AIAA Education Series

VLADIMIR A. CHOBOTOV, EDITOR

ORBITAL MECHANICS
THIRD EDITION

AIAA

Education Series

J.S. PRZEMIENIECKI
EDITOR-IN-CHIEF

Orbital Mechanics
Third Edition

Edited by
Vladimir A. Chobotov

EDUCATION SERIES
J. S. Przemieniecki
Series Editor-in-Chief
Air Force Institute of Technology
Wright-Patterson Air Force Base, Ohio

Published by
American Institute of Aeronautics and Astronautics, Inc.
1801 Alexander Bell Drive, Reston, Virginia 20191–4344

American Institute of Aeronautics and Astronautics, Inc., Reston, Virginia

Library of Congress Cataloging-in-Publication Data

Orbital mechanics / edited by Vladimir A. Chobotov.—3rd ed.
p. cm.—(AIAA education series)
Includes bibliographical references and index.
1. Orbital mechanics. 2. Artificial satellites—Orbits.
3. Navigation (Astronautics). I. Chobotov, Vladimir A. II. Series.
TL1050.O73 2002 629.4′113—dc21
ISBN 1-56347-537-5 (hardcover : alk. paper) 2002008309

Copyright © 2002 by the American Institute of Aeronautics and Astronautics, Inc. All rights reserved. Printed in the United States of America. No part of this publication may be reproduced, distributed, or transmitted, in any form or by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

Data and information appearing in this book are for informational purposes only. AIAA and the authors are not responsible for any injury or damage resulting from use or reliance, nor does AIAA or the authors warrant that use or reliance will be free from privately owned rights.

Foreword

The third edition of *Orbital Mechanics* edited by V. A. Chobotov complements five other space-related texts published in the Education Series of the American Institute of Aeronautics and Astronautics (AIAA): *Re-Entry Vehicle Dynamics* by F. J. Regan, *An Introduction to the Mathematics and Methods of Astrodynamics* by R. H. Battin, *Space Vehicle Design* by M. D. Griffin and J. R. French, *Spacecraft Mission Design* and *Spacecraft Propulsion*, the last two written by C. D. Brown. The revised text on *Orbital Mechanics* is specifically designed as a teaching textbook with a significant amount of reference materials and problems for the practicing aerospace engineer, scientist, or mission planner. The revised edition includes now more recent developments in space exploration and the chapter on space debris was expanded to include new developments. Also a new CD-ROM software package is included, a useful trend particularly encouraged in the AIAA Education Series.

The authors of this text were a team of scientists and engineers from The Aerospace Corporation, one of the leading organizations in the U.S. space program. The text covers both the theory and application of earth orbits and interplanetary trajectories, orbital maneuvers, space rendezvous, orbit perturbations, and collision hazards associated with space debris. It represents a complete authoritative exposition of the present knowledge of orbital mechanics applications to the design of space probes and vehicles.

The AIAA Education Series of textbooks and monographs, inaugurated in 1984, embraces a broad spectrum of theory and application of different disciplines in aeronautics and astronautics, including aerospace design practice. The series also includes texts on defense science, engineering, and management. It serves as teaching texts as well as reference materials for practicing engineers, scientists, and managers. The complete list of textbooks published in the series (over seventy-five titles) can be found on the end pages of this volume.

J. S. Przemieniecki
Editor-in-Chief
AIAA Education Series

AIAA Education Series

Editor-in-Chief

John S. Przemieniecki
Air Force Institute of Technology (retired)

Editorial Advisory Board

Aaron R. Byerley
U.S. Air Force Academy

Daniel J. Biezad
California Polytechnic State University

Kajal K. Gupta
NASA Dryden Flight Research Center

John K. Harvey
Imperial College

David K. Holger
Iowa State University

Rakesh K. Kapania
Virginia Polytechnic Institute and State University

Brian Landrum
University of Alabama, Huntsville

Robert G. Loewy
Georgia Institute of Technology

Michael Mohaghegh
The Boeing Company

Dora Musielak
Arlington, Texas

Conrad F. Newberry
Naval Postgraduate School

David K. Schmidt
University of Colorado, Colorado Springs

Peter J. Turchi
Los Alamos National Laboratory

David M. Van Wie
Johns Hopkins University

Preface

An update of *Orbital Mechanics, Second Edition* has been made to include more recent developments in space exploration (e.g. Galileo, Cassini, Mars Odyssey missions). Also, the chapter on space debris was rewritten to reflect new developments in that area. Additional example problems for student exercises are presented in selected chapters.

A new software package is included on a CD-ROM to illustrate text material and to provide solutions to selected problems. The software package is presented in three folders on the CD-ROM. The first folder "HW Solutions" authored by J. Alekshun, written in Microsoft Visual C++, can be run from the CD-ROM. This folder presents a range of viewpoints and guidelines for solving selected problems in the text. In some cases calculators are provided for obtaining numerical results of broader scope than the problem statement. These solutions accept a more generalized span of initial conditions. They are useful in demonstrating cross-sensitivities between variables. Throughout the work, graphical illustrations appear where thought helpful in projecting vector relationships and spatial trajectories. The second folder entitled "Orbital Calculator" by Dr. E. T. Campbell, also written in C++, is automatically unzipped to the "C" drive. It provides an interactive environment for the generation of Keplerian orbits, orbital transfer maneuvers and animation of ellipses, hyperbolas and interplanetary orbits. The third and final software folder "Orbital Mechanics Solutions" by C. G. Johnson is written in C and Fortran. It must be copied to a folder on the "C" drive to run.

The new text material and the enhanced software package provide an up-to-date database and an improved numerical processing capability to facilitate teaching and text problem solutions. It is hoped that the Third Edition will be a useful textbook for students and a ready reference for the practicing professional in orbital mechanics.

V. A. Chobotov
Spring 2002

About the Authors

Vladimir A. Chobotov
Ph.D. in Mechanical Engineering, University of Southern California
Former Manager, Space Hazards Section, Astrodynamics Department
Instructor, UCLA, The Aerospace Corporation

Hans K. Karrenberg
Engineer's Degree in Aerospace Engineering, University of Southern California
Former Director, Astrodynamics Department
Lecturer, UCLA Extension, The Aerospace Corporation

Chia-Chun "George" Chao
Ph.D. in Astrodynamics, University of California, Los Angeles
Senior Engineering Specialist, Astrodynamics Department
Lecturer, UCLA Extension, The Aerospace Corporation

Jimmy Y. Miyamoto
M.S., University of California, Los Angeles
Former Manager, Orbit Analysis Section, Astrodynamics Department
Lecturer, UCLA Extension, The Aerospace Corporation

Thomas J. Lang
M.S. in Aeronautics and Astronautics, Massachusetts Institute of Technology
Director, Astrodynamics Department
Lecturer, UCLA Extension, The Aerospace Corporation

Jean A. Kechichian
Ph.D. in Aeronautics and Astronautics, Stanford University
Engineering Specialist, Astrodynamics Department
The Aerospace Corporation

Software Development

Cassandra G. Johnson
B.S. in Computer Science, California State University at Dominguez Hills
Former Member of the Technical Staff, Astrodynamics Department
The Aerospace Corporation

Joseph Alekshun, Jr.
M.S. in Aeronautics and Astronautics, Massachusetts Institute of Technology
Former Member of the Technical Staff, The Aerospace Corporation

Eric T. Campbell
Ph.D. in Aeronautics and Astronautics, Purdue University
Senior Member of the Technical Staff, Astrodynamics Department
The Aerospace Corporation

Table of Contents

Foreword v
Preface vii
About the Authors ix

Chapter 1. Basic Concepts 1
1.1 A Historical Perspective 1
1.2 Velocity and Acceleration 5
Problems 9
Selected Solutions 10

Chapter 2. Celestial Relationships 11
2.1 Coordinate Systems 11
2.2 Time Systems 17
References 20

Chapter 3. Keplerian Orbits 21
3.1 Newton's Universal Law of Gravitation 21
3.2 General and Restricted Two-Body Problem 21
3.3 Conservation of Mechanical Energy 23
3.4 Conservation of Angular Momentum 24
3.5 Orbital Parameters of a Satellite 25
3.6 Orbital Elements 28
References 31
Problems 31
Selected Solutions 33

Chapter 4. Position and Velocity as a Function of Time 35
4.1 General Relationships 35
4.2 Solving Kepler's Equation 40
4.3 A Universal Approach 55
4.4 Expressions with f and g 59
4.5 Summary of the Universal Approach 60
4.6 The Classical Element Set 61
4.7 The Rectangular Coordinate System 62
4.8 Modified Classical to Cartesian Transformation 62
4.9 Rectangular to Modified Classical Elements Transformation 66
4.10 The Spherical (ADBARV) Coordinate System 67

4.11 Rectangular to Spherical Transformation 68
4.12 Spherical to Rectangular Transformation 69
4.13 The Earth-Relative Spherical (LDBARV) Coordinate System 70
4.14 Geodetic and Geocentric Altitudes 71
4.15 Converting from Perigee/Apogee Radii to Perigee/Apogee Altitudes 76
4.16 Converting from Perigee/Apogee Altitudes to Perigee/ Apogee Radii 77
References 82
Problems 83
Selected Solutions 85

Chapter 5. Orbital Maneuvers 87
5.1 Orbital Energy 87
5.2 Single-Impulse Maneuvers 89
5.3 Single- and Two-Impulse Transfer Comparison for Coplanar Transfers Between Elliptic Orbits That Differ Only in Their Apsidal Orientation 92
5.4 Hohmann Transfer 94
5.5 The Bi-elliptic Transfer 96
5.6 Restricted Three-Impulse Plane Change Maneuver for Circular Orbits 99
5.7 General Three-Impulse Plane Change Maneuver for Circular Orbit 103
5.8 Hohmann Transfer with Split-Plane Change 104
5.9 Bi-elliptic Transfer with Split-Plane Change 107
5.10 Transfer Between Coplanar Elliptic Orbits 107
References 109
Problems 109
Selected Solutions 115

Chapter 6. Complications to Impulsive Maneuvers 117
6.1 N-Impulse Maneuvers 117
6.2 Fixed-Impulse Transfers 117
6.3 Finite-Duration Burns: Gravity Losses 126
6.4 Very Low Thrust Transfers 130
References 132
Problems 132
Selected Solutions 134

Chapter 7. Relative Motion in Orbit 135
7.1 Space Rendezvous 135
7.2 Terminal Rendezvous 155
7.3 Applications of Rendezvous Equations 162

7.4 An Exact Analytical Solution for Two-Dimensional Relative Motion 172
7.5 Optimal Multiple-Impulse Rendezvous 177
References 181
Problems 182
Selected Solutions 183

Chapter 8. Introduction to Orbit Perturbations 185
8.1 A General Overview of Orbit Perturbations 185
8.2 Earth Gravity Harmonics 186
8.3 Lunisolar Gravitational Attractions 187
8.4 Radiation Pressure Effects 188
8.5 Atmospheric Drag 189
8.6 Tidal Friction Effects and Mutual Gravitational Attraction 190
References 192

Chapter 9. Orbit Perturbations: Mathematical Foundations 193
9.1 Equations of Motion 193
9.2 Methods of Solution 195
9.3 Potential Theory 202
9.4 More Definitions of Gravity Harmonics 204
9.5 Perturbations Due to Oblateness (J_2) 207
9.6 Integration of the Equations of Variation 209
References 213

Chapter 10. Applications of Orbit Perturbations 215
10.1 Earth's Oblateness (J_2) Effects 215
10.2 Critical Inclination 217
10.3 Sun-Synchronous Orbits 218
10.4 J_3 Effects and Frozen Orbits 220
10.5 Earth's Triaxiality Effects and East-West Stationkeeping 221
10.6 Third-Body Perturbations and North/South Stationkeeping 222
10.7 Solar-Radiation-Pressure Effects 223
10.8 Atmospheric Drag Effects 227
10.9 Tidal Friction Effects 230
10.10 Long-Term Inclination Variations 233
References 237
Problems 238
Selected Solutions 240

Chapter 11. Orbital Systems 241
11.1 Launch Window Considerations 241
11.2 Time of Event Occurrence 253
11.3 Ground-Trace Considerations 254

11.4 Highly Eccentric, Critically Inclined $Q = 2$ Orbits (Molniya) . . . 256
11.5 Frozen Orbits 259
References 263

Chapter 12. Lunar and Interplanetary Trajectories 265
12.1 Introduction 265
12.2 Historical Background 266
12.3 Important Concepts 274
12.4 Lunar Trajectories 279
12.5 Analytical Approximations 280
12.6 Three-Dimensional Trajectories 287
12.7 Interplanetary Trajectories 287
12.8 Galileo Mission 294
12.9 Cassini-Huygens Mission to Saturn and Titan 296
12.10 Mars Odyssey Mission 298
References 299
Problems 299
Selected Solutions 300

Chapter 13. Space Debris 301
13.1 Introduction 301
13.2 Space Debris Environment: Low Earth Orbit 302
13.3 Debris Measurements 303
13.4 Space Debris Environment: Geosynchronous Equatorial Orbit . . . 307
13.5 Spatial Density 310
13.6 Collision Hazard Assessment Methods 315
13.7 Collision Hazards Associated with Orbit Operations 320
13.8 Debris Cloud Modeling 322
13.9 Lifetime of Nontrackable Debris 327
13.10 Methods of Debris Control 328
13.11 Shielding 329
13.12 Collision Avoidance 330
References 332

Chapter 14. Optimal Low-Thrust Orbit Transfer 335
14.1 Introduction 335
14.2 The Edelbaum Low-Thrust Orbit-Transfer Problem 335
14.3 The Full Six-State Formulation Using Nonsingular Equinoctial Orbit Elements 354
14.4 Orbit Transfer with Continuous Constant Acceleration 372
14.5 Orbit Transfer with Variable Specific Impulse 389
Appendix: The Partials of the M Matrix 399
References 409

Chapter 15. Orbital Coverage **411**
15.1 Coverage from a Single Satellite 411
15.2 Design of Optimal Satellite Constellations for Continuous Zonal and Global Coverage 429
15.3 Considerations in Selecting Satellite Constellations 439
15.4 Nontypical Coverage Patterns 442
References 446
Problems 448
Selected Solutions 450

Index 453

1
Basic Concepts

1.1 A Historical Perspective

One of man's earliest reasons for attempting to understand the motions of the sun, moon, and planets was his belief that they controlled his destiny. Other reasons were his need to measure time and later to use the celestial objects for navigation. Thus, the names assigned to the days of the week are closely related to the names of the celestial bodies: Saturn, Jupiter, Mars, sun, Venus, Mercury, and moon; by taking the first name and skipping two and repeating in this way, we have a partial derivation of the names of the week in French (*mardi, mercredi, jeudi, vendredi* for (Tuesday–Friday) or English (Saturday, Sunday, Monday).

