After a second bounce, Philae finally landed about 7 min later, becoming the first spacecraft to soft-land on a comet. However, at that site there was insufficient illumination of the lander's solar panels. Consequently, the lander ran out of battery power about 58 hr after the landing. During this period it has sent high-resolution pictures as well as information about the upper crust of the nucleus to the orbiter; these data were relayed to Earth by the orbiter. The orbiter will continue observations of the comet's nucleus, and will monitor the 'awakening' of the comet's activity as it comes closer to the Sun and reaches the perihelion of its orbit in October 2015.

18.12. Non-Hohmann low-energy transfer trajectories

In the previous Sections, it was assumed that interplanetary trajectories may, to first-order approximation, be computed by the *patched conic approach* (Section 17.3). It consists in dividing the trajectory in phases and studying the motion as a two-body problem during each phase. These phases are separated at a planet's sphere of influence, which yields some discontinuities in the model. The conic arcs resulting by this analysis are then patched together to obtain an approximation of the whole transfer trajectory. This simplification was used for designing the trajectories for many interplanetary missions, including e.g. NASA's multiple-planet flyby missions, such as Voyager and Galileo. The trajectory computed in this way could be used as an initial guess for a numerical procedure that produced the fully-integrated n -body system solution. But as space missions have become more demanding, other approaches have become necessary. For example, the ISEE-3, Genesis, and outer-planet multi-moon orbiter trajectories resemble solutions of the restricted three- and four-body problems much more than of two-body problems. In fact, methods based on a patched-conic approximation would be inappropriate to find these complicated trajectories, and a three-body decomposition of the solar system is absolutely necessary.

In the *patched three-body approximation* we consider a restricted four-body problem wherein a spacecraft moves under the influence of three massive bodies whose motion is prescribed. For planetary missions, these bodies are Sun, Earth, planet and spacecraft, or Sun, planet, moon and spacecraft. When the spacecraft is relatively close to one of the planets or moons, the motion is governed by the three-body dynamics of the corresponding Sun-planet system or planet-moon system, and the full trajectory is obtained by patching together the three-body trajectories. This

approach still considers some model discontinuities along the whole trajectory, but the dynamics are modeled more accurately during the different mission phases. One of the key ideas of this analysis approach is that the competing gravitational attractions between celestial bodies creates a vast array of passageways that wind around the Sun, planets and moons (Section 3.12). The boundaries of these passageways are realized geometrically as invariant manifolds attached to the collinear Lagrange points and periodic orbits in interlinked three-body problems. The success of this approach depends greatly on the configuration of the four bodies of interest. To enable low-energy transfers, the manifold tubes of the two three-body systems must intersect within a reasonable time; otherwise, the transfer may require an impractically long time of flight. For the Sun-Earth-Moon-spacecraft case, this is not a problem. The overlap of these tube structures provides many opportunities for low-energy transfers between the Earth and the Moon (Section 17.5). For interplanetary trajectories, this is, however, not the case and we sometimes have to wait for a long time before the rotating tube structures, which form an interplanetary transport network, intersect.

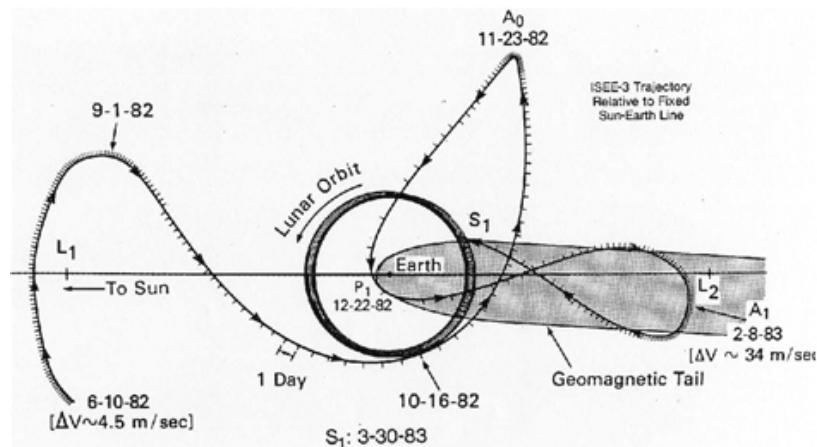


Figure 18.28: Trajectory of ISEE-3 after it left its halo trajectory about the Sun-Earth L_1 point on June 10, 1982, plotted relative to the rotating Sun-Earth reference frame and viewed from the ecliptic north.

The first mission that has used this kind of low-energy non-Hohmann transfer trajectories was ISEE-3. In Section 3.13, the first part of this mission, where the spacecraft was inserted into a halo orbit about the Sun-Earth L_1 point, was discussed. Here, a short description of the extended mission is presented. In 1982, ISEE-3 began the magnetotail and comet encounter phases of its mission. A maneuver was conducted on June 10, 1982, to remove the spacecraft from its halo orbit and to place it on an unstable Sun-Earth manifold transfer trajectory involving a series of passages between the Earth and the Sun-Earth L_2 point (Figure 18.28). After several passes through the Earth's magnetotail, with gravity assists from lunar flybys on March 30, April 23, September 27 and October 21, 1983, a final close lunar flyby (119.4 km altitude) on December 22, 1983, ejected the spacecraft out of the Earth-Moon system and into a heliocentric orbit ahead of the Earth, on a trajectory intercepting that of comet Giacobini-Zinner. At that time, the spacecraft was renamed International Cometary Explorer (ICE). The primary scientific objective of ICE was to study the interaction between the solar wind and a cometary atmosphere. In 1985, three trajectory correction maneuvers were performed totaling a ΔV of 42 m/s, and the spacecraft traversed the plasma tail of the comet on September 11, 1985, at a distance of about 7800 km and made *in situ* measurements of particles, fields, and waves. It also transited between the Sun and comet Halley in late March 1986, when other spacecraft (Giotto, Planet-A, Sakigake, VEGA-1, VEGA-2) were in the vicinity of comet Halley on their early March comet rendez-vous missions.

In 1991, NASA defined a heliospheric mission extension for ICE, consisting of investigations of solar coronal mass ejections in coordination with ground-based observations, continued cosmic ray studies, and special period observations such as when ICE and Ulysses were on the same solar radial line. As of January 1990, ICE is in a 355 days period heliocentric orbit with apohelion distance of 1.03 AU, perihelion distance of 0.93 AU, and inclination of 0.1° . This will bring it back to the vicinity of the Earth-Moon system in August, 2014. On May 5, 1997, operation of ICE was terminated.

Another mission that has used this type of trajectories was Genesis, which was launched by a Delta II 7326 rocket from Cape Canaveral on August 8, 2001, and has swept up specks of the Sun—individual atoms of the solar wind—on five collector arrays and in an ion concentrator. The goal was to return these solar wind particles safely to Earth for study into the origins of the

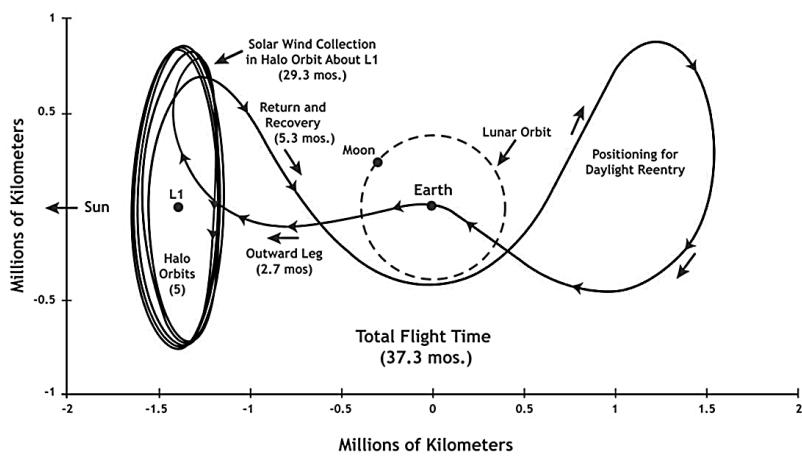


Figure 18.29: Trajectory of Genesis from Earth to the Sun-Earth L_1 point, along a halo orbit about this point, to the region of the Sun-Earth L_2 point, and back to Earth, plotted relative to the rotating Sun-Earth reference frame and viewed from the ecliptic north.

solar system. Its sample-return capsule arrived back on Earth on September 8, 2004. This sample would be the first extraterrestrial material brought back to Earth from beyond the Moon's orbit. For this spacecraft a mission was designed that used non-Hohmann low-energy transfer trajectories. Figure 18.29 shows a sketch of the Genesis trajectory, projected onto the ecliptic and relative to the rotating Sun-Earth reference frame. The spacecraft was launched along a low-energy trajectory to the Sun-Earth L_1 point. On November 16, 2001, it entered a halo orbit about this L_1 point and completed five orbits around this point collecting samples for about two years. The y -amplitude (perpendicular to the Sun-Earth line) of this halo orbit was about 780,000 km; this is about twice the radius of the orbit of the Moon. On April 22, 2004, the spacecraft fired its thrusters to leave this orbit and to begin its journey back toward its home planet. Because of the position of the landing site—the U.S. Air Force Utah Testing and Training Range—and the unique geometry of Genesis' flight path, the spacecraft could not make a direct approach and still make a daytime landing. In order to allow the Genesis helicopter crews an opportunity to capture the return capsule in daylight in the air, Genesis mission navigators designed an orbital detour along a low-energy trajectory toward the Sun-Earth L_2 point. After completing one loop around this point, the spacecraft was set up for a daytime return to Earth on September 8. In dynamical systems theory, this motion between the L_1 and L_2 points is called a *heteroclinic connection* between the L_1 and L_2 regions. The stable manifold, which winds onto the halo orbit about the L_1 point, was used to design the transfer trajectory of the Genesis spacecraft from launch to

insertion into this halo orbit. The unstable manifold, which winds off of the halo orbit, was used to design the return trajectory which brought the spacecraft and its precious samples back to Earth via the nearly-heteroclinic connection. However, when the spacecraft arrived on September 8, 2004, at the Earth and the return capsule entered the Earth's atmosphere, the parachute failed to open and the capsule slammed into the Utah desert. Still, some science objectives of the mission could be met by analyzing the interior of the heavenly damaged capsule. For instance, it was found that the Sun is relatively richer than Earth in the isotope oxygen-16.

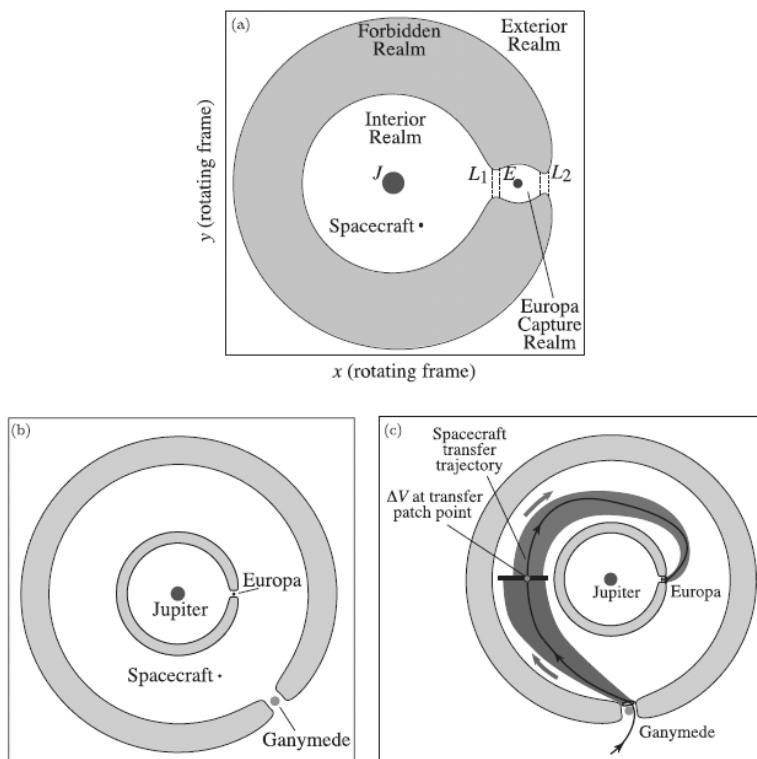


Figure 18.30: Non-Hohmann low-energy transfer along the Jovian moons Ganymede and Europa. (a) The accessible region (schematic) of the Jupiter-Europa three-body system. (b) The four-body system approximated as two nested three-body systems. c) Intersection (schematic) between the tube associated with Ganymede's L_1 periodic orbit unstable manifold and the tube associated with Europa's L_2 periodic orbit stable manifold. [copied from: G. Gómez et al., Connecting orbits and invariant manifolds in the spatial restricted three-body problem, Nonlinearity, Vol. 17, 2004.]

Using low-energy passageways is in no way limited to the inner solar system. For example, consider a spacecraft in the gravity field of Jupiter and its many moons. It is then possible to design a mission through Jupiter's many-moon system, in which a single spacecraft orbits subsequently several moons for any desired duration, allowing long-duration observations instead of flybys lasting only minutes. For example, a multi-moon orbiter could orbit each of the Galilean moons—Callisto, Ganymede, Europa, and Io—one after the other, using a reasonable amount of propellant. Figure 18.30 shows a low-energy non-Hohmann transfer trajectory from an initial Jovian insertion trajectory to Ganymede. After one orbit around Ganymede including a close approach, the spacecraft heads onward to Europa, ending in a high-inclination orbit around this icy moon. Figure 18.30^a schematically shows the region in the Jupiter-Europa system where the spacecraft's motion is possible for a particular value of Jacobi's constant ((3.53)); note the opening near Europa, permitting capture and escape. Figure 18.30^b shows the four-body system

approximated as two nested three-body systems; this picture is only a schematic, as the space-craft's motion conserves Jacobi's integral in only one system at a time. Figure 18.30^c schematically shows the intersection between the tube associated with Ganymede's L_1 periodic orbit unstable manifold and the tube corresponding to Europa's L_2 periodic orbit stable manifold. Integration forward and backward from the patch point (with a ΔV to take into account a velocity discontinuity) yields the desired transfer trajectory between the moons Europa and Ganymede. It was found that for this tour from Ganymede to Europa, the propellant consumption for impulsive burns, as measured by the total ΔV , to be less than half the corresponding Hohmann transfer value. This kind of cleverly designed low-energy transfers will undoubtedly be applied for many more future interplanetary missions.

19. LOW-THRUST TRAJECTORIES

In the previous Chapters we have discussed spacecraft trajectories that are shaped by gravitational forces. If, for a short period of time, a rocket thrust was acting on the spacecraft, we assumed that this powered phase could be considered as an impulsive shot. According to this approximation, the position of the spacecraft does not change during the powered phase and the spacecraft only experiences a discontinuous change in velocity. This approximation is admissible for propulsion systems with a relatively short thrusting time. This is the case for the *chemical rocket propulsion system* that was and is applied for all launch vehicles and for most maneuvers in space. For this type of systems, characteristic values for a powered phase in an orbit about the Earth are: $F/M_0 = 2 - 200 \text{ m/s}^2$, $t_b = 0.1 - 20 \text{ min}$, where F is the thrust, M_0 the initial mass of the spacecraft, and t_b the burning time of the rocket engine. To limit the amount of propellants needed for a certain maneuver, the rocket engine exhaust velocity should be as high as possible. The highest-performance propellants combination that is used today is liquid hydrogen and liquid oxygen, which yields maximum exhaust velocities of up to 4.5 km/s when high-pressure combustion chambers are used. Theoretically, maximum exhaust velocities of up to 6 km/s are possible for exotic chemical propellant combinations.

With a *thermal solid-core nuclear-fission propulsion system*, where the heat produced by a nuclear-fission reactor is transferred to the hydrogen propellant via a heat exchanger, higher exhaust velocities of up to 10 km/s are achievable. However, the additional mass of the nuclear reactor reduces the advantage gained from the higher exhaust velocity. During the period 1959-1968, in the USA ground-based experiments have been performed with solid-core nuclear-fission reactors for rocket engines and with complete nuclear engines, but they were never launched. The first nuclear rocket reactor test was conducted in the Nevada desert in July 1959. This test reached a maximum temperature of 2683 K and a thermal power level of 70 MW. Subsequently, a number of other tests have been performed. The most powerful reactor for a nuclear rocket engine ever constructed, with a design thermal power level of 5000 MW, was tested in June 1968. It was run at a level of 4000 MW for 12.5 min at temperatures of up to 2310 K. The last tests with a complete nuclear rocket engine were conducted in March 1968 with a 1100 MW reactor that accumulated a total of 28 start cycles for a total of 115 min of operations. In the series of tests with this engine, thrust levels of up to 930 kN and exhaust velocities of up to 8.2 km/s have been reached; at the highest thrust level a minimum reactor specific mass of 2.3 kg/MW was achieved. The US nuclear rocket engine development program was terminated in 1972. Since then, only design studies and technology developments have taken place. At present, NASA is studying various advanced solid-core nuclear fission engine concepts, including a 75 MW engine with a thrust of 14 kN and a total mass of 170 kg. This system would heat the hydrogen propellant to 3000 K, leading to an exhaust velocity of 9 km/s. Also, the design of a gas-core nuclear-fission reactor is studied, where radiant energy is transferred from a high-temperature fissioning plasma to the hydrogen propellant. In this concept, the propellant temperature can be significantly higher than the engine structural temperature, yielding exhaust velocities of 30 km/s or higher. In general, with thermal nuclear-fission propulsion systems somewhat longer thrusting periods are applied (10 - 40 min), but these are still short compared to the total flight time. As a result, the concept of the impulsive shot is still applicable to first-order trajectory analyses.

A completely different type of rocket propulsion is *electric propulsion*, where the thrust is very low but the exhaust velocity can reach almost unlimited values. Although this Chapter does not cover the concepts and technologies of these propulsion systems but only the trajectories of electrically propelled spacecraft, we will start with a survey of some characteristic aspects of this kind of propulsion system.

19.1. Electric propulsion systems and missions

Unlike a chemical propulsion system, where the propellants are both the power source and the exhaust medium, an electric propulsion system includes a separate power source and power conversion unit to impart the propulsive energy to the propellant. This power plant may comprise a solar array or a solid-core nuclear-fission reactor. Until now, solar arrays were and are the power source for nearly all electrically propelled spacecraft, but in the future, when higher power levels will be required and regions of space far from the Sun will be traversed, electric propulsion systems will be powered by nuclear-fission reactors.

Nearly all current electric propulsion systems can be grouped into three categories:

- **Electrothermal propulsion.** These systems include *resistojets* and *arcjets*, where the propellant is heated electrically and the hot gas is thermodynamically expanded and accelerated through an exhaust nozzle. Resistojets operate by electrically heating high-resistance metal parts, which heat the propellant flowing over them. Arcjets operate by heating the propellant gas flow by passing it directly through an electric arc discharge.
- **Electrostatic propulsion.** These systems include the *Kaufman thruster*, the *Radiofrequency Ion Thruster* (RIT) and the *Field Emission Electric Propulsion* (FEEP) *thruster*. In the Kaufman and RIT thrusters, the propellant atoms are ionized by electron bombardment. These energetic electrons can be provided from a hot cathode filament and accelerated in the electrical field of the cathode fall to the anode (Kaufman thruster), or can be produced and accelerated by the oscillating electrical field induced by an alternating magnetic field of a coil, which results in a self-sustaining discharge and eliminates the need for a cathode (RIT thruster). In the FEEP thruster, the propellant passes an extremely thin slit in an emitter. An electrical field applied between the emitter and an accelerator ring causes the propellant atoms to become unstable and to ionize. In all three systems, the positive ions are accelerated to high velocity by an electrical field established between the ion source and one or more accelerating electrode(s). After acceleration, the ion beam is electrically neutralized by injecting a stream of electrons into the exhaust beam.
- **Electromagnetic propulsion.** In these systems a highly-ionized propellant plasma is generated and the ions are accelerated by the interaction of electrical and magnetic fields on the plasma. After the acceleration, the ions pick up the electrons to neutralize the beam after the thruster. Stationary plasma thrusters (*Hall thrusters*) use the Hall effect to set up an electrostatic field, which accelerates the propellant ions. *Magnetoplasmodynamic thrusters* use an electric arc discharge, like an arcjet. In this case, the propellant plasma is accelerated by the interaction of the arc and self-induced and applied magnetic fields. *Pulsed plasma thrusters* accelerate the propellant plasma by interaction of an electric arc current with a self-induced magnetic field. *Pulsed inductive thrusters* accelerate the propellant plasma by the interaction of a generalized azimuthal current with a magnetic field from a coil current.

Usually, the Kaufman, RIT, FEEP and Hall thrusters are referred to as members of the family of *ion propulsion systems*. The Hall truster was studied independently in the USA and the former USSR in the 1950s and 1960s. However, the concept was only developed into an efficient propulsion device in the USSR, whereas in the USA scientists focused instead on developing the Kaufman thruster. The RIT and FEEP thrusters were primarily developed in Japan and Europe. Electrothermal propulsion is mainly used for attitude control and stationkeeping of satellites, and to a limited extent for orbit change maneuvers. Up to now, ion propulsion is primarily used for orbit maintenance, correction and transfer maneuvers and for planetary missions, because of the high exhaust velocities achievable, the good exhaust power over mass ratio, and the maturity of these systems. Therefore, in this Section we will focus on ion propulsion systems; in the next

Sections the motion of a spacecraft with a general low-thrust electric propulsion system will be analyzed and discussed.

In ion propulsion systems, electrical power that is generated by solar panels, or by a nuclear reactor and its associated electrical generator, is used to ionize the propellant and to subsequently accelerate the ions by an electrostatic field. So, the electrical energy is converted to directed kinetic energy of the exhausted particles. From the law of conservation of energy and the expression for the thrust of a rocket engine: $F = \dot{m}V_j$ ((1.10)), we find that for this propulsion system the following relation holds:

$$P_j = \epsilon P = \frac{1}{2} \dot{m} V_j^2 = \frac{1}{2} F V_j \quad (19.1)$$

where P is the electrical power that is generated by the solar panels or nuclear reactor and provided to the propulsion system, P_j is the power of the exhaust jet, ϵ is the power conversion efficiency, \dot{m} is the exhausted mass per unit of time (mass flow), V_j is the (effective) exhaust velocity and F is the thrust. For a given engineering design of the propulsion system, ϵ is primarily a function of the exhaust velocity; its value generally drops to low values for low values of the exhaust velocity. With present-day technology, a value of $\epsilon = 0.7$ is achieved for large ion engines. It is emphasized that an electric propulsion system is not limited by a maximum exhaust velocity, but by the maximum power that the powerplant can provide. At this maximum power, the exhaust velocity can be chosen freely. However, according to (19.1), an increase in exhaust velocity means a decrease in thrust. In the past, mercury and cesium were primarily used as propellant for ion propulsion systems, but mercury is toxic and tended to contaminate the spacecraft with the metal, and cesium is very toxic and corrosive. Many current designs use xenon as propellant because of its low ionization energy, reasonably high atomic mass and inert nature. For tests, the most practical propellant is argon, which is a relatively abundant and cheap gas. Other propellants for application in future ion engines are presently being studied.

From (19.1) we find, after elimination of V_j and substitution of the (instantaneous) *thrust acceleration*, $f = F/M$,

$$\frac{\dot{m}}{M^2} = \frac{f^2}{2\epsilon P}$$

where M is the (instantaneous) mass of the spacecraft. With $\dot{m} = -dM/dt$, this equation can be written as

$$\frac{1}{M^2} \frac{dM}{dt} = - \frac{f^2}{2\epsilon P}$$

which, when integrated from the time of engine ignition, t_0 , to the time that the engine stops thrusting, t_e , leads to

$$\frac{1}{M_e} - \frac{1}{M_0} = \frac{1}{2} \int_{t_0}^{t_e} \frac{f^2}{\epsilon P} dt \quad (19.2)$$

The initial mass of an electrically propelled spacecraft, M_0 , may be assumed to consist of four components: the mass of the construction (including the rocket engines and propellant tanks), M_c , the mass of the payload, M_l , the mass of the powerplant, M_{pp} , and the mass of the propellants, M_p . So, $M_0 = M_c + M_l + M_{pp} + M_p$. For the mass at t_e , when all propellants have been consumed, we

may write $M_e = M_c + M_l + M_{pp}$. From these expressions we obtain

$$\frac{1}{M_e} - \frac{1}{M_0} = \frac{1}{M_0} \left(\frac{M_0}{M_e} - 1 \right) = \frac{1}{M_0} \left[\left(\frac{M_l}{M_0} + \frac{M_{pp} + M_c}{M_0} \right)^{-1} - 1 \right] > 1 \quad (19.3)$$

This relation shows that, for given values of M_0 , M_{pp}/M_0 and M_c/M_0 , a maximum payload ratio, M_l/M_0 , corresponds to a minimum value of $(1/M_e - 1/M_0)$, which requires a minimum value of the integral on the right-hand side of (19.2). Electric propulsion systems operate most efficiently at a constant electrical power, P . For planetary missions using solar panels as power source, the power produced will vary with the distance from the Sun. Then we have two alternatives: 1) let the electric propulsion system operate at a variable power level, or 2) let the electric propulsion system operate at the power level corresponding to the largest distance from the Sun reached during the mission and dissipate the excess solar panel power generated at shorter distances from the Sun. We will not analyze this aspect further in this Chapter. If we assume that M_{pp}/M_0 , M_c/M_0 and P are specified, then we conclude from (19.2) and (19.3) that the payload ratio, M_l/M_0 , is maximum if

$$\int_0^{t_e} \frac{f^2}{\epsilon} dt \quad \text{is minimum} \quad (19.4)$$

The value of this integral depends upon the flight time, the mission involved, by which is meant the specification of the kinematic conditions of the spacecraft at t_0 and $t = t_e$ (and at any other time which might be necessary), the force field in which the spacecraft moves, the thrust program applied, and, finally, the engineering design of the system as characterized by the quantity ϵ . Comparison of (19.4) with (1.48) shows that for a chemical or thermal-nuclear propulsion system the integral to be minimized contains the thrust acceleration to the first power, while for an electric propulsion system the corresponding integral contains the thrust acceleration to the second power. To find the optimum thrust program (magnitude and direction of the thrust), we have to integrate the equations of motion for specified initial conditions and required final conditions (including the flight time), and use (19.4) as a minimization criterion. In this book, we will not address the computation of such optimum trajectories.

The computation of the optimum thrust program will generally result in a large variation of the magnitude of the thrust, and consequently of the exhaust velocity, along the trajectory. Such variation is beyond the capability of most thruster concepts, and may introduce unwanted system complexity in those systems that can be designed to change exhaust velocity and thrust. Therefore, the next best option is to use a system with constant power and constant magnitude of the thrust, and consequently constant exhaust velocity. The computation of such trajectories is even a greater numerical challenge than the variable thrust trajectories, because now *coasting arcs* have to be introduced to the analysis. For a given mission and a specified constant thrust level, there exists an optimum exhaust velocity and consequently optimum power level that results in a maximum payload ratio M_l/M_0 . This may be explained as follows. From the perspective of propellant mass consumption, a high exhaust velocity would be preferable. However, according to (19.1), an increase of the exhaust velocity leads, for a given thrust level, to an increase of the required electrical power level. The mass of the powerplant will depend on the electrical power that it has to produce, and we may generally write $M_{pp} = \alpha P$. So, a higher electrical power means a larger mass of the powerplant and consequently a lower payload mass for the same spacecraft initial mass. These two opposite effects lead to an optimum exhaust velocity for which the sum of propellant mass and powerplant mass is a minimum. For state-of-

the-art large solar panels at the distance of the Earth from the Sun: $\alpha = 5.5 \text{ kg/kW}$ can be achieved initially, i.e. before degradation due to solar radiation and high-energy particles has occurred. For missions that require a nuclear powerplant, the critical issue is the development of a suitable nuclear reactor that yields $\alpha < 4 \text{ kg/kW}$. With present-day nuclear reactor technology, values of $\alpha = 6.5 \text{ kg/kW}$ seem realizable. To limit the flight time, future electrically propelled spacecraft will fly with powerplants producing 2 - 10 MW; for a value of $\alpha = 4 \text{ kg/kW}$ this leads to a powerplant mass of 8 - 40 ton.

Optimization studies for future large planetary missions generally yield optimum values of the thrust, exhaust velocity, initial thrust acceleration, and propellant consumption within the ranges $F = 1 - 100 \text{ N}$, $V_e = 25 - 100 \text{ km/s}$, $f_0 = 10^{-4} - 10^{-2} \text{ m/s}^2$, and $\dot{m} = 0.01 - 4 \text{ g/s}$. This example shows that thrust, exhaust velocity and propellant consumption are of a different order of magnitude than for chemical propulsion systems. Because electric propulsion systems provide an acceleration that is much smaller than the gravitational acceleration on Earth, electrically propelled spacecraft can not leave the Earth's surface, but first have to be injected into a parking orbit about the Earth (or directly into an escape trajectory) by chemical (or nuclear) rockets. After the electric propulsion system is activated in Earth orbit, the spacecraft will gradually move away from its parking orbit. The performance of these propulsion systems, expressed in the payload ratio M_l/M_0 of the spacecraft, can be considerably better than will ever be possible for chemical propulsion, or even for thermal-nuclear propulsion. The main reason is that the low thrust can be maintained for a long period of time.

Table 19.1 presents some information about the ion propulsion system of a number of spacecraft; this information includes the maximum thrust and exhaust velocity per thruster. The first US flight experiment with an ion propulsion system was onboard a suborbital launch on December 18, 1962. Due to a number of electrical problems, engine thrusting was not accomplished in this test. The first successful ion propulsion system flight test was SERT I that was launched by a Scout rocket on July 20, 1964. The spacecraft was equipped with two different ion engines. One engine used mercury as propellant; this ion engine was operated at 1.4 kW electrical power with 28 mN of thrust at an exhaust velocity of 48 km/s. It was successfully operated for 31 min. The other engine used cesium as propellant. It was operated at an electrical power of 0.6 kW and was designed to produce a thrust of 5.6 mN at an exhaust velocity of 79 km/s. However, the engine could not be started in space because of a high-voltage electrical short circuit. Two US additional suborbital test flights were subsequently executed. On April 3, 1965, a SNAP-10A nuclear-fission reactor power system was launched onboard the SNAPSHOT satellite into a 1300 km altitude orbit with an ion engine as secondary payload. The SNAP system operated successfully at an electrical power of 500 W for about 43 days, but the ion engine operated at a thrust of 8.5 mN only for a period of less than 1 hour before being commanded off permanently. The reason was that high-voltage breakdowns in the ion engine induced false horizon sensor signals, which created severe attitude control problems of the spacecraft. The first US successful orbital test of an ion engine was on the ATS-4 spacecraft that was launched on August 10, 1968, and was targeted for a geostationary orbit. The launch was a partial failure and the spacecraft did not reach its nominal orbit; it re-entered the Earth's atmosphere on October 17, 1968. The spacecraft carried two ion engines that were each designed to operate at 20 W and provide 0.09 mN at an exhaust velocity of 66 km/s. Both engines operated according to specifications for about 10 hours over a 55 days period. The first long-term operation of an ion engine was accomplished on the SERT II spacecraft that was launched on February 3, 1970, into a 1000 km altitude orbit. The spacecraft carried two identical ion engines. Each thruster was designed to operate at a maximum electrical power level of 850 W, and to provide 28 mN thrust at an exhaust velocity of 41 km/s.

Table 19.1: Ion propulsion system data for a number of spacecraft.

Spacecraft	Sponsor	Launch date	Orbit altitude (km)	Propellant	Thrust (mN)	Exh. velocity (km/s)
SERT I (2)	NASA	20/07/64	suborbital	mercury	28	48
				cesium	5.6	79
Program 661A, B(1)	USAF	29/08/64	suborbital	cesium	8.9	73
SNAPSHOT (1)	USAF	03/04/65	1300	cesium	8.5	50
ATS-5 (2)	USAF/NASA	12/08/69	36,000	cesium	0.089	66
SERT II (2)	NASA	03/02/70	1000	mercury	28	41
ATS-6 (2)	NASA	30/05/74	36,000	cesium	4.5	25
SCATHA P78-2 (1)	USAF/NASA	30/01/79	27.43×10^3	xenon	0.14	3.4
ETS-3 (2)	JAXA	03/09/82	1100	mercury	2	21
EURECA-1 (1)	ESA	31/07/92	510	xenon	10	30
ETS-6 (4)	JAXA	28/08/94	8.39×10^3	xenon	24	28
PAS-5 (4)	PanAmSat	27/08/97	36,000	xenon	18	25
ASTRA-2A (4)	SES	29/08/98	36,000	xenon	18	25
Deep Space 1 (1)	NASA	24/10/98	orbits Sun	xenon	92	30
PAS-6B (4)	PanAmSat	21/12/98	36,000	xenon	18	25
XM-2 Rock (4)	XM	18/03/01	36,000	xenon	165	37
ARTEMIS (4)	ESA	12/07/01	36,000	xenon	15	34
				xenon	18	33
Hayabusa (4)	JAXA	09/05/03	orbits Sun	xenon	8	31
SMART-1 (1)	ESA	27/09/03	to Moon	xenon	68	16
Dawn (3)	NASA	27/09/07	orbits Sun	xenon	92	30
GOCE (2)	ESA	17/03/09	270	xenon	20	40

Data listed are per thruster and correspond to maximum values; (..) number of thrusters; SES: SES Global S.A.; XM: XM Satellite Radio Inc.

The engines were operated for 2011 hours and 3781 hours, respectively.

Between the early 1970s and 2005, more than 120 satellites with electric propulsion have been launched by the former USSR. For these satellites, electric propulsion was used for attitude control, stationkeeping and orbit correction. Many American military satellites have used and still use electric propulsion systems for attitude control, stationkeeping and orbit correction purposes. Electric propulsion entered its era of commercial application in the early 1980s. In the early 1990s, electric propulsion was adopted for stationkeeping of many communications satellites in geostationary orbit. In 1997, a new era of ion propulsion began with the deployment of communications satellites using ion propulsion for auxiliary propulsion. PanAmSat's PAS-5 communications satellite, launched on August 27, 1997, was the first successful operational geostationary communications satellite employing ion propulsion. Since then, many more communications satellites using ion propulsion systems have followed. XM-2 Rock, owned by XM Satellite Radio and launched on March 18, 2001, is one out of the series of very-large Boeing model 702 communications satellites. The solar array of these satellites provide a beginning-of-life electrical power of 15 kW or more. The satellite flies four ion thrusters, each operating at 2.2 or 4.5 kW and producing 79 or 165 mN of thrust, to perform all stationkeeping and spacecraft momentum control. Only two of the four thrusters are required to perform the entire on-orbit mission maneuvers with a series of four daily burns. For these communications satellites, the electric propulsion system is also used in the geostationary transfer trajectory for apogee raising maneuvers, for decreasing the orbital inclination, to circularize the orbit and to position the satellite in its final orbit. In this way, larger payloads can be accommodated with

greater flexibility in the choice and use of a launch vehicle. In total, the number of electrically propelled spacecraft has gone from single digits in the 1960s to double digits in the 1970s and 1980s and has reached the triple-digit mark in the late 1990s. Many more electric propulsion systems on many more communications satellites are planned. These advanced commercial thrusters operate in the 0.4-5 kW range at exhaust velocities of 20-40 km/s and have typical required service life of 5000 to 15,000 hours.

A unique application of electric propulsion was the ESA GOCE satellite, which was launched by a Russian Rockot from Plesetsk on March 17, 2009, into a near-circular orbit at a mean altitude of 284 km and an inclination of 96.7°. The satellite, with an initial total mass of 1050 kg, carried a gradiometer that measured the gradient of the Earth's gravity field. From these data the spatial and temporal variations of the gravity field can be determined with extreme accuracy. To obtain the strongest possible gravity field signal, the orbit of the satellite should be as low as possible. However, then the atmosphere produces a deceleration of the satellite that would disturb the functioning of the gradiometer. Therefore, the satellite was equipped with a throttleable ion propulsion system, which was powered by solar panels on the satellite body producing 1.3 kW of electrical power. About 600 W was available to the ion propulsion system. This system consisted of two redundant Kaufman type of ion engines with 40 kg of xenon as propellant, generating an exhaust velocity of 40 km/s, and producing a thrust of 1 - 20 mN adjustable in steps of as low as 12 μ N; only one ion engine operated at a time. The accelerometers of the gradiometer measured the deceleration of the satellite due to the atmospheric drag and a control system produced continuously the right amount of thrust to achieve a 'drag-free' trajectory. In this way, the gradiometer measures the 'clean' signal of the Earth's gravity field. The satellite's arrow shape and aerodynamic fins helped keeping GOCE stable as it flew through the upper atmosphere. In its initial orbit, the altitude decayed by about 190 m per day. During the commissioning and calibration phase, both ion engines were ignited for a short time on March 31 and April 2 to check their performance. On September 13, 2009, the satellite reached its operational altitude of 255 km¹. Then, one of the ion engines was ignited and the drag-free part of the mission started. When a telemetry problem was discovered in July 2010, operators raised GOCE's orbital altitude to 263 km. After the telemetry problem was resolved in September 2010, the mean orbital altitude was reduced again to 255 km. Because of the unusually low solar activity in 2010, GOCE used less propellant for its ion engines than predicted. Therefore, the gravity field mission could be extended beyond the originally planned 2.5 years. Starting on August 1, 2012, maneuvers were executed to lower the GOCE orbit by about 230 m/day. After coming down by 9 km in 40 days, the satellite's performance and new environment were assessed. Starting on November 6, the GOCE orbit was again gradually being lowered and reached a mean altitude of 239 km on December 2 while continuing its gravity mapping. Starting on February 3, 2013, the orbit was further being lowered to reach a mean altitude of 234 km by February 16, and on May 21 again to reach a mean altitude of 224 km by May 29. On October 21, 2013, all xenon propellant had been consumed and the mission came to an end. The orbit altitude then gradually decreased and the satellite re-entered the atmosphere over the South Atlantic Ocean, near the Falkland Islands, on November 11, 2013.

The main promise of electric propulsion systems, however, is in their application as primary propulsion systems for lunar, interplanetary or even interstellar flights. Deep Space 1, launched on October 24, 1998, by a Delta II 7326 Star 37FM rocket from Cape Canaveral into an orbit

¹ Here, altitude is defined as the difference between the orbit's mean semi-major axis (Section 23.3) and the Earth's mean equatorial radius.

about the Sun, was the first deep-space mission with a primary ion propulsion system. The spacecraft had an initial total mass of 486 kg and carried an ion engine that operated over a 0.5–2.3 kW electrical power range, providing thrust from 19 mN to 92 mN; the exhaust velocity ranged from 19 km/s at 0.5 kW to 30 km/s at 2.3 kW. It also carried a hydrazine attitude control and maneuvering system. A total of 82 kg of xenon propellant and 31 kg of hydrazine propellant was loaded for the flight. The spacecraft was equipped with an autonomous navigation system that enabled the spacecraft to determine its relative position by optically tracking two or more asteroids, of which the orbits are known, across the star background. The ion engine was started on November 10, 1998; it operated for about 4.5 min and then switched automatically to a standby mode. It is believed that the unplanned termination of the thrusting was the result of a contamination within the engine. Thermal cycling during the subsequent two weeks removed the contaminant and when the engine was commanded on again on November 24, 1998, it operated as desired and continued thrusting at a moderate level for two weeks uninterrupted. From November 24 through November 30, throttle levels were commanded up in various steps to reach maximum thrust. Then, the engine was shut down and turned back on on December 11. On January 5, 1999, the autonomous navigation system turned off the ion engine, completing the first thrust segment of the Deep Space 1 mission. During that period, the engine accumulated more than 850 operating hours. On March 15, the engine was turned on again and on April 27, Deep Space 1 completed a six-weeks period of thrusting with its ion propulsion system. It took less than 5 kg of xenon to provide the steady push for this six-weeks period, during which the spacecraft's speed was increased by nearly 300 m/s. By June 30, 1999, the ion engine had operated for nearly 1800 hours; the propellant consumption was 11.4 kg, and the total velocity increase about 700 m/s. On July 29, 1999, the spacecraft conducted an encounter with asteroid 9969/Braille at a distance of 1.334 AU from the Sun. This 2 km by 1 km by 1 km asteroid moves about the Sun in an orbit with perihelion and apohelion distances of 1.334 AU and 3.356 AU. Deep Space 1 was intended to perform the flyby with a relative velocity of 15.6 km/s at only 240 m from the asteroid's surface. Due to technical difficulties the spacecraft instead passed Braille at a closest distance of 25 km. On September 22, 2001, it passed comet 19P/Borrelly at a closest distance of 2171 km and with a relative velocity of 16.6 km/s. This comet, 4 km wide and 8 km long, moves in a 1.35 AU by 5.83 AU orbit about the Sun. This flyby returned extremely detailed images of the comet's surface. Despite the absence of a debris shield, the spacecraft survived the comet passage intact. Because of limited NASA resources, it was decided to end the spacecraft operations on December 18, 2001. During the course of the flight, the ion propulsion system accumulated 16,265 hours of operation and expended 73 kg of xenon.

On May 9, 2003, the JAXA Muses-C spacecraft was launched by a Japanese M-5 rocket from Uchinoura into an orbit about the Sun. After launch, the spacecraft was renamed as Hayabusa. In addition to a nitrogen tetroxide and hydrazine propulsion system with a peak thrust of 22 mN for maneuvering, the spacecraft was equipped with an electric propulsion system, consisting of four ion engines of which three were operated together and the fourth was a spare. Each ion engine used xenon propellant, was operated at a power level of 380 W and produced a thrust of 8 mN at an exhaust velocity of 31 km/s. The launch mass was 510 kg, including 50 kg of chemical propellant and 65 kg of xenon. The ion engines started on May 27, 2003; full power operation started on June 25, 2003. On March 31, 2004, the ion thruster operation was stopped to prepare for an Earth swingby flight (Section 18.11). This swingby occurred on May 19, when the spacecraft reached a minimum altitude of 3725 above the Earth. The ion thruster operation was resumed on May 27. The force generated by this almost continuously operating propulsion system gradually changed the orbit such that the spacecraft encountered on September 12, 2005, the asteroid 1998SF36 /Itokawa. This asteroid measures 535 m by 294 m by 209 m and orbits the

Sun in a 0.953 AU by 1.695 AU orbit. On August 28, the spacecraft had switched over from its ion engines to the chemical bi-propellant 22 mN thrusters for maneuvering. The ion engines had then consumed 22 kg of xenon and had operated for 25,800 hours. In the middle of September the spacecraft moved into a stationary position relative to the asteroid at a distance of 20 km from the asteroid. From this stationary position, various measurements were taken on the asteroid. On November 3, the spacecraft took position 3 km from the asteroid. It then began a descent, planned to include delivery of a target marker, and release of the 0.59 kg Minerva mini-lander. Due to a number of anomalies, this maneuver was stopped. In the period November 7 to 12, a series of partial descents was performed. On November 12, the spacecraft began a full descent from 1.4 km altitude at 3 cm/s to an altitude of 55 m. The Minerva lander was deployed but, unfortunately, Hayabusa had at that time already started an automatic ascent; so, the release was at a higher altitude than planned. Contact with the lander was lost and it is believed that Minerva floated off into space without landing. On November 19, Hayabusa landed on the asteroid on 20^h40^m UTC with a velocity of roughly 10 cm/s. Due to this relatively high velocity it bounced up; it bounced again at 21^h10^m UTC and then landed at 21^h30^m UTC. At 21^h58^m UTC, it was commanded to make an emergency ascent. Unfortunately, the sampling of surface material did not progress well, but it was believed that some dust may have whirled up into the sampling horn when it touched the asteroid. On November 25, a second landing attempt was performed. It was first thought that the spacecraft had collected a sample during this attempt, but subsequent data seemed to indicate that it had not. During the fly away after this second touching down on the asteroid, a leak in the chemical propulsion system was detected. The ground team managed to stop the leak, but on December 8 the spacecraft suffered from a significant torque disturbance apparently caused by the out-gassing of vapor generated by the evaporation of leaked propellant. Due to the resulting attitude changes of the spacecraft, ground contact was lost. On January 23, 2006, contact with the spacecraft could be reestablished. On March 6, Hayabusa's position was established at about 1300 km ahead of Itokawa in the orbit of the asteroid with a relative velocity of about 3 m/s. On June 1, 2006, it was confirmed that two out of the four ion engines worked normally, which is sufficient for the return journey. On August 29, 2007, a third ion engine was successfully reignited. On February 4, 2009, JAXA reported that all required trajectory maneuvers were successfully executed. Sufficient xenon propellant was available to bring the spacecraft back to Earth by June 2010, three years later than originally planned. On March 27, 2010, Hayabusa was on a trajectory which would pass the Earth at a distance of 20,000 km. After a series of orbit trajectory maneuvers a trajectory was realized that would lead to a landing of the reentry capsule at the Australian Woomera Test Range. The reentry capsule was separated from the spacecraft on June 13 and about 3 hr later the reentry capsule and the spacecraft reentered the Earth's atmosphere. The heat-shielded capsule made a parachute landing in the Woomera desert, while the spacecraft broke up and incinerated in a large fireball. The capsule was successfully retrieved on June 14; the two parts of the heat shield, which were jettisoned during the descent, were also found. On November 16, 2010, JAXA confirmed that approximately 1500 particles found in one of the two compartments inside the capsule came from Itokawa. Their size was mostly less than 10 μm .

The first ESA spacecraft using electric propulsion as the primary propulsion system was SMART-1, which was launched on September 27, 2003, together with two communications satellites, by an Ariane 5G rocket from Kourou into a geostationary transfer orbit. Details about the trajectory flown by this spacecraft are given in Section 17.5. The 367 kg spacecraft was equipped with a Hall ion engine that was powered by a 1.85 kW (beginning of life) solar array. The engine was designed to run on an electrical power of 1.19 kW at beginning of life and to produce then a nominal thrust of 68 mN at an exhaust velocity of 16 km/s. The spacecraft carried

82.5 kg of xenon propellant. The particular trajectory taken by SMART-1 to the Moon required thrusting for about one-third to one-half of every orbit on the perigee part of the orbit. The ion engine was used over a period of 14 months to elongate the orbit to reach the Moon. On October 26, 2004, the ion engine had accumulated a total on-time of nearly 3,648 hours out of a total flight time of 8,000 hours. At that time, the orbit perigee and apogee distances were 179,718 km and 305,214 km, and the inclination was 20.59° ; the spacecraft had then consumed about 59 kg of xenon. Lunar capture occurred on November 14, 2004, at an altitude of 60,000 km above the lunar surface. The ion engine was powered on again on November 15 for a planned burn of 4.5 days to enter fully into lunar orbit. Subsequently, the ion engine was used to lower the orbit to reach a 300 km by 3000 km lunar polar orbit to conduct lunar science. The ion engine was shut down in September 2005 after exhausting its xenon fuel supply. Then, it had operated for almost 5,000 hours and underwent 843 starts and stops. SMART-1 ended its mission by being deliberately crashed onto the Moon's surface on September 3, 2006.

On September 27, 2007, the NASA 1220 kg (740 kg dry) Dawn spacecraft was launched by a Delta II 7925H rocket from Cape Canaveral. The Delta upper stage boosted the spacecraft and a PAM-D third stage into a 185 km by 6,835 km altitude orbit. About one hour later, the PAM-D fired and released Dawn after accelerating it to 11.46 km/s and sending it into a 1.00 AU by 1.62 AU orbit about the Sun, inclined at 0.5° to the ecliptic. This was the start of an eight-year mission to the asteroid Vesta and the dwarf planet Ceres (Figure 19.1). These bodies are part of the asteroid belt; a region between the orbits of Mars and Jupiter where matter was prevented from coalescing into planets by the gravitational perturbations of Jupiter. The principal dimensions of Vesta are 578 km by 560 km by 458 km; it orbits the Sun with perihelion and apohelion distances of 2.15 AU and 2.57 AU in a plane that is inclined at 7.1° to the ecliptic. Ceres is a dwarf planet, whose mass comprises about one-third of the total mass of all bodies in the asteroid belt. It is a round, smooth world with a mean equatorial radius of 487 km and mean polar radius of 455 km, displaying signs of water ice on its surface. It moves in an orbit with perihelion and apohelion distances of 2.55 AU and 2.99 AU; its orbital plane is inclined at 10.6° to the ecliptic. Dawn carries three science instruments: a visible-light camera, a visible-light and infrared mapping

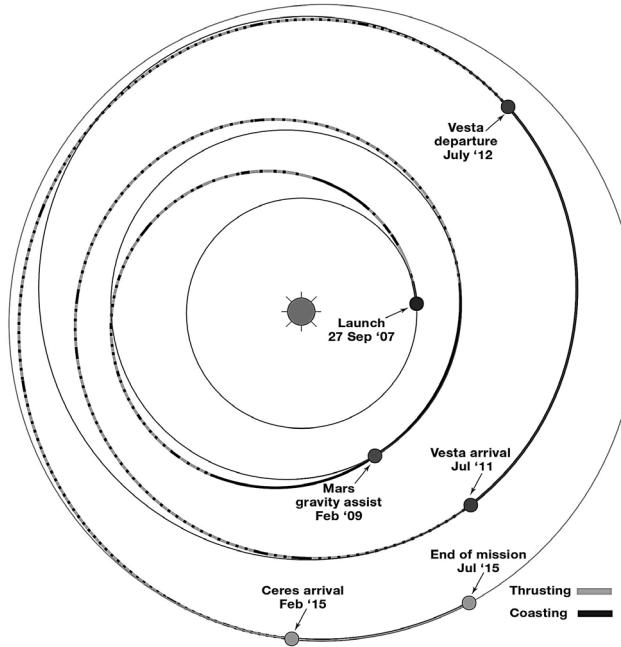


Figure 19.1: Trajectory of the Dawn spacecraft.

spectrometer, and a gamma-ray and neutron spectrometer. In addition to these instruments, radiometric and optical navigation data will provide data about the gravity field and thus of the bulk properties and internal structures of the two bodies. Shortly after launch, Dawn unfolded its two solar panels, each with an area of 18 m^2 and together providing 10.3 kW of electrical power at the start of the mission and 1.3 kW at the end of life at 3 AU . The main propulsion system of the spacecraft consists of three 30 cm diameter ion thrusters; only one engine is thrusting the spacecraft at a time. The thrust of the engines is throttleable between 19 mN and 92 mN , with a corresponding input power of 0.5 kW and 2.6 kW , and an exhaust velocity of 19 km/s and 30 km/s . Dawn carries a total of 425 kg of xenon propellant, which it will use over more than $2,100$ days of thruster operation time. In addition, the spacecraft has a hydrazine reaction control system with two sets of six 0.9 N thrusters and 45 kg of propellant (at start). The ion engines were ignited on October 6, 2007. After initial checkout of all systems and calibrating the thrust levels of the ion engines, during which the engines accumulated more than 11 days of thrust, Dawn began its first long propulsion phase on December 17, 2007. During this first thrusting phase, the ion engine was running for 270 days, or about 85% of the time. In that period it has expended less than 72 kg of xenon propellant. This first electric propulsion phase ended on October 31, 2008, after which the spacecraft coasted to Mars to perform a swingby flight. On November 20, 2008, Dawn performed its first trajectory correction maneuver, firing one of its ion engines for $2 \text{ hr } 11 \text{ min}$. Dawn made its closest approach of 549 km altitude to Mars on February 17, 2009; this swingby increased the spacecraft's heliocentric velocity by 2.6 km/s and changed the orbital inclination by 5.2° . The ion engine was ignited again on June 8, 2009, and, after a flight of about 2 years, Dawn approached Vesta on June 12, 2011. By that time the electric propulsion system had accumulated about $1,100$ days of operation. The ion engine was then used to reduce the spacecraft's velocity relative to Vesta in preparation for its insertion into orbit around Vesta. On July 15, as Dawn continued thrusting, it was gently captured in orbit by Vesta. Note that Figure 19.1 refers to the pre-flight trajectory design and indicates slightly different times for some mission milestones. On August 2, the thrusting needed to reach the circular polar first science observations orbit, at a mean altitude of about 2700 km and with a period of about 69 hr , was completed. From this orbit the entire surface of Vesta, which rotates once in about 5 hr about its polar axis, could be observed within a short period. On August 11, Dawn began its observations program from this orbit. During seven orbital revolutions, the instruments acquired images and spectra when the spacecraft passed over the day side of Vesta and transmitted the results to Earth when the spacecraft was over the night side. On August 31, the ion engine thrusting was resumed, forcing the spacecraft to spiral down to the next science orbit. On September 29, Dawn began its new phase of Vesta observations, circling Vesta every 12.3 hr at an average altitude of 680 km . On October 31, Dawn completed the final mapping cycle in this orbit. It spent another two days transmitting to Earth the last of the science data it had gathered. On November 2, the ion propulsion system began thrusting again and on December 8 Dawn entered the next science orbit, which had an average altitude of 210 km and a period of about 4 hr . Science observations in this orbit began on December 12. After this observation period was completed on May 1, 2012, the ion propulsion system was activated again and Dawn spiraled out to conduct additional observations from a 680 km altitude orbit. The observation mission at Vesta was completed on July 25, 2012, when the propulsion system was ignited again to let the spacecraft spiral out and escape from Vesta's gravity field. It reached the escape velocity at a distance of $17,200 \text{ km}$ and a velocity of 33 m/s relative to Vesta on September 4. Its ion engines will propel it along a trajectory where it will spend nearly three years en route to Ceres, making about three-fourth of an orbital revolution about the Sun as it spirals outward toward the dwarf planet (Figure 19.1). After arrival in February 2015, Dawn will use its ion propulsion system to make a slow approach

and to enter an orbit around Ceres in March 2015. As at Vesta, the spacecraft will fly a series of orbits about the dwarf planet that will allow studying nearly its entire surface. The orbital altitudes will range from 5900 km to 500 km. The mission is planned to end in July 2015. The spacecraft will then be placed in a quarantine orbit about Ceres at an altitude of 700 km. This orbit ensures that the decommissioned spacecraft would not impact Ceres for 50 years, supposedly allowing enough time for action in case some kind of life on Ceres was detected. During the entire mission the three ion propulsion engines will fire cumulatively for about 2,100 days; well in excess of the 678 days for the ion propulsion system of the Deep Space 1 mission.

At present, NASA studies both planetary missions requiring ten to hundred times the power of today's electric propulsion systems and systems providing micro-Newton thrust levels for applications in precision formation-flying missions. For deep-space applications the engine performance goals include a 1 - 7 kW power throttling envelope, an exhaust velocity capability of over 40 km/s and a 70% efficiency. NASA is presently also studying nuclear-electric options for deep-space missions. An advanced ion engine has been designed for a 15 years life, a 20 kW power level and a 75 km/s exhaust velocity. In ground tests, it has been operated at up to 27 kW electrical power at an exhaust velocity of up to 85 km/s.

19.2. Equations of motion

In this Chapter, we will analyze the trajectory of an electrically propelled spacecraft after its departure from a circular initial orbit. This is characteristic for a mission where an electric propulsion system is used to raise the orbit of a spacecraft after it has been injected into a low-Earth parking orbit. In the 1960s and 1970s, one imagined that interplanetary missions with electric propulsion would also start from an initial parking orbit about the Earth. However, up to now all electrically propelled deep space spacecraft were injected by a traditional (chemical) launcher into an escape trajectory, after which the electric propulsion system took over. For those missions we may state that the electric propulsion part of the mission starts, outside the Earth's sphere of influence (Section 4.3), in the Earth's orbit about the Sun. In that case, the velocity of the spacecraft at ignition of the electric propulsion thruster(s) is about tangential to the Earth's orbit, but differs in magnitude somewhat from the Earth's orbital velocity. Nevertheless, the results obtained for a mission that starts in a circular initial orbit about the Sun also provide some insight in the trajectory of a deep space spacecraft that ignites its electric propulsion system in the orbit of the Earth outside the Earth's sphere of influence. Undoubtedly, in the future large interplanetary spacecraft with advanced electric propulsion systems will be assembled in low-Earth orbit, after which they will be propelled out of that orbit and into powered interplanetary transfer trajectories by their electric propulsion system. For these missions, the analyses presented in this Chapter apply very well.

In our analyses we will pay special attention to the phase of the flight between departure from the initial circular orbit and the moment that the spacecraft reaches the local escape velocity. We will assume that the gravity field in which the spacecraft is moving, may be described by a Newton central force field, for which the potential is given by $U = -\mu/r$, and that only this gravitational force and the thrust, \bar{F} , produced by an electric propulsion system act on the spacecraft. Then, the equations of motion may be written as (Chapter 10)

$$\frac{d^2\bar{r}}{dt^2} = -\frac{\mu}{r^3}\bar{r} + \bar{f} \quad (19.5)$$

where $\bar{f} = \bar{F}/M$ and M denotes the instantaneous mass of the spacecraft.

To simplify the analysis, the following dimensionless parameters are introduced:

$$\begin{aligned} \rho &= \frac{r}{r_0} \quad ; \quad \tau = \frac{V_{c_0}}{r_0} t = \sqrt{\frac{\mu}{r_0^3}} t = n_0 t = 2\pi \frac{t}{T_{c_0}} \\ a &= \frac{F/M}{g_0} = \frac{r_0^2}{\mu} f \quad ; \quad w = \frac{V}{V_{c_0}} = \sqrt{\frac{r_0}{\mu}} V \quad ; \quad w_j = \frac{V_j}{V_{c_0}} = \sqrt{\frac{r_0}{\mu}} V_j \end{aligned} \quad (19.6)$$

where the index 0 corresponds to the time that the engine is activated, r_0 is the radius of the initial circular orbit, g_0 is the gravitational acceleration in that orbit ($g_0 = \mu/r_0^2$), V_{c_0} is the velocity in that orbit, n_0 is the mean angular motion in that orbit, and T_{c_0} is the period of that orbit. The local circular velocity and local escape velocity at any point along the spacecraft's trajectory may, according to (6.18), (7.3) and (19.6), be written in dimensionless parameters as

$$w_c = \sqrt{\frac{1}{\rho}} \quad ; \quad w_{esc} = \sqrt{\frac{2}{\rho}} \quad (19.7)$$

With (19.6) we may write for the left-hand side of (19.5):

$$\frac{d^2\bar{r}}{dt^2} = r_0 \left(\frac{d\tau}{dt} \right)^2 \frac{d^2\bar{\rho}}{d\tau^2} = \frac{\mu}{r_0^2} \frac{d^2\bar{\rho}}{d\tau^2} = g_0 \frac{d^2\bar{\rho}}{d\tau^2}$$

Substitution of this expression and (19.6) into (19.5) gives

$$\frac{d^2\bar{\rho}}{d\tau^2} + \frac{1}{\rho^3} \bar{\rho} = \bar{a} \quad (19.8)$$

Vectorial multiplication of (19.8) by $\bar{\rho}$ and scalar multiplication of (19.8) by $d\bar{\rho}/d\tau$ yield

$$\begin{aligned} \bar{\rho} \times \frac{d^2\bar{\rho}}{d\tau^2} &= \bar{\rho} \times \bar{a} \\ \frac{d\bar{\rho}}{d\tau} \cdot \frac{d^2\bar{\rho}}{d\tau^2} + \frac{1}{\rho^3} \frac{d\bar{\rho}}{d\tau} \cdot \bar{\rho} &= \frac{d\bar{\rho}}{d\tau} \cdot \bar{a} \end{aligned} \quad (19.9)$$

Using the method described in Section 5.1, we may write (19.9) as

$$\begin{aligned} \frac{d}{d\tau} (\bar{\rho} \times \bar{w}) &= \frac{d\bar{H}}{d\tau} = \bar{\rho} \times \bar{a} \\ \frac{d}{d\tau} \left(\frac{1}{2} w^2 - \frac{1}{\rho} \right) &= \frac{d\mathcal{E}}{d\tau} = \bar{w} \cdot \bar{a} \end{aligned} \quad (19.10)$$

where $\bar{w} \equiv d\bar{\rho}/d\tau$ is the velocity vector of the spacecraft; \bar{H} and \mathcal{E} are the orbital angular momentum vector and the orbital energy of the spacecraft, both per unit of mass and in dimensionless parameters. We now assume that the thrust vector, \bar{F} , acts continuously in the initial plane of motion. In that case, the motion of the spacecraft about the Earth is two-dimensional. So, the very interesting application of using electric propulsion to change the orbital plane is not discussed here. Using the flight path angle, γ , and the thrust angle, δ , as defined in Figure 1.5, (19.10) can be written as

$$\frac{dH}{d\tau} = a \rho \cos \delta \quad (19.11)$$

$$\frac{d\mathcal{E}}{d\tau} = a w \cos(\delta - \gamma)$$

Note that the instantaneous rate of change of the orbital energy is maximum if $\delta = \gamma$, i.e. if the thrust is acting tangentially to the trajectory. The time-derivative of the orbital angular momentum is maximum if $\delta = 0^\circ$, i.e. if the thrust is directed perpendicular to the radius vector. Sometimes, it is concluded from (19.11-2) that for an optimum escape trajectory from a low-altitude Earth orbit the thrust should be directed continuously along the instantaneous velocity vector, i.e. $\delta = \gamma$. This conclusion is, however, in general not correct, since a minimum total propellant consumption for an escape mission does not necessarily require a maximum instantaneous energy change. The computation of optimum thrust profiles is a difficult topic that is beyond the scope of this book. Many studies have, however, revealed that, when the thrust is small with respect to the acceleration due to gravity, a tangential thrust of constant magnitude is close to the optimum thrust profile for escape trajectories. In Section 19.4, it will be shown that the trajectory then is a spiral with continuously increasing distance between its windings. This implies that the flight path angle gradually increases. For orbit raising missions the spacecraft's flight path angle should eventually become small (or zero) and therefore a component of the thrust should be oriented radially inwards during the last part of the transfer. But also in that case, a tangential thrust profile during the larger part of the mission is a valuable first-order approximation of the true thrust profile. The condition of a relatively small thrust is certainly satisfied in a parking orbit about the Earth. With present-day technology (Table 19.1) we find $f_0 = 10^{-5} - 10^{-3}$ m/s² and $V_j = 15 - 50$ km/s, which corresponds for low orbits about the Earth to $a_0 = 10^{-6} - 10^{-4}$ and $w_j = 2 - 7$. When the same spacecraft moves in interplanetary space, we get a somewhat different picture. For example, at the distance of the Earth from the Sun, the acceleration due to the gravitational attraction by the Sun is approximately $5.9 \cdot 10^{-3}$ m/s². Then, we find for the same range of values for f_0 and V_j : $a_0 = 2 \cdot 10^{-3} - 2 \cdot 10^{-1}$ and $w_j = 0.5 - 2$; so, the dimensionless acceleration is significantly larger and the dimensionless exhaust velocity is smaller. Still, the optimum trajectories for such vehicles resemble a spiral with relatively large distances between its windings, as is illustrated by Figure 19.1.

For first-order analyses of escape flights in the Earth's gravity field and of missions to planets, asteroids or comets the trajectory is sometimes approximated by an appropriately selected analytical function, such as an exponential sinusoid or another type of spiral generating function. Such a function is then used as an approximative solution of the equations of motion. By selecting appropriate values for the various parameters occurring in that analytical function, it is possible to let the trajectory satisfy the specified initial and required final conditions. The analytical expression for the trajectory then can be used to compute the instantaneous values of the required direction and magnitude of the thrust, the velocity of the spacecraft, the flight path angle, etc.

Future more-powerful electric propulsion systems will definitely provide higher thrust levels, which will produce accelerations of interplanetary spacecraft that will not be small anymore relative to the acceleration due to solar gravity. These spacecraft, with characteristic values of $f_0 = 10^{-3} - 10^{-1}$ m/s² and $V_j = 30 - 100$ km/s will fly different types of trajectories that will not be discussed in this Chapter. We will also not discuss optimum powered trajectories for spacecraft that have to encounter another spacecraft or celestial body with a specified velocity.

For our analysis, we assume that the magnitudes of the thrust and of the exhaust velocity, and

consequently also propellant mass consumption per unit of time, are constant. In that case, we find with (1.10) for the dimensionless thrust acceleration

$$\frac{F}{Mg_0} = \frac{F/g_0}{M_0 - \dot{m}t} = \frac{F/(M_0 g_0)}{1 - (F/M_0)(t/V_j)}$$

or, with (19.6),

$$a = \frac{a_0}{1 - a_0 \tau/w_j} \quad (19.12)$$

With the standard rocket engine equation $F = \dot{m}V_j$ we conclude that $a_0 \tau/w_j = M_p(t)/M_0$, where $M_p(t)$ is the mass of propellant consumed until time t . Note that for a constant thrust and exhaust velocity, the thrust acceleration slowly increases with time. If the spacecraft's initial state vector and mass, and the exhaust velocity are known, then the spacecraft trajectory can be computed for any profile of the thrust, \bar{F} , by numerical integration of (19.8).

To gain some qualitative insight, in the following Sections two special cases will be considered for which the equations of motion can partly be solved analytically. These are the cases of *radial thrust* and of *tangential thrust*.

19.3. Trajectories with radial thrust

The motion of a spacecraft with a thrust acting continuously in the radial direction was first studied by H.S. Tsien (1911-2009) around 1953. If we assume that the thrust is directed radially outward ($\delta = 90^\circ$), then (19.11) simplifies to

$$\begin{aligned} \frac{dH}{d\tau} &= 0 \\ \frac{d\mathcal{E}}{d\tau} &= a w \sin \gamma \end{aligned} \quad (19.13)$$

Thus, the angular momentum per unit of mass is constant during the powered flight, just as it is for a Keplerian orbit. From (19.13-1) we find with $H = \rho^2 \dot{\phi}$:

$$\dot{\phi} = \frac{c}{\rho^2}$$

where the notation \cdot indicates a differentiation to τ , ϕ is the angular position of the spacecraft measured from the point of engine ignition, and c is an integration constant. When we assume that the spacecraft starts in a circular orbit, the value of c can be determined from the initial conditions: $\rho = 1$, $\dot{\phi} = w/\rho = 1$, resulting in $c = 1$. So, for the entire propelled flight:

$$\dot{\phi} = \frac{1}{\rho^2} \quad (19.14)$$

Substitution of

$$w^2 = \dot{\rho}^2 + \rho^2 \dot{\phi}^2 \quad (19.15)$$

and (19.14) into (19.10-2) gives

$$\frac{d}{d\tau} \left(\dot{\rho}^2 + \frac{1}{\rho^2} - \frac{2}{\rho} \right) = 2a \frac{d\rho}{d\tau} \quad (19.16)$$

This differential equation can be integrated analytically for $a = constant = a_0$, which, according to (19.12), means that the exhaust velocity is assumed to be infinitely large. We then find

$$\dot{\rho}^2 + \frac{1}{\rho^2} - \frac{2}{\rho} = 2a_0 \rho + c$$

The integration constant c can be determined from the initial conditions: $\rho = 1$, $\dot{\rho} = 0$, resulting in $c = -(1+2a_0)$. After substitution of this relation, we find

$$\dot{\rho}^2 = 2a_0(\rho - 1) + \frac{2}{\rho} - \frac{1}{\rho^2} - 1 \quad (19.17)$$

Although we cannot find an analytical solution for the trajectory of the spacecraft, (19.17) can be used to determine the boundaries of the space that is accessible to the spacecraft. Differentiation of (19.17) with respect to τ yields

$$2\ddot{\rho}\dot{\rho} = 2a_0\dot{\rho} - \frac{2\dot{\rho}}{\rho^2} + \frac{2\dot{\rho}}{\rho^3}$$

So, we may write for the radial acceleration of the spacecraft

$$\ddot{\rho} = a_0 - \frac{1}{\rho^2} + \frac{1}{\rho^3} \quad (19.18)$$

except for the boundaries of the region accessible to the spacecraft, where $\dot{\rho} = 0$. Since the radial acceleration $\ddot{\rho}$ cannot be discontinuous, the acceleration at the boundaries can be found by taking the limit of the acceleration given by (19.18) for $\rho \rightarrow \rho_i$, where the index i denotes a boundary in space where $\dot{\rho} = 0$.

According to (19.17), the radial velocity of the spacecraft equals zero, if

$$(\rho - 1)(2a_0\rho^2 - \rho + 1) = 0$$

This equation has three roots:

$$\rho_1 = 1 \quad ; \quad \rho_{2,3} = \frac{1 \pm \sqrt{1 - 8a_0}}{4a_0} \quad (19.19)$$

The first root, of course, corresponds to the initial circular orbit. The second solution only yields real roots for ρ_2 and ρ_3 if $a_0 \leq 1/8$. This means that for $a_0 > 1/8$ the radial velocity will only be zero in the initial orbit; in any other point of the trajectory the radial velocity is positive. When we consider a spacecraft starting in a circular parking orbit about the Earth, this means that for $a_0 > 1/8$ the spacecraft will continuously move away from the Earth and will eventually escape from the gravity field of the Earth, if the thrust can be maintained long enough. For $a_0 < 1/8$, we find that if the root ρ_2 corresponds to the plus-sign in (19.19) and ρ_3 to the minus-sign: $\rho_1 < \rho_3 < \rho_2$. A numerical analysis of (19.17) shows that $\dot{\rho}^2 < 0$ for $\rho_3 < \rho < \rho_2$, which, of course, is physically impossible; and that $\dot{\rho}^2 > 0$ for $\rho_1 < \rho < \rho_3$ or $\rho > \rho_2$. As we find from (19.18) $\ddot{\rho} > 0$ at $\rho = \rho_1$ and $\ddot{\rho} < 0$ at $\rho = \rho_3$, we conclude that if $a_0 < 1/8$ the spacecraft's distance will continuously vary between ρ_1 and ρ_3 . If $a_0 = 1/8$, the distances ρ_2 and ρ_3 are equal and $\ddot{\rho}$ at that distance is zero. In that case, the spacecraft will asymptotically approach a circular orbit with

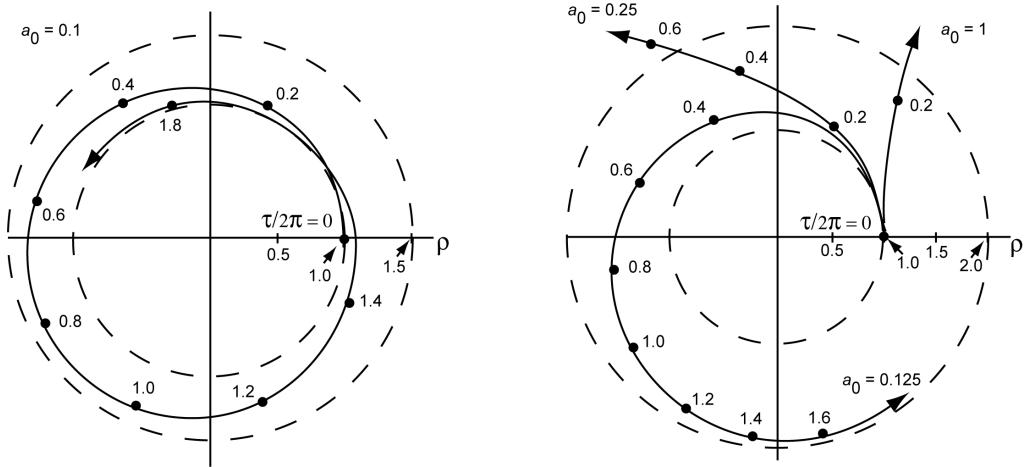


Figure 19.2: Examples of trajectories of spacecraft with a constant radial thrust acceleration of $a_0 = 0.1$, 0.125 , 0.25 and 1 .

radius $\rho = 2$. According to (19.7) and (19.14), the velocity of the spacecraft in that circular orbit is one-half of its value in the initial parking orbit, and $\frac{1}{2}\sqrt{2}$ times the local circular velocity. As an example, Figure 19.2 shows the trajectory of a spacecraft with a constant radial thrust acceleration of $a_0 = 0.1$, 0.125 , 0.25 and 1 . We clearly see the oscillating character of the trajectory for $a_0 = 0.1$; in the trajectories with $a_0 = 0.25$ and 1 , the flight path angle will continuously increase and the trajectory will asymptotically approach the radial outward direction.

The variation of velocity with distance can be found by substitution of (19.14) and (19.17) into (19.15):

$$w^2 = 2a_0(\rho - 1) + \frac{2}{\rho} - 1 \quad (19.20)$$

In Figure 19.3 the value of w is plotted as a function of ρ for some values of $a_0 \geq 1/8$; the Figure also shows curves that represent the variation of the local circular velocity, w_c , and of the local escape velocity, w_{esc} , as a function of ρ . We note that if $a_0 < 1$, the velocity of the spacecraft first

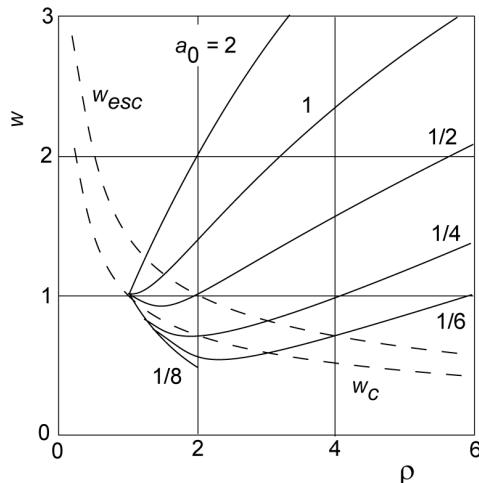


Figure 19.3: Velocity versus distance for a spacecraft with a constant radial thrust acceleration, and curves for the local circular velocity and local escape velocity.

decreases to a minimum value and for $a_0 > 1/8$ increases again at larger distances. For $a_0 \geq 1/2$, the velocity is always higher than the local circular velocity. The distance at which the spacecraft reaches the local escape velocity (for $a_0 > 1/8$) follows by substitution of (19.7-2) into (19.20):

$$\rho_{esc} = \frac{1}{2a_0} + 1 \quad (19.21)$$

This equation shows that the escape velocity is reached at a distance $\rho_{esc} < 5$.

For a numerical example, let we take for the initial thrust acceleration a value of $f_0 = 9.8 \cdot 10^{-4}$ m/s²; this is a large value for present-day technology. For a spacecraft that starts in a low parking orbit about the Earth, we then find $a_0 = 10^{-4} < 1/8$. So, this spacecraft cannot escape from the Earth's gravity field and its distance from the Earth will oscillate between $\rho = 1$ and $\rho = 1.0002$. If the same spacecraft would start in the orbit of the Earth about the Sun, we find $a_0 = 1.66 \cdot 10^{-1} > 1/8$. So, if there is sufficient propellant onboard, the spacecraft will escape from the solar gravity field and will, according to (19.21), reach the local escape velocity at $\rho = 4.0$; between the orbits of Mars and Jupiter.

Though radial thrust may be of some interest for certain orbit maneuvers, it will certainly not play a significant role for orbit raising maneuvers or for lunar and planetary missions. The reasons are that the value of $a_0 > 1/8$, which is needed to escape from the Earth's gravity field, cannot be realized with present-day technology, and that e.g. tangential thrust yields shorter flight times and a lower propellant consumption for these missions.

19.4. Trajectories with tangential thrust

In this Section, the case of a continuous constant tangential thrust in the direction of motion ($\delta = \gamma$) is considered. Then, (19.11) can be written as

$$\begin{aligned} \frac{dH}{d\tau} &= a\rho \cos\gamma \\ \frac{d\varnothing}{d\tau} &= aw \end{aligned} \quad (19.22)$$

For the radial velocity and angular position rate of change we may write

$$\frac{d\rho}{d\tau} = w \sin\gamma \quad (19.23-1)$$

$$\frac{d\varphi}{d\tau} = \frac{w}{\rho} \cos\gamma \quad (19.23-2)$$

where φ is the angular position of the spacecraft measured from the point of engine ignition. From (19.10-2) and (19.22-2) we obtain

$$w \frac{dw}{d\tau} + \frac{1}{\rho^2} \frac{d\rho}{d\tau} = aw$$

Substitution of (19.23-1) into this relation gives

$$\frac{dw}{d\tau} = a - \frac{\sin \gamma}{\rho^2} \quad (19.23-3)$$

From (19.22-1) follows, after substitution of (19.23-2),

$$\frac{d}{d\tau} (\rho w \cos \gamma) = a \rho \cos \gamma$$

Evaluation of this relation leads to

$$w \cos \gamma \frac{d\rho}{d\tau} + \rho \cos \gamma \frac{dw}{d\tau} - \rho w \sin \gamma \frac{d\gamma}{d\tau} = a \rho \cos \gamma$$

Substitution of (19.23-1) and (19.23-3) into this relation yields

$$\frac{d\gamma}{d\tau} = \frac{1}{\rho w} \left(w^2 - \frac{1}{\rho} \right) \cos \gamma \quad (19.23-4)$$

The set of equations of motion (19.23) cannot be solved analytically. However, it can be used to obtain a qualitative insight into the character of the motion. When we assume that the spacecraft starts in a circular parking orbit about the Earth, the initial conditions are $\rho = 1$, $\varphi = 0^\circ$, $w = 1$, $\gamma = 0^\circ$. Equations (19.23-3) and (19.23-1) show that the velocity w will initially increase, while the distance ρ will initially remain constant. Consequently, according to (19.23-4) the flight path angle γ will initially increase. When the flight path angle has reached a non-zero value, the velocity will, according to (19.23-3), increase less rapidly. If $a > 1$, the velocity will keep increasing. However, if the thrust acceleration is sufficiently small, which in practice is the case for spacecraft that start in a circular orbit about the Earth, then the value of $dw/d\tau$ will change sign. From that moment on, the velocity of the spacecraft will decrease. When the distance becomes large enough, a will become larger than $\sin \gamma / \rho^2$ and the value of $dw/d\tau$ will become positive again, which means that the velocity will increase again. When we consider the flight of a deep-space spacecraft that starts its propelled flight in the orbit of the Earth about the Sun, outside the Earth's sphere of influence, the situation is very similar. Then, the initial conditions are $\rho \approx 1$, $\varphi = 0^\circ$, $w > 1$, $\gamma \approx 0^\circ$, while a is, for the same electric propulsion system, significantly larger than for a propelled flight about the Earth. According to (19.23-4), γ will immediately increase and, according to (19.23-1), ρ will also start increasing if $\gamma > 0^\circ$. If $\gamma > 5^\circ$ then, for most interplanetary missions with state-of-the-art electric propulsion systems, $\sin \gamma / \rho^2 > a$, which means, according to (19.23-3), that the velocity of the spacecraft will start decreasing.