The earliest evidence of man's interest in the universe dates back to 1650 B.C. in Babylon and Egypt (e.g., the Ahmes Papyrus). This evidence shows an elaborate system of numeration in which positional or place-value notation was used. For example, clay tablets with cuneiform writing show that the following sexagesimal (base 60) system notation was used:

$$P = 1, \qquad PP = 2, \qquad PPP = 3$$

$$\sphericalangle = 10 \qquad \sphericalangle\sphericalangle P = 2 \times 10 + 1 = 21$$

But

$$P\sphericalangle\sphericalangle = 1,20$$
$$= 60 + 20 = 80$$

or

$$P\sphericalangle = 60 + 10 = 70, \text{ etc.}$$

Thus, in this notation, 3; 3,45 would mean $3 + 3/60 + 45/3600 = 3.0625$, etc. This system of numeration remained in use through the time of Copernicus, who would write 2, 9: 17, 22, 36 to mean 129th day ($2 \times 60 + 9$), 17th h, 22nd min, and 36th s. The sexagesimal system is, of course, still the basis for modern timekeeping (i.e., 60 s/min, etc.).

In 300 B.C., Aristarchus developed a theory in which the sun and stars were fixed and the Earth revolved in a circular orbit about the sun. Unfortunately, the leading philosophers of the time treated Aristarchus' theory with contempt. The most popular theory was that the Earth was a fixed center of the universe and that the planets moved around the Earth. In 130 B.C., Hipparchus introduced the epicyclical motion of the planets, which was further developed by Ptolemy in A.D. 150 as the principal theory for predicting the motions of the planets. Although there were no physical principles on which to base the motions, some of the results obtained by this theory (e.g., the rise and set of the planets) were very accurate. This picture remained virtually unchanged through the Middle Ages.

It is interesting to note that the early Greeks speculated and theorized about the size, shape, and composition of the Earth. Pythagoras and Aristotle supported the spherical figure for the Earth. Anaximenes, however, believed strongly that the Earth was rectangular.

Problems of geometry (e.g., finding areas of surfaces) required the knowledge of π. The Babylonians considered the area of a circle to be $(8/9D)^2$, where D was the diameter. This corresponds to $\pi \approx 3.16049$, which was not as accurate as the figure arrived at by Archimedes, who was the first to have a method for calculating π to any desired degree of accuracy. It was based on the fact that the perimeter of a regular polygon of n sides inscribed in a circle is smaller than the circumference of the circle, whereas the perimeter of a similar polygon circumscribed about the circle is greater than its circumference. By making n sufficiently large, the two perimeters will closely approach the circumference of the circle arbitrarily. Archimedes started with a hexagon and, progressively doubling the number of sides up to 96, obtained

$$3\frac{10}{71} < \pi < 3\frac{1}{7}$$

which was far better than that of the Babylonians. The Babylonians also regarded the perimeter of a hexagon as being equal to six times the radius of the circumscribed circle, which appears to be the reason they chose to divide the circle into 360 deg, a rule that we still live by.

Early in the sixteenth century, Copernicus (1467–1543) put forward a scheme that put the sun at the center of the universe. The planets were seen as moving in epicycles around the sun, with the moon moving around the Earth. Copernicus also hypothesized that the stars lay on a sphere of very large radius. His theory was not very well received in his day.

In 1601, Johann Kepler (1571–1630) became the director of the Prague Observatory on the death of Tycho Brahe (1546–1601), who had observed for 13 years the relative motion of the planet Mars. By 1609, Kepler had formulated his first two laws and, in 1619, published the third law, which he dedicated to James I of England. Kepler stated that

1) Every planet moves in an orbit that is an ellipse, with the sun at one focus of the ellipse.

2) The radius vector drawn from the sun to any planet sweeps out equal areas in equal times.

3) The squares of the periods of revolution of the planets are proportional to the cubes of the semimajor axes of their orbits.

As the result of Galileo's (1564–1642) observations of the four moons of Jupiter, the Copernican heliocentric theory was accepted, and Newton's theory of gravitation produced a theoretical principle that explained the motions of the planets and laid the foundation for modern space flight.

Isaac Newton (1642–1727)

Although unexceptional as a child, Newton began to produce immediately following his studies at Trinity College, Cambridge, in 1665. In that year, the plague was ravishing Cambridge, and Newton retired to the countryside to work

in safety. He had already devised the binomial theorem and was beginning to grasp the calculus when the apple incident occurred (the tale sounds like an incredible legend, but it's true—it comes from Newton's own records).

Newton formulated his famous inverse-square law to describe the behavior of celestial bodies. Furthermore, he intuitively theorized that the full force of the Earth's gravity could be considered as emanating from a point source at its center. But when he checked his theory with calculations using an imprecise figure for the Earth's radius, he found he was off by enough to make him doubt his point-source assumption, and he set aside the problem indefinitely.

It was also during this two-year absence from Cambridge that he performed experiments with prisms, proving that white light is actually a combination of all the colors of the rainbow. The prism experiments made Newton famous, and he returned to Cambridge in 1667 as a Fellow of Trinity College. In 1669, when Newton was just 27, the Cambridge mathematics lecturer Isaac Barrows resigned in his student's favor, and Newton was appointed to the Lucasian Chair of Mathematics.

Three years later, Newton was elected to the Royal Society, to which he promptly reported all his findings in light. Although his papers quickly brought him international recognition, they also embroiled him in troublesome controversy, chiefly with Robert Hooke, who had performed some similar but far less brilliant optical experiments.

Hooke was not Newton's only adversary. Leibnitz developed the calculus simultaneously with, but quite independently of, Newton. Although the two men were friends, mathematicians and scientists in Germany and England soon fell into a chauvinistic debate over the true origin. Secretly, Newton, who was neurotically sensitive to any criticism, urged his supporters on.

His distinguished contributions in optics and mathematics notwithstanding, Newton's greatest discovery by far was the universal law of gravitation. In 1684, Hooke boasted to Wren and Halley that he had worked out the laws governing the motions of the heavenly bodies. Wren was intrigued with the problem but unimpressed with Hooke's explanation, and so he immediately put up a prize for the correct solution. Halley, a friend of Newton, took the problem to his colleague and asked how the planets would move if there was a force between all bodies decreasing as the square of the distance between them. Without a pause, Newton replied, "In ellipses." When Halley asked him how he knew, Newton said, "Why, I have calculated it." Then Newton told Halley about his earlier work during the plague years. Halley urged Newton to try his calculations once again. This time, Newton had a much more accurate figure for the radius of the Earth, and he knew from his calculations that the point-source assumption was correct.

Newton began to record his discoveries, and the result was *Philosophiae Naturalis Principia Mathematica.* One of the greatest scientific works ever published, *Principia Mathematica* climaxed the scientific revolution begun by Copernicus. The three laws of motion were established and, from them, the law of gravitation.

Afterward, honors came to Newton: elected to Parliament in 1689; appointed to the prestigious position of Warden of the Mint in 1696; elected president of the Royal Society in 1703; knighted in 1705.

But decades of bitter controversy and his own neurotic temperament took their toll in his later life. At one point he collapsed from a nervous breakdown and was forced to retire for two years. He had always been an ardent believer in alchemy and, in his later years, he wasted much of his time chasing recipes for gold. He

became a mystic, too, producing a vast amount of writing on the more abstruse passages of the Bible.

Yet, in the end, he was a modest man, who said of his profound contributions, "I seem to have been only like a boy playing on the seashore, and diverting myself in now and then finding a smoother pebble or a prettier shell, while the great ocean of truth lay all undiscovered before me."

Sir Isaac Newton died in 1727 and was buried in Westminster Abbey along with the greatest of England's heroes.

Newton's Laws

The following three laws of motion given by Newton are considered the axioms of mechanics:

1) Every particle persists in a state of rest or of uniform motion in a straight line (i.e., with constant velocity) unless acted on by a force.

2) If $\boldsymbol{F}$ is the (external) force acting on a particle of mass m which, as a consequence, is moving with velocity $\boldsymbol{v}$, then

$$\boldsymbol{F} = \frac{\mathrm{d}}{\mathrm{d}t}(m\boldsymbol{v}) = \frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}t} \tag{1.1}$$

where $\boldsymbol{p} = m\boldsymbol{v}$ is called the momentum. If m is independent of time t, this becomes

$$\boldsymbol{F} = m\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} = m\boldsymbol{a}$$

where $\boldsymbol{a}$ is the acceleration of the particle.

3) If particle 1 acts on particle 2 with a force $\boldsymbol{F}_{12}$ in a direction along the line joining the particles while particle 2 acts on particle 1 with a force $\boldsymbol{F}_{21}$, then $\boldsymbol{F}_{21} = -\boldsymbol{F}_{12}$. In other words, to every action there is an equal and opposite reaction.

It is assumed that m is a positive quantity and that it is large compared to the atomic particles. Furthermore, velocity $\boldsymbol{v}$ is regarded as small compared to the velocity of light (in nonrelativistic mechanics only).

Newton's laws define a force $\boldsymbol{F}$ in terms of the mass, velocity, and time (derivative with respect to time). Newton's definition of mass can be presented as follows:

> If two bodies A and B, which differ in material, size, and shape, give the same results when used in any experiment performed first on A and then on B, they are mechanically equivalent. The mass is a number assigned to a body A based on a comparison of mechanical equivalency with another body B. A distinction can be made between a gravitational mass of a body obtained by weighing and an inertial mass obtained from experiments (such as the measurement of F/acceleration) not affected by the gravitational fields. The equivalence of the inertial and gravitational masses was a subject of investigation by scientists from Newton (who found them equivalent to one part in 10^3) to Baron von Eotvos and R. H. Dicke. Eotvos performed his basic experiment in 1890 with a torsional balance, in which he determined the equivalency of the two masses to within the accuracy of his experiment (one part in 10^8). Dicke used a more refined apparatus,

employing three bodies, and found the equivalency to be one part in 10^{10}. The equivalency of the inertial and gravitational masses is, of course, the reason why all freely falling bodies have the same acceleration, as is required by the principle of equivalence in the general theory of relativity.

It must be emphasized that Newton's laws are postulated under the assumption that all measurements or observations are taken with respect to an "inertial" coordinate system or frame of reference that is fixed in space or is moving with a constant velocity (but not rotating). This is the so-called assumption that space or motion is absolute. It is quite clear, however, that a particle can be at rest or in uniform motion in a straight line with respect to one frame of reference and be traveling in a curve and accelerating with respect to another frame of reference.

The Earth is not exactly an inertial system but, for many practical purposes, it can be considered as one so long as motion takes place with speeds that are not too large. For speeds comparable to the speed of light (300,000 km/s), Newton's laws of mechanics must be replaced by Einstein's laws of relativity of relativistic mechanics.

1.2 Velocity and Acceleration

Kinematics and dynamics are two branches of physics that are concerned with the motion of material bodies and the effects produced on them by forces. Kinematics is the study of the geometry of motion, whereas dynamics is concerned with the physical causes of motion. The axiomatic foundations of these two branches of physics (mechanics) are based on

1) Undefined terms or concepts, such as a line in the Euclidian geometry.
2) Postulates or assertions based on experimental observations, such as Kepler's laws.
3) Definitions (displacement, velocity, accelerations, etc.).
4) Theorems or proved assertions based on definitions and postulates.

The study of the kinematics and dynamics of motion usually requires a reference coordinate frame in which the position, velocity, and acceleration of a material mass point can be specified. A nonrotating and nonaccelerating reference frame is known as the inertial frame in which the laws of mechanics are valid and can be most conveniently expressed. If an orthogonal inertial frame is used to denote the vector position $\boldsymbol{r}$ of a point, then its velocity and accelerations are $\boldsymbol{v} = \mathrm{d}\boldsymbol{r}/\mathrm{d}t$ and $\boldsymbol{a} = \mathrm{d}\boldsymbol{v}/\mathrm{d}t$, respectively. If the position $\boldsymbol{r}$ is measured in a frame that has an angular velocity $\boldsymbol{\omega}$ and the translational velocity $\boldsymbol{v}_0$, then the absolute or inertial velocity of the point is

$$\boldsymbol{v}_{\text{inertial}} = \dot{\boldsymbol{r}}' + \boldsymbol{\omega} \times \boldsymbol{r} + \boldsymbol{v}_0 \tag{1.2}$$

where $\dot{\boldsymbol{r}}'$ = translational velocity of the point relative to the rotating frame.

The acceleration of the point in inertial space is then

$$\begin{aligned} \boldsymbol{a} &= \frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} \\ &= \ddot{\boldsymbol{r}}' + \boldsymbol{\omega} \times \dot{\boldsymbol{r}}' + \dot{\boldsymbol{\omega}} \times \boldsymbol{r} + \boldsymbol{\omega} \times (\dot{\boldsymbol{r}}' + \boldsymbol{\omega} \times \boldsymbol{r}) + \dot{\boldsymbol{v}}_0 \\ &= \ddot{\boldsymbol{r}}' + 2\boldsymbol{\omega} \times \dot{\boldsymbol{r}}' + \dot{\boldsymbol{\omega}} \times \boldsymbol{r} + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \boldsymbol{r}) + \boldsymbol{a}_0 \end{aligned} \tag{1.3}$$

where $\ddot{\boldsymbol{r}}'$ = observed acceleration of the point relative to the rotating frame, and $\boldsymbol{a}_0$ = acceleration of the rotating frame.

The second and fourth terms in Eq. (1.3) are the Coriolis and centripetal accelerations, respectively. The $\dot{\boldsymbol{\omega}} \times \boldsymbol{r}$ term is the tangential (Euler) acceleration. The Coriolis acceleration $\boldsymbol{a}_c = 2\boldsymbol{\omega} \times \dot{\boldsymbol{r}}'$ term exists only when there is an observed velocity $\dot{\boldsymbol{r}}'$ in a rotating frame.

Projectile on a Rotating Earth

The coordinate frames convenient for the use of Eq. (1.3) to compute the Coriolis acceleration terms for a particle ejected with a velocity $\dot{\boldsymbol{r}}'$ at latitude λ on the Earth's surface are illustrated in Fig. 1.1.

The orthogonal frame E_1, E_2, E_3 is an Earth-centered inertial (ECI) frame in Fig. 1.1. The e_1, e_2, e_3 is a frame rotating with the Earth angular velocity $\boldsymbol{\omega}$, in which the particle p is given an easterly velocity $\dot{\boldsymbol{r}}'$ at an elevation angle α. The acceleration of the particle p relative to the E_1, E_2, E_3 frame is

$$\begin{aligned} \boldsymbol{a} &= \boldsymbol{a}_0 + \ddot{\boldsymbol{r}}' + 2\boldsymbol{\omega} \times \dot{\boldsymbol{r}}' + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \boldsymbol{r}) \\ &= \boldsymbol{g} = \text{gravitational acceleration} \end{aligned} \tag{1.4}$$

Here,

$$\boldsymbol{a}_0 = \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \boldsymbol{R}) = \text{centripetal acceleration of } 0' \text{ relative to } 0$$

$$\ddot{\boldsymbol{r}}' = \text{apparent (relative) acceleration}$$

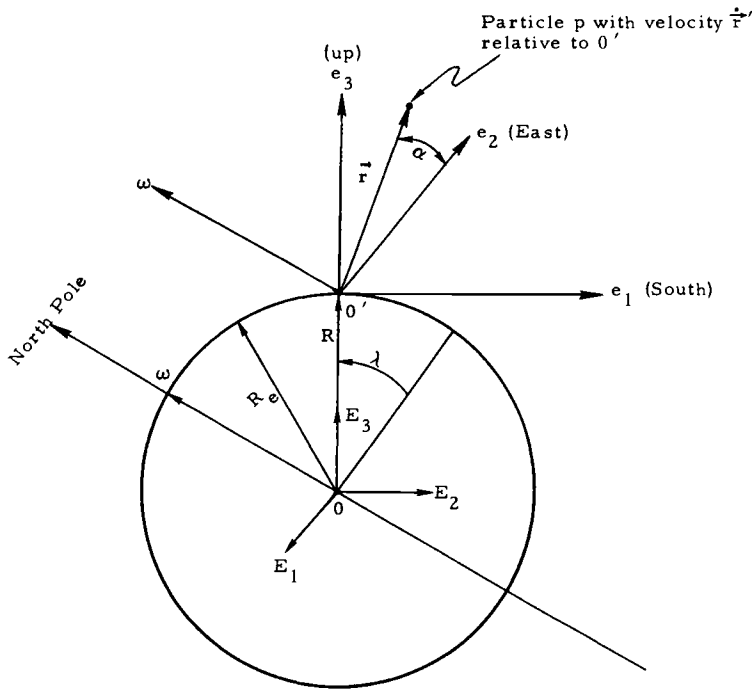


Fig. 1.1 Projectile on the rotating Earth.

If $\boldsymbol{a}_0$ and $\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \boldsymbol{r}) \ll 2\boldsymbol{\omega} \times \dot{\boldsymbol{r}}'$, Eq. (1.4) can be solved approximately for the acceleration of p relative to $0'$ as

$$\ddot{\boldsymbol{r}}' \approx \boldsymbol{g} - 2\boldsymbol{\omega} \times \dot{\boldsymbol{r}}'$$

Rocket Equation

An important application of Newton's second law of motion is in the derivation of the rocket equation. For example, let a mass m with a velocity v collide inelastically with an element of mass Δm, which has a velocity v_1. Then, prior to impact, the initial system momentum (in scalar form) is

$$P_1 = mv + \Delta m v_1 \tag{1.5}$$

At a time Δt later, the two masses are joined, and their combined momentum is

$$P_2 = (m + \Delta m)(v + \Delta v) \tag{1.6}$$

Newton's laws state that the impulse is equal to the change in momentum or

$$\begin{aligned} \text{Impulse} &= F\Delta t \\ &= P_2 - P_1 \\ &= m\Delta v + \Delta m(v - v_1) + \Delta m \Delta v \end{aligned} \tag{1.7}$$

where, in the limit as $\Delta t \to 0$ and Δm, $\Delta v \to 0$, the force F can be expressed as

$$F = m\frac{\mathrm{d}v}{\mathrm{d}t} + \frac{\mathrm{d}m}{\mathrm{d}t}(v - v_1) \tag{1.8}$$

If $v_1 = 0$, then this result is equivalent to the case of a falling raindrop, for example, which absorbs moisture (stationary) from the air. If, on the other hand, $u = v - v_1$, then Eq. (1.8) can be written as

$$F = m\frac{\mathrm{d}v}{\mathrm{d}t} - T \tag{1.9}$$

where $T = -u(\mathrm{d}m/\mathrm{d}t) =$ momentum thrust of a rocket. Equation (1.9) is known as the rocket equation, where F is the external force acting on the rocket (e.g., gravity, aerodynamic, etc.), and u is the exhaust velocity.

A basic application of Eq. (1.9) is in the calculation of velocity impulses (ΔV) for rocket motion. A somewhat different form of this equation can be obtained by assuming that $F = 0$ and rewriting Eq. (1.9) as

$$\begin{aligned} \mathrm{d}v &= \frac{T\,\mathrm{d}t}{m} \\ &= \frac{-u\,\mathrm{d}m}{m} \end{aligned}$$

which, after integration, becomes

$$\begin{aligned}\Delta v &= u \ln \frac{m_i}{m_f} \\ &= g_o I_{sp} \ln \frac{m_i}{m_f}\end{aligned} \tag{1.10}$$

where

m_i = initial mass of rocket

m_f = final mass of rocket

I_{sp} = propellant specific impulse (thrust/propellant weight flow rate)

g_o = gravitational constant at sea level

Equation (1.10) can be used to compute the mass of propellant for a given value of ΔV. It is also known as the Tsiolkovsky formula.

The specific impulse I_{sp} is a parameter of propellant quality and generally varies from 60 to 3000 + s for cold gas or ion propellants, respectively. It is a measure of propellant thrust obtained per unit of propellant weight flow. For example, Eq. (1.10) can be solved for the final mass remaining after a burn of velocity impulse ΔV as

$$m_f = m_i \exp(-\Delta V / g_o I_{sp}) \tag{1.11}$$

in terms of the initial mass m_i and the specific impulse I_{sp}.

If there are two or more impulses $\Delta V_i (i = 1, 2, \ldots, n)$ with a coast period between the impulses, the final mass remaining m_f is given by the equation

$$m_f = m_i \exp(-\Delta V_T / g_o I_{sp}) \tag{1.12}$$

where

$\Delta V_T = \Delta V_1 + \Delta V_2 + \cdots + \Delta V_n$
$m_i = m_{bo} + m_p + m_{pl}$ = initial rocket mass
$m_f = m_{bo} + m_{pl}$ = final rocket mass
m_{bo} = burnout mass (structural mass)
m_p = propellant mass
m_{pl} = payload mass

The propellant mass ratio $\varepsilon = m_p/(m_p + m_{bo})$. The structural factor is $1 - \varepsilon$. The mass of the propellant burned is

$$m_p = m_i[1 - \exp(-\Delta V_T / g_o I_{sp})] \tag{1.13}$$

Example. Compute the propellant mass m_p required to deliver a payload of 1000 kg to a mission orbit requiring a velocity impulse $\Delta V = 4354$ m/s. Assume that the burnout mass of the rocket (structure) is 2000 kg and that the specific impulse of the propellant (LO_2/LH_2) is 460 s.

Using Eq. (1.11), the mass ratio

$$\frac{m_i}{m_f} = \frac{m_{\text{pl}} + m_{\text{bo}} + m_p}{m_{\text{pl}} + m_{\text{bo}}}$$
$$= \exp(\Delta V / g I_{\text{sp}})$$

Solving for $m_p = 4881$ kg.

Problems

1.1. A bar $AB = 3$ m is moving in a plane. At a given instant, the velocities of A and B are

$$V_A = 2 \text{ m/s}, \text{ 60 deg clockwise from line } A \text{ to } B$$
$$V_B = 30 \text{ deg clockwise from line } A \text{ to } B$$

Determine the angular velocity ω of the bar, stating whether it is clockwise or counterclockwise.

1.2. Determine the radial and normal acceleration components for a particle moving in a plane.

1.3. A single-stage rocket of mass $m_0 = m_{\text{bo}} + m_p$ carries a payload of mass m_{pl}. If the propellant mass ratio $\varepsilon = 0.8$, find the maximum payload that can be given a final speed of 1.3 u, where u is the exhaust velocity. Here $m_{\text{bo}} =$ burnout mass, and $m_p =$ propellant mass.