In Figure 19.4, distance, position angle, velocity and flight path angle are plotted as a function of time for one orbital revolution. These parameters were computed by numerical integration of (19.23) with $a_0 = 10^3$ and $w_i = 10$. The notation Δ stands for the difference between the actual value of a parameter and its value at $\tau = 0$. Because in the initial circular orbit the position angle increases linearly with time, the quantity $\varphi - \tau$ is plotted to indicate the variation of φ due to the thrust. We see that in the first part of the trajectory the velocity increases, as was already concluded above on basis of a qualitative analysis. After about one-third of a revolution, the velocity falls below the circular velocity at departure, and there is a gain of potential energy at the expense of kinetic energy; the velocity oscillates about a decreasing mean value. We thus have the apparent contradiction that a thrust acting in the direction of motion yields a decreasing velocity! In this context, it is interesting to mention that for atmospheric drag acting on a satellite (Section 21.4) we have the opposite situation. The drag, opposing the direction of motion, leads

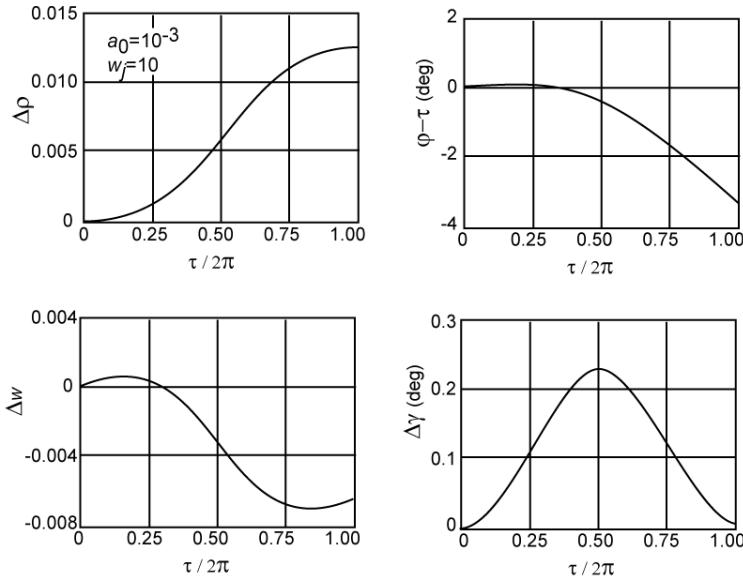


Figure 19.4: Variation of trajectory parameters during the first revolution of a trajectory with constant tangential thrust, and $a_0 = 10^{-3}$, $w_j = 10$.

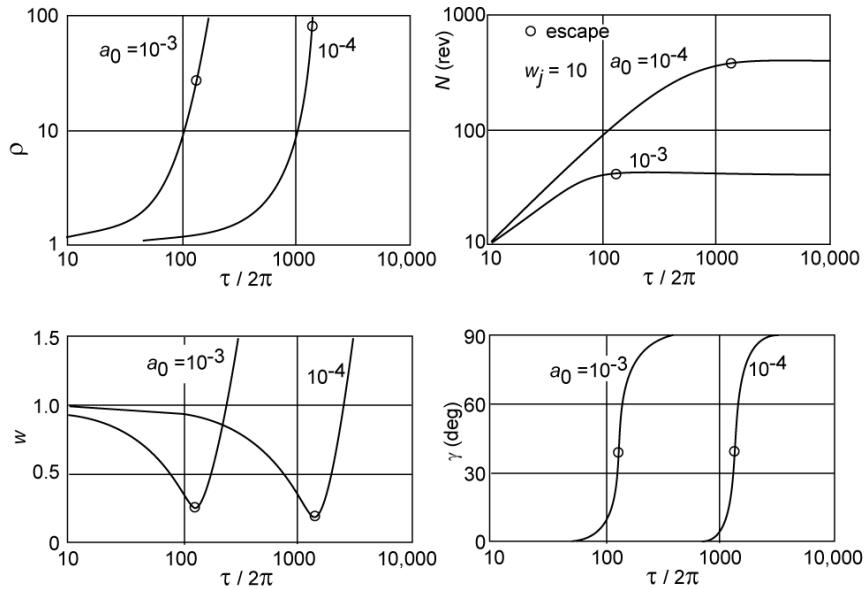


Figure 19.5: Variation of trajectory parameters during two escape flights with constant tangential thrust.

to a decreasing orbital altitude and an increasing velocity! The quantity $\varphi - \tau$ first slightly increases, but starts decreasing after about one-fifth of a revolution. The flight path angle oscillates and reaches a maximum value after about half a revolution; after one complete revolution, it has a small positive value.

The variation of position and velocity over a long period of time are shown in Figure 19.5 for two values of a_0 and for $w_j = 10$. In this case, the number of revolutions, N , defined as $N = \varphi / 2\pi$, is plotted instead of the position angle, φ . Also indicated in the plots is the point where the escape velocity ($\mathcal{E} = 0$) is reached. It should be realized that on the scale applied, the oscillations shown in Figure 19.4 are not visible. Aside from the oscillations, we note that the velocity decreases until some time before escape takes place. This is a major difference with trajectories of spacecraft with chemical propulsion, where the thrust yields an immediate velocity

increase. Electric propulsion systems lead to an increase of the distance and a decrease of the velocity for a significant part of the spiral-shaped trajectory. The escape velocity is reached at a large distance and after the spacecraft has completed many revolutions. Another important fact shown in Figure 19.5 is that the flight path angle remains close to zero ($< 5^\circ$) for a long period of time, but then increases rapidly to $\gamma = 90^\circ$.

Table 19.2 lists the values of the orbital parameters at the moment that the spacecraft reaches the escape velocity, and the ratio between the mass of propellant used until escape and the initial mass, M_p/M_0 , for three values of a_0 . This Table clearly shows that the propellant consumption for escape flights is less than 9% of the initial mass of the spacecraft. This low propellant consumption, which is characteristic for electric propulsion, will be discussed in Section 19.6. As an example, we consider an escape flight where the electric propulsion system starts in a 500 km altitude orbit about the Earth, with $a_0 = 10^4$ and $w_j = 10$. Note that the value taken for the thrust acceleration is large for present-day technology. From the values listed in Table 19.2 we then find $\tau_{esc} = 91.6$ day and $r_{esc} \approx 576,812$ km or about 1.5 times the distance to the Moon. As a second example, we consider an escape flight where the electric propulsion system starts in the Earth's orbit about the Sun, with $a_0 = 10^{-2}$ and $w_j = 10$. Note that in this case w_j is unrealistically large. From the values listed in Table 19.2 we find $t_{esc} \approx 11.4$ year and $r_{esc} \approx 8.5$ AU, which corresponds to a distance between the orbits of Jupiter and Saturn. Figure 19.6 shows an example of an escape trajectory for $a_0 = 10^{-3}$ and $w_j = 10$. Note the increasing distance between the windings of the spiral. Because the inner windings are very close together, only the last few revolutions are shown.

Table 19.2: Trajectory data at the moment of escape for constant tangential thrust and $w_j = 10$.

a_0	10^{-4}	10^{-3}	10^{-2}
τ	8765	818.0	71.50
ρ	83.9	26.6	8.5
N (rev)	390.2	39.12	4.02
w	0.154	0.274	0.485
γ ($^\circ$)	39.2	39.1	38.8
M_p/M_0	0.088	0.083	0.072

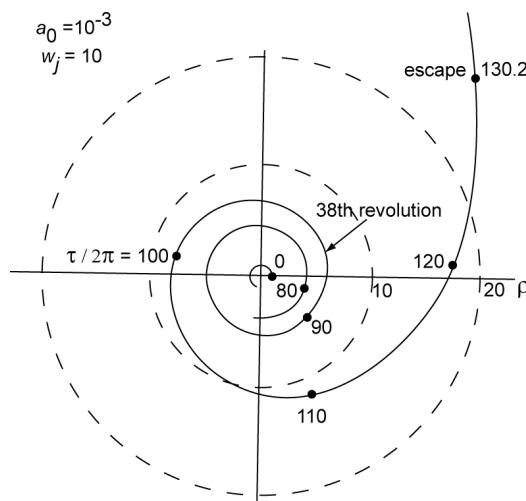


Figure 19.6: Example of an escape trajectory with constant tangential thrust for $a_0 = 10^{-3}$, $w_j = 10$.

19.5. Initial part of a trajectory with tangential thrust

In this Section, we will analyze the initial part of a trajectory with constant tangential thrust, just after the electric engine has been ignited in the original circular orbit. That initial orbit will be considered as the reference orbit, relative to which orbit deviations are measured. In this part of the trajectory the deviations from the circular orbit are small and we may apply *small-perturbation theory*. Since we consider a short time interval, we may assume $a = a_0$. For the virtual position of the spacecraft in the reference orbit at a certain time, we may write $\rho^* = 1$, $\varphi^* = \tau$, $w^* = 1$, $\gamma^* = 0$. Using these reference values, the deviations (Δ) in distance, position angle, velocity and flight path angle in the true orbit may be written as

$$\rho = 1 + \Delta\rho ; \quad \varphi = \tau + \Delta\varphi ; \quad w = 1 + \Deltaw \quad (19.24)$$

Substitution of (19.24) into (19.23), and subsequent linearization of the resulting expressions, leads to

$$\begin{aligned} \frac{d\Delta\rho}{d\varphi^*} &= \Delta\gamma \\ \frac{d\Delta\varphi}{d\varphi^*} &= \Deltaw - \Delta\rho \\ \frac{d\Deltaw}{d\varphi^*} &= a_0 - \Delta\gamma \\ \frac{d\Delta\gamma}{d\varphi^*} &= 2\Deltaw + \Delta\rho \end{aligned} \quad (19.25)$$

We now omit for simplicity the index $*$; the angle φ , however, still indicates the virtual position that the spacecraft would have had in the reference orbit at the same moment of time. Summation of (19.25-1) and (19.25-3) gives

$$\frac{d\Delta\rho}{d\varphi} + \frac{d\Deltaw}{d\varphi} = a_0$$

Integration of this equation results in

$$\Delta\rho + \Deltaw = a_0 \varphi + c$$

The integration constant c can be determined from the initial conditions $\varphi = 0$, $\Delta\rho = 0$, $\Deltaw = 0$, which yields $c = 0$. So, we obtain

$$\Delta\rho + \Deltaw = a_0 \varphi \quad (19.26)$$

Differentiation of (19.25-1) with respect to φ yields

$$\frac{d^2\Delta\rho}{d\varphi^2} = \frac{d\Delta\gamma}{d\varphi}$$

Substitution of (19.25-4) and (19.26) into this relation gives

$$\frac{d^2\Delta\rho}{d\varphi^2} + \Delta\rho = 2a_0 \varphi$$

The solution of this differential equation is

$$\Delta\rho = c_1 \sin\varphi + c_2 \cos\varphi + 2a_0\varphi ; \quad \frac{d\Delta\rho}{d\varphi} = c_1 \cos\varphi - c_2 \sin\varphi + 2a_0$$

The integration constants c_1 and c_2 can be determined from the initial conditions $\varphi = 0$, $\Delta\rho = 0$, $\Delta\gamma = 0$, and (19.25-1). This yields $c_1 = -2a_0$, $c_2 = 0$. Substitution of these values gives

$$\Delta\rho = 2a_0(\varphi - \sin\varphi) \quad (19.27-1)$$

Substitution of this relation into (19.26) gives

$$\Delta w = -a_0(\varphi - 2\sin\varphi) \quad (19.27-2)$$

Substitution of (19.27-1) and (19.27-2) into (19.25-2) leads to

$$\frac{d\Delta\varphi}{d\varphi} = -a_0(3\varphi - 4\sin\varphi)$$

which, after integration, gives

$$\Delta\varphi = -\frac{3}{2}a_0\varphi^2 - 4a_0\cos\varphi + c$$

The integration constant c can be determined from the initial conditions $\varphi = 0$, $\Delta\varphi = 0$, which yields $c = 4a_0$. So, we obtain

$$\Delta\varphi = -\frac{1}{2}a_0[3\varphi^2 - 8(1 - \cos\varphi)] \quad (19.27-3)$$

Substitution of (19.27-1) and (19.27-2) into (19.25-4) leads to

$$\frac{d\Delta\gamma}{d\varphi} = 2a_0\sin\varphi$$

which, after integration, gives

$$\Delta\gamma = -2a_0\cos\varphi + c$$

The value of the integration constant c follows from the initial conditions $\varphi = 0$, $\Delta\gamma = 0$, which yield $c = 2a_0$. So, we obtain

$$\Delta\gamma = 2a_0(1 - \cos\varphi) \quad (19.27-4)$$

Equations (19.27) describe a first-order approximation of the deviations of the powered trajectory from the initial circular orbit as a function of the spacecraft's virtual position angle in that reference orbit. When we analyze the expressions (19.27) for $0^\circ < \varphi < 360^\circ$ we may draw some interesting conclusions:

- The variations of $\Delta\rho$, $\Delta\varphi$, Δw and $\Delta\gamma$ are proportional to the dimensionless thrust acceleration a_0 and show an oscillatory character.
- $\Delta\rho$ is always positive and increases continuously; $\Delta\varphi$ increases until $\varphi \approx 73^\circ$, after which it starts decreasing and becomes negative at $\varphi \approx 105^\circ$; Δw increases until $\varphi \approx 60^\circ$, then starts decreasing and becomes negative at $\varphi \approx 109^\circ$, and starts increasing again at $\varphi = 300^\circ$; $\Delta\gamma$ is always positive, increases until $\varphi = 180^\circ$ and then starts decreasing.

Comparing these results with the results from a numerical integration of the equations of motion ((19.23)) as shown in Figure 19.4, we conclude that the approximative expressions (19.27) quite

accurately describe the initial part of the trajectory.

19.6. Approximate solution of an escape trajectory with tangential thrust

In this Section we will focus on an escape trajectory with constant tangential thrust, starting from a circular parking orbit about the Earth. In Figure 19.7, the values of ρ and w , as obtained from the numerical integration procedure mentioned in Section 19.4, are plotted versus each other

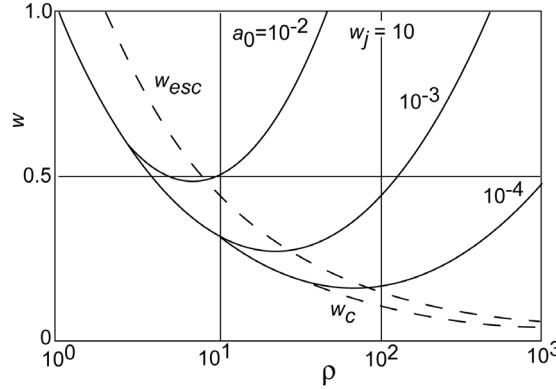


Figure 19.7: Velocity versus distance for trajectories with constant tangential thrust.

for the entire trajectory. Curves for the local circular velocity, w_c , and the local escape velocity, w_{esc} , are also shown. Note that for a small dimensionless thrust acceleration ($a_0 < 10^{-3}$), the velocity of the spacecraft is close to the local circular velocity for a large part of the escape trajectory. Since for present-day technology: $a_0 = 10^{-6} - 10^{-4}$, we conclude that $w \approx w_c$ is a realistic approximation for almost the entire trajectory. Only during the last couple of revolutions before escape, the velocity deviates considerably from the local circular velocity. Consequently, for the largest part of the escape trajectory: 1) the radial velocity component may be neglected with respect to the normal velocity component; 2) the windings of the spiral trajectory are very close to each other. For larger values of a_0 , as is the case for deep space missions where the trajectory is spiraling out about the Sun, the condition $w \approx w_c$ is certainly not valid for the entire escape trajectory but does approximately hold for the initial part of the trajectory (Figure 19.7), in particular for missions with $a_0 < 10^{-2}$. That property is very attractive for missions to solar system bodies moving about the Sun in low-eccentricity orbits at distances of up to Jupiter's distance from the Sun. It means that a spacecraft with electric propulsion approaches that body with a heliocentric velocity that differs little from the body's velocity. This is in sharp contrast to trajectories flown by spacecraft with high-thrust chemical propulsion systems, where the spacecraft's heliocentric velocity upon arrival at the target differs considerably from the velocity of that body, and a large ΔV has to be applied to enter the orbit of that body about the Sun or an orbit about that body.

Returning to the case of an escape trajectory in the Earth's gravity field, it must be noted that the above mentioned properties of the trajectory form the basis of an approximative analytical theory for the escape trajectory of a spacecraft with a small (constant) tangential thrust. This theory is based on the explicit assumption that the velocity of the spacecraft is anywhere (about) equal to the local circular velocity: $w^2 = 1/\rho$. In that case, (19.22-2) leads to

$$\frac{d\mathcal{E}}{d\tau} = \frac{a}{\sqrt{\rho}} \quad (19.28)$$

and we can write for the total orbital energy (per unit of mass):

$$\mathcal{E} = \frac{1}{2} w^2 - \frac{1}{\rho} = -\frac{1}{2\rho} \quad (19.29)$$

Differentiation of (19.29) with respect to τ and subsequently substitution of (19.28) yields

$$\frac{d\rho}{d\tau} = 2a\sqrt{\rho^3}$$

or, with (19.12),

$$\frac{d\rho}{d\tau} = \frac{2a_0}{1 - a_0 \tau/w_j} \sqrt{\rho^3} \quad (19.30)$$

This expression may also be written as

$$\frac{d\rho}{\sqrt{\rho^3}} = -2w_j \frac{d(1 - a_0 \tau/w_j)}{1 - a_0 \tau/w_j}$$

which, after integration, gives

$$\frac{1}{\sqrt{\rho}} = w_j \ln(1 - a_0 \tau/w_j) + c$$

The integration constant c can be determined from the initial conditions $\tau = 0, \rho = 1$, which gives $c = 1$. So, we obtain

$$\rho = \left[\frac{1}{1 + w_j \ln(1 - a_0 \tau/w_j)} \right]^2 \quad (19.31)$$

This equation shows that the distance ρ will initially increase only very slowly.

Table 19.2 and Figure 19.7 show that for present-day technology values of $a_0 \leq 10^{-4}$ and $w_j = 10$, escape will take place at $\rho > 80$. According to (19.31), this means that at the time of escape: $-1 < w_j \ln(1 - a_0 \tau_{esc}/w_j) < -0.89$. When we approximate this relation by $w_j \ln(1 - a_0 \tau_{esc}/w_j) = -1$, which corresponds to $\rho_{esc} = \infty$, then we obtain

$$\tau_{esc} = \frac{1}{a_0} \frac{1 - \exp(-1/w_j)}{1/w_j} \quad (19.32)$$

According to this expression, the escape time is inversely proportional to the initial thrust acceleration, a_0 . For a given value of a_0 , the escape time increases at increasing values of the exhaust velocity, w_j . This can be explained as follows. A higher exhaust velocity at a given value of a_0 implies a decrease of the mass flow rate, \dot{m} , and thus a larger mass of the spacecraft at a given time, and consequently a lower thrust acceleration at that time and thus an increase of the time it takes to reach the escape velocity. Equation (19.32), of course, gives a somewhat too high value of the escape time. The actual escape time can be expressed by

$$\tau_{esc} = \frac{\beta}{a_0} \frac{1 - \exp(-1/w_j)}{1/w_j} \approx \frac{\beta}{a_0} \left[1 - \frac{1}{2} \frac{1}{w_j} + O\left(\frac{1}{w_j}\right)^2 \right] \quad (19.33)$$

where β is a correction factor, which is a function of a_0 , that can be found from a comparison with the results of a numerical integration of the equations of motion. The truncation of the series

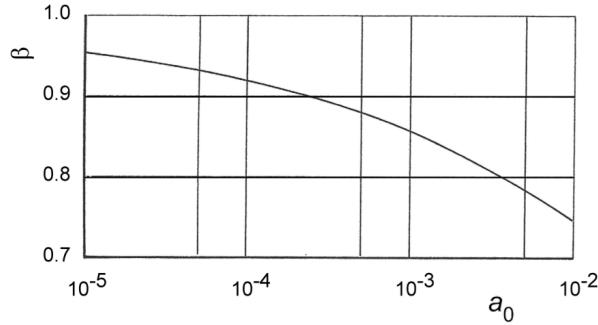


Figure 19.8: The correction factor β as a function of the initial tangential thrust acceleration, a_0 .

expansion gives a reasonable accurate result if $w_j > 2$. Figure 19.8 shows the value of β as a function of a_0 . Note that, for the range of interest for spacecraft with electric propulsion systems escaping from the Earth's gravity field, i.e. $a_0 = 10^{-6} - 10^{-3}$: $\beta = 0.85 - 0.95$. The distance at the time of escape can be found by substitution of (19.33) into (19.31):

$$\rho_{esc} = \left[\frac{1}{1 + w_j \ln \{1 - \beta(1 - \exp(-1/w_j))\}} \right]^2 \approx \left[1 + w_j \ln \left\{ 1 - \beta \left(\frac{1}{w_j} + O\left(\frac{1}{w_j}\right)^2 \right) \right\} \right]^{-2} \quad (19.34)$$

This relation has a poor accuracy, since it is the result of the assumption that the spacecraft's velocity is always (about) equal to the local circular velocity. Note that according to (19.34) ρ_{esc} is not a direct function of a_0 ; it is only indirectly dependent on a_0 through the value of β . For $w_j = 10$ and values of $a_0 = 10^{-2} - 10^{-5}$ a good approximation of the true distance at escape is found by multiplying the value of ρ_{esc} obtained from (19.34) by a factor 0.58.

A relation between position angle, φ , and distance, ρ , may be obtained by using the fact that during the major part of the escape trajectory the radial velocity of the spacecraft can be neglected with respect to the normal velocity. So, we may write to first-order approximation

$$\frac{d\varphi}{d\tau} = \frac{w}{\rho} \quad (19.35)$$

Substitution of (19.30) and (19.35) into

$$\frac{d\varphi}{d\rho} = \frac{d\varphi}{d\tau} \frac{d\tau}{d\rho}$$

yields

$$\frac{d\varphi}{d\rho} = w \left[\frac{1 - a_0 \tau / w_j}{2 a_0 \rho^{5/2}} \right]$$

Using the assumption that the velocity is (about) equal to the local circular velocity ((19.7-1)), we find

$$\frac{d\varphi}{d\rho} = \frac{1 - a_0 \tau / w_j}{2 a_0 \rho^3}$$

Combination of this relation and (19.31) gives

$$d\varphi = \frac{1}{2a_0} \exp\left[\frac{\rho^{-1/2} - 1}{w_j}\right] \frac{d\rho}{\rho^3} \quad (19.36)$$

If the notation

$$x = \frac{1}{\sqrt{\rho}} \quad (19.37)$$

is introduced, we find

$$d\rho = -\frac{2}{x^3} dx ; \quad \rho^3 = \frac{1}{x^6}$$

With these expressions, (19.36) can be written as

$$d\varphi = -\frac{1}{a_0} \exp\left[\frac{x-1}{w_j}\right] x^3 dx$$

This equation can be integrated to

$$\varphi = -\frac{w_j^4}{a_0} \exp\left[\frac{x-1}{w_j}\right] \left[\left(\frac{x}{w_j}\right)^3 - 3\left(\frac{x}{w_j}\right)^2 + 6\left(\frac{x}{w_j}\right) - 6 \right] + c$$

The integration constant c can be determined from the initial conditions $\rho = 1, x = 1, \varphi = 0$, which yields

$$c = \frac{w_j^4}{a_0} \left[\left(\frac{1}{w_j}\right)^3 - 3\left(\frac{1}{w_j}\right)^2 + 6\left(\frac{1}{w_j}\right) - 6 \right]$$

So, we finally obtain

$$\begin{aligned} \varphi = & -\frac{w_j^4}{a_0} \exp\left[\frac{x-1}{w_j}\right] \left[\left(\frac{x}{w_j}\right)^3 - 3\left(\frac{x}{w_j}\right)^2 + 6\left(\frac{x}{w_j}\right) - 6 \right] \\ & + \frac{w_j^4}{a_0} \left[\left(\frac{1}{w_j}\right)^3 - 3\left(\frac{1}{w_j}\right)^2 + 6\left(\frac{1}{w_j}\right) - 6 \right] \end{aligned} \quad (19.38)$$

Equations (19.37) and (19.38) describe the relation between φ and ρ , for given values of a_0 and w_j .

We know that when the spacecraft reaches the escape velocity: $\rho \gg 1, x \ll 1$. Substitution of the series expansion for $\exp[(x-1)/w_j]$ into (19.38), and evaluating the resulting expression for $x \rightarrow 0$ leads to

$$\varphi_{esc} \approx \frac{1}{4a_0} \left[1 - \frac{1}{5w_j} + O\left(\frac{1}{w_j}\right)^2 \right] \quad (19.39)$$

Since for Earth escape trajectories: $w_j > 2$, we find to first-order approximation $\varphi_{esc} \approx 1/(4a_0)$. The number of revolutions that the spacecraft completes about the Earth before escaping from the Earth's gravity field, is then given by

$$N_{esc} \approx \frac{1}{8\pi a_0} \left[1 - \frac{1}{5w_j} + O\left(\frac{1}{w_j}\right)^2 \right] \quad (19.40)$$

or to first-order approximation: $N_{esc} \approx 1/(8\pi a_0)$. Note that φ_{esc} , N_{esc} and τ_{esc} are inversely proportional to a_0 . So, for a small value of a_0 the values of φ_{esc} , N_{esc} and τ_{esc} will be very large.

Because thrust and exhaust velocity are assumed to be constant, the mass of propellant consumed during the escape flight follows from (19.32) and (19.34)

$$\left(\frac{M_p}{M_0} \right)_{esc} \approx \beta [1 - \exp(-1/w_j)] \approx \frac{\beta}{w_j} \left[1 - \frac{1}{2} \frac{1}{w_j} + O\left(\frac{1}{w_j}\right)^2 \right] \quad (19.41)$$

or, to first-order approximation: $(M_p/M_0)_{esc} \approx \beta/w_j$. As shown in Figure 19.8, the correction factor β is for small thrust accelerations only weakly dependent on the initial thrust acceleration, a_0 . So, (19.41) shows that the propellant consumption during the escape flight is to first-order approximation inversely proportional to the exhaust velocity. This conclusion is confirmed by Figure 19.9, where the propellant mass fraction required for escape, as obtained from numerical integration of the equations of motion, is plotted as a function of a_0 for two values of w_j . Note that for $w_j = 10$ ($V_j \approx 77$ km/s), the required propellant mass is less than 9% of the initial mass of the spacecraft.

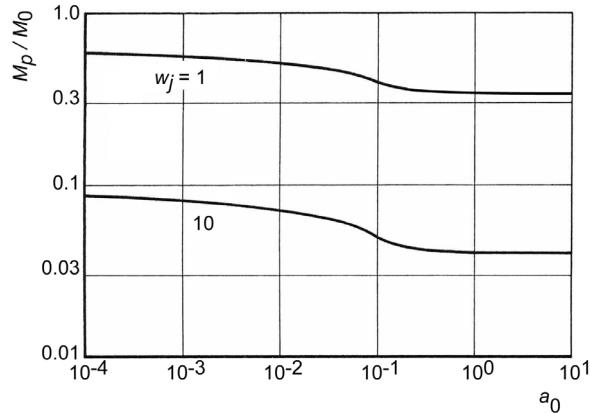


Figure 19.9: Propellant mass fraction for an escape flight with constant tangential thrust as a function of the dimensionless thrust acceleration and exhaust velocity.

As an example, we consider a spacecraft with $a_0 = 10^{-4}$ and $w_j = 10$. From Figure 19.8 we read the corresponding value $\beta = 0.919$. We then compute from (19.7), (19.33), (19.34), (19.39), (19.40), (19.41): $\rho_{esc} = 139$, $\varphi_{esc} = 2,450$ rad, $N_{esc} = 390$, $\tau_{esc} = 8,745$, $w_{esc} = 0.120$ and $(M_p/M_0)_{esc} = 0.0873$. A comparison of these values and the corresponding values obtained from numerical integration of the equations of motion as listed in Table 19.2, shows that the escape time, the number of revolutions until escape, and the propellant consumption until escape can be computed accurately by using the approximative relations. However, the values of the velocity at escape differ about 22% and the values of the escape distance (without applying the correction factor 0.58 given above) even about 66%. This is a direct result of the fact that in the last part of the trajectory before the escape velocity is reached, the approximation $w = w_c$ is not valid anymore and that in this part of the trajectory the radial velocity is quite large (Figure 19.5).

From the viewpoint of propellant consumption, it is interesting to compare an escape flight of a

spacecraft with a low-thrust electric propulsion system and that of a spacecraft with a hypothetical high-thrust rocket engine with the same high exhaust velocity as the low-thrust engine. We assume that the spacecraft with this hypothetical rocket engine departs from the same circular parking orbit about the Earth and that the (large) thrust is (also) acting tangentially to the trajectory. For the spacecraft with the high-thrust engine the impulsive shot concept may be used. Then, the impulsive velocity increase that is required to reach the escape velocity is, according to (7.3-2), given by

$$\Delta V = V_{esc_0} - V_{c_0} = (\sqrt{2} - 1) V_{c_0}$$

or, in the dimensionless notation used in this Chapter,

$$\Delta w = \sqrt{2} - 1$$

The mass consumption then follows from *Tsiolkovski's law* ((1-49)):

$$\left(\frac{M_p}{M_0} \right)_{esc} = 1 - \exp \left[- \frac{(\sqrt{2} - 1)}{w_j} \right]$$

For a high exhaust velocity, we may approximate this expression by

$$\left(\frac{M_p}{M_0} \right)_{esc} \approx \frac{\sqrt{2} - 1}{w_j} \quad (19.42)$$

As an example, we find for $w_j = 10$: $(M_p/M_0)_{esc} = 0.041$. Note that this value compares very well with the values indicated in Figure 19.9 by the curve for $w_j = 10$ at high thrust levels ($a_0 = 1 - 10$). Equation (19.41) gives the value of $(M_p/M_0)_{esc}$ for a spacecraft with a low-thrust engine (electric propulsion). When we compare the propellant consumption of the spacecraft with the high-thrust engine and that of the spacecraft with the low-thrust engine, for the case that both engines have the same high exhaust velocity, we find from (19.41) and (19.42)

$$\frac{(M_p)_{esc,cont}}{(M_p)_{esc,imp}} = \frac{\beta}{\sqrt{2} - 1} \approx 1.9 - 2.3$$

where the index 'cont' refers to the continuous thrusting electric propulsion system and the index 'imp' to the impulsive-shot high-thrust system. So, we conclude that the low-thrust propulsion system requires, for the same exhaust velocity and initial mass, approximately twice the propellant required by the high-thrust system to reach the escape velocity. This is a direct consequence of the *gravity loss* (Section 1.7) experienced during the low-thrust flight. During the long thrusting period of the low-thrust system, the propellant that will be expelled later in the mission is transported away from the Earth, against the action of gravity, which costs energy that has to be provided by the propulsion system.

As a more-realistic example, let us compare the escape flight of a spacecraft with an electric propulsion system with $a_0 = 10^{-4}$, $w_j = 10$, and of a spacecraft with a chemical propulsion system with $a_0 = 1$, $w_j = 0.6$. We have already found that for the electric propulsion system the propellant mass required to achieve the escape velocity is only about 9% of the spacecraft's initial mass, despite the unfavorable low acceleration level. From Figure 19.9 we read that for the chemical propulsion system the required propellant mass is about 40% of the spacecraft's initial mass, despite the favorable high acceleration level. A high-thrust propulsion system with $w_j = 10$ would, according to Figure 19.9, have a propellant consumption of only about 4% of the initial

mass. However, at present, this comparison is not realistic. There is no choice between a high-thrust and a low-thrust system with the same high exhaust velocity. We can only reach a high thrust by chemical or solid-core thermal nuclear fission propulsion with their inherently relatively low exhaust velocities. Nowadays, high exhaust velocities can only be achieved by low-thrust electric propulsion systems. Hopefully, in the future other space propulsion systems that combine a high thrust and a high exhaust velocity will become available; such systems will probably be based on nuclear fusion. Only then, short-duration spaceflight with a relatively low propellant consumption may become feasible.

20. PERTURBING FORCES AND PERTURBED SATELLITE ORBITS

The orbit of a satellite about the Earth and the orbit of an interplanetary spacecraft about the Sun are primarily determined by the gravity field of the Earth, respectively by that of the Sun. In Chapter 5 it was proved that when the mass distribution of the primary body (Earth or Sun) is radially symmetric, then the orbit is a conic section that is called a *Keplerian orbit*. In reality, the mass distribution of celestial bodies is not perfectly radially symmetric and also other forces act on the spacecraft. These additional forces, including those that are the result of the non-radially-symmetric part of the mass distribution of the primary body, cause the orbit of the spacecraft to deviate from a Keplerian orbit. That these deviations are quite small is obvious, for we know that a Keplerian orbit is a good approximation of the actual orbit of a satellite about the Earth, or of a planet or spacecraft about the Sun. After all, Kepler has found from observations of the (true) motion of the planets that the orbit of a planet can be described by a conic section with the Sun at a focal point (Section 5.4). Therefore, these additional forces are called *perturbing forces* and one speaks of *perturbed Keplerian orbits*.

In classical celestial mechanics, where primarily the motion of planets about the Sun is considered, the attraction force between an individual planet and the other planets is the only important perturbation. As a result of the enormous distances between planets, they may be considered as point masses, allowing their gravitational potential to be described by $-\mu/r$. Earth satellites, however, move relatively close to the Earth and the actual mass distribution of the Earth produces significant perturbing forces on a satellite. Many satellites orbit the Earth at low altitudes and this leads to a noticeable effect of the atmosphere on the orbit of those satellites. In addition, there are other perturbing forces acting on a satellite, such as gravitational attraction by other celestial bodies, and in particular by the Sun and the Moon; radiation pressure produced by direct sunlight, sunlight reflected by the Earth (*albedo*) and Earth's infrared radiation; and electromagnetic forces. Which perturbing forces have to be included in the orbit computation process is determined by the accuracy required for the computed orbit. Generally, one will include only those perturbing forces that have a magnitude above some chosen threshold. However, it is emphasized that one always has to be careful and verify whether resonance phenomena occur. Then, very small forces may have large effects. In the following, the most important perturbing forces will be discussed.

20.1. Earth's gravitational force

In Section 1.6, we have derived expressions for the external gravitational potential of an arbitrary body with finite dimensions. We have found that the external potential of a body with a mass distribution that is axially symmetric about the body's rotation axis can, to first-order approximation, be described by (1.45). In this Section, we consider the general case of a body with arbitrary shape and internal mass density distribution and assume that this body is the Earth. We also assume that the external potential is static, and thus neglect the effects of solid-Earth, ocean and pole tides. General potential theory shows that the gravitational potential of the Earth (*geopotential*) at a point outside the Earth may be written as

$$U = -\frac{\mu}{r} \left[1 + \sum_{n=2}^{\infty} \sum_{m=0}^n \left(\frac{R}{r} \right)^n P_{n,m}(\sin \phi) \{ C_{n,m} \cos m\Lambda + S_{n,m} \sin m\Lambda \} \right] \quad (20.1-1)$$

where r, ϕ, Λ are the spherical coordinates of the point that is considered, relative to the geocentric rotating reference frame; r is the distance from the mass center of the Earth, ϕ is the

geocentric latitude, and Λ is the geographic longitude (Section 11.1). The quantity μ is the Earth's gravitational parameter, R is the mean equatorial Earth radius, and the parameters $C_{n,m}$ and $S_{n,m}$ are (constant) model parameters. The terms $P_{n,m}(\sin\phi)$ are *associated Legendre functions of the first kind* of degree n and order m . In order to prevent that in computer calculations very large and very small numbers appear together, often normalized Legendre functions, and normalized coefficients $\bar{C}_{n,m}, \bar{S}_{n,m}$ are used. In this book, we will not apply any normalization and will not discuss this topic.

When we separate the terms with $m = 0$ from the terms with $m \neq 0$, (20.1-1) can be written as

$$U = -\frac{\mu}{r} \left[1 + \sum_{n=2}^{\infty} C_{n,0} \left(\frac{R}{r} \right)^n P_{n,0}(\sin\phi) + \sum_{n=2}^{\infty} \sum_{m=1}^n \left(\frac{R}{r} \right)^n P_{n,m}(\sin\phi) \{ C_{n,m} \cos m\Lambda + S_{n,m} \sin m\Lambda \} \right]$$

When we define

$$\begin{aligned} J_n &\equiv J_{n,0} = -C_{n,0} ; \quad P_n(\sin\phi) \equiv P_{n,0}(\sin\phi) \\ C_{n,m} &= J_{n,m} \cos m\Lambda_{n,m} ; \quad S_{n,m} = J_{n,m} \sin m\Lambda_{n,m} \end{aligned} \quad (20.1-2)$$

the expression for the gravitational potential can be written as

$$U = -\frac{\mu}{r} \left[1 - \sum_{n=2}^{\infty} J_n \left(\frac{R}{r} \right)^n P_n(\sin\phi) + \sum_{n=2}^{\infty} \sum_{m=1}^n J_{n,m} \left(\frac{R}{r} \right)^n P_{n,m}(\sin\phi) \{ \cos m(\Lambda - \Lambda_{n,m}) \} \right] \quad (20.1-3)$$

where the terms $P_n(\sin\phi)$ are *Legendre polynomials* of degree n , and $J_{n,m}$ and $\Lambda_{n,m}$ replace the model parameters $C_{n,m}$ and $S_{n,m}$. The definition $J_n \equiv -C_{n,0}$ is merely to make J_2 positive; so, it is a matter of convention. When we compare (1.45-1) and (20.1-3), we note that (1.45-1) represents the first term of the first series of perturbing terms in (20.1-3). For the Legendre polynomials and associated Legendre function of the first kind the following expressions hold:

$$x = \sin\phi$$

$$P_n(x) = \frac{1}{(-2)^n n!} \frac{d^n}{dx^n} (1-x^2)^n ; \quad P_{n,m}(x) = (1-x^2)^{m/2} \frac{d^m P_n(x)}{dx^m} \quad (20.2-1)$$

Equations (20.1) only hold when r is larger than the maximum radius of the body (*external gravity field*). The summation index n begins at two, because, by definition, the center of the coordinate system coincides with the Earth's center of mass; then, the coefficients $C_{1,0}, C_{1,1}, S_{1,1}, J_1$, and $J_{1,1}$ are all zero. The coefficients $C_{2,1}, S_{2,1}, J_{2,1}$ would vanish if the Z-axis would be aligned with the Earth's main axis of inertia. A gravity model, e.g. developed by processing satellite tracking data, essentially consists of the values of the $C_{n,m}$ and $S_{n,m}$ coefficients, together with adopted values of μ and R . We then can compute the values of $J_{n,m}$ and $\Lambda_{n,m}$ from

$$J_{n,m} \equiv \sqrt{C_{n,m}^2 + S_{n,m}^2} \quad ; \quad \Lambda_{n,m} = \frac{1}{m} \arctan\left(\frac{S_{n,m}}{C_{n,m}}\right)$$

where the signs of $C_{n,m}$ and $S_{n,m}$ determine the quadrant of $\Lambda_{n,m}$. Note that the coefficients $J_{n,m}$ are all positive. Modern gravity models contain coefficients up to degree and order 2150, and some even up to higher degree and order. However, for most applications in astrodynamics truncated versions of these models are used, retaining only terms that contribute significant orbit perturbations. In geodesy, geophysics and advanced astrodynamics, generally, (20.1-1) is used. In this book, we will mostly use (20.1-3); only in Section 23.5 we will use the $C_{n,m}$ and $S_{n,m}$ coefficients as gravity model parameters.



Figure 20.1: Zonal harmonics ($n \neq 0, m = 0$; left), sectorial harmonics ($m = n \neq 0$; center), and tesseral harmonics ($m \neq n \neq 0$; right). White areas represent elevation above and black areas represent elevation below a mean spherical surface.

Table 20.1: Some coefficients of the GRACE GGM02C Earth gravity field model¹.

n	m	$J_{n,m} (10^{-6})$	$\Lambda_{n,m} (\text{°})$	n	m	$J_{n,m} (10^{-6})$	$\Lambda_{n,m} (\text{°})$
<i>Zonal harmonics</i>							
2	0	1082.6357	----	6	0	0.5406168	----
3	0	-2.5324737	----	7	0	-0.3505229	----
4	0	-1.6199743	----	8	0	-0.2040168	----
5	0	-0.2279051	----	9	0	-0.1221502	----
<i>Tesseral and sectorial harmonics</i>							
2	1	0.0018225	98.3325	5	2	0.1177945	-13.1874
2	2	1.8155628	-14.9287	5	3	0.0165311	-51.5184
3	1	2.2094849	6.9684	5	4	0.0023321	42.6063
3	2	0.3744510	-17.1887	5	5	0.0017034	-15.0730
3	3	0.2213884	20.9932	6	1	0.0632600	160.7431
4	1	0.6786576	-138.5480	6	2	0.0468934	-41.2906
4	2	0.1675835	31.0578	6	3	0.0012011	2.9578
4	3	0.0604207	-3.8219	6	4	0.0018140	-25.0853
4	4	0.0076442	30.3487	6	5	0.0004837	-23.2946
5	1	0.0971238	-123.6853	6	6	0.0000554	-14.6194

Associated constants: $\mu = 398600.4415 \text{ km}^3/\text{s}^2$; $R = 6378.1363 \text{ km}$.

¹ B. Tapley, J. Ries, S. Bettadpur, D. Chambers, M. Cheng, F. Condi, B. Gunter, Z. Kang, P. Nagel, R. Pastor, T. Pekker, S. Poole, F. Wang, *GGM02 - An improved Earth gravity field model from GRACE*, Journal of Geodesy, Vol. 79, No. 8, November 2005, pp. 467-478.

The first term in (20.1-3) represents the gravitational potential of a spherical body with radially symmetric mass density distribution: the Newton potential, while the second term represents the influence of deviations of the shape and mass density distribution in north-south direction (*zonal harmonics*) and the third term represents the influence of deviations of the shape and mass density distribution in north-south and east-west direction (*tesseral and sectorial harmonics*). For the tesseral harmonics holds $m \neq n$; for the sectorial harmonics $m = n$, but frequently the name tesseral harmonics is used to include all $m \neq 0$ harmonics, irrespective of their order. When projected onto a sphere, the tesseral harmonics have zeros along $n-m$ parallels of latitude and along $2m$ meridians (Figure 20.1). As may be expected, (20.1) shows that the effect of the non-radially-symmetric mass distribution of the Earth on the orbit of a satellite is larger when the satellite is closer to the Earth. The value of the coefficient of the first perturbing term is $J_2 = 1.083 \cdot 10^{-3}$, while all other J -coefficients are smaller than $2.6 \cdot 10^{-6}$ (Table 20.1). So, the first perturbing term produces by far the largest perturbing force.

In practice, the values of the Legendre polynomials and functions are not computed from (20.2-1), but from recurrence relations that yield these values for all values of n and m ($m \leq n$):

$$\begin{aligned} (n+1)P_{n+1,0}(x) &= (2n+1)xP_{n,0}(x) - nP_{n-1,0}(x) \\ P_{n,n}(x) &= (2n-1)\sqrt{1-x^2}P_{n-1,n-1}(x) \\ P_{n,n-1}(x) &= (2n-1)xP_{n-1,n-1}(x) \\ P_{n,m}(x) &= \frac{2n-1}{n-m}xP_{n-1,m}(x) - \frac{n+m-1}{n-m}P_{n-2,m}(x) \end{aligned} \quad (20.2-2)$$

with the starting values $P_{0,0}(x) = 1$, $P_{1,0}(x) = x$. Successively, the first of these expressions yields the values of all zonal harmonics polynomials, the second one the values of all sectorial harmonics functions, and the third and fourth ones the values of all tesseral harmonics functions.

The perturbing acceleration \bar{f} is given by

$$\bar{f} = -\bar{\nabla}\left(U + \frac{\mu}{r}\right) \quad (20.3)$$

As an example, expressions for the perturbing acceleration will be derived for the cases that only the J_2 -term or the $J_{2,2}$ -term of the gravitational potential is taken into account.

When we consider only the J_2 -term, we may write

$$\bar{f} = -\bar{\nabla}\left[\frac{\mu}{r}J_2\left(\frac{R}{r}\right)^2P_2(\sin\phi)\right] \quad (20.4)$$

From (20.2) we find

$$P_2(\sin\phi) = \frac{1}{2}(3\sin^2\phi - 1)$$

Substitution of this relation into (20.4) leads to

$$\bar{f} = -\bar{\nabla}\left[\frac{1}{2}\mu J_2 \frac{R^2}{r^3}(3\sin^2\phi - 1)\right] \quad (20.5)$$

With $\sin\phi = z/r$, (20.5) becomes

$$\bar{f} = -\nabla \left[\frac{1}{2} \mu J_2 \frac{R^2}{r^3} \left(3 \frac{z^2}{r^2} - 1 \right) \right]$$

From this relation, we obtain for the rectangular components of the perturbative acceleration

$$\begin{aligned} f_x &= -\frac{3}{2} \mu J_2 \frac{R^2}{r^5} x \left(1 - 5 \frac{z^2}{r^2} \right) \\ f_y &= -\frac{3}{2} \mu J_2 \frac{R^2}{r^5} y \left(1 - 5 \frac{z^2}{r^2} \right) \\ f_z &= -\frac{3}{2} \mu J_2 \frac{R^2}{r^5} z \left(3 - 5 \frac{z^2}{r^2} \right) \end{aligned} \quad (20.6)$$

We also can express the perturbing acceleration in spherical coordinates by taking the relevant partial derivatives of (20.5). We then find

$$\begin{aligned} f_r &= \frac{3}{2} \mu J_2 \frac{R^2}{r^4} (3 \sin^2 \phi - 1) \\ f_\phi &= -\frac{3}{2} \mu J_2 \frac{R^2}{r^4} \sin 2\phi \\ f_\Lambda &= 0 \end{aligned} \quad (20.7)$$

From (20.7) we find that

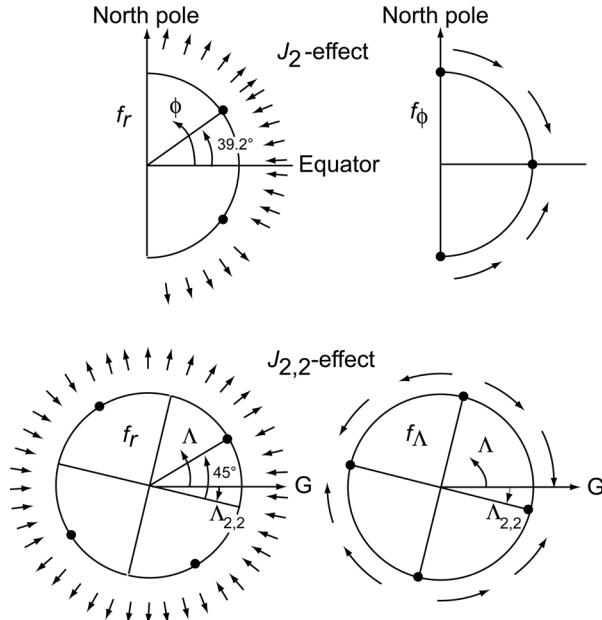


Figure 20.2: Diagrams for the direction of the J_2 -accelerations in the meridional plane (top) and for the direction of the $J_{2,2}$ -accelerations in the equatorial plane (bottom). Points where an acceleration component is zero are indicated by •. The direction to the Greenwich meridian is indicated by G .

$$f_r = 0 \quad \text{if} \quad \phi = -39.18^\circ, 39.18^\circ$$

$$f_\phi = 0 \quad \text{if} \quad \phi = -90^\circ, 0^\circ, 90^\circ$$

Since $J_2 > 0$ (Table 20.1), we conclude that f_r is positive if $\phi > 39.2^\circ$ or $\phi < -39.2^\circ$; f_r is negative if $-39.2^\circ < \phi < 39.2^\circ$. The latitudinal acceleration, f_ϕ , is negative if $0^\circ < \phi < 90^\circ$ and positive if $-90^\circ < \phi < 0^\circ$. The direction of the perturbative accelerations is shown in Figure 20.2 (top). The diagrams suggest that the J_2 -effect may be visualized as the gravitational effect of an additional ring of mass around the Earth's equator and a shortage of mass at the polar regions. However, this conclusion is, in general, not correct, since (20.1) does not describe the shape of the Earth but the Earth's gravitational potential. So, (20.7) would also hold for a purely spherical Earth with a band of higher internal mass densities around the equator!

For a specified distance, r , the maximum values of the perturbing acceleration components are

$$|f_r|_{max} = 3\mu J_2 \frac{R^2}{r^4} \quad \text{at} \quad \phi = -90^\circ, 90^\circ$$

$$|f_\phi|_{max} = \frac{3}{2}\mu J_2 \frac{R^2}{r^4} \quad \text{at} \quad \phi = -45^\circ, 45^\circ$$

For a satellite at an altitude of 250 km, we find: $|f_r|_{max} = 2|f_\phi|_{max} = 2.73 \text{ cm/s}^2$, which is about 0.3% of the local gravitational acceleration (9.08 m/s^2). For higher altitudes, the perturbing accelerations are smaller both in an absolute and in a relative sense. For a geostationary satellite ($i = 0^\circ$, $r = 42,164 \text{ km}$), the J_2 -term produces a constant acceleration of $-8.33 \times 10^{-6} \text{ m/s}^2$ in radial direction, which is about 0.004% of the local gravitational acceleration (22 cm/s^2). This acceleration has the same effect on the orbit as a slight increase of the value of μ would have and consequently ((6.19)) increases the value of the true geostationary radius by about 2 km (Section 23.7).

The Earth completes one revolution about its axis in one day, while the orbital plane of a satellite keeps (in first approximation) a fixed orientation in space. As a result, in most cases the effects of deviations in the shape and mass density distribution in east-west direction will largely average out over periods longer than a day, and we may then neglect the effects of the $J_{n,m}$ -terms for many applications. An important exception to this rule is a *geostationary satellite*. This satellite is continuously located above the same point on the Earth's equator and its orbit is therefore strongly perturbed by $J_{n,m}$ -terms. However, because for a geostationary satellite $R/r \approx 0.15$, in most cases only the first few $J_{n,m}$ -terms have to be taken into account.

As an example, we will derive expressions for the perturbing accelerations due to the $J_{2,2}$ -term; for geostationary satellites this is the dominant term of the series of tesseral and sectorial harmonics. We may write according to (20.1-3) and (20.3):

$$\bar{f} = -\bar{\nabla} \left[-\frac{\mu}{r} J_{2,2} \left(\frac{R}{r} \right)^2 P_{2,2}(\sin \phi) \{ \cos 2(\Lambda - \Lambda_{2,2}) \} \right] \quad (20.8)$$

From (20.2), we find

$$P_{2,2}(\sin \phi) = 3 \cos^2 \phi$$

Substitution of this relation into (20.8) leads to

$$\bar{f} = \bar{\nabla} \left[3\mu J_{2,2} \frac{R^2}{r^3} \cos^2 \phi \cos 2(\Lambda - \Lambda_{2,2}) \right] \quad (20.9)$$

We obtain the disturbing accelerations in spherical coordinates by taking the relevant partial derivatives of (20.9), and find

$$\begin{aligned} f_r &= -9\mu J_{2,2} \frac{R^2}{r^4} \cos^2 \phi \cos 2(\Lambda - \Lambda_{2,2}) \\ f_\phi &= -3\mu J_{2,2} \frac{R^2}{r^4} \sin 2\phi \cos 2(\Lambda - \Lambda_{2,2}) \\ f_\Lambda &= -6\mu J_{2,2} \frac{R^2}{r^4} \cos \phi \sin 2(\Lambda - \Lambda_{2,2}) \end{aligned} \quad (20.10)$$

We now confine ourselves to equatorial satellites, i.e. $\phi = 0^\circ$. In that case (20.10) shows that

$$f_r = 0 \quad \text{if } \Lambda - \Lambda_{2,2} = 45^\circ, 135^\circ, 225^\circ, 315^\circ$$

$$f_\phi = 0 \quad \text{for all values of } \Lambda$$

$$f_\Lambda = 0 \quad \text{if } \Lambda - \Lambda_{2,2} = 0^\circ, 90^\circ, 180^\circ, 270^\circ$$

Since $J_{2,2} > 0$ (Table 20.1), we conclude that the radial component, f_r , is negative if $-45^\circ < \Lambda - \Lambda_{2,2} < 45^\circ$ and $135^\circ < \Lambda - \Lambda_{2,2} < 225^\circ$; and positive if $45^\circ < \Lambda - \Lambda_{2,2} < 135^\circ$ and $225^\circ < \Lambda - \Lambda_{2,2} < 315^\circ$. The longitudinal component, f_Λ , is negative if $0^\circ < \Lambda - \Lambda_{2,2} < 90^\circ$ and $180^\circ < \Lambda - \Lambda_{2,2} < 270^\circ$; and positive if $90^\circ < \Lambda - \Lambda_{2,2} < 180^\circ$ and $270^\circ < \Lambda - \Lambda_{2,2} < 360^\circ$. The direction of the perturbative accelerations is shown in Figure 20.2 (bottom).

From (20.10) we find for the maximum values of the perturbing acceleration components for a circular equatorial orbit with a specified radius:

$$|f_r|_{max} = 9\mu J_{2,2} \frac{R^2}{r^4} \quad \text{at } \Lambda - \Lambda_{2,2} = 0^\circ, 90^\circ, 180^\circ, 270^\circ$$

$$|f_\Lambda|_{max} = 6\mu J_{2,2} \frac{R^2}{r^4} \quad \text{at } \Lambda - \Lambda_{2,2} = 45^\circ, 135^\circ, 225^\circ, 315^\circ$$

For a satellite at an altitude of 250 km, we find $|f_r|_{max} = 1.5 |f_\Lambda|_{max} = 138 \mu\text{m/s}^2$, which is about 1.5×10^{-5} of the local gravitational acceleration (9.08 m/s^2). For a geostationary satellite, the values of the maximum acceleration components are $|f_r|_{max} = 1.5 |f_\Lambda|_{max} = 0.084 \mu\text{m/s}^2$, which is about 3.8×10^{-7} of the local gravitational acceleration (22 cm/s^2). From (20.7-1) and (20.10-1) we find that for a satellite in a circular equatorial orbit the ratio between the maximum radial acceleration due to the $J_{2,2}$ -term and the radial acceleration due to the J_2 -term is given by

$$\frac{|f_{r_{2,2}}|_{max}}{|f_{r_2}|} \approx 6 \frac{J_{2,2}}{J_2} \approx 10^{-2}$$

independent of altitude.

For most equatorial satellites, the satellite's orbital angular velocity differs significantly from the Earth's rotational velocity and therefore the satellite will traverse the full equatorial force

field produced by the $J_{2,2}$ -term. As mentioned before, an exception is a geostationary satellite, which rotates about the Earth at the same angular velocity as the Earth rotates about its axis. Therefore, a geostationary satellite keeps (almost) a fixed position above the equator relative to the rotating gravity field. In that case, the $J_{2,2}$ -effect produces an almost constant force on the satellite for an extended period of time, which results in *orbital resonance* and leads to significant orbit perturbations. Resonance orbit perturbations may occur for any orbital altitude; the phenomenon will be discussed in some detail in Section 23.5. In essence, orbital gravity field resonance occurs when two frequencies in the disturbing acceleration—those associated with the Earth's rotation and with the orbital period—are nearly commensurate; i.e. if their ratio is the ratio of two integers. This causes the satellite to sample a particular aspect of the Earth's gravity field over and over, causing dynamical effects that build up over time. The result is a variation in the satellite's motion with very long periods. These periods may range from weeks for *shallow resonance* to months or even years for *deep resonance* (Section 23.5). The amplitudes range from meters to tens or hundreds of kilometers, depending on the orbit.

Because J_2 is about a thousand times larger than the other J_n - and $J_{n,m}$ -coefficients, for first-order orbit computations of non-geostationary satellites only the J_2 -term has to be included in the computations. Table 20.1 shows that for $n \geq 3$ the magnitude of the J_n - and $J_{n,m}$ -coefficients is not always very small compared to the magnitude of J_3 . For satellites in low orbits, for which $R/r \approx 1$, it is therefore not sufficient to add only the J_3 -term if one wants to compute the orbit more accurately, but one has to include a series of J_n - and $J_{n,m}$ -terms. This is a major reason why an accurate orbit computation of satellites below 800 km altitude is so complicated.

20.2. Atmospheric drag

In analogy with common practice in aeronautics, we write for the acceleration of a satellite due to atmospheric drag

$$\bar{f} = -C_D \frac{1}{2} \rho \frac{A}{M} |\bar{v}| \bar{v} \quad (20.11)$$

Here, ρ is the atmospheric density, \bar{v} is the velocity of the satellite relative to the (rotating) Earth's atmosphere, M is the mass of the satellite and C_D is the drag coefficient related to a reference surface A . The satellite's *ballistic coefficient* is defined as $B = C_D A/M$. In general, the motion through the upper atmosphere will also generate a force component perpendicular to the velocity vector. In most cases, this component is small when compared to the drag component and we will therefore neglect it. The mass of the satellite is, of course, known accurately. When certain assumptions are made about the rotation of the atmosphere and about the speed and direction of (horizontal) winds high in the atmosphere, then also the velocity of the satellite with respect to the atmosphere can be determined. The determination of the values of the parameters C_D , A and ρ , however, introduces serious problems.

At altitudes above 180 km, the molecular mean free path is larger than 120 m and one may therefore state that in the altitude range where satellites move about the Earth, the aerodynamic flow around the satellite is of the free molecular flow type. When we know the shape and attitude of the satellite, then, in principle, we can derive the orientation of each surface element of the satellite relative to the incoming flow of atmospheric particles at every instant of time. For high-accuracy analyses such detailed satellite models are used, where each surface element has specific surface characteristics, to compute the atmospheric forces on each surface element. An additional problem is that for complicated shapes individual particles may hit the satellite more

than once. From the atmospheric forces on all surface elements, it is possible to compute the atmospheric drag, lift and torque on a spacecraft of arbitrary complexity. In most cases, however, the analysis is simplified by taking the cross-sectional area of the satellite perpendicular to the velocity vector as a reference surface. For non-spherical satellites, this reference surface will, generally, vary with time. The drag coefficient is primarily dependent on the kind of atmospheric particles that strike the satellite, the kinetic energy of these particles, the velocity of the satellite, the kind and characteristics of the surface (elements) of the satellite, and the temperature of this surface. When all these aspects are taken into account, we can theoretically compute the drag coefficient for different shapes of a satellite and we then usually obtain a value of $C_D = 2-3$; where for lower altitudes C_D tends to be closer to 2 and for higher altitudes closer to 3. The value of the atmospheric density has to be taken from an atmospheric model.

At altitudes below 90 km the atmosphere is well mixed; i.e. the concentrations of the major species are independent of height and location and the atmosphere consists of about 78% N₂, 21% O₂, and 1% Ar. At altitudes above 100 km molecular collisions are much less frequent and the atmospheric constituents can separate through gaseous diffusion more rapidly than they are mixed by turbulence. As a result, the composition of the atmosphere starts to change with altitude, where the mean molecular weight of the atmospheric constituents decreases with altitude. In the *thermosphere*, which extends from about 90 km altitude to about 500 km altitude, the (kinetic gas) temperature is increasing asymptotically with altitude due to the absorption of solar extreme UV and X-ray radiation. At altitudes above about 500 km (*exosphere*) atmospheric particles may escape from the atmosphere. The region above about 600 km altitude is called the *magnetosphere*; it extends several tens of thousands of kilometers into space. In this region the atmosphere is practically completely ionized and the motion of the charged particles is primarily controlled by the Earth's magnetic field.

The first observations of satellite orbit perturbations in the late 1950s paved the way for the first generation of empirical density models in the 1960s and early 1970s. Well known is the series of Jacchia models (L.G. Jacchia, 1911-1996) developed at the US Smithsonian Astrophysical Observatory. The Jacchia-71 model was adopted as the Committee on Space Research (COSPAR) International Reference Atmosphere (CIRA) in 1972, and is therefore also known as the CIRA-72 reference atmosphere. Over the years, the Jacchia models have been improved and this has eventually resulted in the Jacchia-Bowman 2006 and 2008 models. In France, the first of a series of DTM (Drag Temperature Model) atmospheric models was developed at the Groupe de Recherche de Géodésie Spatiale, in 1978. Like the Jacchia models, the early DTM models were based solely on observations of satellite drag and neutral atmospheric temperatures. A major update, named DTM-94, was published in 1998. The database for this model included satellite mass spectrometer data and accelerometer data measured by the CACTUS accelerometer on board the Castor satellite. The DTM-2009 model, published in 2012, also incorporated accelerometer data acquired by the CHAMP and GRACE satellites. The problem with this kind of atmospheric models is that the observed orbit perturbations or onboard accelerometer data basically provide information about the product $C_D \rho S$. In the construction of the atmospheric model certain assumptions about S and C_D of each satellite of which the measurements are used in the construction of that model have been made. Traditionally, a constant reference surface perpendicular to the satellite's velocity vector relative to the rotating atmosphere and a certain profile of C_D versus altitude were adopted. This profile was selected in such a way that for altitudes below 400 km: $C_D \approx 2.2$, while for higher altitudes C_D gradually increases and at 800 km altitude asymptotically approaches a specific value, e.g. 2.6. When one uses such an atmospheric model in the computation of satellite orbit perturbations due to atmospheric forces, then it is important that the same definition of the reference surface and the

same C_D - h profile are used and not, e.g. a value of C_D computed from theoretical gas dynamics models.

Starting in the late 1970s, a new class of atmospheric density models, named MSIS, was developed at the NASA Goddard Space Flight Center. These models were based solely on satellite mass spectrometer data and ground-based incoherent scatter radar observations. The main advantage of these models over the drag-derived density models is that they provide observations of both temperature and number densities for the atmospheric constituents independent of assumptions on the values of C_D and S . The MSIS-86 model replaced the Jacchia-71 model as the COSPAR International Reference Atmosphere, and is therefore also known as the CIRA-86 model. An extension of this model was published as MSIS-90. At the end of the 1990s, the development of MSIS-class models was continued at the US Naval Research Laboratory (NRL). The resulting NRLMSISE-00 model, published in 2002, includes additional mass spectrometer and incoherent scatter radar data, as well as satellite accelerometer data and the DTM and Jacchia satellite orbit decay databases. The CIRA-2012 model is a state-of-the-art model based on updated versions of all relevant empirical models that have been developed by means of the considerable advances in direct measurements by appropriate space-based instrumentation of the 1980s, 1990s and early-mid 2000s.

The atmosphere is a highly dynamical system that is driven by the solar heat input, and we may recognize various types of variations in the density at a location with specified latitude, longitude and height. Examples are the *diurnal variation* due to the rotation of the Earth relative to the direction of the solar heat input; the *27-day variation* (actually 26-30 days) due to the rotation of the Sun about its axis and resulting in a fluctuation that is correlated to the radiation produced by active regions on the Sun; the *11-year solar activity cycle*, which is related to the sunspot cycle (actually, it is a 22-year variation); the *semi-annual variation*, which is related to the varying distance of the Earth from the Sun and to the variation of the Sun's declination during a year. In addition, there are various types of (semi-)regular and irregular variations. As a result, the conditions in the higher atmosphere are, apart from being a function of altitude, also dependent on the local time of the day, the period in the year (season) and the *solar activity*. These variables lead to variations in the heating of the atmosphere. A larger heat input results in an expansion of the atmosphere, causing the atmospheric density at a certain altitude to increase. Each of the variations mentioned has been found to be related to one or more observable parameters, and empirical formulas have been constructed to compute the so-called *exospheric temperature*, T_∞ , i.e. the kinetic gas temperature at altitudes above 500 km. Once the exospheric temperature has been computed, atmospheric composition, density and related quantities can be found for any altitude above 150 km. Of the three types of variations, solar activity yields the largest effect on atmospheric density for altitudes higher than 300 km. This solar activity is connected to processes in the Sun, which show a cycle of 11 years. During an active period, regions of the Sun radiate a considerably larger amount of UV, X-ray and corpuscular radiation. If this radiation hits the Earth, then a large part of the additional radiation is absorbed in the higher atmosphere and leads there to an additional heating of the atmosphere. Since the Sun is a gaseous body, it does not rotate rigidly like the solid planets and moons do. The time for a fixed feature on the Sun to rotate to the same apparent position as viewed from Earth varies from 26 days for a feature at the equator to more than 30 days for a feature near the poles. Because of this solar rotation, the additional heating of the Earth's atmosphere produced by a long-living feature on the Sun may show a variation with a period of 26 to 30 days. The level of solar activity can be correlated to the values of two parameters that can be measured on Earth: the *solar 10.7-cm*

radiation flux ($F_{10.7}$) in units of $10^{-22} \text{ W m}^{-2} \text{ Hz}^{-1}$ (1 sfu or 10^4 Jy)², which serves as a measure of UV and X-ray radiation entering the atmosphere, and the *3-hour geomagnetic index* (a_p) in units of nanotesla³, which serves as a measure of corpuscular radiation entering the atmosphere. The $F_{10.7}$ flux is measured at the Dominion Radio Astrophysical Observatory in British Columbia, while the a_p index is derived from geomagnetic field measurements made at several locations around the world. Traditionally, only these two parameters were used to model the fluctuations in the solar heat input. Nowadays, however, also satellite observations in the (extreme) UV and X-ray part of the spectrum are used to better model the heat input from specific solar irradiance sources to major thermospheric layers. The radiation at these wavelengths does not penetrate to the Earth's surface and measurements on the intensity of this radiation can only be provided by satellites. In addition to the a_p index, also a parameter is used that is an indicator of the strength of the storm-time ring current in the inner magnetosphere generated by high-energy solar particles entering the atmosphere. This current produces Joule heating in which the drift energy of ions turns into thermal and kinetic energy of neutral particles. This current produces magnetic field perturbations, which are measured by four observatories.

For illustration, Figure 20.3 (top panel) shows the atmospheric density derived from the drag experienced by the Explorer 9 satellite during the period June to October, 1961, and reduced to an altitude of 730 km. This was a period of medium solar activity. The satellite was launched on February 16, 1961, by a Scout-X1 rocket into a 38.6° inclination orbit with perigee and apogee altitudes of 634 km and 2583 km, respectively. The satellite was in fact a 3.7 m diameter balloon used for atmospheric density studies. The drag was determined from precise satellite position measurement on photographs taken with Baker-Nunn cameras. The second panel from top shows the variation of the exospheric temperature, T_∞ , derived from the atmospheric density. In the other

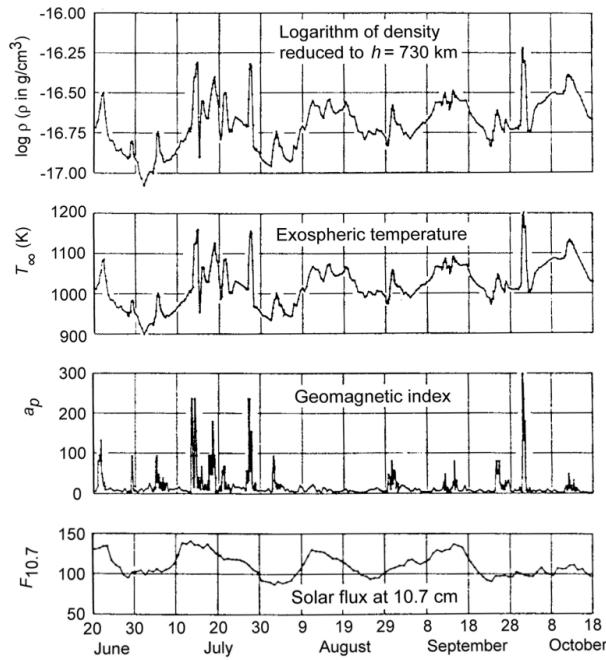


Figure 20.3: Effect of solar activity on atmospheric density at an altitude of 730 km, as deduced from the tracking of Explorer 9 in 1961. Initial orbit: $h_p = 634 \text{ km}$, $h_a = 2583 \text{ km}$, $i = 38.6^\circ$. [copied from US Standard Atmosphere Supplements, 1966, US Government Printing Office, 1966.]

² 1 solar flux unit (sfu) = $10^{-22} \text{ W m}^{-2} \text{ Hz}^{-1}$; 1 Jansky (Jy) = $10^{-26} \text{ W m}^{-2} \text{ Hz}^{-1}$.

³ 1 tesla (T) = 1 Wb m⁻² = 1 V m⁻² s = 1 N A⁻¹ m⁻¹.

two panels the actual variations of $F_{10.7}$ and a_p are plotted. Note the strong correlation of the variations of atmospheric density and exospheric temperature with the variations of $F_{10.7}$ and a_p .

Solar radiation at ultraviolet and shorter wavelengths leads to dissociation of atmospheric gas molecules into atoms and to ionization of these atoms. Already at an altitude of 80 km free electrons and ions can exist for some period of time. Figure 20.4 shows the number density of individual atmospheric constituents as a function of altitude, for a low level ($T_\infty = 600$ K) and a high level ($T_\infty = 2000$ K) of solar activity. This Figure clearly shows that at high altitudes the atmosphere primarily consists of gasses with a low atomic mass. Figure 20.5 shows the variation of atmospheric pressure, density and mean molecular weight as a function of altitude for a minimum and a maximum solar activity level. Note that in periods of maximum solar activity the atmospheric density at a given altitude above 500 km can be more than two hundred times the atmospheric density at that altitude in periods of minimum solar activity. The Figure also shows the decrease of the mean molecular weight with altitude, which is a result of the dominance of atmospheric constituents with a low atomic mass at high altitudes (Figure 20.4).

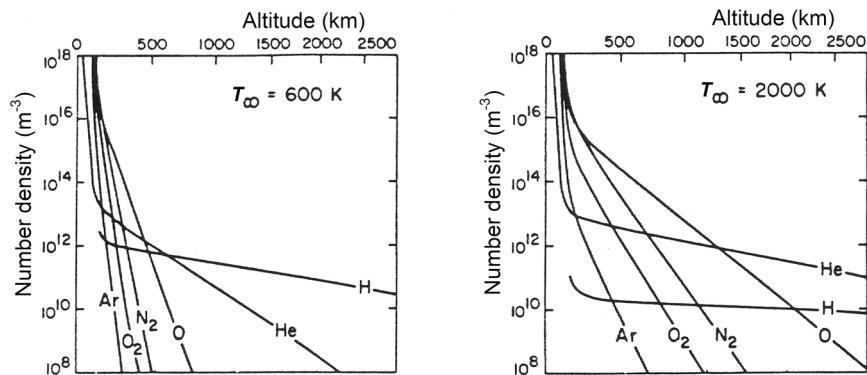


Figure 20.4: Number density of individual atmospheric constituents as a function of altitude for two solar activity levels. $T_\infty = 600$ K corresponds to a low solar activity level; $T_\infty = 2000$ K to a high solar activity level. [copied from L.G. Jacchia, Thermospheric temperature, density, and composition: New models, SAO Special Report No.375, 1977.]

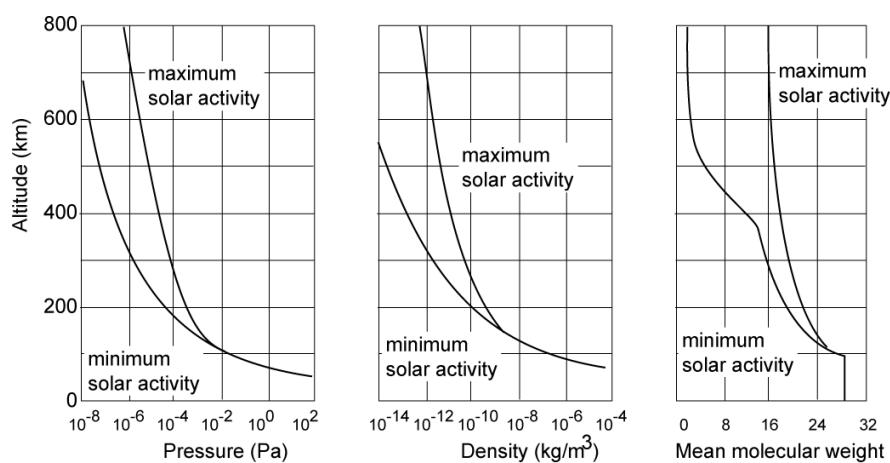


Figure 20.5: Variation of pressure, density and mean molecular weight with altitude, according to the CIRA 1972 Reference Atmosphere (Committee on Space Research, International Council of Scientific Unions, Akademie-Verlag, 1972).

As an example of the effects of atmospheric forces, Figure 20.6 shows the history of three mean orbital elements⁴ of the *Astronomical Netherlands Satellite* (ANS; Section 16.6) over a period of about 700 days, as obtained from tracking data. The decrease of the semi-major axis, a , and eccentricity, e , are caused by atmospheric drag; the decrease of the inclination, i , is the result of a force generated by the rotation of the atmosphere. The oscillation superimposed on the steadily decreasing eccentricity is caused primarily by the J_3 -term of the Earth's gravity field (Section

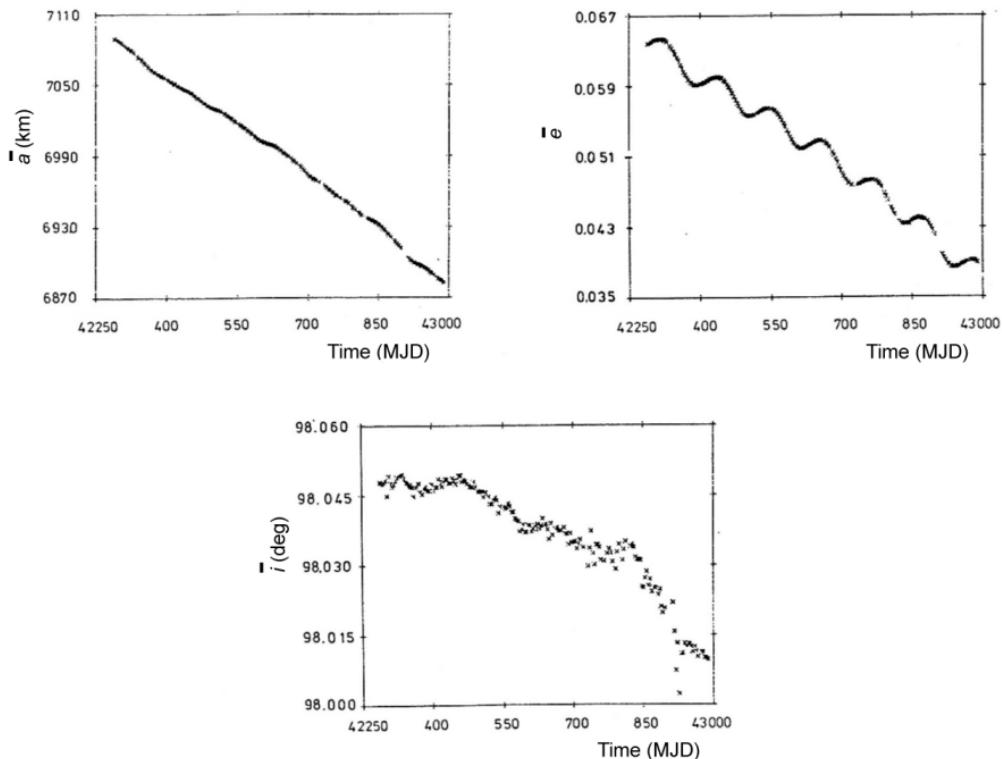


Figure 20.6: Evolution of three mean orbital elements of ANS. Initial orbit: $h_p = 266 \text{ km}$, $h_a = 1175 \text{ km}$, $i = 98.0^\circ$.

20.1), while the peculiar variation of the inclination between MJD 42750 and MJD 42880 is a result of orbital resonance produced by gravity field model terms of order 15 (Section 20.1, Section 23.5). For a numerical example of the relative magnitude of atmospheric drag, let us consider a satellite in a circular orbit at 160 km altitude. Then, the average atmospheric density is $\rho \approx 1.24 \times 10^{-9} \text{ kg/m}^3$ and the satellite's velocity is 7.81 km/s. For a balloon-type satellite with a very large area-to-mass ratio, A/M , we may adopt a value of $C_D A / 2M \approx 13 \text{ m}^2/\text{kg}$. This leads to $f_D \approx 0.98 \text{ m/s}^2$, while the central gravitational acceleration is 9.33 m/s^2 . So, the ratio between the perturbing acceleration and the main acceleration is about 0.1; this ratio will be (much) smaller for 'ordinary' satellites with (much) smaller values of A/M or for satellites at higher altitudes. In fact, atmospheric drag is the most important perturbing force below 200 km altitude, and can, in most cases, completely be neglected at altitudes above 1000 km. Atmospheric drag is the cause of the decay of Earth satellites. Through the friction with the atmosphere, they dissipate their orbital energy and their orbital altitude decreases, through which they enter denser parts of the atmosphere, where the drag becomes larger and larger. Ultimately, they burn up in

⁴ The concept of *mean orbital elements* is introduced in Section 23.3; the concept of *Modified Julian Date* (MJD) in Section 11.4.

the low Earth's atmosphere. Sometimes, parts of the satellite survive the reentry and impact on the Earth's surface.

20.3. Gravitational attraction by other celestial bodies

The presence of other celestial bodies, and in particular of the Sun and the Moon, will also lead to a perturbing force that affects the orbit of a satellite about the Earth. In the discussion on relative motion in the many-body problem (Chapter 4) it was found that the perturbing potential expressing the gravitational perturbations of a number of bodies j on the motion of body i relative to a non-rotating reference frame fixed at body k may be written as ((4.6))

$$\mathbf{R} = -G \sum_{j \neq k, i} \mathbf{m}_j \left(\frac{1}{r_{ij}} - \frac{\bar{\mathbf{r}}_i \cdot \bar{\mathbf{r}}_j}{r_j^3} \right) \quad (20.12)$$

where $\bar{\mathbf{r}}_i$ and $\bar{\mathbf{r}}_j$ are the position vectors from body k to body i and from body k to the perturbing body j , respectively; r_{ij} is the distance between body i and body j . The parameter G denotes the universal gravitational constant and m_j is the mass of the perturbing body. Now, assume that body k is the Earth, body i is a satellite moving about the Earth, and body j is a single perturbing body (Sun, Moon, planet). Then, the perturbing acceleration of the satellite caused by the gravitational attraction between the satellite and the perturbing body may be written as

$$\bar{\mathbf{f}} = -\bar{\nabla} \left[-\mu_j \left(\frac{1}{r_{ij}} - \frac{x_i x_j + y_i y_j + z_i z_j}{r_j^3} \right) \right] \quad (20.13)$$

where $\mu_j = G m_j$. With

$$r_{ij}^2 = (x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2$$

we find from (20.13) for the rectangular components of the perturbing acceleration:

$$f_x = \mu_j \left(\frac{x_j - x_i}{r_{ij}^3} - \frac{x_j}{r_j^3} \right) ; \quad f_y = \mu_j \left(\frac{y_j - y_i}{r_{ij}^3} - \frac{y_j}{r_j^3} \right) ; \quad f_z = \mu_j \left(\frac{z_j - z_i}{r_{ij}^3} - \frac{z_j}{r_j^3} \right) \quad (20.14)$$

The components of the perturbing acceleration can be computed if the positions of the satellite and the disturbing body, both relative to a non-rotating reference frame with origin at the center of the Earth, are known. For the maximum value of the ratio between the magnitude of the perturbing acceleration and the magnitude of the acceleration due to the central force field of the Earth, we may write according to (4.17) with our present notation:

$$\left(\frac{f_d}{f_E} \right)_{max} \approx 2 \frac{m_d}{m_E} \left(\frac{r_i}{r_d} \right)^3 \quad (20.15)$$

where f_d and f_E are the accelerations due to the gravitational attraction by the perturbing body and by the Earth, respectively, and m_d and m_E are the masses of the perturbing body and of the Earth, respectively. Note that the relative perturbing acceleration increases with increasing orbital altitude of the satellite, causing the attractions by Sun and Moon to become the most important perturbing forces above the geostationary altitude.

In Table 4.2, the maximum relative perturbing accelerations of a geostationary satellite are listed for a number of perturbing bodies. The magnitude of the maximum relative perturbing

acceleration due to solar attraction is, according to (20.15), for a satellite at an altitude of 400 km and for a geostationary satellite $6.2 \cdot 10^{-8}$ and $1.5 \cdot 10^{-5}$, respectively. For the corresponding values of the maximum perturbing acceleration we then find $0.54 \mu\text{m}/\text{s}^2$ and $3.3 \mu\text{m}/\text{s}^2$, respectively. In certain cases, the attractions by the Sun and the Moon lead to very severe orbit perturbations. Figure 20.7 shows the evolution of the perigee distance of Explorer 28 as a result of primarily the perturbing forces by Sun and Moon. This satellite was launched by a Delta 31 rocket on May 25, 1965, into a 33.9° inclination orbit with perigee and apogee altitudes of 202 km and 263,900 km, respectively. Both the modeled perigee distance variation as obtained by accounting for all orbit perturbations and the actual variation as determined from tracking data are shown. Due to the perturbing forces, the perigee altitude first increased to about 35,600 km and then decreased again such that the satellite entered the Earth's atmosphere already after only about 3 years.

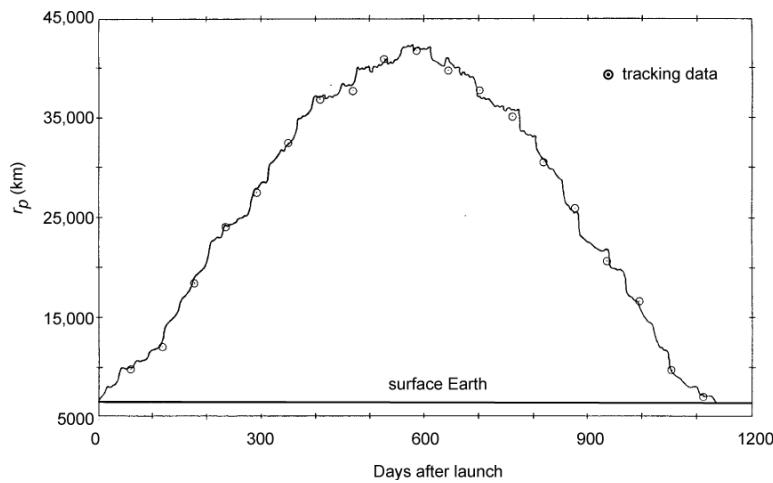


Figure 20.7: Evolution of the perigee distance of Explorer 28. Initial orbit: $h_p = 202 \text{ km}$, $h_a = 263,900 \text{ km}$, $i = 33.9^\circ$.

20.4. Radiation force

An Earth satellite will, in general, experience a radiation force produced by direct sunlight, sunlight reflected by the Earth (*albedo radiation*), and thermal *infrared radiation* emitted by the Earth (sometimes called *terrestrial radiation*). In addition, radiation from the Sun, Earth and internal sources will heat the exposed portions of a spacecraft, and any imbalance in the reradiated heat pattern will generate a *photon thrust*. However, for most satellites the uncertainties of atmospheric drag and solar radiation force tend to mask the effects of this unbalanced thermal radiation force and we will therefore not discuss it further.

In Section 5.11 it was mentioned that the solar radiation force is very difficult to model accurately for satellites with complex shapes. In that Section we have derived the following approximative expression ((5.67)) for the direct solar radiation force acting on a satellite:

$$F = C_R \frac{WA}{c}$$

where W is the energy flux (sometimes called power density) of the incoming solar radiation, A is the effective cross-sectional area of the satellite, c is the speed of light, C_R is the satellite's reflectivity, and $C_R W/c$ is the radiation pressure. For the acceleration of the satellite due to solar radiation we then write

$$\bar{f} = -C_R \frac{WA}{Mc} \bar{e}_s \quad (20.16)$$

where M is the mass of the satellite and \bar{e}_s is the unit vector from the satellite to the Sun. So, the force acts along the Sun-satellite vector. The *solar constant* is defined as the incoming solar electromagnetic radiation flux (at all wavelengths) at a distance of 1 AU (the mean distance of the Earth from the Sun). Its long-term average value was thought to be 1366 W/m^2 , but an analysis of recalibrations of relevant satellite observations has shown in 2011 that a value of 1361 W/m^2 is more realistic. The solar constant has varied by approximately 0.1% over the last three 11-year sunspot cycles. The actual direct solar irradiance close to the Earth, and above the top of the atmosphere, W_s , fluctuates during a year due to the Earth's varying distance from the Sun; from 1408 W/m^2 at perihelion to 1317 W/m^2 at aphelion. This variation can be modeled by

$$W_s = \frac{1361}{1 + 0.0334 \cos(2\pi D_{ap}/365)} \text{ W/m}^2$$

where D_{ap} is the number of days from when the Earth is at aphelion. This aphelion crossing occurs in the period July 3 to July 7 each year; for instance, in the year 2010 it occurred on July 6, 12^h UTC.

Albedo radiation is produced by reflection and scattering of incident solar radiation on the daylight side of the Earth's atmosphere and surface. This type of radiation has essentially the same spectral distribution as direct solar radiation, and is mostly diffusive reflection with only about 1 - 10% of the radiation reflected specularly by the oceans. Its intensity varies widely due to spatial and temporal variations of the reflection and scattering characteristics of the Earth's surface and cloud cover. In fact, the local albedo can vary from 95 W/m^2 over a cloudless ocean to 1250 W/m^2 over large and thick cumulonimbus clouds. The albedo radiation flux near the satellite is also dependent on the Sun-Earth-satellite angle and the altitude of the satellite. Often, a very simple model is used where it is assumed that 32% of the solar radiation is reflected by the Earth, that the albedo radiation is diffusive, that the energy flux just above the clouds (on the daylight side) has a yearly global average value of 435 W/m^2 , and that the flux falls off approximately quadratically with altitude. In contrast to albedo radiation, the Earth's *infrared radiation* is a near isotropic re-emission of direct solar radiation absorbed by the Earth and its atmosphere. The spectrum of the global average thermal radiation approximates that of a black body with an effective temperature of -18° C (255 K), leading to an average energy flux of 240 W/m^2 . It is noted that the Earth's global average surface temperature is about 14° C (287 K); the mechanism that produces the higher surface temperature is due to the atmosphere and is known as the *greenhouse effect*. A temperature of -18° C actually occurs at an altitude of about 5 km in the atmosphere. The local infrared energy flux can vary between 150 W/m^2 and 350 W/m^2 due to spatial and temporal variations of the emission characteristics of the Earth's surface and cloud cover. However, its variation across the globe is less than that of albedo radiation. Often, a simple model is used where it is assumed that the effective Earth's infrared radiation flux just above the cloud cover is 240 W/m^2 , and that the flux falls off approximately quadratically with altitude.

We conclude that the energy flux of direct solar radiation is at least three times the flux of albedo and infrared radiation, and that the albedo and infrared radiation flux decreases with increasing altitude. To compute the albedo and infrared radiation forces acting on a satellite, the method mentioned in Section 5.11 for the computation of the direct solar radiation force may be used. However, it is noticed that the satellite's effective cross-sectional area, A , will, generally, be different for direct solar radiation, albedo radiation and thermal infrared radiation, and that the satellite's reflectivity, C_R , for infrared radiation will be different from the reflectivity for direct

solar radiation and albedo radiation. In this book we will not discuss the albedo and infrared radiation forces any further.

The solar radiation force may produce significant orbit perturbations for very large and light satellites, such as balloon-type satellites like Echo I (1960-I1). This was an aluminized 30.5 m diameter balloon satellite that was launched by a Delta rocket on August 12, 1960, into a 47.2° inclination orbit with perigee and apogee altitudes of 1528 km and 1687 km, respectively. When we adopt for this satellite the values $A/M = 12 \text{ m}^2/\text{kg}$, $C_R = 1.9$, we find for the acceleration due to direct solar radiation $f = 0.10 \text{ mm/s}^2$. This value corresponds to 1.5×10^{-5} times the local gravitational acceleration (6.3 m/s^2) at the orbital altitude of Echo I. Figure 20.8 shows that the perigee altitude of Echo I decreased about 600 km in 140 days, primarily as a result of solar radiation pressure, and thereafter increased again. Both the modeled perigee altitude variation when accounting for all perturbing forces and the actual variation as determined from tracking data are shown.

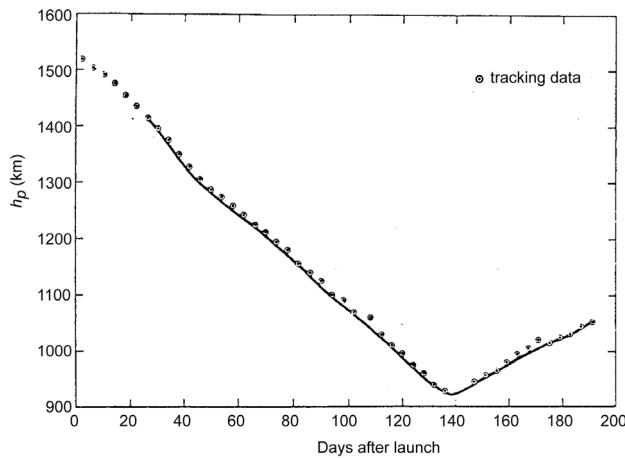


Figure 20.8: Evolution of the perigee altitude of Echo I. Initial orbit: $h_p = 1520 \text{ km}$, $h_a = 1695 \text{ km}$, $i = 47.2^\circ$.

The Sun produces a shadow cone ‘behind’ the Earth with its axis along the Sun-Earth line. A simple geometric analysis, based on the radii of Sun and Earth and the distance between Sun and Earth, shows that this cone extends to a distance of $1.383 \times 10^6 \text{ km}$ from the Earth ‘above’ the dark side of the Earth. Beyond this distance, there is no *umbra* (full shadow) of the Earth and an observer would see the Sun partially obscured by the Earth. This is called the *penumbra* (partial shadow) region. The sub-region of the penumbra from which the Earth appears entirely contained within the disc of the Sun is called the *antumbra* region; an observer in this region experiences an annular eclipse. The Sun-Earth L_2 point (Section 3.7) is located at a distance of about $1.508 \times 10^6 \text{ km}$ from the Earth; at that point the radius of the penumbra region perpendicular to the Sun-Earth line is about 13,446 km. So, a spacecraft positioned at that L_2 point is outside the umbra region but inside the antumbra region; the spacecraft ‘sees’ 84% of the solar disk obscured by the Earth. However, a spacecraft that moves about that point at a distance of more than 13,446 km and in a plane perpendicular to the Sun-Earth line would ‘see’ the full solar disk continuously, which is very useful for the satellite’s power generation. So, it is possible to fly Lissajous or halo orbits (Section 3.10) about the L_2 point for which the spacecraft is continuously in full sunlight.

The Earth’s shadow cone has a top angle of only 0.53° . To simplify the analysis it is therefore often assumed that, for the entire range of Earth satellite orbital altitudes, the solar radiation field can be approximated by a parallel beam of sunlight along the Sun-Earth line with

a constant energy flux (power density), and that the Sun produces a shadow cylinder behind the Earth with a radius equal to the radius of the Earth. When computing the effect of solar radiation pressure on a satellite orbit, we have, of course, to account for the fact that the satellite may periodically pass through the Earth's shadow where the solar radiation force is absent (Section 21.5). It is noted that in our simplified model the entrance into and exit out of the shadow of the Earth are discontinuous processes, while in reality these processes are more gradual due to the existence of a penumbra region.

When we compare the expressions for the satellite's accelerations due to atmospheric drag and due to solar radiation force, we conclude that we may write for the ratio of the magnitudes of both accelerations:

$$\frac{f_{rad}}{f_{atm}} = 2 \frac{C_R}{C_D} \frac{W}{c \rho} \frac{A_{rad}}{A_{atm}} \frac{1}{V^2}$$

where A_{rad} and A_{atm} are the effective cross-sectional areas for the radiation force and the atmospheric drag, respectively. When we assume that $A_{rad} = A_{atm}$, which is e.g. the case for a spherical satellite, and that the satellite moves in a circular orbit with radius r , we can write the relation given above as

$$\frac{f_{rad}}{f_{atm}} = 2 \frac{C_R}{C_D} \frac{W}{c \mu} \frac{r}{\rho}$$

where μ is the gravitational parameter of the Earth. The parameters W , c and μ are constants, while the values of C_R and C_D are known for a particular satellite. At altitudes below 300 km $f_{rad}/f_{atm} \ll 1$. Since the value of ρ decreases with increasing altitude (Figure 20.6), we conclude that the ratio f_{rad}/f_{atm} increases with increasing orbital altitude. In fact, f_{rad} becomes larger than f_{atm} for orbital altitudes of more than 400 - 1600 km, depending on the level of solar activity and the precise values of C_D and C_R .

20.5. Electromagnetic force

At high altitudes the Earth's atmosphere is partly ionized. Because at a particular altitude the mean thermal velocity of electrons (≈ 200 km/s at 1000 km altitude) is much larger than the mean thermal velocity of ions (≈ 1 km/s at 1000 km altitude), the satellite's surface will acquire a negative electrical potential of several volts to several tens of volts with respect to its plasma environment. But there are also other electric charging mechanisms. In sunlight, high-energy photons strike the satellite's surface and this may lead to surface degradation resulting in a positive electrical surface potential of several tens of volts. When a satellite crosses atmospheric regions of high concentrations of high-energy positively or negatively charged particles its surface may be charged to positive or negative potentials of several tens to thousands of volts. If the satellite structure is electrically conductive, the entire satellite will be charged to the same potential. Finally, there is internal charging, i.e. the buildup of charge on and within dielectric materials or on insulated floating conductors inside the satellite. All these effects together may result in an overall positive or negative potential of the satellite, which potential may vary significantly during an orbital revolution. An electrically charged satellite interacts with the Earth's magnetic field, which results in a *Lorentz force* (H.A. Lorentz; 1853-1928) acting on the satellite. This force can be expressed by

$$\bar{F} = q \bar{v} \times \bar{B} \quad (20.17)$$

where q is the electric charge of the satellite, \bar{v} is the satellite's velocity relative to the magnetic field, and \bar{B} is the magnetic induction (also called magnetic field, magnetic field vector or magnetic flux density) of the geomagnetic field. So, the Lorentz force is always directed perpendicular to both the velocity of the satellite and the magnetic field vector.

Already in 1600, W. Gilbert (also: Gilberde or Gylberde; 1544-1603) recognized that the Earth's magnetic field is similar to that of a dipole. C.F. Gauss (1777-1855) studied the magnetic field in detail and proved in 1838 that 95% of the geomagnetic field is of internal origin. Nowadays, we know that the geomagnetic field extends from the Earth's inner core to where it meets the solar plasma, a stream of energetic charged particles emanating from the Sun and called the *solar wind*. During quiet solar activity periods the point where the solar wind dynamic pressure and the pressure exerted by the Earth's magnetic field are in balance is at about 12 Earth radii in the sunward direction along the Earth-Sun line; during high solar activity levels this distance may decrease to about 9 Earth radii. The geomagnetic field forms a cavity in the shape of a blunt body with its 'nose' at this position and an extended 'tail'; this region of the atmosphere is called the *magnetosphere*. The cavity flares to about 15 Earth radii on either side of the Earth above the dawn-dusk terminator, and the solar wind stretches the terrestrial magnetic field for perhaps 1000 Earth radii in the antisolar direction. In the magnetosphere there exist various regions of high concentrations of energetic charged particles and the motion of these particles is governed by the magnetic field.

The geomagnetic field consists of three components: 1) *main magnetic field* that is generated by electric currents in the Earth's outer core due to the motion of molten iron alloys in the outer core; 2) *external magnetic field* that is produced by the motion of ionized particles and electrons in the upper atmosphere; 3) *crustal magnetic field* that is the result of permanent magnetization of rocks (and other magnetic materials) in the Earth's crust and upper mantle or by magnetization of those rocks induced by the main field and the external field. The main magnetic field, with an intensity at the Earth's surface ranging from about 24,000 nT over Brazil to maxima of about 65,000 nT over northern Canada, Siberia and the coast of Antarctica south of Australia, is steady on time scales of days, but is variable over time scales of years. The intensity of the crustal field at the Earth's surface is generally less than 200 nT and may be regarded as essentially constant, but there are locations where much larger intensities exist. The external field is highly variable; during solar storms field intensity variations at the Earth's surface of more than 1,000 nT on time scales of seconds to hours may occur. Although irregular, the external field variations exhibit a correlation with the 11-year solar cycle and with the 27-day recurrence of some solar storms related to the 27-day rotation period of the Sun as seen from Earth. In addition, the external field shows smaller and more-regular diurnal and seasonal variations associated with variations of the position of the Sun relative to a particular geographic location. Below, we will only consider the main magnetic field.

The *magnetic poles*, also called *dip poles*, are defined as positions on the Earth's surface where the magnetic field is vertical, i.e. perpendicular to the (mean) surface of the Earth; these poles are not necessarily antipodal. The intensity of the main magnetic field at a particular location and the locations of the magnetic poles slowly vary; the geomagnetic intensity has declined over the last 2,000 years almost continuously from a maximum of 35% above the current value. Over the last 400 years the intensity has monotonically decreased by about 20%; over the last 150 years it has decreased by 10 - 15%, and this decrease has further accelerated in the past several years. The Earth's magnetic north pole was drifting from northern Canada towards Siberia at a rate of 10 km/yr at the beginning of the twentieth century, 40 km/yr in 2003

and 55 km/yr in 2009. In 2010, the magnetic north pole was situated at 85° N, 133° W; the magnetic south pole was situated at 64° S, 137° E. Moreover, the poles periodically reverse their orientation in a process called *geomagnetic reversal*. Reversals occur at intervals ranging from less than 0.1 million years to as much as 50 million years; a reversal happens over hundreds or thousands of years and is certainly not exactly a clean ‘back flip’. The most recent reversal occurred about 780,000 years ago. Today, the magnetic pole that is in the northern hemisphere actually is a south pole; it attracts the north pole of a magnet.

The main magnetic field is connected to the Earth’s interior and rotates with the Earth. So, it can very well be described relative to the geocentric rotating reference frame that we also have used for describing the Earth’s gravity field (Section 20.1). In magnetic models the local magnetic induction (magnetic field), \bar{B} , above the Earth’s surface is written as the negative gradient of a geomagnetic potential, which is expressed by a spherical harmonics series expansion similar to the series expansion used for describing the gravity field,

$$\begin{aligned} \bar{B} &= -\bar{\nabla}U \\ U &= R \left[\sum_{n=1}^{\infty} \sum_{m=0}^n \left(\frac{R}{r} \right)^{n+1} P_{n,m}(\sin\phi) \{ g_{n,m} \cos m\Lambda + h_{n,m} \sin m\Lambda \} \right] \end{aligned} \quad (20.18)$$

where r is the distance from the center of the Earth, Λ is the geographic longitude measured eastward from the Greenwich meridian, ϕ is the geocentric latitude, $P_{n,m}(\sin\phi)$ is the Schmidt semi-normalized associated Legendre function of degree n and order m with argument $\sin\phi$, and $g_{n,m}$ and $h_{n,m}$ are the *Gauss geomagnetic coefficients*. It is noticed that most authors use $\cos\Phi$ as the argument of the associated Legendre functions, where Φ is the *geocentric co-latitude*. This angle is measured along a meridian from the north pole; at the north pole $\Phi = 0^\circ$, at the equator $\Phi = 90^\circ$, at the south pole $\Phi = 180^\circ$. Since $\Phi = 90^\circ - \phi$, we find $\cos\Phi = \sin\phi$, which justifies the use of $\sin\phi$ as argument of the associated Legendre functions in (20.18). It is emphasized that for the geomagnetic field we usually take the mean Earth radius ($R = 6371.2$ km) as scaling factor, while for the Earth’s gravity field we take the mean equatorial Earth radius ($R = 6378.14$ km) as scaling factor. The values of B , $g_{n,m}$ and $h_{n,m}$ are generally reported in nanotesla. Because the main magnetic field slowly varies, the values of the Gauss geomagnetic coefficients also slowly change with time. However, over a period of less than five years, these coefficients may be considered as ‘semi-constant’. Therefore, geomagnetic models often provide the values of the coefficients per five-year interval. Modern geomagnetic field models contain about 200 Gauss coefficients.

The absolute value of the first Gauss coefficient, $g_{1,0}$, is at least six times the absolute value of all other coefficients. This means that in first approximation the magnetic field can be described by a dipole field. The coefficients $g_{1,0}$, $g_{1,1}$ and $h_{1,1}$ together determine the orientation of the dipole axis in the Earth-fixed reference frame. Therefore, in first approximation, the main geomagnetic field can be modeled as a dipole, placed at the center of the Earth and with its axis tilted with respect to the Earth’s axis of rotation. This dipole approximation is reasonably accurate up to an altitude of about 20,000 km; above that altitude the geomagnetic field is significantly distorted by the interaction with the magnetized plasma flowing from the Sun. The poles defined by the dipole model are generally called *geomagnetic poles*; the geomagnetic poles are antipodal. The co-latitude of the geomagnetic north pole has varied from a low of 3.1° in 1550 to $11.5^\circ - 11.8^\circ$ in the period 1880–1960, and is since then decreasing again at a rate of about 0.025° per year. In 2010, the geomagnetic north pole was situated at 80.1° N, 72.2° W; so, its co-latitude was 9.9° . For our analyses, we further simplify the dipole model by ignoring

the tilt and thus assume that the dipole is oriented along the Earth's rotation axis. Then, the geomagnetic field is axially symmetric about the Earth's rotation axis, and $g_{1,0} \approx -29,496$ nT (2010) is the only non-zero Gauss coefficient. For $n = 1, m = 0$, the Schmidt semi-normalized associated Legendre function is identical to the non-normalized associated Legendre function and we then find from (20.2-1): $P_{1,0}(\sin\phi) = \sin\phi$. So, the geomagnetic potential of the non-tilted dipole field can be expressed as

$$U = g_{1,0} \frac{R^3}{r^2} \sin\phi \quad (20.19)$$

From this relation we find for the radial and latitudinal components of $\bar{\mathbf{B}}$, and for the magnitude of \mathbf{B} :

$$\begin{aligned} B_r &= 2g_{1,0} \left(\frac{R}{r} \right)^3 \sin\phi \quad ; \quad B_\phi = -g_{1,0} \left(\frac{R}{r} \right)^3 \cos\phi \\ B &= \sqrt{B_r^2 + B_\phi^2} = g_{1,0} \left(\frac{R}{r} \right)^3 \sqrt{1 + 3 \sin^2\phi} \end{aligned}$$

From these expressions we conclude that: 1) the magnetic dipole field falls off as the cube of the distance from the Earth's center; 2) $B_\phi = 0$ at the north and south poles; 3) \mathbf{B} is directed downward (inward) at the north pole and upward (outward) at the south pole; 4) $B_r = 0$ at the equator; 5) at the equator \mathbf{B} is directed parallel to the vector from the Earth's center to the north pole; 6) for a particular altitude the magnetic field at the poles is twice as strong as at the equator. On the Earth's surface ($r = R$), the dipole field at the equator was 29,496 nT and at the poles 58,992 nT in 2010.

Combination of (20.17) to (20.19) gives for the acceleration of the satellite due to the Lorentz force:

$$\bar{\mathbf{f}} = -\frac{q}{M} g_{1,0} R^3 \bar{\mathbf{v}} \times \bar{\nabla} \left(\frac{\sin\phi}{r^2} \right) \quad (20.20)$$

where M is the mass of the satellite. The vector $\bar{\mathbf{f}}$, obviously, depends on the position and velocity of the satellite in its motion about the Earth. For an order-of-magnitude analysis, we now assume that the satellite moves in a circular equatorial orbit with radius $r < 8,000$ km. Then, $\mathbf{v} \approx V_c \hat{\mathbf{e}}_r$, where V_c is the circular velocity relative to the geocentric non-rotating reference frame. In that case, (20.20) can be written approximately as

$$\bar{\mathbf{f}} = -\frac{q}{M} g_{1,0} \left(\frac{R}{r} \right)^3 V_c \hat{\mathbf{e}}_r \quad (20.21)$$

where $\hat{\mathbf{e}}_r$ is the unit vector in the radial direction. Realizing that $g_{1,0} < 0$, we conclude that for a satellite with a negative electrical potential the Lorentz force is pointing continuously towards the center of the Earth, and for a satellite with a positive potential always radially away from the center of the Earth. Now, we may write

$$q = U^* C \quad (20.22)$$

where U^* is the electrical potential difference between the satellite and the surrounding plasma, and C is the satellite's capacitance. The capacitance of a spherical conducting satellite is given by

$$C = 4\pi \epsilon_0 R_s \quad (20.23)$$

where C is expressed in farad (F)⁵, $\epsilon_0 \approx 8.854 \cdot 10^{-12}$ F/m is the electric constant (also called permittivity of free space) and R_s is the satellite radius. Substitution of (6.18), (20.22) and (20.23) into (20.21) then leads to

$$\bar{f} \approx -4\pi \epsilon_0 g_{1,0} \frac{U^* R_s}{M} \left(\frac{R}{r} \right)^3 \sqrt{\frac{\mu}{r}} \bar{e}_r \quad (20.24)$$

where μ is the gravitational parameter of the Earth. With the classical relation for the local gravitational acceleration, $g = \mu/r^2$, we find for the magnitude of the relative acceleration from (20.24):

$$\frac{|\bar{f}|}{g} \approx 4\pi \epsilon_0 g_{1,0} \frac{U^* R_s}{M} \frac{R^3}{\sqrt{\mu r^3}} \quad (20.25)$$

Note that, for a given satellite with a given electrical potential, both the perturbing acceleration and the relative perturbing acceleration decrease with increasing altitude; the acceleration decreases with distance by a factor of $r^{7/2}$, the relative acceleration by a factor of $r^{3/2}$. When we adopt a rather large value of $R_s/M = 0.3$ m/kg and $U^* = \pm 100$ V, we obtain from (20.24) and (20.25) for an orbital altitude of 500 km: $|\bar{f}| = 6.0 \cdot 10^{-10}$ m/s² and $|\bar{f}|/g = 7.1 \cdot 10^{-11}$. These numbers demonstrate that, under normal conditions, electromagnetic forces produce only very small orbit perturbations.

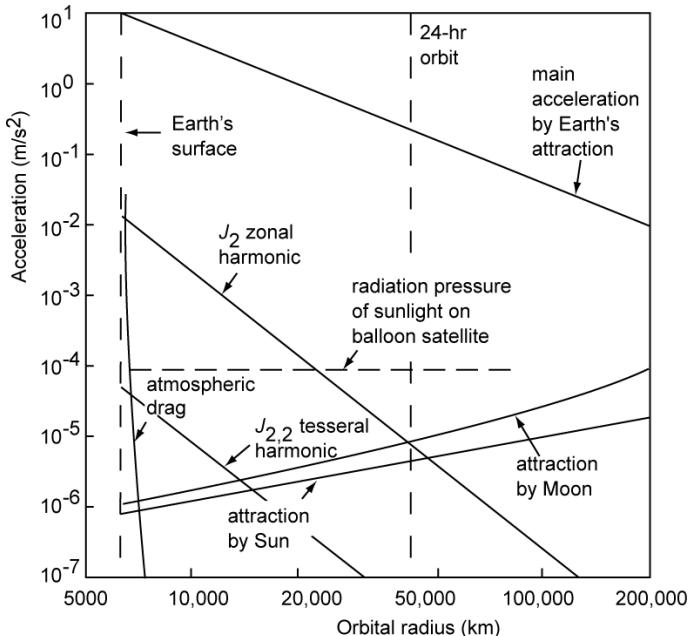


Figure 20.9: Sketch of the magnitude of some types of perturbing forces, as a function of the orbital radius.

In addition to the forces mentioned above, there are other perturbing forces acting on a satellite; however, these will not be discussed here. Figure 20.9 presents a schematic picture of the magnitude of the acceleration due to the most important perturbing forces as a function of the

⁵ 1 farad (F) = 1 C V⁻¹ = 1 A V⁻¹ s = 1 N⁻¹ A² m⁻¹ s².

orbital radius. Note that for low orbits only the perturbations due to the J_2 -term in the Earth's gravitational potential and atmospheric drag play an important role. For geostationary satellites, the attracting forces of Sun and Moon, and possibly also the direct solar radiation force, become very significant.

It is emphasized that the character of the perturbing accelerations produced by the Earth's gravity field and by the Sun and the Moon differs significantly from the character of the perturbing accelerations produced by the atmospheric, radiation and electromagnetic forces. The accelerations due to gravitational forces are independent of the mass of the satellite, while the accelerations resulting from the last three forces mentioned are inversely proportional to the satellite's mass. As a consequence, the Earth and an Earth satellite experience almost the same acceleration due to the gravitational attraction by the Sun and the Moon, and it is the differential acceleration that counts for the perturbed motion of the satellite about the Earth (Section 4.1). However, for the perturbed motion of a satellite due to solar radiation pressure we may neglect the acceleration of the Earth due to this radiation force. The reason is that the term A/M in (20.16) is for the Earth about $2 \cdot 10^{-11} \text{ m}^2/\text{kg}$, which is many orders of magnitude smaller than for a satellite.

In the following, we will assume that the perturbing forces acting on a satellite are known and we will discuss some methods that can be used to compute their effects on the orbit of the satellite.

20.6. Special and general perturbations methods

When the motion of a satellite is described relative to a non-rotating geocentric reference frame and perturbing forces are taken into account, then the equation of motion can be written in the form (Chapter 10)

$$\frac{d^2\bar{r}}{dt^2} + \frac{\mu}{r^3}\bar{r} = -\bar{\nabla}R + \bar{f} \quad (20.26)$$

where R , the *perturbing or disturbing potential*, describes all perturbing forces that can be expressed by a potential function, and \bar{f} stands for all perturbing forces that cannot be written as the gradient of a scalar function of the satellite's coordinates. In general, (20.26) cannot be solved analytically and we have to resort to numerical integration techniques or to approximative analytical methods. In the first case, we speak of *special perturbations methods*; in the second case, we deal with *general perturbations methods*.

Special perturbations methods generate just one special trajectory for a particular satellite, given specific initial conditions. The computation of the orbit is done by means of a numerical step-by-step integration process. The various methods of special perturbations are usually classified according to the formulation of the equations to be integrated. The three classical methods are known as *Cowell's method*, *Encke's method* and the *method of variation of orbital elements*. These methods will be described below. An important topic in special perturbations methods is the choice of the most suitable numerical integration technique. This highly specialized subject is, however, beyond the scope of this book.

General perturbations methods cover analytical methods in which the perturbing accelerations are expanded into series and integrated analytically term by term. In practice, only a limited number of terms of the series is taken into account. The solution of the differential equations of motion is obtained in the form of analytical expressions, describing the approximate change of the orbit as a function of time for a particular perturbing force. These methods thus yield solu-

tions that are generally applicable to all satellites for all initial conditions. The most important classical general perturbations method is the *method of variation of orbital elements*, which was already mentioned in connection with special perturbations methods. For the analytical integration, many techniques are known; for example: the *method of successive approximations* (Section 5.10), *Taylor series expansions*, *multi-variable asymptotic expansions* and *averaging* (Section 23.3). The main problem of all general perturbations methods always was the tremendous amount of analytical labor involved, especially when higher-order approximations are required. Nowadays, most of the analysis associated with the development of higher-order theories can be performed by a computer.

The reader should be aware that, in addition to the special and general perturbations methods mentioned, there exist numerous other specialized, and in most cases very complicated, methods. Comparing the special and general perturbations methods, some remarks can be made about their advantages and disadvantages. Obviously, the methods of special perturbations are directly applicable to any orbit and to any perturbing force. As these are purely numerical methods, we have the problem of accumulated integration errors (Section 10.1) and these methods are not well suited to long-term orbit predictions. Moreover, these methods require the computation of the coordinates and velocity components of the satellite at all intermediate epochs prior to the epoch of interest, leading to long integration times. An advantage of general perturbations methods is that once the analytical expressions that describe the effects of the perturbing force are available, the computation of a perturbed orbit for various initial conditions is much faster. Their main advantage, however, is that general perturbations methods, in particular the method of variation of orbital elements, reveal the source of perturbation from orbital data. Major achievements of the methods of general perturbations were the discovery of the planet Neptune in 1846 by J.G. Galle (1812-1910) at the position predicted by U.J.J. Le Verrier (1811-1877) on basis of his analysis of the observed orbit perturbations of Uranus; and the demonstration of the Earth's *pear shape* in 1959, through the analysis of the observed orbit perturbations of the Vanguard 1 satellite, by J.A. O'Keefe (1916-2000).

Method of Cowell

The simplest method for the computation of perturbed satellite orbits is the *method of Cowell*. Although this name originally refers to a certain integration scheme in rectangular coordinates, it nowadays refers to all direct numerical integration procedures for the equations of motion of celestial bodies. In this method, we write these equations in the form

$$\frac{d^2\bar{r}}{dt^2} = \bar{f}_t \quad (20.27)$$

where the total acceleration, \bar{f}_t , is given by (20.26)

$$\bar{f}_t = -\frac{\mu}{r^3}\bar{r} - \bar{\nabla}R + \bar{f} \quad (20.28)$$

This technique was developed by P.H. Cowell (1870-1949) at the beginning of the twentieth century in order to compute the orbit of the eighth satellite of Jupiter. Later, it was applied by Cowell and A.C. de la Cherois Crommelin (1865-1939) to compute the periodic return of the comet of Halley during the period 1759 through 1910. It is obvious that this method can always be applied, even when f is not a perturbing acceleration, but, for instance, is the result of a large force produced by a rocket engine. This method has the important advantage that, because of the simple formulas, it can easily be programmed and it always works; no 'clever' assumptions and

approximations have been made and used, which, in certain cases, may lead to problems. However, the method has a fundamental drawback: one does not make use of the fact that one knows that $\bar{\nabla}R$ and \bar{f} are perturbing accelerations, which means that the orbit, in first approximation, resembles a conic section. Because in this method the integration has to account for both the variation of the satellite's state vector components produced by the central Newtonian gravity field (Keplerian orbit) and the orbit perturbations produced by the various perturbing forces, small integration steps have to be applied in the numerical integration process. This implies long computation times and a steadily growing numerical integration error (Section 10.1).

Method of Encke

The *method of Encke*, which was introduced by J.F. Encke (1791-1865) in 1852, eliminates some drawbacks of Cowell's method. It makes use of a reference orbit and only the deviations from that orbit are integrated numerically. For simplicity, we will assume that this reference orbit is a Keplerian one, but this restriction is by no means a necessity. For the analysis we start from (20.26), and write for the Keplerian reference orbit

$$\frac{d^2\bar{\rho}}{dt^2} + \frac{\mu}{\rho^3}\bar{\rho} = 0 \quad (20.29)$$

where $\bar{\rho}$ denotes the position vector of the satellite if the satellite would follow the unperturbed reference orbit. We now assume that at some instant of time the following conditions hold:

$$t = t_0 : \quad \bar{r} = \bar{\rho} \quad ; \quad \frac{d\bar{r}}{dt} = \frac{d\bar{\rho}}{dt}$$

Because the reference orbit is a Keplerian one, we may solve (20.29) analytically to obtain position and velocity in the reference orbit at any time. For the deviation of the actual trajectory from the reference orbit at a time t , we write

$$\Delta\bar{r} = \bar{r} - \bar{\rho} \quad (20.30)$$

By differentiating this relation twice with respect to time and subsequent substitution of (20.26) and (20.29), we obtain

$$\frac{d^2\Delta\bar{r}}{dt^2} = \mu \left(\frac{\bar{\rho}}{\rho^3} - \frac{\bar{r}}{r^3} \right) - \bar{\nabla}R + \bar{f} \quad (20.31)$$

Integration of this equation leads to the deviations of position and velocity relative to their values in the reference orbit. The latter are computed analytically, and the actual position and velocity at a time t are obtained by adding $\bar{\rho}$ and $\Delta\bar{r}$, and $d\bar{\rho}/dt$ and $d\Delta\bar{r}/dt$, respectively.

A practical numerical difficulty arises from the form of the first term on the right-hand side of (20.31). This term is the difference of two nearly equal and small quantities, which is always undesirable in numerical computations. Therefore, (20.31) is mostly written in another form. First, we transform this equation into

$$\frac{d^2\Delta\bar{r}}{dt^2} = \frac{\mu}{\rho^3} \left[\bar{\rho} - \bar{r} + \bar{r} \left(1 - \frac{\rho^3}{r^3} \right) \right] - \bar{\nabla}R + \bar{f} \quad (20.32)$$

We then write

$$\frac{r^2}{\rho^2} = \frac{(\bar{\rho} + \Delta\bar{r}) \cdot (\bar{\rho} + \Delta\bar{r})}{\rho^2} \quad (20.33)$$

Defining a quantity q as

$$q = \frac{\Delta\bar{r} \cdot (\bar{\rho} + \frac{1}{2}\Delta\bar{r})}{\rho^2} \quad (20.34)$$

(20.33) may also be written as

$$\frac{r^2}{\rho^2} = 1 + 2q \quad (20.35)$$

Using this relation, we find

$$1 - \frac{\rho^3}{r^3} = 1 - (1 + 2q)^{-3/2} = \mathcal{F}(q) \quad (20.36)$$

where $\mathcal{F}(q)$ indicates that \mathcal{F} is a function of q . The term $\mathcal{F}(q)$ can be expanded into a binomial series, as was first done by Encke:

$$\mathcal{F}(q) = 3q \left(1 - \frac{5}{2}q + \frac{35}{6}q^2 - \frac{105}{8}q^3 + \dots \right) \quad (20.37)$$

but it can also be written in a closed form

$$\mathcal{F}(q) = \frac{2q}{1+2q} \left[1 + \frac{1}{1+2q+\sqrt{1+2q}} \right] \quad (20.38)$$

As, according to (20.35), q is small, we conclude that $\mathcal{F}(q) \approx 3q$. By substituting (20.30) and (20.36) into (20.32), we finally obtain

$$\frac{d^2\Delta\bar{r}}{dt^2} = \frac{\mu}{\rho^3} [(\bar{\rho} + \Delta\bar{r}) \mathcal{F}(q) - \Delta\bar{r}] - \bar{\nabla}R + \bar{f} \quad (20.39)$$

where the term $\mathcal{F}(q)$ should be computed either from (20.37) or from (20.38). This form of the equation of relative motion is mostly used when Encke's method is applied.

As in Encke's method only the perturbing accelerations are integrated numerically to obtain deviations from a reference orbit, usually the integration step can be chosen larger than in Cowell's method. However, at each integration step Encke's method involves more computing time. Nevertheless, for small but strongly varying perturbing forces Encke's method yields in most cases a more-efficient computing process than the method of Cowell. If the perturbations accumulate, eventually $\Delta\bar{r}$ may become large and the Encke formulation loses its advantage over the simpler Cowell formulation. Then, the reference orbit should be *rectified*; i.e. a new reference orbit must be selected, such that at the instant of rectification again $\Delta\bar{r} = 0$ and $d\Delta\bar{r}/dt = 0$. As a very rough criterion for rectification we may use $q = 0.01$. This need for rectification is a drawback of Encke's method.

Variation of orbital elements

A third method for the computation of perturbed satellite orbits is the *method of variation of orbital elements*, which was published by J.L. Lagrange (1736-1813) in 1808. In this method,

which is also called the *method of variation of parameters* or the *method of variation of constants*, the actual perturbed orbit is regarded as a continuous sequence of Keplerian orbits, each of which is at a certain point of the perturbed orbit tangent to that orbit. This very important method will be treated in Chapter 22.

20.7. Historical development of the theory of orbit perturbations

The theory of orbit perturbations, as applied to lunar motion, was developed from a geometric standpoint by I. Newton (1643-1727). The memoirs of A.C. Clairaut (1713-1765) and J.B. le Rond d'Alembert (1717-1783) contain important advances, making the solutions depend upon the integration of the differential equations in series. Clairaut soon had occasion to apply his processes of integration to the orbit perturbations of Halley's comet by the planets Jupiter and Saturn. This comet had been observed in 1531, 1607 and 1682. If its period were constant, it would pass its perihelion again about the middle of 1759. Clairaut computed the perturbations of the comet's trajectory and found that the perihelion passage would be April 13, 1759. He remarked that the time was uncertain to the extent of a month, because of the uncertainties in the masses of Jupiter and Saturn and the possibility of perturbations from unknown planets beyond these two. The comet passed its perihelion by March 13, giving a striking proof of the value of Clairaut's method.

The theory of the perturbations of the motion of planets was begun by L. Euler (1707-1783), whose memoirs on the mutual perturbations of Jupiter and Saturn gained the prizes of the French Academy of Sciences in 1748 and 1752. In these memoirs the first analytical development of the method of variation of orbital elements (Chapter 22) was given. The equations were not entirely general, as he had not considered the elements as being all simultaneously variables. The first steps in the development of the perturbing function were also given by Euler. Lagrange wrote his first memoir in 1766 on the perturbations of Jupiter and Saturn. In this work he developed still further the method of variation of orbital elements, leaving his final equations, however, still incorrect by regarding the major axis and the epoch of perihelion passage as constants in deriving the equations for the variations. The equations for the inclination, longitude of the node, and longitude of the perihelion from the node were correct. In the expressions for the mean longitudes of the planets there were terms proportional to the first- and second-order powers of time. These were entirely due to the imperfections of the method, their true form being that of long-period terms, as was shown by P.S. Laplace (1749-1827) in 1784 by considering terms of the third order in the eccentricities. The method of variation of orbital elements was completely developed for the first time in 1782 by Lagrange in a prize memoir on the perturbations of comets moving in elliptical orbits. By far the most extensive use of this method is due to C.E. Delaunay (1816-1872), whose lunar theory is essentially a long succession of the application of the process, each step of it removing a term from the perturbation function. J.C.F. Gauss (1777-1855), P.A. Hansen (1795-1874), G.B. Airy (1801-1892), U.J.J. Le Verrier, J.C. Adams (1819-1892), S. Newcomb (1835-1909), G.W. Hill (1838-1914), and many others have made further important contributions to the theory of planetary motion. Adams and Le Verrier are noteworthy for having computed in 1843 and 1846, independently of each other, the existence and apparent position of Neptune from the unexplained irregularities in the motion of Uranus. It should be mentioned here that Le Verrier predicted the position of Neptune much more accurate than Adams did. J.H. Poincaré (1854-1912) applied to the problem all the resources of modern mathematics with unrivaled genius; he brought into the investigation such a wealth of ideas, and he devised methods of such immense power that the subject in its theoretical aspects has been entirely revolutionized in his hands.

The method of variation of orbital elements was commonly applied by astronomers up to the middle of the nineteenth century for the computation of perturbed planetary orbits. In the nineteenth century, more and more planetoids were discovered for which the orbits had to be computed, and the need arose for a simpler method that could also be used by non-astronomers lacking specialized knowledge. As a result, the complex method of variation of orbital elements lost ground to the method of Encke and later to the very simple method of Cowell. After the introduction of the computer, especially the method of Cowell became popular, because it is by far the easiest one to be programmed on a computer. This method is still popular for the computation of perturbed orbits of spacecraft over not too long time intervals, because it does not demand information about an approximated orbit. Since the 1960s, the developments in the computation of perturbed orbits of Earth satellites got well under way and for the first time one encountered new types of perturbing forces, such as the ones that are generated by the spatial variations of the Earth's gravity field (Section 20.1) and by atmospheric drag (Section 20.2). Moreover, the magnitude and direction of these forces change relatively fast, contrary to the classical perturbing forces in celestial mechanics: the attracting forces by other planets. After all, within about 95 min a satellite traverses the complete perturbing force field; this means that the stepsize during the numerical integration of satellite orbits usually has to be of the order of 1° - 10° along the orbit. For intermediate- and long-term orbit computations, where one considers hundreds to several thousands of revolutions about the Earth, a numerical integration procedure is usually not suited, because of the long computation times and the associated accumulating integration errors. Then, one has to resort to the general perturbations methods, of which Lagrange's method of variation of orbital elements, which will be treated in Chapter 22, is the classical example.

21. ELEMENTARY ANALYSIS OF ORBIT PERTURBATIONS

The major perturbing forces acting on a satellite were introduced and discussed in Chapter 20. This Chapter presents an elementary analysis of some characteristic orbital effects of these forces. For simplicity, we will restrict ourselves to orbits that would be circular if the perturbing force would not act on the satellite.

21.1. Basic equations

The equation of motion of a satellite relative to a non-rotating geocentric equatorial reference frame can, according to (20.26), be written as

$$\frac{d^2\bar{r}}{dt^2} + \frac{\mu}{r^3}\bar{r} = \bar{f} \quad (21.1)$$

where \bar{r} is the position vector of the satellite, μ is the gravitational parameter of the Earth and \bar{f} is the acceleration due to all perturbing forces acting on the satellite. Just as we did in Section 5.1 and Section 19.2, we now take the scalar product of (21.1) and $d\bar{r}/dt$, and the vector product of (21.1) and \bar{r} . We then obtain in our present notation:

$$\begin{aligned} \frac{d}{dt} \left(\frac{1}{2} V^2 - \frac{\mu}{r} \right) &= \frac{d\mathcal{E}}{dt} = \bar{V} \cdot \bar{f} \\ \frac{d}{dt} (\bar{r} \times \bar{V}) &= \frac{d\bar{H}}{dt} = \bar{r} \times \bar{f} \end{aligned} \quad (21.2)$$

where $\bar{V} \equiv d\bar{r}/dt$ is the velocity vector of the satellite, and \mathcal{E} and \bar{H} are the orbital energy and orbital angular momentum of the satellite, both per unit of mass. According to (6.14), we may write

$$\mathcal{E} = -\frac{\mu}{2a} \quad (21.3)$$

where a is the semi-major axis of the orbit. Differentiation of (21.3) to time yields

$$\frac{d\mathcal{E}}{dt} = \frac{\mu}{2a^2} \frac{da}{dt} \quad (21.4)$$

Substitution of this relation into (21.2-1) leads with (21.2-2) to the set of equations:

$$\begin{aligned} \frac{da}{dt} &= \frac{2a^2}{\mu} \bar{V} \cdot \bar{f} \\ \frac{d\bar{H}}{dt} &= \bar{r} \times \bar{f} \end{aligned} \quad (21.5)$$

These relations show that for a perturbed orbit the semi-major axis and angular momentum are, generally, not constant. The rate of change of a or \bar{H} is zero when the perturbing force is directed perpendicular to the velocity vector or along the radius vector, respectively. The rates are maximum when the force is directed along the velocity vector or perpendicular to the radius vector, respectively. These results were already found in Section 1.7 and Section 19.2.

Equations (21.5) form the basis for the analysis in this Chapter. The position vector, \bar{r} , and the velocity vector, \bar{V} , are functions of the true (perturbed) orbital elements; the acceleration, \bar{f} , is, in general, a function of the true (perturbed) position and velocity of the satellite, and so of the true (perturbed) orbital elements. As a first-order approximation, we will express \bar{r} , \bar{V} and \bar{f} in terms of the elements of the *unperturbed* orbit and of the *unperturbed* position angle of the satellite in this orbit, and we will substitute for a on the right-hand side of (21.5-1) its *unperturbed* value. The errors in the computed orbit perturbations thus committed are obviously small, as long as the perturbing accelerations themselves are sufficiently small. This technique is widely used in astrodynamics and is described in more detail in Sections 22.6 and 23.2.

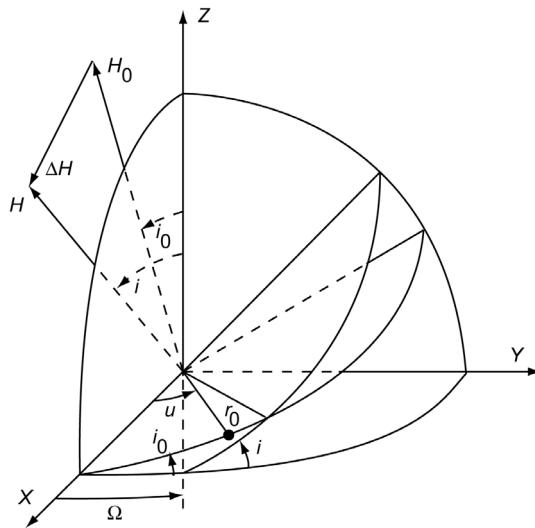


Figure 21.1: Unperturbed and perturbed orbital planes and orbital angular momentum vectors. The unperturbed true anomaly is measured in the unperturbed orbital plane.

Figure 21.1 shows the unperturbed orbital plane and the unperturbed angular momentum vector, H_0 , relative to the non-rotating geocentric equatorial reference frame. For simplicity, the right ascension of the ascending node of the initial orbital plane is chosen zero, i.e. $\Omega_0 = 0^\circ$. As a consequence, H_0 lies in the YZ -plane. Also shown are the perturbed orbital plane and the perturbed angular momentum vector, H . We now select the line of nodes of the unperturbed orbit as the reference line from which the *unperturbed* satellite position angle is measured. This angle is the argument of latitude, u (Section 11.5), and we can write for the unperturbed circular orbit:

$$\frac{du}{dt} = n_0 = \sqrt{\frac{\mu}{r_0^3}} \quad (21.6)$$

where n_0 is the *unperturbed* mean motion of the satellite and r_0 is the radius of the unperturbed circular orbit. With (21.6) we can write (21.5) as

$$\begin{aligned} \frac{da}{du} &= \frac{2}{r_0 n_0^3} \bar{V} \cdot \bar{f} \\ \frac{d\bar{H}}{du} &= \frac{1}{n_0} (\bar{r} \times \bar{f}) \end{aligned} \quad (21.7)$$

or

$$\Delta\alpha = \frac{2}{r_0 n_0^3} \int (\bar{V} \cdot \bar{f}) du \quad ; \quad \Delta\bar{H} = \frac{1}{n_0} \int (\bar{r} \times \bar{f}) du$$

For the integration of these equations we need expressions for \bar{r} and \bar{V} in terms of the elements of the *unperturbed* orbit. For the position components we have

$$x = r_0 \cos u \quad ; \quad y = r_0 \sin u \cos i_0 \quad ; \quad z = r_0 \sin u \sin i_0 \quad (21.8)$$

or

$$\bar{r} = r_0 (\cos u \bar{e}_x + \sin u \cos i_0 \bar{e}_y + \sin u \sin i_0 \bar{e}_z) \quad (21.9)$$

where \bar{e}_x , \bar{e}_y and \bar{e}_z are unit vectors along the axes of the reference frame and i_0 is the inclination of the *unperturbed* orbit. Differentiation of (21.9) to time leads, with (21.6), to

$$\bar{V} = \frac{d\bar{r}}{dt} = r_0 n_0 (-\sin u \bar{e}_x + \cos u \cos i_0 \bar{e}_y + \cos u \sin i_0 \bar{e}_z) \quad (21.10)$$

Now, it will be shown that the rectangular components of $\Delta\bar{H}$ provide information about the changes in inclination, right ascension of the ascending node, and magnitude of the angular momentum. As is shown in Figure 21.1, the angle between \bar{H} and \bar{e}_z is equal to i . Further, the vector $\bar{e}_z \times \bar{H}$ is perpendicular to \bar{e}_z and \bar{H} , and consequently is oriented along the line of nodes and points towards the ascending node of the perturbed orbit. So, we may write

$$\bar{e}_z \cdot \bar{H} = H \cos i \quad ; \quad (\bar{e}_z \times \bar{H}) \times \bar{e}_x = -H \sin i \sin \Omega \bar{e}_z \quad ; \quad (\bar{e}_z \times \bar{H}) \cdot \bar{e}_x = H \sin i \cos \Omega \quad (21.11)$$

where i and Ω are elements of the perturbed orbit. For a given $\Delta\bar{H}$ with components ΔH_x , ΔH_y , ΔH_z along the axes of the reference frame, we may write for the adopted geometry:

$$\bar{H} = (\Delta H_x) \bar{e}_x + (-H_0 \sin i_0 + \Delta H_y) \bar{e}_y + (H_0 \cos i_0 + \Delta H_z) \bar{e}_z \quad (21.12)$$

We now write

$$i = i_0 + \Delta i \quad ; \quad \Omega = \Delta \Omega \quad ; \quad H = H_0 + \Delta H \quad (21.13)$$

If the perturbing force is small, we may assume that also the orbit perturbations Δi , $\Delta \Omega$ and ΔH are small. Substitution of (21.12) and (21.13) into (21.11), and carrying out the vector operations, lead, after neglecting terms which are small of the second order, to

$$\begin{aligned} \Delta H_x &= H_0 \sin i_0 \Delta \Omega \\ \Delta H_y &= -H_0 \cos i_0 \Delta i - \sin i_0 \Delta H \\ \Delta H_z &= -H_0 \sin i_0 \Delta i + \cos i_0 \Delta H \end{aligned}$$

From these relations, we obtain

$$\begin{aligned} \Delta i &= -\frac{\Delta H_y \cos i_0 + \Delta H_z \sin i_0}{H_0} \\ \Delta \Omega &= \frac{\Delta H_x}{H_0 \sin i_0} \end{aligned} \quad (21.14)$$

$$\Delta H = -\Delta H_y \sin i_0 + \Delta H_z \cos i_0 \quad (21.14)$$

with, for the adopted unperturbed circular orbit,

$$H_0 = \sqrt{\mu r_0} = n_0 r_0^2 \quad (21.15)$$

Note that the expression for $\Delta\Omega$ becomes singular for $i_0 = 0^\circ$. This does not introduce a problem since at $i_0 = 0^\circ$ the ascending node is not defined. The equations (21.7), (21.9), (21.10) and (21.14) form the basis for our analysis. In Sections 21.2 to 21.7, we will derive expressions for $V \cdot f$ and $\bar{r} \times f$ for a particular perturbing force. Substitution of these expressions into (21.7) and subsequent integration, for example from $u = 0$ to $u = 2\pi$, yields for that force expressions for Δa , ΔH_x , ΔH_y , ΔH_z . Substitution of the latter three expressions into (21.14) yields expressions for Δi , $\Delta\Omega$, ΔH .

21.2. Perturbations due to the J_2 -term of the gravity field

In Section 20.1 we have found the following expressions for the accelerations in the X -, Y - and Z -directions due to the J_2 -term of the Earth's (static) gravitational potential:

$$\begin{aligned} f_x &= -\frac{3}{2}\mu J_2 \frac{R^2}{r^5} x \left(1 - 5 \frac{z^2}{r^2}\right) \\ f_y &= -\frac{3}{2}\mu J_2 \frac{R^2}{r^5} y \left(1 - 5 \frac{z^2}{r^2}\right) \\ f_z &= -\frac{3}{2}\mu J_2 \frac{R^2}{r^5} z \left(3 - 5 \frac{z^2}{r^2}\right) \end{aligned} \quad (21.16)$$

where R is the mean equatorial Earth radius and J_2 is the second-degree zonal harmonic coefficient.

To illustrate the computation process, the derivation of an expression for Δa will be given in full detail. Substitution of (21.8) into (21.16) gives

$$\begin{aligned} f_x &= -\frac{3}{2}\mu J_2 \frac{R^2}{r_0^4} \cos u \left(1 - 5 \sin^2 u \sin^2 i_0\right) \\ f_y &= -\frac{3}{2}\mu J_2 \frac{R^2}{r_0^4} \sin u \cos i_0 \left(1 - 5 \sin^2 u \sin^2 i_0\right) \\ f_z &= -\frac{3}{2}\mu J_2 \frac{R^2}{r_0^4} \sin u \sin i_0 \left(3 - 5 \sin^2 u \sin^2 i_0\right) \end{aligned} \quad (21.17)$$

From (21.10) and (21.17) we obtain

$$\begin{aligned}\frac{d\bar{r}}{dt} \cdot \bar{f} = & -\frac{3}{2} \mu J_2 n_0 \frac{R^2}{r_0^3} \left[-\sin u \bar{e}_x + \cos u \cos i_0 \bar{e}_y + \cos u \sin i_0 \bar{e}_z \right] \cdot \\ & \cdot \left[\cos u (1 - 5 \sin^2 u \sin^2 i_0) \bar{e}_x + \sin u \cos i_0 (1 - 5 \sin^2 u \sin^2 i_0) \bar{e}_y \right. \\ & \left. + \sin u \sin i_0 (3 - 5 \sin^2 u \sin^2 i_0) \bar{e}_z \right]\end{aligned}$$

or

$$\begin{aligned}\frac{d\bar{r}}{dt} \cdot \bar{f} = & -\frac{3}{2} \mu J_2 n_0 \frac{R^2}{r_0^3} \left[-\sin u \cos u (1 - 5 \sin^2 u \sin^2 i_0) + \right. \\ & \left. \sin u \cos u \cos^2 i_0 (1 - 5 \sin^2 u \sin^2 i_0) + \sin u \cos u \sin^2 i_0 (3 - 5 \sin^2 u \sin^2 i_0) \right]\end{aligned}$$

or

$$\frac{d\bar{r}}{dt} \cdot \bar{f} = -3 \mu J_2 n_0 \frac{R^2}{r_0^3} \sin^2 i_0 \sin u \cos u \quad (21.18)$$

Substitution of (21.18) into (21.7-1) leads with (21.6) to

$$\frac{da}{du} = -6 J_2 \frac{R^2}{r_0} \sin^2 i_0 \sin u \cos u \quad (21.19)$$

Integration from $u = 0$ to $u = u_e$ yields

$$\Delta a = -3 J_2 \frac{R^2}{r_0} \sin^2 i_0 \sin^2 u_e \quad (21.20)$$

Note that Δa is independent of μ and is always negative ($J_2 > 0$) or zero. We conclude that $|\Delta a|$ decreases with decreasing inclination and decreases inversely proportionally to the orbital radius. We also conclude that Δa shows a periodic variation with a period of half the orbital period. In astrodynamics, variations of orbital parameters with periods of the order of the orbital period are called *short-period variations*. Variations that show a linear, quadratic, etc., dependence on time are called *secular variations*. In Section 23.2 these variations, and the so-called *long-period variations*, will be discussed in more detail. Equation (21.20) shows that $\Delta a = 0$ if $u_e = k\pi$, where k is an integer. This means that, to first-order approximation, Δa does not show a long-term trend. So, the J_2 -term of the rotating Earth's gravity field does not produce a secular or long-period change of the satellite's orbital energy.

For a given orbit, the maximum value of $|\Delta a|$ over an orbital revolution is, according to (21.20), given by

$$|\Delta a|_{max} = 3 J_2 \frac{R^2}{r_0} \sin^2 i_0 \quad (21.21)$$

and occurs at $u_e = 90^\circ, 270^\circ$; so, at points where the satellite reaches a maximum or minimum declination. For a satellite at an altitude of 500 km and an inclination of 90° , we find for the maximum and mean values of Δa over an orbital revolution:

$$|\Delta a|_{max} = 19.2 \text{ km} ; \Delta a_{mean} = -9.6 \text{ km}$$

Similarly, we find for a satellite at geostationary altitude ($h = 35,786$ km) and $i = 90^\circ$:

$$|\Delta a|_{max} = 3.1 \text{ km} ; \quad \Delta a_{mean} = -1.6 \text{ km}$$

For the determination of Δi , $\Delta\Omega$ and ΔH , we start from (21.7-2). Substitution of (21.9) and (21.17) into (21.7-2) gives after some algebraic manipulation:

$$\frac{d\bar{H}}{du} = -3\mu J_2 \frac{R^2}{n_0 r_0^3} [\sin i_0 \cos i_0 \sin^2 u \bar{e}_x - \sin i_0 \sin u \cos u \bar{e}_y] \quad (21.22)$$

Integration from $u = 0$ to $u = u_e$ results in

$$\begin{aligned} \Delta H_x &= -\frac{3}{2}\mu J_2 \frac{R^2}{n_0 r_0^3} \sin i_0 \cos i_0 (u_e - \frac{1}{2} \sin 2u_e) \\ \Delta H_y &= \frac{3}{2}\mu J_2 \frac{R^2}{n_0 r_0^3} \sin i_0 \sin^2 u_e \\ \Delta H_z &= 0 \end{aligned} \quad (21.23)$$

The third relation shows that the J_2 -term does not affect the component of the angular momentum about the Z -axis. In Section 23.1, it will be shown that this characteristic holds for any term of the series expansion for the potential of a gravity field that is rotationally symmetric about the Z -axis.

Substitution of (21.23) into (21.14) leads with (21.15) to

$$\begin{aligned} \Delta i &= -\frac{3}{4}J_2 \frac{R^2}{r_0^2} \sin 2i_0 \sin^2 u_e \\ \Delta\Omega &= -\frac{3}{2}J_2 \frac{R^2}{r_0^2} \cos i_0 (u_e - \frac{1}{2} \sin 2u_e) \\ \Delta H &= -\frac{3}{2}J_2 \frac{R^2}{r_0^2} H_0 \sin^2 i_0 \sin^2 u_e \end{aligned} \quad (21.24)$$

We conclude that Δi and $\Delta\Omega$ are negative if $i_0 < 90^\circ$ and positive if $i_0 > 90^\circ$, that ΔH is negative if $i_0 \neq 0^\circ$, and that $|\Delta i|$ and $|\Delta\Omega|$ decrease inversely proportionally to the square of the orbital radius. We also conclude that, for a specified orbit, Δi varies periodically along the orbit with a maximum value of

$$|\Delta i|_{max} = \frac{3}{4}J_2 \frac{R^2}{r_0^2} \sin 2i_0 \quad (21.25)$$

The absolute maximum value of Δi occurs at $i_0 = 45^\circ, 135^\circ$. The parameter $\Delta\Omega$ shows a secular trend and superimposed on it a short-period variation. The maximum value of the periodic variation in Ω is given by

$$|\Delta\Omega_{per}|_{max} = \frac{3}{4}J_2 \frac{R^2}{r_0^2} \cos i_0 \quad (21.26)$$

For a satellite at an altitude of 500 km and an inclination of 45° , we find for these maximum values and for the mean values of Δi and ΔH :

$$|\Delta i|_{max} = 2.4' ; |\Delta\Omega_{per}|_{max} = 1.7' ; \Delta i_{mean} = -1.2' ; \Delta H_{mean} = -18.3 \text{ km}^2/\text{s}$$

Similarly, we find for a satellite at geostationary altitude and for $i_0 = 45^\circ$:

$$|\Delta i|_{max} = 4.0'' ; |\Delta\Omega_{per}|_{max} = 2.7'' ; \Delta i_{mean} = -1.9'' ; \Delta H_{mean} = -1.20 \text{ km}^2/\text{s}$$

For a specified inclination, the values of Δi and ΔH are, according to (21.24), zero when $u_e = k\pi$, where k is an integer. So, i and H do not show secular variations. However, Ω does show a secular variation, which is called the *regression of the nodes*. According to (21.24-2), the value of $\Delta\Omega$ after one orbital revolution of the satellite ($u_e = 2\pi$) is given by

$$\Delta_{2\pi}\Omega = -3\pi J_2 \frac{R^2}{r_0^2} \cos i_0 \quad (21.27)$$

So, $\Delta_{2\pi}\Omega < 0$ if $0^\circ < i < 90^\circ$ and $\Delta_{2\pi}\Omega > 0$ if $90^\circ < i < 180^\circ$. The mean rate of this regression is given by

$$\dot{\Omega}_{mean} = \frac{\Delta_{2\pi}\Omega}{T} \quad (21.28)$$

where T is the period of the unperturbed orbit:

$$T = 2\pi \sqrt{\frac{r_0^3}{\mu}} \quad (21.29)$$

Substitution of (21.27) and (21.29) into (21.28) gives

$$\dot{\Omega}_{mean} = -\frac{3}{2}J_2 R^2 \sqrt{\frac{\mu}{r_0^7}} \cos i_0 \quad (21.30)$$

For a satellite at an altitude of 500 km and an inclination of 45° , we find $\dot{\Omega}_{mean} = -0.23^\circ/\text{hr}$; for a satellite at geostationary altitude and the same inclination: $\dot{\Omega}_{mean} = -34''/\text{day}$.

In Chapter 23 we will treat the orbital effects of the J_2 -term in the Earth's gravitational potential in more detail.

21.3. Perturbations due to lunar and solar attraction

In Section 20.3 we have found the following expressions for the accelerations in the X -, Y - and Z -direction due to the gravitational attraction by a third body, e.g. the Moon or the Sun:

$$f_x = \mu_d \left(\frac{x_d - x_s}{r_{sd}^3} - \frac{x_d}{r_d^3} \right) \quad (21.31)$$

$$\begin{aligned} f_y &= \mu_d \left(\frac{y_d - y_s}{r_{sd}^3} - \frac{y_d}{r_d^3} \right) \\ f_z &= \mu_d \left(\frac{z_d - z_s}{r_{sd}^3} - \frac{z_d}{r_d^3} \right) \end{aligned} \quad (21.31)$$

where we have used a slightly different notation: μ_d is the gravitational parameter of the perturbing body (d), \bar{r}_s is the position vector of the satellite (s) relative to the non-rotating geocentric reference frame, \bar{r}_d is the position vector of the perturbing body relative to the non-rotating geocentric reference frame, and \bar{r}_{sd} is the vector from the satellite to the perturbing body. Elaboration of (21.31-1) leads to

$$f_x = \mu_d \left[\frac{x_d - x_s}{\{(x_d - x_s)^2 + (y_d - y_s)^2 + (z_d - z_s)^2\}^{3/2}} - \frac{x_d}{r_d^3} \right]$$

or

$$f_x = \frac{\mu_d}{r_d^3} \left[\frac{x_d - x_s}{\left\{ 1 - 2 \frac{(x_d x_s + y_d y_s + z_d z_s)}{r_d^2} + \frac{r_s^2}{r_d^2} \right\}^{3/2}} - x_d \right]$$

When the perturbing body is the Sun, then $r_s/r_d \ll 1$ for all satellite orbital altitudes; when the perturbing body is the Moon, then $r_s/r_d \ll 1$ if $r_s < 45,000$ km. Realizing that the large majority of Earth satellites have orbital radii of much less than 45,000 km, for those satellites we may approximate the above equation, both for the perturbations due to solar attraction and for the perturbations due to lunar attraction, to

$$f_x = -\frac{\mu_d}{r_d^3} \left[x - 3x_d \frac{(x_d x + y_d y + z_d z)}{r_d^2} \right] \quad (21.32)$$

where we have omitted the index s . For the other two position components, we obtain similar approximative relations:

$$\begin{aligned} f_y &= -\frac{\mu_d}{r_d^3} \left[y - 3y_d \frac{(x_d x + y_d y + z_d z)}{r_d^2} \right] \\ f_z &= -\frac{\mu_d}{r_d^3} \left[z - 3z_d \frac{(x_d x + y_d y + z_d z)}{r_d^2} \right] \end{aligned} \quad (21.32)$$

It is evident that the perturbing effects depend on the position of the perturbing body relative to the satellite's orbital plane. To simplify the analysis, we choose $z_d = 0$, i.e. we assume that the perturbing body is situated in the equatorial plane. In reality, the Sun moves about the Earth in the ecliptic, which is inclined by 23.4° to the equatorial plane and the inclination of the Moon's orbital plane relative to the equator varies between 18.3° and 28.6° (Section 17.1). To compensate for these non-zero inclinations of the orbits of the perturbing bodies, one may use a corrected value of the inclination of the satellite orbits. We will not address that aspect in the following

analysis.

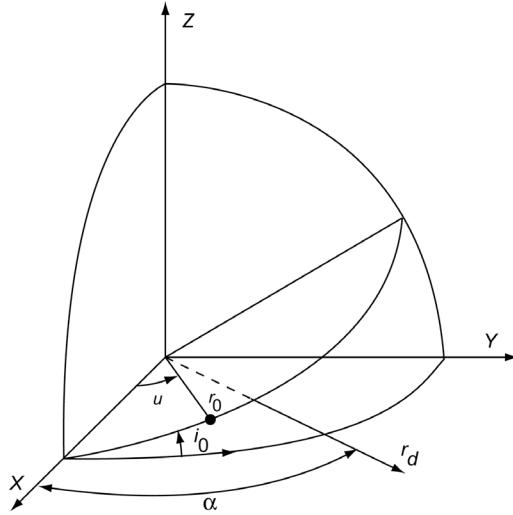


Figure 21.2: Geometry used for the analysis of luni-solar attraction perturbations, and the initial orbital plane and the position vector of the perturbing body.

According to Figure 21.2:

$$x_d = r_d \cos \alpha \quad ; \quad y_d = r_d \sin \alpha \quad ; \quad z_d = 0 \quad (21.33)$$

where α is the right ascension of the perturbing body. Substitution of (21.8) and (21.33) into (21.32) leads, after some algebraic manipulation, to

$$\begin{aligned} f_x &= -\frac{\mu_d}{r_d^3} r_0 \left[(1 - 3 \cos^2 \alpha) \cos u - 3 \sin \alpha \cos \alpha \cos i_0 \sin u \right] \\ f_y &= -\frac{\mu_d}{r_d^3} r_0 \left[(1 - 3 \sin^2 \alpha) \cos i_0 \sin u - 3 \sin \alpha \cos \alpha \cos u \right] \\ f_z &= -\frac{\mu_d}{r_d^3} r_0 \sin i_0 \sin u \end{aligned} \quad (21.34)$$

From (21.10) and (21.34) we obtain, after some algebraic manipulation,

$$\frac{d\bar{r}}{dt} \cdot \bar{f} = \frac{3}{2} \frac{\mu_d}{r_d^3} n_0 r_0^2 \left[\sin 2\alpha \cos i_0 \cos 2u + (\sin^2 \alpha \cos^2 i_0 - \cos^2 \alpha) \sin 2u \right] \quad (21.35)$$

Substitution of this relation and (21.6) into (21.7-1) yields

$$\frac{da}{du} = 3 \frac{\mu_d}{\mu} \frac{r_0^4}{r_d^3} \left[\sin 2\alpha \cos i_0 \cos 2u + (\sin^2 \alpha \cos^2 i_0 - \cos^2 \alpha) \sin 2u \right] \quad (21.36)$$

We now assume that the position of Sun and Moon may be considered stationary during an orbital revolution of the satellite. This assumption is quite realistic for low-altitude satellites; even for a satellite at an altitude of 5000 km the right ascensions of Moon and Sun change only

by 1.8° and 0.14° , respectively, during an orbital revolution of that satellite. The assumption is less accurate for the analysis of lunar attraction perturbations of high-altitude satellites; e.g. for a geostationary satellite we find that the right ascensions of the Moon changes by 13° during an orbital revolution of the satellite. With the assumption $\alpha = \alpha_0$, where α_0 indicates the value of α at $u = 0^\circ$, integration of (21.36) from $u = 0$ to $u = u_e$ yields

$$\Delta\alpha = \frac{3}{2} \frac{\mu_d}{\mu} \frac{r_0^4}{r_d^3} [\sin 2\alpha_0 \cos i_0 \sin 2u_e + 2(\sin^2 \alpha_0 \cos^2 i_0 - \cos^2 \alpha_0) \sin^2 u_e] \quad (21.37)$$

We conclude that $|\Delta\alpha|$ increases proportionally to the fourth power of the orbital radius. Note that $\Delta\alpha$ varies periodically over an orbital revolution of the satellite and that $\Delta\alpha$ is zero if $u_e = k\pi$, where k is an integer. So, when the satellite has completed an orbital revolution, i.e. $u_e = 2\pi$: $\Delta_{2\pi}\alpha = 0$. This implies that a fixed ('not moving') celestial body will not produce any long-term change in the total orbital energy of the satellite. The change in energy produced by swingby maneuvers (Section 18.11), therefore is produced by the 'motion' of the gravity field of the swingby planet in its orbit about the Sun!

If the position vector of the perturbing body is perpendicular to the orbital plane of the satellite, i.e. $i_0 = 90^\circ$ and $\alpha_0 = 90^\circ, 270^\circ (\Omega_0 = 0^\circ)$, we find from (21.37): $\Delta\alpha = 0$, which could be expected. To find the inclination for which the maximum and minimum values of $\Delta\alpha$ occur, we differentiate (21.37) partially with respect to i_0 :

$$\frac{\partial \Delta\alpha}{\partial i_0} = -6 \frac{\mu_d}{\mu} \frac{r_0^4}{r_d^3} [\sin \alpha_0 \sin i_0 \sin u_e \{ \sin \alpha_0 \cos i_0 \sin u_e + \cos \alpha_0 \cos u_e \}]$$

So, $\partial \Delta\alpha / \partial i_0 = 0$ if

$$i_0 = 0^\circ \quad \vee \quad \cos i_0 = -\frac{1}{\tan \alpha_0 \tan u_e}$$

Since we are only interested in solutions that hold for any value of u_e , only the first solution is relevant. For $i_0 = 0^\circ$, (21.37) simplifies to

$$\Delta\alpha = 3 \frac{\mu_d}{\mu} \frac{r_0^4}{r_d^3} \sin u_e \sin(2\alpha_0 - u_e) \quad (21.38)$$

Note that $\Delta\alpha$ is maximum if $u_e = \frac{1}{2}\pi + 2k\pi$ and $2\alpha_0 - u_e = \frac{1}{2}\pi + 2k\pi$, where k is an integer. These conditions can also be written as $u_e - \alpha_0 = k\pi$. This means that the maximum value of $\Delta\alpha$ occurs when the satellite crosses the line connecting the Earth and the disturbing body (Sun or Moon). In a similar way, we find that the minimum value of $\Delta\alpha$ occurs when $u_e - \alpha_0 = \frac{1}{2}\pi + k\pi$; so, if the satellite is 90° ahead or behind the line connecting the Earth and the disturbing body. We also find that $\Delta\alpha = 0$ if $u_e = k\pi$ or $u_e = 2\alpha_0 + k\pi$. When $\alpha_0 = k\pi$, the product of the two sine-terms on the right-hand side of (21.38) can be written as $-\sin^2 u_e$; for $\alpha_0 = \frac{1}{4}\pi + k\pi$ as $\frac{1}{2}\sin 2u_e$; for $\alpha_0 = \frac{1}{2}\pi + k\pi$ as $\sin^2 u_e$; for $\alpha_0 = \frac{3}{4}\pi + k\pi$ as $-\frac{1}{2}\sin 2u_e$. So, when $\alpha_0 = k\pi$ then $\Delta\alpha \leq 0$; when $\alpha_0 = \frac{1}{2}\pi + k\pi$ then $\Delta\alpha \geq 0$; when $\alpha_0 = \frac{1}{4}\pi + k\pi$ or $\alpha_0 = \frac{3}{4}\pi + k\pi$ then $\Delta\alpha$ can be positive and negative. In all cases, $\Delta\alpha$ experiences a periodic variation with a period equal to $\frac{1}{2}T$, where T is the (unperturbed) orbital period of the satellite.

For our simplified analysis we have assumed that the perturbing body is stationary during an orbital revolution of the satellite. In reality, this is not the case and the angular positions of Sun and Moon in their orbital planes change (almost linearly) with time. In our model, where the

perturbing body is assumed to move in the equatorial plane, this implies that the right ascensions of Sun and Moon increase about linearly with time. To account for this change in right ascension, we may, to first-order approximation, assume that for each subsequent revolution α_0 increases discontinuously by an amount of $n_d T$, where n_d is the mean angular motion of the perturbing body about the Earth. An analysis of (21.38) then shows that Δa experiences a short-period variation with amplitude, A , and period, P :

$$A(\Delta a_{sh}) = \frac{3}{2} \frac{\mu_d}{\mu} \frac{r_0^4}{r_d^3} ; \quad P_{sh} = \frac{1}{2} T = \pi \sqrt{\frac{r_0^3}{\mu}} \quad (21.39)$$

superimposed on a long-period variation with amplitude and period:

$$A(\Delta a_l) = \frac{3}{2} \frac{\mu_d}{\mu} \frac{r_0^4}{r_d^3} ; \quad P_l = \frac{\pi}{n_d} \quad (21.40)$$

Note that the period of the long-period variation is just half the orbital period of the perturbing body in its motion about the Earth; so, half a year for the Sun as perturbing body and half a month for the Moon as perturbing body. The long-term average of Δa is zero, so there is no secular variation of Δa . It is recalled that this analysis only holds for the case that $i_0 = 0^\circ$; i.e. for the case that the perturbing body is situated in the satellite's orbital plane. The separation of orbit perturbations into short-period, long-period, and secular components will be discussed further in Section 23.2.