1.4. Determine the remaining propellant in the orbit-to-orbit shuttle after deployment of a 1000-kg payload to the circular synchronous equatorial orbit. Assume a start from, and a return to, a low-altitude circular parking orbit. Assume the following:

1) Ascent transfer-velocity impulse
   $\Delta V_1 = 2524$ m/s
2) Synchronous equatorial orbit injection-velocity impulse
   $\Delta V_2 = 1830$ m/s
3) Descent transfer-velocity impulse
   $\Delta V_3 = \Delta V_2$
4) Parking orbit injection velocity impulse
   $\Delta V_4 = \Delta V_1$
5) Propellant (LO_2/LH_2)
   $I_{\text{sp}} = 460$ s

Also, let

$$m_{\text{bo}} = 3041 \text{ kg} = \text{burnout mass}$$
$$m_p = 26500 \text{ kg} = \text{propellant mass}$$
$$m_{\text{pl}} = 1000 \text{ kg} = \text{payload mass}$$

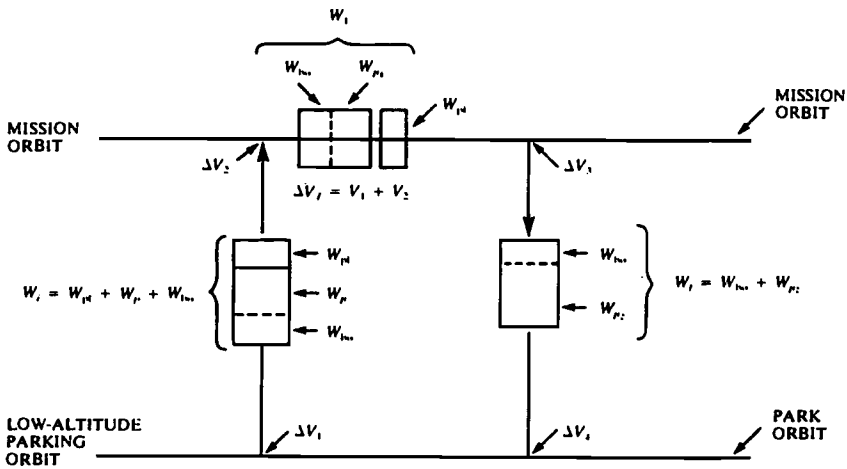
HINT: Given W_i, Find W_{p_2}.
Use

$$\frac{W_i}{W_1} = \exp[(\Delta V_1 + \Delta V_2)/gI_{\text{sp}}]$$

$$\frac{W_2}{W_f} = \exp[(\Delta V_3 + \Delta V_4)/gI_{\text{sp}}]$$

where

$$W_1 = W_{\text{bo}} + W_{p_1} + W_{\text{pl}}$$
$$W_2 = W_1 - W_{\text{pl}}$$
$$W_2 = W_1 - W_{\text{pl}}$$



Selected Solutions

1.1. $\omega = 0.3849$ rad/s counterclockwise

1.3. $m_{\text{pl}} = m_0/10$

1.4. $W_{p_2} = 1004$ kg

2
Celestial Relationships

2.1 Coordinate Systems

Several coordinate systems are used in the study of the motions of the Earth and other celestial bodies. The heliocentric-ecliptic coordinate system, illustrated in Fig. 2.1, has its origin at the center of the sun.[1] The $X_\varepsilon Y_\varepsilon$ fundamental plane coincides with the ecliptic plane, which is the plane of the Earth's revolution around the sun. The line of intersection of the ecliptic plane and the Earth's equatorial plane defines the direction of the X_ε axis. On the first day of spring, a line joining the center of the Earth and the center of the sun points in the direction of the positive X_ε axis. This is called the vernal equinox direction and is denoted by the symbol of a ram's head by astronomers because it points in the general direction of the constellation Aries. All Earth locations experience identical durations of daylight and darkness. The Earth's spin axis wobbles slightly and shifts in direction slowly over centuries. This effect is known as precession, and it causes the line of intersection of the Earth's equatorial plane and the ecliptic plane to shift slowly. As a result, the heliocentric-ecliptic system is not an inertial reference frame. Where extreme precision is required, it is necessary to specify the coordinates of an object based on the vernal equinox direction of a particular year or epoch.

ECI System

The geocentric-equatorial coordinate system, on the other hand, has its origin at the Earth's center. The fundamental plane is the equator, and the positive X axis points in the vernal equinox direction. The Z axis points in the direction of the North Pole. (At equinox all Earth locations experience identical durations of daylight and darkness.) This is shown in Fig. 2.2. It is important to keep in mind, when looking at Fig. 2.2, that the X, Y, Z system is not fixed to the Earth and turning with it; rather, the geocentric-equatorial frame is nonrotating with respect to the stars (except for precession of the equinoxes), and the Earth turns relative to it.

The two angles needed to define the location of an object along some direction from the origin of the celestial sphere are defined as follows:

α (right ascension) = the angle measured eastward in the plane of the equator from a fixed inertial axis in space (vernal equinox) to a plane normal to the equator (meridian), which contains the object; $0 \text{ deg} \leq \alpha \leq 360 \text{ deg}$.

δ (declination) = the angle between the object and equatorial plane measured (positive above the equator) in the meridional plane, which contains the object; $-90 \text{ deg} \leq \delta \leq 90 \text{ deg}$.

r (radial distance) = the distance between the origin of the coordinate system and the location of a point (object) within the coordinate system; $r \geq 0$.

To an observer fixed in an inertial frame (i.e., one that is at rest or moves with constant velocity relative to distant galaxies), the Earth revolves about the sun

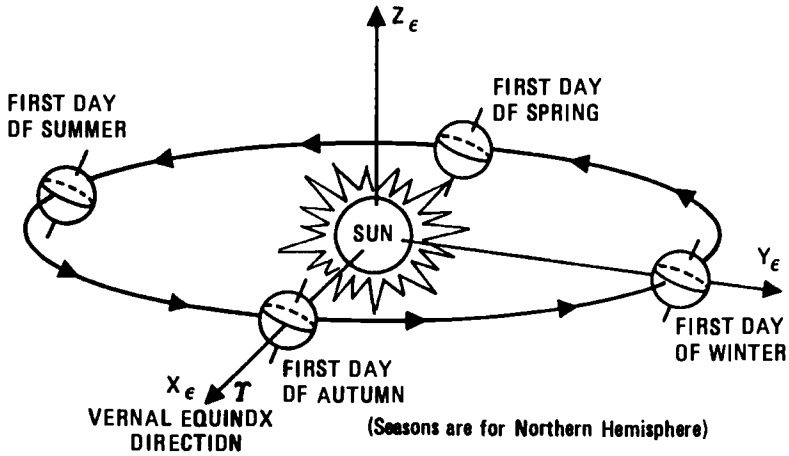


Fig. 2.1 Heliocentric ecliptic coordinate system (from Ref. 1).

in a nearly circular orbit in the ecliptic plane. The Earth rotates at an essentially constant rate about its polar axis. However, that axis of rotation is tilted away from the normal to the ecliptic. This tilt causes the equatorial and ecliptic planes to form a dihedral angle that is conventionally known as the obliquity of the ecliptic and is denoted by $\varepsilon(\approx 23.5$ deg).

On account of this angle between the two planes, the sun, as viewed by an observer on Earth, appears to shuttle northward and southward across the equatorial plane (called the celestial equator). This gives rise to the changing seasons during the course of the Earth's yearly revolution about the sun.

Of particular interest is the time each year when the sun moves northward through the equatorial plane; this event, called the vernal equinox, marks the beginning of spring. It is customary to associate a direction in inertial space with

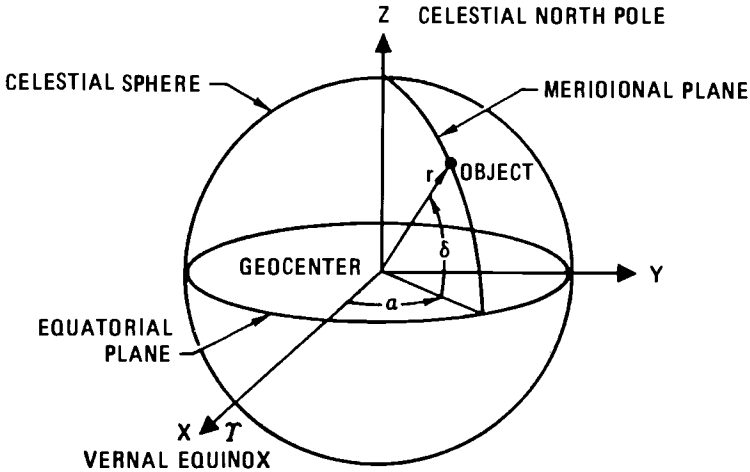


Fig. 2.2 Earth-centered inertial system (from Ref. 2).

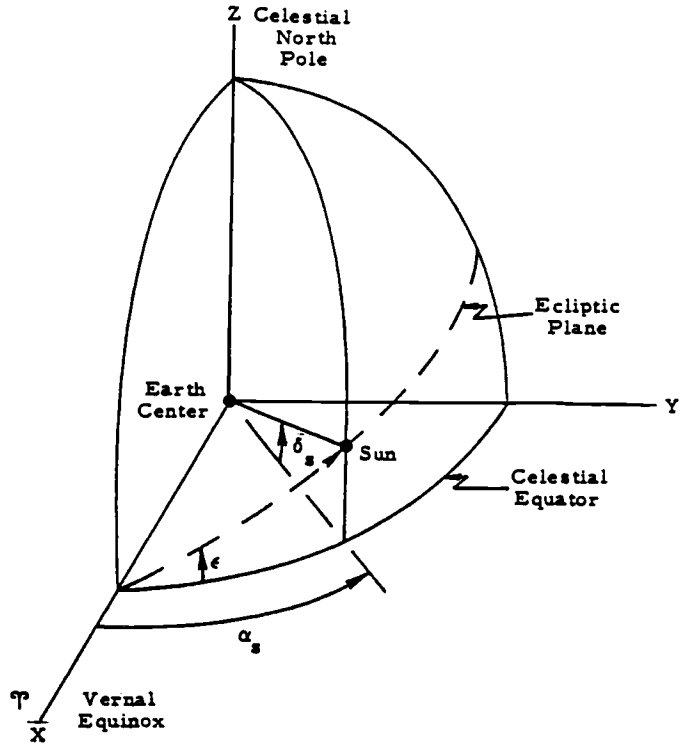


Fig. 2.3 Motion of the sun in the ECI reference frame (from Ref. 3).

this event by noting the position of the sun as seen from Earth against the field of infinitely distant background galaxies. Figure 2.3 illustrates the motion of the sun in the ECI frame frequently used in defining celestial relationships. Note that the X axis is directed toward the vernal equinox and that the Z axis lies along the Earth's polar axis of rotation; the Y axis completes the right-handed triad. The three arcs drawn in the three references planes bound an octant of the celestial sphere that is centered at the Earth and of arbitrary radius. The path of the sun's motion is shown by the dashed arc on the celestial sphere.

About three months after the vernal equinox, the sun reaches its northernmost position (and, at local noon, is directly overhead to observers on the Tropic of Cancer at about 23.5°N); this event is called the summer solstice, since the sun appears to stand still momentarily as it reverses its direction of motion from northward to southward. Then, about three months after the summer solstice, the sun crosses the celestial equator from north to south at the autumnal equinox. The sun reaches its southernmost position about three months later at the winter solstice, which marks the start of winter. At that time, the sun is directly overhead to observers, at local noon, on the Tropic of Capricorn at about 23.5°S.

Referring again to Fig. 2.3, it is customary to indicate the sun's position by a pair of angles similar to the familiar longitude and latitude system used for terrestrial measurements. The right ascension of the sun is the angle measured

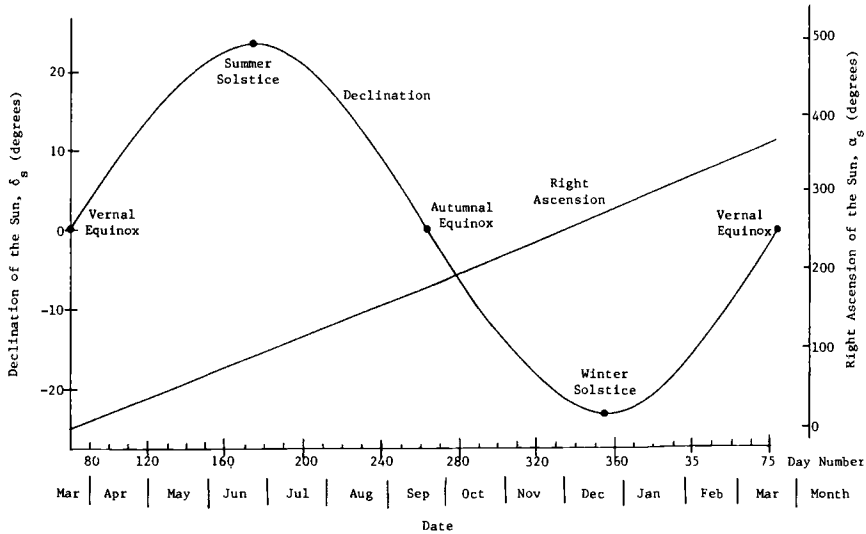


Fig. 2.4 Solar declination and right ascension vs date (from Ref. 3)

positively eastward along the celestial equator to the meridional plane (through the celestial polar axis), which passes through the sun; this angle, which is the counterpart of terrestrial longitude, is denoted as α_s. The second angle used to indicate the sun's position is the declination measured positively northward from the celestial equator; this angle, which is the counterpart of geocentric latitude, is denoted by δ_s. Figure 2.4 shows the variation with time of solar declination and right ascension.

Geographic Coordinate System

It is customary to locate an object relative to the Earth by two angular coordinates (latitude-longitude) and altitude above (or perhaps below) the adopted reference ellipsoid. The present discussion considers the Earth to be spherical. The origin of the latitude-longitude coordinate system is the geocenter (Fig. 2.5). The fundamental plane is the equator, and the principal axis in the fundamental plane points toward the Greenwich meridian.

The two angles required to define the location of a point along some ray from the geocenter are defined as follows:

ϕ (geocentric latitude) = the acute angle measured perpendicular to the equatorial plane between the equator and a ray connecting the geocenter with a point on the Earth's surface; -90 deg $\leq \phi \leq 90$ deg.

λ_E (east longitude) = the angle measured eastward from the prime meridian in the equatorial plane to the meridian containing the surface point; 0 deg $\leq \lambda_E \leq$ 360 deg.