In the equations given above, the disturbing body is characterized by the ratio μ_d/r_d^3 . For the Sun (index S) and the Moon (index M), we find

$$\frac{\mu_M}{r_M^3} \approx 2.2 \frac{\mu_S}{r_S^3} \quad (21.41)$$

This relation shows that for the same values of α_0 and u_e , the magnitude of the solar effect is about half that of the lunar effect (Section 4.2). For a satellite at an altitude of 500 km and $i_0 = 0^\circ$ or a geostationary satellite, we find, to first-order approximation, for the amplitude and period of the variation of Δa due to lunar attraction:

$$h = 500 \text{ km: } A(\Delta a_{sh}) = 0.72 \text{ m} ; P_{sh} = 47 \text{ min} ; A(\Delta a_l) = 0.72 \text{ m} ; P_l = 13.7 \text{ day}$$

$$h = 35,786 \text{ km: } A(\Delta a_{sh}) = 1.0 \text{ km} ; P_{sh} = 0.50 \text{ day} ; A(\Delta a_l) = 1.0 \text{ km} ; P_l = 13.7 \text{ day}$$

Note that for low-altitude orbits Δa is relatively small, but that for geostationary satellites lunar and solar attraction lead to major orbit perturbations.

In principle, we can apply the same methodology to analyze the long-period variations of the other orbital element. However, it is emphasized that the analysis presented for the long-period variation of Δa only served to indicate some major features of this type of variation, and that it is only of limited practical value. The reason is that we have assumed that the elements of the satellite orbit are constant and that the perturbing body moves in the equatorial plane. As mentioned before, in reality, Sun and Moon move in planes inclined to the equatorial plane, and the satellite's orbital elements, in particular the right ascension of the ascending node, change continuously due to various orbit perturbations. These orbital changes become significant over periods comparable to the period of the long-period variations.

We now continue with the computation of Δi , $\Delta\Omega$ and ΔH . Combination of (21.9) and (21.34) yields for the components of the vector product $\bar{r} \times \bar{f}$:

$$\begin{aligned} (\bar{r} \times \bar{f})_x &= -3 \frac{\mu_d}{r_d^3} r_0^2 \sin i_0 \sin \alpha_0 [\cos i_0 \sin \alpha_0 \sin^2 u + \cos \alpha_0 \sin u \cos u] \\ (\bar{r} \times \bar{f})_y &= 3 \frac{\mu_d}{r_d^3} r_0^2 \sin i_0 \cos \alpha_0 [\cos \alpha_0 \sin u \cos u + \cos i_0 \sin \alpha_0 \sin^2 u] \\ (\bar{r} \times \bar{f})_z &= -3 \frac{\mu_d}{r_d^3} r_0^2 [\cos i_0 \cos 2\alpha_0 \sin u \cos u + \sin \alpha_0 \cos \alpha_0 (\cos^2 i_0 \sin^2 u - \cos^2 u)] \end{aligned} \quad (21.42)$$

Substitution into (21.7-2) and subsequent integration from $u = 0$ to $u = 2\pi$ leads to

$$\begin{aligned} \Delta_{2\pi} H_x &= -3 \frac{\mu_d}{r_d^3} \frac{r_0^2}{n_0} \pi \sin i_0 \cos i_0 \sin^2 \alpha_0 \\ \Delta_{2\pi} H_y &= 3 \frac{\mu_d}{r_d^3} \frac{r_0^2}{n_0} \pi \sin i_0 \cos i_0 \sin \alpha_0 \cos \alpha_0 \\ \Delta_{2\pi} H_z &= 3 \frac{\mu_d}{r_d^3} \frac{r_0^2}{n_0} \pi \sin^2 i_0 \sin \alpha_0 \cos \alpha_0 \end{aligned} \quad (21.43)$$

Substitution of (21.6), (21.15) and (21.43) into (21.14) yields

$$\begin{aligned} \Delta_{2\pi} i &= -\frac{3}{2} \pi \frac{\mu_d}{\mu} \left(\frac{r_0}{r_d} \right)^3 \sin i_0 \sin 2\alpha_0 \\ \Delta_{2\pi} \Omega &= -3 \pi \frac{\mu_d}{\mu} \left(\frac{r_0}{r_d} \right)^3 \cos i_0 \sin^2 \alpha_0 \\ \Delta_{2\pi} H &= 0 \end{aligned} \quad (21.44)$$

We conclude that $|\Delta_{2\pi} i|$ and $|\Delta_{2\pi} \Omega|$ increase proportionally to the cube of the orbital radius. Note that according to (21.44-1) the inclination change after an orbital revolution is zero when the perturbing body lies in the plane of the satellite orbit ($i_0 = 0^\circ$). If $i_0 \neq 0^\circ$, the inclination decreases when α_0 lies in the first or third quadrant, and increases when α_0 lies in the second or fourth quadrant. Since in reality the perturbing body moves about the Earth and consequently, on the long run, α_0 will take all values between zero and 2π , the changes in inclination will cancel over long intervals of time. So, the inclination will show only short-period and long-period variations. On the other hand, (21.44-2) shows that the right ascension of the ascending node changes without limit, unless $i_0 = 90^\circ$. Averaging $\sin^2 \alpha_0$ over the range $\alpha_0 = 0$ to $\alpha_0 = 2\pi$ yields for the long-term mean rate of precession

$$\dot{\Omega}_{mean} = \frac{\Delta\Omega_{av}}{T} = -\frac{3}{4} \frac{\mu_d}{r_d^3} \sqrt{\frac{r_0^3}{\mu}} \cos i_0 \quad (21.45)$$

Equation (21.44-3) shows that a perturbing body does not produce any secular or long-periodic effect in the magnitude of the orbital angular momentum.

For a satellite at an altitude of 500 km and $i_0 = 10^\circ$, we find for the maximum effects of the lunar attraction and for the resulting mean rate of precession:

$$|\Delta_{2\pi}i|_{max} = 0.012'' \quad ; \quad |\Delta_{2\pi}\Omega|_{max} = 0.13'' \quad ; \quad \dot{\Omega}_{mean} = -1.02''/\text{day}$$

Similarly, we find for a satellite at geostationary altitude and $i_0 = 10^\circ$:

$$|\Delta_{2\pi}i|_{max} = 2.7'' \quad ; \quad |\Delta_{2\pi}\Omega|_{max} = 31'' \quad ; \quad \dot{\Omega}_{mean} = -15.6''/\text{day}$$

When we compare these values of $\dot{\Omega}_{mean}$ with the values obtained in Section 21.2 for the perturbations due to the J_2 -term of the Earth's gravity field, and ignore the fact that the values presented in Section 21.2 hold for a different inclination, we conclude that for a 500 km altitude orbit the orbital mean rate of precession produced by lunar attraction is much smaller than that produced by the J_2 -term, but that at geostationary altitude both rates of precession are of similar magnitude.

21.4. Perturbations due to atmospheric drag

In Section 20.2, we have used the following expression for the acceleration of the satellite due to atmospheric drag:

$$\bar{f} = -C_D \frac{1}{2} \rho \frac{A}{M} |\bar{v}| \bar{v} \quad (21.46)$$

where ρ is the atmospheric density, \bar{v} is the velocity of the satellite relative to the rotating Earth's atmosphere, M is the mass of the satellite, and C_D is the drag coefficient related to a reference surface A . When \bar{V} denotes the inertial velocity of the satellite and \bar{V}_a the local atmospheric inertial velocity, we have

$$\bar{v} = \bar{V} - \bar{V}_a \quad (21.47)$$

When we assume that the atmosphere rotates with the same angular velocity as the Earth, $\dot{\Theta}$, we can write

$$V_a = r_0 \dot{\Theta} \cos \delta \quad (21.48)$$

where δ is the declination of the satellite and V_a is directed due east. For the components of \bar{V}_a we write according to Figure 21.3:

$$V_{a_x} = -r_0 \dot{\Theta} \cos \delta \sin \alpha \quad ; \quad V_{a_y} = r_0 \dot{\Theta} \cos \delta \cos \alpha \quad (21.49)$$

where α is the right ascension of the satellite. For the orbit geometry selected in this Chapter ($\Omega_0 = 0^\circ$) and using the first and second cosine rule of spherical geometry (Section 11.8), we may write

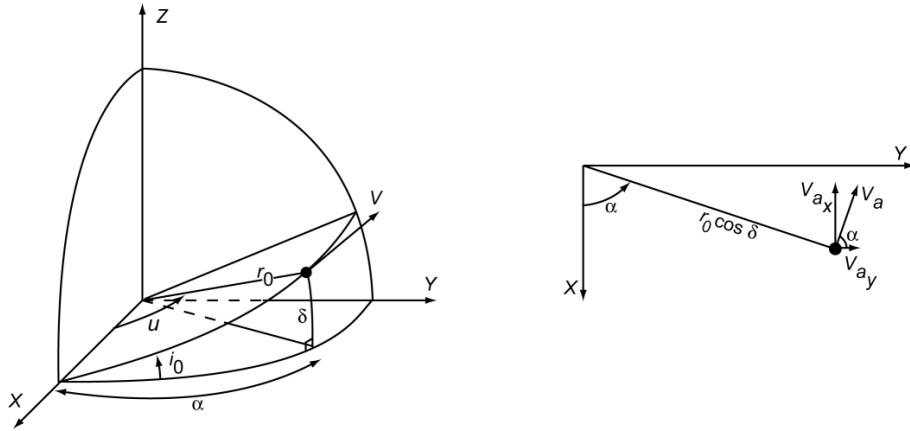


Figure 21.3: Geometry used for the analysis of atmospheric drag perturbations, and the initial orbital plane and satellite inertial velocity vector (left), and the atmospheric inertial velocity vector and its components (right).

$$V_{ax} = -r_0 \dot{\theta} \cos i_0 \sin u \quad ; \quad V_{ay} = r_0 \dot{\theta} \cos u \quad (21.50)$$

Substitution of (21.10) and (21.50) into (21.47) gives for the relative velocity

$$\bar{v} = -n_0 r_0 \left(1 - \frac{\dot{\theta}}{n_0} \cos i_0\right) \sin u \bar{e}_x + n_0 r_0 \left(\cos i_0 - \frac{\dot{\theta}}{n_0}\right) \cos u \bar{e}_y + n_0 r_0 \sin i_0 \cos u \bar{e}_z \quad (21.51)$$

From this expression, we find for the magnitude of the relative velocity

$$v^2 = n_0^2 r_0^2 \left[1 - 2 \frac{\dot{\theta}}{n_0} \cos i_0 + \left(\frac{\dot{\theta}}{n_0} \right)^2 (\cos^2 i_0 \sin^2 u + \cos^2 u) \right] \quad (21.52)$$

Since for orbital altitudes where atmospheric drag yields an important orbit perturbation: $\dot{\theta}/n_0 \ll 1$, we may linearize (21.52) and obtain

$$v \approx n_0 r_0 \left(1 - \frac{\dot{\theta}}{n_0} \cos i_0\right) \quad (21.53)$$

Substitution of (21.51) and (21.53) into (21.46), and neglecting terms of the second order in $\dot{\theta}/n_0$, leads to the following relations for the components of the drag acceleration:

$$\begin{aligned} f_x &= \frac{C_D A}{2M} \rho n_0^2 r_0^2 \left(1 - 2 \frac{\dot{\theta}}{n_0} \cos i_0\right) \sin u \\ f_y &= -\frac{C_D A}{2M} \rho n_0^2 r_0^2 \left[\cos i_0 - \frac{\dot{\theta}}{n_0} (1 + \cos^2 i_0)\right] \cos u \\ f_z &= -\frac{C_D A}{2M} \rho n_0^2 r_0^2 \sin i_0 \left(1 - \frac{\dot{\theta}}{n_0} \cos i_0\right) \cos u \end{aligned} \quad (21.54)$$

Combination of (21.10) and (21.54) results in

$$\frac{d\bar{r}}{dt} \cdot \bar{f} = -\frac{C_D A}{2M} \rho n_0^3 r_0^3 (1 - 2 \frac{\dot{\theta}}{n_0} \cos i_0) \quad (21.55)$$

Substitution of (21.55) into (21.7-1) results in

$$\frac{da}{du} = -\frac{C_D A}{M} \rho r_0^2 (1 - 2 \frac{\dot{\theta}}{n_0} \cos i_0) \quad (21.56)$$

Now, as a first-order approximation, we assume that ρ is constant (ρ_0) during an orbital revolution and that $C_D A/M$ is constant too. We then conclude from (21.56) that da/du is constant and thus that the semi-major axis decreases linearly during an orbital revolution. Integration of (21.56) from $u = 0$ to $u = 2\pi$ then yields

$$\Delta_{2\pi}a = -2\pi \frac{C_D A}{M} \rho_0 r_0^2 (1 - 2 \frac{\dot{\theta}}{n_0} \cos i_0) \quad (21.57)$$

To first-order approximation, the atmospheric density falls off about exponentially with increasing altitude:

$$\rho_1 \approx \rho_0 \exp[-(h_1 - h_0)/H] \quad (21.58)$$

where h is the altitude above the surface of the Earth and H is a scaling factor that is called the *density scale height*. The value of this scaling factor increases with increasing altitude and is a function of the exospheric temperature (Section 20.2). It ranges from 30 km to 150 km at altitudes between 200 km and 600 km, which is the altitude range where atmospheric drag yields significant orbit perturbations. We now consider two successive orbital revolutions. For the first one the initial conditions are indicated by the index 0; for the second one by the index 1. At the end of the first revolution the semi-major axis has decreased by an amount specified by (21.57). Consequently, for the second revolution: $\rho_1 > \rho_0$, but $r_1 < r_0$. If we assume that all other parameters in (21.57) are constant, realize that for the region of interest $(h_0 - h_1)/H$, h_0/R and h_1/R are (rather) small quantities, and linearize the expressions for the terms $\exp((h_0 - h_1)/H)$ and $((R + h_1)/(R + h_0))^2$, we find

$$\frac{(\Delta_{2\pi}a)_1}{(\Delta_{2\pi}a)_0} = \frac{\rho_1}{\rho_0} \left(\frac{r_1}{r_0} \right)^2 \approx 1 + \left(\frac{1}{H} - \frac{2}{R} \right) (h_0 - h_1)$$

Since $2/R < 1/H$, we conclude that the right-hand side of this equation is larger than one, which means that after the second revolution the decrease of the semi-major axis is larger than after the first revolution. When we repeat this analysis for successive revolutions, this results in the well-known *spiral decay* of orbits that are subjected to atmospheric drag. It is emphasized that in reality the situation is much more complicated, since: 1) the atmosphere is oblate and the atmospheric density therefore varies with the declination of the satellite; 2) the atmospheric density at a specific location exhibits a diurnal variation because of the heating of the atmosphere by solar radiation; 3) the density varies explicitly with time as a result of solar activity (Section 20.2).

We now continue with the computation of Δi , $\Delta\Omega$ and ΔH . Combination of (21.9) and (21.54) gives for the components of the vector product $\bar{r} \times \bar{f}$:

$$\begin{aligned}
 (\bar{r} \times \bar{f})_x &= -\frac{C_D A}{2M} \rho n_0^2 r_0^3 \dot{\theta} \sin i_0 \sin u \cos u \\
 (\bar{r} \times \bar{f})_y &= \frac{C_D A}{2M} \rho n_0^2 r_0^3 \sin i_0 \left[1 - \frac{\dot{\theta}}{n_0} \cos i_0 (1 + \sin^2 u) \right] \\
 (\bar{r} \times \bar{f})_z &= -\frac{C_D A}{2M} \rho n_0^2 r_0^3 \left[\cos i_0 - \frac{\dot{\theta}}{n_0} \{ \cos^2 i_0 (1 + \sin^2 u) + \cos^2 u \} \right]
 \end{aligned} \tag{21.59}$$

Substitution into (21.7-2) and integration from $u = 0$ to $u = 2\pi$ leads, for constant values of the density and $C_D A/M$, to

$$\begin{aligned}
 \Delta_{2\pi} H_x &= 0 \\
 \Delta_{2\pi} H_y &= \frac{C_D A}{2M} \pi \rho_0 n_0 r_0^3 \sin i_0 \left[2 - 3 \frac{\dot{\theta}}{n_0} \cos i_0 \right] \\
 \Delta_{2\pi} H_z &= -\frac{C_D A}{2M} \pi \rho_0 n_0 r_0^3 \left[2 \cos i_0 - \frac{\dot{\theta}}{n_0} (3 \cos^2 i_0 + 1) \right]
 \end{aligned} \tag{21.60}$$

Substitution of these relations and (21.15) into (21.14) gives

$$\begin{aligned}
 \Delta_{2\pi} i &= -\frac{1}{2} \pi \frac{C_D A}{M} \rho_0 r_0 \frac{\dot{\theta}}{n_0} \sin i_0 \\
 \Delta_{2\pi} \Omega &= 0 \\
 \Delta_{2\pi} H &= -\pi \frac{C_D A}{M} \rho_0 n_0 r_0^3 \left(1 - 2 \frac{\dot{\theta}}{n_0} \cos i_0 \right)
 \end{aligned} \tag{21.61}$$

Note that: 1) if the rotation rate of the atmosphere would be zero then $\Delta_{2\pi} i$ is zero, while $\Delta_{2\pi} a$ and $\Delta_{2\pi} H$ are non-zero; 2) the inclination change is linearly dependent on the rotation rate of the atmosphere. This is the reason why the history of Δi values as obtained from tracking data can be used, in combination with an atmospheric density model, to determine the rotation rate of the atmosphere at a particular altitude. The change in the right ascension of the ascending node is zero, and the change in the magnitude of the angular momentum is only weakly dependent on the inclination.

For a satellite at an altitude of 300 km, we may assume an average atmospheric density of $2 \cdot 10^{-11}$ kg/m³ at a high level of solar activity. Adopting $C_D A/M = 0.5$ m²/kg and $i_0 = 45^\circ$, we find for the orbital change after one orbital revolution:

$$\Delta_{2\pi} a = -2.55 \text{ km} ; \quad \Delta_{2\pi} i = -0.97'' ; \quad \Delta_{2\pi} H = -9.86 \text{ km}^2/\text{s}$$

For an orbital altitude of 500 km and the same high level of solar activity, we may assume an average density of 10^{-12} kg/m³ and find for the same satellite:

$$\Delta_{2\pi}a = -135 \text{ m} ; \quad \Delta_{2\pi}i = -0.050'' ; \quad \Delta_{2\pi}H = -0.54 \text{ km}^2/\text{s}$$

Note the large increase in Δa when the orbital altitude decreases from 500 km to 300 km.

21.5. Perturbations due to the solar radiation force

In Section 20.4, we have found the following expression for the acceleration of a satellite due to the solar radiation force:

$$\bar{f} = -C_R \frac{WA}{Mc} \bar{e}_s \quad (21.62)$$

where W is the power density of the solar radiation near the Earth (about 1360 W/m^2), which is for Earth satellites practically independent of orbital altitude, M is the mass of the satellite, c is the speed of light (in vacuum), A is the effective cross-sectional area of the satellite, C_R is the satellite's reflectivity, and \bar{e}_s is a unit vector from the satellite to the Sun. The quantity $C_R W/c$ is the solar radiation pressure. We can write (21.62) as

$$\bar{f} = -f \bar{e}_s ; \quad f = C_R \frac{WA}{Mc} \quad (21.63)$$

Because the distance to the Sun is much larger than the orbital altitude of the satellite, the vector \bar{e}_s is (about) parallel to the vector from the Earth to the Sun. When the satellite crosses the shadow cylinder (Section 20.4) produced by the Earth's interception of solar radiation (Figure 21.4), the

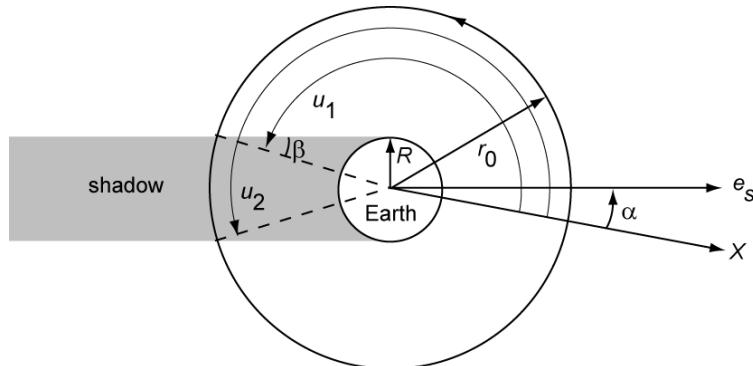


Figure 21.4: Geometry of an eclipse when the Sun is in the satellite's orbital plane.

satellite experiences an eclipse and $\bar{f} = 0$. In general, the value of A , and so of f , will change with the satellite's position in its orbit, because of the changing geometry of the satellite when viewed from the Sun. However, for the analysis given below we assume that, if the satellite is in sunlight, the value of f is constant during an orbital revolution. When we assume for simplicity that the Sun lies in the XY -plane and use the geometry depicted in Figure 21.4, we may write for the components of the acceleration:

$$f_x = -f \cos \alpha ; \quad f_y = -f \sin \alpha ; \quad f_z = 0 \quad (21.64)$$

where α is the right ascension of the Sun (Figure 21.2). To allow for the shadow effect (Figure 21.4), we assume that the satellite enters an eclipse at $u = u_1$ and emerges into sunlight again at $u = u_2$.

From (21.10) and (21.64) we obtain

$$\frac{d\bar{r}}{dt} \cdot \bar{f} = -f n_0 r_0 (-\cos \alpha \sin u + \cos i_0 \sin \alpha \cos u) \quad (21.65)$$

Substitution of this relation into (21.7-1) gives

$$\frac{da}{du} = -2 \frac{f}{n_0^2} (-\cos \alpha \sin u + \cos i_0 \sin \alpha \cos u) \quad (21.66)$$

For satellites below the geostationary altitude, the right ascension of the Sun hardly varies during an orbital revolution of the satellite (Section 21.3). Then, we may write $\alpha = \alpha_0$, where α_0 is the value of α at $u = 0$. Integration of (21.66) from $u = 0$ to $u = u_1$ and from $u = u_2$ to $u = 2\pi$ results in

$$\Delta_{2\pi}a = 2 \frac{f}{n_0^2} [\cos \alpha_0 (\cos u_2 - \cos u_1) + \cos i_0 \sin \alpha_0 (\sin u_2 - \sin u_1)] \quad (21.67)$$

We conclude that $|\Delta_{2\pi}a|$ increases proportionally to the cube of the orbital radius.

We now continue with the computation of Δi , $\Delta\Omega$ and ΔH . Combination of (21.9) and (21.64) yields, with $\alpha = \alpha_0$, for the components of the vector product $\bar{r} \times \bar{f}$:

$$\begin{aligned} (\bar{r} \times \bar{f})_x &= fr_0 \sin i_0 \sin \alpha_0 \sin u \\ (\bar{r} \times \bar{f})_y &= -fr_0 \sin i_0 \cos \alpha_0 \sin u \\ (\bar{r} \times \bar{f})_z &= fr_0 (\cos i_0 \cos \alpha_0 \sin u - \sin \alpha_0 \cos u) \end{aligned} \quad (21.68)$$

Substitution into (21.7-2) and integration from $u = 0$ to $u = u_1$ and from $u = u_2$ to $u = 2\pi$ yields

$$\begin{aligned} \Delta_{2\pi}H_x &= f \frac{r_0}{n_0} \sin i_0 \sin \alpha_0 (\cos u_2 - \cos u_1) \\ \Delta_{2\pi}H_y &= -f \frac{r_0}{n_0} \sin i_0 \cos \alpha_0 (\cos u_2 - \cos u_1) \\ \Delta_{2\pi}H_z &= f \frac{r_0}{n_0} [\cos i_0 \cos \alpha_0 (\cos u_2 - \cos u_1) + \sin \alpha_0 (\sin u_2 - \sin u_1)] \end{aligned} \quad (21.69)$$

Substitution of these relations and (21.15) into (21.14) gives

$$\begin{aligned} \Delta_{2\pi}i &= -\frac{f}{n_0^2 r_0} \sin i_0 \sin \alpha_0 (\sin u_2 - \sin u_1) \\ \Delta_{2\pi}\Omega &= \frac{f}{n_0^2 r_0} \sin \alpha_0 (\cos u_2 - \cos u_1) \\ \Delta_{2\pi}H &= f \frac{r_0}{n_0} [\cos \alpha_0 (\cos u_2 - \cos u_1) + \cos i_0 \sin \alpha_0 (\sin u_2 - \sin u_1)] \end{aligned} \quad (21.70)$$

Note that $|\Delta_{2\pi}i|$ and $|\Delta_{2\pi}\Omega|$ increase proportionally to the square of the orbital radius. From (21-67) and (21.70) we conclude that when the satellite does not experience an eclipse during an orbital revolution, i.e. $u_2 = u_1$,

$$\Delta_{2\pi}a = \Delta_{2\pi}i = \Delta_{2\pi}\Omega = \Delta_{2\pi}H = 0 \quad (21.71)$$

So, the change of the orbital parameters after an orbital revolution is zero when the satellite is in sunlight continuously. This implies that these parameters then only show a short-period variation due to solar radiation pressure.

To analyze the short-period variations, we consider the case that the satellite is in sunlight during an entire orbital revolution. It is emphasized that this happens only for a particular range of values of α and i_0 . An elementary analysis of the geometry shown in Figure 21.2 reveals that for a satellite orbital altitude of 500 km and $i_0 = 90^\circ$ the satellite is continuously in sunlight if $68.0^\circ < \alpha < 112.0^\circ$ or $248.0^\circ < \alpha < 292.0^\circ$; for the same orbital altitude and $\alpha = 90^\circ$ the satellite is continuously in sunlight if $68.0^\circ < i_0 < 112.0^\circ$. For an orbit at geostationary altitude and $i_0 = 90^\circ$ the constraints are $8.7^\circ < \alpha < 171.3^\circ$ or $188.7^\circ < \alpha < 351.3^\circ$, and for $\alpha = 90^\circ$ the constraint is $8.7^\circ < i_0 < 171.3^\circ$. To simplify the analysis, we now assume that $i_0 = 90^\circ$. Substituting again $\alpha = \alpha_0$ and integrating (21.66) from $u = 0$ to $u = u_e$ then yields

$$\Delta\alpha = 2 \frac{f}{n_0^2} \cos \alpha_0 (1 - \cos u_e) \quad (21.72)$$

Substitution of (21.68) into (21.7-2), and integration from $u = 0$ to $u = u_e$ yields

$$\Delta H_x = f \frac{r_0}{n_0} \sin \alpha_0 (1 - \cos u_e) \quad (21.73)$$

$$\Delta H_y = -f \frac{r_0}{n_0} \cos \alpha_0 (1 - \cos u_e) \quad (21.73)$$

$$\Delta H_z = -f \frac{r_0}{n_0} \sin \alpha_0 \sin u_e$$

Substitution of these expressions into (21.14) gives

$$\begin{aligned} \Delta i &= \frac{f}{n_0^2 r_0} \sin \alpha_0 \sin u_e \\ \Delta \Omega &= \frac{f}{n_0^2 r_0} \sin \alpha_0 (1 - \cos u_e) \\ \Delta H &= f \frac{r_0}{n_0} \cos \alpha_0 (1 - \cos u_e) \end{aligned} \quad (21.74)$$

Note that $(1 - \cos u_e) \geq 0$ and that this expression has a mean value of one over an orbital revolution. From (21.72) and (21.74) we find that for a specified value of α_0 and $i_0 = 90^\circ$ the extreme values of the short-period variations during an orbital revolution are given by

$$\begin{aligned} (\Delta a)_{extr} &= 4 \frac{f}{n_0^2} \cos \alpha_0 & ; & (\Delta i)_{extr} = \frac{f}{n_0^2 r_0} \sin \alpha_0 \\ (\Delta \Omega)_{extr} &= 2 \frac{f}{n_0^2 r_0} \sin \alpha_0 & ; & (\Delta H)_{extr} = 2f \frac{r_0}{n_0} \cos \alpha_0 \end{aligned} \quad (21.75)$$

These expressions show that when the Earth-Sun vector is perpendicular to the satellite's orbital plane, i.e. $\alpha_0 = 90^\circ, 270^\circ$, the parameters a and H do not show a short-period variation, while the short-period variations of the orbital elements i and Ω are maximum.

For a numerical example, we adopt $A/M = 0.2 \text{ m}^2/\text{kg}$, $W = 1360 \text{ W/m}^2$, and $C_R = 1.8$, which yields with (21.63) $f = 1.63 * 10^{-6} \text{ m/s}^2$. When we assume $\alpha_0 = 70^\circ$, we find from (21.6) and (21.75) for the maximum values of the short-period variations for a circular orbit at an altitude of 500 km:

$$(\Delta a)_{max} = 1.82 \text{ m} ; (\Delta i)_{max} = 0.037'' ; (\Delta \Omega)_{max} = 0.075'' ; (\Delta H)_{max} = 0.0069 \text{ km}^2/\text{s}$$

and for a circular orbit at geostationary altitude:

$$(\Delta a)_{max} = 419 \text{ m} ; (\Delta i)_{max} = 1.41'' ; (\Delta \Omega)_{max} = 2.82'' ; (\Delta H)_{max} = 0.645 \text{ km}^2/\text{s}$$

Just as we did for the orbit perturbations due to solar gravitational attraction (Section 21.3), we can use (21.72) and (21.74) to analyze the long-period variation of the orbital parameters. However, such an analysis is only valid for a limited period of time, because the variation of α will lead to a situation where the satellite experiences an eclipse again. An exception is a *Sun-synchronous orbit*, which will be discussed in Section 23.7. For that type of orbit, the inclination is selected such that the gravitational perturbation due to the J_2 -term of the Earth's gravity field (Section 21.2) makes the orbit's right ascension of the ascending node to increase at the same rate as the angular motion of the Sun about the Earth. As will be shown in Section 23.7, a satellite in such an orbit may remain in sunlight for several months.

We now return to the analysis of the orbital changes after an orbital revolution and assume that the satellite experiences an eclipse during that revolution. To simplify the analysis, we assume that the inclination of the orbital plane is very small ($i_0 \approx 0^\circ$). As the Sun is assumed to be in the XY -plane, this implies that we assume the Sun to move in the satellite's orbital plane, but that the right ascension of the ascending node of the satellite orbit is still defined. Then, we find from Figure 21.4:

$$u_1 = \alpha_0 + \pi - \beta \quad ; \quad u_2 = \alpha_0 + \pi + \beta \quad (21.76)$$

From these relations we find

$$\begin{aligned} \cos u_2 - \cos u_1 &= 2 \sin \alpha_0 \sin \beta \\ \sin u_2 - \sin u_1 &= -2 \cos \alpha_0 \sin \beta \end{aligned} \quad (21.77)$$

From Figure 21.4 we find

$$\sin \beta = \frac{R}{r_0}$$

where R is the radius of the Earth. Substitution of this relation into (21.77) leads to

$$\begin{aligned}\cos u_2 - \cos u_1 &= 2 \frac{R}{r_0} \sin \alpha_0 \\ \sin u_2 - \sin u_1 &= -2 \frac{R}{r_0} \cos \alpha_0\end{aligned}\tag{21.78}$$

Substitution of (21.78) into (21.67) and (21.70) then yields

$$\begin{aligned}\Delta_{2\pi}a &= \Delta_{2\pi}i = \Delta_{2\pi}H = 0 \\ \Delta_{2\pi}\Omega &= 2 \frac{f}{n_0^2} \frac{R}{r_0^2} \sin^2 \alpha_0\end{aligned}\tag{21.79}$$

So, for this eclipse geometry the semi-major axis, inclination and angular momentum do not experience a change after an orbital revolution of the satellite, as could be expected. The maximum value of $\Delta_{2\pi}\Omega$ occurs for $\alpha_0 = 90^\circ, 270^\circ$ and amounts

$$(\Delta_{2\pi}\Omega)_{max} = 2 \frac{f}{n_0^2} \frac{R}{r_0^2}\tag{21.80}$$

Assuming again $f = 1.63 \cdot 10^{-6} \text{ m/s}^2$, we find for an orbital altitude of 500 km: $(\Delta_{2\pi}\Omega)_{max} = 0.07''$ and for the geostationary altitude: $(\Delta_{2\pi}\Omega)_{max} = 0.45''$.

The very fact that solar radiation produces changes in the orbit of a satellite has led to the concept of *solar sailing*. For such missions, the spacecraft is equipped with a very large lightweight highly reflective solar sail, of which the orientation is modulated as a function of time in such a way that the appropriate time-varying solar radiation force is generated to achieve prescribed changes of the spacecraft's orbital elements, without spending any propellant. The idea was proposed as early as 1924 by F. Zander (1887-1933). The Japanese space agency JAXA has successfully unfurled a solar sail in space with the IKAROS mission. This 307 kg spacecraft was launched on May 21, 2010, aboard a Japanese H-IIA rocket, together with Akatsuki (Venus Climate Orbiter), into a trajectory towards Venus. The 14 kg solar sail consisted of four trapezoid petals of a highly-reflective aluminized polyimide membrane only $7.5 \mu\text{m}$ thick. When expanded it formed a square shape measuring 20 m along the diagonal. Thin-film solar cells were attached to about 5% of the solar sail surface, capable of generating about 500 W of electrical power. Attached on the edge of the solar sail were a series of thin film optical devices of which the reflectance could be varied by applying an appropriate voltage to the electrodes. This system was used to fine-tune the attitude control of the solar sail. The sail was deployed by spinning the spacecraft at up to 20 rpm and using four 0.5 kg tip masses; once unfurled the sail was held flat by the spacecraft's spinning motion. IKAROS finished unfurling its sail on June 11, 2010. On July 9, JAXA confirmed that IKAROS was accelerated by its solar sail, and on July 23 JAXA announced successful attitude control. IKAROS continued to spin at approximately 2 rpm, requiring the variable reflectance devices to be cycled at that rate for attitude control. On December 8, 2010, IKAROS passed by Venus at a distance of about 80,800 km, completing the planned mission successfully.

21.6. Perturbations due to the electromagnetic force

In Section 20.5 we have found the following first-order expression for the perturbing acceleration

of an electrically charged satellite due to its motion through the Earth's magnetic field:

$$\bar{f} = -\frac{q}{M} g_{1,0} R^3 \bar{v} \times \bar{\nabla} \left(\frac{\sin \phi}{r^2} \right) \quad (21.81)$$

where q is the electric charge of the satellite, M is the mass of the satellite, $g_{1,0}$ is the first Gauss geomagnetic coefficient, R is the mean radius of the Earth, v is the satellite's velocity relative to the rotating geomagnetic field, ϕ is the satellite's latitude and r is the distance of the satellite from the center of the Earth. With $\delta = \phi$, where δ is the satellite's declination, and (11.19), we find

$$\bar{\nabla} \left(\frac{\sin \delta}{r^2} \right) = -\frac{1}{r^5} (3xz\bar{e}_x + 3yz\bar{e}_y - (r^2 - 3z^2)\bar{e}_z) \quad (21.82)$$

When we define a parameter γ as

$$\gamma = -\frac{q}{M} g_{1,0} R^3 \quad (21.83)$$

assume that the value of q is (about) constant during an orbital revolution, substitute (21.51), (21.82) and (21.83) into (21.81), perform the vector operation, and use (21.8), we find:

$$\begin{aligned} f_x &= \gamma \frac{n_0}{r_0^2} \left(\cos i_0 - \frac{\dot{\theta}}{n_0} (1 - 3 \sin^2 i_0 \sin^2 u) \right) \cos u \\ f_y &= \gamma \frac{n_0}{r_0^2} \left((1 - 3 \sin^2 i_0) - \frac{\dot{\theta}}{n_0} \cos i_0 (1 - 3 \sin^2 i_0 \sin^2 u) \right) \sin u \\ f_z &= 3\gamma \frac{n_0}{r_0^2} \left(\cos i_0 - \frac{\dot{\theta}}{n_0} (1 - \sin^2 i_0 \sin^2 u) \right) \sin i_0 \sin u \end{aligned} \quad (21.84)$$

where $\dot{\theta}$ is the Earth's (mean) rotation rate. As an illustration of the direction of the perturbing acceleration (Lorentz force), we consider the cases that the satellite moves in a circular low-altitude equatorial or polar orbit and we neglect the rotation of the geomagnetic field about the polar axis. We then find from (21.84):

$$\text{equatorial orbit: } f_x = \gamma \frac{n_0}{r_0^2} \cos u ; \quad f_y = \gamma \frac{n_0}{r_0^2} \sin u ; \quad f_z = 0$$

$$\text{polar orbit: } f_x = 0 ; \quad f_y = -2\gamma \frac{n_0}{r_0^2} \sin u ; \quad f_z = 0$$

Note that these results hold for the assumption made in this Chapter that the right ascension of the ascending node of the orbit is zero. Remembering that $g_{1,0} < 0$ (Section 20.5), we conclude from (21.83) that $\gamma > 0$ when $q > 0$ and that $\gamma < 0$ when $q < 0$. So, for the equatorial orbit the perturbing acceleration is always acting in the radial outward direction if the satellite has a positive potential and in the radial inward direction if the satellite has a negative potential. We already have found that result in Section 20.5. For the polar satellite the perturbing acceleration

is zero when the satellite passes the ascending and descending nodes, is at all other points along the orbit acting perpendicular to the orbital plane, is maximum when the satellite passes over the poles, and is pointing in opposite directions over the north pole and over the south pole.

From (21.10) and (21.84) we obtain

$$\frac{d\bar{r}}{dt} \cdot \bar{f} = -2 \frac{\gamma \dot{\theta} n_0}{r_0} \sin^2 i_0 \sin u \cos u \quad (21.85)$$

which leads with (21.7-1) and (21.83), and after integration from $u = 0$ to $u = u_e$, to

$$\Delta a = 2 \frac{q g_{1,0}}{M n_0} \frac{\dot{\theta}}{n_0} \frac{R^3}{r_0^2} \sin^2 i_0 \sin^2 u_e \quad (21.86)$$

We conclude that $|\Delta a|$ increases proportionally to the orbital radius, and that a shows a short-period variation during an orbital revolution.

We now continue with the computation of Δi , $\Delta \Omega$ and ΔH . Combination of (21.9) and (21.84) yields for the components of the vector product $\bar{r} \times \bar{f}$:

$$\begin{aligned} (\bar{r} \times \bar{f})_x &= 2 \gamma \frac{n_0}{r_0} \sin i_0 \left(1 - \frac{\dot{\theta}}{n_0} \cos i_0 \right) \sin^2 u \\ (\bar{r} \times \bar{f})_y &= -2 \gamma \frac{n_0}{r_0} \sin i_0 \left(\cos i_0 - \frac{\dot{\theta}}{n_0} \right) \sin u \cos u \\ (\bar{r} \times \bar{f})_z &= -2 \gamma \frac{n_0}{r_0} \sin^2 i_0 \sin u \cos u \end{aligned} \quad (21.87)$$

Note that in the last expression no term with $\dot{\theta}$ occurs. Substitution of (21.87) into (21.7-2) and integration from $u = 0$ to $u = u_e$ yields

$$\begin{aligned} \Delta H_x &= \frac{\gamma}{r_0} \sin i_0 \left(1 - \frac{\dot{\theta}}{n_0} \cos i_0 \right) \left(u_e - \frac{1}{2} \sin 2u_e \right) \\ \Delta H_y &= -\frac{\gamma}{r_0} \sin i_0 \left(\cos i_0 - \frac{\dot{\theta}}{n_0} \right) \sin^2 u_e \\ \Delta H_z &= -\frac{\gamma}{r_0} \sin^2 i_0 \sin^2 u_e \end{aligned} \quad (21.88)$$

Substitution of these relations, (21.15) and (21.83) into (21.14) results in

$$\Delta i = -\frac{q g_{1,0}}{M n_0} \left(\frac{R}{r_0} \right)^3 \sin i_0 \left(1 - \frac{\dot{\theta}}{n_0} \cos i_0 \right) \sin^2 u_e \quad (21.89)$$

$$\begin{aligned}\Delta\Omega &= -\frac{q g_{1,0}}{M n_0} \left(\frac{R}{r_0} \right)^3 \left(1 - \frac{\dot{\theta}}{n_0} \cos i_0 \right) \left(u_e - \frac{1}{2} \sin 2u_e \right) \\ \Delta H &= \frac{q g_{1,0}}{M} \left(\frac{R^3}{r_0} \right) \frac{\dot{\theta}}{n_0} \sin^2 i_0 \sin^2 u_e\end{aligned}\quad (21.89)$$

Note that a and H only change because of the rotation of the geomagnetic field, while i and Ω would also change if the magnetic field would not rotate. Since $g_{1,0} < 0$, we find that: 1) for $q < 0$, Δa and ΔH are always positive (or zero), while Δi and $\Delta\Omega$ are always negative (or zero); 2) for $q > 0$, Δa and ΔH are always negative (or zero), while Δi and $\Delta\Omega$ are always positive (or zero). Equations (21.86) and (21.89) show that the variations in a , i and H are zero for $i_0 = 0^\circ$. The value of $\Delta\Omega$ is, generally, non-zero at $i_0 = 0^\circ$, but it should be realized that Ω is undefined at $i_0 = 0^\circ$. When we consider a spherical conducting satellite that moves in a circular 500 km altitude orbit with $i = 45^\circ$, apply (20.22) and (20.23), and adopt the values $U^* = -100$ V, $R_s/M = 0.1$ m/kg, we find for the maximum values of Δa , Δi , and ΔH during an orbital revolution:

$$\Delta a_{max} = 11 \mu\text{m} ; \quad |\Delta i|_{max} = 3.3 * 10^{-6} " ; \quad \Delta H_{max} = 4.1 * 10^{-2} \text{ m}^2/\text{s}$$

This demonstrates that the orbital perturbations due to the geomagnetic field are extremely small.

From (21.86) and (21.89) we find that after an orbital revolution ($u_e = 2\pi$):

$$\begin{aligned}\Delta_{2\pi} a &= 0 ; \quad \Delta_{2\pi} i = 0 ; \quad \Delta_{2\pi} H = 0 \\ \Delta_{2\pi} \Omega &= -2\pi \frac{q g_{1,0}}{M n_0} \left(\frac{R}{r_0} \right)^3 \left(1 - \frac{\dot{\theta}}{n_0} \cos i_0 \right)\end{aligned}\quad (21.90)$$

So, for our simplified geomagnetic field model, only the satellite orbit's right ascension of the ascending node shows a secular perturbation, which is only weakly dependent on the orbital inclination. For a satellite in a circular 500 km altitude orbit at $i_0 = 45^\circ$, and $U^* = -100$ V, $R_s/M = 0.1$ m/kg, we find $\Delta_{2\pi} \Omega = -2.9 * 10^{-5} "$. For a geostationary satellite $i_0 = 0^\circ$, $n_0 = \dot{\theta}$, and thus $\Delta_{2\pi} \Omega = 0$. The physical explanation is that this satellite is stationary with respect to the rotating geomagnetic field and thus does not generate a Lorentz force.

The very fact that the Lorentz force produces changes in the orbit of a satellite has led to the concept of a *Lorentz spacecraft*. In such a spacecraft the electrostatic charge of a capacitor is modulated as a function of time in such a way that an appropriate time-varying Lorentz force is generated to achieve prescribed changes of the orbital elements without spending any propellant. To realize an appreciable Lorentz force, of course, high levels of the electric charge are required.

21.7. Perturbations due to the $J_{2,2}$ -term of the gravity field

In Section 20.1 we have found the following expressions for the radial, latitudinal and longitudinal components of the acceleration of a satellite due to the $J_{2,2}$ -term of the Earth's gravity field potential:

$$\begin{aligned}
 f_r &= -9\mu J_{2,2} \frac{R^2}{r^4} \cos^2 \phi \cos 2(\Lambda - \Lambda_{2,2}) \\
 f_\phi &= -3\mu J_{2,2} \frac{R^2}{r^4} \sin 2\phi \cos 2(\Lambda - \Lambda_{2,2}) \\
 f_\Lambda &= -6\mu J_{2,2} \frac{R^2}{r^4} \cos \phi \sin 2(\Lambda - \Lambda_{2,2})
 \end{aligned} \tag{21.91}$$

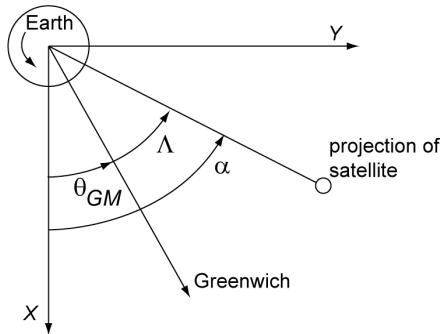


Figure 21.5: Relation between Greenwich mean sidereal time, θ_{GM} , longitude, Λ , and right ascension, α .

where μ is the Earth's gravitational parameter, $J_{2,2} = 1.82 \times 10^{-6}$ and $\Lambda_{2,2} = -14.9^\circ$ (Table 20.1) are constants specifying the $J_{2,2}$ sectorial harmonic of the geopotential, and R is the mean equatorial radius of the Earth. The position of a satellite relative to the geocentric non-rotating reference frame can be described by its radial distance, r , declination, δ , and right ascension, α (Figure 21.3). Now, we may write $\delta = \phi$, and, according to Figure 21.5 and (11.2),

$$\Lambda = \alpha - \theta_{GM} \tag{21.92}$$

where θ_{GM} is the Greenwich mean sidereal time (Section 11.4), which can be calculated as a function of time from

$$\theta_{GM} = \theta_{GM_0} + \dot{\theta} t \tag{21.93}$$

where $\dot{\theta}$ is the (mean) angular (rotational) velocity of the Earth, θ_{GM_0} is the Greenwich mean sidereal time at $0^{\text{h}}0^{\text{m}}0^{\text{s}}$ UT1, and t is counted from midnight UT1. Since $\dot{\theta} = 360.985647^\circ/\text{day}$, the longitude of a hypothetical observer at a fixed position in inertial space (above the Earth's surface) would decrease at a rate of $15.04^\circ/\text{hr}$ and the observer would experience a $J_{2,2}$ -force field that varies *explicitly* with time. A satellite encircling the Earth will experience a second time variation of the force field, because of the variation of the satellite's right ascension, α , and declination, δ , during an orbital revolution. It is emphasized that, in contrast to the analysis of luni-solar attraction and solar radiation pressure perturbations, the explicit time variation of the $J_{2,2}$ -force field can in first approximation not be neglected during an orbital period of the satellite. The reason is that even for a low-altitude satellite the rotation of the Earth during an orbital revolution of the satellite is at least 22° and for higher orbital altitudes this rotation angle further increases. It is the occurrence of these two different periods, i.e. the orbital period of the satellite and the period of the rotation of the Earth about its axis, that makes the analysis of orbit perturbations due to tesseral and sectorial harmonics in the model of the Earth's gravity field quite complicated. A linearized analysis is presented in Section 23.6. A detailed analysis is

beyond the scope of this book. Note that in our analysis of the orbital perturbations due to the J_2 -term of the geopotential, atmospheric drag and the geomagnetic field, this complication did not occur because the models which we have used to describe these forces are rotationally symmetric about the Earth's rotation axis.

However, for a geostationary satellite ($r = 42,164$ km) a simplified analysis of the effects of the $J_{2,2}$ -term can be made due to the fact that such a satellite keeps an almost fixed position above the equator relative to the (rotating) Earth. In general, a geostationary satellite has to remain within a deadband of 0.4° or less, centered at its nominal longitude, Λ_0 . The main reason for a precise position allocation is that nowadays the geostationary orbit houses a few hundred satellites and one does not want that satellites approach each other too closely. At the geostationary altitude, the width of a 0.4° deadband corresponds to an along-track distance of 294 km. It is the task of the thrusters of the satellite's orbit control system to counteract any position drift outside the deadband due to initial position and velocity errors, and perturbing forces. This requires rocket engine pulses when the spacecraft tends to cross one of the boundaries of the deadband. When electric propulsion systems are used, these pulses, of course, are replaced by finite-time engine operations. The smaller the deadband, generally, the more frequently stationkeeping maneuvers must be performed. So, for a geostationary satellite we may assume that $\Lambda - \Lambda_0$ and δ are very small and that the acceleration due to the $J_{2,2}$ -term is about constant for a considerable period of time; the magnitude and direction of this acceleration depends on the satellite's nominal position.

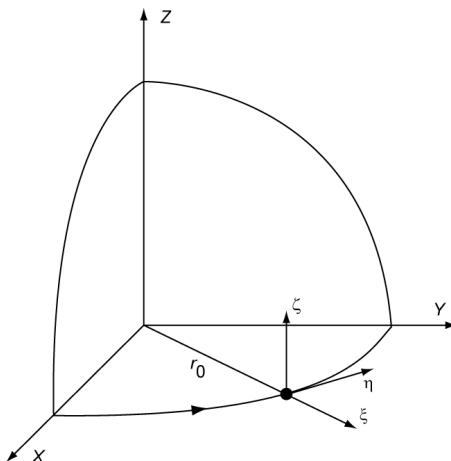


Figure 21.6: The reference frame $\xi\eta\zeta$ that is centered at a geostationary satellite.

In the following, we assume that a fictitious ‘ideal’ geostationary satellite is located at the position $\Lambda = \Lambda_0$, $\delta = 0^\circ$, and that the ‘true’ geostationary satellite is always close to the ‘ideal’ one. When we adopt the relative coordinates ξ , η , ζ for the position of the ‘true’ satellite relative to the ‘ideal’ satellite (Figure 21.6), where ξ is measured in the radial direction, η in the along-track (longitudinal) direction and ζ in the cross-track (latitudinal) direction, we can approximate the $J_{2,2}$ -acceleration components according to (21.91) by

$$f_\xi = -9\mu J_{2,2} \frac{R^2}{r_0^4} \cos 2(\Lambda_0 - \Lambda_{2,2}) \quad (21.94)$$

$$f_\eta = -6\mu J_{2,2} \frac{R^2}{r_0^4} \sin 2(\Lambda_0 - \Lambda_{2,2})$$

$$f_\zeta = 0 \quad (21.94)$$

For the maximum values of f_ξ and f_η we find from (21.94):

$$|f_\xi|_{max} = 8.4 * 10^{-8} \text{ m/s}^2 \text{ at } \Lambda_0 - \Lambda_{2,2} = 0^\circ, 90^\circ, 180^\circ, 270^\circ$$

$$|f_\eta|_{max} = 5.6 * 10^{-8} \text{ m/s}^2 \text{ at } \Lambda_0 - \Lambda_{2,2} = 45^\circ, 135^\circ, 225^\circ, 315^\circ$$

which confirms the results found in Section 20.1. Since the accelerations given by (21.94) are constant, we may apply the theory for relative motion presented in Section 9.2, and write (9.12) as

$$\begin{aligned} \xi &= \xi_0 (4 - 3 \cos \dot{\theta} t) + \frac{\dot{\xi}_0}{\dot{\theta}} \sin \dot{\theta} t + 2 \frac{\dot{\eta}_0}{\dot{\theta}} (1 - \cos \dot{\theta} t) + \frac{f_\xi}{\dot{\theta}^2} (1 - \cos \dot{\theta} t) \\ &\quad + 2 \frac{f_\eta}{\dot{\theta}^2} (\dot{\theta} t - \sin \dot{\theta} t) \\ \eta &= \eta_0 - \frac{\dot{\eta}_0}{\dot{\theta}} (3 \dot{\theta} t - 4 \sin \dot{\theta} t) - 6 \xi_0 (\dot{\theta} t - \sin \dot{\theta} t) - 2 \frac{\dot{\xi}_0}{\dot{\theta}} (1 - \cos \dot{\theta} t) \quad (21.95) \\ &\quad - 2 \frac{f_\xi}{\dot{\theta}^2} (\dot{\theta} t - \sin \dot{\theta} t) + 2 \frac{f_\eta}{\dot{\theta}^2} \left(2 - \frac{3}{4} \dot{\theta}^2 t^2 - 2 \cos \dot{\theta} t\right) \\ \zeta &= \zeta_0 \cos \dot{\theta} t + \frac{\dot{\zeta}_0}{\dot{\theta}} \sin \dot{\theta} t \end{aligned}$$

where we have used the fact that the mean motion of the ‘ideal’ satellite is equal to $\dot{\theta}$, and where t is the time since t_0 for which the initial conditions are $\xi_0, \dot{\xi}_0, \eta_0, \dot{\eta}_0, \zeta_0, \dot{\zeta}_0$. According to Section 9.2, these relations provide accurate results over a period of one day ($\dot{\theta} t \leq 2\pi$) and if the relative position components satisfy the following conditions: $|\xi|/r_0 < 8 * 10^{-3}$, $|\eta|/r_0 < 6 * 10^{-2}$, $|\zeta|/r_0 < 6 * 10^{-2}$. For the geostationary orbit these relative motion requirements correspond to 337 km, 2530 km and 2530 km, respectively. Expressions for the velocity components can be obtained by differentiation of (21.95).

The last expression of (21.95) shows that ζ performs a pure harmonic oscillation with a period of one day and an amplitude that is determined by the initial conditions ζ_0 and $\dot{\zeta}_0$. The first two equations can be written as

$$\begin{aligned} \xi &= \xi_0 + \left(3\xi_0 + 2 \frac{\dot{\eta}_0}{\dot{\theta}} + \frac{f_\xi}{\dot{\theta}^2}\right) (1 - \cos \dot{\theta} t) + \left(\frac{\dot{\xi}_0}{\dot{\theta}} - 2 \frac{f_\eta}{\dot{\theta}^2}\right) \sin \dot{\theta} t + 2 \frac{f_\eta}{\dot{\theta}} t \\ \eta &= \eta_0 - 2 \left(\frac{\dot{\xi}_0}{\dot{\theta}} - 2 \frac{f_\eta}{\dot{\theta}^2}\right) (1 - \cos \dot{\theta} t) + 2 \left(3\xi_0 + 2 \frac{\dot{\eta}_0}{\dot{\theta}} + \frac{f_\xi}{\dot{\theta}^2}\right) \sin \dot{\theta} t \quad (21.96) \\ &\quad - \left(6\xi_0 + 3 \frac{\dot{\eta}_0}{\dot{\theta}} + 2 \frac{f_\xi}{\dot{\theta}^2}\right) \dot{\theta} t - \frac{3}{2} f_\eta t^2 \end{aligned}$$

Note that these expressions for the relative motion of the satellite contain constant terms, periodic

terms and secular terms. The periodic terms have a period of one day and amplitudes that are a function of the initial relative position and velocity components and of the $J_{2,2}$ -acceleration components. The factors $|f_\xi|/\dot{\theta}^2$ and $|f_\eta|/\dot{\theta}^2$ are always smaller than 16 m and 11 m, respectively, which means that in most cases the contribution of the $J_{2,2}$ -acceleration components to the amplitude of the periodic terms is small. When the satellite crosses a boundary of the deadband, the appropriate thrusters are activated to reverse the direction of the satellite's relative velocity. The resulting relative position and velocity components after the impulse should be taken as the new initial conditions when using (21.96) again.

For the secular change of the radial and along-track position components we find from (21.96):

$$\begin{aligned} (\Delta\xi)_s &= 2 \frac{f_\eta}{\dot{\theta}} t \\ (\Delta\eta)_s &= -\left(6\xi_0 + 3 \frac{\dot{\eta}_0}{\dot{\theta}} + 2 \frac{f_\xi}{\dot{\theta}^2}\right) \dot{\theta} t - \frac{3}{2} f_\eta t^2 \end{aligned} \quad (21.97)$$

We conclude that the secular change of the radial position component is linear, while the secular change of the along-track position component also includes a quadratic term. If the satellite is positioned at a point where f_η is positive, (21.97-1) shows that the linear radial displacement is in the positive direction. At a higher altitude the mean angular motion of the satellite is lower than the rotational velocity of the Earth. This results in an along-track displacement component in the negative direction, which is expressed by the second term on the right-hand side of (21.97-2).

The value of f_η , of course, varies along the equator and it is interesting to determine the direction of the radial and along-track secular motions induced by f_η as a function of the longitudinal position of the satellite. In Section 20.1, we have analyzed the directions of the radial and along-track accelerations along the equator (Figure 20.2, bottom). In the two plots at the top of Figure 21.7 these results have been reproduced, using the notation applied in this Section. Note that f_η is positive if $90^\circ < \Lambda - \Lambda_{2,2} < 180^\circ$ and $270^\circ < \Lambda - \Lambda_{2,2} < 360^\circ$; and negative if $0^\circ < \Lambda - \Lambda_{2,2} < 90^\circ$ and $180^\circ < \Lambda - \Lambda_{2,2} < 270^\circ$. Substitution of these results for f_η into (21.97) leads to

$$\begin{aligned} (\Delta\xi)_s > 0 &: 90^\circ < \Lambda - \Lambda_{2,2} < 180^\circ \text{ and } 270^\circ < \Lambda - \Lambda_{2,2} < 360^\circ \\ (\Delta\xi)_s < 0 &: 0^\circ < \Lambda - \Lambda_{2,2} < 90^\circ \text{ and } 180^\circ < \Lambda - \Lambda_{2,2} < 270^\circ \\ (\Delta\eta)_s > 0 &: 0^\circ < \Lambda - \Lambda_{2,2} < 90^\circ \text{ and } 180^\circ < \Lambda - \Lambda_{2,2} < 270^\circ \\ (\Delta\eta)_s < 0 &: 90^\circ < \Lambda - \Lambda_{2,2} < 180^\circ \text{ and } 270^\circ < \Lambda - \Lambda_{2,2} < 360^\circ \end{aligned}$$

These results are depicted in the two plots at the bottom of Figure 21.7. We conclude that the points $\Lambda = 90^\circ + \Lambda_{2,2} \approx 75.1^\circ$ and $\Lambda = 270^\circ + \Lambda_{2,2} \approx 255.1^\circ$ are *stable equilibrium points* for the along-track acceleration due to the $J_{2,2}$ -term, while the points $\Lambda = \Lambda_{2,2} \approx -14.9^\circ$ and $\Lambda = 90^\circ + \Lambda_{2,2} \approx 104.9^\circ$ are *unstable equilibrium points*. The location of these equilibrium points changes slightly when the effects of all sectorial and tesseral harmonics are taken into account. At the stable equilibrium points any small longitude deviation from that position would induce a drift back towards the stable point. A geostationary spacecraft at any other longitude is accelerated towards the nearest of the two stable equilibrium points. If no stationkeeping maneuvers are executed, the longitude of the satellite would eventually swing back and forth

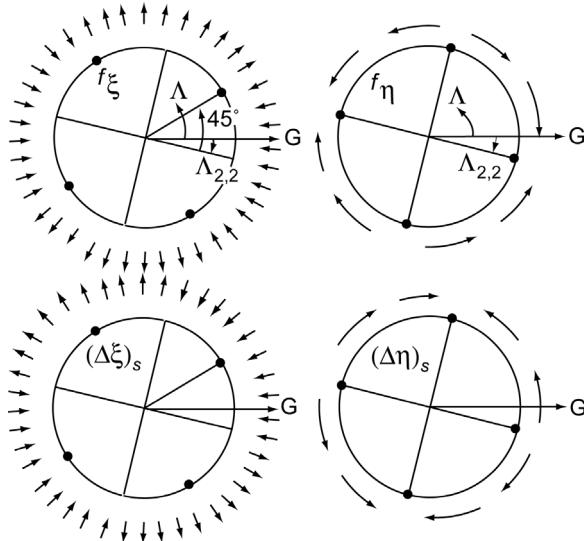


Figure 21.7: Diagrams for the direction of the $J_{2,2}$ -acceleration components f_ξ and f_η in the equatorial plane (top) and of the f_η -induced secular motion components in the ξ and η directions, in the equatorial plane (bottom). Points where an acceleration component is zero are indicated by •. The direction to the Greenwich meridian is indicated by G .

symmetrically around this stable equilibrium point with a period of more than two years. Since there is no damping, the swing will continue indefinitely with constant amplitude. For geostationary satellites that have to stay within a specified deadband, the effects of the $J_{2,2}$ -term require periodic thruster firings when the satellite reaches the east-boundary of the deadband if $0^\circ < \Lambda - \Lambda_{2,2} < 90^\circ$ or $180^\circ < \Lambda - \Lambda_{2,2} < 270^\circ$, or the west-boundary of the deadband if $90^\circ < \Lambda - \Lambda_{2,2} < 180^\circ$ or $270^\circ < \Lambda - \Lambda_{2,2} < 360^\circ$.

To get an impression of the evolution of the position deviations of a geostationary satellite, these deviations were computed from (21.96) for a period of one day. For this analysis, it was assumed that the satellite's nominal position is at 30° E, which is very close to the position where the longitudinal acceleration due to the $J_{2,2}$ -term is a maximum and the radial acceleration due to the $J_{2,2}$ -term is zero. For the initial position and velocity deviations quite arbitrary values of +100 m and +1 cm/s were adopted. According to (21.97-2), the initial position deviation ξ_0 and the initial velocity deviation $\dot{\eta}_0$ yield a negative linear secular variation $(\Delta\eta)_s$, the perturbing acceleration f_ξ (which is negative at 30° E) yields a positive linear secular variation $(\Delta\eta)_s$, and the perturbing acceleration f_η (which is also negative at 30° E) yields a positive quadratic secular variation $(\Delta\eta)_s$. According to (21.97-1), the perturbing acceleration f_η yields a negative linear secular variation $(\Delta\xi)_s$. Superimposed upon these secular variations are the periodic variations expressed by (21.96). Table 21.1 lists the minimum and maximum values of $\Delta\xi$ and $\Delta\eta$ during that day for a few combinations of the initial position and velocity deviations. We conclude that the $J_{2,2}$ -term produces a maximum value of 626 m in the along-track direction and a minimum value of -132 m in the radial direction. Although this is a significant orbit perturbation, we conclude that the effects of deviations in the initial conditions lead to even larger positions deviations. This illustrates that the orbit control of geostationary satellites has to be performed very accurately. An extended analysis of the motion of the satellite over some days, for the case that the initial position and velocity deviations have values of +100 m and +1 cm/s, has shown that: 1) the radial position deviation grows during the first 12 hours to about 1.2 km due to the initial position and velocity deviations; 2) after about 1 day the radial position deviation starts decreasing due to the

Table 21.1: Range of radial (ξ) and longitudinal (η) position deviations of a geostationary satellite positioned at $30^\circ E$ for a period of 1 day, due to initial position and velocity deviations and the $J_{2,2}$ -term of the Earth's gravity field.

Initial deviations		Position deviations within 1 day (km)			
ξ_0, η_0 (m)	$\dot{\xi}_0, \dot{\eta}_0$ (cm/s)	$(\Delta\xi)_{min}$	$(\Delta\xi)_{max}$	$(\Delta\eta)_{min}$	$(\Delta\eta)_{max}$
0	0	-0.132	0	-0.004	0.626
100	0	-0.032	0.637	-3.159	0.100
0	1	-0.164	0.536	-2.455	0.027
100	1	-0.048	1.210	-5.916	0.123

acceleration by the f_η -term in the $-\eta$ direction; 3) for the first five days the longitudinal position deviation grows to about -20 km due to the initial position and velocity deviations; 4) in the long run, the westward acceleration due to the f_η -term forces the spacecraft to start moving in the eastward direction. Of course, in reality, other perturbations act on the satellite, which will affect the evolution of position deviations. An important perturbation is produced by the gravitational attraction of Sun and Moon. These attractions cause north-south perturbations, which, if not corrected, would make the inclination of a geostationary satellite to vary over a range of about 15° with a period of about 55 years. Therefore, north-south stationkeeping is necessary and thrusters have to be fired when the spacecraft crosses the boundaries of a north-south deadband.

22. METHOD OF VARIATION OF ORBITAL ELEMENTS

In the discussion on Keplerian orbits and the introduction of the concept of orbital elements (Section 11.5) it was found that when at a certain moment of time position and velocity of a satellite are known, then the orbit is completely determined and is specified by the orbital elements a , e , i , ω , Ω and τ . These orbital elements are constants of motion; i.e. for every point along the orbit they have the same numerical value. In this Chapter we deal with the computation of perturbed elliptical satellite orbits. We then could apply the same transformation relations from Chapter 11 in order to compute from the momentary position and velocity the orbital elements of a *fictitious* momentary Keplerian elliptical orbit. This is the orbit that the satellite would follow if from that moment on no perturbation would act anymore on the satellite. This Keplerian orbit touches ('osculates') the true (perturbed) orbit of the satellite at that moment of time, t , and we speak of the *osculating Keplerian orbit* for that moment of time. In the point of contact of both orbits, position and velocity are identical for both orbits. Because perturbing forces act on the satellite, another osculating ellipse will be found for another moment of time, $t + \Delta t$.

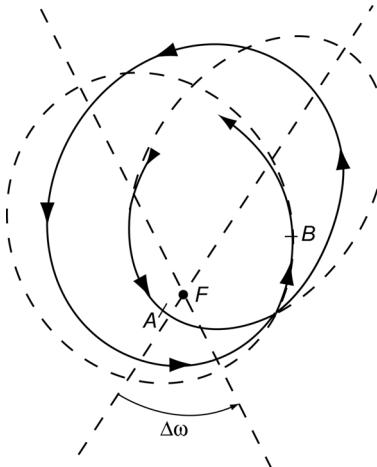


Figure 22.1: The true orbit and two osculating orbits.

Figure 22.1 shows a cartoon of two osculating Keplerian orbits touching the true (perturbed) orbit in points A and B. We now generalize this 'model' and consider the motion of the satellite as a continuous transition from one osculating Keplerian orbit to another osculating Keplerian orbit. So, the true, perturbed, satellite orbit is viewed as a sequence of small parts of a series of osculating Keplerian orbits, of which the orbital elements vary continuously. This concept forms the basis of the *method of variation of orbital elements*, also called the *method of variation of constants* or the *method of variation of parameters*, which yields a set of first-order differential equations for the variation of the osculating Keplerian orbital elements with time. Integration of these equations yields the values of the osculating orbital elements at any moment of time, from which we can determine the instantaneous position and velocity of the satellite in the perturbed orbit by using the transformations that hold for the pure Keplerian orbit (Section 11.8). When we numerically integrate the differential equations, then, in fact, we only integrate the differences between the perturbed orbit and a Keplerian reference orbit. So, the approach is similar to the method of Encke discussed in Section 20.6. However, the advantage of the method of variation of orbital elements is that the variation of the orbital elements provides a much clearer picture of the geometric/physical characteristics of the orbit perturbations than the variation of the rectangular position and velocity components. In addition, the method is very convenient for

obtaining approximative analytical solutions of the differential equations.

22.1. Lagrange's planetary equations

In this Section, we will limit ourselves to perturbing forces that can be described by a perturbing potential, R . The main force is expressed by the Newton potential

$$U = -\frac{\mu}{r} \quad (22.1)$$

In perturbation theory, we usually work with so-called *force functions*, which are defined as

$$\tilde{U} = -U \quad ; \quad \tilde{R} = -R \quad (22.2)$$

For the motion of the satellite under the influence of forces described by the potentials U and R , we then may write

$$\frac{d^2\bar{r}}{dt^2} + \frac{\mu}{r^3}\bar{r} = \bar{\nabla}\tilde{R} \quad (22.3)$$

When no perturbations would occur ($\tilde{R}=0$), then the solution of (22.3) is a pure Keplerian orbit, for which

$$\bar{r} = \bar{r}(\alpha_i, t) \quad ; \quad i = 1, 2, \dots, 6 \quad (22.4)$$

where α_i indicates an arbitrary (constant) orbital element. The relations (22.4) are known analytical expressions that have been derived in Section 11.8. Since for the unperturbed Keplerian orbit the orbital elements are constant, we may write

$$\frac{d\bar{r}}{dt} = \frac{\partial \bar{r}}{\partial t} \quad (22.5)$$

In that case, (22.3) can be written as

$$\frac{\partial^2\bar{r}}{\partial t^2} + \frac{\mu}{r^3}\bar{r} = 0 \quad (22.6)$$

For the perturbed orbit the problem can be formulated as: "How can the equations of motion (22.3) be satisfied, while using the relations (22.4) that hold for the (unperturbed) Keplerian orbit?" It will be clear that then the orbital elements in (22.4) can no longer be constant, but have to vary with time. The question then is: "How do the orbital elements have to vary with time?" In the following, that question will be addressed.

Differentiation of (22.4) with respect to time yields for the perturbed orbit

$$\frac{d\bar{r}}{dt} = \frac{\partial \bar{r}}{\partial t} + \sum_i \frac{\partial \bar{r}}{\partial \alpha_i} \frac{d\alpha_i}{dt} \quad (22.7)$$

We now use the fact that the instantaneous osculating orbit is a Keplerian one. Consequently, (22.4) and (22.5) apply to the instantaneous osculating orbit and a comparison between (22.5) and (22.7) shows that the following relation must hold for the osculating orbit:

$$\sum_i \frac{\partial \bar{r}}{\partial \alpha_i} \frac{d\alpha_i}{dt} = 0 \quad (22.8)$$

This is sometimes called the *Lagrange constraint* (J.A. Lagrange; 1736-1813). Differentiation of (22.5) to time results in

$$\frac{d^2 \bar{r}}{dt^2} = \frac{\partial^2 \bar{r}}{\partial t^2} + \sum_i \frac{\partial^2 \bar{r}}{\partial \alpha_i \partial t} \frac{d\alpha_i}{dt} \quad (22.9)$$

From this equation we conclude that, contrary to the velocities, the accelerations in the perturbed orbit and the osculating orbit are not equal at the point of osculation. Substitution of (22.9) into (22.3) yields

$$\frac{\partial^2 \bar{r}}{\partial t^2} + \sum_i \frac{\partial^2 \bar{r}}{\partial \alpha_i \partial t} \frac{d\alpha_i}{dt} + \frac{\mu}{r^3} \bar{r} = \bar{\nabla} \tilde{R}$$

Because the osculating orbit is a Keplerian one, for which (22.6) holds, the osculating orbit must satisfy the following condition:

$$\sum_i \frac{\partial^2 \bar{r}}{\partial \alpha_i \partial t} \frac{d\alpha_i}{dt} = \bar{\nabla} \tilde{R} \quad (22.10)$$

The relations (22.8) and (22.10) represent six scalar first-order differential equations in the six osculating orbital elements. This set of six equations is, in fact, equivalent to the original set of three second-order differential equations (22.3). Only a transformation of the time-derivatives of \bar{r} to the time-derivatives of α_i has taken place. In passing, it is remarked that the introduction of the Lagrange constraint is basically arbitrary. In principle, the right-hand side of (22.8) may be an arbitrary function of α_i , $d\alpha_i/dt$ and t , although the calculations then become much more complicated. The use of such functions results in a type of non-osculating orbital elements, which can be attractive for the analysis of orbit perturbations. In this book, we will, however, follow the classical Lagrange approach.

The form of the relations (22.8) and (22.10) is rather inconvenient; it is a set of six linear equations for the six terms $d\alpha_i/dt$. It would be more convenient to have explicit relations for each of the terms $d\alpha_i/dt$. Langrange has developed around 1780 a method to achieve that goal. In that method, we start by forming the scalar product of (22.8) and $\partial^2 \bar{r} / \partial \alpha_j \partial t$, and the scalar product of (22.10) and $\partial \bar{r} / \partial \alpha_j$, where α_j is an arbitrary orbital element. We then obtain

$$\sum_i \frac{\partial \bar{r}}{\partial \alpha_i} \cdot \frac{\partial^2 \bar{r}}{\partial \alpha_j \partial t} \frac{d\alpha_i}{dt} = 0 \quad (22.11)$$

$$\sum_i \frac{\partial \bar{r}}{\partial \alpha_j} \cdot \frac{\partial^2 \bar{r}}{\partial \alpha_i \partial t} \frac{d\alpha_i}{dt} = \bar{\nabla} \tilde{R} \cdot \frac{\partial \bar{r}}{\partial \alpha_j}$$

When (22.11-1) is subtracted from (22.11-2), we find

$$\sum_i \left[\frac{\partial \bar{r}}{\partial \alpha_j} \cdot \frac{\partial^2 \bar{r}}{\partial \alpha_i \partial t} - \frac{\partial^2 \bar{r}}{\partial \alpha_j \partial t} \cdot \frac{\partial \bar{r}}{\partial \alpha_i} \right] \frac{d\alpha_i}{dt} = \bar{\nabla} \tilde{R} \cdot \frac{\partial \bar{r}}{\partial \alpha_j} \quad (22.12)$$

The term within brackets is called the *Lagrange bracket* $[\alpha_j, \alpha_i]$:

$$[\alpha_j, \alpha_i] = \frac{\partial \bar{r}}{\partial \alpha_j} \cdot \frac{\partial^2 \bar{r}}{\partial \alpha_i \partial t} - \frac{\partial^2 \bar{r}}{\partial \alpha_j \partial t} \cdot \frac{\partial \bar{r}}{\partial \alpha_i} \quad (22.13)$$

with which (22.12) may be written as

$$\sum_i [\alpha_j, \alpha_i] \frac{d\alpha_i}{dt} = \bar{\nabla} \tilde{R} \cdot \frac{\partial \bar{r}}{\partial \alpha_j} \quad (22.14)$$

Since the partial derivative $\partial \bar{r} / \partial \alpha_j$ is computed from the relations that hold for an unperturbed Keplerian orbit, we may write

$$\bar{\nabla} \tilde{R} \cdot \frac{\partial \bar{r}}{\partial \alpha_j} = \frac{\partial \tilde{R}}{\partial x} \frac{\partial x}{\partial \alpha_j} + \frac{\partial \tilde{R}}{\partial y} \frac{\partial y}{\partial \alpha_j} + \frac{\partial \tilde{R}}{\partial z} \frac{\partial z}{\partial \alpha_j} \equiv \frac{\partial \tilde{R}}{\partial \alpha_j}$$

So, (22.14) can be written as

$$\sum_i [\alpha_j, \alpha_i] \frac{d\alpha_i}{dt} = \frac{\partial \tilde{R}}{\partial \alpha_j} \quad (22.15)$$

Since there are six orbital elements, (22.15) represents a set of six linear equations for the six unknowns $d\alpha_i/dt$. The Lagrange brackets $[\alpha_j, \alpha_i]$ have to be computed for each combination of α_i and α_j . This means that in total thirty-six brackets have to be computed.

The computation of these brackets can be simplified significantly, if we use three properties of Lagrange brackets. Two properties follow directly from the definition of the Lagrange bracket:

$$[\alpha_j, \alpha_j] = 0 \quad ; \quad [\alpha_j, \alpha_i] = -[\alpha_i, \alpha_j] \quad (22.16)$$

By using these properties, only fifteen brackets remain to be computed. A third property of a Lagrange bracket is that the partial derivative of a bracket to time is equal to zero. To prove this property, we differentiate (22.13) to time and find

$$\frac{\partial}{\partial t} [\alpha_j, \alpha_i] = \frac{\partial \bar{r}}{\partial \alpha_j} \cdot \frac{\partial^3 \bar{r}}{\partial \alpha_i \partial t^2} - \frac{\partial \bar{r}}{\partial \alpha_i} \cdot \frac{\partial^3 \bar{r}}{\partial \alpha_j \partial t^2} \quad (22.17)$$

For the unperturbed Keplerian orbit we may write according to (22.1), (22.2) and (22.6)

$$\frac{\partial^2 \bar{r}}{\partial t^2} = \bar{\nabla} \tilde{U}$$

Substitution of this relation into (22.17) yields

$$\frac{\partial}{\partial t} [\alpha_j, \alpha_i] = \frac{\partial \bar{r}}{\partial \alpha_j} \cdot \frac{\partial \bar{\nabla} \tilde{U}}{\partial \alpha_i} - \frac{\partial \bar{r}}{\partial \alpha_i} \cdot \frac{\partial \bar{\nabla} \tilde{U}}{\partial \alpha_j}$$

This expression can also be written as

$$\frac{\partial}{\partial t} [\alpha_j, \alpha_i] = \frac{\partial}{\partial \alpha_i} \left\{ \frac{\partial \bar{r}}{\partial \alpha_j} \cdot \bar{\nabla} \tilde{U} \right\} - \frac{\partial}{\partial \alpha_j} \left\{ \frac{\partial \bar{r}}{\partial \alpha_i} \cdot \bar{\nabla} \tilde{U} \right\}$$

After evaluation of the scalar products we find

$$\frac{\partial}{\partial t} [\alpha_j, \alpha_i] = \frac{\partial}{\partial \alpha_i} \left\{ \frac{\partial \tilde{U}}{\partial \alpha_j} \right\} - \frac{\partial}{\partial \alpha_j} \left\{ \frac{\partial \tilde{U}}{\partial \alpha_i} \right\} = 0 \quad (22.18)$$

This relation shows that a Lagrange bracket does not explicitly depend on time. This is a very important property for computing the remaining fifteen brackets efficiently. Because the partial derivative of a bracket to time is zero, we may compute the bracket at any point along the instantaneous osculating Keplerian orbit. The easiest way to compute a Lagrange bracket is to express all elements of that bracket as a function of the orbital elements, α_i , and the eccentric anomaly, E . Subsequently, the relations obtained are evaluated at pericenter, where $E = 0$ and $t = \tau$, after which the results are substituted in (22.13). There exists a number of other analytical techniques, which partially date back to F.F. Tisserand (1845-1896), to further reduce the amount of computational effort needed to find expressions for the Lagrange brackets. However, these will not be treated in this Chapter.

To compute the brackets, we start from the relations between $x, y, z, \dot{x}, \dot{y}, \dot{z}$ and the orbital elements, derived in Section 11.8, and write

$$\begin{aligned} x &= r(l_1 \cos \theta + l_2 \sin \theta) \\ y &= r(m_1 \cos \theta + m_2 \sin \theta) \\ z &= r(n_1 \cos \theta + n_2 \sin \theta) \\ \dot{x} &= \frac{\mu}{H} (-l_1 \sin \theta + l_2 (e + \cos \theta)) \\ \dot{y} &= \frac{\mu}{H} (-m_1 \sin \theta + m_2 (e + \cos \theta)) \\ \dot{z} &= \frac{\mu}{H} (-n_1 \sin \theta + n_2 (e + \cos \theta)) \end{aligned} \quad (22.19)$$

in which

$$\begin{aligned} l_1 &= \cos \omega \cos \Omega - \sin \omega \sin \Omega \cos i \\ m_1 &= \cos \omega \sin \Omega + \sin \omega \cos \Omega \cos i \\ n_1 &= \sin \omega \sin i \\ l_2 &= -\sin \omega \cos \Omega - \cos \omega \sin \Omega \cos i \\ m_2 &= -\sin \omega \sin \Omega + \cos \omega \cos \Omega \cos i \\ n_2 &= \cos \omega \sin i \end{aligned} \quad (22.20)$$

In Section 11.8, we have found that the following relations hold:

$$l_1^2 + m_1^2 + n_1^2 = l_2^2 + m_2^2 + n_2^2 = l_3^2 + m_3^2 + n_3^2 = 1 \quad (22.21)$$

$$l_1 l_2 + m_1 m_2 + n_1 n_2 = l_1 l_3 + m_1 m_3 + n_1 n_3 = l_2 l_3 + m_2 m_3 + n_2 n_3 = 0 \quad (22.21)$$

In Section 6.5, we have derived for an elliptical orbit the following expressions:

$$\begin{aligned} r &= a(1 - e \cos E) \\ r \sin \theta &= a \sqrt{1 - e^2} \sin E ; \quad r \cos \theta = a(\cos E - e) \\ E - e \sin E &= \sqrt{\frac{\mu}{a^3}} (t - \tau) \end{aligned} \quad (22.22)$$

Substitution of these relations into (22.19) yields

$$\begin{aligned} x &= a \left[l_1 (\cos E - e) + l_2 \sqrt{1 - e^2} \sin E \right] \\ y &= a \left[m_1 (\cos E - e) + m_2 \sqrt{1 - e^2} \sin E \right] \\ z &= a \left[n_1 (\cos E - e) + n_2 \sqrt{1 - e^2} \sin E \right] \\ \dot{x} &= \sqrt{\frac{\mu}{a}} \left[-l_1 \frac{\sin E}{1 - e \cos E} + l_2 \sqrt{1 - e^2} \frac{\cos E}{1 - e \cos E} \right] \\ \dot{y} &= \sqrt{\frac{\mu}{a}} \left[-m_1 \frac{\sin E}{1 - e \cos E} + m_2 \sqrt{1 - e^2} \frac{\cos E}{1 - e \cos E} \right] \\ \dot{z} &= \sqrt{\frac{\mu}{a}} \left[-n_1 \frac{\sin E}{1 - e \cos E} + n_2 \sqrt{1 - e^2} \frac{\cos E}{1 - e \cos E} \right] \end{aligned} \quad (22.23)$$

By using the relations (22.20) to (22.23) we can derive expressions for all Lagrange brackets.

As an example, we will now evaluate the bracket $[a, \tau]$. For simplicity, we will divide this bracket into three parts: $[a, \tau]_x, [a, \tau]_y, [a, \tau]_z$, where

$$[a, \tau]_x = \frac{\partial x}{\partial a} \frac{\partial^2 x}{\partial \tau \partial t} - \frac{\partial^2 x}{\partial a \partial t} \frac{\partial x}{\partial \tau}, \quad \text{etc.} \quad (22.24)$$

So, we have to determine expressions for the partial derivatives $\partial x / \partial a, \partial \dot{x} / \partial a, \partial x / \partial \tau, \partial \dot{x} / \partial \tau$. From (22.23-1) we find

$$\frac{\partial x}{\partial a} = l_1 (\cos E - e) - l_1 a \sin E \frac{\partial E}{\partial a} + l_2 \sqrt{1 - e^2} \sin E + l_2 a \sqrt{1 - e^2} \cos E \frac{\partial E}{\partial a} \quad (22.25)$$

From (22.22-4) we find

$$\frac{\partial E}{\partial a} = -\frac{3}{2} \sqrt{\frac{\mu}{a^5}} \frac{t - \tau}{1 - e \cos E} \quad (22.26)$$

Substitution of this relation into (22.25) and subsequently substituting $E = 0, t = \tau$ yields

$$\frac{\partial x}{\partial a} = l_1(1 - e) \quad (22.27)$$

From (22.23-4) we find

$$\begin{aligned} \frac{\partial^2 x}{\partial a \partial t} &= \frac{1}{2} l_1 \sqrt{\frac{\mu}{a^3}} \frac{\sin E}{1 - e \cos E} - l_1 \sqrt{\frac{\mu}{a}} \frac{(1 - e \cos E) \cos E + e \sin^2 E}{(1 - e \cos E)^2} \frac{\partial E}{\partial a} \\ &\quad - \frac{1}{2} l_2 \sqrt{\frac{\mu}{a^3} (1 - e^2)} \frac{\cos E}{1 - e \cos E} \\ &\quad - l_2 \sqrt{\frac{\mu}{a} (1 - e^2)} \frac{(1 - e \cos E) \sin E + e \sin E \cos E}{(1 - e \cos E)^2} \frac{\partial E}{\partial a} \end{aligned}$$

Substitution of (22.26) into this equation and evaluation of the result for $E = 0, t = \tau$ yields

$$\frac{\partial^2 x}{\partial a \partial t} = -\frac{1}{2} l_2 \sqrt{\frac{\mu}{a^3} \left(\frac{1+e}{1-e} \right)} \quad (22.28)$$

From (22.23-1) we find

$$\frac{\partial x}{\partial \tau} = -l_1 a \sin E \frac{\partial E}{\partial \tau} + l_2 a \sqrt{1 - e^2} \cos E \frac{\partial E}{\partial \tau} \quad (22.29)$$

From (22.22-4) we find

$$\frac{\partial E}{\partial \tau} - e \cos E \frac{\partial E}{\partial \tau} = -\sqrt{\frac{\mu}{a^3}} \quad (22.30)$$

Substitution of this relation into (22.29) and evaluation of the result for $E = 0$ leads to

$$\frac{\partial x}{\partial \tau} = -l_2 \sqrt{\frac{\mu}{a} \left(\frac{1+e}{1-e} \right)} \quad (22.31)$$

From (22.23-4) we find

$$\begin{aligned} \frac{\partial^2 x}{\partial \tau \partial t} &= -l_1 \sqrt{\frac{\mu}{a}} \frac{(1 - e \cos E) \cos E - e \sin^2 E}{(1 - e \cos E)^2} \frac{\partial E}{\partial \tau} \\ &\quad - l_2 \sqrt{\frac{\mu}{a} (1 - e^2)} \frac{(1 - e \cos E) \sin E + e \sin E \cos E}{(1 - e \cos E)^2} \frac{\partial E}{\partial \tau} \end{aligned}$$

Substitution of (22.30) into this relation and evaluation of the result for $E = 0$ gives

$$\frac{\partial^2 x}{\partial \tau \partial t} = l_1 \frac{\mu}{a^2} \frac{1}{(1-e)^2} + l_2 \frac{\mu}{a^2} \frac{\sqrt{1-e^2}}{(1-e)^3} \quad (22.32)$$

Substitution of (22.27), (22.28), (22.31) and (22.32) into (22.24) yields

$$[a, \tau]_x = l_1^2 \frac{\mu}{a^2} \left(\frac{1}{1-e} \right) + l_1 l_2 \frac{\mu}{a^2} \frac{\sqrt{1-e^2}}{(1-e)^2} - \frac{1}{2} l_2^2 \frac{\mu}{a^2} \left(\frac{1+e}{1-e} \right)$$

Similarly, we obtain

$$[a, \tau]_y = m_1^2 \frac{\mu}{a^2} \left(\frac{1}{1-e} \right) + m_1 m_2 \frac{\mu}{a^2} \frac{\sqrt{1-e^2}}{(1-e)^2} - \frac{1}{2} m_2^2 \frac{\mu}{a^2} \left(\frac{1+e}{1-e} \right)$$

$$[a, \tau]_z = n_1^2 \frac{\mu}{a^2} \left(\frac{1}{1-e} \right) + n_1 n_2 \frac{\mu}{a^2} \frac{\sqrt{1-e^2}}{(1-e)^2} - \frac{1}{2} n_2^2 \frac{\mu}{a^2} \left(\frac{1+e}{1-e} \right)$$

With (22.21) we finally obtain the very simple relation

$$[a, \tau] = \frac{\mu}{2a^2}$$

In the same way, the other fourteen Lagrange brackets can be determined. We then find

$$[a, e] = [a, i] = [e, i] = [e, \tau] = 0$$

$$[i, \omega] = [i, \tau] = [\omega, \Omega] = [\omega, \tau] = [\Omega, \tau] = 0$$

$$[a, \omega] = -\frac{1}{2} \sqrt{\frac{\mu}{a} (1-e^2)} ; \quad [a, \Omega] = -\frac{1}{2} \sqrt{\frac{\mu}{a} (1-e^2) \cos i} ; \quad [a, \tau] = \frac{\mu}{2a^2}$$

$$[e, \omega] = \sqrt{\frac{\mu a}{1-e^2}} e ; \quad [e, \Omega] = \sqrt{\frac{\mu a}{1-e^2}} e \cos i$$

$$[i, \Omega] = \sqrt{\mu a (1-e^2)} \sin i$$

(22.33)

Note that nine of these fifteen Lagrange brackets have the value zero.

When these expressions are substituted into the set of linear equations (22.15), then the following equations are obtained:

$$\begin{aligned} -\frac{1}{2} \sqrt{\frac{\mu}{a} (1-e^2)} \frac{d\omega}{dt} - \frac{1}{2} \sqrt{\frac{\mu}{a} (1-e^2) \cos i} \frac{d\Omega}{dt} + \frac{\mu}{2a^2} \frac{d\tau}{dt} &= \frac{\partial \tilde{R}}{\partial a} \\ \sqrt{\frac{\mu a}{1-e^2}} e \frac{d\omega}{dt} + \sqrt{\frac{\mu a}{1-e^2}} e \cos i \frac{d\Omega}{dt} &= \frac{\partial \tilde{R}}{\partial e} \\ \sqrt{\mu a (1-e^2)} \sin i \frac{d\Omega}{dt} &= \frac{\partial \tilde{R}}{\partial i} \\ \frac{1}{2} \sqrt{\frac{\mu}{a} (1-e^2)} \frac{da}{dt} - \sqrt{\frac{\mu a}{1-e^2}} e \frac{de}{dt} &= \frac{\partial \tilde{R}}{\partial \omega} \end{aligned} \quad (22.34)$$

$$\begin{aligned} \frac{1}{2} \sqrt{\frac{\mu}{a} (1 - e^2)} \cos i \frac{da}{dt} - \sqrt{\frac{\mu a}{1 - e^2}} e \cos i \frac{de}{dt} - \sqrt{\mu a (1 - e^2)} \sin i \frac{di}{dt} &= \frac{\partial \tilde{R}}{\partial \Omega} \\ - \frac{\mu}{2a^2} \frac{da}{dt} &= \frac{\partial \tilde{R}}{\partial \tau} \end{aligned} \quad (22.34)$$

From this set of equations, explicit expressions for $d\alpha_i/dt$ can be obtained. From (22.34-6), we directly find a relation for da/dt . Substitution of that relation into (22.34-4) yields an expression for de/dt . Substitution of these expressions into (22.34-5) yields an expression for di/dt , etc. In this way, we finally obtain the following relations:

$$\begin{aligned} \frac{da}{dt} &= -2 \frac{a^2}{\mu} \frac{\partial \tilde{R}}{\partial \tau} \\ \frac{de}{dt} &= -\frac{a(1 - e^2)}{\mu e} \frac{\partial \tilde{R}}{\partial \tau} - \frac{1}{e} \sqrt{\frac{1 - e^2}{\mu a}} \frac{\partial \tilde{R}}{\partial \omega} \\ \frac{di}{dt} &= \frac{\cot i}{\sqrt{\mu a (1 - e^2)}} \frac{\partial \tilde{R}}{\partial \omega} - \frac{1}{\sqrt{\mu a (1 - e^2)} \sin i} \frac{\partial \tilde{R}}{\partial \Omega} \\ \frac{d\omega}{dt} &= \frac{1}{e} \sqrt{\frac{1 - e^2}{\mu a}} \frac{\partial \tilde{R}}{\partial e} - \frac{\cot i}{\sqrt{\mu a (1 - e^2)}} \frac{\partial \tilde{R}}{\partial i} \\ \frac{d\Omega}{dt} &= \frac{1}{\sqrt{\mu a (1 - e^2)} \sin i} \frac{\partial \tilde{R}}{\partial i} \\ \frac{d\tau}{dt} &= 2 \frac{a^2}{\mu} \frac{\partial \tilde{R}}{\partial a} + \frac{a(1 - e^2)}{\mu e} \frac{\partial \tilde{R}}{\partial e} \end{aligned} \quad (22.35)$$

So, our goal has been achieved: we have obtained six simultaneous first-order differential equations that express the effect of a perturbing force on each of the osculating orbital elements. These differential equations are called *Lagrange's planetary equations*. Note that the right-hand sides of these equations contain the orbital elements a, e, i , and the partial derivatives of \tilde{R} to all orbital elements. Also note that we have derived these equations for elliptical orbits. For hyperbolic trajectories a similar set of planetary equations can be derived; we then find that in these equations $\sqrt{1 - e^2}$ is replaced by $\sqrt{e^2 - 1}$ and that some terms have an opposite sign when compared to the same terms in the equations for elliptical motion. In this book we will not discuss the set of equations for hyperbolic motion.

It is interesting to note that in the derivation of (22.35) it was nowhere assumed that the perturbing force has to be small. These equations are therefore generally valid. However, when the perturbing accelerations are in magnitude comparable to the main acceleration (μ/r^2), then the orbital elements will show large variations over an orbital revolution and the method of Lagrange will lose its advantage over the much simpler method of Cowell.

When the perturbing function \tilde{R} is known as a function of x, y, z , then the partial derivative $\partial \tilde{R}/\partial \alpha_j$ in (22.35) can be obtained from

$$\frac{\partial \tilde{R}}{\partial \alpha_j} = \frac{\partial \tilde{R}}{\partial x} \frac{\partial x}{\partial \alpha_j} + \frac{\partial \tilde{R}}{\partial y} \frac{\partial y}{\partial \alpha_j} + \frac{\partial \tilde{R}}{\partial z} \frac{\partial z}{\partial \alpha_j} \quad (22.36)$$

With this equation, the set of differential equations (22.35) can be solved numerically and the osculating orbital elements can be computed as a function of time. Then, with the Keplerian transformation relations between position and velocity on the one hand and orbital elements on the other hand, the instantaneous position and velocity of the satellite in its perturbed orbit can be computed. Although the equations that have to be integrated are more complicated than, for instance, the equations in the method of Encke, a numerical integration of Lagrange's planetary equations has a number of advantages. Often, it turns out that the variation of the osculating orbital elements with time is somewhat smoother than the variation of Δx , Δy , Δz , which allows larger integration steps and therefore a faster computation process. The variation of the osculating orbital elements also provides a better physical insight into the nature of the perturbations. In addition, no rectification process is needed; we may also say that in the method of variation of orbital elements the orbit is rectified continuously.

In general, the set (22.35) cannot be solved in a closed analytical way. However, if we deal with small perturbations, we can obtain an approximative analytical solution for these equations, because the orbital elements only experience a rather small variation if the perturbing force is small. In Section 22.6, a method to find an approximative analytical solution will be discussed in some detail.

22.2. Canonic form of the planetary equations

Lagrange's planetary equations express the change of the osculating classical orbital elements as a result of perturbing forces. In principle, one may derive such equations for any other set of six parameters that describe a Keplerian orbit, as long as these parameters form a set of six independent combinations of the classical orbital elements. Examples of such sets of parameters are already given in Section 11.5 and will be given in Section 22.4. In astrodynamics, also the following independent set of orbital constants is frequently used:

$$\begin{aligned} L &= \sqrt{\mu a} & ; \quad l &= M \\ G &= \sqrt{\mu a(1 - e^2)} & ; \quad g &= \omega \\ H &= \sqrt{\mu a(1 - e^2)} \cos i & ; \quad h &= \Omega \end{aligned} \quad (22.37)$$

These variables were first used by C.E. Delaunay (1816-1872) in his studies on the motion of the Moon. The parameter L is directly linked to the semi-major axis and thus to the total orbital energy per unit of mass, while the parameters G and H represent the orbital angular momentum and the component of the orbital angular momentum about the Z-axis, both per unit of mass. When the orbit is described by these elements, the Lagrange brackets have a very simple form:

$$\begin{aligned} [L, l] &= -[l, L] = -1 \\ [G, g] &= -[g, G] = -1 \\ [H, h] &= -[h, H] = -1 \end{aligned} \quad (22.38)$$

while all other brackets are zero. This characteristic simplifies the construction of perturbation solutions. Then, Lagrange's planetary equations take an exceptionally simple form:

$$\begin{aligned}\frac{dL}{dt} &= \frac{\partial \mathcal{H}}{\partial l} & ; \quad \frac{dl}{dt} &= -\frac{\partial \mathcal{H}}{\partial L} \\ \frac{dG}{dt} &= \frac{\partial \mathcal{H}}{\partial g} & ; \quad \frac{dg}{dt} &= -\frac{\partial \mathcal{H}}{\partial G} \\ \frac{dH}{dt} &= \frac{\partial \mathcal{H}}{\partial h} & ; \quad \frac{dh}{dt} &= -\frac{\partial \mathcal{H}}{\partial H}\end{aligned}\tag{22.39}$$

where

$$\mathcal{H} = \frac{\mu^2}{2L^2} + \tilde{R} \quad ; \quad \tilde{R} = \tilde{R}(L, G, H, l, g, h)\tag{22.39}$$

The set of orbital parameters (22.37) is called a *canonical set* and (22.39) represents the *canonical form* or *Hamilton form* of the planetary equations. Note that the *Hamiltonian* \mathcal{H} equals minus the total orbital energy per unit of mass ($\mu^2/2L^2 = \mu/2a$) of a Keplerian orbit plus the perturbing force function, which may also be called the *perturbing Hamiltonian*, and that this force function, of course, has to be expressed in terms of the canonical elements. The Delaunay elements can be considered as the canonical counterpart of the classical orbital elements. As such, they contain singularities for small values of eccentricity and inclination. There are many more canonical sets of orbital elements known. Some of them, like the set of Poincaré elements (J.H. Poincaré, 1854-1912), are non-singular and may well be applied for small eccentricities and inclinations. These sets may therefore be considered as the canonical counterpart of equinoctial elements that will be introduced in Section 22.4.

W.R. Hamilton (1805-1865) developed the canonical form of the equations of motion and showed that if the parameters used to represent the motion are canonical, the equations of motion for a conservative force become particularly simple. His canonical form of the equations of motion can be extended for non-conservative perturbations. The application of canonical variables to perturbed motion is embedded in the generalized approach to classical mechanics, which originates with Lagrange's work in analytical dynamics and his use of generalized coordinates and momenta. The canonical form of the planetary equations is widely used for computing higher-order effects of perturbing forces and therefore it forms the basis of what could be called 'advanced orbit perturbation methods'. We will not discuss this topic further.

22.3. Use of the mean anomaly

The equation for $d\tau/dt$ introduces a serious problem. As (22.35) shows, the equation for $d\tau/dt$ is the only one for which $\partial\tilde{R}/\partial a$ has to be computed. This means that one has to determine $\partial x/\partial a$, $\partial y/\partial a$ and $\partial z/\partial a$, and thus one has to compute $\partial E/\partial a$. From (22.26) and (6.36-1) we obtain

$$\frac{\partial E}{\partial a} = -\frac{3}{2} \sqrt{\frac{\mu}{a^5}} \frac{t - \tau}{1 - e \cos E} = -\frac{3}{2a} \frac{E - e \sin E}{1 - e \cos E}$$

The appearance of E outside trigonometric arguments results in a complicated expression for $\partial\tilde{R}/\partial a$, which complicates the numerical integration (special perturbations techniques) of the planetary equations. For general perturbations techniques, the situation is even worse. We then end up with integrals of the kind

$$\int \frac{E \sin^k E \cos^m E}{(1 - e \cos E)^n}$$

for which in most cases no closed-form analytical solutions exist. However, there exists a classical method with which these problems can be circumvented.

In that method, we start by rewriting (22.35-6) as

$$\frac{d\tau}{dt} = \frac{2}{n^2 a} \frac{\partial \tilde{R}}{\partial a} + \frac{1}{n^2 a^2} \frac{1 - e^2}{e} \frac{\partial \tilde{R}}{\partial e} \quad (22.40)$$

where the relation between n and a is given by $n = \sqrt{\mu/a^3}$. From this relation we obtain

$$\frac{\partial n}{\partial a} = -\frac{3}{2} \frac{n}{a} ; \quad \frac{dn}{dt} = -\frac{3}{2} \frac{n}{a} \frac{da}{dt} \quad (22.41)$$

We now consider $n(a)$ as a separate variable and write

$$\frac{\partial \tilde{R}}{\partial a} = \left(\frac{\partial \tilde{R}}{\partial a} \right)_n + \frac{\partial \tilde{R}}{\partial n} \frac{\partial n}{\partial a} \quad (22.42)$$

where $(\partial \tilde{R}/\partial a)_n$ signifies that the derivative has to be taken with respect to a without considering the dependence of n on a . From the definition of the mean anomaly, M ,

$$M = n(t - \tau) \quad (22.43)$$

we find

$$\frac{\partial \tilde{R}}{\partial n} = \frac{\partial \tilde{R}}{\partial M} \frac{\partial M}{\partial n} = (t - \tau) \frac{\partial \tilde{R}}{\partial M} ; \quad \frac{\partial \tilde{R}}{\partial \tau} = \frac{\partial \tilde{R}}{\partial M} \frac{\partial M}{\partial \tau} = -n \frac{\partial \tilde{R}}{\partial M}$$

or

$$\frac{\partial \tilde{R}}{\partial n} = -\frac{t - \tau}{n} \frac{\partial \tilde{R}}{\partial \tau} \quad (22.44)$$

Differentiation of (22.43) to time yields

$$\frac{dM}{dt} = n(1 - \frac{d\tau}{dt}) + (t - \tau) \frac{dn}{dt}$$

Substitution of (22.35-1), (22.40), (22.41), (22.42) and (22.44) into this relation results in

$$\frac{dM}{dt} = n - \frac{2}{na} \left(\frac{\partial \tilde{R}}{\partial a} \right)_n - \frac{1 - e^2}{na^2 e} \frac{\partial \tilde{R}}{\partial e} \quad (22.45)$$

This equation, which replaces the equation for $d\tau/dt$ ((22.35-6)), has the important property that $\partial \tilde{R}/\partial a$, and so $\partial E/\partial a$, has to be evaluated without considering the dependence of n on a . In that case we find from (22.22-4): $(\partial E/\partial a)_n = 0$. This means that we do not get anymore expressions in which E appears outside a trigonometric argument. This is the reason why for the computation of perturbed satellite orbits we mostly use (22.45) instead of (22.35-6). It is then logical to replace the expression $\partial \tilde{R}/\partial \tau$ in the planetary equations for a and e by

$$\frac{\partial \tilde{R}}{\partial M} = -\frac{1}{n} \frac{\partial \tilde{R}}{\partial \tau} \quad (22.46)$$

It is emphasized that, as already stated in Section 11.5, the orbital parameter M varies linearly with time in an unperturbed Keplerian orbit, while the orbital elements $a, e, i, \omega, \Omega, \tau$ are constants for such an orbit.

22.4 Singularities of the planetary equations

Inspection of the relations (22.35) and (22.45) reveals that a problem arises if $e = 0$ or $\sin i = 0$, unless the corresponding term $\partial \tilde{R}/\partial \alpha_i$ contains factors with e or $\sin i$ in the numerator. Then, singularities will occur, causing the time-derivatives of some orbital elements to become infinitely large. Also, when e and i are very small, problems will arise. Then, the time-derivatives of some orbital elements may become so large that an analytical solution of the equations (22.35) and (22.45) is not possible. It is emphasized that these problems have nothing to do with the Lagrange method itself, but are a consequence of the selection of the set of constants that describe the orbit: the orbital elements. This can be illustrated as follows. Consider an elliptical orbit with a very small eccentricity and assume that a hypothetical perturbing force tends to shorten the major axis and to elongate the minor axis. At a certain moment, the eccentricity goes through zero and the ‘old’ major axis becomes the ‘new’ minor axis. At that moment, the argument of perigee, ω , changes discontinuously by 90° and also the orbital element τ changes discontinuously. This makes the time-derivative of these orbital elements to become infinitely large. A similar problem occurs when at $i \approx 0^\circ$ a perturbing force tends to decrease the inclination and let it pass through zero; this will result in a discontinuous change of i, Ω and ω . One should realize that this kind of problems can always be prevented by using another set of appropriately selected elements to describe the orbit.

For example, let we consider a (near-)circular orbit. Then, (22.35) shows that the expressions for $de/dt, d\omega/dt$ and $d\tau/dt$ may become (nearly) singular. When we replace the elements e, ω, τ by three other elements:

$$l = e \cos \omega \quad ; \quad m = e \sin \omega \quad ; \quad \chi = \omega - n \tau \quad (22.47-1)$$

develop expressions for $dl/dt, dm/dt, d\chi/dt$ to replace the expressions for $de/dt, d\omega/dt, d\tau/dt$, and write the right-hand sides of (22.35) in terms of the elements a, l, m, i, Ω, χ , we obtain

$$\begin{aligned} \frac{da}{dt} &= 2 \sqrt{\frac{a}{\mu}} \frac{\partial \tilde{R}}{\partial \chi} \\ \frac{dl}{dt} &= -\sqrt{\frac{1-l^2-m^2}{\mu a}} \frac{\partial \tilde{R}}{\partial m} + \frac{m \cot i}{\sqrt{\mu a(1-l^2-m^2)}} \frac{\partial \tilde{R}}{\partial i} - \frac{l \sqrt{1-l^2-m^2}}{\sqrt{\mu a(1+\sqrt{1-l^2-m^2})}} \frac{\partial \tilde{R}}{\partial \chi} \\ \frac{dm}{dt} &= \sqrt{\frac{1-l^2-m^2}{\mu a}} \frac{\partial \tilde{R}}{\partial l} - \frac{l \cot i}{\sqrt{\mu a(1-l^2-m^2)}} \frac{\partial \tilde{R}}{\partial i} - \frac{m \sqrt{1-l^2-m^2}}{\sqrt{\mu a(1+\sqrt{1-l^2-m^2})}} \frac{\partial \tilde{R}}{\partial \chi} \\ \frac{di}{dt} &= \frac{\cot i}{\sqrt{\mu a(1-l^2-m^2)}} \left(l \frac{\partial \tilde{R}}{\partial m} - m \frac{\partial \tilde{R}}{\partial l} + \frac{\partial \tilde{R}}{\partial \chi} \right) - \frac{1}{\sqrt{\mu a(1-l^2-m^2)} \sin i} \frac{\partial \tilde{R}}{\partial \Omega} \end{aligned} \quad (22.48)$$

$$\begin{aligned}\frac{d\Omega}{dt} &= \frac{1}{\sqrt{\mu a(1-l^2-m^2)} \sin i} \frac{\partial \tilde{R}}{\partial i} \\ \frac{d\chi}{dt} &= \frac{\sqrt{1-l^2-m^2}}{\sqrt{\mu a(1+\sqrt{1-l^2-m^2})}} \left(l \frac{\partial \tilde{R}}{\partial l} + m \frac{\partial \tilde{R}}{\partial m} \right) - 2\sqrt{\frac{a}{\mu}} \frac{\partial \tilde{R}}{\partial a} - \frac{\cot i}{\sqrt{\mu a(1-l^2-m^2)}} \frac{\partial \tilde{R}}{\partial i}\end{aligned}\quad (22.48)$$

Note that the elements l and m are the projections (in the orbital plane) of the eccentricity vector introduced in Section 5.7 on the line of nodes and normal to the line of nodes, respectively. We conclude that the variation of these orbital elements due to a perturbing force expressed by \tilde{R} can be evaluated without difficulty, because there is no singularity if $e^2 = l^2 + m^2 \downarrow 0$. Note that these expressions should not be used if $e^2 = l^2 + m^2 \uparrow 1$, $i \downarrow 0^\circ$, or $i \uparrow 180^\circ$, since then the equations will exhibit singularities.

For orbits that are both (near-)circular and (near-)equatorial, the set of classical orbital elements may be replaced by the set

$$\begin{aligned}a &; \quad l = e \cos(\Omega + \omega) \quad ; \quad m = e \sin(\Omega + \omega) \\ p &= \tan \frac{1}{2} i \sin \Omega \quad ; \quad q = \tan \frac{1}{2} i \cos \Omega \quad ; \quad \lambda_{m,0} = \Omega + \omega + M_0\end{aligned}\quad (22.47-2)$$

where the elements Ω and ω , of course, are measured in different planes, and M_0 is the mean anomaly at epoch. This set does not lead to singularities in the planetary equations for any elliptical orbit, including orbits with $i = 0^\circ$ and $e = 0$, except for virtually non-existent orbits with $i = 180^\circ$, and was proposed by R.A. Broucke (1932-2005) and P.J. Cefola (-) in 1972. Several other closely related sets of non-singular elements have been studied in the literature. These types of elements were named *equinoctial elements* by J.L. Arsenault (-) in 1970, but there are indications that S. Herrick (1911-1974) has first suggested the name prior to 1967. Unfortunately, the benefits of the equinoctial orbital elements do not come without a price. Working with these elements can quickly become very complicated mathematically. Various authors have proposed alternative sets of elements for which the Lagrange planetary equations do not contain singularities. We will not discuss this issue any further. In the following, we will only use the set of classical orbital elements.

22.5. Gauss' form of the planetary equations

Lagrange's planetary equations (22.35) and (22.45) can only be used for perturbing forces that can be expressed by a perturbing potential. However, the only place in the derivation of these equations where this assumption has been used, is when we applied in Section 22.1 the relation

$$\frac{\partial \tilde{R}}{\partial x} \frac{\partial x}{\partial \alpha_j} + \frac{\partial \tilde{R}}{\partial y} \frac{\partial y}{\partial \alpha_j} + \frac{\partial \tilde{R}}{\partial z} \frac{\partial z}{\partial \alpha_j} = \frac{\partial \tilde{R}}{\partial \alpha_j}$$

We may interpret the substitution of $\partial \tilde{R}/\partial \alpha_j$ for the three terms on the left-hand side of this equation as a kind of shorthand notation. When this notation would not have been applied, then in the right-hand sides of (22.15) the following terms would have appeared:

$$f_x \left(\frac{\partial x}{\partial \alpha_j} \right) + f_y \left(\frac{\partial y}{\partial \alpha_j} \right) + f_z \left(\frac{\partial z}{\partial \alpha_j} \right)$$

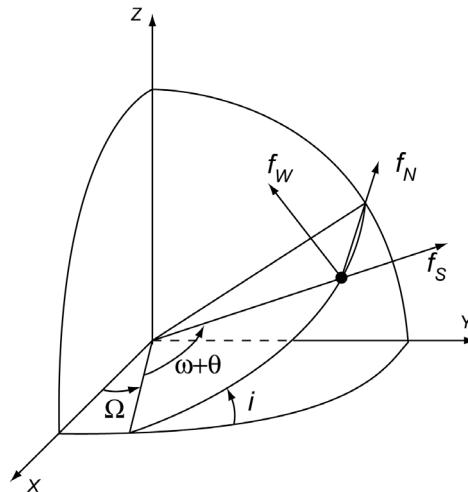


Figure 22.2: Geometry of the f_s, f_N, f_w acceleration frame relative to the orbit and to the inertial reference frame.

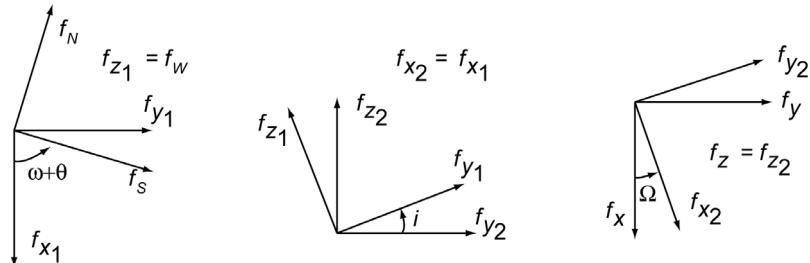


Figure 22.3: Geometry of the acceleration frames rotations.

where f_x, f_y, f_z are the components of the perturbing acceleration, regardless whether these can be derived or not from a potential function. The partial derivatives $\partial x/\partial \alpha_j, \partial y/\partial \alpha_j, \partial z/\partial \alpha_j$ can be elaborated further. J.C.F. Gauss (1777-1855) has done this around 1813 for the first time, when he computed the perturbations that Jupiter exerts on the orbit of the asteroid Pallas. He decomposed the three components of the perturbing acceleration into three other orthogonal directions: a radial component, f_s , a component in the orbital plane perpendicular to the radius vector, f_N , and a component perpendicular to the orbital plane, f_w . The component f_N points in the direction of motion; the component f_w in the direction of the orbital angular momentum vector. Figure 22.2 shows the geometry of these acceleration components in inertial space. To express f_x, f_y, f_z in terms of f_s, f_N, f_w , we can apply the method of rotation matrices. First, we rotate the frame f_s, f_N, f_w about the f_w -axis over an angle $-(\omega+\theta)$, then we apply a rotation about the line of nodes over an angle $-i$, and finally we apply a rotation about the Z-axis over an angle $-\Omega$. These rotations are indicated in Figure 22.3. We then can write for the relations between the coordinates:

$$\begin{pmatrix} f_{x_1} \\ f_{y_1} \\ f_{z_1} \end{pmatrix} = \begin{bmatrix} \cos(\omega+\theta) & -\sin(\omega+\theta) & 0 \\ \sin(\omega+\theta) & \cos(\omega+\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} f_s \\ f_N \\ f_w \end{pmatrix}$$

$$\begin{pmatrix} f_{x_2} \\ f_{y_2} \\ f_{z_2} \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos i & -\sin i \\ 0 & \sin i & \cos i \end{bmatrix} \begin{pmatrix} f_{x_1} \\ f_{y_1} \\ f_{z_1} \end{pmatrix}$$

$$\begin{pmatrix} f_x \\ f_y \\ f_z \end{pmatrix} = \begin{bmatrix} \cos \Omega & -\sin \Omega & 0 \\ \sin \Omega & \cos \Omega & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} f_{x_2} \\ f_{y_2} \\ f_{z_2} \end{pmatrix}$$

Multiplication of the rotation matrices leads to

$$\begin{aligned} f_x &= (\cos \Omega \cos(\omega + \theta) - \cos i \sin \Omega \sin(\omega + \theta)) f_S \\ &\quad - (\cos \Omega \sin(\omega + \theta) + \cos i \sin \Omega \cos(\omega + \theta)) f_N + (\sin i \sin \Omega) f_W \\ f_y &= (\sin \Omega \cos(\omega + \theta) + \cos i \cos \Omega \sin(\omega + \theta)) f_S \\ &\quad - (\sin \Omega \sin(\omega + \theta) - \cos i \cos \Omega \cos(\omega + \theta)) f_N - (\sin i \cos \Omega) f_W \\ f_z &= (\sin i \sin(\omega + \theta)) f_S + (\sin i \cos(\omega + \theta)) f_N + (\cos i) f_W \end{aligned}$$

or, using the relations (11.41),

$$\begin{aligned} f_x &= (l_1 \cos \theta + l_2 \sin \theta) f_S - (l_1 \sin \theta - l_2 \cos \theta) f_N + l_3 f_W \\ f_y &= (m_1 \cos \theta + m_2 \sin \theta) f_S - (m_1 \sin \theta - m_2 \cos \theta) f_N + m_3 f_W \\ f_z &= (n_1 \cos \theta + n_2 \sin \theta) f_S - (n_1 \sin \theta - n_2 \cos \theta) f_N + n_3 f_W \end{aligned} \quad (22.49)$$

where $l_1, l_2, l_3, m_1, m_2, m_3, n_1, n_2, n_3$ are functions of i, Ω, ω only ((11.41)).

As an example, we now derive an expression for $\partial \tilde{R} / \partial a$ in terms of the perturbing accelerations f_S, f_N, f_W . We start from (22.36) and write

$$\frac{\partial \tilde{R}}{\partial a} = f_x \frac{\partial x}{\partial a} + f_y \frac{\partial y}{\partial a} + f_z \frac{\partial z}{\partial a} \quad (22.50)$$

From (22.19-1) we find

$$\frac{\partial x}{\partial a} = (l_1 \cos \theta + l_2 \sin \theta) \frac{\partial r}{\partial a} \quad (22.51)$$

From the relation for a Keplerian orbit:

$$r = \frac{a(1-e^2)}{1+e \cos \theta}$$

we find

$$\frac{\partial r}{\partial a} = \frac{r}{a}$$

Substitution of this relation into (22.51) yields

$$\frac{\partial x}{\partial a} = \frac{r}{a} (l_1 \cos \theta + l_2 \sin \theta)$$

In the same way, we can derive

$$\frac{\partial y}{\partial a} = \frac{r}{a} (m_1 \cos \theta + m_2 \sin \theta)$$

$$\frac{\partial z}{\partial a} = \frac{r}{a} (n_1 \cos \theta + n_2 \sin \theta)$$

After substitution of these relations and (22.49) into (22.50), and using the relations (22.21), we obtain after some algebraic manipulation

$$\frac{\partial \tilde{R}}{\partial a} = \frac{r}{a} f_s$$

Similarly, we can derive expressions for the other partial derivatives of \tilde{R} . With these expressions, we obtain from (22.35) and (22.45) after some algebraic manipulation the so-called *Gauss' form of Lagrange's planetary equations*:

$$\begin{aligned} \frac{da}{dt} &= 2 \frac{a^2}{\sqrt{\mu p}} \left[f_s e \sin \theta + f_N \frac{p}{r} \right] \\ \frac{de}{dt} &= \sqrt{\frac{p}{\mu}} \left[f_s \sin \theta + f_N (\cos E + \cos \theta) \right] \\ \frac{di}{dt} &= f_w \frac{r}{\sqrt{\mu p}} \cos u \\ \frac{d\omega}{dt} &= - \sqrt{\frac{p}{\mu}} \left[f_w \frac{r}{p} \cot i \sin u + \frac{1}{e} \left\{ f_s \cos \theta - f_N \left(1 + \frac{r}{p} \right) \sin \theta \right\} \right] \\ \frac{d\Omega}{dt} &= f_w \frac{r}{\sqrt{\mu p} \sin i} \sin u \\ \frac{dM}{dt} &= n - f_s \left[\frac{2r}{\sqrt{\mu a}} - \frac{1-e^2}{e} \sqrt{\frac{a}{\mu}} \cos \theta \right] - f_N \frac{1-e^2}{e} \sqrt{\frac{a}{\mu}} \left(1 + \frac{r}{p} \right) \sin \theta \end{aligned} \quad (22.52)$$

These differential equations contain components of the perturbing accelerations and are therefore applicable to the computation of orbits that are perturbed by forces which cannot be expressed through a perturbing potential, such as atmospheric drag. They lend themselves very well to numerical integration and have been used extensively for computing by numerical integration perturbations in the elements of comets and asteroids.

The expressions (22.52), of course, suffer from the same singularities as Lagrange's form of the planetary equations in the classical orbital elements. For small values of e and/or i the Gauss' form of the planetary equations should therefore be written in terms of non-singular elements. As an example, we conclude that only (22.52-4) and (22.52-6) are singular for $e = 0$. Again, this

is the result of the fact that perigee is not defined for circular orbits and that therefore the parameters ω and M are then inappropriate for describing the satellite's position. When we use for $e \downarrow 0$ the relation $du/dt = d\omega/dt + d\theta/dt \approx d\omega/dt + dM/dt$, substitute (22.52-4) and (22.52-6), and neglect terms of order e , we find

$$\left(\frac{du}{dt} \right)_{e \downarrow 0} \approx \frac{d\omega}{dt} + \frac{dM}{dt} = n - \sqrt{\frac{a}{\mu}} (2f_S + f_W \cot i \sin u) \quad (22.52)$$

Clearly, the use of the argument of latitude, u , which describes the angular position of the satellite along its orbit relative to the ascending node, leads to a non-singular differential equation for $e \downarrow 0$, as could be expected.

Equation (22.52-1) may be written in another form. To that end, we first write (22.52-1) as

$$\frac{da}{dt} = 2 \frac{a^2}{\sqrt{\mu p}} [f_S e \sin \theta + f_N (1 + e \cos \theta)] \quad (22.53)$$

From (16.11) we obtain

$$e \sin \theta = \frac{HV \sin \gamma}{\mu} ; \quad 1 + e \cos \theta = \frac{HV \cos \gamma}{\mu}$$

where $H = \sqrt{\mu p}$ is the angular momentum (per unit of mass), V is the velocity and γ is the flight path angle. Substitution of these relations into (22.53) yields

$$\frac{da}{dt} = 2 \frac{a^2}{\mu} V [f_S \sin \gamma + f_N \cos \gamma]$$

Inspection of Figure 1.5 shows that the term in brackets is just the acceleration in the along-track direction (along the velocity vector), f_T , and thus we can write

$$\frac{da}{dt} = 2f_T \frac{a^2}{\mu} V \quad (22.52)$$

The set of first-order differential equations (22.52) can be used very well for a qualitative analysis of perturbed orbits. For instance, we conclude that the semi-major axis of an orbit can be changed most effectively by an impulse that is tangentially to the orbit, and that the impulse has to be applied at perigee of the orbit, where the velocity is maximum. This result was already found in Section 1.7 and Section 12.1 through different analyses. If we assume that the atmospheric force on a satellite is directed oppositely to the velocity vector of the satellite with respect to the atmosphere (atmospheric drag) and if we neglect atmospheric rotation, then (22.52) shows that all orbital elements except i and Ω will exhibit changes. Only the rotation of the atmosphere and the possible occurrence of out-of-plane lift forces may induce changes in i and Ω . When we again neglect atmospheric rotation and consider the orbital effects of atmospheric drag ($f_W = 0$) at perigee and apogee, where $f_S = 0$ and $f_N < 0$, we conclude from (22.53) and (22.52-2) that at perigee ($\theta = E = 0^\circ$): $da/dt < 0$ and $de/dt < 0$, while at apogee ($\theta = E = 180^\circ$): $da/dt < 0$ and $de/dt > 0$. So, the drag experienced at perigee will lead to a decrease of the semi-major axis and eccentricity, while the drag experienced at apogee will lead to a decrease of the semi-major axis, but to an increase of the eccentricity. Because the atmospheric density at apogee is (much) smaller than the density at perigee, the increase of the eccentricity at apogee will be (much) smaller than the decrease of the eccentricity at perigee, which leads to a net decrease of the eccentricity after an orbital revolution.

22.6. Approximate analytical solution of the planetary equations

Lagrange's planetary equations can be solved numerically to obtain the orbit perturbations due to a specific perturbing force. They are, however, particularly suited to generate approximate analytical expressions for the effects of perturbing forces on the orbit. For such a general perturbations approach (Section 20.6), we may use various analytical techniques. In the following, we will apply the *method of successive approximations*. This simple technique, which we have also used in Section 5.10 for the analysis of relativistic effects on Keplerian orbits, suffices to explain the essentials of the derivation of approximative analytical solutions of the planetary equations.

First of all, it is emphasized that the planetary equations for the orbital elements a, e, i, ω, Ω may be written in the following symbolical form:

$$\frac{d\alpha_i}{dt} = \epsilon f_i(\alpha_j(t), t) ; \quad \alpha_i = 1, \dots, 5 ; \quad \alpha_j = 1, \dots, \alpha_i, \dots, 6$$

where ϵ is a small factor that indicates the order of magnitude of the perturbing acceleration. Integration of this relation yields

$$\alpha_i(t) = \alpha_{i_0} + \epsilon \int f_i(\alpha_j(t), t) dt \quad (22.54)$$

where α_{i_0} is the value of α_i at $t = 0$. It is assumed that when the magnitude of a perturbing acceleration is small with respect to μ/r^2 , then also the effect of that perturbing acceleration over not too long a period of time will be small. Then, for a first-order solution of (22.54) it is allowed to substitute for the orbital elements in the integral their initial values:

$$\alpha_i^{(1)}(t) = \alpha_{i_0} + \epsilon \int f_i(\alpha_{j_0}, t) dt \quad (22.55)$$

For a given perturbing force this integral can be evaluated and we obtain an analytical expression for the first-order approximation of the variation of α_i with time. When such expressions for the first-order variation of the orbital elements are substituted into the integral in (22.54), we find to second-order approximation:

$$\alpha_i^{(2)}(t) = \alpha_{i_0} + \epsilon \int f_i(\alpha_j^{(1)}(t), t) dt \quad (22.56)$$

In a similar way, we may obtain higher-order analytical approximations. Obviously, it is assumed that the higher the order of the approximation the closer it approximates the true solution of the differential equations.

The differential equation for M , (22.45), introduces a problem. This relation may be written symbolically as

$$\frac{dM}{dt} = n(t) + \epsilon f_M(\alpha_j(t), t)$$

from which we find after integration:

$$M(t) = M_0 + \int n(t) dt + \epsilon \int f_M(\alpha_j(t), t) dt \quad (22.57)$$

The second term on the right-hand side of this equation does not contain a small multiplicative factor. For this reason, it is not justified to substitute the initial value of n into the integral in order to obtain a first-order solution for this term. Instead, we have to substitute the first-order variation of n and write

$$M^{(1)}(t) = M_0 + \int n^{(1)}(t) dt + \epsilon \int f_M(\alpha_{j_0}, t) dt \quad (22.58)$$

For the second term on the right-hand side of (22.57) we may write

$$\int n(t) dt = \int \sqrt{\frac{\mu}{\left(a_0 + \int \frac{da}{dt} dt\right)^3}} dt \quad (22.59)$$

To first-order approximation, we may write

$$\sqrt{\frac{\mu}{\left(a_0 + \int \frac{da}{dt} dt\right)^3}} = n_0 \left[1 - \frac{3}{2a_0} \int \frac{da}{dt} dt \right] \quad (22.60)$$

Substitution of (22.59) and (22.60) into (22.58) results in

$$M^{(1)}(t) = M_0 + n_0 \int dt - \frac{3n_0}{2a_0} \int \left\{ \int \frac{da}{dt} dt \right\} dt + \epsilon \int f_M(\alpha_{j_0}, t) dt \quad (22.61)$$

where the initial values of the elements α_j have to be substituted into the expression for da/dt . This method can also be used to obtain higher-order approximations of $M(t)$. Note that the price we have to pay for replacing the expression for $d\tau/dt$ by the expression for dM/dt is that we have to evaluate a double integral in (22.61).

A serious problem for the approximate analytical solution of Lagrange's planetary equations is that, in practice, a number of perturbing forces act simultaneously on the satellite. Generally, it is impossible to take these multiple perturbing forces into account simultaneously when evaluating the integral in (22.54) and (22.61) analytically. Therefore, one usually computes the effect of each perturbing force on the orbit over a certain time interval separately and these orbit perturbations are subsequently added. It is emphasized that this approach is theoretically not correct, since a perturbing force will lead to a change of the position and velocity of the satellite relative to their values for the (unperturbed) Keplerian orbit. These deviations will affect the computation of another perturbing force, because, in general, this force is a function of the instantaneous position and velocity of the satellite. However, it is clear that the effect of an orbit deviation on the magnitude of a computed perturbing force will be small of the second or higher order as long as the orbit deviations are small. But when one starts working with higher-order approximations of (22.54) and (22.61), this type of effects has to be taken into account, which makes the evaluation of the integral in (22.54) and (22.61) even more difficult.

22.7. Application of Gauss' equations to orbit maneuvers

The Gauss' form of Lagrange's planetary equations can very well be applied to the analysis of small orbit maneuvers. In this Section, we restrict ourselves to the case of small impulsive shot maneuvers (Section 1.7) executed by the satellite's onboard rocket propulsion system. We will consider only the effects of a maneuver on the orbital elements a, e, i, ω and Ω , which means that we will only consider changes of the shape and orientation of the orbit, and changes of the orbital plane.

The change of an orbital element due to a small impulsive shot can be found by integration

of (22.52), where the orbital parameters on the right-hand side of the equations may, in first-order approximation, be kept constant, and

$$\lim_{t_e \rightarrow 0} \left(\int_0^{t_e} f_i dt \right) = \Delta V_i$$

where the index i refers to any of the indices S, N, W in the equations. We then find from (22.52-1) to (22.52-5), after substituting a series of classical expressions for Keplerian orbits,

$$\begin{aligned} \Delta a &= 2 \sqrt{\frac{a^3}{\mu(1-e^2)}} [\Delta V_S e \sin \theta + \Delta V_N (1 + e \cos \theta)] \\ \Delta e &= \sqrt{\frac{a(1-e^2)}{\mu}} \left[\Delta V_S \sin \theta + \Delta V_N \left\{ \frac{2 \cos \theta + e(1+\cos^2 \theta)}{1+e \cos \theta} \right\} \right] \\ \Delta i &= \sqrt{\frac{a(1-e^2)}{\mu}} \Delta V_W \frac{\cos(\omega + \theta)}{1+e \cos \theta} \\ \Delta \omega &= \sqrt{\frac{a(1-e^2)}{\mu}} \left[-\Delta V_S \frac{\cos \theta}{e} + \Delta V_N \left\{ \frac{\sin \theta (2+e \cos \theta)}{e(1+e \cos \theta)} \right\} - \Delta V_W \frac{\cot i \sin(\omega + \theta)}{1+e \cos \theta} \right] \\ \Delta \Omega &= \sqrt{\frac{a(1-e^2)}{\mu}} \Delta V_W \frac{\sin(\omega + \theta)}{\sin i (1+e \cos \theta)} \end{aligned} \tag{22.62}$$

where $\Delta V_S, \Delta V_N$ and ΔV_W are small impulsive shots in the S, N and W directions. It is important to realize that (22.62) are linearized relations between small impulsive shots and the resulting small changes of the elements of the original orbit.

Equations (22.62) show that each of the impulsive shots $\Delta V_S, \Delta V_N, \Delta V_W$, in general, produces changes in three orbital elements simultaneously. From (22.62) we can determine the relation between an impulsive shot in each of the S, N and W directions and the resulting change of an orbital element, and thus the magnitude of the ΔV in that direction required for a prescribed change of the orbital element:

$$\begin{aligned} \Delta V_S &= \frac{1}{2} \frac{\Delta a}{a} \sqrt{\frac{\mu(1-e^2)}{a}} \frac{1}{e \sin \theta} \\ \Delta V_S &= \Delta e \sqrt{\frac{\mu}{a(1-e^2)}} \frac{1}{\sin \theta} \\ \Delta V_S &= -\Delta \omega \sqrt{\frac{\mu}{a(1-e^2)}} \frac{e}{\cos \theta} \end{aligned} \tag{22.63}$$

$$\begin{aligned}
 \Delta V_N &= \frac{1}{2} \frac{\Delta a}{a} \sqrt{\frac{\mu(1-e^2)}{a}} \frac{1}{1+e \cos \theta} \\
 \Delta V_N &= \Delta e \sqrt{\frac{\mu}{a(1-e^2)}} \frac{1+e \cos \theta}{2 \cos \theta + e(1+\cos^2 \theta)} \\
 \Delta V_N &= \Delta \omega \sqrt{\frac{\mu}{a(1-e^2)}} \frac{e(1+e \cos \theta)}{(2+e \cos \theta) \sin \theta} \\
 \Delta V_W &= \Delta i \sqrt{\frac{\mu}{a(1-e^2)}} \frac{1+e \cos \theta}{\cos(\omega+\theta)} \\
 \Delta V_W &= -\Delta \omega \sqrt{\frac{\mu}{a(1-e^2)}} \tan i \frac{1+e \cos \theta}{\sin(\omega+\theta)} \\
 \Delta V_W &= \Delta \Omega \sqrt{\frac{\mu}{a(1-e^2)}} \sin i \frac{1+e \cos \theta}{\sin(\omega+\theta)}
 \end{aligned} \tag{22.63}$$

Note that the right-hand side of (22.63-3) and (22.63-6) is zero if $e = 0$, and that the right-hand side of (22.63-8) and (22.63-9) is zero if $i = 0^\circ$. This is a consequence of the fact that ω is undefined for $e = 0$, and that ω and Ω are undefined for $i = 0^\circ$. As was shown in Section 22.4, this kind of problems can be prevented by selecting other, more appropriate, orbital elements to describe the orbit. In the following, we will assume that $e \neq 0$, $i \neq 0^\circ$, and that the linearized expressions (22.63) are accurate enough for small values of e and i .

We conclude from (22.63-1) and (22.63-2) that an impulsive shot in the S direction yields for any value of θ a simultaneous change in a and e , where

$$\Delta e = \frac{1-e^2}{2ae} \Delta a$$

and from (22.63-8) and (22.63-9) that an impulsive shot in the W direction yields for any value of θ a simultaneous change in ω and Ω , where

$$\Delta \omega = -\cos i \Delta \Omega$$

Note that the right-hand side of this equation expresses the change in ω due to the change in the line of nodes, which is the reference for measuring ω . So, it is impossible to apply an impulsive shot in the S direction to change only a or e , or an impulsive shot in the W direction to change only ω or Ω ; in both cases the two orbital elements will always change simultaneously.

The equations (22.63) show that, for given values of the orbital elements, the ΔV required to produce a prescribed change of an orbital element depends on the position along the orbit where the impulsive shot is executed. They also show that a prescribed change of the orbital elements a , e and ω can be effectuated by impulsive shots in different directions. Therefore, an important question is: “What is the most efficient way to produce a prescribed change of an orbital element and where along the orbit should the impulse be applied?” To answer this question, we will analyze each of the equations (22.63) separately in order to determine the minimum value of ΔV .

required to produce a specified change in an orbital element. In our analysis we will focus on low-eccentricity orbits with $e < 0.01$. We will assume that the impulsive shot is acting in one of the directions S , N or W and do not consider the possibility of an impulsive shot in any other direction.

Semi-major axis

A change of the semi-major axis can be realized by an impulsive shot in the S or N direction. For the optimum location and the minimum magnitude of these impulsive shots we find from (22.63-1) and (22.63-4)

$$\Delta V_S : \theta_{opt} = \frac{1}{2}\pi, \frac{3}{2}\pi ; \quad |\Delta V_S|_{min} = \frac{1}{2} \frac{|\Delta a|}{a} \frac{1}{e} \sqrt{\frac{\mu(1-e^2)}{a}}$$

$$\Delta V_N : \theta_{opt} = 0 ; \quad (\Delta V_N)_{min} = \frac{1}{2} \frac{\Delta a}{a} \sqrt{\frac{\mu(1-e)}{a(1+e)}}$$

From the first expression we conclude that when e is small, a specified value of Δa requires a very large ΔV_S . This means that for low-eccentricity orbits a small impulse ΔV_S hardly affects the orbit's semi-major axis. Comparing both expressions for ΔV , we find that for a specified Δa the following relation holds: $|\Delta V_S|_{min} = (1+e)|\Delta V_N|_{min}/e$. Since $(1+e)/e > 2$ for $e < 1$, the optimum maneuver ΔV_N is more efficient than the optimum maneuver ΔV_S , for any value of e . The optimum maneuver ΔV_N has to be executed at perigee, which means that it is oriented along the satellite's local velocity vector. Using (6.23) we then may write

$$(\Delta V_N)_{min} = \frac{1}{2} \frac{\mu}{V_p} \frac{\Delta a}{a^2}$$

where V_p is the velocity at perigee. These results fully agree with the results obtained in Section 1.7 and Section 22.5. For $e < 0.01$, we may write

$$(\Delta V_N)_{min} \approx \frac{1}{2} \frac{\Delta a}{a} \sqrt{\frac{\mu}{a}}$$

Eccentricity

A change of the eccentricity can be realized by an impulsive shot in the S or N direction. For the optimum location and minimum magnitude of these impulsive shots we find from (22.63-2) and (22.63-5)

$$\Delta V_S : \theta_{opt} = \frac{1}{2}\pi, \frac{3}{2}\pi ; \quad |\Delta V_S|_{min} = |\Delta e| \sqrt{\frac{\mu}{a(1-e^2)}}$$

$$\Delta V_N : \theta_{opt} = 0, \pi ; \quad |\Delta V_N|_{min} = \frac{1}{2} |\Delta e| \sqrt{\frac{\mu}{a(1-e^2)}}$$

We conclude that the optimum maneuver ΔV_N is twice as efficient as the optimum maneuver ΔV_S , and that the maneuver ΔV_N has to be executed at perigee or apogee. For $e < 0.1$, we may write

$$|\Delta V_N|_{\min} \approx \frac{1}{2} |\Delta e| \sqrt{\frac{\mu}{a}}$$

Inclination

A change of the inclination can only be realized by an impulsive shot in the W direction. From (22.63-7) we find that the location where $|\Delta V_W|$ reaches a (local) minimum value is given by

$$\theta = -\omega - \arcsin(e \sin \omega)$$

This expression yields two solutions for θ , which depend on the eccentricity and argument of perigee of the orbit. Substitution of these values into (22.63-7) yields two values of $|\Delta V_W|$. The smaller one corresponds to the absolute minimum $|\Delta V_W|$, and the associated value of θ corresponds to the optimum location. For $e < 0.01$, the two values of $|\Delta V_W|$ are nearly identical and we find

$$\theta_{opt} \approx -\omega, \pi - \omega ; \quad |\Delta V_W|_{\min} \approx |\Delta i| \sqrt{\frac{\mu}{a}}$$

So, the maneuver has to be executed at (near) the ascending ($u = \omega + \theta = 0^\circ$) or the descending ($u = 180^\circ$) node. If we could freely choose the value of ω , (22.63-7) shows that, for any value of e , the optimum location of the impulsive shot and the absolute minimum value of $|\Delta V_W|$ are

$$\omega_{opt} = 0, \pi ; \quad \theta_{opt} = \pi ; \quad |\Delta V_W|_{\min} = |\Delta i| \sqrt{\frac{\mu(1-e)}{a(1+e)}}$$

So, the maneuver ΔV_W has to be executed at the ascending or descending node, and at apogee. In that case, we may write with (6.22)

$$|\Delta V_W|_{\min} = |\Delta i| V_a$$

where V_a is the satellite's velocity at apogee. These results fully agree with the results obtained in Section 13.3.

Argument of perigee

The argument of perigee can be changed by applying an impulsive shot in the S, N or W direction. For the optimum ΔV_S maneuver, we find from (22.63-3)

$$\theta_{opt} = 0, \pi ; \quad |\Delta V_S|_{\min} = |\Delta \omega| \sqrt{\frac{\mu}{a(1-e^2)}} e$$

So, the maneuver has to be executed at perigee or apogee. For a particular $\Delta \omega$, $|\Delta V_S|_{\min}$ decreases with decreasing values of e .

For the ΔV_N maneuver, we find from (22.63-6) for the location where $|\Delta V_N|$ reaches a (local) minimum value the following condition:

$$e^2 \cos^3 \theta + 2e \cos^2 \theta + 2 \cos \theta + e = 0$$

A numerical analysis shows that this expression has one negative (real) root for $\cos \theta$, and thus two roots for θ , for arbitrary values of e . For $e < 0.3$, these roots lie within the ranges $90^\circ < \theta < 99^\circ$ and $261^\circ < \theta < 270^\circ$; for smaller values of e the roots are closer to 90° and 270° . Substitu-

tion of these values into (22.63-6) yields two equal values of $|\Delta V_N|$. For $e < 0.01$ we find

$$\theta_{opt} \approx \frac{1}{2}\pi, \frac{3}{2}\pi ; \quad |\Delta V_N|_{min} \approx \frac{1}{2}|\Delta\omega| \sqrt{\frac{\mu}{a}} e$$

For a particular $\Delta\omega$, $|\Delta V_N|_{min}$ decreases with decreasing values of e .

For the ΔV_W maneuver, we find from (22.63-8) for the location where $|\Delta V_W|$ reaches a (local) minimum value the following condition:

$$\theta = -\omega + \arccos(-e \cos \omega)$$

This expression yields two solutions for θ , which depend on the eccentricity and argument of perigee of the orbit. Substitution of these values into (22.63-8) yields two values of $|\Delta V_W|$. The smaller one corresponds to the absolute minimum $|\Delta V_W|$ and the associated value of θ corresponds to the optimum location. For $e < 0.01$ the two values of $|\Delta V_W|$ are nearly identical and we find

$$\theta_{opt} \approx \frac{1}{2}\pi - \omega, \frac{3}{2}\pi - \omega ; \quad |\Delta V_W|_{min} \approx |\Delta\omega| \sqrt{\frac{\mu}{a}} |\tan i|$$

So, the maneuver has to be executed at (near) $u = 90^\circ, 270^\circ$. For a particular $\Delta\omega$, $|\Delta V_W|_{min}$ increases for decreasing values of $|90^\circ - i|$.

When we compare the results for $e < 0.01$, we find that an optimum ΔV_N maneuver is twice as efficient as an optimum ΔV_S maneuver, and that, unless the inclination is very small ($|\tan i| < \frac{1}{2}e$), an optimum ΔV_N maneuver is more efficient than an optimum ΔV_W maneuver. If we could freely choose the value of ω , (22.63-8) shows that, for any value of e , the optimum location of an impulsive shot ΔV_W and the absolute minimum value of $|\Delta V_W|$ are

$$\omega_{opt} = \frac{1}{2}\pi, \frac{3}{2}\pi ; \quad \theta_{opt} = \pi ; \quad |\Delta V_W|_{min} = |\Delta\omega| \sqrt{\frac{\mu(1-e)}{a(1+e)}} |\tan i|$$

or, with (6.22),

$$|\Delta V_W|_{min} = |\Delta\omega \tan i| V_a$$

where V_a is the satellite's velocity at apogee.

Right ascension of the ascending node

A change of the right ascension of the ascending node can only be realized by a maneuver ΔV_W . For the location where $|\Delta V_W|$ reaches a (local) minimum value, we find from (22.63-9)

$$\theta = -\omega + \arccos(-e \cos \omega)$$

This expression yields two solutions for θ , which depend on the eccentricity and argument of perigee of the orbit. Substitution of these values into (22.63-9) yields two values of $|\Delta V_W|$. The smaller one corresponds to the absolute minimum $|\Delta V_W|$, and the associated value of θ corresponds to the optimum location. For $e < 0.01$ the two values of $|\Delta V_W|$ are nearly identical and we find

$$\theta_{opt} \approx \frac{1}{2}\pi - \omega, \frac{3}{2}\pi - \omega ; \quad |\Delta V_W|_{min} \approx |\Delta\Omega| \sqrt{\frac{\mu}{a}} \sin i$$

So, the maneuver has to be executed at (near) $u = 90^\circ, 270^\circ$. For a particular $\Delta\Omega$, $|\Delta V_W|_{min}$ increases for decreasing values of $|90^\circ - i|$. If we could freely choose the value of ω , (22.63-9) shows that, for any value of e , the optimum location of an impulsive shot ΔV_W and the absolute minimum value of $|\Delta V_W|$ are

$$\omega_{opt} = \frac{1}{2}\pi, \frac{3}{2}\pi ; \quad \theta_{opt} = \pi ; \quad |\Delta V_W|_{min} = |\Delta\Omega| \sqrt{\frac{\mu(1-e)}{a(1+e)}} \sin i$$

So, the maneuver ΔV_W has to be executed at $u = 90^\circ, 270^\circ$ and at apogee. In that case, we may write with (6.22)

$$|\Delta V_W|_{min} = |\Delta\Omega| \sin i V_a$$

These results fully agree with the results obtained in Section 13.2.

From the results given above, we conclude that a ΔV_S maneuver is always less efficient than a ΔV_N maneuver. This is the reason why radial impulsive shots are hardly applied for satellite maneuvers. We also conclude that the magnitude of an impulsive shot executed to change a particular orbital element by a specified amount decreases for increasing values of the semi-major axis. As an example, Table 22.1 lists the direction and magnitude of the optimum impulsive shot to produce a specified change of an orbital element, and the (approximate) location along the orbit where this impulsive shot has to be applied. For this example, it has been assumed that the satellite's initial orbital elements are: $a = 7000$ km, $e = 0.005$, $i = 80^\circ$.

Table 22.1: Location, direction and magnitude of the optimum impulsive shot to produce a specified change of an orbital element. Initial orbit: $a = 7000$ km, $e = 0.005$, $i = 80^\circ$.

Orbit change	Optimum impulsive shot		
	location	direction	ΔV (m/s)
$\Delta a = 2$ km	$\theta = 0$	N	1.07
$\Delta e = 0.001$	$\theta = 0, \pi$	N	3.78
$\Delta i = 0.1^\circ$	$u = 0, \pi$	W	13.1
$\Delta\omega = 10^\circ$	$\theta = \pi/2, 3\pi/2$	N	3.29
$\Delta\Omega = 0.1^\circ$	$u = \pi/2, 3\pi/2$	W	12.9

We have analyzed the case of optimum maneuvers by impulsive shots in the S , N or W direction. In reality, these cannot always be applied. For instance, we already have concluded that an impulsive shot of given direction and magnitude, generally, will produce specific changes of three orbital elements simultaneously. In general, this is an undesirable situation; instead we often wish that an impulsive shot in a particular direction produces prescribed changes in two selected orbital elements simultaneously, using an acceptably small amount of propellant. Below, we will analyze a few elementary aspects of this kind of maneuvers.

First, we will address the question: "Is it possible to execute an impulsive shot ΔV_S , ΔV_N or ΔV_W such that (instead of three orbital elements) only one or two orbital elements will change?" To answer this question, we may determine the value of θ for which the denominator of the right-

hand side of each of the equations (22.63) is zero. In that case, a prescribed change of an orbital element would require an infinitely large ΔV , which means that the particular orbital element is practically insensitive to a small ΔV . This implies that if a maneuver is executed at such a point along the orbit, the impulsive shot only affects one or two other orbital element(s). Of course, such a maneuver will be non-optimal for a particular orbital element in terms of fuel consumption. Inspection of (22.63) shows that a particular orbital element is not affected by an impulsive shot in the S , N or W direction, if that shot is applied at the true anomaly listed in Table 22.2.

Table 22.2: Positions along the orbit where an impulsive shot in the S , N or W direction does not affect a particular orbital element. The position is specified by its true anomaly.

Orbit change	Position of the impulsive shot		
	ΔV_S	ΔV_N	ΔV_W
$\Delta a = 0$	$0, \pi$	never	---
$\Delta e = 0$	$0, \pi$	$\arccos((-1 + \sqrt{1 - e^2})/e)$	---
$\Delta i = 0$	---	---	$\pi/2 - \omega, 3\pi/2 - \omega$
$\Delta \omega = 0$	$\pi/2, 3\pi/2$	$0, \pi$	$-\omega, \pi - \omega$
$\Delta \Omega = 0$	---	---	$-\omega, \pi - \omega$

Note that for $e < 0.3$, the expression $\arccos((-1 + \sqrt{1 - e^2})/e)$ can be approximated by $\arccos(-e/2)$ and for $e < 0.01$ by $\pi/2, 3\pi/2$. From this Table we conclude that if ΔV_S is applied at $\theta = 0^\circ, 180^\circ$, the elements a and e will not change, but the element ω will change; if ΔV_S is applied at $\theta = 90^\circ, 270^\circ$, then ω will not change, but a and e will change. It is not possible to apply ΔV_S such that a will change but e not, and vice versa. This was already found before. We also conclude that ΔV_N will always lead to a change of a , whatever the value of θ , and that it is possible to apply ΔV_N such that either e or ω will change. For the ΔV_W impulse we conclude that it is possible to apply that impulse such that either i or ω and Ω will change. It is not possible to apply the impulse such that ω will change but not Ω , and vice versa. This was also already found before.

Another question is: “Is it possible to execute an impulsive shot ΔV_S , ΔV_N or ΔV_W such that two or three orbital elements will change simultaneously, each by a prescribed amount?” Of course, such a maneuver will be non-optimal for a particular orbital element. To answer this question, we set the three expressions for ΔV_S , ΔV_N or ΔV_W in (22.63) equal to each other and analyze the results obtained. We then find:

- It is impossible to change three orbital elements by a specified amount simultaneously through a single impulsive shot in the S , N or W direction.
- An impulsive shot in the S direction yields simultaneous changes of prescribed magnitude in a and ω , or e and ω , if the impulse is applied at a point along the orbit with the true anomaly

$$a \text{ and } \omega: \theta = -\arctan\left(\frac{1 - e^2}{2ae^2} \frac{\Delta a}{\Delta \omega}\right) ; \quad e \text{ and } \omega: \theta = -\arctan\left(\frac{1}{e} \frac{\Delta e}{\Delta \omega}\right)$$

These expressions yield a solution of θ for any value of Δa , Δe and $\Delta \omega$. A ΔV_S maneuver cannot produce simultaneous changes of prescribed magnitude in a and e . Instead, as already

- stated before, the simultaneous changes Δa and Δe have a constant ratio for any value of θ .
- An impulsive shot in the W direction yields simultaneous changes of prescribed magnitude in i and ω , or i and Ω , if the impulse is applied at a point with the true anomaly

$$i \text{ and } \omega: \theta = -\arctan\left(\tan i \frac{\Delta\omega}{\Delta i}\right) - \omega ; \quad i \text{ and } \Omega: \theta = \arctan\left(\sin i \frac{\Delta\Omega}{\Delta i}\right) - \omega$$

These expressions yield a solution of θ for any value of Δi , $\Delta\omega$ and $\Delta\Omega$. A ΔV_w maneuver cannot produce simultaneous changes of prescribed magnitude in ω and Ω . Instead, as already stated before, the simultaneous changes $\Delta\omega$ and $\Delta\Omega$ have a constant ratio for any value of θ .

- An impulsive shot in the N direction yields simultaneous changes of prescribed magnitude in e and ω , if the impulse is applied at a point with the true anomaly

$$\theta = \arctan\left(e \frac{\Delta\omega}{\Delta e}\right)$$

This expression yields a solution of θ for any value of Δe and $\Delta\omega$. The analytical analysis of a ΔV_N maneuver that yields simultaneous changes of prescribed magnitude in a and e , or a and ω , is quite complicated. A numerical analysis of the relevant expressions shows that a solution of θ only exists for certain ranges of the values of a , e , Δa , Δe and $\Delta\omega$. Outside these ranges, it is impossible to produce changes of prescribed magnitude in a and e , or a and ω , simultaneously. As an example, for an orbit with $a = 7000$ km, $e = 0.005$, we find that solutions of θ only exist if $\Delta a > 4$ km, $\Delta e < 0.0005$, $\Delta\omega < 5^\circ$.

In all cases, the corresponding magnitude of the required impulsive shot can be found by substituting the particular value of θ in the relevant expression of (22.63).

23. ORBIT PERTURBATIONS DUE TO THE EARTH'S GRAVITY FIELD

As an application of Lagrange's planetary equations, which have been derived and discussed in Chapter 22, in this Chapter we will derive approximative analytical solutions for the orbit perturbations due to the non-radially-symmetric mass distribution of the Earth and we will discuss this class of orbit perturbations.

In Section 20.1 we have found that to first-order approximation the (static) gravitational potential of the Earth (geopotential) can be written as

$$U = -\frac{\mu}{r} \left[1 - J_2 \left(\frac{R}{r} \right)^2 P_2(\sin \phi) \right] = -\frac{\mu}{r} \left[1 - \frac{1}{2} J_2 \left(\frac{R}{r} \right)^2 (3 \sin^2 \phi - 1) \right] \quad (23.1)$$

where r is the distance and ϕ is the geocentric latitude of the satellite, μ is the Earth's gravitational parameter, R is the mean equatorial Earth radius, and J_2 is a coefficient of the series expansion for the Earth's gravity field (Table 20.1). When we compute the orbit of a satellite relative to a geocentric equatorial reference frame, then $\delta = \phi$, where δ is the declination of the satellite. In Section 11.9, we have found

$$\sin \delta = \sin i \sin(\omega + \theta) = \sin i \sin u$$

where i is the inclination, ω the argument of perigee, θ the true anomaly, and u the argument of latitude. So, we may write for the perturbing force function, \tilde{R} , (Section 22.1) that describes the first-order perturbing term in the expression for the gravity field

$$\tilde{R} = \frac{1}{2} J_2 \mu R^2 \frac{1 - 3 \sin^2 i \sin^2 u}{r^3} \quad (23.2)$$

The equations (23.1) and (23.2) describe an *axially-symmetric gravity field* that accounts for the first-order J_2 -term. Now, we will first derive two conservation laws that hold for any axially-symmetric gravity field.

23.1. Application of conservation laws

For an axially-symmetric gravity field, \tilde{R} is a function of r and ϕ ($= \delta$) but not of the geographic longitude, Λ . So, at a fixed position in space, relative to a non-rotating geocentric equatorial reference frame, the gravitational potential is constant although the Earth rotates. We may therefore conclude that the potential of this gravity field is not *explicitly* dependent on time. Since such a force field is conservative (Section 1.4), the sum of kinetic and potential energy (per unit of mass) of the satellite is constant:

$$\mathcal{E} = \frac{1}{2} V^2 - \frac{\mu}{r} - \tilde{R} = \text{constant} \quad (23.3)$$

or, with (6.21),

$$\mathcal{E} = -\frac{\mu}{2a} - \tilde{R} = \text{constant} \quad (23.4)$$

We will apply this result for the gravity field described by (23.1).

For a near-circular orbit, the energy integral can be evaluated in a simple manner. When we use the index 0 to denote the (osculating) orbital parameters at the moment that the satellite passes the ascending node ($u = 0^\circ$), we obtain from (23.2) and (23.4)

$$\frac{\mu}{2a} + \frac{1}{2}J_2\mu \frac{R^2}{r^3}(1 - 3\sin^2 i \sin^2 u) = \frac{\mu}{2a_0} + \frac{1}{2}J_2\mu \frac{R^2}{r_0^3}$$

where a , r and i denote the (osculating) semi-major axis, distance and inclination at the argument of latitude u . Realizing that $J_2 \approx 10^{-3}$ and writing $a = a_0 + \Delta a$, $r = r_0 + \Delta r$, $i = i_0 + \Delta i$, where $\Delta a \ll a_0$, $\Delta r \ll r_0$, $\Delta i \ll i_0$, linearization of this equation gives

$$\frac{\Delta a}{a_0} = -3J_2 \left(\frac{R}{r_0} \right)^2 \sin^2 i_0 \sin^2 u \quad (23.5)$$

This equation is in agreement with (21.20) derived in Section 21.2, and shows that Δa is always negative or zero. The maximum value of a occurs at the nodal crossing points and the minimum value occurs when the satellite reaches its maximum or minimum declination.

From (23.2) and (23.3) we obtain

$$V^2 - \frac{2\mu}{r} - J_2\mu \frac{R^2}{r^3}(1 - 3\sin^2 i \sin^2 u) = V_0^2 - \frac{2\mu}{r_0} - J_2\mu \frac{R^2}{r_0^3}$$

where V denotes the velocity of the satellite at the argument of latitude u , and V_0 at the passage through the ascending node. With $V = V_0 + \Delta V$, $\Delta V \ll V_0$, and the notation given above, linearization of this equation leads to

$$\frac{\Delta V}{V_0} = \frac{3}{2}J_2 \frac{\mu}{V_0^2 r_0} \left(\frac{R}{r_0} \right)^2 \sin^2 i_0 \sin^2 u$$

For a near-circular orbit $\mu/V_0^2 r_0 \approx 1$, and the expression simplifies to

$$\frac{\Delta V}{V_0} = \frac{3}{2}J_2 \left(\frac{R}{r_0} \right)^2 \sin^2 i_0 \sin^2 u \quad (23.6)$$

Note that ΔV is always positive or zero. The minimum value of V occurs at the nodal crossing points and the maximum value occurs when the satellite reaches its maximum or minimum declination. Equations (23.5) and (23.6) show that for near-circular orbits the semi-major axis and the velocity of the satellite exhibit a variation with a period equal to half the orbital period. The maximum values of $|\Delta a|$ and ΔV increase with increasing orbital inclination. When, as an example, we consider a polar circular orbit at an altitude of 500 km, we find for the maximum value of $|\Delta a|$ and ΔV over an orbital revolution: $|\Delta a| = 19.2$ km, $\Delta V = 10.6$ m/s. The value of $|\Delta a|$ agrees with the result obtained in Section 21.2.

A second integral of motion can be found as follows. According to (22.37), the component of the angular momentum about the Z -axis is given by

$$H_z = \sqrt{\mu a(1 - e^2)} \cos i$$

From this relation we find

$$\frac{d}{dt} \left(\frac{H_z^2}{\mu} \right) = (1 - e^2) \cos^2 i \frac{da}{dt} - 2ae \cos^2 i \frac{de}{dt} - a(1 - e^2) \sin 2i \frac{di}{dt}$$

Substitution of (22.35-1) to (22.35-3) into this equation gives

$$\frac{d}{dt} \left(\frac{H_z^2}{\mu} \right) = 2 \sqrt{\frac{p}{\mu}} \cos i \frac{\partial \tilde{R}}{\partial \Omega} \quad (23.7)$$

For any axially-symmetric gravity field, \tilde{R} is not a function of Ω . So, $\partial \tilde{R} / \partial \Omega = 0$ and we find from (23.7)

$$H_z = \text{constant} \quad (23.8)$$

This is a generalization of the result found in Section 21.2. The other two components of the angular momentum vector, in general, do vary with time for an axially-symmetric gravity field.

Also this integral can be evaluated in a simple manner for near-circular orbits. From (23.8) we find to first-order approximation

$$a \cos^2 i = a_0 \cos^2 i_0$$

Linearization leads to

$$\Delta i = \frac{1}{2} \frac{\cos i_0}{\sin i_0} \frac{\Delta a}{a_0}$$

Substitution of (23.5) into this equation yields

$$\Delta i = -\frac{3}{4} J_2 \left(\frac{R}{r_0} \right)^2 \sin 2i_0 \sin^2 u \quad (23.9)$$

This equation, which is identical to (21.24-1), shows that also the inclination exhibits a variation with a period of half the orbital period. The maximum value of $|\Delta i|$ occurs at $u = 90^\circ, 270^\circ$ and $i = 45^\circ, 135^\circ$. In Section 21.2, we have found that the maximum value of $|\Delta i|$ is $2.4'$ for an orbit with an altitude of 500 km and inclination of 45° . Although this value is rather small, it certainly is not negligible.

23.2. Characteristics of the variation of the orbital elements

In the previous Section, we have analyzed some general aspects of the effects of the J_2 -term in the expression for the geopotential on a near-circular satellite orbit. We have found that the orbital elements a and i show variations with a period of half the orbital period; these are the so-called *short-period variations* introduced in Section 21.2. The J_2 -term of the gravity field also leads to *secular* and *long-period variations* of the orbital elements. To demonstrate that, we will analyze in this Section the behavior of the orbital elements Ω and ω during an orbital revolution, starting from (22.35-4) and (22.35-5), and applying the *method of successive approximations* (Section 22.6).