The observer's coordinate frame is related to the previously discussed inertial frame through the angle Θ, the sidereal time; Θ is measured eastward in the equatorial plane from the vernal equinox to the observer's meridian and ranges from 0^h to 24^h (or, equivalently, from 0 to 360 deg). The distance from the geocenter to the object is called the range and is denoted by r.

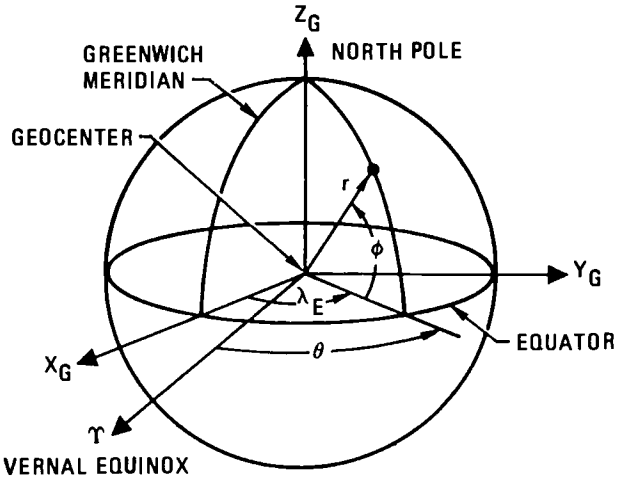


Fig. 2.5 Latitude-longitude coordinate system (from Ref. 2).

Azimuth-Elevation Coordinate System

An observer standing at a particular point on the surface of a rotating planet sees objects in a rotating coordinate system. In this system, the observer is at the origin of the system, and the fundamental plane is the local horizon (Fig. 2.6). Such a coordinate system is referred to as a topocentric system. Generally, the principal axis or direction is taken as pointing due south. Relative to an observer, the object is in a meridional plane that contains the object and passes through the zenith of the observer.

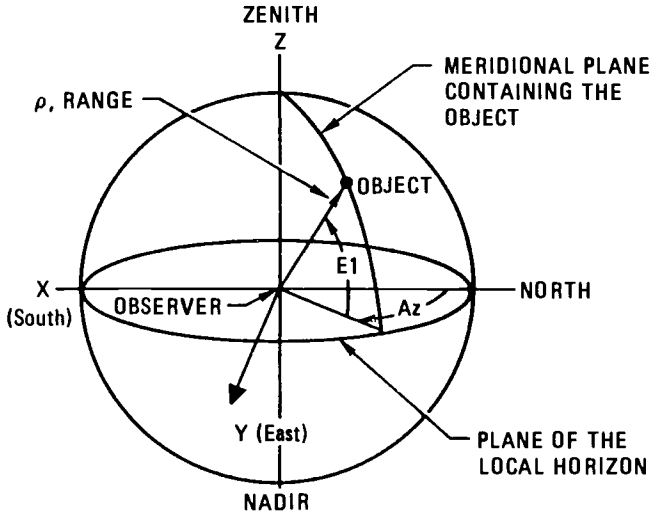


Fig. 2.6 Azimuth-elevation topocentric coordinate system (from Ref. 2).

Table 2.1 Constants and conversion factors

Earth Gravitational Constant:[a]

$$\begin{aligned}\mu &= 5.53042885 \times 10^{-3}\ \mathrm{ER}^3\ (\text{equatorial Earth Radius})^3/\mathrm{min}^2\\ &= 3.986005 \times 10^5\ \mathrm{km}^3/\mathrm{s}^2\\ &= 1.40764438 \times 10^{16}\ \mathrm{ft}^3/\mathrm{s}^2\\ &= 6.27501680 \times 10^4\ (\mathrm{n.mi.})^3/\mathrm{s}^2\end{aligned}$$

Mean Equatorial Earth Radius:

$$\begin{aligned}\mathrm{ER} &= 6.37813700 \times 10^3\ \mathrm{km}\\ &= 3.44391847 \times 10^3\ \mathrm{n.mi.}\end{aligned}$$

Rotational Rate of the Earth:

$$\omega_E = 7.29211515 \times 10^{-5}\ \mathrm{rad/s}$$

Mean Obliquity of the Ecliptic (1 January 1976):

$$\begin{aligned}\varepsilon &= 23^\circ\ 26'\ 32.66''\\ &= 23.442405^\circ\end{aligned}$$

Time Conversions:

1 mean solar day = 86,400 ephemeris s
1 mean sidereal day = 86164.09054 ephemeris s
1 tropical yr = 365.2421988 mean solar days
1 calendar yr = 365 mean solar days
1 mean solar s = 1.1574074×10^{-5} mean solar day
= 1.1605763×10^{-5} sidereal day
= 1.002737 sidereal s
Julian day = continuous count of the number of days since noon, 1 January 4713 B.C.

[a]IAU (1976) Astrodynamics Standards.

The two angles needed to define the location of an object along some ray from the origin are defined as follows:

Az (azimuth) = the angle, eastward from north to the object's meridian, as measured in the local horizontal plane, which is tangent to the sphere at the observer's position; 0 deg $\leq Az <$ 360 deg.

El (elevation) = angular elevation (measured positively upward in the meridional plane) of an object above the local horizontal plane, which is tangent to the sphere at the observer's position; -90 deg $\leq El \leq$ 90 deg.

The distance from the observer to the object is called the range and is usually denoted by ρ. $\rho \geq 0$.

Coordinate Transformations

The satellite state vector at a given time is obtained by integrating the equations of motion that equate the acceleration of the vehicle to the sum of the various accelerations acting on the vehicle. The integration must be performed in an inertial (nonrotating) reference frame. However, the principal acceleration due to gravity and aerodynamic drag is expressed mostly in the rotating (body-fixed) systems. Transformations from body-fixed coordinates to inertial and back are therefore required. These transformations involve the motion of the equinox, which is due to the combined motions of the Earth's equatorial plane and the ecliptic plane, the equinox being defined as the intersection of these planes. The motion of the equatorial plane is due to the gravitational attraction of the sun and moon on the Earth's equatorial bulge. It consists of the lunisolar precession and nutation. The former is the smooth, long-period westward motion of the equator's mean pole around the ecliptic pole, with an amplitude of about 23.5 deg and a period of about 26,000 yr. The latter (nutation) is a relatively short-period motion that carries the actual (or true) pole around the mean pole in a somewhat irregular curve, with an amplitude of approximately 9 s of arc and a period of about 18.6 yr. The motion of the ecliptic (i.e., the mean plane of the Earth's orbit) is due to the planet's gravitational attraction on the Earth and consists of a slow rotation of the ecliptic. This motion is known as planetary precession and consists of an eastward movement of the equinox of about 12 s of arc a century and a decrease of the obliquity of the ecliptic, the angle between the ecliptic and the Earth's equator, of approximately 47 s of arc a century. The "true" equator and equinox are obtained by correcting the mean equator and equinox for nutation.

2.2 Time Systems

Astronomers also specify the location of stars by their right ascensions and declinations, which are essentially invariant. (Very gradual changes in star locations occur as a result of the precession of the equinoxes, which is caused by coning of the Earth's spin axis in inertial space.) The angular coordinates of the sun, however, vary considerably in the course of each year, as shown in Fig. 2.4 or by the tabulations of sun position given in a solar ephemeris. As indicated earlier, the sun moves at an irregular rate (as a result of noncircularity of the Earth's orbit about the sun) along the dashed arc in Fig. 2.3. The time between two successive passages through the vernal equinox is called a tropical year (about 365.25 days). Because of the precession of the equinoxes, the tropical year is about 20 min shorter than the orbital period of the Earth relative to the fixed-star or sidereal year.[4]

On account of the irregular motion of the true sun along the ecliptic, astronomers have introduced a fictitious mean sun, which moves along the celestial equator at a uniform rate and with exactly the same period as the true sun. It is relative to this fictitious mean sun that time is reckoned. This can best be explained by the geometry illustrated in Fig. 2.7, in which the celestial sphere is viewed by an observer at the celestial North Pole looking in the direction of the negative Z axis of the Earth-centered inertial (ECI) frame.

The rotational orientation of the Earth is conventionally determined by specification of the Greenwich hour angle of the vernal equinox (GHAVE), which is measured positively westward along the celestial equator from the Greenwich

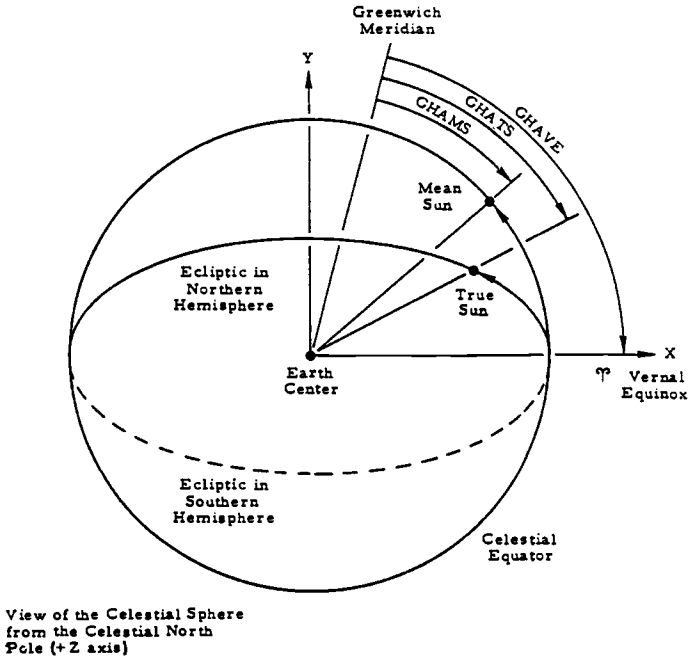


Fig. 2.7 Orientation of the mean and true sun and the Greenwich meridian in the ECI frame (from Ref. 3).

meridian to the vernal equinox. (Greenwich hour angle is measured either in degrees or in hours, the conversion factor being exactly 15 deg/h). Similarly, the instantaneous locations of the mean and the true suns are determined by Greenwich hour angle of the mean sun (GHAMS) and Greenwich hour angle of the true sun (GHATS), respectively. Incidentally, the difference (GHATS − GHAMS), also known as the equation of time, can be appreciated as the discrepancy between time reckoned by a sundial (based on the true sun position) and conventionally determined time (based on the position of the mean sun). Greenwich mean time (GMT) is linearly related to the GHAMS as follows[3]:

$$\mathrm{GMT(h)} = [12 + \mathrm{GHAMS(deg)}/15]\mathrm{mod}\ 24 \tag{2.1}$$

Noon GMT occurs when the mean sun is at upper transit of the Greenwich meridian (that is, the mean sun is at maximum elevation to an observer located anywhere on the Greenwich meridian); when the mean sun is 90 deg west of Greenwich, the GMT is 18 h or 6 p.m., and so on. The notation "mod 24" in the preceding equation simply means that integer multiples of 24 are to be added or subtracted as necessary to give GMT a value in the range 0–24 h.

A mean solar day is defined as the time during which the Earth makes a complete rotation relative to the mean sun; the mean sun, of course, moves eastward along the celestial equator at a rate of 360 deg in about 365.25 mean solar days or slightly less than 1 deg/day. The length of a mean solar day is 24 mean solar hours. A sidereal day is the time during which the Earth makes a complete rotation relative to a fixed direction, e.g., the vernal equinox; clearly, it is shorter than a mean solar

day by about 4 min, since the sidereal reference meridian is fixed in inertial space. The length of a sidereal day is 24 sidereal hours or about 23 h, 56 min, of mean solar time.

Another useful term carried over from astronomy is epoch, an instant of time or a date selected as a point of reference. Quantities defining an orbit or some other celestial relationship are often specified at a particular instant. For example, the GHAVE is tabulated in ephemerides for zero hours on every day of a given year. Since GHAVE increases linearly with mean solar time, it is possible to establish an important relationship between time, longitude, and right ascension of some celestial event.

Refer to Fig. 2.8, and suppose that an observer at longitude $\lambda°$E observes a star at upper transit of his meridian at t h GMT on a particular date; let this event be regarded as epoch and the day on which it occurs as epoch date. The right ascension of the star is just the sum of the Greenwich hour angle at midnight of the epoch date plus the Earth rotation occurring in time t plus the east longitude of the observer. Symbolically, this becomes

$$\alpha^* = \mathrm{GHAVE}_0 + \omega_E t + \lambda \tag{2.2}$$

where the Earth's rotational rate (ω_E) is about 15.04 deg/mean solar hour. Knowledge of any three of the quantities (α, GHAVE, t, and λ) enables the remaining quantity to be determined. If the star is recognized based on its relation to a known constellation, one can consult a star catalog to find its right ascension. Similarly, on a given epoch date, there is a unique value of GHAVE at midnight at the start of that day. Finally, if the time t of upper transit is known, the observer can infer his longitude. This is, in fact, the way mariners who are far from known landmarks use celestial observations to determine their longitude.