We start with the analysis of the effect of the J_2 -term on the orbital element Ω . We then need, according to (22.35-5), an expression for the partial derivative of \tilde{R} to i . To obtain that expression, we rewrite (23.2) in the form

$$\tilde{R} = \frac{1}{2} J_2 \frac{\mu R^2}{a^3} \left(\frac{a}{r} \right)^3 (1 - 3 \sin^2 i \sin^2 u) \quad (23.10)$$

As r and u are not a function of i , we immediately find from (23.10):

$$\frac{\partial \tilde{R}}{\partial i} = -3J_2 \frac{\mu R^2}{a^3} \left(\frac{a}{r} \right)^3 \sin i \cos i \sin^2 u \quad (23.11)$$

Substitution of this relation into (22.35-5) and using

$$n = \sqrt{\frac{\mu}{a^3}} \quad (23.12)$$

where n is the mean angular (orbital) motion, yields

$$\frac{d\Omega}{dt} = -3J_2 \frac{\mu R^2}{n a^5} \left(\frac{a}{r} \right)^3 \frac{\cos i \sin^2 u}{\sqrt{1 - e^2}} \quad (23.13)$$

To find a first-order solution of this differential equation, we apply the method described in Section 22.6 and consider all orbital elements on the right-hand side of (23.13) as constants. Note that what makes this a first-order approximation is not the appearance of the first-order geopotential term J_2 , but rather the holding of the orbital elements as constants; i.e. at their initial values. For the analysis, it is convenient to change from t to the argument of latitude, u , as independent variable. Therefore, we write

$$\frac{d\Omega}{dt} = \frac{d\Omega}{du} \frac{du}{dt}$$

and

$$\frac{du}{dt} = \frac{d(\omega + \theta)}{dt} \approx \frac{d\theta}{dt} = \frac{H}{r^2} = \frac{\sqrt{\mu a (1 - e^2)}}{r^2}$$

where θ is the satellite's true anomaly. It is emphasized that, because we consider ω as a constant, the parameter u denotes the *unperturbed argument of latitude*, corresponding to pure Keplerian motion. With these relations and (23.13), we can write (23.13) as

$$\frac{d\Omega}{du} = -3J_2 \left(\frac{R}{a} \right)^2 \frac{a}{r} \frac{\cos i \sin^2 u}{1 - e^2}$$

Substitution of

$$r = \frac{a(1 - e^2)}{1 + e \cos \theta} ; \quad p = a(1 - e^2) ; \quad \theta = u - \omega$$

yields

$$\frac{d\Omega}{du} = -3J_2 \left(\frac{R}{p} \right)^2 \cos i (\sin^2 u + e \cos \omega \sin^2 u \cos u + e \sin \omega \sin^3 u) \quad (23.14)$$

To compute the change of Ω from $u = 0$ to $u = u_e$, we have to integrate the three terms in brackets in (23.14). We then find

$$\int_0^{u_e} \sin^2 u \, du = \left(\frac{1}{2}u - \frac{1}{4}\sin 2u \right) \Big|_0^{u_e} = \frac{1}{2}u_e - \frac{1}{4}\sin 2u_e$$

$$\int_0^{u_e} \sin^2 u \cos u \, du = \left(\frac{1}{3} \sin^3 u \right) \Big|_0^{u_e} = \frac{1}{3} \sin^3 u_e$$

$$\int_0^{u_e} \sin^3 u \, du = \left(\frac{1}{3} \cos^3 u - \cos u \right) \Big|_0^{u_e} = \frac{1}{3} \cos^3 u_e - \cos u_e + \frac{2}{3}$$

Integration of (23.14), substitution of the results listed above, and some trigonometric manipulation leads to

$$\begin{aligned} \Delta\Omega = & -\frac{3}{2} J_2 \left(\frac{R}{p} \right)^2 \cos i \left[u_e - \frac{1}{2} \sin 2u_e + \right. \\ & \left. + \frac{1}{6} e \{ 8 \sin \omega - 3 \sin(\omega + u_e) - 6 \sin(\omega - u_e) + \sin(\omega - 3u_e) \} \right] \end{aligned} \quad (23.15)$$

where the orbital elements on the right-hand side are considered as constants, and Δ denotes the difference relative to the value of Ω at $u = 0^\circ$. It is emphasized that, because u is the unperturbed argument of latitude, $u = 2\pi$ will, in general, not correspond to an ascending node crossing of the satellite in its perturbed orbit.

For the analysis of the effect of the J_2 -term on the orbital element ω we need, according to (22.35-4), in addition to the expression for the partial derivative of $\tilde{\mathbf{R}}$ to i , also an expression for the partial derivative of $\tilde{\mathbf{R}}$ to e . This expression can be obtained from (23-10) and we find

$$\frac{\partial \tilde{\mathbf{R}}}{\partial e} = \frac{3}{2} J_2 \frac{\mu R^2}{a^3} \left(\frac{a}{r} \right)^2 (1 - 3 \sin^2 i \sin^2 u) \left[\frac{2e + (1 + e^2) \cos(u - \omega)}{(1 - e^2)^2} \right] \quad (23.16)$$

Substitution of (23.11) and (23.16) into (22.35-4) and using the same methodology as described above for the derivation of (23.14), we find

$$\frac{d\omega}{du} = \frac{3}{2} J_2 \left(\frac{R}{p} \right)^2 \left[\left\{ 2 + \frac{1+e^2}{e} \cos(u - \omega) \right\} (1 - 3 \sin^2 i \sin^2 u) + 2 \cos^2 i (1 + e \cos(u - \omega)) \sin^2 u \right] \quad (23.17)$$

which, after integration from $u = 0$ to $u = u_e$ and considerable analytical manipulation, yields

$$\begin{aligned} \Delta\omega = & \frac{3}{2} J_2 \left(\frac{R}{p} \right)^2 \left[(3 - 4 \sin^2 i) u_e - \frac{1}{2} (1 - 4 \sin^2 i) \sin 2u_e + \right. \\ & + e \left\{ \sin \omega - \sin(\omega - u_e) + \frac{1}{12} (2 - 5 \sin^2 i) (8 \sin \omega - 3 \sin(\omega + u_e) - 6 \sin(\omega - u_e) + \sin(\omega - 3u_e)) \right\} + \\ & \left. + \frac{1}{e} \left\{ \sin \omega - \sin(\omega - u_e) - \frac{1}{4} \sin^2 i (8 \sin \omega - 3 \sin(\omega + u_e) - 6 \sin(\omega - u_e) + \sin(\omega - 3u_e)) \right\} \right] \end{aligned} \quad (23.18)$$

where Δ denotes the difference relative to the value of ω at $u = 0^\circ$. Equation (23.15), which is identical to (21.24-2) for $e = 0$, shows that $\Delta\Omega$ exhibits various types of variations. First, there is a linear term in the argument of latitude. This kind of terms correspond to ever-increasing or ever-decreasing changes from the epoch values, and are called *secular variations* (Section 21.2). Secondly, there are *periodic variations* expressed by trigonometric functions of the argument of

latitude. These terms have periods equal to $1/k$ times the orbital period, where k is an integer value, and express the so-called *short-period variations*. Equation (23.18) shows that $\Delta\omega$ also exhibits a secular variation. This linearly changing value of ω will result in a *long-period variation* of Ω . Because this long-period variation is induced by the secular variation of ω that is produced by the J_2 -term of the geopotential, it may be considered as a ‘second-order’ effect of the J_2 -term. Since J_2 is of order 10^{-3} and subsequent zonal harmonics coefficients are of order 10^{-6} or smaller (Table 20.1), a ‘second-order’ effect of the J_2 -term (Section 23.3) may have the same magnitude as ‘first-order’ effects of the J_3 -term, J_4 -term, etc. This implies that when analyzing long-period perturbations the effects of additional zonal harmonics have to be included in the analysis. The period of the long-period variations is $1/m$ times the time it takes perigee to complete one rotation about the Earth, where m is an integer value. In Section 23.3, an expression for the secular variation of the argument of perigee will be derived ((23.35-3)). From that expression, it may be concluded that for near-circular orbits at altitudes between 500 km and 1000 km, the period of the long-period variations is about 25 days for low-inclination orbits and about 100 days for polar orbits; for $i \approx 63^\circ, 117^\circ$ this period becomes very long. From (23.18) we conclude that ω also experiences short- and long-period variations. Note that for this orbital element a series of short- and long-period terms are multiplied by the term $1/e$. This means that for low-eccentricity orbits ($e \downarrow 0$) these short- and long-period terms will have large amplitudes. This phenomenon is a direct result of the fact that for circular orbits ($e = 0$) the argument of perigee is not defined. If for low-eccentricity orbits an appropriate set of other orbital elements (Section 22.4) is used, then the amplitudes of the short- and long-period terms will remain small.

The separation of orbit perturbations into secular and various periodic terms may be applied for all orbital elements, although a particular element does not necessarily exhibits all types of variations. To visualize the overall variation of an orbital element, Figure 23.1 may serve. The secular change of a generic orbital element, α , is depicted by the straight line. It is emphasized,

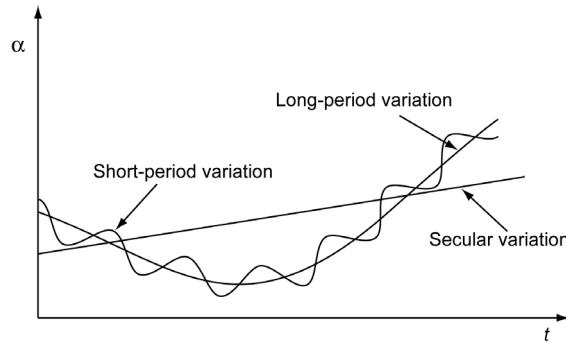


Figure 23.1: Secular, long-period and short-period variations of an orbital element.

however, that the term ‘secular’ by no means is limited to a linear change. In general, secular variations are associated with a steady, non-oscillatory type, continuous increasing deviation of an orbital element from its epoch value. The long-period variations are bounded fluctuations about the secular-variation curve. As already mentioned above, these long-period variations are caused by the secular change of ω . The short-period variations, which are superimposed on the long-period variations, are caused by the secular change of θ (or M). Various analytical orbit perturbation theories are available which yield expressions for the secular, long-period and short-period variations of the orbital elements due to specific terms of the gravitational potential. For operational orbit analyses the variations of the orbital elements are usually written as

$$\begin{aligned}\alpha_s(t) &= \alpha_s(t_0) + \dot{\alpha}_s(t - t_0) + \ddot{\alpha}_s(t - t_0)^2 + \dots \\ \Delta\alpha_l(t) &= \mathcal{F}_{\alpha}^1(\alpha_s(t)) \\ \Delta\alpha_{sh}(t) &= \mathcal{F}_{\alpha}^2(\alpha_s(t) + \Delta\alpha_l(t))\end{aligned}\tag{23.19}$$

where the indices s , l en sh denote secular, long-period and short-period, respectively, and \mathcal{F}_{α}^1 and \mathcal{F}_{α}^2 denote specific, sometimes very complicated, analytical expressions for a particular orbital element. Values of the parameters α_s , $\dot{\alpha}_s$, $\ddot{\alpha}_s$, ... are distributed by organizations like NASA, US Air Force Space Command¹ and ESA, where these parameters are computed on basis of the actual orbit of the satellite as determined from tracking data. The value of an osculating orbital element at any moment of time is then found from

$$\alpha(t) = \alpha_s(t) + \Delta\alpha_l(t) + \Delta\alpha_{sh}(t)\tag{23.20}$$

It is emphasized that this approach is somewhat ‘artificial’, because the basic quantities in the theory are the secularly changing ‘averaged’ orbital elements, while the tracking data on the position and velocity of the satellite, of course, correspond to the osculating orbital elements. So, with these methods usually some type of iterative searching procedure has to be applied to determine the secularly changing averaged elements at epoch from the osculating elements at epoch.

The various analytical orbit perturbations theories do not always give identical results and differ subtly in the way they separate secular, short-period, and long-period variations. This is of little consequence within a theory, but if we mix formulas from different theories, as is often done, their varying assumptions may lead to unintended results. Today, often semi-analytical theories are applied for modeling the orbital perturbations due to a variety of perturbing forces. These theories combine the best features of analytical and numerical techniques. The underlying approach is to separate the short-period contributions from the long-period and secular effects, and to propagate the secular and long-period variations due to all kinds of perturbing forces numerically. This is desirable because the short-period contributions seriously constrain the integrator step size; their separation from the long-period and secular effects allows us to numerically integrate the equations of motion governing the long-period and secular effects with large step sizes. The short-period contributions are then modeled analytically. They are recovered at the integrator’s step-size times and then combined with the secular and long-period contributions. If the short-period variations are needed at times other than the integrator times, an interpolation scheme is used.

23.3. First-order secular variation of the orbital elements

In the previous Section, we have derived expressions for the variation of the osculating orbital elements Ω and ω , relative to their values at a passage of the ascending node ($u = 0^\circ$). In practice, we are often primarily interested in the averaged long-term evolution of the orbit and then we want to know the secular variations of the orbital elements. The reason is that these represent the major long-term variations, while the long-period and short-period variations are bounded fluctuations about the secular changes. Therefore, these secular variations lead to a type of orbital elements that may be considered as *doubly-averaged* or *mean-mean* orbital elements, where the

¹ The well-known ‘two-line orbital elements’ are distributed by Space Track, US Air Force Space Command, Colorado Springs, USA.

long- and short-period variations have been averaged out. The sum of the secular and long-period variations leads to a type of orbital elements which are usually called *averaged* or *mean* orbital elements. It is emphasized that the term ‘mean element’ does not refer to any numerical average (i.e. mean) of a sampling of the variation of an osculating element and is not related to statistics at all. Mean and mean-mean orbital elements theory began with the work of J.L. Lagrange (1736-1813) and has been developed further by many experts over many years. In essence, it is a mathematical technique for approximating motion by separating the effects of fast variations from those that are more slowly varying. In order to sort out such terms, we should remember that short-period perturbations result from the variation of Θ around the orbit, while the long-period perturbations arise from the secular variation of ω , caused by e.g. the J_2 -term of the geopotential.

A traditional method in celestial mechanics and astrodynamics is to apply Lagrange's planetary equations ((22-35), (22-45)), which have been derived to compute the variation of the osculating orbital elements, also for the propagation of mean or mean-mean orbital elements from their initial values and for a particular perturbing force function, \tilde{R} . Of course, the orbital elements on the right-hand side of these equations should then be considered as mean or mean-mean orbital elements, and for the derivations of the partial derivatives of \tilde{R} to the orbital elements only those parts of the perturbing force function should be taken which lead to the long-period or secular orbit perturbations. In this Section, we will only consider the perturbations due to the J_2 -term of the gravity field. Then, \tilde{R} is given by (23.10) and, using the relation $u = \omega + \theta$, may also be written as:

$$\tilde{R} = \frac{3}{2} J_2 \frac{\mu R^2}{a^3} \left(\frac{a}{r} \right)^3 \left[\frac{1}{3} - \frac{1}{2} \sin^2 i + \frac{1}{2} \sin^2 i \cos 2(\omega + \theta) \right]$$

The true anomaly, θ , is easily transformed to the mean anomaly, M (Section 6.5), which itself is a linear function of time in unperturbed motion. The quantities a/r and θ in the perturbing force function are then functions of e and M only (Section 6.5) and are periodic with respect to M . With this in mind, we conclude from the expression for \tilde{R} that terms depending neither on M nor on ω lead to secular variations of the orbital elements, terms depending on ω but not on M lead to long-period variations of the elements, and terms depending on M lead to short-period variations of the elements. So, when we want to derive expressions for the long-period variations of the orbital elements, we take the mean value of \tilde{R} with respect to M , where we use the time-averaged value of $(a/r)^3$ and retain only terms of the expression in square brackets which are dependent on ω . We then find that this averaged \tilde{R} is zero, but this is certainly not the case when we would have considered odd-degree zonal harmonics (Section 23.7). When we want to derive expressions for the secular variations of the orbital elements, we average \tilde{R} with respect to M , where we use the time-averaged value of $(a/r)^3$ and retain only terms of the expression in square brackets which are dependent neither on M nor on ω . Various authors have exploited the method of averaging in order to derive analytical expressions for higher-order secular and periodic variations of the orbital elements. However, the derivation of these expressions is often rather vaguely formulated and then lacks a rigorous mathematical foundation. Some authors have even misused the method and some results of their higher-order analyses are wrong as the consequence of that misuse.

We now assume that the mean-mean orbital elements at a certain time (epoch) are known and that we want to compute the first-order secular variations of the orbital elements due to the J_2 -term. As explained above, we then should start from the following averaged form of the perturbing force function:

$$\tilde{R}_{mm} = \frac{3}{2} J_2 \frac{\mu R^2}{a^3} \left(\frac{a}{r} \right)_{mm}^3 \left[\frac{1}{3} - \frac{1}{2} \sin^2 i \right] \quad (23.21)$$

where $(a/r)^3_{mm}$ is the time-averaged value of $(a/r)^3$ over an orbital revolution. Note that \tilde{R}_{mm} is a function of the doubly-averaged (mean-mean) orbital elements a , e and i only. So, only the partial derivatives of \tilde{R} to a , e and i are non-zero. Inspection of Lagrange's planetary equations (22.35) and (22.45) reveals that the elements a , e and i therefore will not undergo a secular variation; only the elements ω , Ω and M (or τ) will exhibit secular variations. Now:

$$\left(\frac{a}{r} \right)_{mm}^3 = \frac{1}{2\pi} \int_0^{2\pi} \left(\frac{a}{r} \right)^3 dM = \frac{1}{2\pi} \int_0^{2\pi} \left(\frac{a}{r} \right)^3 \frac{dM}{dE} \frac{dE}{d\theta} d\theta \quad (23.22)$$

From (6.3), (6.33), (6.36-3) and (6.43) we find

$$\frac{r}{a} = 1 - e \cos E = \frac{1 - e^2}{1 + e \cos \theta} ; \quad \frac{dM}{dE} = 1 - e \cos E ; \quad \frac{dE}{d\theta} = \frac{1 - e \cos E}{\sqrt{1 - e^2}}$$

Substitution of these relations into (23.22) yields, since e does not show a secular variation,

$$\left(\frac{a}{r} \right)_{mm}^3 = (1 - e^2)^{-3/2}$$

Substitution of this relation into (23.21) gives

$$\tilde{R}_{mm} = \frac{3}{2} J_2 \frac{\mu R^2}{a^3} (1 - e^2)^{-3/2} \left(\frac{1}{3} - \frac{1}{2} \sin^2 i \right) \quad (23.23)$$

To compute the secular variations of ω , Ω and M , the partial derivatives of \tilde{R}_{mm} with respect to a , e and i have to be determined. From (23.23) we find

$$\begin{aligned} \frac{\partial \tilde{R}_{mm}}{\partial a} &= -\frac{9}{2} J_2 \frac{\mu R^2}{a^4} (1 - e^2)^{-3/2} \left(\frac{1}{3} - \frac{1}{2} \sin^2 i \right) \\ \frac{\partial \tilde{R}_{mm}}{\partial e} &= \frac{9}{2} J_2 \frac{\mu R^2}{a^3} e (1 - e^2)^{-5/2} \left(\frac{1}{3} - \frac{1}{2} \sin^2 i \right) \\ \frac{\partial \tilde{R}_{mm}}{\partial i} &= -\frac{3}{2} J_2 \frac{\mu R^2}{a^3} (1 - e^2)^{-3/2} \sin i \cos i \end{aligned} \quad (23.24)$$

Equations (22.35) and (22.45) show that the Lagrange planetary equations for ω , Ω , τ and M contain terms with e and/or $\sin i$ in the denominator, and in Section 22.4 it was mentioned that this may lead to problems for small values of e and i , unless the associated partial derivatives of the perturbing force function $\partial \tilde{R}/\partial e$ and $\partial \tilde{R}/\partial i$ contain factors with e or $\sin i$ in the numerator. From (23.24) we conclude that this is indeed the case for the doubly-averaged force function that describes the secular J_2 -perturbations. Starting with the Lagrange planetary equation (22.45), we write, after substitution of (23.24-1) and (23.24-2),

$$\frac{dM}{dt} = n + \frac{3}{4} J_2 \frac{\mu R^2}{n a^5} \frac{3 \cos^2 i - 1}{(1 - e^2)^{3/2}}$$

where n is given by (23.12). Because we want to find a first-order analytical solution for the secular variation of the orbital elements, all (mean-mean) orbital elements on the right-hand side of this expression should be considered as constants. Substitution of (23.12) then yields for the first-order secular variation of M :

$$\frac{dM}{dt} = n \left[1 + \frac{3}{4} J_2 \left(\frac{R}{a} \right)^2 (1 - e^2)^{-3/2} (3 \cos^2 i - 1) \right] \quad (23.25)$$

We now define

$$\tilde{n} \equiv \frac{dM}{dt} \quad (23.26)$$

where \tilde{n} is the *perturbed mean motion*, that is related to the unperturbed (constant) mean motion, n , through

$$\tilde{n} = n \left[1 + \frac{3}{4} J_2 \left(\frac{R}{a} \right)^2 (1 - e^2)^{-3/2} (3 \cos^2 i - 1) \right] \quad (23.27)$$

Note that \tilde{n} is a function of a , e and i , while n is a function of a only. Since a , e , and i do not exhibit a secular variation, \tilde{n} is also constant. Adopting some convenient epoch $t = t_0$, where $M = M_0$, the integral of (23.25) can be written as

$$M = M_0 + \tilde{n}(t - t_0) = M_0 + n \left[1 + \frac{3}{4} J_2 \left(\frac{R}{a} \right)^2 (1 - e^2)^{-3/2} (3 \cos^2 i - 1) \right] (t - t_0) \quad (23.28)$$

This equation may also be written as

$$M = M_0 + n(t - t_0) + \frac{3}{4} n J_2 \left(\frac{R}{a} \right)^2 (1 - e^2)^{-3/2} (3 \cos^2 i - 1) (t - t_0)$$

The first two terms on the right-hand side express the variation of M in an unperturbed orbit corresponding to the mean-mean orbital elements at epoch; the third term expresses the perturbation in M due to the J_2 -term.

In astrodynamics, often the semi-major axis, a , and the mean angular motion, n , are used together. If osculating orbital elements were used, there would be no difficulty; n would even not be necessary as a separate parameter and is given by the relation for Keplerian orbits (23.12): $n^2 a^3 = \mu$. However, when using mean-mean orbital elements, the physically appropriate definition of the relation between a and n requires attention. It should be realized that the semi-major axis does not experience a secular nor a long-period variation (Section 21.2) and we thus may write $a = a_s + \Delta a_{sh}$, where a is the osculating value, a_s the (constant) mean-mean value and Δa_{sh} the short-period contribution. It is immediately apparent that the split in a_s and Δa_{sh} is not uniquely defined; the origin of a_s is essentially arbitrary as long as the sum of a_s and Δa_{sh} equals the value of the osculating semi-major axis. Instead of the mean-mean semi-major axis, a , we may define a modified mean-mean semi-major axis, \tilde{a} , by a relation similar to the one for a Keplerian orbit: $\tilde{n}^2 \tilde{a}^3 = \mu$. Substitution of (23.27) and subsequent linearization, retaining only

terms of the first order in J_2 , then leads to

$$\tilde{a} = a \left[1 - \frac{1}{2} J_2 \left(\frac{R}{a} \right)^2 (1 - e^2)^{-3/2} (3 \cos^2 i - 1) \right] \quad (23.29)$$

Note that for $54.7^\circ < i < 125.3^\circ$: $\tilde{a} > a$, and that for other values of i : $\tilde{a} < a$. In the literature, another modified mean-mean semi-major axis, introduced by Y. Kozai (1928-) in 1959, is often used:

$$\tilde{a} = a \left[1 - \frac{3}{4} J_2 \left(\frac{R}{a} \right)^2 (1 - e^2)^{-3/2} (3 \cos^2 i - 1) \right] \quad (23.30)$$

which leads, with (23.27) and after linearization, to

$$\tilde{n}^2 \tilde{a}^3 = \mu \left[1 - \frac{3}{4} J_2 \left(\frac{R}{a} \right)^2 (1 - e^2)^{-3/2} (3 \cos^2 i - 1) \right] \quad (23.31)$$

Note that in this case the expressions in brackets for \tilde{a} and $\tilde{n}^2 \tilde{a}^3$ are identical. It is emphasized that the distance, r , is a function of the orbital elements a and e and the in-orbit position angle (θ or E), and that the mean value of the short-period variation of r is not necessarily zero. As will be shown in Section 23.4, Kozai added an approximate expression for the mean value of the short-period variation of r over an orbital revolution to the (constant) 'classical' mean-mean (mean) semi-major axis. In this way he obtained the expression for his modified mean-mean semi-major axis, which is nearly equal to the mean value of the geocentric distance of the satellite in a J_2 -perturbed orbit. In astronomy this has been a conventional definition of the semi-major axis adopted for natural satellites. For orbital altitudes between 300 km and 1000 km, the difference between Kozai's mean-mean semi-major axis and the 'classical' mean-mean semi-major axis varies from about -10 km at $i = 0^\circ$ to about 5 km for $i = 90^\circ$; this difference is zero for $i = 54.74^\circ, 125.26^\circ$.

For the first-order secular variation of Ω , we start from the Lagrange planetary equation (22.35-5), substitute (23.12), (23.24-3) and $p = a(1 - e^2)$, and find

$$\frac{d\Omega}{dt} = -\frac{3}{2} J_2 \frac{\mu R^2}{n a^5 (1 - e^2)^2} \cos i = -\frac{3}{2} J_2 n \left(\frac{R}{p} \right)^2 \cos i \quad (23.32)$$

For the integration of this expression, the (mean-mean) orbital elements on the right-hand side of (23.32) should be considered as constants. Then, there is a subtle problem with the mean (angular) motion, n . This parameter is linked to the unperturbed semi-major axis through (23.12), while in the secularly changing orbit the perturbed mean motion, \tilde{n} , is linked to the mean anomaly, M , which describes the true perturbed along-track position of the satellite. Therefore, traditionally, most authors replace n by \tilde{n} in (23.32). Mathematically, this replacement is not necessary, because the difference between \tilde{n} and n is of order J_2 and n or \tilde{n} in (23.32) is multiplied by a term that is of order J_2 . So the difference in the value of $\Omega - \Omega_0$ obtained from (23.32) by using n or \tilde{n} is of order J_2^2 , which is negligible since (23.32) provides a first-order solution of the secular change of Ω . Most modern authors therefore do not replace n by \tilde{n} . In this Section, however, we do replace n by \tilde{n} , because this is more consistent with our method of deriving expressions for the secular change of the orbital elements. Integration of (23.32) then gives

$$\Omega = \Omega_0 - \left[\frac{3}{2} \tilde{n} J_2 \left(\frac{R}{p} \right)^2 \cos i \right] (t - t_0) \quad (23.33)$$

In an analogous fashion, the expression for the first-order secular change of ω may be obtained:

$$\omega = \omega_0 + \left[\frac{3}{4} \tilde{n} J_2 \left(\frac{R}{p} \right)^2 (5 \cos^2 i - 1) \right] (t - t_0) \quad (23.34)$$

From (23.28), (23.33) and (23.34) we find for the secular rate of change of the orbital parameter M , Ω and ω :

$$\begin{aligned} \dot{M} &= \tilde{n} \\ \dot{\Omega} &= -\frac{3}{2} \tilde{n} J_2 \left(\frac{R}{p} \right)^2 \cos i \\ \dot{\omega} &= \frac{3}{4} \tilde{n} J_2 \left(\frac{R}{p} \right)^2 (5 \cos^2 i - 1) \end{aligned} \quad (23.35)$$

These secular rates of the orbital parameters M , Ω , ω play an important role in the selection of an appropriate satellite orbit for a particular mission; this issue is discussed in Section 23.7. In Section 23.5, it will be shown that only the even zonal harmonics terms of the Earth's gravity field will produce secular variations in M , Ω , ω . In practice, often the contributions of a limited number of low-degree even zonal harmonics terms, in addition to the J_2 -term, are included to obtain accurate expressions for the secular variations of M , Ω , ω . However, if we do so we have to realize that J_2 is of order 10^{-3} , and that J_4 , J_6 , etc. are of order 10^{-6} or smaller (Table 20.1). When we derive, for instance by applying the method of successive approximations discussed in Section 22.6, expressions for the second-order contributions of the J_2 -term we find that these expressions contains a coefficient J_2^2 , which is of order 10^{-6} . This means that the second-order contribution of the J_2 -term is of the same order of magnitude as the first-order contributions of the higher-degree zonal harmonics terms and thus has to be included. For the analysis presented in this book we will, however, use (23.35) to describe the secular variations of M , Ω , ω .

For pure Keplerian motion, we have found that the orbital period can be expressed by

$$T = \frac{2\pi}{n}$$

For perturbed orbits there are various definitions for the orbital period, where the expression derived for a particular definition often depends on the analytical theory adopted and on the type of orbital elements involved; i.e. osculating, mean or mean-mean elements. The *anomalistic period*, T_A , is defined as the period between two successive passages through perigee; i.e. the time to advance 2π in mean anomaly. Expression (23.26) shows that, according to the analytical theory used in the Section, we thus can write for the mean-mean anomalistic period:

$$T_A = \frac{2\pi}{\tilde{n}} \quad (23.36)$$

With this relation we find for the secular change of an orbital element after one anomalistic period:

$$\begin{aligned}\Delta_{2\pi} \alpha &= \Delta_{2\pi} e = \Delta_{2\pi} i = 0 \\ \Delta_{2\pi} \Omega &= -3\pi J_2 \left(\frac{R}{p} \right)^2 \cos i \\ \Delta_{2\pi} \omega &= \frac{3}{2}\pi J_2 \left(\frac{R}{p} \right)^2 (5 \cos^2 i - 1) \\ \Delta_{2\pi} M &= 2\pi\end{aligned}\tag{23.37}$$

When we compare the secular term in (23.18) with (23.37-3) we note that the secular change of ω after an orbital revolution differs for both equations. In (23.18) the secular change is proportional to the factor $(3 - 4\sin^2 i)$; in (23.37-3) this change is proportional to the factor $(5\cos^2 i - 1)$. This difference is a result of the fact that both equations hold for different conditions. In (23.18) the orbital elements are osculating ones, their variation is expressed as a function of the unperturbed argument of latitude, and an orbital revolution is defined as the time it takes the satellite to move from $u = 0$ to $u = 2\pi$; in (23.37) the orbital elements are secular ones, their variation is a function of time, and an orbital revolution is defined as the time it takes the satellite to move from perigee passage to the next perigee passage. Note that this difference in modeling the secular change of Ω after an orbital revolution does not lead to differences between (23.15) and (23.37-2).

Inspection of (23.37-2) reveals that for $i \neq 90^\circ$ the orbital plane shows a secular rotation about the polar axis. That means a motion of the line of nodes in the equatorial plane. For $i < 90^\circ$, the line of nodes rotates in a westerly direction; for $i > 90^\circ$, in an easterly direction. Equation (23.37-3) shows that, unless $i = 63.435^\circ$ or $i = 116.565^\circ$ (*critical inclinations*), the argument of perigee exhibits a secular variation. This is equivalent to a rotation of the major axis in the orbital plane. For $0^\circ \leq i < 63.435^\circ$ and $116.565^\circ < i < 180^\circ$, the major axis rotates in the direction of motion of the satellite; for $63.435^\circ < i < 116.565^\circ$, the rotation of the major axis is in the opposite direction. In Figure 23.2 the secular rate of change of Ω and of ω are plotted as a function of i and p/R for $i \leq 90^\circ$. Note that the changes in the elements are at maximum about one degree per revolution at $i \approx 0^\circ$, and that, for a particular inclination, the changes are larger for

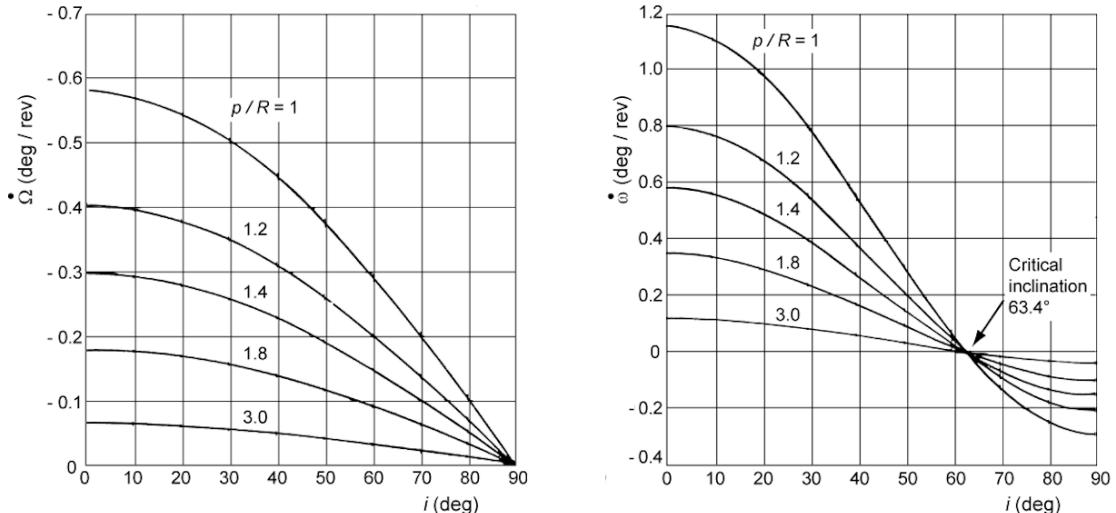


Figure 23.2: Secular rate of change of the right ascension of the ascending node (left) and of the argument of perigee (right), as a function of the orbital inclination and semi-latus rectum.

satellites at lower altitudes. For $90^\circ < i < 180^\circ$, the curves for the secular changes of Ω and ω are a mirror image of the ones shown in Figure 23.2.

In the literature, many discussions are devoted to the physical meaning and consequences of the existence of the critical inclinations. We have found that in the expression for the first-order secular change of ω ((23.34)) the factor $(5\cos^2 i - 1)$ occurs in the numerator, but in analytical theories for the long-period variations of the orbital elements due to zonal harmonics this term often appears in the denominator. This seems to indicate that the orbit experiences extremely large perturbations around the critical inclinations. However, this is not true! The troubles in the analysis of orbits at, or close to, the critical inclinations only emerge in some analytical approaches, whereas there are no difficulties in the numerical integration of orbits at the critical inclinations. It is only a mathematical misapplication of perturbations theory that leads to the appearance of infinities in the analytical expressions for the orbit perturbations, and the small denominator problem can be avoided by using special ways for handling the critical terms. It has been shown by A. Deprit (1926-2006) and his colleagues around 1983 that the critical inclinations are associated with unstable equilibria in the phase space of J_2 -perturbed satellite dynamics. It is noted that the second-order contribution of the J_2 -term and the first-order contributions of other even zonal harmonics to the secular variation of Ω are also zero at $i = 90^\circ$. These additional terms also affect the values of the critical inclinations. As an example, when the effects of the J_2 -, J_2^2 -, J_4 -, and J_6 -terms are accounted for the critical inclinations are 63.441° and 116.559° when $a = 6800$ km and $e < 0.01$; for $a = 7400$ km and $e < 0.01$ the critical inclinations are 63.438° and 116.562° . For $a = 6800 - 7400$ km and $e < 0.1$, the secular variation of ω due to the J_2^2 -, J_4 - or J_6 -term is larger than 10% of the secular variation of ω due to the J_2 -term when $i = 63.2^\circ - 63.7^\circ$ and $i = 116.3^\circ - 116.8^\circ$. So, within these ranges of inclinations the first-order analysis accounting for the J_2 -term only does not produce accurate results.

In addition to the (mean-mean) anomalistic period defined above, in astrodynamics the *draconic (draconitic) period* or *nodal period* is often used. This period is defined as the time interval between two successive passages through the ascending node, so the period it takes to travel from $u = 0$ to $u = 2\pi$. In orbit computations its is often more convenient to use the nodal period then the anomalistic period. Disadvantages of the anomalistic period are that the perigee location changes with time, and that for near-circular orbits perigee is hard to define from observations. Use of the nodal period eliminates these two disadvantages, since the equator crossing is a well-defined point even in (near-)circular satellite orbits. The mean-mean nodal period can be expressed by $2\pi/\dot{u}$, where \dot{u} denotes the secular variation of the argument of latitude, u . Now, $\dot{u} = \dot{\omega} + \dot{\Theta}$, where $\dot{\omega}$ and $\dot{\Theta}$ also denote secular variations. Equation (6.45) presents an expression for $\dot{\Theta}$ as a function of e and M . When we apply that expression for the mean-mean orbital parameters, differentiate it with respect to time, realize that e does not experience a secular variation, and average the resulting expression with respect to M , we find for the relation between the secular variations of $\dot{\Theta}$ and \dot{M} : $\dot{\Theta} = \dot{M} = \tilde{n}$. We then find for the mean-mean nodal period

$$T_N = \frac{2\pi}{\tilde{n} + \dot{\omega}} \quad (23.38)$$

Substitution of (23.27) into (23.36), and of (23.27) and (23.35-2) into (23.38), yields the following approximative expressions for the mean-mean anomalistic and nodal periods of near-circular orbits:

$$\begin{aligned} T_A &= 2\pi \sqrt{\frac{a^3}{\mu}} \left[1 - \frac{3}{4} J_2 \left(\frac{R}{a} \right)^2 (3 \cos^2 i - 1) \right] \\ T_N &= 2\pi \sqrt{\frac{a^3}{\mu}} \left[1 - \frac{3}{2} J_2 \left(\frac{R}{a} \right)^2 (4 \cos^2 i - 1) \right] \end{aligned} \quad (23.39)$$

where $2\pi\sqrt{a^3/\mu}$ is the period of the (unperturbed) Keplerian orbit. When we use (23.30) to write the expressions (23.39) as a function of Kozai's mean-mean semi-major axis, \tilde{a} , we find the following approximate relations:

$$\begin{aligned} T_A &= 2\pi \sqrt{\frac{\tilde{a}^3}{\mu}} \left[1 + \frac{3}{8} J_2 \left(\frac{R}{\tilde{a}} \right)^2 (3 \cos^2 i - 1) \right] \\ T_N &= 2\pi \sqrt{\frac{\tilde{a}^3}{\mu}} \left[1 - \frac{3}{8} J_2 \left(\frac{R}{\tilde{a}} \right)^2 (7 \cos^2 i - 1) \right] \end{aligned}$$

where $2\pi\sqrt{\tilde{a}^3/\mu}$ is the period of a Keplerian orbit with semi-major axis \tilde{a} . In the literature both sets of expressions are found. In this Chapter we will use the expressions (23.39). Note that the expression for the nodal period contains the factor $(4\cos^2 i - 1)$. This factor may also be written as $(3 - 4\sin^2 i)$, which is just the factor that is present in the secular term of (23.18). This confirms the conclusion reached above about the difference in the definitions of the secular terms in (23-18) and (23.37-3). From the approximative relations for T_A and T_N given in (23.39) we obtain:

$$T_A - T_N = \frac{3}{2}\pi J_2 \sqrt{\frac{a^3}{\mu}} \left(\frac{R}{a} \right)^2 (5 \cos^2 i - 1) \quad (23.39)$$

Equations (23.39) show that, for a specified value of a , the anomalistic period is equal to the Keplerian period if $\cos^2 i = 1/3$, or $i = 54.7^\circ$ and $i = 125.3^\circ$. For $i < 54.7^\circ$ and $i > 125.3^\circ$, the anomalistic period is shorter than the Keplerian period; for $54.7^\circ < i < 125.3^\circ$, the anomalistic period is longer than the Keplerian period. Equations (23.39) also show that the nodal period is equal to the Keplerian period if $\cos^2 i = 1/4$, or $i = 60^\circ$ and $i = 120^\circ$. For $i < 60^\circ$ and $i > 120^\circ$, the nodal period is shorter than the Keplerian period; in the intermediate interval, the nodal period is longer than the Keplerian period. The anomalistic period is, for a specified value of a , equal to the nodal period if $\cos^2 i = 1/5$. This is just the case at the critical inclinations ($i = 63.435^\circ$, 116.565°), as could be expected. Note that the differences between the anomalistic or nodal period and the Keplerian period are largest for low-altitude near-equatorial orbits. For an orbital altitude of 300 - 1000 km and an inclination of about 0° , the differences between the anomalistic or nodal period and the Keplerian period are about -8 s and -24 s, respectively; for an inclination of 90° the corresponding values are 4 s and 8 s.

In addition to the anomalistic and nodal orbital periods discussed above, the (mean-mean) *nodal period of Greenwich*, $T_{\Theta G}$, or (mean-mean) *nodal day* is often used in astrodynamics. This quantity is defined as

$$T_{\Theta G} = \frac{2\pi}{\dot{\Theta} - \dot{\Omega}} \quad (23.40)$$

where $\dot{\Theta}$ is the Earth's mean (sidereal) rotation rate. The nodal day is the time it takes the Earth to complete one revolution about its axis relative to the precessing orbital plane, or, specified

more accurately, relative to the line of nodes. A first-order approximation of $\dot{\Omega}$ is given by (23.35-2). If $i < 90^\circ$: $\dot{\Omega} < 0$ and thus $T_{\theta G} < 2\pi/\dot{\Omega}$, where $2\pi/\dot{\Omega}$ is the length of a sidereal day (Section 11.4). If $i > 90^\circ$: $\dot{\Omega} > 0$ and thus $T_{\theta G} > 2\pi/\dot{\Omega}$. For a circular orbit at an altitude of more than 300 km, the difference between a nodal day and a sidereal day is less than 35 min, depending on the orbital inclination; for near-polar orbits with $80^\circ < i < 100^\circ$, this difference is less than 5.9 min. For a particular inclination the difference between a nodal day and a sidereal day decreases with increasing orbital altitude; for $i = 90^\circ$ the difference is zero for all orbital altitudes.

23.4. Periodic variations of the orbital elements

In this Section we start with some examples of short- and long-period variations of orbital elements due to zonal harmonics (Section 20.1) of the geopotential. For clarity, the parameters describing the mean orbit will be indicated as $\bar{\alpha}$, where α is an arbitrary orbital parameter. Figure 23.3 shows the effects of the J_2 -term of the gravitational potential of the Earth on the osculating elements a , e , i , ω and Ω of the nominal orbit of the ANS satellite (Section 16.6) during 1.5 orbital revolution. The variations shown refer to a nominal orbit that differs slightly from the nominal orbit listed in Table 16.2. The difference is in particular noticeable in the eccentricity, which is about 0.008 in this case and about 0.003 in Table 16.2. In this Chapter we consider the orbit with an eccentricity of about 0.008 as the nominal one. Figure 23.3 shows that all five orbital elements exhibit a short-period variation during an orbital revolution. The amplitude of the oscillation in a is about 9 km; e varies between 0.0067 and 0.0091; i shows an oscillation with an amplitude of about 0.005° ; and ω varies between 153° and 179° . In Section 23.3 it was shown that the orbital elements ω and Ω also experience a secular variation. As a result of the applied scale, this secular variation is only visible for Ω in Figure 23.3. The plot also shows that the amplitude of the short-period variation of Ω is small with respect to the secular change of Ω after an orbital revolution. The amplitude of the short-period variation of ω is much larger than the secular change of ω after an orbital revolution, which is a direct result of the small value of the nominal eccentricity, as was mentioned in Section 23.2. Figure 23.4 shows similar plots for the

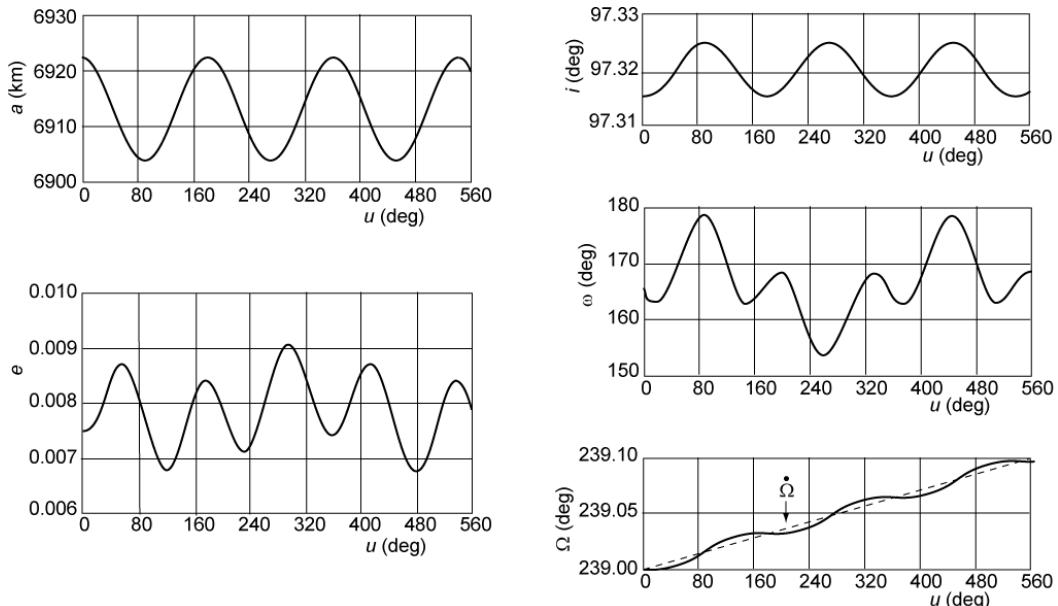


Figure 23.3: Variation of the osculating elements of the nominal ANS orbit during 1½ orbital revolution, due to the J_2 -term of the Earth's gravitational potential.

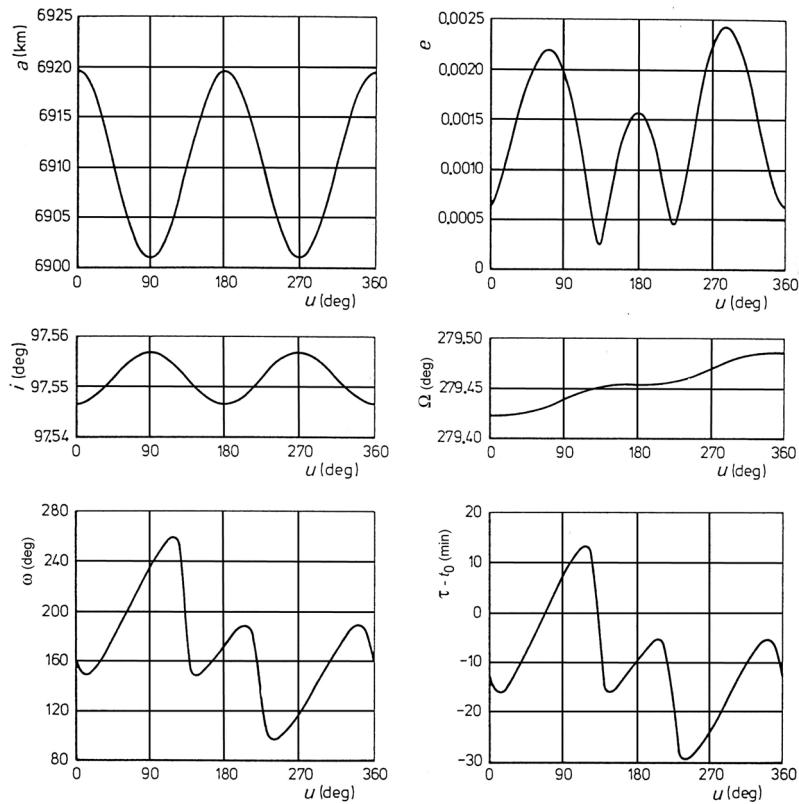


Figure 23.4: Variation of the TD-1A osculating orbital elements due to the J_2 -term of the Earth's gravitational potential, during the 392nd orbital revolution of the satellite.

short-period variation of the orbital elements of the ESA satellite TD-1A. This satellite was launched by a Thor-Delta rocket from the Western Test Range on March 12, 1972, and was injected into a near-circular orbit at an altitude of about 540 km, and with an inclination of 97.6° . Due to the very small value of the eccentricity of this orbit, the amplitude of the short-period variation of ω is even larger than that of the ANS orbit.

Figure 23.5 shows the effects of a number of higher-degree zonal harmonics on the osculating semi-major axis of the nominal ANS orbit during an orbital revolution. Note that the J_3 -term gives rise to short-period oscillations with an amplitude of about 25 m; the J_4 -term produces a variation of the order of 10 m. Together, the J_3 - through J_7 -term and J_9 -term produce a variation with a maximum of about 40 m. Figure 23.6 shows the history of the nominal mean orbital elements \bar{e} , $\bar{\omega}$ and $\bar{\Omega}$ of the ANS satellite over a period of 200 days. Here, the effects of the J_2 -through J_9 -term of the gravitational potential have been taken into account. Note that these elements show a long-period oscillation with a period of about 105 days, which corresponds to the period it takes perigee to complete one revolution about the Earth. Figure 23.7 (left) shows the variation of the mean eccentricity of the orbit of the TD-1A satellite. The solid curve represents the predicted variation due to the J_2 - to J_9 -term of the gravitational potential. The symbol \odot indicates the values of the eccentricity as computed from actual tracking data. Figure 23.7 (right) shows the variation of the mean eccentricity of the (true) orbit of the ANS satellite as computed from tracking data. As mentioned in Section 16.6, this satellite was injected into an elliptical orbit with perigee and apogee altitude of 266 km and 1175 km, respectively. Due to the low perigee altitude, the atmospheric drag led to a continuous decrease of the mean eccentricity.

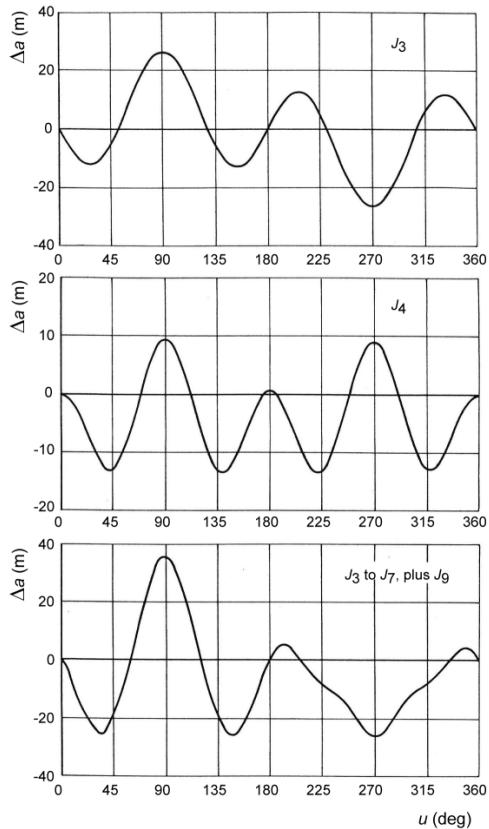


Figure 23.5: Variation of the osculating semi-major axis of the nominal ANS orbit during an orbital revolution, due to the J_3 -term (top), J_4 -term (center), and J_3 - to J_7 , plus J_9 -term (bottom) of the Earth's gravitational potential.

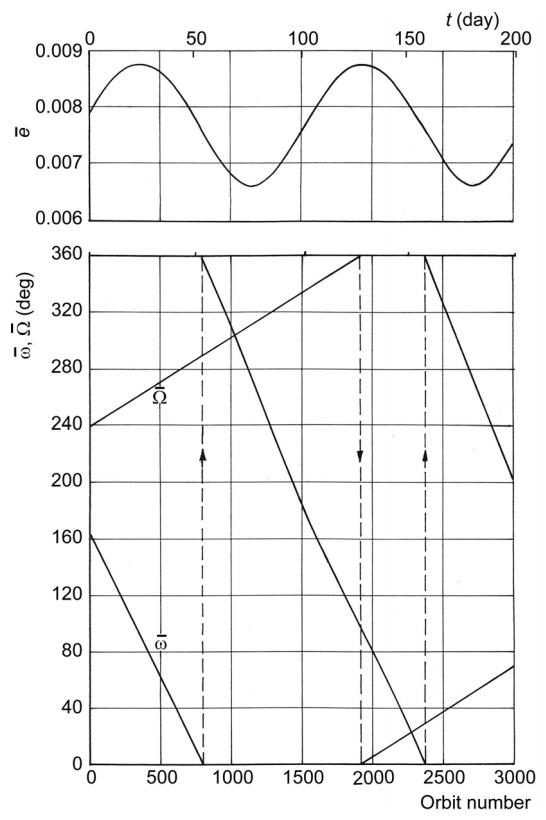


Figure 23.6: Variation of the mean elements of the nominal ANS orbit over a period of 200 days, due to the J_2 - to J_9 -term of the Earth's gravitational potential.

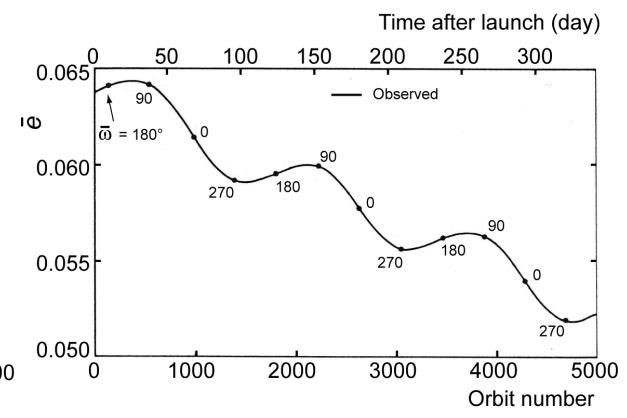
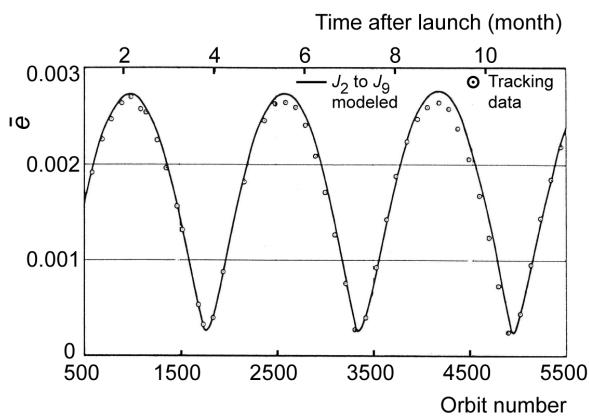


Figure 23.7: Variation of the mean eccentricity of the TD-IA (left) and ANS (right) orbits.

When the mean orbital elements are known, the osculating orbital elements can be obtained by adding the short-period variations. For the computation of these short-period variations, a number of analytical theories is available. As an example, Figure 23.8 (left) shows for an orbit with the mean orbital parameters $\bar{a} = 7500 \text{ km}$, $\bar{e} = 0.001$, $i = 45^\circ$, $\Omega = \bar{\omega} = M = 0^\circ$ the short-period variation of the orbital elements i and Ω , and of the argument of latitude, $u = \omega + \theta$, as computed

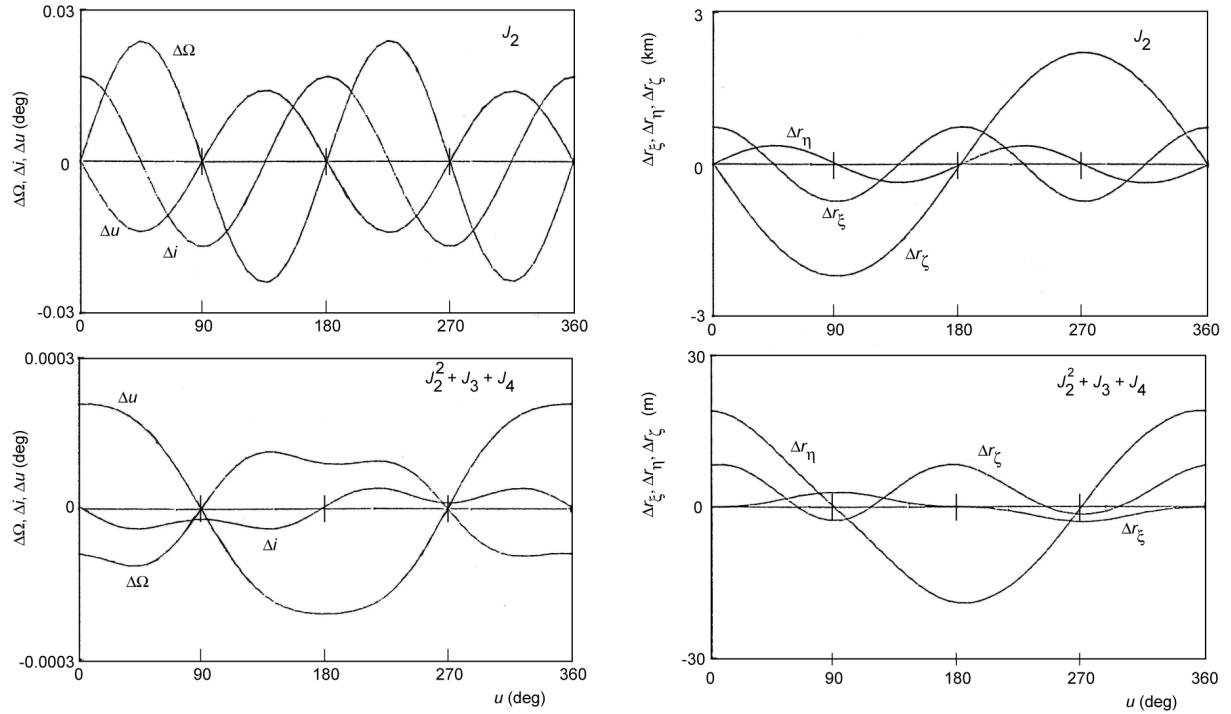


Figure 23.8: Short-period variation of the argument of latitude, u , and of the orbital elements i and Ω (left), and of the radial (ξ), normal (η) and cross-track (ζ) satellite position components (right) during an orbital revolution, due to the J_2 - to J_4 -term of the Earth's gravitational potential, according to Kozai's analytical theory. The mean orbital parameters are $\bar{a} = 7500$ km, $\bar{e} = 0.001$, $i = 45^\circ$, $\bar{\Omega} = \bar{\omega} = M = 0$

from the analytical theory of Y. Kozai². Both the variations due to the J_2 -term only, and the variations due to the second-order effect of the J_2 -term (J_2^2) and of the first-order effect of the J_3 -term and of the J_4 -term are depicted. The variations of Δi and $\Delta\Omega$ shown for the J_2 -term are in good agreement with the short-period variations that can be computed from (23.9) and (23.15). One has to realize that orbital elements are only auxiliary quantities, which are used for the computation of the position and velocity of the satellite. Figure 23.8 (right) shows the short-period variation of the components of the satellite position vector in the radial direction (ξ), in the normal direction (in the orbital plane and perpendicular to the radial direction, η) and in the direction perpendicular to the orbital plane (cross-track, ζ). For this case, the position of the satellite was computed in two ways. First, the orbit was computed from the mean orbital elements that experience secular and long-period variation during an orbital revolution. Then, the orbit was recomputed, but now on basis of the osculating orbital elements. The differences between the position vectors obtained from both orbits, decomposed into the three orthogonal components, are depicted in Figure 23.8 (right), where the orbit described by the mean orbital elements served as reference. Note that the short-period variation due to the J_2 -term produces a component in the cross-track direction with an amplitude of about 2 km; the amplitudes in the radial and normal directions are less than 1 km. The short-period variations due to the second-order effect of the J_2 -term, the J_3 -term and J_4 -term together yield perturbations of up to 20 m. From (23.5) we find that the amplitude of the short-period variation of the semi-major axis due to the J_2 -term is for this orbit about 4.4 km. So, the amplitude of the short-period variation of the semi-major axis due

² Y. Kozai, Second-order solution of artificial satellite theory without air drag, The Astronomical Journal, Vol. 67, No. 7, pp. 446-461, 1962.

to the J_2 -term is significantly larger than the amplitude of the short-period variations of the position components due to the J_2 -term. The reason is that the distance, r , of the satellite is a function of the semi-major axis and the eccentricity of the orbit, and that the short-period variations in both elements counteract each other to a large extent. This phenomenon is characteristic for low-eccentricity orbits.

An efficient first-order method to compute the osculating orbital elements from the mean orbital elements is given by J.H. Kwok³ and is based on the analytical theories of Y. Kozai, D. Brouwer (1902-1966), and I.G. Izsák (1929-1965). This method, which accounts only for the effects of the J_2 -term in the gravitational potential of the Earth, avoids problems that occur for $\bar{e} \approx 0$, and circumvents numerical difficulties in situations where the perturbations of the semi-major axis are large. In this method, a special set of orbital parameters is used and the short-period variation of these parameters is expressed by

$$\begin{aligned}
\Delta i_{sh} &= \frac{3}{8} J_2 \left(\frac{R}{\bar{p}} \right)^2 \sin 2\bar{i} \left[\cos 2(\bar{\omega} + \bar{\theta}) + \bar{e} \cos(2\bar{\omega} + \bar{\theta}) + \frac{1}{3} \bar{e} \cos(2\bar{\omega} + 3\bar{\theta}) \right] \\
\Delta p_{sh} &= \frac{3}{2} J_2 \frac{R^2}{\bar{p}} \sin^2 \bar{i} \left[\cos 2(\bar{\omega} + \bar{\theta}) + \bar{e} \cos(2\bar{\omega} + \bar{\theta}) + \frac{1}{3} \bar{e} \cos(2\bar{\omega} + 3\bar{\theta}) \right] \\
\Delta \Omega_{sh} &= -\frac{3}{2} J_2 \left(\frac{R}{\bar{p}} \right)^2 \cos \bar{i} \left[\bar{\theta} - \bar{M} + \bar{e} \sin \bar{\theta} - \frac{1}{2} \sin 2(\bar{\omega} + \bar{\theta}) - \frac{1}{2} \bar{e} \sin(2\bar{\omega} + \bar{\theta}) - \frac{1}{6} \bar{e} \sin(2\bar{\omega} + 3\bar{\theta}) \right] \\
\Delta r_{sh} &= -\frac{1}{4} J_2 \frac{R^2}{\bar{p}} \left\{ (3 \cos^2 \bar{i} - 1) \left[\frac{2\sqrt{1-\bar{e}^2}}{1+\bar{e}\cos\bar{\theta}} + \frac{\bar{e}\cos\bar{\theta}}{1+\sqrt{1-\bar{e}^2}} + 1 \right] - \sin^2 \bar{i} \cos 2(\bar{\omega} + \bar{\theta}) \right\} \\
\Delta \dot{r}_{sh} &= \frac{1}{4} J_2 \sqrt{\frac{\mu}{\bar{p}}} \left(\frac{R}{\bar{p}} \right)^2 \left\{ (3 \cos^2 \bar{i} - 1) \bar{e} \sin \bar{\theta} \left[\sqrt{1-\bar{e}^2} + \frac{(1+\bar{e}\cos\bar{\theta})^2}{1+\sqrt{1-\bar{e}^2}} \right] \right. \\
&\quad \left. - 2 \sin^2 \bar{i} (1 + \bar{e} \cos \bar{\theta})^2 \sin 2(\bar{\omega} + \bar{\theta}) \right\} \\
\Delta u_{sh} &= -\frac{1}{8} J_2 \left(\frac{R}{\bar{p}} \right)^2 \left\{ 6(1 - 5 \cos^2 \bar{i})(\bar{\theta} - \bar{M}) + 4\bar{e} \sin \bar{\theta} \left[(1 - 6 \cos^2 \bar{i}) + \frac{1 - 3 \cos^2 \bar{i}}{1 + \sqrt{1 - \bar{e}^2}} \right] \right. \\
&\quad + \frac{1 - 3 \cos^2 \bar{i}}{1 + \sqrt{1 - \bar{e}^2}} \bar{e}^2 \sin 2\bar{\theta} + 2\bar{e}(5 \cos^2 \bar{i} - 2) \sin(2\bar{\omega} + \bar{\theta}) \\
&\quad \left. + (7 \cos^2 \bar{i} - 1) \sin 2(\bar{\omega} + \bar{\theta}) + 2\bar{e} \cos^2 \bar{i} \sin(2\bar{\omega} + 3\bar{\theta}) \right\}
\end{aligned} \tag{23.41}$$

where all parameters have their usual meaning. The osculating orbital elements may then be computed from

$$\begin{aligned}
i &= \bar{i} + \Delta i_{sh} \quad ; \quad p = \bar{p} + \Delta p_{sh} \quad ; \quad \Omega = \bar{\Omega} + \Delta \Omega_{sh} \\
r &= \bar{r} + \Delta r_{sh} \quad ; \quad \dot{r} = \bar{r} + \Delta \dot{r}_{sh} \quad ; \quad u = \bar{u} + \Delta \bar{u}_{sh}
\end{aligned} \tag{23.42}$$

³ J.H. Kwok, *The long-term orbit predictor (LOP)*, JPL report EM 312/86-151, Jet Propulsion Laboratory, Pasadena, 30 June 1986.

$$C_1 = \frac{p}{r} - 1 \quad ; \quad C_2 = \sqrt{\frac{p}{\mu}} \dot{r} \quad ; \quad e = \sqrt{C_1^2 + C_2^2} \quad ; \quad \theta = \arctan\left(\frac{C_2}{C_1}\right) \quad (23.42)$$

$$\alpha = \frac{p}{1 - e^2} \quad ; \quad \omega = u - \theta$$

Note that in this method the short-period variations are expressed as a function of the true anomaly, θ , while for the analyses in Sections 21.2, 23.1 and 23.2 the argument of latitude, $u = \theta + \omega$, was used as the independent variable. For near-circular orbits, we may neglect all terms that are proportional to \bar{e} and \bar{e}^2 in (23.41), and we may substitute $\bar{\Theta} - M = 0$ (Section 6.6). Then, we arrive at the expressions

$$\begin{aligned} \Delta i_{sh} &= \frac{3}{8} J_2 \left(\frac{R}{\bar{a}} \right)^2 \sin 2\bar{i} \cos 2\bar{u} \\ \Delta p_{sh} &= \frac{3}{2} J_2 \frac{R^2}{\bar{a}} \sin^2 \bar{i} \cos 2\bar{u} \\ \Delta \Omega_{sh} &= \frac{3}{4} J_2 \left(\frac{R}{\bar{a}} \right)^2 \cos \bar{i} \sin 2\bar{u} \\ \Delta r_{sh} &= -\frac{3}{4} J_2 \frac{R^2}{\bar{a}} \left[2 - 3 \sin^2 \bar{i} \left\{ 1 + \frac{1}{9} \cos 2\bar{u} \right\} \right] \\ \Delta \dot{r}_{sh} &= -\frac{1}{2} J_2 \sqrt{\frac{\mu}{\bar{a}}} \left(\frac{R}{\bar{a}} \right)^2 \sin^2 \bar{i} \sin 2\bar{u} \\ \Delta u_{sh} &= -\frac{1}{8} J_2 \left(\frac{R}{\bar{a}} \right)^2 (7 \cos^2 \bar{i} - 1) \sin 2\bar{u} \end{aligned} \quad (23.43)$$

These expressions show that, for given values of \bar{a} and \bar{i} , the short-period variations of the orbital parameters have, except for Δr_{sh} , a zero mean value over an orbital revolution. The extreme values of Δi_{sh} and Δp_{sh} occur at $\bar{u} = 0^\circ, 90^\circ, 180^\circ, 270^\circ$; the extreme values of $\Delta \Omega$, $\Delta \dot{r}_{sh}$ and Δu_{sh} occur at $\bar{u} = 45^\circ, 135^\circ, 225^\circ, 315^\circ$. The mean value (index m) of the variation of Δr_{sh} over an orbital revolution is given by

$$(\Delta r_{sh})_m = -\frac{3}{4} J_2 \frac{R^2}{\bar{a}} (3 \cos^2 \bar{i} - 1)$$

This mean value is zero if $\bar{i} = 54.7^\circ, 125.3^\circ$. If $0^\circ \leq \bar{i} < 54.7^\circ$ or $125.3^\circ < \bar{i} \leq 180^\circ$, then the mean value of Δr_{sh} is negative; if $54.7^\circ < \bar{i} < 125.3^\circ$, then the mean value is positive. When we compare these results with the results obtained in Section 23.3, we conclude that these cases correspond to the cases where the anomalistic orbital period is equal to, shorter than, or longer than the Keplerian orbital period, respectively. Note that the expression for $(\Delta r_{sh})_m$ is, for circular orbits, just equal to the correction applied by Kozai to the 'classical' mean-mean semi-major axis, as expressed by (23.30). So, we may write (23.30) as $\tilde{a} \approx a + (\Delta r_{sh})_m$, where \tilde{a} is Kozai's mean-mean semi-major axis and a is the 'classical' mean-mean semi-major axis. Since a does not exhibit a long-period variation and in a Keplerian orbit the time-averaged distance of a satellite

during an orbital revolution is, for near-circular orbits, nearly equal to a (Section 6.6), \tilde{a} closely approximates the mean distance of a satellite in a J_2 -perturbed orbit. For a (near-)circular orbit at an altitude of 500 km and with an inclination of 45° , we find for the extreme values $\Delta i_{sh} = \pm 1.2'$, $\Delta p_{sh} = \pm 4.8$ km, $\Delta\Omega_{sh} = \pm 1.7'$, $\Delta\dot{r}_{sh} = \pm 1.8$ m/s, $\Delta u_{sh} = \pm 1.0'$. For the mean value of Δr_{sh} we find a value of -2.4 km, and for the extreme values of Δr_{sh} we find -1.6 km and -3.2 km. For a circular equatorial orbit at an altitude of 500 km, we find for all values of \bar{u} : $\Delta r_{sh} = -9.6$ km, $\Delta\dot{r}_{sh} = 0$. So, in this case the short-period variation of r degenerates to a constant decrease of the orbital radius by 9.6 km.

In Section 21.2, we have derived expressions for the variation of the inclination and right ascension of the ascending node of circular orbits during an orbital revolution due to the J_2 -term of the Earth's gravitational potential, and it is interesting to compare those expressions with the expressions derived in this Section. We then should realize that (21.24-1) and (21.24-2) express the change of the osculating inclination and right ascension of the ascending node relative to their values at the ascending node; consequently, (21.24-2) accounts for the secular variation of Ω . The equations (23.43-1) and (23.43-3) express the short-period perturbations of i and Ω relative to the instantaneous values of the mean inclination and mean right ascension of the ascending node. When we account for these differences in definition, we conclude that both equations for Δi and both equations for $\Delta\Omega$ fully agree.

23.5 Full spectrum of orbit perturbations

In the previous Sections, we have discussed some elementary aspects of the variation of orbital elements due to a few zonal harmonics of the geopotential, and we have found that the orbital elements show secular, long-period and short-period variations. In reality, each harmonic of the geopotential will contribute to the variation of orbital elements and these elements show a rich spectrum of perturbations with many different frequencies. W.M. Kaula (1926-2000) has developed around 1961 an analytical linear perturbations theory⁴ to determine the effects of individual harmonics of the geopotential on a satellite orbit. His theory is quite complicated and the derivation of the relevant expressions is beyond the scope of this book. Nevertheless, below some aspects of his theory will be presented, and we will apply it in Section 23.6 for four characteristic classes of orbit perturbations. We will concentrate our analysis on low- to medium-altitude ($h < 5000$ km) nearly circular orbits.

Equation (20.1-1) describes the gravitational potential as a function of the satellite's position components r , ϕ , and Λ . Kaula used the relations between spherical coordinates and orbital elements (Section 11.9) to obtain an expression for the perturbing force function (Section 22.1) of the geopotential in terms of orbital elements:

$$\tilde{R} = \frac{\mu}{a} \sum_{n=2}^{\infty} \left(\frac{R}{a} \right)^n \sum_{m=0}^n \sum_{p=0}^n F_{n,m,p}(i) \sum_{q=-\infty}^{\infty} G_{n,p,q}(e) Z_{n,m,p,q}(\omega, \Omega, M, \theta_{GM}) \quad (23.44)$$

where

$$Z_{n,m,p,q} = X_{n,m} \cos \Theta_{n,m,p,q} + Y_{n,m} \sin \Theta_{n,m,p,q} \quad (23.44)$$

$n - m$ even: $X_{n,m} = C_{n,m}$, $Y_{n,m} = S_{n,m}$; $n - m$ odd: $X_{n,m} = -S_{n,m}$, $Y_{n,m} = C_{n,m}$

with the phase angle

⁴ W.M. Kaula, *Theory of Satellite Geodesy*, Dover Publications, Mineola, New York, 2000.

$$\Theta_{n,m,p,q} = (n - 2p)\omega + (n - 2p + q)M + m(\Omega - \theta_{GM}) \quad (23.45)$$

The parameters a , e , i , ω , Ω and M are the osculating orbital elements, θ_{GM} is the Greenwich mean sidereal time (Section 11.4), μ and R are the gravitational parameter and mean equatorial radius of the Earth, respectively, n and m are the degree and order of the spherical harmonic, respectively, and $C_{n,m}$ and $S_{n,m}$ are gravity field model coefficients (Section 20.1). For a specified orbit and time, a particular set of $\{n,m,p,q\}$ determines the value of $\Theta_{n,m,p,q}$.

Each $F_{n,m,p}(i)$ is a finite series of trigonometric functions of the inclination. An important characteristic of $F_{n,m,p}(i)$ for zonal harmonics ($m = 0$) is

$$F_{n,0,p}(i) = F_{n,0,n-p}(i) \text{ if } n \text{ is even} ; \quad F_{n,0,p}(i) = -F_{n,0,n-p}(i) \text{ if } n \text{ is odd}$$

Each $G_{n,p,q}(e)$ is an infinite power series of the eccentricity. These eccentricity series represent a subset of the *Hansen coefficients* (P.A. Hansen, 1795-1874) and were published in 1855. Each term $G_{n,p,q}(e)$ is of the order of the $|q|$ power of the eccentricity ($e^{|q|}$), except for the terms $G_{2,0,-2}(e)$ and $G_{2,2,2}(e)$ which have the value zero. Consequently, while the summation over q ranges in theory from $-\infty$ to $+\infty$, for near-circular orbits only a few terms need to be included in the summation. For high-eccentricity orbits, many terms have to be included and the convergence of the function is then very slow. An important characteristic of $G_{n,p,q}(e)$ is

$$G_{n,p,q} = G_{n,n-p,-q}$$

For nearly circular orbits we may use the approximate relations

$$G_{n,p,0}(e) = 1 + O(e^2)$$

$$G_{n,p,1}(e) = \frac{1}{2}e(3n - 4p + 1) + O(e^3)$$

$$G_{n,p,-1}(e) = \frac{1}{2}e(-n + 4p + 1) + O(e^3)$$

where the notation O stands for ‘order of’.

From (23.44) we may derive the partial derivatives of $\tilde{\mathbf{R}}$ with respect to the orbital elements. These expressions can be substituted into the right-hand sides of the Lagrange planetary equations ((22.35), (22.45)). For a first-order solution of these equations, it is assumed that the orbital elements a , e , and i in (23.44) and on the right-hand sides of (22.35) and (22.45) are constant, that the orbital elements ω , Ω , and M in the expressions for the partial derivatives of $\tilde{\mathbf{R}}$ exhibit the first-order secular variations according to (23.27) and (23.35), and that the Greenwich mean sidereal time, θ_{GM} , varies linearly with time. With these assumptions, integration of (22.35) and (22.45) leads, for each combination $\{n,m,p,q\}$ in (23.44), to first-order expressions for the perturbations of the orbital elements. Noteworthy is that the time-dependent parameters ω , Ω , M and θ_{GM} are all contained in the argument $\Theta_{n,m,p,q}$, which significantly eases the integration of the Lagrange planetary equations. Kaula found the following expressions for the perturbations of the orbital elements:

$$\begin{aligned} \Delta\alpha &= \sum_{n=2}^{\infty} \sum_{m=0}^n \sum_{p=0}^n \sum_{q=-\infty}^{\infty} \Delta\alpha_{n,m,p,q} \\ \Delta\alpha_{n,m,p,q} &= 2\sqrt{\frac{\mu}{a}} \left(\frac{R}{a} \right)^n F_{n,m,p}(i) G_{n,p,q}(e) (n - 2p + q) \frac{Z_{n,m,p,q}}{\dot{\Theta}_{n,m,p,q}} \end{aligned} \quad (23.46)$$

$$\begin{aligned}
\Delta e_{n,m,p,q} &= \sqrt{\frac{\mu}{a^3}} \left(\frac{R}{a} \right)^n \frac{\sqrt{1-e^2}}{e} F_{n,m,p}(i) G_{n,p,q}(e) \left[\sqrt{1-e^2} (n-2p+q) - (n-2p) \right] \frac{Z_{n,m,p,q}}{\dot{\Theta}_{n,m,p,q}} \\
\Delta i_{n,m,p,q} &= \sqrt{\frac{\mu}{a^3}} \left(\frac{R}{a} \right)^n \frac{1}{\sqrt{1-e^2} \sin i} F_{n,m,p}(i) G_{n,p,q}(e) [(n-2p) \cos i - m] \frac{Z_{n,m,p,q}}{\dot{\Theta}_{n,m,p,q}} \\
\Delta \omega_{n,m,p,q} &= \sqrt{\frac{\mu}{a^3}} \left(\frac{R}{a} \right)^n \left[\frac{\sqrt{1-e^2}}{e} F_{n,m,p}(i) \frac{\partial G_{n,p,q}(e)}{\partial e} - \frac{\cot i}{\sqrt{1-e^2}} \frac{\partial F_{n,m,p}(i)}{\partial i} G_{n,p,q}(e) \right] \frac{Z_{n,m,p,q}^*}{\dot{\Theta}_{n,m,p,q}} \\
\Delta \Omega_{n,m,p,q} &= \sqrt{\frac{\mu}{a^3}} \left(\frac{R}{a} \right)^n \frac{1}{\sqrt{1-e^2} \sin i} \frac{\partial F_{n,m,p}(i)}{\partial i} G_{n,p,q}(e) \frac{Z_{n,m,p,q}^*}{\dot{\Theta}_{n,m,p,q}} \\
\Delta M_{n,m,p,q} &= \sqrt{\frac{\mu}{a^3}} \left(\frac{R}{a} \right)^n F_{n,m,p}(i) \cdot \\
&\quad \cdot \left[2(n+1) G_{n,p,q}(e) - \frac{1-e^2}{e} \frac{\partial G_{n,p,q}(e)}{\partial e} - 3(n-2p+q) G_{n,p,q}(e) \frac{\sqrt{\mu/a^3}}{\dot{\Theta}_{n,m,p,q}} \right] \frac{Z_{n,m,p,q}^*}{\dot{\Theta}_{n,m,p,q}}
\end{aligned} \tag{23.46}$$

where α is a generic orbital element, $Z_{n,m,p,q}^*$ is the integral of $Z_{n,m,p,q}$ with respect to its argument:

$$Z_{n,m,p,q}^* = \int Z_{n,m,p,q} d\Theta_{n,m,p,q} = X_{n,m} \sin \Theta_{n,m,p,q} - Y_{n,m} \cos \Theta_{n,m,p,q} \tag{23.47}$$

and

$$\dot{\Theta}_{n,m,p,q} = (n-2p)\dot{\omega} + (n-2p+q)\dot{M} + m(\dot{\Omega} - \dot{\theta}) \tag{23.47}$$

where $\dot{\theta}$ is the Earth's mean (sidereal) rotation rate. As a result of the assumptions made, (23.46) describes, for each combination $\{n,m,p,q\}$, the orbit perturbations relative to a slowly rotating ellipse. This rotating reference ellipse corresponds to the evolution of the secularly changing, or mean-mean, orbital elements introduced in Section 23.3. Therefore, the orbital elements on the right-hand sides of (23.46) are mean-mean elements, and the addition of all terms $\Delta\alpha_{n,m,p,q}$ to the mean-mean elements yields the osculating elements. From (23.46) we conclude that the change of an orbital element due to a particular term of the perturbing force function, so a specific combination $\{n,m,p,q\}$, is a function of: 1) the orbital elements a , e , and i ; 2) the phase angle $\Theta_{n,m,p,q}$, which is a function of the orbital elements ω , Ω , M and time; 3) the frequency $\dot{\Theta}_{n,m,p,q}$, which is a function of $\dot{\theta}$ and of the (constant) secular rates of ω , Ω and M ; 4) the coefficients $C_{n,m}$ and $S_{n,m}$, and the constants μ and R , which specify the gravity field. The period of an orbit perturbation produced by a particular combination $\{n,m,p,q\}$ is given by $P_{n,m,p,q} = 2\pi/|\dot{\Theta}_{n,m,p,q}|$.