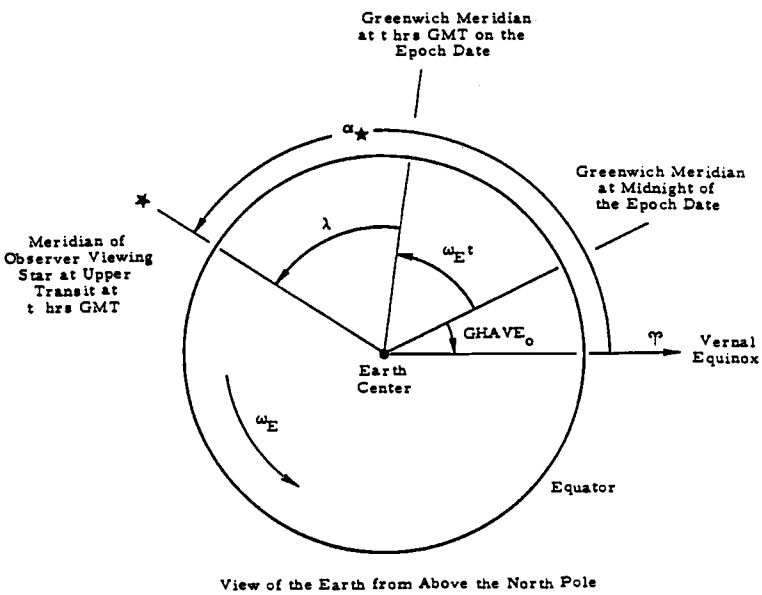


Fig. 2.8 Relation among time, longitude, and celestial position (from Ref. 3).

Computer programs use the atomic time system A.1 in the integration of the equations of motion. However, the input-output data sets are often referenced to ephemeris time (ET) for solar/lunar ephemerides, universal time (UT) for computing Greenwich sidereal time, and universal time coordinated (UTC) for input-output epochs and tracking data. The unit of ET time is defined as the ephemeris second, which is a fraction 1/31, 556, 925.9747 of the tropical year for 12^h ET of Jan 0^d, 1900. The A.1 time is one of several types of atomic time. In 1958, the U.S. Naval Observatory established the A.1 system based on an assumed frequency of 9, 192, 631, 770 oscillations of the isotope 133 of cesium atom/A.1 s. The UTC time scale is kept in synchronism with the rotation of the Earth to within ± 0.9 s by step-time adjustments of exactly 1 s, when needed. It is often referred to as the GMT, or ZULU time. It is defined by the atomic second and leap second (± 1) to approximate UT1, the fundamental unit of which is the mean solar day corrected for seasonal variation of Earth's rotation. An error in UT1 translates directly into an error in Earth orientation relative to vernal equinox. Thus, a 1-s error in UT1 may mean a 400- to 500-mile error in satellite position. This is not likely to occur, however, since the time adjustments to UT1 are known in advance.[4]

In the selection of reference coordinate frames for orbital mechanics calculations, a compromise is often made between extreme accuracy and ease of computation. A frame completely free of known errors and approximations is computationally time-consuming and is unnecessary for many applications. This is the reason why there are often numerous options available to the user, depending on his needs. However, it is essential to understand these options if the resultant positional errors are to be resolved in application to specific missions.

The coordinates used in all orbit determination programs consist of the fundamental astronomical reference systems. For example, the input ephemerides of the planets are heliocentric and refer to the mean equator and equinox of 1950.0. The notation .0 after the year refers to the beginning of the Besselian solar year at the instant at which the right ascension of the fictitious mean sun is $18^h\ 40^m$.

As of 1 January 1984, the new unit of time has been designated as the Julian Ephemeris Century, which equals 36,525 days or 3.15576×10^9 s of International Atomic Time (TAI). The origin of time for precessions and nutation is noon, 1 January A.D. 2000 or 2000.0 exactly.

The input observational data are usually in a topocentric coordinate system. The integration is done in either geocentric, heliocentric, or other planetary coordinates referred to the mean equator and equinox of 2000.0 or a specified epoch. The force model includes terms referred to a coordinate system that is fixed in the rotating Earth and terms that are referred to the moon and planets.

References

[1]Bate, R. R., Mueller, D., and White, J. E., *Fundamentals of Astrodynamics,* Dover, New York, 1971.

[2]Cantafio, L. (ed.), *Space-Based Radar Handbook,* Artech House, Norwood, MA, 1989.

[3]Ginsberg, L. J., and Luders, R. D., *Orbit Planner's Handbook,* The Aerospace Corporation Technical Memorandum, 1976.

[4]Wertz, J. R. (ed.), *Spacecraft Attitude Determination and Control,* D. Reidel Publishing, Dordrecht, Holland, 1980.

3
Keplerian Orbits

3.1 Newton's Universal Law of Gravitation

Newton's law of gravitation states that any two particles of mass m_1 and m_2, distance r apart, are attracted toward each other with a force

$$F = \frac{Gm_1m_2}{r^2} \tag{3.1}$$

where G is a universal constant called the gravitational constant. The value of G is $(6.6726 \pm 0.0005) \times 10^{-11}$ m^3/kg-s^2(1σ), which is known to about 1 part in 13,000.[1] The value of $\mu = GM_e$, on the other hand, where M_e is the Earth mass, is known to a much higher accuracy, and is therefore used for astrodynamical calculation. Accurate measurement of the gravitational attraction between small masses is limited by the accuracy to which G is known.

There are several laboratory techniques available for the determination of G. One of the most accurate experiments to determine G is the "time of swing method" of Heyl. This approach is illustrated in Fig. 3.1, where two large masses M and two small masses m are placed on coaxially suspended torsional springs.

When two balances are aligned in parallel, the period of oscillation is less than when they are aligned at right angles. In the former position, the gravitational attraction adds to the torsional spring whereas, in the latter position, the attraction subtracts from the spring. The periods are on the order of 0.5 h and can be measured to 0.1 s of accuracy.

A constant torque is thus maintained on the free balance. The angular displacement, after many hours, determines G.

Other methods for improving the accuracy of G include a dynamically resonant torsional balance system, with a possible accuracy of 1 part in 10^6. Miniature orbital systems with equal or unequal masses have also been proposed that could be established in space to determine G to a high accuracy.

The gravitational force equation (3.1) can be expressed in terms of the gradient ∇ of a scalar gravitational potential function V of a body. Thus, one can write

$$\boldsymbol{F} = \nabla V \tag{3.2}$$

where $V = \mu/r$ for a spherical body. For a nonspherical body, V contains additional terms that can be used to determine $\boldsymbol{F}$, as will be shown later.

3.2 General and Restricted Two-Body Problem

A more useful and general expression of the gravitational law is in its vector formulation. Thus, referring to Fig. 3.2, the masses m and M are moving in an inertial reference frame E_1, E_2, E_3. It is desired to determine the motion of m relative to the larger mass M.

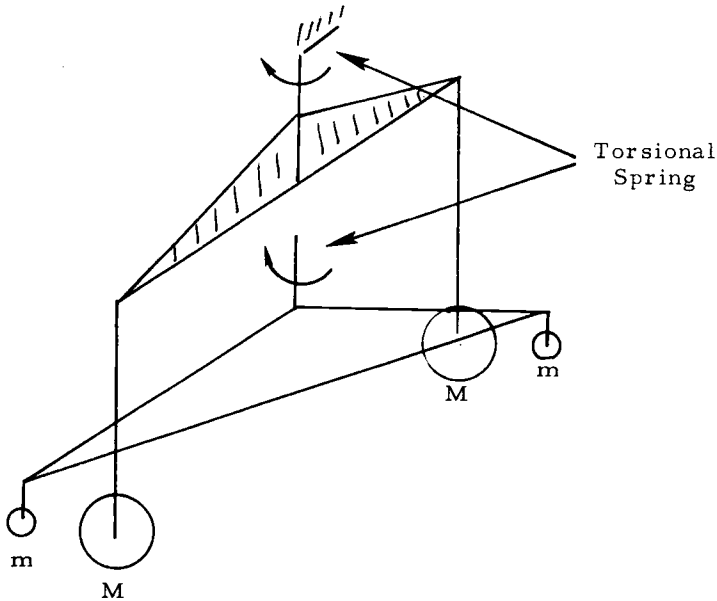


Fig. 3.1 Torsional balance system.

The general two-body problem results if $\boldsymbol{a} \neq 0$. Then, for the mass M,

$$M\ddot{\boldsymbol{a}} = \frac{GMm\boldsymbol{r}}{r^3} \tag{3.3}$$

and, for the mass m,

$$m\ddot{\boldsymbol{\rho}} = -\frac{GMm\boldsymbol{r}}{r^3} \tag{3.4}$$

Subtracting Eq. (3.4) from (3.3) yields

$$\begin{aligned} \ddot{\boldsymbol{r}} &= -\frac{G(M+m)\boldsymbol{r}}{r^3} \\ \ddot{\boldsymbol{r}} + \frac{G(M+m)\boldsymbol{r}}{r^3} &= 0 \end{aligned} \tag{3.5}$$

In a restricted two-body problem, the principal mass M is assumed fixed in inertial space. This implies that $M \gg m$, so that m does not affect the motion of M. Letting M be at the origin O (i.e., $\boldsymbol{a} = O$), then the force on the mass m is

$$\boldsymbol{F} = -\frac{GMm\boldsymbol{r}}{r^3} \tag{3.6}$$

or, in accordance with Eq. (3.4) with $\boldsymbol{\rho} = \boldsymbol{r}$,

$$\boldsymbol{F} = -\frac{GMm\boldsymbol{r}}{r^3} = m\ddot{\boldsymbol{r}} \tag{3.7}$$

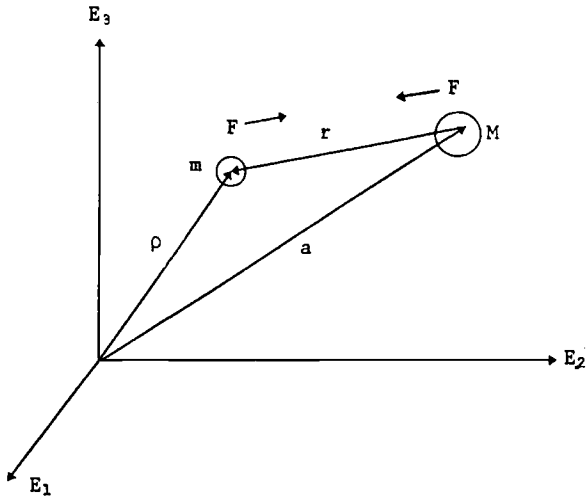


Fig. 3.2 Two-body system.

Therefore,

$$\ddot{\boldsymbol{r}} + \frac{GM\boldsymbol{r}}{r^3} = 0 \tag{3.8}$$

where $\ddot{\boldsymbol{r}} = \mathrm{d}^2\boldsymbol{r}/\mathrm{d}t^2$ = acceleration of mass m relative to the inertial frame. Equation (3.8) represents the motion of mass m in a gravitational field of mass M, assumed spherically symmetric and, therefore, concentrated at the origin of the reference system.

Equation (3.8) differs from Eq. (3.5) only in the gravitational constant term. The motion of the restricted two-body problem is therefore similar to that of the general two-body system and is affected by the magnitude of the gravitational term. The latter is a negligible effect when $m \ll M$, which is true for the satellites of the Earth and other planetary bodies.

3.3 Conservation of Mechanical Energy

Consider Eq. (3.8) in the form

$$\ddot{\boldsymbol{r}} + \frac{\mu\boldsymbol{r}}{r^3} = 0 \tag{3.9}$$

where $\mu = GM$, and m is assumed negligibly small compared to M.

Scalar multiplication of Eq. (3.9) by $\dot{\boldsymbol{r}}$ results in

$$\dot{\boldsymbol{r}} \cdot \ddot{\boldsymbol{r}} + \frac{\mu\boldsymbol{r} \cdot \dot{\boldsymbol{r}}}{r^3} = 0 \tag{3.10}$$

or

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\dot{\boldsymbol{r}} \cdot \dot{\boldsymbol{r}}}{2}\right) + \frac{\mu}{r^3}\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\boldsymbol{r} \cdot \boldsymbol{r}}{2}\right) = 0 \tag{3.11}$$

since $\mathrm{d}(\boldsymbol{r} \cdot \boldsymbol{r})/\mathrm{d}t = 2r\dot{r}$, etc. This equation can be integrated to yield

$$\frac{(\dot{r})^2}{2} - \frac{\mu}{r} = \varepsilon$$
$$= \text{specific mechanical energy} \tag{3.12}$$

Here $(\dot{r})^2/2 = v^2/2 =$ specific kinetic energy and $-\mu/r$ is the specific potential energy of the satellite. The specific potential energy is also equal to the gravitational potential function per unit mass.

3.4 Conservation of Angular Momentum

The specific angular momentum H of a satellite (angular momentum per unit mass) can be obtained by vector-multiplying Eq. (3.9) by $\boldsymbol{r}$. Then,

$$\boldsymbol{r} \times \ddot{\boldsymbol{r}} + \boldsymbol{r} \times \frac{\mu \boldsymbol{r}}{r^3} = 0 \tag{3.13}$$

which shows that

$$\begin{aligned} \boldsymbol{r} \times \ddot{\boldsymbol{r}} &= \frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{r} \times \dot{\boldsymbol{r}}) \\ &= \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{H} \\ &= 0 \end{aligned} \tag{3.14}$$

Consequently, $\boldsymbol{H} =$ const. This means that $\boldsymbol{r}$ and $\dot{\boldsymbol{r}}$ are always in the same plane. The actual solution for the satellite motion can be obtained by cross-multiplying Eq. (3.9) by $\boldsymbol{H}$. Then,

$$\ddot{\boldsymbol{r}} \times \boldsymbol{H} = \frac{\mu}{r^3}(\boldsymbol{H} \times \boldsymbol{r}) \tag{3.15}$$

or

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}(\dot{\boldsymbol{r}} \times \boldsymbol{H}) &= \frac{\mu}{r^3}(r^2\dot{\theta})r\hat{\theta} \\ &= \mu\dot{\theta}\hat{\theta} \\ &= \mu\frac{\mathrm{d}}{\mathrm{d}t}(\hat{r}) \end{aligned} \tag{3.16}$$

where the magnitude of the specific angular momentum $H = r^2\dot{\theta}$, $\hat{\theta}$ is a unit vector normal to the unit vector $\hat{r}$ along the $\boldsymbol{r}$ vector, and $\dot{\theta}$ is the angular rate of the $\boldsymbol{r}$ vector.

Integrating Eq. (3.16), one obtains

$$\dot{\boldsymbol{r}} \times \boldsymbol{H} = \mu\hat{r} + \boldsymbol{B} \tag{3.17}$$

where $\boldsymbol{B}$ is a constant of integration.

Furthermore, since

$$\begin{aligned} \boldsymbol{r} \cdot (\dot{\boldsymbol{r}} \times \boldsymbol{H}) &= \boldsymbol{r} \cdot (\mu \hat{r} + \boldsymbol{B}) \\ &= (\boldsymbol{r} \times \dot{\boldsymbol{r}}) \cdot \boldsymbol{H} \\ &= \boldsymbol{H} \cdot \boldsymbol{H} \\ &= H^2 \end{aligned}$$

$$H^2 = \mu r + rB\cos\theta \tag{3.18}$$

therefore,

$$r = \frac{H^2/\mu}{1 + (B/\mu)\cos\theta} \tag{3.19}$$

where

$$H^2/\mu = p = \text{semilatus rectum}$$

$$B/\mu = e = \left(1 + \frac{2\varepsilon H^2}{\mu^2}\right)^{1/2} = \text{eccentricity}$$

$$\theta = \text{true anomaly}$$

The general equation for the radius r in Eq. (3.9) is therefore of the form

$$r = \frac{p}{1 + e\cos\theta} \tag{3.20}$$

This is an equation for a conic section, an example of which is the ellipse illustrated in Fig. 3.3.

3.5 Orbital Parameters of a Satellite

The orbit ellipse geometry is shown in Fig. 3.3. The following notation is used:

a = semimajor axis = $(r_a + r_p)/2$
b = semiminor axis
e = eccentricity = $(r_a - r_p)/(r_a + r_p)$
θ = true anomaly
r_a = apogee radius = $a(1 + e)$
r_p = perigee radius = $a(1 - e)$
p = $a(1 - e^2) = b^2/a = r_p(1 + e) = r_a(1 - e)$ = semilatus rectum
γ = flight-path angle
= $\pi/2 - \beta$

The radial velocity component v_r can be found as follows:

$$\begin{aligned} v_r &= \frac{\mathrm{d}r}{\mathrm{d}t} = \frac{\mathrm{d}r}{\mathrm{d}\theta}\frac{\mathrm{d}\theta}{\mathrm{d}t} = \frac{\mathrm{d}r}{\mathrm{d}\theta}\dot{\theta} \\ &= \frac{\mathrm{d}r}{\mathrm{d}\theta} \cdot \frac{H}{r^2} \end{aligned} \tag{3.21}$$

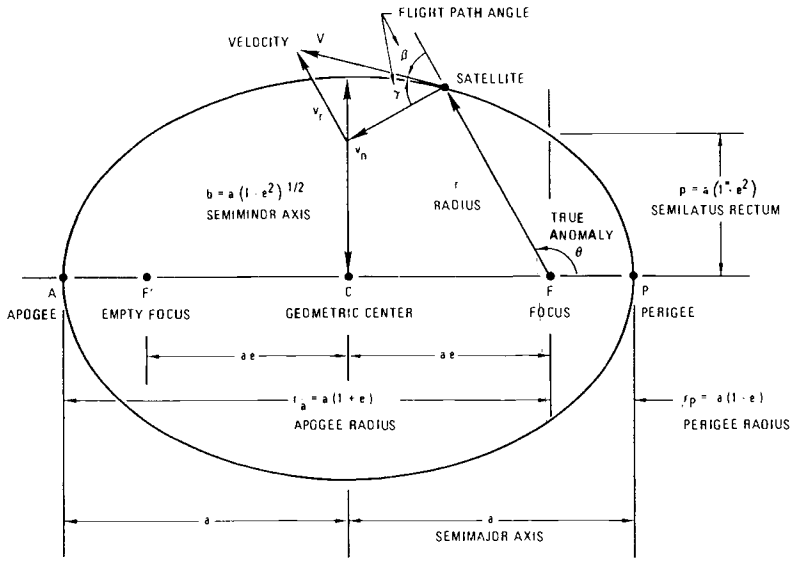


Fig. 3.3 Ellipse geometry (from Ref. 2).

but, since

$$\frac{\mathrm{d}r}{\mathrm{d}\theta} = \frac{\mathrm{d}}{\mathrm{d}\theta}[p(1 + e\cos\theta)^{-1}]$$

$$v_r = \frac{He\sin\theta}{p}$$

$$= \sqrt{\frac{\mu}{p}}e\sin\theta \tag{3.22}$$

The normal component v_n is found as

$$v_n = r\dot{\theta} = r\left(\frac{H}{r^2}\right) = \sqrt{\frac{\mu}{p}}(1 + e\cos\theta) \tag{3.23}$$

The flight-path angle γ is given by

$$\gamma = \cos^{-1}\left(\frac{v_n}{v}\right) = \tan^{-1}\left(\frac{v_r}{v_n}\right) \tag{3.24}$$

where

$$v = \left(v_n^2 + v_r^2\right)^{1/2}$$

$$= \left[\frac{\mu}{p}(1 + e^2 + 2e\cos\theta)\right]^{1/2} \tag{3.25}$$

The velocity at perigee v_p is

$$v_p = \sqrt{\frac{\mu}{p}}(1+e) \tag{3.26}$$

and, at apogee,

$$v_a = \sqrt{\frac{\mu}{p}}(1-e) \tag{3.27}$$

These are the maximum and minimum values of v in orbit, respectively.

The velocity v at any position in orbit is found from the energy or vis-viva equation (3.12) as follows:

$$\varepsilon = \frac{v^2}{2} - \frac{\mu}{r} \tag{3.28}$$

as $e \to 1$, $v_a \to 0$, $r_a \to 2a$; therefore, $\varepsilon = -\mu/2a$ since the energy ε remains constant. From Eq. (3.28),

$$v = \sqrt{\mu\left(\frac{2}{r} - \frac{1}{a}\right)} \tag{3.29}$$

For a circular orbit $r = a$ and, therefore, the velocity $v_c = \sqrt{\mu/r}$. For the escape trajectory, $a = \infty$ (parabola) and the velocity $v_{\text{esc}} = \sqrt{2}v_c$. The classification of the different possible conic sections of orbits in terms of the eccentricity e is as follows:

e	Orbit
0	Circle ($a = r$)
<1	Ellipse ($a > 0$)
1	Parabola [$a \approx \infty$ (undefined)]
>1	Hyperbola ($a < 0$)

The period of an elliptic orbit is $P = 2\pi/n$, where $n = \sqrt{\mu/a^3}$ = mean motion.

If the position of a satellite is desired at a specified time t, then it can be found from

$$M = n(t - \tau) \tag{3.30}$$

where τ = time of perigee passage, and M = mean anomaly, and from Kepler's equation,

$$M = E - e\sin E \tag{3.31}$$

which can be solved for E.

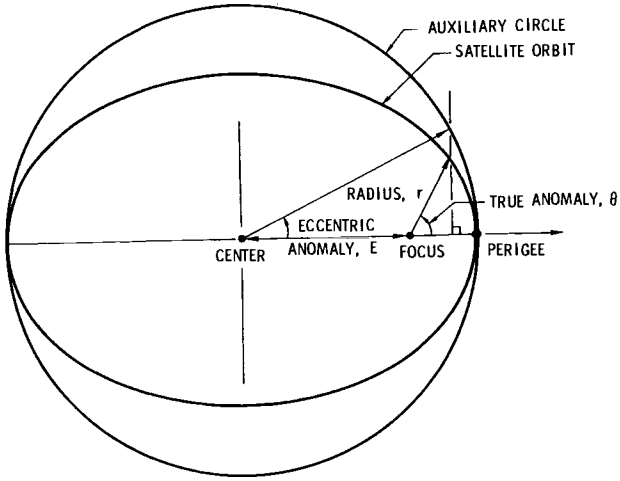


Fig. 3.4 Definition of eccentric anomaly (from Ref. 2).

The definition of the eccentric anomaly E is shown in Fig. 3.4. The true anomaly θ can be determined from

$$\tan\frac{\theta}{2} = \left[\frac{1+e}{1-e}\right]^{1/2} \tan\frac{E}{2} \tag{3.32}$$

The trigonometric arguments $\theta/2$ and $E/2$ are not always in the same quadrant.

Conversely, if the time t of travel from one point on the ellipse to another point is desired, then it can be found from Eq. (3.30), where M is given in Eq. (3.31).

3.6 Orbital Elements

The motion of a satellite around the Earth may be described mathematically by three scalar second-order differential equations. The integration of these equations of motion yields six constants of integration. It is these constants of integration that are known as the orbital elements.

The Keplerian orbital elements are often referred to as classical or conventional elements and are the simplest and easiest to use. This set of orbital elements can be divided into two groups: the dimensional elements and the orientation elements.

The dimensional elements specify the size and shape of the orbit and relate the position in the orbit to time (Fig. 3.3). They are as follows:

a = semimajor axis, which specifies the size of the orbit.

e = eccentricity, which specifies the shape of the orbit.

τ = time or perigee passage, which relates position in orbit to time (τ is often replaced by M, the mean anomaly at some arbitrary time t; the mean anomaly is a uniformly varying angle)

The orientation elements specify the orientation of the orbit in space (Fig. 3.5). They are as follows:

i = inclination of the orbit plane with respect to the reference plane, which is taken to be the Earth's equator plane for satellite orbits. 0 deg $\leq i \leq$ 180 deg).

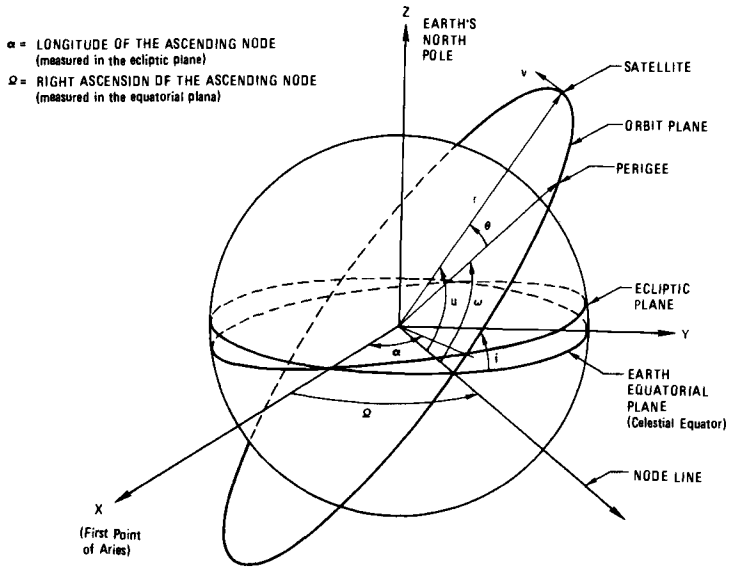


Fig. 3.5 Orientation of orbit in space (from Ref. 2).

For 0 deg $\leq i \leq$ 90 deg, the motion is "posigrade" or "direct"; for 90 deg $< i \leq$ 180 deg, the motion is termed "retrograde"

Ω = right ascension of the ascending node (often shortened to simply "node"); Ω is measured counterclockwise in the equator plane, from the direction of the vernal equinox to the point at which the satellite makes its south-to-north crossing of the equator (0 deg $\leq \Omega <$ 360 deg)

ω = argument of perigee; ω is measured in the orbit plane in the direction of motion, from the ascending node to perigee (0 deg $\leq \omega <$ 360 deg)

The angles i and Ω specify the orientation of the orbit plane in space. The angle ω then specifies the orientation of the orbit in its plane. The argument of latitude u defines the position of the satellite relative to the node line.