Equation (23.47) shows that the frequencies of the perturbations of a given orbit are also dependent on the values of $n-2p$, $n-2p+q$, and m . Thus, for any m there are many combinations of n , p and q that produce orbit perturbations with the same frequency $\dot{\Theta}$. For example, the terms $\{3,2,1,0\}$ and $\{11,2,5,0\}$ both yield the frequency $\dot{\Theta} = (\dot{\omega} + \dot{M}) + 2(\dot{\Omega} - \dot{\theta})$. This means that the gravity field model coefficient pairs $C_{3,2}, S_{3,2}$ and $C_{11,2}, S_{11,2}$ create orbit perturbations that overlap in frequency. In general, there is a complete family of combinations $\{n,m,p,q\}$ that produce the same frequency. The gravity field model coefficients that fall under this family are called *lumped*

coefficients. The consequence is that when we use tracking data of a satellite to determine a certain frequency in the satellite orbit perturbations, it is not directly possible to pinpoint which gravity field model coefficient is responsible for that perturbation frequency. We can only figure out which set of lumped coefficients produce that frequency. By using tracking data from many satellites in different orbits, we can combine various sets of lumped coefficients, separate the effects of individual coefficients and determine the value of those coefficients.

Note that in the expression for $\Delta e_{n,m,p,q}$ the term in brackets reduces to $\approx q$ for nearly circular orbits, and that for near-polar orbits the term in brackets in the expression for $\Delta i_{n,m,p,q}$ reduces to $\approx m$. The latter result implies that zonal harmonics ($m = 0$) do not produce inclination perturbations for polar orbits; this result is in agreement with (21.24) and (23.9). The expressions for $\Delta \alpha_{n,m,p,q}$, except for $\alpha = a$, contain terms with the eccentricity and/or the term $\sin i$ in the denominator, and we therefore may expect that these terms lead to very large perturbations for low-eccentricity or low-inclination orbits. However, in the expression for $\Delta e_{n,m,p,q}$ the term $G_{n,p,q}(e)$ constitutes for $|q| = 1$ a power series of e that compensates for the e in the denominator. If $q = 0$ the term in square brackets produces a power series of e that compensates for the e in the denominator. When we consider only zonal harmonics, it turns out that the $\sin i$ term in the denominator of the expression for $\Delta i_{n,m,p,q}$ is mostly compensated for by $\sin i$ terms in the series $F_{n,m,p}(i)$, and that for those values of n and p where this is not the case the term in square brackets is zero, leading to $\Delta i_{n,m,p,q} = 0$. So, the problem of having e and/or $\sin i$ in the denominator may occur only in the expressions for $\Delta \omega_{n,m,p,q}$, $\Delta \Omega_{n,m,p,q}$, $\Delta M_{n,m,p,q}$, and for tesseral harmonics also in the expression for $\Delta i_{n,m,p,q}$. It is emphasized that this kind of problems is a direct result of the fact that perigee is not defined for circular orbits and that the ascending node is not defined for zero-inclination orbits, and thus that the orbital elements ω , Ω , M are not appropriate to describe the position of satellites in low-eccentricity or low-inclination perturbed orbits (Section 22.4). Of course, Kaula's linear perturbations theory is not valid anymore when the orbit perturbations become unrealistically large.

From (23.46) we obtain, neglecting terms of order e^2 relative to one, the following expressions:

$$\begin{aligned} \left(\frac{1}{2a} \Delta a_{n,m,p,q} - e \Delta e_{n,m,p,q} \right) \cos i - \sin i \Delta i_{n,m,p,q} &= \sqrt{\frac{\mu}{a^3}} \left(\frac{R}{a} \right)^n m F_{n,m,p}(i) G_{n,p,q}(e) \frac{Z_{n,m,p,q}}{\dot{\Theta}_{n,m,p,q}} \\ \Delta \omega_{n,m,p,q} + \cos i \Delta \Omega_{n,m,p,q} + \Delta M_{n,m,p,q} &= \sqrt{\frac{\mu}{a^3}} \left(\frac{R}{a} \right)^n F_{n,m,p}(i) \cdot \\ &\cdot \left[\left(2(n+1) - 3(n-2p+q) \frac{\sqrt{\mu/a^3}}{\dot{\Theta}_{n,m,p,q}} \right) G_{n,p,q}(e) + \frac{1}{2} e \frac{\partial G_{n,p,q}(e)}{\partial e} \right] \frac{Z_{n,m,p,q}^*}{\dot{\Theta}_{n,m,p,q}} \end{aligned}$$

These expressions can be used to partially check the correctness of the expressions for $\Delta \alpha_{n,m,p,q}$ derived from (23.46) for near-circular orbits and for a particular combination $\{n,m,p,q\}$. Note that the right-hand sides of these expressions do not contain e and/or $\sin i$ in the denominator anymore. Further note that the right-hand side of the first expression is zero if $m = 0$, and that the expression can then be written as

$$\cos^2 i \Delta a_{n,m,p,q} - 2a e \cos^2 i \Delta e_{n,m,p,q} - a \sin 2i \Delta i_{n,m,p,q} = 0$$

Remembering that the component of the orbital angular momentum along the Z-axis, H_z , is given by $H_z^2 = \mu a(1-e^2) \cos^2 i$, the above expression can be written as $(\Delta H_z^2)_{n,m,p,q} = 0$, when we

neglect terms of order e^2 . This result fully agrees with our finding in Section 23.1 that for an axially-symmetric gravity field, which requires $m = 0$, H_z is constant. For $q = 0$: $G_{n,p,q}(e) = 1 + O(e^2)$, $e \partial G_{n,p,q}(e) / \partial e = O(e^2)$. In that case, we may substitute $G_{n,p,q}(e) = 1$ and may neglect the term $e \partial G_{n,p,q}(e) / \partial e$ on the right-hand side of the second expression.

For the orbit perturbations analysis to remain ‘manageable’, it is important that the perturbing influence of successive terms of the perturbing force function, \tilde{R} , converges with increasing degree n . This convergence is ensured by a combination of characteristics of both the perturbing force function and the perturbations equations (23.46). First, the factor $(R/a)^n$ decreases with increasing n , albeit slowly for low-altitude satellites. Secondly, the gravity field coefficients themselves have the tendency to decrease slowly with increasing n according to an empirically observed rule, known as *Kaula's rule*:

$$O\{C_{n,m}, S_{n,m}\} \approx 0.85 \frac{10^{-5}}{n^2}$$

A third contribution to the convergence results from the frequency $\dot{\Theta}_{n,m,p,q}$ in the denominator of the orbit perturbations equations (23.46). For large values of n and m many high-frequency perturbations occur (see below) and, generally, the higher the frequency the smaller the amplitude of a perturbation. So, many terms may be neglected in practice. In this respect, the satellite may be viewed as a ‘low pass filter’ that makes the orbit only sensitive to perturbing components with longer, and in particular very long, periods.

From (23.46) we conclude that, for a particular orbit, the smaller the value of $\dot{\Theta}_{n,m,p,q}$ the larger the associated perturbation components in the orbital elements. This first-order analysis breaks down when $\dot{\Theta}_{n,m,p,q}$ approaches or equals zero. This occurs: 1) for the secular perturbations due to zonal harmonics (see below); 2) for the long-period perturbations due to zonal harmonics (see below) when the orbital inclination approaches the critical inclination (Section 23.3) or when the satellite moves in a so-called *frozen orbit* (Section 23.7); 3) when the satellite's orbit is in resonance with the Earth's gravity field. In these cases, long-period, large-amplitude perturbations occur in elements other than the angular variables what makes the integration of the Lagrange planetary equations more complex. The case of orbital resonance will be discussed below separately.

Since the parameter $\dot{\Theta}_{n,m,p,q}$ controls the frequency, and thus the type and even the magnitude, of the perturbation, it is interesting to study the characteristics of (23.47). From (23.27), (23.35), and (11.3-2), we find for a satellite in a nearly circular orbit at an altitude of less than 5000 km:

$$|\dot{\Omega}| = 0 - 8^\circ/\text{day} ; |\dot{\omega}| = 0 - 17^\circ/\text{day} ; \dot{M} = 2600 - 5700^\circ/\text{day} ; \dot{\theta} = 361^\circ/\text{day}$$

depending on the altitude and inclination of the orbit. So, $\dot{M} \gg \dot{\theta} \gg |\dot{\omega}| \approx |\dot{\Omega}|$. This clear distinction in angular rates makes it possible to identify various classes of perturbations, such as short-period variations, long-period variations and orbital resonance. In general, the smaller the rate of change of a particular variable, the larger will be the amplitude of the perturbation that it produces. So, long-period variations, which originate from $\dot{\omega}$, will, generally, have larger amplitudes than short-period variations, which originate from \dot{M} . Resonance, and in particular deep resonance, produces even larger effects than the long-period terms do. A strict separation in different classes of perturbation frequencies is, of course, a simplification of the true situation. Each type of periodic variation will occur in frequency bands. For low degrees and orders, the three frequency bands associated with \dot{M} , $\dot{\theta}$, and $\dot{\omega}$ are essentially disjoint. Band overlapping starts to occur at values of m approximately equal to the number of orbital revolutions per day. Although this complicates the appearance of the spectrum, it has, of course, no effect on the

amplitudes of the individual perturbations. A useful categorization of the various classes of perturbations on basis of the values of individual terms in (23.47) is given by D. A. Vallado⁵. The following description is partly based on his analysis.

Zonal harmonics ($m = 0$)

Secular

$$(n - 2p) = 0 ; (n - 2p + q) = 0 \Rightarrow p = n/2 ; q = 0 ; \dot{\Theta}_{n,m,p,q} = 0$$

In Section 23.3 it was shown that the elements a , e , and i do not show a secular variation. This is confirmed by the presence of the terms $n - 2p = 0$ and $n - 2p + q = 0$ in (23.46-2) to (23.46-4). Since $p = n/2$ and both p and n are integers, n must be even and we conclude that the secular variations in Ω , ω , and M result from even zonal harmonics. Note that Kaula's method breaks down for secular perturbations, because of the occurrence of $\dot{\Theta}_{n,m,p,q} = 0$ in the denominator of (23.46). However, this is not a problem, because we already know the (first-order) secular variation of the orbital elements ω , Ω and M ((23.35)) and the availability of these expressions, in fact, forms an essential element of Kaula's theory.

Long-period

$$(n - 2p) \neq 0 ; (n - 2p + q) = 0 \Rightarrow q = 2p - n \neq 0 ; \dot{\Theta}_{n,m,p,q} = -q\dot{\omega}$$

From these conditions we conclude that odd zonal harmonics produce the major long-period variations with $|q| = 1$. Even zonal harmonics contribute smaller long-period variations with $|q| \geq 2$. The long-period variations are driven by the apsidal rotation rate and the fundamental period of the long-period variations is $2\pi/|\dot{\omega}|$, which corresponds for a satellite below 5000 km altitude to 20–1600 days, depending on the orbital altitude and inclination, if the inclination stays more than 5° from the critical inclination and if the satellite's orbit is non-frozen. The variation of the mean eccentricity of the ANS orbit shown in Figure 23.6 is a clear example of a long-period variation due to odd zonal harmonics. Close to the critical inclination and for frozen orbits $\dot{\omega}$ goes to zero and, consequently, the period goes to infinity; then the first-order expressions (23.46) are not valid anymore. Harmonics of the fundamental period exist at $1/2, 1/3, \dots$ of this period for $|q| = 2, 3, \dots$. These harmonics have arguments $2\omega, 3\omega, \dots$; their amplitudes are smaller than the amplitude for $|q| = 1$. The semi-major axis does not exhibit a long-period variation, since the expression for $\Delta a_{n,m,p,q}$ in (23.46) contains the term $(n - 2p + q) = 0$ as a multiplicative factor.

Short-period

$$(n - 2p) = 0 ; (n - 2p + q) \neq 0 \Rightarrow p = n/2 ; q \neq 0 ; \dot{\Theta}_{n,m,p,q} = q\dot{M}$$

The expressions for this class of perturbations contain sine- and cosine-terms with argument qM and are therefore 'pure' short-period variations. Since $p = n/2$, and n and p are integers, these variations are caused only by even zonal harmonics. The main term ($|q| = 1$) leads to a fundamental period of the perturbations equal to the anomalistic orbital period: $2\pi/\dot{M}$; additional harmonics with periods of $2\pi/(|q|\dot{M})$, $|q| = 2, 3, \dots$ exist. From (23.46) we conclude that the inclination does not show a pure short-period variation, since the expression for $\Delta i_{n,m,p,q}$ contains the term $(n - 2p) = 0$ as a multiplicative factor.

⁵ D.A. Vallado, *Fundamentals of Astrodynamics and Applications*, Microcosm Press and Kluwer Academic Publishers, 2001.

Long-period/short-period

$$(n - 2p) \neq 0 ; (n - 2p + q) \neq 0 \Rightarrow \dot{\Theta}_{n,m,p,q} = (n - 2p)\dot{\omega} + (n - 2p + q)\dot{M}$$

This component of the motion is caused by the interaction of short-period and long-period variations, and is produced by all zonal harmonics. The expressions for this class of perturbations contain sine- and cosine-terms with the argument $(n - 2p)\omega + (n - 2p + q)M$. Since $|\dot{\omega}| \ll \dot{M}$, the period of this class of perturbations may be approximated by $2\pi/|(n - 2p + q)|\dot{M}$ or $1/|(n - 2p + q)|$ times the anomalistic orbital period. This leads to a variety of perturbation periods. Since short-period variations are only produced by even zonal harmonics, the contributions of the odd zonal harmonics to the variation of the semi-major axis of the ANS orbit that are plotted in Figure 23.5 do not belong to the class of 'pure' short-period perturbations, but to the class of long-period/short-period perturbations. Similarly, as the inclination does not show a pure short-period variation, the variation of the inclinations of the ANS and TD-1A orbits that are plotted in Figure 23.3 and 23.4 also belong to the class of long-period/short-period perturbations. However, we should realize that the separation of orbit perturbations into a class of 'pure' short-period variations and a class of long-period/short-period variations is somewhat artificial. This separation was introduced in this Section; in the foregoing Sections we called both classes of variations simply 'short-period variations'.

Tesseral and sectorial harmonics ($m \neq 0$)

m -daily

$$(n - 2p) = 0 ; (n - 2p + q) = 0 \Rightarrow p = n/2 ; q = 0 ; \dot{\Theta}_{n,m,p,q} = m(\dot{\Omega} - \dot{\theta})$$

This component of the motion is caused by the interaction of the rotation of the Earth's gravity field and the secular variation of the orbit's right ascension of the ascending node. Since both n and p are integers, only even values of n apply, with all values of $m \leq n$. Since $|\dot{\Omega}| \ll \dot{\theta}$, the period of this class of perturbations may be approximated by $2\pi/m\dot{\theta} = 1/m$ day. So, all terms of even degree n and of order $m = 1$ contribute oscillations with a period of about 24 hr, whereas all $m = 2$ terms contribute 12 hr oscillations, and so on. Higher-order harmonics (e.g. $m = 14$) contribute oscillations with periods roughly the same as of the basic zonal harmonics short-period variations for low-altitude satellites. Thus, there is little distinction between the short-period variations from the zonal harmonics and the m -daily variations from higher-order tesseral and sectorial terms. From (23.46) we conclude that the semi-major axis and eccentricity do not show m -daily perturbations.

Long-period/ m -daily

$$(n - 2p) \neq 0 ; (n - 2p + q) = 0 \Rightarrow q = 2p - n \neq 0 ; \dot{\Theta}_{n,m,p,q} = m(\dot{\Omega} - \dot{\theta}) - q\dot{\omega}$$

This class of perturbations is produced by all terms of even and odd degree n , with all values of $m \leq n$. For low-eccentricity orbits the major perturbations are associated with $|q| = 1$ and $p = (n \pm 1)/2$. Since n and p are integers, these major perturbations are produced by terms with odd values of n . From (23.46) we conclude that the semi-major axis does not show long-period/ m -daily perturbations. Since $\dot{\theta} \gg |\dot{\Omega}| \approx |\dot{\omega}|$, the period of the major perturbations is approximately given by $2\pi/m\dot{\theta} = 1/m$ day. The source of this component of the motion is the interaction between the secular variation of the argument of perigee, which leads to long-period variations, the rotation of the gravity field, and the secular variation of the right ascension of the ascending node.

m-daily/short-period

$$(n - 2p) = 0 ; (n - 2p + q) \neq 0 \Rightarrow p = n/2 ; q \neq 0 ; \dot{\Theta}_{n,m,p,q} = q \dot{M} + m(\dot{\Omega} - \dot{\theta})$$

Since both n and p are integers, only even values of n apply, with all values of $m \leq n$. Since $\dot{\theta} \gg |\dot{\Omega}|$, the period of the perturbations is approximately given by $2\pi/|(m\dot{\theta} - q\dot{M})|$. For low-eccentricity orbits, only terms with $0 < |q| \leq 2$ are significant. If m is small the period is approximately given by $2\pi/(|q|\dot{M}) = T/|q|$, where T is the anomalistic orbital period. If m is large ($> 7 - 15$) then $m\dot{\theta} - q\dot{M}$ may become (almost) zero and the perturbation period may become very large, which corresponds to resonance. For very large values of m (> 50) the period is about $1/m$ day.

Long-period/m-daily/short-period

$$(n - 2p) \neq 0 ; (n - 2p + q) \neq 0 ; \dot{\Theta}_{n,m,p,q} = (n - 2p)\dot{\omega} + (n - 2p + q)\dot{M} + m(\dot{\Omega} - \dot{\theta})$$

For $q = 0$, the period of the perturbations is approximately given by $2\pi/|(m\dot{\theta} - (n-2p)\dot{M})|$. If $m\dot{\theta} \gg |(n-2p)\dot{M}|$ then the period is about $1/m$ day; if $m\dot{\theta} \ll |(n-2p)\dot{M}|$ then the period is about $T/|(n-2p)|$; if $m\dot{\theta} \approx (n-2p)\dot{M}$ then resonance will occur and the period becomes very large. In other cases, all kinds of periods may occur. The long-period/m-daily/short-period terms are the most common (numerous) ones in (23.46). They represent the interaction between all fundamental periods in (23.47).

Resonance

When $\dot{\Theta}_{n,m,p,q}$ approaches zero and, consequently, $P_{n,m,p,q}$ becomes very large, the satellite's orbit is in resonance with the gravity field and large orbit perturbations will occur. Theoretically, certain combinations $\{n,m,p,q\}$ and orbital elements may even yield $\dot{\Theta}_{n,m,p,q} = 0$ and $P_{n,m,p,q} = \infty$, which corresponds to *exact resonance*. However, in practice this situation will never occur, because it requires very specific values of the orbital elements a , e , and i ; any deviation from these values will result in *deep resonance* or *shallow resonance*. Deep resonance corresponds to very small values of $|\dot{\Theta}_{n,m,p,q}|$, which cause large-amplitude oscillations with very long periods; shallow resonance corresponds to somewhat larger values of $|\dot{\Theta}_{n,m,p,q}|$, which lead to oscillations with smaller amplitudes and shorter periods. When resonance occurs, the first-order Kaula theory for the orbit perturbations is not valid anymore, but we still can use (23.47) to determine the occurrence of resonance perturbations. In this book, we define deep resonance as resonance for which $|\dot{\Theta}_{n,m,p,q}| < 4*10^{-7}$ rad/s, corresponding to resonance periods of longer than 6 months; we define shallow resonance as resonance for which $4*10^{-7}$ rad/s $< |\dot{\Theta}_{n,m,p,q}| < 5*10^{-6}$ rad/s, corresponding to resonance periods of 15 days to 6 months.

According to (23.47), the condition for orbital resonance is

$$(n - 2p)\dot{\omega} + (n - 2p + q)\dot{M} + m(\dot{\Omega} - \dot{\theta}) \approx 0 \quad (23.48)$$

Now, $\dot{\theta}$ is known, and the first-order expressions for \dot{M} , $\dot{\Omega}$ and $\dot{\omega}$ are functions of the orbital elements a , e , and i . For a given orbit, we can evaluate (23.47) numerically in order to determine whether there are combinations $\{n,m,p,q\}$ that lead to orbital resonance. As stated before, for nearly circular orbits the largest perturbations occur for $|q| = 0, 1$. From (23.48) we can identify a number of special classes of orbital resonance; the two most important classes will be discussed below.

Repeat ground track resonance

$$m \neq 0 ; n - 2p \neq 0 ; q = 0$$

For this class, we find from (23.48)

$$\frac{\dot{\omega} + \dot{M}}{\dot{\theta} - \dot{\Omega}} \approx \frac{m}{n - 2p} \quad (23.49)$$

Because $\dot{M} \gg |\dot{\omega}|$, $\dot{M} > 0$ and $\dot{\theta} \gg |\dot{\Omega}|$, $\dot{\theta} > 0$, and $m > 0 : p < n/2$. The left-hand side of this expression is the ratio of the nodal day ((23.40)) and the nodal or draconic (draconitic) orbital period ((23.38)); the right-hand side is the ratio of two integers. So, this kind of resonance occurs when the satellite completes α nodal revolutions in β nodal days, where α and β are positive mutually prime integers; i.e. the ratio α/β must be irreducible. The source of the resonance are those terms in the Kaula perturbing force function of which the values of $\{n,m,p\}$ satisfy the condition $m = j\alpha$, $n - 2p = j\beta$, $j = 1, 2, 3, \dots$. This class of resonance arises when successive ground tracks⁶ (Section 23.7) of the satellite are separated by an interval in longitude equal to the wavelength of the geopotential harmonics, so that after m (or $2m, 3m, \dots$) orbital revolutions of the satellite the ground track sequence repeats itself exactly. Therefore, this class is called *repeat ground track resonance*. In this orbit, the satellite will cross about the same position in the (rotating) gravity field after β nodal days, and the satellite's trajectory through the gravity field (almost) repeats every β nodal days. Consequently, the satellite's motion is perturbed in an identical manner every β nodal days, enhancing the earlier perturbation. The amplitude of the orbital resonance perturbations will, generally, be larger if β is smaller; major perturbations in the orbital elements may be expected if $\beta < 5$.

A special case is a geostationary satellite, for which $e = 0$, $\dot{M} = \dot{\theta}$, $\alpha = \beta = 1$. For this orbit resonance occurs primarily when $m = n - 2p$, and the main resonance term is $\{n,m,p,q\} = 2,2,0,0$. Further resonances occur with all coefficients for which $n - m$ is even (e.g. $\{n,m,p,q\} = 3,1,1,0$; $4,2,1,0$; $6,2,2,0$; etc.), but because of the attenuation factor $(R/a)^n$ in the perturbing force function, the higher-degree resonances will be of lesser importance. For satellites in lower altitude orbits resonance occurs at higher degrees. For example, for orbital altitudes below 5000 km, appreciable resonance will occur if the satellite completes 8-15 orbital revolutions in 1 day, 15-29 revolutions in 2 days, or 22-46 revolutions in 3 days, where the larger the number of days the smaller the resonance perturbation amplitude. This implies that, for this range of orbital altitudes, appreciable resonance effects require $m = 8-46$, $n - 2p = 1-3$. For a satellite orbiting the Earth at e.g. about 1260 km altitude, such that $\dot{M} \approx 13$ rev/day, $\alpha = 13$, $\beta = 1$ and major resonances occur when $\{n,m,p,q\} = 13,13,6,0$; $15,13,7,0$; $17,13,8,0$;

Anomalistic resonance

$$m \neq 0 \quad ; \quad n - 2p = 0 \quad ; \quad q \neq 0$$

Since $p = n/2$ and both p and n are integers, this class of resonance is produced by terms with even n and all $m \leq n$. From (23.48) we find

$$\frac{\dot{M}}{\dot{\theta} - \dot{\Omega}} \approx \frac{m}{q}$$

The left-hand side of this expression is the ratio of the nodal day ((23.40)) and the anomalistic orbital period ((23.36)); the right-hand side is the ratio of two integers. This class is called *anomalistic resonance* and occurs if the satellite performs (about) α anomalistic revolutions in

⁶ Sub-satellite point: the point on the Earth's surface directly between the satellite and the center of the Earth.
Ground track: the path of the sub-satellite point on the Earth's surface.

β nodal days. Those values of $\{m, q\}$ that satisfy the condition $m = j\alpha, q = j\beta, j = 1, 2, 3, \dots$ identify the terms of the Kaula perturbing force function which are the source of this kind of resonance. Because $q \neq 0, \dot{M} > 0, \dot{\Theta} \gg |\dot{\Omega}|, \dot{\Theta} > 0$, and $m > 0$, only terms with $q = +1, +2$ produce appreciable effects for nearly circular orbits. So, significant anomalistic resonance occurs when the satellite performs α anomalistic revolutions per day or per 12 hr. However, these resonance amplitudes will, generally, be smaller than those for repeat ground track resonance, since no terms with $q = 0$ are present. Because the ratio $\dot{M}/(\dot{\Theta} - \dot{\Omega})$ varies from 7 to 16 for altitudes of up to 5000 km, but is as low as 1 or 2 for geosynchronous or 12 hr orbits, the lowest resonance order may vary from $m = 1$ to $m = 16$, depending on the satellite orbit.

In the discussion above we have found that significant cases of resonance occur if the ratio of \dot{M} , or a linear combination of \dot{M} and $\dot{\omega}$, and a linear combination of $\dot{\Theta}$ and $\dot{\Omega}$ (about) equals the ratio of two integers. It is emphasized that the ratio of these linear combinations can always be written as the ratio of two integers, because the computed rates of the parameters involved have a finite decimal representation. However, this usually results in the ratio of two large integers, which does not identify a resonance contribution of real significance since it will be associated with large values of n and m . The high-degree and high-order harmonics have smaller gravity field coefficients, making their contributions less significant. In addition, the factor $(R/a)^n$ in (23.46) will significantly attenuate the amplitude of the orbit perturbation.

To get an impression of the various classes of perturbations a numerical analysis was performed for one of the orbits flown by the ESA ERS-1 satellite that was launched by an Ariane-4 rocket from Kourou on July 17, 1991. This satellite has flown a number of ground track repeat orbits (Section 23.7); for our example we consider the orbit in which ERS-1 completed 43 nodal revolutions in 3 nodal days ($\alpha/\beta = 43/3$). In Section 23.7 the corresponding values of semi-major axis, eccentricity and inclination are given: $a = 7153.14$ km, $e = 0.00106$, $i = 98.523^\circ$. For this numerical experiment, the maximum value of the degree, n , in the gravity field series expansion was set at 50. All perturbations for $|q| = 0, 1$ were analyzed. For the Earth's gravitational parameter, mean equatorial Earth radius and J_2 -coefficient of the Earth's gravity field, the following values were adopted: $\mu = 398600.44$ km 3 /s 2 , $R = 6378.136$ km, $J_2 = 1082.636 * 10^{-6}$. The secular rates of ω , Ω , and M were computed from (23.27) and (23.35). The results are listed in Table 23.1 in the order of decreasing periods of the perturbation component, and for periods longer than 12 hr. To limit the number of cases, for a particular perturbation period only the results for the term with the lowest value of $|q|$ and, for that value of q , the lowest values of n and m are listed. Note that repeat ground track deep resonance leads to orbit perturbations with a period of 5991 days (16.4 yr). The other classes of orbit perturbations have periods within bands around 120, 3, 1.5, 1.0, 0.75, 0.60, and 0.50 days, with various classes of perturbations overlapping per band.

As a second example, a numerical search process was executed to find the value of the semi-major axis of the 43/3 ERS-1 orbit for which exact resonance would have occurred. The semi-major axis was varied from 7148 km to 7158 km, in steps of 0.1 m, the eccentricity and inclination were kept at their nominal values, and the computation was performed for $|q| = 0, 1$. Three cases of exact resonance were found in this semi-major axis band; Table 23.2 lists the results. The case with $q = 0$ corresponds to exact repeat ground track resonance, while the two cases with $|q| = 1$ correspond to side-bands of exact resonance; the case with $q = 0$ will lead to the largest orbit perturbations. From this Table it may be concluded that the value of $a = 7153.14$ km listed above is close to the value for which exact repeat ground track resonance would have occurred: $a = 7153.1584$ km. It was found that deep resonance will occur if the semi-major axis

satisfies the condition $7151.6 \text{ km} < a < 7154.7 \text{ km}$.

Table 23.1: Examples of ERS-1 orbit perturbations with periods longer than 12 hr and $|q| = 0, 1$. Results are listed for the 43/3 ground track repeat orbit and in the order of decreasing periods. Per perturbation period only the results for the lowest value of $|q|$, and, for that value of q , the lowest values of n and m , are listed.

$P_{n,m,p,q}$ (day)	n	m	p	q	Type
∞	2	0	1	0	secular
5991	43	43	20	0	repeat ground track resonance
123.8	44	43	20	-1	long-period/ m -daily/short-period
121.3	3	0	1	-1	long-period
118.9	44	43	21	1	long-period/ m -daily/short-period
3.08	14	14	6	-1	long-period/ m -daily/short-period
3	15	14	7	0	long-period/ m -daily/short-period
2.93	14	14	7	1	m -daily/short-period
1.52	16	15	8	1	m -daily/short-period
1.5	15	15	7	0	long-period/ m -daily/short-period
1.48	16	15	7	-1	long-period/ m -daily/short-period
1.01	3	1	2	1	long-period/ m -daily
1	2	1	1	0	m -daily
0.99	3	1	1	-1	long-period/ m -daily
0.75	13	13	6	0	long-period/ m -daily/short-period
0.74	14	13	7	1	m -daily/short-period
0.6	17	16	8	0	long-period/ m -daily/short-period
0.5	2	2	1	0	m -daily

Table 23.2: Three cases of exact orbital resonance for the 43/3 ERS-1 ground track repeat orbit with adjusted values of the semi-major axis, and for $|q| = 0, 1$.

a (km)	n	m	p	q	$n - 2p$	$n - 2p + q$
7152.2486	44,46,48,50	43	20,21,22,23	-1	4	3
7153.1584	43,45,47,49	43	20,21,22,23	0	3	3
7154.0678	44,46,48,50	43	21,22,23,24	1	2	3

For most satellites the resonance orbit perturbations will not fully build up. Satellites with a perigee altitude of less than 600 km experience substantial atmospheric drag, which leads to a steadily decreasing semi-major axis. Consequently, the satellite passes through the narrow altitude region for which a case of (deep) resonance occurs, before the resonance perturbations in the orbital elements fully develop. Satellites in higher orbits may experience significant lunisolar gravitational attraction perturbations, which also drive the orbital parameters outside the range of resonance. Satellites for geodetic and oceanographic research are sometimes deliberately kept in an orbit of which the altitude and inclination maintain specified values (Section 23.7). These orbits may be in resonance with the rotating Earth's gravity field, but the periodic firings of the satellite's thrusters to keep the semi-major axis and inclination at their nominal values destroy the development of resonance perturbations in the orbital elements.

23.6. Application of Kaula's theory

To illustrate the characteristics of some important classes of orbit perturbations, we will apply Kaula's theory to derive first-order linearized analytical expressions for the effects of some selected terms of the gravity field series expansion on the orbital elements of a satellite in a near-circular orbit. We will only consider terms with $|q| = 0, 1$ and will assume again that the orbital altitude is less than 5000 km, which means that we may use the condition $\dot{M} \gg \dot{\Theta} \gg |\dot{\omega}| \approx |\dot{\Omega}|$. In the analytical developments we will neglect terms of order e^2 and of order J_2 relative to one. Numerical results will be presented for a satellite in an orbit with the following (mean-mean) orbital parameters: $h = 500$ km, $e = 0.005$, $i = \omega = \Omega = 45^\circ$, $M = 30^\circ$.

J_2 short-period variations

Since the conditions for this class of orbital element variations are $n = 2$, $m = 0$, $n - 2p = 0$, $n - 2p + q \neq 0$, only two terms $\{2,0,p,q\}$ contribute: $\{2,0,1,-1\}$ and $\{2,0,1,1\}$. When we substitute the relevant values of n , m , p , q into (23.44), (23.45) and (23.47), subsequently substitute $-J_2 = C_{2,0}$ ((20.1-2)), and use (23.27) and (23.35), we obtain

$$\Theta_{2,0,1,1} = -\Theta_{2,0,1,-1} = M \quad ; \quad \dot{\Theta}_{2,0,1,1} = -\dot{\Theta}_{2,0,1,-1} = \dot{M} = \sqrt{\frac{\mu}{a^3}} (1 + O(J_2))$$

$$Z_{2,0,1,1} = Z_{2,0,1,-1} = -J_2 \cos M \quad ; \quad Z_{2,0,1,1}^* = -Z_{2,0,1,-1}^* = -J_2 \sin M$$

where the notation O stands for ‘order of’. For both terms the following expressions hold⁷:

$$F_{2,0,1}(i) = \frac{3}{4} \sin^2 i - \frac{1}{2} \quad ; \quad G_{2,1,1}(e) = G_{2,1,-1}(e) = \frac{3}{2} e + O(e^3)$$

Inspection of (23.46) shows that, when applying the relevant expressions for the terms $\{2,0,1,1\}$ and $\{2,0,1,-1\}$, for a generic orbital element α : $\Delta\alpha_{2,0,1,1} = \Delta\alpha_{2,0,1,-1}$. When we evaluate the six equations (23.46) for the term $\{2,0,1,1\}$, linearize the resulting expressions by neglecting terms of order e^2 or of order J_2 relative to one, and multiply the then resulting expressions by a factor two, we obtain

$$\begin{aligned} \Delta a_{2,0} &= \frac{3}{2} J_2 \frac{R^2}{a} e (3 \cos^2 i - 1) \cos M \\ \Delta e_{2,0} &= \frac{3}{4} J_2 \left(\frac{R}{a} \right)^2 (3 \cos^2 i - 1) \cos M \\ \Delta i_{2,0} &= 0 \\ \Delta \omega_{2,0} &= \frac{3}{4} J_2 \left(\frac{R}{a} \right)^2 \frac{1}{e} (3 \cos^2 i - 1) \sin M \\ \Delta \Omega_{2,0} &= -\frac{9}{2} J_2 \left(\frac{R}{a} \right)^2 e \cos i \sin M \end{aligned} \tag{23.50}$$

⁷ The expressions for $F_{n,m,p}(i)$ and $G_{n,p,q}(e)$ were taken from: W.M. Kaula, *Theory of Satellite Geodesy*, Dover Publications, Mineola, New York, 2000.

$$\Delta M_{2,0} = -\frac{3}{4} J_2 \left(\frac{R}{a} \right)^2 \frac{1}{e} (3 \cos^2 i - 1) \sin M \quad (23.50)$$

The orbital elements experience an oscillation with a period equal to the anomalistic orbital period. Variations in a , e , ω and M are zero for $i = 54.7^\circ, 125.3^\circ$; these values correspond to the situation where the anomalistic orbital period is equal to the Keplerian period (Section 23.3). The variation in Ω is zero for $i = 90^\circ$. The amplitudes of the variations are proportional to the term $J_2(R/a)^2$, except for the semi-major axis for which the amplitude is proportional to the term $J_2 R^2/a$. The amplitudes of the variation of a and Ω are also proportional to e , which means that these amplitudes are very small for near-circular orbits; the amplitudes of the variation of ω and M are proportional to $1/e$, indicating that these variations become very large for $e \downarrow 0$. However, then the expressions lose their validity since in a first-order theory the values of $\Delta\alpha$ should not become unrealistically large. This phenomenon is a direct result of the fact that the parameters ω and M are inappropriate to describe the position of a satellite in a near-circular orbit (Section 22.4), as the location of perigee is not defined in a circular orbit. According to (6.45): $\theta \rightarrow M$ if $e \downarrow 0$, which means that then for the argument of latitude, u , can be written $u = \omega + \theta \approx \omega + M$. The argument of latitude describes the position of the satellite without any reference to the location of perigee. Indeed, when we combine (23.50-4) and (23.50-6) we find that the perturbation $\Delta\omega_{2,0} + \Delta M_{2,0}$ is zero; in fact, a more-detailed analysis, keeping more terms in the expressions for $\Delta\omega_{2,0}$ and $\Delta M_{2,0}$, shows that the perturbation $\Delta\omega_{2,0} + \Delta M_{2,0}$ actually is of order e . For the satellite orbit specified before, we find $\Delta a_{2,0} = 20.8$ m, $\Delta e_{2,0} = 3.02 \times 10^{-4}$, $\Delta\omega_{2,0} = 2.00^\circ$, $\Delta\Omega_{2,0} = -1.53''$, $\Delta M_{2,0} = -2.00^\circ$. The period of the perturbations is 1.576 hr.

J_2 long-period/short-period variations

Since the conditions for this class of orbital element variations are $n = 2$, $m = 0$, $n - 2p \neq 0$, $n - 2p + q \neq 0$, a total of six terms $\{2,0,p,q\}$ contribute: $\{2,0,0,-1\}$, $\{2,0,0,0\}$, $\{2,0,0,1\}$, $\{2,0,2,-1\}$, $\{2,0,2,0\}$, $\{2,0,2,1\}$. When we substitute the relevant values of n, m, p, q into (23.44), (23.45) and (23.47), subsequently substitute $-J_2 = C_{2,0}$ ((20.1-2)), and use (23.27) and (23.35), we obtain

$$\begin{aligned} \Theta_{2,0,0,-1} &= -\Theta_{2,0,2,1} = 2\omega + M \quad ; \quad \dot{\Theta}_{2,0,0,-1} = -\dot{\Theta}_{2,0,2,1} = 2\dot{\omega} + \dot{M} = \sqrt{\frac{\mu}{a^3}} (1 + O(J_2)) \\ \Theta_{2,0,0,0} &= -\Theta_{2,0,2,0} = 2(\omega + M) \quad ; \quad \dot{\Theta}_{2,0,0,0} = -\dot{\Theta}_{2,0,2,0} = 2(\dot{\omega} + \dot{M}) = 2\sqrt{\frac{\mu}{a^3}} (1 + O(J_2)) \\ \Theta_{2,0,0,1} &= -\Theta_{2,0,2,-1} = 2\omega + 3M \quad ; \quad \dot{\Theta}_{2,0,0,1} = -\dot{\Theta}_{2,0,2,-1} = 2\dot{\omega} + 3\dot{M} = 3\sqrt{\frac{\mu}{a^3}} (1 + O(J_2)) \end{aligned}$$

$$Z_{2,0,0,-1} = Z_{2,0,2,1} = -J_2 \cos(2\omega + M) \quad ; \quad Z_{2,0,0,-1}^* = -Z_{2,0,2,1}^* = -J_2 \sin(2\omega + M)$$

$$Z_{2,0,0,0} = Z_{2,0,2,0} = -J_2 \cos 2(\omega + M) \quad ; \quad Z_{2,0,0,0}^* = -Z_{2,0,2,0}^* = -J_2 \sin 2(\omega + M)$$

$$Z_{2,0,0,1} = Z_{2,0,2,-1} = -J_2 \cos(2\omega + 3M) \quad ; \quad Z_{2,0,0,1}^* = -Z_{2,0,2,-1}^* = -J_2 \sin(2\omega + 3M)$$

$$F_{2,0,0}(i) = F_{2,0,2}(i) = -\frac{3}{8} \sin^2 i \quad ; \quad G_{2,0,-1}(e) = G_{2,2,1}(e) = -\frac{1}{2} e + O(e^3)$$

$$G_{2,0,0}(e) = G_{2,2,0}(e) = 1 - \frac{5}{2}e^2 + O(e^4) ; \quad G_{2,0,1}(e) = G_{2,2,-1}(e) = \frac{7}{2}e + O(e^3)$$

Inspection of (23.46) shows that, when applying the relevant expressions for all six terms, for a generic orbital element α : $\Delta\alpha_{2,0,0} = \Delta\alpha_{2,0,2,0}$, $\Delta\alpha_{2,0,0,-1} = \Delta\alpha_{2,0,2,1}$, $\Delta\alpha_{2,0,0,1} = \Delta\alpha_{2,0,2,-1}$. So, we have to determine $\Delta\alpha$ for only three cases. When we evaluate the six equations (23.46) for these three cases, linearize the resulting expressions by neglecting terms of order e^2 or of order J_2 relative to one, add the expressions for each particular orbital element, and multiply the then resulting expressions by a factor two, we obtain

$$\begin{aligned}\Delta a_{2,0} &= \frac{3}{2} J_2 \frac{R^2}{a} \sin^2 i \left[\cos 2(\omega + M) - \frac{1}{2} e \{ \cos(2\omega + M) - 7 \cos(2\omega + 3M) \} \right] \\ \Delta e_{2,0} &= \frac{3}{8} J_2 \left(\frac{R}{a} \right)^2 \sin^2 i \left[\cos(2\omega + M) + \frac{7}{3} \cos(2\omega + 3M) - e \cos 2(\omega + M) \right] \\ \Delta i_{2,0} &= \frac{3}{8} J_2 \left(\frac{R}{a} \right)^2 \sin 2i \left[\cos 2(\omega + M) - e \left\{ \cos(2\omega + M) - \frac{7}{3} \cos(2\omega + 3M) \right\} \right] \\ \Delta \omega_{2,0} &= -\frac{3}{8} J_2 \left(\frac{R}{a} \right)^2 \left[(2 + 3 \sin^2 i) \sin 2(\omega + M) + \frac{1}{e} \sin^2 i \left\{ \sin(2\omega + M) - \frac{7}{3} \sin(2\omega + 3M) \right\} \right] \\ \Delta \Omega_{2,0} &= \frac{3}{4} J_2 \left(\frac{R}{a} \right)^2 \cos i \left[\sin 2(\omega + M) - e \left\{ \sin(2\omega + M) - \frac{7}{3} \sin(2\omega + 3M) \right\} \right] \\ \Delta M_{2,0} &= \frac{3}{8} J_2 \left(\frac{R}{a} \right)^2 \sin^2 i \left[8 \sin 2(\omega + M) + \frac{1}{e} \left\{ \sin(2\omega + M) - \frac{7}{3} \sin(2\omega + 3M) \right\} \right]\end{aligned}\tag{23.51}$$

The expressions contain sine- and cosine-terms with the arguments $2\omega + jM$, where $j = 1, 2, 3$. Since we have assumed that ω , Ω and M vary linearly with time, these arguments can be written as $(2\dot{\omega} + j\dot{M})t$. So, the period of the sine- and cosine-terms is $P = 2\pi/(2\dot{\omega} + j\dot{M})$. With the first-order expressions (23.27) and (23.35) we then find

$$P = \frac{2\pi}{\left(1 + \frac{2}{j} \frac{\dot{\omega}}{\dot{M}} \right) j \dot{M}} \approx \frac{2\pi}{\left[1 + \frac{1}{j} \frac{3}{2} J_2 \left(\frac{R}{a} \right)^2 (5 \cos^2 i - 1) \right] j \dot{M}}$$

For $h > 250$ km, $e < 0.03$, and any value of i : $|\dot{\omega}| < 3.53 \times 10^{-6}$ rad/s, $\dot{M} < 1.17 \times 10^{-3}$ rad/s, and $2|\dot{\omega}|/\dot{M} < 6.03 \times 10^{-3}$. So, the term in square brackets is slightly larger or slightly smaller than 1. Consequently, the period of the sine- and cosine-terms in (23.51) is about $1/j$ times the anomalistic orbital period. The variations in a , e , i , M are zero for $i = 0^\circ$; the variation in Ω is zero for $i = 90^\circ$. Just as for the class of short-period variations, the amplitudes of the variations are proportional to the term $J_2(R/a)^2$ or $J_2 R^2/a$. Again, the terms with the factor $1/e$ in the expressions for $\Delta\omega_{2,0}$ and $\Delta M_{2,0}$ cancel when we consider the perturbation $\Delta\omega_{2,0} + \Delta M_{2,0}$:

$$\Delta\omega_{2,0} + \Delta M_{2,0} = \frac{3}{8} J_2 \left(\frac{R}{a} \right)^2 [(5 \sin^2 i - 2) \sin 2(\omega + M) + O(e)]$$

For the satellite orbit specified before, we find $\Delta a_{2,0} = -4.24$ km, $\Delta e_{2,0} = -4.94 \times 10^{-4}$, $\Delta i_{2,0} = -1.05'$, $\Delta\omega_{2,0} = -1.767^\circ$, $\Delta\Omega_{2,0} = 0.841'$, $\Delta M_{2,0} = 1.772^\circ$. The periods of the perturbations are 1.573 hr,

0.787 hr, and 0.525 hr. Note that the value of Δa is much larger than for the pure J_2 short-period variation.

J_3 long-period variations

Since the conditions for this class of orbital element variations are $n = 3$, $m = 0$, $n - 2p \neq 0$, $n - 2p + q = 0$, only two terms $\{3,0,p,q\}$ contribute: $\{3,0,1,-1\}$ and $\{3,0,2,1\}$. When we apply the same methodology as described above for the analysis of the other classes of orbit perturbations, we obtain

$$\begin{aligned}\Theta_{3,0,1,-1} &= -\Theta_{3,0,2,1} = \omega \\ \dot{\Theta}_{3,0,1,-1} &= -\dot{\Theta}_{3,0,2,1} = \dot{\omega} = \frac{3}{4} J_2 \sqrt{\frac{\mu}{a^3}} \left(\frac{R}{a} \right)^2 (5 \cos^2 i - 1) (1 + O(J_2)) \\ Z_{3,0,1,-1} &= -Z_{3,0,2,1} = -J_3 \sin \omega \quad ; \quad Z_{3,0,1,-1}^* = Z_{3,0,2,1}^* = J_3 \cos \omega \\ F_{3,0,1}(i) &= -F_{3,0,2}(i) = \frac{3}{16} \sin i (5 \sin^2 i - 4) \quad ; \quad G_{3,1,-1}(e) = G_{3,2,1} = e + O(e^3)\end{aligned}$$

Inspection of (23.46) shows that, when applying the relevant expressions for the terms $\{3,0,1,-1\}$ and $\{3,0,2,1\}$, for a generic orbital element α : $\Delta\alpha_{3,0,1,-1} = \Delta\alpha_{3,0,2,1}$. When we evaluate the six equations (23.46) for the term $\{3,0,1,-1\}$, linearize the resulting expressions by neglecting terms of order e^2 or of order J_2 relative to one, and multiply the then resulting expressions by a factor two, we obtain

$$\begin{aligned}\Delta a_{3,0} &= 0 \\ \Delta e_{3,0} &= -\frac{1}{2} \frac{J_3}{J_2} \frac{R}{a} \sin i \sin \omega \\ \Delta i_{3,0} &= \frac{1}{2} \frac{J_3}{J_2} \frac{R}{a} e \cos i \sin \omega \\ \Delta \omega_{3,0} &= -\frac{1}{2} \frac{J_3}{J_2} \frac{R}{a} \frac{1}{e} \sin i \cos \omega \quad (23.52) \\ \Delta \Omega_{3,0} &= -\frac{1}{2} \frac{J_3}{J_2} \frac{R}{a} e \frac{\cos i (15 \cos^2 i - 11)}{\sin i (5 \cos^2 i - 1)} \cos \omega \\ \Delta M_{3,0} &= \frac{1}{2} \frac{J_3}{J_2} \frac{R}{a} \frac{1}{e} \sin i \cos \omega\end{aligned}$$

The value of J_2 is of order 10^{-3} , while the value of J_3 is of order 10^{-6} (Table 20.1). This means that the value of J_3/J_2 is of order 10^{-3} , i.e. is of the same order of magnitude as J_2 . The attenuation of the magnitude of the perturbations with increasing altitude is less than for the perturbations produced by the J_2 -term, since in this case the ratio R/a occurs to the first power while in the expressions for the perturbations produced by the J_2 -term it appears to the second power. The variations in the orbital elements have a period equal to the time it takes perigee to complete one revolution about the Earth (Section 23.3). The variations of e , ω and M are zero for $i = 0^\circ$, while

the variations of i and Ω are zero for $i = 90^\circ$; the variation of Ω is also zero for $i = 31.1^\circ, 148.9^\circ$. Note that the expression for Ω contains the terms $\sin i$ and $(5\cos^2 i - 1)$ in the denominator. This would lead to a singularity if $i = 63.435^\circ, 116.565^\circ$; i.e. at the critical inclinations (Section 23.3), and if $i = 0^\circ$. However, as stated before, Kaula's theory is not valid for inclinations close to the critical inclinations because these inclinations lead to a resonance phenomenon, and Ω is not defined for $i = 0^\circ$. A numerical analysis shows that for $e < 0.03$ the absolute value of the term $\dot{\theta}i$ in the expression for Ω is larger than 100 if $0^\circ < i < 0.0172^\circ, 63.417^\circ < i < 63.453^\circ, 116.547^\circ < i < 116.583^\circ, 179.982^\circ < i < 180^\circ$. So, outside these narrow ranges of forbidden inclinations, we may apply the expression for Ω . The variations of i and Ω are very small for near-circular orbits. Just as we have found for the J_2 short-period variations of ω and M , and for two components of the J_2 long-period/short-period variations of ω and M , the amplitudes of the J_3 long-period variation of ω and M are proportional to $1/e$, which leads to large amplitudes for near-circular orbits. Again, we conclude that this is due to the fact that the parameters ω and M are inappropriate to describe the position of a satellite in a near-circular orbit. When we add (23.52-4) and (23.52-6), we find that the perturbation $\Delta\omega_{3,0} + \Delta M_{3,0}$ is zero; in fact, a more-detailed analysis shows that the perturbation $\Delta\omega_{3,0} + \Delta M_{3,0}$ actually is of order e . For the satellite orbit specified before, we find $\Delta e_{3,0} = 5.42 \cdot 10^{-4}$, $\Delta i_{3,0} = -0.559''$, $\Delta\omega_{3,0} = 6.21^\circ$, $\Delta\Omega_{3,0} = -1.85''$, $\Delta M_{3,0} = -6.21^\circ$. The period of the perturbations is 62.71 day.

$J_{2,2}$ m-daily variations

Since the conditions for this class of orbital element variations are $n = 2$, $m = 2$, $n - 2p = 0$, $n - 2p + q = 0$, only one term contributes: $\{2,2,1,0\}$. When we apply the same methodology as described above for the analysis of the other classes of orbit perturbations, we obtain

$$\Theta_{2,2,1,0} = -2(\theta_{GM} - \Omega) \quad ; \quad \dot{\Theta}_{2,2,1,0} = -2(\dot{\theta} - \dot{\Omega})$$

$$Z_{2,2,1,0} = J_{2,2} \cos 2(\theta_{GM} - \Omega + \Lambda_{2,2}) \quad ; \quad Z_{2,2,1,0}^* = -J_{2,2} \sin 2(\theta_{GM} - \Omega + \Lambda_{2,2})$$

$$F_{2,2,1}(i) = \frac{3}{2} \sin^2 i \quad ; \quad G_{2,1,0}(e) = 1 + \frac{3}{2} e^2 + O(e^4)$$

where $\Lambda_{2,2}$ is a constant. When we apply (11.2) to the orbit's ascending node, we conclude that the following relation holds (Figure 21.5):

$$\Omega - \theta_{GM} = \Lambda_\Omega$$

where Λ_Ω is the geographic longitude of the ascending node, i.e. the position of the ascending node relative to the location of Greenwich. As the Earth rotates and the orbital plane precesses, Λ_Ω varies linearly with time. When we substitute these expressions into the six equations (23.46), linearize the resulting expressions by neglecting terms of order e^2 relative to one, we obtain

$$\Delta a_{2,2} = \Delta e_{2,2} = 0$$

$$\Delta i_{2,2} = \frac{3}{2} J_{2,2} \left(\frac{R}{a} \right)^2 \frac{\sqrt{\mu/a^3}}{\dot{\theta} - \dot{\Omega}} \sin i \cos 2(\Lambda_\Omega - \Lambda_{2,2}) \quad (22.53)$$

$$\Delta\omega_{2,2} = -\frac{3}{4} J_{2,2} \left(\frac{R}{a} \right)^2 \frac{\sqrt{\mu/a^3}}{\dot{\theta} - \dot{\Omega}} (5 \sin^2 i - 2) \sin 2(\Lambda_\Omega - \Lambda_{2,2})$$

$$\begin{aligned}\Delta\Omega_{2,2} &= -\frac{3}{2} J_{2,2} \left(\frac{R}{a} \right)^2 \frac{\sqrt{\mu/a^3}}{\dot{\theta} - \dot{\Omega}} \cos i \sin 2(\Lambda_\Omega - \Lambda_{2,2}) \\ \Delta M_{2,2} &= -\frac{9}{4} J_{2,2} \left(\frac{R}{a} \right)^2 \frac{\sqrt{\mu/a^3}}{\dot{\theta} - \dot{\Omega}} \sin^2 i \sin 2(\Lambda_\Omega - \Lambda_{2,2})\end{aligned}\quad (23.53)$$

The amplitudes of the variations are proportional to the term $J_{2,2}(R/a)^2$. Since $J_{2,2} \approx 1.7 \cdot 10^{-3} J_2$, generally, we may expect rather small perturbations when compared to the perturbations associated with the J_2 zonal harmonics term. Note that the orbital eccentricity does not appear in the expressions; it only enters through $\dot{\Omega}$, which is a weak function of the eccentricity. Also note that the elements a and e do not experience any variation. The variations in i and M are zero for $i = 0^\circ$, while the variation in Ω is zero for $i = 90^\circ$, and the variation in ω is zero for $i = 39.2^\circ, 140.8^\circ$. The variation of ω, Ω and M is zero for $\Lambda_\Omega - \Lambda_{2,2} = 0^\circ, 90^\circ, 180^\circ, 270^\circ$; the variation of i is zero for $\Lambda_\Omega - \Lambda_{2,2} = 45^\circ, 135^\circ, 225^\circ, 315^\circ$. When we add (23.53-3) and (23.53-5), we find

$$\Delta\omega_{2,2} + \Delta M_{2,2} = -\frac{3}{2} J_{2,2} \left(\frac{R}{a} \right)^2 \frac{\sqrt{\mu/a^3}}{\dot{\theta} - \dot{\Omega}} (4 \sin^2 i - 1) \sin 2(\Lambda_\Omega - \Lambda_{2,2})$$

Now, for $h > 250$ km, $e < 0.03$ and any value of i : $|\dot{\Omega}| < 1.76 \cdot 10^{-6}$ rad/s, while $|\dot{\theta}| \approx 7.29 \cdot 10^{-5}$ rad/s. Since $|\dot{\Omega}| \ll \dot{\theta}$: $\sqrt{\mu/a^3}/(\dot{\theta} - \dot{\Omega}) \approx n/\dot{\theta}$, where n is the mean angular (orbital) motion of the satellite. For satellites in orbits with an altitude of up to 5000 km, the ratio $n/\dot{\theta}$ varies from 7 to 16. For a geosynchronous satellite the ratio $n/\dot{\theta}$ would be equal to one, but this case corresponds to orbital resonance and, as stated before, then Kaula's linear theory is not valid.

Since $J_{2,2}$ is positive, we find for an equatorial orbit ($i = 0^\circ$) that $\Delta\omega_{2,2} + \Delta M_{2,2} > 0$ for $0^\circ < \Lambda_\Omega - \Lambda_{2,2} < 90^\circ$ and $180^\circ < \Lambda_\Omega - \Lambda_{2,2} < 270^\circ$, and that $\Delta\omega_{2,2} + \Delta M_{2,2} < 0$ for $90^\circ < \Lambda_\Omega - \Lambda_{2,2} < 180^\circ$ and $270^\circ < \Lambda_\Omega - \Lambda_{2,2} < 360^\circ$. These results agree with what we have found in Section 21.7 for the secular variation of the along-track position component. However, it is noted that the analysis in Section 21.7 holds for a geostationary orbit, while the analysis in this Section is not valid for a geostationary orbit. Since $\Lambda_{2,2}$ is a constant and Ω and θ_{GM} are assumed to vary linearly with time, we can write for the argument of the sine- and cosine-functions: $2((\dot{\Omega} - \dot{\theta})t - \Lambda_{2,2})$. So, for the period of these functions we may write with (23.27) and (23.35-2)

$$P = \frac{2\pi}{2 \left(1 - \frac{\dot{\Omega}}{\dot{\theta}} \right) \dot{\theta}} \approx \frac{2\pi}{2 \left[1 + \frac{3}{2} \frac{J_2}{\dot{\theta}} \sqrt{\frac{\mu}{a^3}} \left(\frac{R}{a} \right)^2 \cos i \right] \dot{\theta}}$$

For the given range of $|\dot{\Omega}|$ and the known value of $\dot{\theta}$ we find that the ratio $|\dot{\Omega}|/\dot{\theta} < 2.42 \cdot 10^{-2}$, and thus $1 - \dot{\Omega}/\dot{\theta}$ is slightly larger or slightly smaller than one. Consequently, the period of the sine-and cosine-terms in (23.53) is close to half a day. For the satellite in the orbit specified before and $\Lambda_\Omega = 10^\circ$, we find $\Delta i_{2,2} = 3.29''$, $\Delta\omega_{2,2} = -1.38''$, $\Delta\Omega_{2,2} = -3.90''$, $\Delta M_{2,2} = -4.14''$. The period of the perturbations is 11.790 hr.

23.7. Specialized orbits

In the design of satellite missions appropriate orbits have to be selected. For such specialized orbits, the effects of specific orbit perturbations are often exploited explicitly. In this Section, a few examples of such orbits will be presented and discussed.

Up to this point, we have made a clear distinction between osculating, mean and mean-mean orbital elements. However, it should be realized that the relations between these types of elements are mathematically only defined for a particular mathematical model and analysis scheme. For instance, when we considered in Section 23.6 the effect of the J_3 -term of the Earth's gravity field, we have derived analytical expressions for the long-period variations, which link the mean-mean to the mean elements for that particular perturbing force and according to Kaula's computation scheme. However, in reality, many perturbing forces act on a satellite simultaneously and the mathematical definition of mean and mean-mean elements becomes a serious problem, and is also somewhat dependent on the computation scheme applied. In addition, the question can be raised "how do we obtain the values of the orbital elements for a particular satellite?". The answer is that these elements are derived from tracking data acquired on the motion of that satellite. If we would have a measuring concept that would provide us with an accurate measurement of the six components of the satellite's state vector at a single moment of time, we could easily derive the osculating orbital elements of that satellite at that time. However, such a measurement concept does not exist and information about the satellite orbit is derived from a series of measurements on the satellite's motion taken during some period of time. Most modern orbit computation programs integrate the satellite's equations of motion numerically, accounting for many perturbing forces and applying precise geometric and observation models. The observations are used to estimate the satellite's state vector at epoch such that the observations fit the computed orbit as good as possible. The output of such computer programs are satellite state vectors at selected times. From these state vectors the instantaneous osculating orbital elements can be derived (Section 11.10). However, for satellite mission analyses one is particularly interested in the long-term smoothed variation of the orbital elements and then deliberately filters out all kinds of short-period variations, with periods of some tens of minutes to a few days, of these elements. This can be done by applying some type of numerical filtering method or an analytical averaging technique based on an orbit dynamics theory. This results in a kind of 'operational' mean orbital elements; i.e. osculating elements minus short-period variations. These mean elements represent the secular and long-period variations due to all perturbing forces acting on a satellite. In this Section, we will analyze the variations of these mean orbital elements. For simplicity, we will indicate a generic mean orbital element by α and not by $\bar{\alpha}$ as we did before.

For satellites at altitudes between 200 km and 15,000 km the Earth gravity field produces the largest orbit perturbations. Because the absolute value of the J_2 -coefficient of the Earth gravity field is more than 480 times the absolute value of the other even zonal harmonics coefficients and of all tesseral harmonics coefficients, and the absolute value of the J_3 -coefficient is more than seven times the absolute value of the other odd zonal harmonics coefficients (Table 20.1), the J_2 -term of the geopotential produces the dominant secular perturbations and the J_3 -term produces the dominant long-period perturbations. Therefore, in the following analyses we will approximate the true long-term averaged variation of an orbital element by the secular variation due to the J_2 -term of the gravity field model, and by the long-period variation due to the J_3 -term of the gravity field model. Based on the analyses given in Sections 23.3 and 23.5, it may then be stated that only the elements M , Ω , ω show secular variations and that the element a does not show a long-period variation. For low-eccentricity orbits it may be further stated that: 1) the long-period variations of i and Ω are very small, because the relevant expressions contain a factor eJ_3/J_2 ; 2) the long-period variations of ω and M may be relatively large, because the relevant expressions contain a factor $J_3/(J_2 e)$; 3) the long-period variation of e may be of the same order of magnitude as the mean-mean eccentricity.

Earth observation orbits

Most Earth observation satellites encircle the Earth in near-circular orbits at altitudes of 500 km to 2000 km and at orbital inclinations of 70° to 110° . This guarantees both a world coverage and a high resolution of the observations. Exceptions are geostationary satellites which have the advantage of observing continuously the same region on Earth. As a satellite orbits the Earth, the Earth spins beneath it in an eastward direction, and the satellite will pass over the equator at a different longitude on each of its orbits. During a 12-hour period, the satellite can observe the entire Earth's surface, provided that its instruments have a sufficiently wide field of view. For an inclination of exactly 90° , (23.35-2) shows that, to first-order approximation, the secular rate of change of the right ascension of the ascending node is zero. In fact, for $i = 90^\circ$ all zonal harmonics secular and periodic terms in Ω and i vanish, indicating that the orbital plane remains fixed in inertial space, while the Earth rotates underneath and revolves about the Sun. This causes the nature of the illumination at the sub-satellite point to vary, through its range, over a three-month period. So, the same region can be observed every six months under the same lighting conditions. The fact that the satellite observes any spot on the Earth's surface under continuously changing solar lighting conditions can be disadvantageous for the observation instruments and for the interpretation of the acquired observations. Therefore, these satellites often fly so-called *Sun-synchronous orbits*, which are also very attractive for the satellite power supply. This type of orbits is described below.

It should be realized that, in general, the altitude of a satellite in its orbit varies both locally and globally. The primary causes of these variations are: 1) the oblate ellipsoidal shape of the Earth, which results in a variable altitude even for a circular orbit; 2) the eccentricity of the nominal orbit; 3) the short-period variation of the orbital semi-major axis due to the even zonal harmonics of the Earth's gravity field; 4) the long-period variation of the orbital eccentricity primarily due to odd zonal harmonics; 5) the precession of the argument of perigee, which results in a variable configuration of the elliptical orbit relative to the oblate ellipsoid. The overall altitude variation for a near-circular high-inclination orbit may be as large as 50 km or more.

Sun-synchronous orbits

As shown in Section 23.2, but also by (23.52-5), the amplitude of the long-period variation in Ω is, in particular for near-circular orbits, quite small. Consequently, the secular variation of Ω , as produced by the J_2 -term of the geopotential, describes the long-term variation of the mean orbital element Ω very well. The secular rate of change of Ω is given by (23.35-2), which can also be written as

$$\dot{\Omega} = -\frac{3}{2}J_2 \sqrt{\frac{\mu}{a^3}} \left(\frac{R}{a(1-e^2)} \right)^2 \cos i \quad (23.54)$$

where we have neglected terms of order J_2^2 . The orbital elements on the right-hand side of this expression formally are mean-mean orbital elements. However, for the near-circular orbits that we consider in this Section, we may also substitute the mean orbital elements, because: 1) a does not show a long-period variation; 2) the long-period variation of i is negligible; 3) e^2 may be neglected relative to one. It is possible to select, for given values of a and e , the inclination such that the line of nodes rotates at the same rate, and in the same direction, as the Sun in its motion about the Earth. These orbits are called *Sun-synchronous orbits*. The Earth completes one revolution about the Sun (relative to the stars) in one sidereal year. Consequently, the mean motion of the Sun about the Earth, when viewed from the center of the Earth, is $n_S = 0.98565^\circ/\text{day}$, or $1.99107 \times 10^{-7} \text{ rad/s}$, in an eastward direction. Substitution of n_S for $\dot{\Omega}$ in (23.54)

leads to

$$\cos i_{ss} = -n_s \left(\frac{3}{2} J_2 \sqrt{\frac{\mu}{a^3}} \left(\frac{R}{a} \right)^2 (1 - e^2)^{-2} \right)^{-1} \quad (23.55)$$

After substitution of the numerical values of n_s , μ , R and J_2 into (23.55), and evaluation of that expression for nearly circular orbits, we find for the Sun-synchronous inclination, i_{ss} ,

$$\cos i_{ss} \approx -0.098916 \left(\frac{r}{R} \right)^{\frac{7}{2}} \quad (23.56)$$

where $r \approx a$. Of course, other even zonal harmonics of the Earth's geopotential will also affect the secular change of Ω and should be taken into account for a more-accurate computation of i_{ss} . From (23.56) we conclude that the Sun-synchronous inclination is always larger than 90° , and that its value increases at increasing orbital altitudes. We also conclude that Sun-synchronous orbits only exist if $(r/R)^{7/2} < 10.110$. So, the maximum orbital altitude for which a Sun-synchronous orbit exists is $h_{max} = 5975$ km; where $i_{ss} = 180^\circ$. However, this is a purely theoretical case. The real value of Sun-synchronism shows up for low-altitude orbits, where $r/R \approx 1$ and consequently $i_{ss} \approx 95^\circ - 100^\circ$.

The fact that low-altitude, near-circular, Sun-synchronous orbits are quasi-polar forms the basis of an important application of these orbits. By selecting an appropriate launch time, one can create a situation where the vector Earth-Sun is (about) perpendicular to the line of nodes of the satellite orbit. Because in this situation the satellite passes the nodes around 6 a.m. and 6 p.m. local solar time, this orbit is called a *twilight orbit* or *dawn/dusk orbit*. Since the orbit is Sun-synchronous, the Earth-Sun vector will remain (about) perpendicular to the line of nodes during the entire flight of the satellite. As the orbit is nearly polar and the declination of the Sun only varies between 23.5° and -23.5° , it is then possible to create a situation in which the satellite does not enter the Earth's shadow cone for several months, which is ideal for the satellite power supply through solar cell arrays. In Figure 23.9, four configurations of a Sun-synchronous twi-

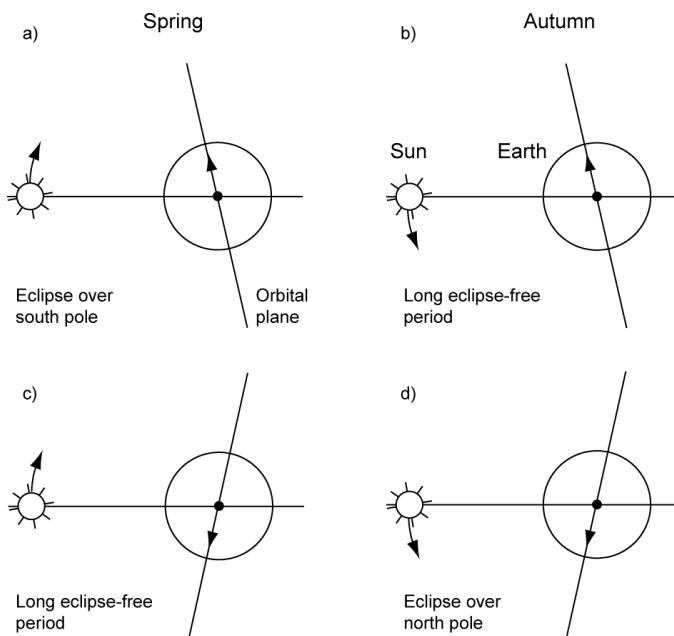


Figure 23.9: Four configurations of a Sun-synchronous twilight orbit.

light orbit are shown schematically. In the graphs the Earth is ‘behind’ the velocity vectors indicated. Note that the terms spring and autumn in this Figure hold for an observer in the northern hemisphere. For the upper two graphs, the local solar time at the ascending node is about 6 p.m.; for the lower two graphs about 6 a.m. It will be clear that when launch takes place in such a way that just after launch the geometry indicated in the configurations b) or c) occurs, then the satellite will be continuously illuminated by the Sun for a long period of time (long eclipse-free period); in configurations a) or d) the satellite may soon enter the Earth’s shadow cone over the south pole or north pole, respectively.

Many Earth observation missions benefit from the characteristics of Sun-synchronous orbits, although the line of nodes does not always need to be about perpendicular to the Earth-Sun vector. By choosing a particular value of the angle between the projection of the Earth-Sun vector on the equatorial plane and the line of nodes, it is possible to reach a situation in which the satellite passes over a specific area always at the same local solar time. Then, the sensors on board observe that area every time under the same lighting conditions, which is very attractive for the interpretation of the measurements. An example is the ERS-1 satellite for which a Sun-synchronous near-circular orbit with perigee and apogee altitudes of about 770 km and 785 km was selected. Equation (23.56) then yields for the required inclination $i = 98.504^\circ$. The value of Ω , in combination with the launch time, was selected such that for moderate northern latitudes the northbound ERS-1 overflights took place at about 22.5 hr local solar time and the southbound passes at about 10.5 hr local solar time.

Since for a Sun-synchronous orbit the secular rate of change of the right ascension of the ascending node, $\dot{\Omega}$, is equal to the angular motion of the Sun as viewed from the Earth, the satellite’s orbit is in deep resonance with respect to the Sun. This leads to a slowly growing oscillation in the orbital inclination. So, the solar gravitational attraction affects in an indirect way the secular rate of change of the ascending node ((23.54)) and thus the Sun-synchronism of the orbit. Ignoring the effects of changes in a and e , we may write

$$\Delta\dot{\Omega} \approx \frac{\partial\dot{\Omega}}{\partial i} \Delta i$$

where we have neglected terms of order $(\Delta i)^2$. With (23.35-2) and (23.54) this relation may be written as

$$\Delta\dot{\Omega} \approx \frac{3}{2} \tilde{n} J_2 \left(\frac{R}{p} \right)^2 \sin i \Delta i = -\dot{\Omega} \tan i \Delta i$$

So, a change in i will lead to an unwanted drift of the angle between the (projection of the) Sun-Earth line and the satellite orbit’s line of nodes. If the effect of Δi on $\dot{\Omega}$, and so on Ω , becomes too large, an orbit maintenance maneuver has to be applied. A possibility is (Section 22.7) to apply a ΔV_w impulsive shot (perpendicular to the orbital plane) at $u = 90^\circ$ or 270° to correct the ascending node back to its nominal position, and a ΔV_w impulsive shot at $u = 0^\circ$ or 180° to correct the inclination back to its nominal value. However, a more desirable approach is to take advantage of the fact that Δi causes the change of the nodal drift. Assume that an impulsive shot with magnitude $(\Delta V_w)_0$ executed at $u = 0^\circ$ or 180° leads to a correction of i and so of $\dot{\Omega}$ back to their nominal values. Then, an impulsive shot with magnitude $2(\Delta V_w)_0$, applied at the same position and in the same direction, will result in an ‘over-corrected’ value of i , but also of $\dot{\Omega}$, and thus in a drift of Ω towards its nominal value. Due to the continuous acting solar gravitational attraction, the values of i and $\dot{\Omega}$ will gradually change again, resulting in a drift of Ω away from its nominal value, and a similar maneuver has to be applied periodically to keep the value of Ω within accepted limits. This maintenance strategy allows to adjust only the inclination and is

more efficient than the previously mentioned dual-impulse maneuver.

Frozen orbits

In general, the major axis of a satellite orbit will precess under the influence of the J_2 zonal harmonic of the Earth's gravity field. In Section 23.3, it was found that if the orbital inclination is equal to one of the critical inclinations (63.435° , 116.565°), this J_2 -induced precession is zero for an arbitrary orbital eccentricity and so the location of the argument of perigee will be stationary. This means that the satellite will always pass over a certain region of the Earth at the same altitude, which is attractive for various types of Earth observation missions, in particular when the satellite acquires radar measurements. However, there is another possibility to keep the location of the argument of perigee stationary at an arbitrary orbital inclination; this is the so-called *frozen orbit*. This type of orbit is realized through the balancing of the secular perturbations due to the even zonal harmonics with the long-period perturbations due to the odd zonal harmonics of the Earth's gravity field, resulting in a cancellation of the rates of change of the mean eccentricity and mean argument of perigee. From a more abstract perspective, frozen orbits arise from bifurcations or singularities in the relevant system of differential equations. In this Section, we will perform a first-order analysis and only account for the first-order effects of the J_2 - and J_3 -term of the Earth's gravity field potential. It will be shown that a frozen orbit is nearly circular ($e \leq 0.001$).

We start from the Lagrange planetary equations (Section 22.1) for the osculating orbital elements e and ω to derive expressions for the time derivatives of the mean orbital elements e and ω . These time derivatives include a secular contribution due to the J_2 -term and a long-period contribution due to the J_3 -term. In Section 23.3 an expression for the perturbing force function, \tilde{R} , that accounts for the J_2 -term is given. From (20.1-3) we can derive an expression for the perturbing force function that account for the J_3 -term. When we combine both expressions we find

$$\begin{aligned} \tilde{R} = & \frac{3}{2} J_2 \frac{\mu R^2}{a^3} \left(\frac{a}{r} \right)^3 \left[\frac{1}{3} - \frac{1}{2} \sin^2 i + \frac{1}{2} \sin^2 i \cos 2(\omega + \theta) \right] \\ & - J_3 \frac{\mu R^3}{a^4} \left(\frac{a}{r} \right)^4 \sin i \left[\left(\frac{15}{8} \sin^2 i - \frac{3}{2} \right) \sin(\omega + \theta) - \frac{5}{8} \sin^2 i \sin 3(\omega + \theta) \right] \end{aligned} \quad (23.57)$$

In Section (23.3) we have derived (23.23), which represents the averaged form of the first part of (23.57) that accounts for the secular variations of the orbital elements due to the J_2 -term. Using the method described in Section 23.3, we should take the mean value of the second part of (23.57) with respect to M , where we substitute the time-averaged value of $(a/r)^4$, and retain only terms of the expression in square brackets which are dependent on ω , to obtain an expression that accounts for the long-period variations of the orbital elements due to the J_3 -term. We then need the following expressions:

$$\begin{aligned} \left(\left(\frac{a}{r} \right)^4 \cos \theta \right)_m &= e (1 - e^2)^{-5/2} \\ \left(\left(\frac{a}{r} \right)^4 \sin \theta \right)_m &= \left(\left(\frac{a}{r} \right)^4 \cos 3\theta \right)_m = \left(\left(\frac{a}{r} \right)^4 \sin 3\theta \right)_m = 0 \end{aligned}$$

where the index m refers to the time-average (average with respect to M) over an orbital revo-

lution. These expressions can be derived by the computation scheme given in Section 23.3. We then find for the relevant averaged perturbing force function

$$\begin{aligned}\tilde{R}_m &= \frac{3}{2} J_2 \frac{\mu R^2}{a^3} (1 - e^2)^{-3/2} \left(\frac{1}{3} - \frac{1}{2} \sin^2 i \right) \\ &\quad - J_3 \frac{\mu R^3}{a^4} e (1 - e^2)^{-5/2} \sin i \left(\frac{15}{8} \sin^2 i - \frac{3}{2} \right) \sin \omega\end{aligned}\tag{23.58}$$

The Lagrange planetary equations for the eccentricity and the argument of perigee contain the partial derivatives of \tilde{R} with respect to $a, e, i, \tau(M)$. When we derive from (23.58) expressions for the partial derivatives of \tilde{R}_m with respect to these elements, substitute these expressions into (22.35-2) and (22.35-4), and neglect terms of order e^2 relative to one, we find after some algebraic manipulation for the two averaged Lagrange planetary equations

$$\begin{aligned}\frac{de}{dt} &= -\frac{3}{8} J_3 \sqrt{\frac{\mu}{a^3}} \left(\frac{R}{a} \right)^3 \sin i (5 \cos^2 i - 1) \cos \omega \\ \frac{d\omega}{dt} &= \frac{3}{4} J_2 \sqrt{\frac{\mu}{a^3}} \left(\frac{R}{a} \right)^2 (5 \cos^2 i - 1) \\ &\quad + \frac{3}{8} J_3 \sqrt{\frac{\mu}{a^3}} \left(\frac{R}{a} \right)^3 \frac{1}{e} \left(\sin i (5 \cos^2 i - 1) - e^2 \frac{\cos^2 i}{\sin i} (15 \cos^2 i - 11) \right) \sin \omega\end{aligned}$$

where the orbital elements, formally, are mean elements. Note that at the critical inclinations, where $5 \cos^2 i = 1$, $d\omega/dt \neq 0$ due to the J_3 -term. As already stated in Section 23.2, J_3 is of order 10^{-6} while J_2 is of order 10^{-3} . Therefore, formally, we should include the second-order contribution of the J_2 -term in the expression for $d\omega/dt$, which would lead to a J_2^2 -term (Section 23.3). For simplicity, we will neglect this J_2^2 -term because it produces an argument of perigee rate that is of order J_2^2 , while the J_3 contribution is much larger and is of order J_3/e .

We now divide part of the expression for $d\omega/dt$ by $\sin i$ and $(5 \cos^2 i - 1)$. This is allowed since: 1) $i = 0^\circ, 180^\circ$ is excluded from our analysis because then ω is not defined; 2) we are not looking for the critical inclinations solution. We then obtain

$$\begin{aligned}\frac{de}{dt} &= -\frac{3}{8} J_3 \sqrt{\frac{\mu}{a^3}} \left(\frac{R}{a} \right)^3 \sin i (5 \cos^2 i - 1) \cos \omega \\ \frac{d\omega}{dt} &= \frac{3}{4} J_2 \sqrt{\frac{\mu}{a^3}} \left(\frac{R}{a} \right)^2 (5 \cos^2 i - 1) \left[1 + \frac{1}{2} \frac{J_3}{J_2} \frac{R}{a} \frac{\sin i \sin \omega}{e} \left(1 - e^2 \frac{\cos^2 i}{\sin^2 i} \frac{(15 \cos^2 i - 11)}{(5 \cos^2 i - 1)} \right) \right]\end{aligned}\tag{23.59}$$

Note that the absolute value of the last term in (23.59-2), $e^2 f(i)$, can become large relative to one when $i \approx 0^\circ, 63.435^\circ, 116.565^\circ, 180^\circ$, even for small values of e . Equation (23.59-2) shows that, in addition to the already discussed case of the critical inclinations, the location of the mean argument of perigee is also stationary when the term in square brackets is zero; this is the requirement for a frozen orbit. From this requirement, the eccentricity of a frozen orbit can be computed for specified values of $R/a, \omega$ and i . According to (23.59-1), e will generally change with time, unless $\cos \omega = 0^\circ$, i.e. $\omega = 90^\circ, 270^\circ$. Because $J_2 > 0$ and $J_3 < 0$ (Table 20.1), we conclude from (23.59-2) that only $\omega = 90^\circ$ satisfies the requirement $e \geq 0$. A numerical analysis of

the term in square brackets shows that, generally, for $\omega = 90^\circ$, $R/a = 0.75 - 0.95$ ($h \approx 340 - 2120$ km): $e < 0.0012$. So, a frozen orbit is a nearly circular orbit indeed! Only for $i = 0^\circ, 180^\circ$, and inclinations within narrow bands of less than $4*10^{-4}$ ° centered at the critical inclinations, e becomes (much) larger than 0.0012, which leads to orbits that are not suited for Earth observation missions. The numerical analysis also shows that, for $R/a = 0.75 - 0.95$ and inclinations outside the bands $0^\circ < i < 0.573^\circ, 63.432^\circ < i < 63.438^\circ, 116.562^\circ < i < 116.568^\circ, 179.427^\circ < i < 180^\circ$, the term in round brackets differs less than 0.01 from one. Consequently, outside these narrow bands of forbidden inclinations and for $\omega = 90^\circ$, we may approximate (23.59-2) by

$$\frac{d\omega}{dt} = \frac{3}{4} J_2 \sqrt{\frac{\mu}{a^3}} \left(\frac{R}{a} \right)^2 (5 \cos^2 i - 1) \left[1 + \frac{1}{2} \frac{J_3}{J_2} \frac{R}{a} \frac{\sin i}{e} \right] \quad (23.60)$$

and the conditions for a frozen orbit are

$$\omega = 90^\circ \quad ; \quad e = -\frac{1}{2} \frac{J_3}{J_2} \frac{R}{a} \sin i \quad (23.61)$$

For orbital inclinations between 64° and 116° , which corresponds to the range of inclinations used for Earth observation missions, and orbital altitudes of 340 - 2120 km, we find $e = 0.00079 - 0.00111$. A frozen orbit is surprisingly stable, suffering little decay due to other natural perturbations—even drag. However, special care must be taken to avoid that the usual orbit maintenance impulsive shots disrupt the frozen-orbit geometry.

The frozen orbit phenomenon was first pointed out by Y. Kozai in 1959 as a peculiarity for small eccentricity orbits and without discussion of its significance for mission planning and operations. The term ‘frozen orbit’ was probably first used in the literature in the mid 1970s in connection with mission analysis studies for SEASAT, a NASA satellite for ocean studies that was launched by an Atlas-Agena rocket from Vandenberg on June 26, 1978, into a non-Sun-synchronous near-circular orbit. Since then, a frozen orbit has been applied for numerous other Earth-orbiting satellites, as well as for Martian, Venusian, and lunar orbiters.

As an example, we consider the orbit selected for SEASAT. Its mission asked for a frozen orbit with $a = 7172.3$ km, $i = 108.0^\circ$. Equation (23.61) then yields $e = 0.000988$. In reality, the satellite flew a near-frozen orbit with $a = 7172.3$ km, $e = 0.0008$, $i = 108.02^\circ$. For that orbit, the perigee librates in the (e, ω) phase plane (Figure 23.10, top); the instantaneous values of e and ω move over a closed contour in this Figure in a counter-clockwise direction at the apsidal period (210 days for SEASAT). The other perturbing forces acting on the satellite make the actual motion of e versus ω to deviate somewhat from this pure libration, but the curve remains nearly closed. For orbits with a value of e that differs more from its frozen-orbit value the trajectories in the (e, ω) phase plane become wider and ultimately they appear as sinusoidal waves. This is of little concern though because orbit adjust maneuvers can restore the frozen-orbit geometry long before any significant decay has occurred.

The frozen orbit geometry minimizes both the local and global altitude variation. The global variation is diminished because the long-period motion in the mean eccentricity is damped out. The local variation is diminished because the motion of the mean argument of perigee is damped out. Furthermore, all sub-satellite points at a given latitude have the same altitude variation. Adoption of a frozen orbit eccentricity results in a global altitude variation of about 35 km, most of this due to the Earth's flattening. Figure 23.10 (bottom) shows that for a SEASAT-type of orbit but with a slightly higher eccentricity, the variation of the orbital altitude over the northern hemisphere is small. This is characteristic for frozen orbits and is attractive for satellites that carry a radar altimeter or a Synthetic Aperture Radar (SAR).

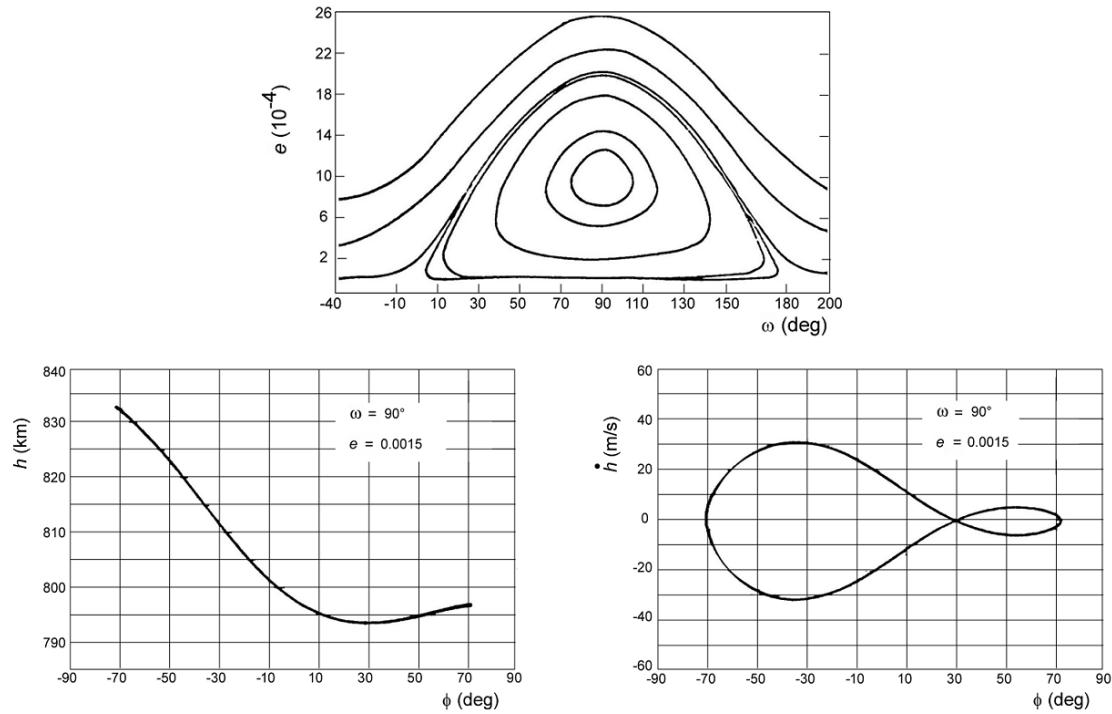


Figure 23.10: Libration of the mean argument of perigee in the (e, ω) phase plane for near-frozen orbits (top), and the variation of altitude (bottom left) and altitude rate (bottom right) as a function of latitude over an orbital revolution for a SEASAT-type of frozen orbit. [copied from: E. Cutting, G.H. Born, J.C. Frautnick, *Orbit Analysis for SEASAT-A*, The Journal of the Astronautical Sciences, Vol. XXVI, No. 4, pp. 315-342, December 1978.]

As a second example, the orbit of the US/French TOPEX/Poseidon satellite may serve. This oceanography satellite was launched on August 10, 1992, by an Ariane-42P rocket from Kourou. The mission asked for a non-Sun-synchronous frozen orbit with $a = 7714.429$ km and $i = 66.041^\circ$. From (23.61) we obtain $e = 0.000884$. However, this value holds for the case that we only account for the effects of the J_2 and J_3 zonal harmonics. Close to the critical inclination a frozen orbit is very sensitive to higher-degree zonal harmonics and to tesseral and sectorial harmonics of the Earth's geopotential (Section 20.1). To illustrate this, it is noted that when, instead of only the J_2 and J_3 zonal harmonics, the J_2 to J_9 zonal harmonics would have been taken into account, the inclination bands where $e > 0.0012$ would have broadened from less than 4×10^{-4} to $57.1^\circ < i < 65.6^\circ$ and $114.4^\circ < i < 122.9^\circ$. So, these broad bands should be avoided for Earth observation missions. When we take the additional gravitational perturbations into account, we find that the value of ω should be either 90° or 270° , depending on the gravity field terms included in the analysis. It can be shown that when for TOPEX/Poseidon only terms up to $n,m = 7$ are accounted for, its frozen orbit requires $\omega = 90^\circ$; when terms up to $n,m = 11$ are accounted for, its frozen orbit requires $\omega = 270^\circ$; and when terms up to $n,m = 180$ are accounted for, its frozen orbit again requires $\omega = 90^\circ$. For the latter case, the value of the mean eccentricity turns out to be $e = 0.000095$. The conditions actually achieved by the orbit acquisition maneuvers were $e = 0.000143$ and $\omega = 90.6^\circ$.

Ground track repeat orbits

A satellite ground track is defined as the curve on the Earth's surface that connects successive crossing points of the satellite's geocentric position vector and the Earth's surface (Section 3.5).

In general, the satellite's ground track will gradually fill all longitudes within the latitude range that is determined by the orbital inclination. Note that the angle at which the ground track intersects the equator is not equal to the inclination, because the Earth is rotating under the orbit. For some missions it is required that the ground track exactly repeats after a certain period of time: the *repeat period*, P_{rep} . All repeat ground tracks are characterized by a set of fixed equator crossing points, which 'tie' the ground tracks to the Earth. For our analysis we will consider only ascending equator crossing points. An important parameter in the analysis of repeat ground tracks is the precession of the orbital plane. Equation (23.15) shows that the amplitude of the long-period variation in Ω is, in particular for near-circular orbits, much smaller than the secular variation of Ω . Consequently, the secular variation of Ω , as produced by the J_2 -term of the geopotential, describes the long-term precession of the orbital plane very well.

If the orbital plane would be fixed in space, the longitude of a consecutive crossing of the ascending node would be shifted over an angle $T_N \dot{\theta}$, where T_N is the nodal period (Section 23.3) and $\dot{\theta} = 7.29212 \times 10^{-5}$ rad/s is the mean (sidereal) rotational velocity of the Earth, towards the west relative to the longitude of the previous crossing of the ascending node. When the secular precession of the orbital plane is accounted for, the geocentric angle (longitude difference) between consecutive crossings of the ascending node is $S = T_N(\dot{\theta} - \dot{\Omega})$; where S is the *fundamental interval*. Now, suppose that the satellite completes precisely α nodal revolutions while the Earth rotates exactly β times relative to the precessing orbital plane. Then, the ground track will repeat after α nodal revolutions of the satellite and we may write

$$\alpha T_N(\dot{\theta} - \dot{\Omega}) = 2\beta \pi$$

or

$$\alpha T_N = \frac{2\beta \pi}{\dot{\theta} - \dot{\Omega}} \quad (23.62)$$

where α and β are positive irreducible integers. The quantity $2\pi/(\dot{\theta} - \dot{\Omega})$ defines the nodal day introduced in Section 23.3. So, (23.62) indicates that the satellite completes α nodal revolutions in β nodal days, and the repeat period, P_{rep} , is given by

$$P_{rep} = \frac{2\beta \pi}{\dot{\theta} - \dot{\Omega}} \quad (23.63)$$

Substitution of (23.38) into (23.62) yields

$$\alpha \frac{2\pi}{\tilde{n} + \dot{\omega}} = \beta \frac{2\pi}{\dot{\theta} - \dot{\Omega}}$$

or

$$\frac{\tilde{n} + \dot{\omega}}{\dot{\theta} - \dot{\Omega}} = \frac{\alpha}{\beta} \quad (23.64)$$

where $\tilde{n} = \dot{M}$ ((23.35)). As we have assumed that only the J_2 -term of the gravity field produces the secular variations of the mean orbital elements, $\dot{\omega}$ and $\dot{\Omega}$ are given by (23.27) and (23.35). In reality, also other zonal harmonics of the Earth's gravity field and the gravitational attraction of the Sun and the Moon contribute to $\dot{\Omega}$ (Section 21.3). However, the contributions by the other zonal harmonics is much smaller than the contribution by the J_2 -term, and, generally, luni-solar perturbations may be ignored for first-order analyses.

During a repeat period, a complete grid of ground tracks will be produced. On each day, ground tracks will be laid down at evenly spaced intervals (fundamental interval) toward the west; ground tracks on successive days will fill in the intervals between the tracks laid down the first day. It will be clear that for a repeat orbit the equatorial longitudinal distance between consecutive ground tracks is $2\pi R\beta/\alpha$, where R is the (mean) equatorial radius of the Earth; the equatorial distance between adjacent ground tracks then is $2\pi R/\alpha$.

Note that (23.64) is identical to the condition given in Section 23.5 for repeat ground track resonance. Consequently, ground track repeat orbits are subject to large-amplitude resonance perturbations that might become unacceptable to the mission profile and therefore need periodic orbit maintenance maneuvers. Comparing (23.49) and (23.64) we find $m = \alpha$ and $n - 2p = \beta$, where n and m are the degree and order of a harmonic of the Earth's gravity field, respectively, and p is a coefficient defined in Section 23.5. Most Earth observation satellites circle the Earth at altitudes between 500 km and 2000 km, and these satellites complete 11.3 - 15.2 revolutions about the Earth per day. As an example, let us consider again the ERS-1 satellite. It was mentioned in Section 23.5 that one of the ground track repeat orbits flown by this satellite had a semi-major axis, eccentricity and inclination of 7153.14 km, 0.00106 and 98.523° , respectively, and that the satellite completed in this orbit 43 nodal revolutions in 3 nodal days; so, $\alpha/\beta = 43/3$ (≈ 14.33). With the theory developed in Section 23.5 we find that for this orbit the major resonance-producing terms of the gravity field are $\{43,43,20,0\}$, $\{45,43,21,0\}$, $\{47,43,22,0\}$, etc., where the notation introduced in Section 23.5 is used. It is emphasized that the larger the values of β , the weaker the resonance effects (Section 23.5). So, polar circular orbits for which e.g. $\alpha/\beta = 127/10$ ($h \approx 1360$ km), $501/35$ ($h \approx 770$ km), $2411/168$ ($h \approx 755$ km) will experience a milder resonance than a polar circular orbit for which $\alpha/\beta = 43/3$ ($h \approx 760$ km). The most intense resonance perturbations may be expected for satellites in orbits for which $\alpha/\beta = 2/1$ or $1/1$. However, for near-circular orbits this situation occurs only when $r/R \approx 4.2$ or $r/R \approx 6.6$, respectively, and then the term $(R/a)^n$ in the expressions for the orbit perturbations due to the Earth's gravity field leads to a weakening of the resonance perturbations in the orbital elements.

Substitution of (23.12), (23.27) and (23.35) into (23.64) yields

$$\sqrt{\frac{\mu}{\mu}} \frac{\alpha}{\beta} \dot{\theta} = \left[1 + \frac{3}{4} J_2 \left(\frac{R}{a} \right)^2 \frac{3 \cos^2 i - 1}{(1 - e^2)^{3/2}} \right] \cdot \left[1 + \frac{3}{4} J_2 \left(\frac{R}{a} \right)^2 \frac{1}{(1 - e^2)^2} \left(5 \cos^2 i - 2 \frac{\alpha}{\beta} \cos i - 1 \right) \right]$$

Because the term α/β may be quite large, we do not evaluate the product of both terms in square brackets and subsequently retain only terms of the first order of J_2 , but rewrite the expression as

$$\begin{aligned} a &= \left\{ \frac{\mu}{\dot{\theta}^2} \left(\frac{\beta}{\alpha} \right)^2 \right\}^{1/3} \left[1 + \frac{3}{4} J_2 \left(\frac{R}{a} \right)^2 \frac{3 \cos^2 i - 1}{(1 - e^2)^{3/2}} \right]^{2/3} \cdot \\ &\quad \cdot \left[1 + \frac{3}{4} J_2 \left(\frac{R}{a} \right)^2 \frac{1}{(1 - e^2)^2} \left(5 \cos^2 i - 2 \frac{\alpha}{\beta} \cos i - 1 \right) \right]^{2/3} \end{aligned} \tag{23.65}$$

Note that when α/β , e and i are specified, this equation can be used to solve for a in an iterative way. As a zeroth-order approximation, we may use

$$a^{(0)} = \left\{ \frac{\mu}{\dot{\theta}^2} \left(\frac{\beta}{\alpha} \right)^2 \right\}^{1/3} \quad (23.66)$$

Substitution of this value in the terms in brackets on the right-hand side of (23.65) yields an improved value of the mean semi-major axis, a , which we substitute on the right-hand side of (23.65), and so on. If the orbit also has to be Sun-synchronous, we have to solve (23.54) and (23.65) simultaneously, which yields the required values of a and i for a specified value of α/β and e .

In reality, various perturbing forces will cause the actual ground track to drift from that generated by the reference orbit (nominal ground track), which will require periodic orbit adjustment maneuvers. Such deviations are, for example, caused by the long-period variation of the mean orbital elements due to odd zonal harmonics of the Earth's gravity field, by a nodal drift due to the gravitational attraction by the Sun and the Moon, and by a decay of the semi-major axis due to atmospheric drag. The first effect leads to small periodic deviations of the ground track, the second effect is quite small, although it does grow over time and eventually causes an unacceptable error in the repeat ground track pattern. The third effect can be significant for low-Earth orbits. For that reason we will discuss it briefly.

Although T_N depends on a , e , and i , the primary dependence is on the semi-major axis, through the mean motion, n . Basically, a change in the nodal period, ΔT_N , will cause a change in the ground track shift, ΔS , per revolution of

$$\Delta S = (\dot{\theta} - \dot{\Omega}) \Delta T_N - T_N \Delta \dot{\Omega} \quad (23.67)$$

The last term on the right-hand side is really a higher-order effect and can be ignored for a first-order analysis. The change in T_N due to a decaying semi-major axis is to first-order approximation equal to the change in the Keplerian period and can be found from (6.25):

$$\Delta T_N \approx \frac{\partial T_N}{\partial a} \Delta a = 3\pi \sqrt{\frac{a}{\mu}} \Delta a$$

Substitution of this relation into (23.67) gives

$$\Delta S = 3\pi (\dot{\theta} - \dot{\Omega}) \sqrt{\frac{a}{\mu}} \Delta a$$

Consequently, a decay in the semi-major axis causes the nodal period to decrease, which means that the satellite crosses the ascending node sooner than desired. This results in an eastward shift of the ground track, which continues to grow as the semi-major axis decays. In fact, the ground track shift grows quadratically with the number of revolutions. Consequently, periodic orbit adjustment burns have to be made to restore the semi-major axis. To maximize the interval between such ground track maintenance maneuvers, we take advantage of a margin (deadband) on each side of the nominal ground track pattern; when the actual ground tracks stay within this deadband, the mission is considered successful. When the actual ground track pattern nears the eastern deadband boundary, an impulsive shot in the orbital plane and perpendicular to the position vector is executed at perigee (Section 22.7) to increase the value of the semi-major axis and to reverse the ground track drift. The value of Δa is chosen such that by the time the ground track has drifted to the western deadband boundary, the effect of drag has reduced a to its

nominal value for the repeat ground track, and the ground track drift stops. As the decay effects continue, the ground track will start again to drift in the eastward direction and eventually the next maintenance maneuver at the eastern deadband boundary has to be executed.

Oceanography satellite orbits

Oceanography satellites require near-circular orbits of which the ground track repeats itself after some time within a deadband of about 1 km. This requirement guarantees that measurements are made repeatedly at the same geographic location, which e.g. is crucial to monitor variations in sea level through satellite altimetry. Then, a series of sea-level height observations at the same geographic location is obtained and all kinds of geographically-correlated errors that are present in the models applied for the processing of the altimeter data cancel. The period in between successive measurements at a certain location, the ground track repeat period, may range from days to months, depending on the type of features to be sampled. A dense grid of ground tracks provides a good spatial sampling of the ocean surface, but inevitably leads to a long repeat period. This means that it takes a long time before the same geographic location is measured again, which leads to a bad temporal sampling of ocean phenomena. So, for processes that do not, or only very slowly, change with time, a long repeat period (dense ground track pattern) can be chosen, while for observing fast changing ocean phenomena one has to select a short repeat period and consequently a wide ground track pattern. In practice, the selection of the orbit repeat period is a trade-off between various requirements.

SEASAT originally flew in a 17-days near-repeat orbit, which was adjusted to a 3-days repeat frozen orbit during the last month of its 3-months mission. In that orbit the satellite completed 43 nodal revolutions in 3 nodal days: $\alpha/\beta = 43/3$. Using this value, we find from (23.66): $a^{(0)} = 7145.7$ km; using the values of $e = 0.0008$ and $i = 108^\circ$ given before, we find from (23.65) after a number of iterations: $a = 7169.0$ km. This value is close to the value $a = 7172.3$ km given before. From (23.27), (23.35) and (23.63) we find that the repeat period is 3.0088 (mean solar) days. ERS-1 has flown in a number of different Sun-synchronous repeat orbits for which the following α/β -combinations held: 43/3, 501/35, 2411/168. It will be clear that the last combination resulted in a very dense pattern of ground tracks; however, in this orbit it took 168 nodal days before the same geographic location was passed over again. Figure 23.11 shows the ERS-1 ground track for the 43/3 orbit, while Figure 23.12 shows the ERS-1 ground track over the North Atlantic both for the 43/3 (left) and 501/35 (right) orbit. Clearly visible is the much denser pattern for the 501/35 orbit; the equatorial spacing between adjacent ground tracks is 932 km for the 43/3 orbit and 80 km for the 501/35 orbit. For TOPEX/Poseidon an orbit with $\alpha/\beta = 127/10$ was selec-

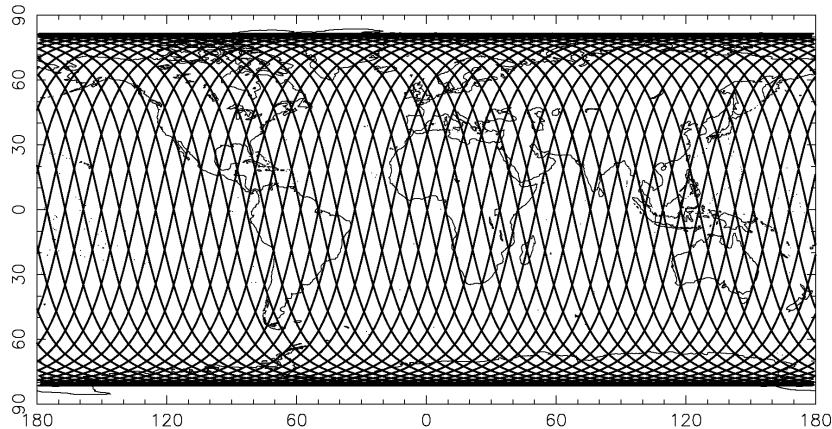


Figure 23.11: ERS-1 ground track for its 43/3 repeat orbit.

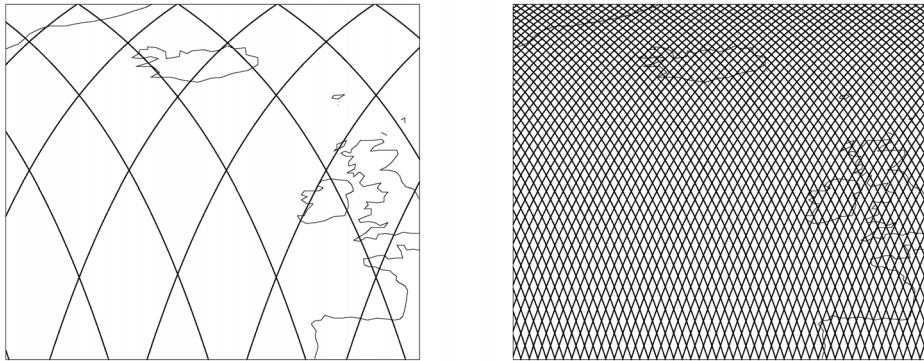


Figure 23.12: ERS-1 ground track over the North Atlantic for its 43/3 (left) and 501/35 (right) repeat orbits.

ted, which means that after 10 nodal days, when the satellite has encircled the Earth 127 times, the same geographic location was passed over again. Using $\alpha/\beta = 127/10$ and the values of e and i given before: $e = 0.000883$, $i = 66.041^\circ$, we find from (23.27), (23.35), (23.63), (23.65) and (23.66): $a = 7714.4$ km, $P_{rep} = 9.9156$ day; the equatorial spacing between adjacent ground tracks is 316 km.

It is remarkable that during the mission of a satellite various α/β -combinations can be realized by executing only small orbit maneuvers. This can be illustrated by comparing two orbits selected for the ERS-1 mission. For the 43/3 orbit, the mean semi-major axis, inclination, and eccentricity have the values 7153.16 km, 98.523° , and 0.00106, respectively. For the 501/35 orbit, the mean elements have the values 7159.51 km, 98.549° , and 0.00106, respectively. So, the semi-major axes of these two orbits differ by only 6.4 km and the inclination by only 0.026° . These small orbit changes can easily be realized by the satellite's onboard propulsion system and therefore ERS-1 could be maneuvered into different repeat orbits during its operational life.

GPS orbits

The GPS satellite navigation system originally called for 24 satellites in circular orbits in six orbital planes, with four satellites in each plane. The six orbital planes have an inclination of 55° and are separated in the right ascension of the ascending node by 60° . The objective was that at least six satellites are always within line of sight from almost everywhere on the Earth's surface, allowing an accurate position determination globally. The result of this objective is that the four satellites are not evenly spaced apart within each orbital plane; the angular distance between the satellites in each plane is 30° , 105° , 120° , and 105° . The orbital altitude is approximately 20,200 km; the orbital period is about $11^{\text{h}}58^{\text{m}}$. This means that a satellite completes about two revolutions each sidereal day, approximately repeating the same ground track each day. As of April 2014, there are 30 actively broadcasting satellites in the GPS constellation. With the increased number of satellites, the constellation was changed to a non-uniform arrangement. Such an arrangement was shown to improve reliability and availability of the system, relative to a uniform system, when multiple satellites fail. In this arrangement about eight satellites are visible from any point on the ground at any one time. When we substitute $\alpha/\beta = 2/1$, $e = 0$ and $i = 55^\circ$ into (23.27), (23.35), (23.63), (23.65) and (23.66) we find $a = 26,560.4$ km, $P_{rep} = 23^{\text{h}}55^{\text{m}}55^{\text{s}}$. Figure 23.13 shows a typical GPS ground track; the tick marks indicate time intervals of 30 min. As mentioned before, the near-repeat ground track leads to near resonance with those terms of the Earth's gravity field for which $m = j\alpha$, $n - 2p = j\beta$, $j = 1, 2, 3, \dots$. Realizing that, in particular for high-altitude orbits, the largest perturbation amplitudes are produced by terms with the lowest

value of n , we find that the dominant resonance term is $\{3,2,1,0\}$; here the notation introduced in Section (23.5) is used.

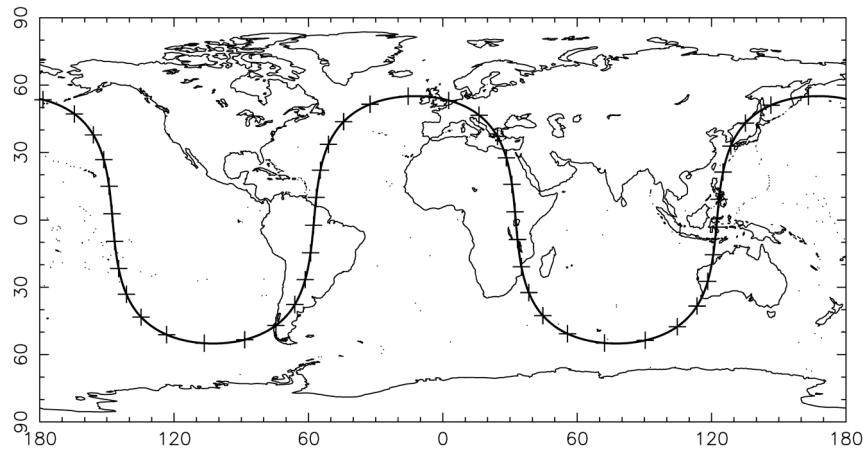


Figure 23.13: Ground track of a GPS repeat orbit.

Molniya orbits

In Section 23.3 the concept of the critical inclination was introduced; i.e. the inclination for which the secular variation of the mean argument of perigee vanishes, if we account only for the effect of the J_2 -term of the Earth's gravitational potential. It was found that the value of the critical inclination is 63.435° or 116.565° . For that inclination, the orientation of an elliptical orbit in its orbital plane is stationary, while the orbital plane itself rotates about the polar axis. This can be an attractive characteristic for certain missions, such as the USSR/Russian Molniya satellite communications system. In the west, the geostationary orbit has been broadly adopted for satellite communications systems. This was possible because the USA, and later ESA, have a launch site at (fairly) low latitudes (Cape Canaveral, 28.5° N; Kourou, 5.2° N). The Soviet Union, however, did not have the disposal of a low-latitude launch site (Table 16.1) and, in the early days of spaceflight, did not have rockets powerful enough to launch communications satellites from their launch sites into geostationary orbit. Therefore, they used a system of Molniya communications satellites, which move about the Earth in an eccentric orbit with an argument of perigee $\omega = 270^\circ$ and of which the ground track repeats itself after 1 nodal day when the satellite has completed 2 nodal revolutions about the Earth. This means that the orbital period is close to 12 hr. The orbit geometry is such that once every two revolutions the apogee is over Russia. The early satellite orbits had an eccentricity $e \approx 0.74$. When we substitute $\alpha/\beta = 2/1$, $e = 0.74$ and $i = 63.435^\circ$ into (23.27), (23.35), (23.63), (23.65) and (23.66) we find $a = 26,561.8$ km, $P_{rep} = 23^{\text{h}}55^{\text{m}}29^{\text{s}}$. The resulting perigee and apogee altitudes are 526 km and 39,825 km, respectively. Since the satellite moves very slowly at its high apogee (Section 6.5) it appears to 'hover' for hours at a time over northern latitudes. The elevation of the spacecraft is over 5° from all points north of latitude 49.2° N for a period of 6 hr centered at apogee. As the whole of Russia is located above 42° N, an appropriate selection of the orbit's right ascension of the ascending node guarantees that the satellite is visible during this period from most locations within Russia. If four satellites are used, which are appropriately phased in similar orbits but in four orbital planes with the right ascension of the ascending nodes separated by 90° , the satellite system may provide continuous communication links between Russian ground stations. Each satellite is then active for a period of 6 hours. In practice, for operational reasons a system with more satellites in more orbital planes, but all with $i = 63.435^\circ$, was used. Of course, perturbing forces acting on the satellites gradually move the orbital elements away from their nominal values. Therefore, the

spacecraft are equipped with an orbit correction propulsion system, which is periodically activated to correct the orbital elements. Nowadays, the Russian launchers are powerful enough to bring large telecommunications satellites into geostationary orbit and Russia frequently uses their geostationary satellites for national and international communications. These satellites are visible at elevations above 5° at latitudes below 76.3° . So, they cover almost the entire Russian territory, although, for a major part of this territory they are visible at low elevations. This means that their signals may easily be obstructed by buildings and irregular terrain topography, while the signals from the Molniya satellites reach the user at higher elevations.

The first operational Molniya satellite was launched in April 1965 and since then more than 160 satellites have been launched, in three series of increasing communications performance, to replace old or failed satellites and to replenish the configuration. Figure 23.14 shows the ground track of a typical Molniya orbit; in this example the apogee is over Finland and over Alaska. The tick marks indicate time intervals of 30 min and clearly show that the satellite is most of the time over northern latitudes. As for the Molniya orbits the same value of α/β holds as for the GPS orbits, also for these orbits the dominant resonance term is $\{3,2,1,0\}$. The low perigee altitude of about 526 km of the early Molniya satellites made them vulnerable to the deleterious effect of the Van Allen radiation belts. Therefore, later generations of Molniya satellites flew in orbits with the same inclination, argument of perigee and semi-major axis, but with an eccentricity of about 0.72, resulting in perigee and apogee altitudes of 1057 km and 39,295 km, respectively. The last successful launch of a Molniya satellite occurred in February 2004. In the period December 2006 to August 2014 five Russian Meridian military communications satellites have been launched in Molniya-type orbits.

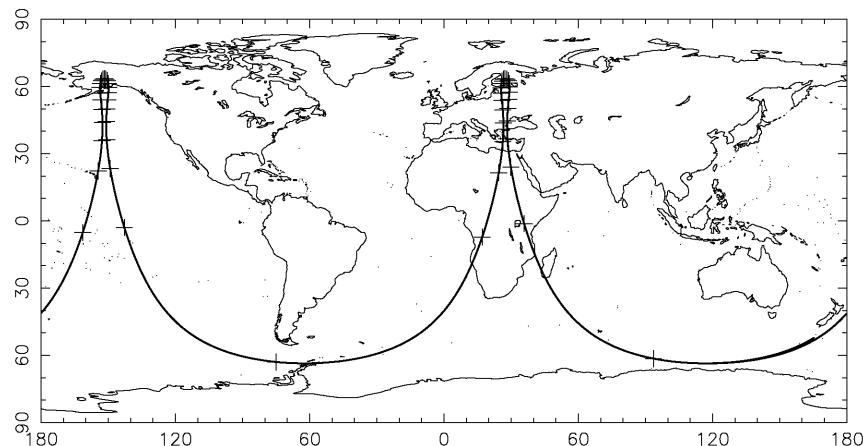


Figure 23.14: Ground track of a Molniya repeat orbit.

Tundra orbits

A Tundra orbit is a modern variant of the Molniya orbit. The ground track repeats itself after 1 nodal day when the satellite has completed 1 nodal revolution ($\alpha/\beta = 1/1$), which leads to an orbital period of about 24 hr (geosynchronous satellite); the inclination of the orbit is equal to the critical inclination (63.435°), the eccentricity is about 0.3, and the argument of perigee is 90° or 270° . When we substitute $\alpha/\beta = 1/1$, $e = 0.3$, and $i = 63.435^\circ$ into (23.27), (23.35), (23.63), (23.65) and (23.66) we find $a = 42,163.4$ km, $P_{rep} = 23^{\text{h}}56^{\text{m}}2^{\text{s}}$. Figure 23.15 (left) shows the ground track of a Tundra orbit with apogee over Canada. Again, the tick marks indicate time intervals of 30 min. For a $\alpha/\beta = 1$ repeat ground track those terms of the Earth's gravity field for which $m = j$, $n - 2p = j$, $j = 1, 2, 3, \dots$ lead to resonance. Realizing that the largest perturbation

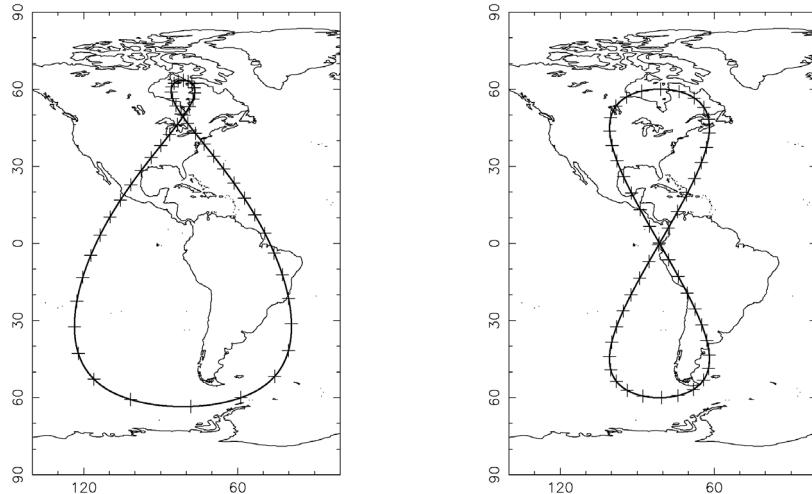


Figure 23.15: Ground track of a Tundra repeat orbit (left) and of a circular geosynchronous orbit with an inclination of 30° (right).

amplitudes are produced by terms with the lowest value of n , we find for the dominant resonance term $\{2,2,0,0\}$. Typically, a Tundra orbits constellation consists of two satellites in each of three orbital planes; all orbits have the same semi-major axis, eccentricity and inclination. The selection of a larger orbital eccentricity would result in a higher apogee, which is attractive for an extended communications period, but would also lead to a lower perigee, which would bring the satellite periodically in the Van Allen belts and would shorten the lifetime of the satellite. Therefore, generally, the value of about $e = 0.3$ is applied that yields, in combination with $\alpha/\beta = 1/1$, perigee and apogee altitudes of 23,140 km and 48,440 km, respectively. The high apogee in combination with the 24 hr period guarantee a visibility for more than 12 hr under the apogee. This type of orbit has many attractive features to commercial communications operators. The geosynchronous period of the orbit allows each satellite to maintain its ‘visible’ position over the same region of the world, allowing operators to tailor their services to a specific market region. The high inclination is attractive because major dense population centers of the world are located at higher latitudes, and allows the satellite to broadcast with higher elevation angles into that desired region, when compared to geostationary satellites. As the satellite travels into the higher latitudes, the signals are received on the ground from overhead rather than from a low elevation angle. This feature mitigates the ‘urban jungle’ effect in large cities as well as providing coverage into remote mountainous or heavily forested regions.

Geosynchronous orbits

As a last example, we consider the cases of a circular geosynchronous orbit with $i = 30^\circ$ (Figure 23.15, right) and a geostationary orbit. For both orbits: $\alpha/\beta = 1/1$. When we substitute $e = 0$, $i = 30^\circ$, $\alpha/\beta = 1/1$ into (23.27), (23.35), (23.63), (23.65) and (23.66), we find $a = 42,165.4$ km, $P_{rep} = 23^{h}56^{m}1^{s}$. Note that the mean semi-major axis is about 1.2 km larger than the value computed on basis of pure Keplerian motion (Section 6.2). Again, the tick marks in Figure 23.15 (right) are plotted at 30 min time intervals. When we substitute $e = 0$, $i = 0$, $\alpha/\beta = 1/1$ into (23.27), (23.35), (23.63), (23.65) and (23.66), we find $a = 42,166.3$ km, $P_{rep} = 23^{h}56^{m}1^{s}$. For this geostationary orbit the mean semi-major axis is about 2.1 km larger than the value of 42,164.2 km computed on basis of pure Keplerian motion. As for a geosynchronous orbit the same value of α/β holds as for a Tundra orbit, also for this type of orbit the dominant resonance term is

{2,2,0,0}. A satellite in a geostationary orbit will appear motionless to an observer on Earth with no change of elevation or azimuth during the day. Therefore, this orbit is attractive for global communications satellites, because it eliminates the need for tracking mechanisms so that communication with the satellite can be maintained with a fixed antenna on Earth.

APPENDIX A: CONSULTED AND RECOMMENDED BOOKS

- V.I. Arnol'd, V.V. Kozlov, A.I. Neishtadt, *Mathematical Aspects of Classical and Celestial Mechanics*, Springer Verlag, New York, 1997.
- R.M.L. Baker, *Astrodynamics, Applications and Advanced Topics*, Academic Press, New York, 1967.
- R.R. Bate, D.D. Mueller, J.E. White, *Fundamentals of Astrodynamics*, Dover Publications, Mineola, New York, 1971.
- R.H. Battin, *An introduction to the Mathematics and Methods of Astrodynamics*, American Institute of Aeronautics and Astronautics, New York, 1987.
- D. Brouwer, G.M. Clemence, *Methods of Celestial Mechanics*, Academic Press, New York, 1961.
- V.A. Chobotov, *Orbital Mechanics*, third edition, American Institute of Aeronautics and Astronautics, New York, 2002.
- J.W. Cornelisse, H.F.R. Schöyer, K.F. Wakker, *Rocket Propulsion and Spaceflight Dynamics*, Pitman Publishing Ltd., London, 1979.
- J.M.A. Danby, *Fundamentals of Celestial Mechanics*, The Macmillan Company, New York, 1962.
- R. Deutsch, *Orbital Dynamics of Space Vehicles*, Prentice-Hall, Englewood Cliffs, New Jersey, 1963.
- K.A. Ehricke, *Space Flight, Vol. 1: Environment and Celestial Mechanics; Vol.2: Dynamics; Vol. 3, Operations*, Van Nostrand, Princeton, 1960, 1962, 1964.
- P.R. Escobal, *Methods of Orbit Determination*, John Wiley & Sons, New York, 1976.
- P.M. Fitzpatrick, *Principles of Celestial Mechanics*, Academic Press, New York, 1970.
- F.T. Geyling, H.R. Westerman, *Introduction to Orbital Mechanics*, Addison-Wesley, Reading, 1971.
- Y. Hagihara, *Celestial Mechanics, Vols. 1-2*, Massachusetts Institute of Technology Press, Cambridge, 1970-1972.
- S. Herrick, *Astrodynamics, Vol. 1 and Vol. 2*, Van Nostrand Reinhold, London, 1971 and 1972.
- J. Jensen, G. Townsend, J. Kork, D. Kraft, *Design Guide to Orbital Flight*, McGraw-Hill, New York, 1962.
- W.M. Kaula, *Theory of Satellite Geodesy*, Blaisdell Publishing Company, Waltham, Massachusetts, 1966; reprint: Dover Publications, Mineola, New York, 2000.
- D.G. King-Hele, *Theory of Satellite Orbits in an Atmosphere*, Butterworths, London, 1964.
- C. Marchal, *The Three-Body Problem*, Elsevier Science Publishers, Amsterdam, 1990.
- S.W. McCuskey, *Introduction to Celestial Mechanics*, Addison-Wesley, Reading, Massachusetts, 1963.
- J.E. McGolrick, *Launch Vehicle Estimating Factors for Advance Mission Planning*, NHB 7100 .5B, NASA, Washington, 1973.
- W.G. Melbourne, *Interplanetary Trajectories and Payload Capabilities of Advanced Propulsion Vehicles*, Technical Report No. 32-68, Jet Propulsion Laboratory, Pasadena, 1961.
- O. Montenbruck, E. Gill, *Satellite Orbits - Models, Methods, Applications*, Springer Verlag, Berlin, 2001.
- F.R. Moulton, *An Introduction to Celestial Mechanics*, The Macmillan Company, New York, 1923; reprint: Dover Publications, Mineola, New York, 1970.
- H.C. Plummer, *An Introductory Treatise on Dynamical Astronomy*, Cambridge University Press, Cambridge, 1918; reprint: Dover Publications, Mineola, New York, 1960.
- H. Pollard, *Mathematical Introduction to Celestial Mechanics*, Prentice-Hall, Englewood Cliffs, New Jersey, 1966.

- J.E. Prussing, B.A. Conway, *Orbital Mechanics*, Oxford University Press, New York, 1993.
- F.P.J. Rimrott, *Introductory Orbit Dynamics*, Friedrich Vieweg & Sohn, Braunschweig, 1989.
- A.E. Roy, *The Foundations of Astrodynamics*, The Macmillan Company, New York, 1965.
- A.E. Roy, *Orbital Motion*, John Wiley & Sons, New York, 1978.
- W.M. Smart, *Celestial Mechanics*, Longmans, Green and Co., London, 1960.
- T.E. Sterne, *An Introduction to Celestial Mechanics*, Interscience Publishers, New York, 1960.
- E.L. Stiefel, G. Scheifele, *Linear and Regular Celestial Mechanics*, Springer Verlag, New York, 1971.
- E. Stuhlinger, *Ion Propulsion for Space Flight*, McGraw-Hill, New York, 1964.
- V. Szebehely, *Theory of Orbits, The Restricted Problem of Three Bodies*, Academic Press, New York, 1967.
- V. Szebehely, *Adventures in Celestial Mechanics, A First Course in the Theory of Orbits*, University of Texas Press, Austin, 1991.
- L.G. Taff, *Celestial Mechanics, A Computational Guide for the Practitioner*, John Wiley & Sons, New York, 1985.
- W.T. Thomson, *Introduction to Space Dynamics*, Dover Publications, Mineola, New York, 1986.
- D.A. Vallado, *Fundamentals of Astrodynamics and Applications*, Microcosm Press, El Segundo, and Kluwer Academic Publishers, Dordrecht, 2001.
- J.F. White, *Flight Performance Handbook for Powered Flight Operations*, John Wiley & Sons, New York, 1963.
- E.T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies*, Cambridge University Press, Cambridge, 1917; reprint 1988.
- W.E. Wiesel, *Spaceflight Dynamics*, McGraw-Hill, New York, 1989.
- Wikipedia, Various articles on the historical development of dynamics, celestial mechanics and cosmology.

APPENDIX B: ASTRONOMICAL AND GEOPHYSICAL CONSTANTS AND PARAMETERS¹

Table B.1: Selected astronomical constants and parameters.

Mean solar day = msd	$24^{\text{h}} 00^{\text{m}} 00^{\text{s}} = 86,400 \text{ s}$
Mean sidereal day	$23^{\text{h}} 56^{\text{m}} 4.091^{\text{s}} = 86,164.091 \text{ s}$
Synodic month (msd) (new Moon to new Moon)	29.530589
Tropical month (msd) (equinox to equinox)	27.321582
Sidereal month (msd) (fixed star to fixed star)	27.321661
Anomalistic month (msd) (perigee to perigee)	27.554550
Draconic month (msd) (node to node)	27.212220
Tropical year (msd) (equinox to equinox)	365.242190
Sidereal year (msd) (fixed star to fixed star)	365.256363
Anomalistic year (msd) (perihelion to perihelion)	365.259636
Julian year (msd) (based on Julian calendar)	365.25
Gregorian year (msd) (based on Gregorian calendar)	365.2425
Universal gravitational constant ($\text{km}^3/(\text{kg s}^2)$)	6.67428×10^{-20}
Speed of light (km/s)	2.99792458×10^5
Light-year (km) = ly	$9.460730473 \times 10^{12}$
Mass Sun (kg)	1.9886×10^{30}
Mean radius Sun (km)	6.95508×10^5
Solar luminosity (W)	3.839×10^{26}
Distance Sun from galactic center (ly)	27,000
Velocity Sun about galactic center (km/s)	220
Orbital period Sun about galactic center (yr)	240×10^6
Distance Sun-Earth (km)	$147.10 \times 10^6 - 152.10 \times 10^6$
Astronomical Unit (km) = AU	$149.597870700 \times 10^6$
Solar constant (W/m^2)	1361
Inclination invariable plane of Laplace to ecliptic (deg)	1.5786944
Obliquity of the ecliptic (J2000, deg)	23.439279
Sun/Earth mass ratio	$3.329460487 \times 10^{-5}$
Mass Earth (kg) = M_{\oplus}	5.9736×10^{24}
Mean radius Earth (km)	6371.00
Earth/Moon mass ratio	81.30059
Mass Moon (kg)	7.348×10^{22}
Mean radius Moon (km)	1737.1
Distance Earth-Moon (km)	363,100 - 405,700
Mean distance Earth-Moon (km)	384,401
Mean dist. Earth-Moon barycenter from center Earth (km)	4671
Mean inclination lunar orbit to ecliptic (deg)	5.145396
Inclination lunar equator to lunar orbit (deg)	6.6878
Total mass entire planetary system (M_{\oplus})	446.7
Total mass of planets (M_{\oplus})	446.6
Total angular momentum of planetary system ($\text{kg km}^2/\text{s}$)	3.148×10^{37}
Total translational kinetic energy of planetary system (J)	1.99×10^{35}
Total rotational energy of planets (J)	0.7×10^{35}

¹ Data taken from: A.N. Cox (ed.), *Allen's Astrophysical Quantities*, Springer Verlag, New York, 1999; NASA National Space Science Data Center (NSSDC), <http://nssdc.gsfc.nasa.gov>; NASA JPL Solar System Dynamics (SSD), <http://ssd.jpl.nasa.gov>; The Astronomical Almanac Online, 2014, <http://asa.usno.navy.mil>. Last updated: February 5, 2010; minor corrections: August 2014.

Table B.2: Physical parameters of the Earth.

<i>Mass of layers</i>	
Core (kg)	1.93×10^{24}
Mantle (kg)	4.04×10^{24}
Oceans (kg)	1.4×10^{21}
Atmosphere (kg)	5.1×10^{18}
<i>Surface area</i>	
Land area (km^2)	1.48×10^8
Sea area (km^2)	3.62×10^8
<i>Reference ellipsoid</i>	
Mean equatorial radius, a_e (km)	6378.1366
Polar radius, a_p (km)	6356.7519
Mean radius (km) = $(a_e^2 a_p)^{1/3}$	6371.00
Flattening = $(a_e - a_p)/a_e$	$1/298.2570$
Eccentricity = $(a_e^2 - a_p^2)^{1/2}/a_e$	0.081819
<i>Moments of inertia</i>	
About rotation axis, C (kg km^2)	8.0358×10^{31}
Largest about equatorial axis, B (kg km^2)	8.0096×10^{31}
Smallest about equatorial axis, A (kg km^2)	8.0094×10^{31}
Dynamical flattening = $\{C - (A + B)/2\}/C$	0.00327285
Mantle about rotation axis (kg km^2)	7.0400×10^{31}
Mantle about equatorial axis (kg km^2)	7.0165×10^{31}
Core dynamical flattening	0.002646
<i>Rotation</i>	
Rotational period with respect to fixed stars (s)	86,164.091
Mean angular rotational velocity (rad/s)	7.292115×10^{-5}
Equatorial rotational velocity (km/s)	0.46510
Centrifugal acceleration at equator (m/s^2)	3.39157×10^{-2}
Angular momentum ($\text{km}^2 \text{ kg/s}$)	5.8598×10^{27}
Rotational energy (J)	2.1365×10^{29}
Chandler period (msd)	433.1
Free core nutation period (msd)	430.2
<i>Gravity</i>	
Gravitational parameter (km^3/s^2)	398,600.4418
Equatorial gravity on reference ellipsoid (m/s^2)	9.78033
Polar gravity on reference ellipsoid (m/s^2)	9.83219
<i>Atmosphere at surface</i>	
Standard temperature (K)	273.15
Standard pressure (Pa)	1.013250×10^5
Mass density (kg/m^3)	1.2928
Molecular weight (kg/mole)	28.964×10^{-3}
Molecular rms velocity (m/s)	4.850×10^2
Specific heat at constant pressure, c_p (J/(kg K))	1005
Specific heat at constant volume, c_v (J/(kg K))	717.6
Speed of sound (m/s)	3.313×10^2

Table B.3: Ratio of the mass of the Sun and the mass of a planetary system.

Mercury	6023600
Venus	408523.71
Earth + Moon	328900.5614
Mars system	3098708
Jupiter system	1047.3486
Saturn system	3497.898
Uranus system	22902.98
Neptune system	19412.24
Pluto system	1.366×10^8

Table B.4: Comparison of physical parameters of the Sun, Earth and Moon.

	Sun	Earth	Moon
Mean radius (km)	695,508	6371.00	1737.1
Mass (kg)	1.9886×10^{30}	5.9736×10^{24}	7.348×10^{22}
Mean density (gr/cm ³)	1.408	5.5151	3.341
Gravitational parameter (km ³ /s ²)	$1.327124421 \times 10^{11}$	398,600.4418	4902.801
Rotational period (msd)	25.38	0.997270	27.321661
Equatorial rotational velocity (km/s)	1.9874	0.4651	0.00464
Surface equatorial gravity (m/s ²)	274.0	9.7803	1.622
Escape velocity at equator (km/s)	617.8	11.18	2.376

Solar rotational period refers to 26° N latitude.

Table B.5: Planetary orbital elements. They refer to the mean ecliptic and equinox of January 1, 2000, and describe an ellipse with the Sun at a focal point (heliocentric reference frame).

Planet	<i>a</i> (AU)	<i>e</i>	<i>i</i> (deg)	Ω (deg)	ω (deg)	M (deg)
Mercury	0.387099	0.205636	7.0050	48.3308	29.1270	174.7925
Venus	0.723336	0.006777	3.3947	76.6798	54.9223	50.3766
Earth	1.000000	0.016710	0.0000	348.7393	114.2078	357.5171
Mars	1.523710	0.093394	1.8497	49.5595	286.4968	19.3902
Jupiter	5.202887	0.048386	1.3044	100.4739	274.2545	19.6680
Saturn	9.536676	0.053862	2.4860	113.6624	338.9364	317.3553
Uranus	19.189165	0.047257	0.7726	74.0169	96.9373	142.2838
Neptune	30.069923	0.008590	1.7700	131.7842	273.1805	259.9152
Pluto	39.482117	0.248827	17.1400	110.3039	113.7649	14.8601

Table B.6: Planetary orbital and physical parameters.

	Mercury	Venus	Earth	Mars	Jupiter	Saturn	Uranus	Neptune	Pluto
Apohelion distance (AU)	0.46670	0.72824	1.01671	1.66601	5.45463	10.0503	20.0960	30.3282	49.3063
Perihelion distance (AU)	0.30750	0.71843	0.98329	1.38140	4.95114	9.02301	18.2823	29.8116	29.6579
Orbital period (sidereal year)	0.24084	0.61519	1	1.88085	11.8677	29.4507	84.0591	164.892	248.085
Mean orbital velocity (km/s)	47.872	35.021	29.785	24.129	13.058	9.645	6.799	5.432	4.740
Equatorial radius (km)	2439.7	6051.9	6378.14	3396.2	71,492	60,268	25,559	24,764	1195
Oblateness	0.00000	0.00000	0.00335	0.00589	0.06487	0.09796	0.02293	0.01708	0.00000
Mass (M_{\oplus})	0.055274	0.81500	1	0.10744	317.83	95.152	14.536	17.147	0.00219
Mean density (gr/cm ³)	5.427	5.204	5.5151	3.934	1.326	0.687	1.270	1.638	2.03
Gravity acceleration at equator (g_{\oplus})	0.378	0.905	1	0.379	2.530	1.066	0.905	1.137	0.0674
Equatorial escape velocity (km/s)	4.25	10.36	11.18	5.02	59.53	35.48	21.29	23.49	1.23
Rotation period (msd)	58.6462	-243.0185	0.99727	1.02596	0.41354	0.44402	-0.71833	0.67125	-6.38723
Length of day (hr)	4221.6	2803.2	24	24.6597	9.9259	10.656	17.24	16.11	153.2820
Inclin. equator to orbit plane (deg)	0.01	177.3	23.44	25.19	3.13	26.73	97.77	29.58	119.59
Number of moons	0	0	1	2	65	62	27	14	5
Ring system	no	No	No	no	yes	yes	yes	yes	No

Table B.7: Planetary satellite parameters².

Satellite	<i>a</i> (10 ³ km)	<i>E</i>	<i>i</i> (deg)	<i>T</i> (msd)	<i>R</i> (km)	<i>M/M_{pl}</i>	ρ (gr/cm ³)
<i>Earth</i>							
Moon	384.401	0.0554	18.29-28.58	27.3216	1737.1	1.230×10 ⁻²	3.341
<i>Mars</i>							
Phobos	9.376	0.0151	1.075	0.3189	NS13.4	1.651×10 ⁻⁸	1.872
Deimos	23.458	0.0002	1.788	1.2624	NS7.5	3.739×10 ⁻⁹	1.471
<i>Jupiter</i>							
Io	421.8	0.0041	0.036	1.76914	NS1830	4.70×10 ⁻⁵	3.528
Europa	671.1	0.0094	0.466	3.5512	1561	2.53×10 ⁻⁵	3.013
Ganymede	1070.4	0.0013	0.177	7.1545	2632	7.80×10 ⁻⁵	1.936
Callisto	1882.7	0.0074	0.192	16.689	2410	5.67×10 ⁻⁵	1.834
Amalthea	181.4	0.0032	0.380	0.4982	NS131	1.09×10 ⁻⁹	0.849
Himalia	11461	0.1623	27.496	250.56	85	3.55×10 ⁻⁹	2.6
Elara	11741	0.2174	26.627	259.64	43	4.6×10 ⁻¹⁰	2.6
Pasiphae	23624	0.4090	151.431	R743.63	30	1.6×10 ⁻¹⁰	2.6
Sinope	23939	0.2495	158.109	R758.90	19	3.9×10 ⁻¹¹	2.6
Lysithea	11717	0.1124	28.302	259.20	18	3.3×10 ⁻¹¹	2.6
Carme	23404	0.2533	164.907	R734.17	23	6.9×10 ⁻¹¹	2.6
Ananke	21276	0.2435	148.889	R629.77	14	1.6×10 ⁻¹¹	2.6
Leda	11165	0.1636	27.457	240.92	10	5.8×10 ⁻¹²	2.6
Thebe	221.9	0.0176	1.080	0.6745	NS55	7.9×10 ⁻¹⁰	3.0
Adrastea	129.0	0.0018	0.054	0.2983	NS13	4×10 ⁻¹²	3.0
Metis	128.0	0.0012	0.019	0.2948	21	6×10 ⁻¹¹	3.0
<i>Saturn</i>							
Mimas	185.539	0.0196	1.574	0.9424	NS209	6.60×10 ⁻⁸	1.150
Enceladus	238.037	0.0047	0.009	1.3702	NS256	1.90×10 ⁻⁷	1.608
Tethys	294.672	0.0001	1.091	1.8878	NS536	1.09×10 ⁻⁶	1.15
Dione	377.415	0.0022	0.028	2.7369	562	1.93×10 ⁻⁶	1.476
Rhea	527.068	0.0010	0.333	4.5175	764	4.06×10 ⁻⁶	1.233
Titan	1221.86	0.0288	0.312	15.9454	2576	2.37×10 ⁻⁴	1.880
Hyperion	1500.93	0.0232	0.615	21.2766	NS185	9.83×10 ⁻⁹	0.542
Iapetus	3560.85	0.0293	8.313	79.3302	736	3.18×10 ⁻⁶	1.083
Phoebe	12947.9	0.1634	175.243	R550.30	107	1.46×10 ⁻⁸	1.634
Janus	151.460	0.0068	0.163	0.6945	NS97	3.34×10 ⁻⁹	0.634
Epimetheus	151.410	0.0098	0.351	0.6942	NS69	9.26×10 ⁻¹⁰	0.689
Helene	377.420	0.0071	0.213	2.7369	NS18	4.5×10 ⁻¹¹	1.5
Telesto	294.710	0.0002	1.180	1.8878	NS15	1.3×10 ⁻¹¹	1.0
Calypso	294.710	0.0005	1.499	1.8878	NS15	6.3×10 ⁻¹²	1.0
Atlas	137.670	0.0012	0.003	0.6019	NS18	1.2×10 ⁻¹¹	0.437
Prometheus	139.380	0.0022	0.008	0.6130	NS74	2.8×10 ⁻¹⁰	0.475
Pandora	141.720	0.0042	0.050	0.6285	NS55	2.4×10 ⁻¹⁰	0.500
Pan	133.580	0.0000	0.001	0.5750	15	8.7×10 ⁻¹²	0.360
<i>Uranus</i>							
Ariel	190.900	0.0012	0.041	2.5204	NS581	1.49×10 ⁻⁵	1.592
Umbriel	266.000	0.0039	0.128	4.1442	585	1.41×10 ⁻⁵	1.459

² For Jupiter to Neptune only a selected set of satellites is included.

Titania	436.300	0.0011	0.079	8.7059	789	4.87×10^{-5}	1.662
Oberon	583.500	0.0014	0.068	13.4632	761	3.32×10^{-5}	1.559
Miranda	129.900	0.0013	4.338	1.4135	NS240	7.6×10^{-7}	1.214
Cordelia	49.80	0.0003	0.085	0.3350	20	5.2×10^{-10}	1.3
Ophelia	53.80	0.0099	0.104	0.3764	21	6.2×10^{-10}	1.3
Bianca	59.17	0.0009	0.193	0.4346	26	1.1×10^{-9}	1.3
Cressida	61.78	0.0004	0.006	0.4636	40	3.95×10^{-9}	1.3
Desdemona	62.68	0.0001	0.113	0.4736	32	2.05×10^{-9}	1.3
Juliet	64.35	0.0007	0.065	0.4931	47	6.42×10^{-9}	1.3
Portia	66.09	0.0001	0.059	0.5132	68	1.94×10^{-8}	1.3
Rosalind	69.94	0.0001	0.279	0.5585	36	2.93×10^{-9}	1.3
Belinda	75.26	0.0001	0.031	0.6235	40	4.11×10^{-9}	1.3
Puck	86.01	0.0001	0.319	0.7618	81	3.33×10^{-8}	1.3
Caliban	7231.1	0.1812	141.53	R579.73	36	3.4×10^{-9}	1.5
Sycorax	12179.4	0.5219	159.42	R1288.38	75	3.1×10^{-8}	1.5
Neptune							
Triton	354.759	0.0000	156.865	R5.8768	1352.6	2.09×10^{-4}	2.061
Nereid	5513.82	0.7507	7.090	360.136	170	3.01×10^{-7}	1.5
Naïad	48.227	0.0003	4.691	0.2944	NS48	1.9×10^{-9}	1.3
Thalassa	50.074	0.0002	0.135	0.3115	NS54	3.6×10^{-9}	1.3
Despina	52.526	0.0002	0.068	0.3347	NS90	2.0×10^{-8}	1.3
Galatea	61.953	0.0001	0.034	0.4287	NS102	3.7×10^{-8}	1.3
Larissa	73.548	0.0014	0.205	0.5546	NS108	4.8×10^{-8}	1.3
Proteus	117.646	0.0005	0.075	1.1223	NS220	4.9×10^{-7}	1.3
Pluto							
Charon	17.536	0.0022	0.001	6.3872	604	1.30×10^{-1}	1.68
Nix	48.708	0.0030	0.195	24.856	44	7×10^{-6}	--
Hydra	64.749	0.0051	0.212	38.206	36	7×10^{-6}	--
Kerberos	59	0	0	32.1	15	--	--
Styx	42	0	0	20.2	10	--	--

R before the orbital period indicates a retrograde orbit.

NS before the radius indicates that the satellite is known to be non-spherical; the largest radius is listed.

Table B.8: Parameters of five dwarf planets.

	a (AU)	e	i (deg)	T (year)	R (km)	ρ (gr/cm ³)
Ceres	2.766	0.0789	10.587	4.599	476	2.08
Pluto	39.482	0.2488	17.140	247.921	1148	2.0
Haumea	43.132	0.1950	28.22	283.3	650	2.6
Makemake	45.792	0.159	28.96	309.9	750	2.0
Eris	67.668	0.4418	44.187	557	1163	2.25

Table B.9: Parameters of the 45 largest asteroids.

Name	No.	Yr. disc.	D (km)	a (AU)	e	i (deg)	P (hr)
Ceres ³	1	1801	975	2.766	0.0789	10.587	9.074
Pallas	2	1802	582	2.773	0.2310	34.840	7.811
Vesta	4	1807	578	2.362	0.0887	7.134	5.342
Hygieia	10	1849	530	3.140	0.1169	3.841	18.4
Sylvia	87	1866	385	3.485	0.0812	10.858	5.183
Hektor	624	1907	370	5.237	0.0223	18.182	6.921
Europa	52	1858	360	3.096	0.1062	7.482	5.631
Eunomia	15	1851	357	2.642	0.1881	11.739	6.083
Davida	511	1903	357	3.167	0.1861	15.937	5.13
Interamnia	704	1910	350	3.061	0.1508	17.290	8.727
Juno	3	1804	320	2.670	0.2550	12.982	7.210
Eugenia	45	1857	305	2.720	0.0813	6.610	5.699
Cybele	65	1861	302	3.434	0.1082	3.562	6.07
Camilla	107	1868	285	3.478	0.0768	10.050	4.84
Chariklo	10199	1997	258	15.778	0.1710	23.377	--
Euphrosyne	31	1854	256	3.150	0.2248	26.310	5.531
Themis	24	1853	249	3.129	0.1313	0.759	8.38
Iris	7	1847	240	2.387	0.2309	5.523	7.139
Psyche	16	1852	240	2.923	0.1384	3.099	4.196
Amphitrite	29	1854	233	2.554	0.0733	6.097	5.39
Chiron	2060	1977	233	13.704	0.3788	6.930	5.9
Thisbe	88	1866	232	2.768	0.1655	5.215	6.042
Bamberga	324	1892	229	2.684	0.3371	11.108	29.43
Fortuna	19	1852	225	2.442	0.1576	1.573	7.445
Aurora	94	1867	225	3.161	0.0883	7.965	7.22
Patientia	451	1899	225	3.059	0.0775	15.223	9.727
Doris	48	1857	222	3.107	0.0750	6.556	11.89
Herculina	532	1904	222	2.772	0.1781	16.312	9.405
Egeria	13	1850	217	2.576	0.0858	16.544	7.045
Hermione	121	1872	217	3.445	0.1355	7.599	6.1
Ursula	375	1893	216	3.123	0.1074	15.950	16.83
Diotima	423	1896	209	3.066	0.0397	11.233	4.622
Alauda	702	1910	202	3.193	0.0213	20.611	8.36
Palma	372	1893	195	3.144	0.2618	23.864	6.58
Nemesis	128	1872	194	2.750	0.1247	6.249	39
Hebe	6	1847	192	2.425	0.2025	14.754	7.274
Bertha	154	1875	192	3.188	0.0832	21.033	5.11
Freia	76	1862	190	3.416	0.1632	2.117	9.98
Elektra	130	1873	189	3.123	0.2090	22.873	5.225
Kalliope	22	1852	187	2.909	0.1024	13.707	4.147
Aletheia	259	1886	185	3.143	0.1242	10.801	15
Berbericia	776	1914	183	2.934	0.1615	18.245	7.672
Daphne	41	1856	182	2.764	0.2732	15.788	5.988
Metis	9	1848	179	2.387	0.1222	5.574	5.078
Lachesis	120	1872	178	3.116	0.0584	6.954	--

D: diameter; for non-spherical objects the largest diameter is listed.*P*: rotational period.³ Since 2006, Ceres is classified as a dwarf planet (Table B.8).

Table B.10: Parameters of 45 near-Earth asteroids.

Name	No.	D (km)	a (AU)	e	r _p (AU)	i (deg)
Eros	433	23	1.458	0.2228	1.133	10.829
Albert	719	2	2.627	0.5528	1.175	11.561
Alinda	887	4	2.479	0.5669	1.074	9.353
Ganymed	1036	38	2.665	0.5340	1.242	26.686
Amor	1221	1	1.920	0.4343	1.086	11.877
Icarus	1566	1.3	1.078	0.8269	0.187	22.837
Betulia	1580	4.6	2.196	0.4877	1.125	52.092
Geographos	1620	3	1.246	0.3355	0.828	13.338
Ivar	1627	9	1.863	0.3969	1.124	8.448
Toro	1685	3	1.367	0.4358	0.771	9.379
Apollo	1862	1.4	1.470	0.5599	0.647	6.353
Antinous	1863	2	2.259	0.6066	0.889	18.401
Daedalus	1864	3	1.461	0.6146	0.563	22.200
Cerberus	1865	1	1.080	0.4670	0.576	16.094
Sisyphus	1866	9	1.894	0.5386	0.873	41.183
Quetzalcoatl	1915	0.4	2.542	0.5716	1.089	20.417
Boreas	1916	3	2.274	0.4491	1.253	12.877
Cuyo	1917	5	2.151	0.5042	1.066	23.940
Anteros	1943	2	1.430	0.2559	1.064	8.704
Tezcatlipoca	1980	7	1.710	0.3649	1.086	26.858
Midas	1981	2	1.776	0.6500	0.622	39.838
Baboquivari	2059	3	2.644	0.5300	1.243	11.039
Anza	2061	2	2.265	0.5373	1.048	3.772
Aten	2062	1	0.967	0.1826	0.790	18.933
Bacchus	2063	1	1.078	0.3495	0.701	9.433
Ra-Shalom	2100	2.3	0.832	0.4365	0.469	15.759
Adonis	2101	1	1.874	0.7640	0.442	1.334
Tantalus	2102	3	1.290	0.2990	0.904	64.006
Aristaeus	2135	1	1.599	0.5033	0.794	23.057
Oljato	2201	2	2.175	0.7132	0.624	2.514
Pele	2202	1	2.290	0.5120	1.118	8.742
Hephaistos	2212	3	2.167	0.8338	0.360	11.736
Orthos	2329	4	2.404	0.6575	0.823	24.436
Hathor	2340	0.3	0.844	0.4497	0.464	5.855
Beltrovata	2368	2.3	2.105	0.4138	1.234	5.236
Seneca	2608	1	2.504	0.5762	1.061	14.991
Krok	3102	1.6	2.151	0.4485	1.186	8.420
Eger	3103	2.5	1.405	0.3544	0.907	20.933
Florence	3122	2.5	1.768	0.4229	1.020	22.165
Nefertiti	3199	2	1.574	0.2840	1.127	32.971
Phaethon	3200	5	1.271	0.8899	0.140	22.183
Ul	3271	2	2.102	0.3957	1.270	25.040
Seleucus	3288	3	2.032	0.4575	1.102	5.934
McAuliffe	3352	2	1.878	0.3696	1.184	4.774
Syrinx	3360	2	2.467	0.7438	0.632	21.416

D: diameter; for non-spherical objects the mean diameter is listed.

Table B.11: Parameters of 45 selected comets having more than one known apparition.

Name	No.	τ (ymd)	T (year)	r_p (AU)	e	ω (deg)	Ω (deg)	i (deg)
Encke	2P	19970524	3.28	0.33	0.849	186.4	334.7	11.9
Wilson-Harrington	107P	20091022	4.29	0.99	0.624	91.2	270.6	2.8
Grigg-Skjellerup	26P	19970830	5.10	1.00	0.664	359.3	213.3	21.1
du Toit-Hartley	79P	20030215	5.21	1.23	0.594	253.1	308.0	2.9
Machholz	96P	20020109	5.24	0.12	0.959	14.6	94.6	60.2
Schwassmann-Wachmann 3	73P	19950923	5.35	0.93	0.695	198.8	69.9	11.4
Tuttle-Giacobini-Kresak	41P	20010107	5.46	1.05	0.659	62.2	141.1	9.2
Tempel 2	10P	19990908	5.48	1.48	0.523	195.0	118.2	12.0
Tempel 1	9P	20050705	5.50	1.51	0.517	178.9	68.9	10.5
Howell	88P	19930226	5.58	1.41	0.552	234.8	57.7	4.4
Tempel-Swift-LINEAR	11P	20011231	6.37	1.58	0.539	163.6	240.7	13.5
Russell 1	83P	19850705	6.10	1.61	0.517	0.4	230.8	22.7
Forbes	37P	20050802	6.13	1.57	0.541	329.3	315.1	9.0
Hartley 2	103P	19971222	6.26	1.03	0.700	180.7	219.9	13.6
Wild 2	81P	19970507	6.37	1.58	0.540	41.8	136.1	3.2
d'Arrest	6P	20020204	6.51	1.35	0.613	178.1	138.9	19.5
West-Kohoutek-Ikemura	76P	20061120	6.41	1.60	0.538	0.1	84.1	30.5
Kopff	22P	20021212	6.46	1.58	0.543	162.8	120.9	4.7
Giacobini-Zinner	21P	19981121	6.61	1.03	0.706	172.5	195.4	31.9
Kohoutek	75P	19940630	6.65	1.78	0.496	175.8	269.7	5.9
Arend-Rigaux	49P	19980713	6.82	1.37	0.611	330.6	121.7	18.3
Gunn	65P	20100302	6.84	2.44	0.319	196.6	68.4	10.4
Borrelly	19P	20010915	6.86	1.36	0.624	353.4	75.4	30.3
Wild 3	86P	19940721	6.91	2.30	0.366	179.3	72.6	15.5
Longmore	77P	20020905	7.00	2.31	0.358	196.4	15.0	24.4
Shoemaker 1	102P	20060607	7.26	1.97	0.472	18.5	339.9	26.2
Russell 2	89P	19941027	7.38	2.28	0.400	249.2	42.5	12.0
Shajn-Schaldach	61P	20010509	7.49	2.33	0.389	216.6	166.9	6.1
Harrington-Abell	52P	19990128	7.59	1.76	0.543	138.9	337.3	10.2
Gehrels 2	78P	20041027	7.22	2.01	0.462	193.0	210.5	6.2
Whipple	36P	20030707	8.53	3.09	0.259	202.2	182.4	9.9
Denning-Fujikawa	72P	19781002	9.01	0.78	0.820	334.3	41.5	8.6
Swift-Gehrels	64P	20090614	9.21	1.38	0.690	96.3	300.7	9.0
Gale	34D	19380618	11.0	1.18	0.761	209.2	67.9	11.7
Wild 1	63P	19991227	13.3	1.96	0.650	167.9	358.5	19.9
Schwassmann-Wachmann 1	29P	20040630	14.6	5.72	0.045	48.2	312.6	9.4
du Toit	66P	20030828	14.7	1.27	0.788	257.2	22.2	18.7
Neujmin 1	28P	20021227	18.2	1.55	0.775	346.9	347.0	14.2
Crommelin	27P	19840220	27.4	0.73	0.919	195.8	250.9	29.1
Tempel-Tuttle	55P	19980228	32.9	0.98	0.906	172.5	235.3	162.5
Westphal	20D	19131127	61.9	1.25	0.920	57.1	348.0	40.9
Olbers	13P	19560619	69.6	1.18	0.930	64.6	86.1	44.6
Halley	1P	19860206	76.0	0.59	0.967	111.3	58.4	162.3
Swift-Tuttle	109P	19921212	135	0.96	0.963	153.0	139.4	113.4
Herschel-Rigollet	35P	19390809	155	0.75	0.974	29.3	356.0	64.2

τ is the perihelion date; first 4 digits indicate the year, next 2 digits the month, last 2 digits the day.

ω and Ω are the longitude of perihelion and the longitude of the ascending node, respectively.

APPENDIX C: COMPILATION OF SOME EXPRESSIONS FOR ELLIPTICAL, PARABOLIC AND HYPERBOLIC ORBITS.

	Ellipse	Parabola	Hyperbola
r	$\frac{p}{1 + e \cos \theta}$	$\frac{p}{1 + \cos \theta}$	$\frac{p}{1 + e \cos \theta}$
e	$0 \leq e < 1$	$e = 1$	$e > 1$
a	$0 \leq a < \infty$	$a = \infty$	$a < 0$
n	$\sqrt{\frac{\mu}{a^3}}$	$\sqrt{\frac{\mu}{p^3}}$	$\sqrt{\frac{\mu}{-a^3}}$
r_{min}	$a(1 - e)$	$p/2$	$a(1 - e)$
r_{max}	$a(1 + e)$	∞	∞
θ	$0^\circ \leq \theta \leq 360^\circ$	$0^\circ \leq \theta \leq 360^\circ$	$\cos \theta \geq -\frac{1}{e}$
γ	$0 \leq \gamma \leq \arctan\left(\frac{e}{\sqrt{1-e^2}}\right)$	$-90^\circ \leq \gamma \leq 90^\circ$	$-90^\circ \leq \gamma \leq 90^\circ$
V^2	$\mu\left(\frac{2}{r} - \frac{1}{a}\right)$	$\frac{2\mu}{r}$	$\mu\left(\frac{2}{r} - \frac{1}{a}\right)$
V_{min}	$\sqrt{\frac{\mu}{a}} \frac{(1-e)}{(1+e)}$	0	$\sqrt{\frac{\mu}{-a}}$
V_{max}	$\sqrt{\frac{\mu}{a}} \frac{(1+e)}{(1-e)}$	$2\sqrt{\frac{\mu}{p}}$	$\sqrt{\frac{\mu}{-a}} \frac{(e+1)}{(e-1)}$
\mathcal{E}	$\mathcal{E} < 0$	$\mathcal{E} = 0$	$\mathcal{E} > 0$
E, F	$\tan \frac{E}{2} = \sqrt{\frac{1-e}{1+e}} \tan \frac{\theta}{2}$	—	$\tanh \frac{F}{2} = \sqrt{\frac{e-1}{e+1}} \tan \frac{\theta}{2}$
r	$a(1 - e \cos E)$	—	$a(1 - e \cosh F)$
M, \bar{M}	$\sqrt{\frac{\mu}{a^3}}(t - \tau)$	$\sqrt{\frac{\mu}{p^3}}(t - \tau)$	$\sqrt{\frac{\mu}{-a^3}}(t - \tau)$
t	$M = E - e \sin E$	$2\bar{M} = \tan \frac{\theta}{2} + \frac{1}{3} \tan^3 \frac{\theta}{2}$	$\bar{M} = e \sinh F - F$