Still another system of specifying the satellite state vector involves the scalar quantities of

v = velocity
r = radius
Ω = node
γ = flight-path angle
δ = geocentric latitude
A_z = azimuth of v from true north

Various sets of elements are used in orbit determination, the inertial rectangular $(x, y, z, \dot{x}, \dot{y}, \dot{z})$ and the spherical $(\alpha, \delta, \beta, A_z, r, v)$, where α is longitude

Table 2.1 Orbital element systems

Type	1	2	3	4	5	6
1 Cartesian (ECI)	x	y	z	$\dot{x}$	$\dot{y}$	$\dot{z}$
2 ADBARV (ADB)	α (right ascension)	δ (declination)	β (flight-path angle measured from vertical	A (azimuth)	r (distance from geocenter)	v (inertial speed)
3 Classical (CLA)	a (semimajor axis)	e (eccentricity)	i (inclination)	Ω (right ascension of ascending node)	ω (argument of perigee)	M (mean anomaly)
4 Geographic (GEO)	λ (longitude)	ϕ (latitude)	β	A	r	v
5 Equinoctial (EQU)	$a_f = e\cos(\Omega + \omega)$	$a_g = e\sin(\Omega + \omega)$	n (mean motion)	L (mean longitude) $L = M + \Omega + \omega$	$= \sin\Omega \tan\frac{i}{2}$	$= \cos\Omega \tan\frac{i}{2}$
6 Spherical (SPH)	r	α	δ	v	$\alpha_v =$ (right ascension of velocity)	δ_v (declination of velocity)

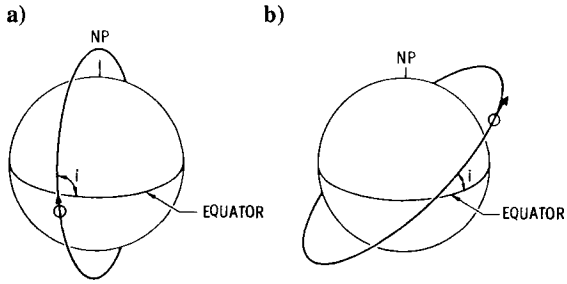


Fig. 3.6 Earth orbits: a) retrograde orbit: $90 < i < 180$ deg, $180 < A_z < 360$ deg; b) posigrade orbit: $0 < i < 90$ deg, $0 < A_z < 180$ deg.

and $\beta = \pi/2 - \gamma$ are examples of such elements. Note that both of these sets of elements give the satellite's position as a point in space at any specific time. Several orbital element systems are listed in Table 3.1. Posigrade and retrograde orbits are illustrated in Fig. 3.6.

References

[1]Luther, G. G., and Tower, W. R., "Redetermination of the Newtonian Gravitational Constant G," *Physical Review Letters,* Vol. 48, No. 3, 18 Jan. 1982, pp. 121–123.

[2]Cantafio, L. (ed.), *Space-Based Radar Handbook,* Artech House, Norwood, MA, 1989.

Selected Bibliography

Battin, R. H., *An Introduction to the Mathematics and Methods of Astrodynamics,* AIAA Education Series, AIAA, Washington, DC, 1987.

Bate, R. R., Mueller, D., and White, J. E., *Fundamentals of Astrodynamics,* Dover, New York, 1971.

Flury, W., *Vorlesung Raumfahrtmechanik,* Technische Hochschule Damstadt, 1994.

Kaplan, M. H., *Modern Spacecraft Dynamics and Control,* Wiley, New York, 1976.

Prussing, J. E., and Conway, B. A., *Orbital Mechanics,* Oxford, 1993.

Roy, A. E., *Orbital Motion,* 3rd ed., Adam Hilger, Bristol and Philadelphia, 1988.

Taff, L. G., *Celestial Mechanics,* Wiley, New York, 1985.

Wertz, J. R., and Larson, W. J., eds., *Space Mission Analysis and Design,* Kluwer Academic Publishers, 1991.

Problems

3.1. The period of revolution of a satellite is 106 min. Find the apogee altitude if the perigee altitude is 200 km.

$$(\mu = 3.986 \times 10^5 \text{ km}^3/\text{s}^2)$$

3.2. Find the period of revolution of a satellite if the perigee and apogee altitudes are 250 and 300 km, respectively.

3.3. Find the maximum and minimum velocity of the Earth if the eccentricity of the Earth's orbit about the sun is 1/60. What is the mean velocity if the mean distance to the sun is 149.6×10^6 km? ($\mu s = 1.327 \times 10^{11}$ km^3/s^3)

3.4. Show that for a satellite moving on an elliptical orbit the velocity at the time of passage through the minor axis is equal in magnitude to the local circular velocity (v_c). (Hint: $v_c = \sqrt{\mu/r}$.)

3.5. A satellite in a circular orbit at an altitude h_c above the Earth's surface is given a velocity v_0 with a flight-path angle γ. What is the magnitude of v_0 if the perigee altitude of the resultant orbit is to be equal to $h_p (h_p < h_c)$?

3.6. A sounding rocket is launched from a planet. Find initial velocity to reach height H above the surface. Assume a spherical planet of radius R and gravity g at surface.

3.7. Find the escape velocity from the moon's surface. Assume that

$$g_{\text{moon}} = \frac{1}{6} g_{\text{Earth}}$$

$$r_{\text{moon}} = \frac{1}{4} r_{\text{Earth}}$$

3.8. Assuming that the period of Mars about the sun is 687 Earth days, find the mean distance of Mars to sun if the Earth distance to sun is 149.5×10^6 km.

3.9. Find the mass of the sun using Kepler's third law.

3.10. For a Keplerian orbit with period $P = 205$ min, eccentricity $e = 0.40$, and true anomaly $\theta = 60$ deg, find the time t since passage of perigee.

3.11. Given the following orbit:

$$h_p = 200 \text{ km}$$

$$h_a = 600 \text{ km}$$

a) What is the time interval over which the satellite remains above an altitude of 400 km? Assume a spherical Earth with radius = 6378 km and $\mu = 3.986 \times 10^5$ km^3/s^2.

b) What additional information is needed to solve part a if the Earth is assumed to be nonspherical?

3.12. A spaceship is in a 200-km circular orbit above a spherical Earth. At $t = 0$, it retrofires its engine, reducing its velocity by 600 m/s. How long (in minutes of time) does it take to impact the Earth?

3.13. An object was observed at a distance of $(1.05)^{2/3}$ Earth radii from the center of the Earth. Sixteen minutes later, the same object was observed at a position 60 deg (measured at the Earth center) from the original position. Show whether the object is in a circular orbit. Assume that $\mu = 0.00553$ (ER^3/min^2).

Selected Solutions

3.1. 1882 km

3.2. 1.5 h

3.3. 29.29, 30.28 km/s

3.10. 14 min, 56 s

4
Position and Velocity as a Function of Time

4.1 General Relationships

In this chapter, we will discuss what is known historically as "Kepler's problem." Succinctly stated, this problem is one of finding the state (position and velocity) of an object in orbit at a specified time t, given the state at some reference time t_0.

For example, let us assume that, at the reference (or initial condition) time, the object is not necessarily at its perifocal point. That is, we will assume the state at t_0 to be available in the following terms:

a = semimajor axis
e = eccentricity
i = inclination
Ω = right ascension of the ascending node
ω = argument of perifocal point
θ = true anomaly (not zero in the present discussion)

To fix ideas firmly, let us review what these terms signify. First, a and e specify the size and shape of the orbit as illustrated in Fig. 4.1. The semiminor axis b is related to the semimajor axis a by the relationship of

$$b = \sqrt{a^2 - c^2}$$

where

$$c = ae$$

or

$$b = a\sqrt{1 - e^2}$$

Second, the orientation of the orbit is specified by i, Ω, and ω, i.e., inclination, right ascension of the ascending node, and argument of perifocus, respectively, as illustrated in Fig. 4.2.

Third, the position of the object is specified by the sixth term, θ, true anomaly, which is measured in the orbit plane from the perifocal axis, as shown in Fig. 4.3. In our discussion, the object is not at perifocus and has a nonzero value for θ_0.

The problem at hand is to find the position θ corresponding to time t, which may be before or after the reference time of t_0. Note that the other terms, $a - \omega$, do not change over this time. Such an orbit is called a "Keplerian" orbit, where the only influence experienced by the object is the gravitational force of the attracting body represented by a spherical potential field. The strength of a spherical field is a function only of the distance from the center of the attracting body.

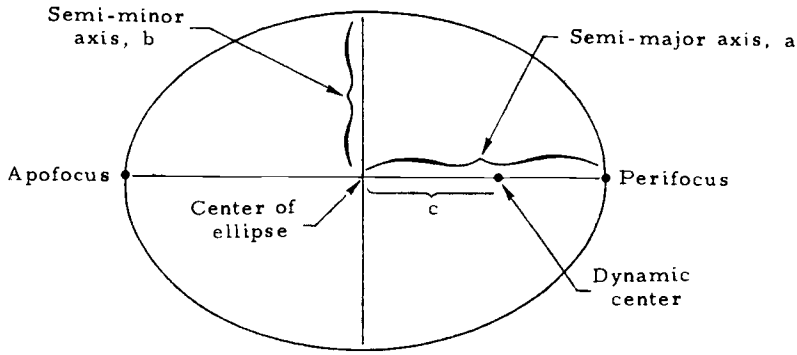


Fig. 4.1 Size and shape of an orbit.

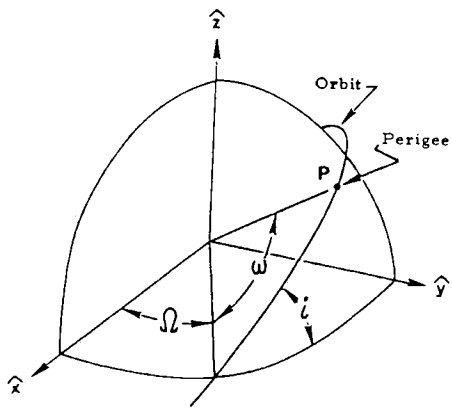


Fig. 4.2 Orientation of an orbit.

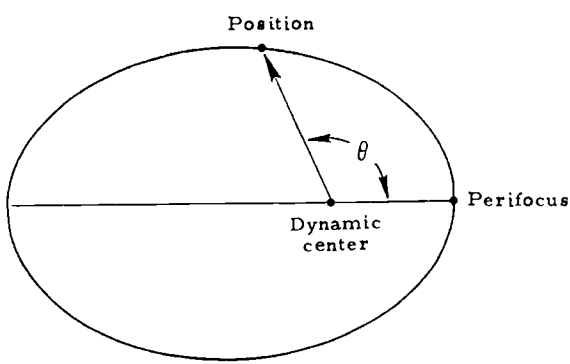


Fig. 4.3 Position of true anomaly.

Generally speaking, the potential field is not spherical, and the first five terms, $a - \omega$, do change as a function of time. There are, however, many problems that can be solved adequately by considering the orbit to be Keplerian. Non-Keplerian influences like the nonspherical potential, atmospheric drag, lunar and planetary effects, and solar pressure are referred to as "perturbative" effects and are discussed more fully in Chapter 8.

Now, let us solve our problem. First, we must relate the initial position θ_0 to the initial time t_0. Equations (4.1–4.3) transform θ_0, the true anomaly at t_0, to its equivalent eccentric anomaly, usually symbolized by the letter E_0.

$$\sin E_0 = \frac{\sqrt{1 - e^2}\sin\theta_0}{1 + e\cos\theta_0} \tag{4.1}$$

$$\cos E_0 = \frac{e + \cos\theta_0}{1 + e\cos\theta_0} \tag{4.2}$$

$$E_0 = \tan^{-1}\left(\frac{\sin E_0}{\cos E_0}\right) \tag{4.3}$$

Kepler's equation then relates the eccentric anomaly E_0 to its mean anomaly M_0.

$$M_0 = E_0 - e\sin E_0 \tag{4.4}$$

Finally, the mean anomaly M_0 is related to time t_0 by

$$M_0 = n(t_0 - T) \tag{4.5}$$

where n is the mean motion and T the time at the last previous perifocus passage. We need not determine the exact value of T but merely note its existence for the time being. The mean motion is determined from

$$n = \frac{2\pi}{P} \tag{4.6}$$

where

$$P = 2\pi\sqrt{a^3/\mu} \tag{4.7}$$

In Eq. (4.7), a is the semimajor axis that was given at the start of this problem, and μ is the product of the universal gravitation constant and the mass of the attracting body. If the attracting body is the Earth, μ has the value of 398,600.8 km^3/s^2. You may recognize Eq. (4.7) as Kepler's third law, which states that the square of the orbit period P is proportional to the cube of the orbit's semimajor axis a.

Before proceeding, let us look at these equations from a geometrical view. Figure 4.4 illustrates the relationship of θ and E as expressed in Eqs. (4.1–4.3). For convenience and to demonstrate generality, the subscript zeros have been temporarily removed in the geometrical discussion.

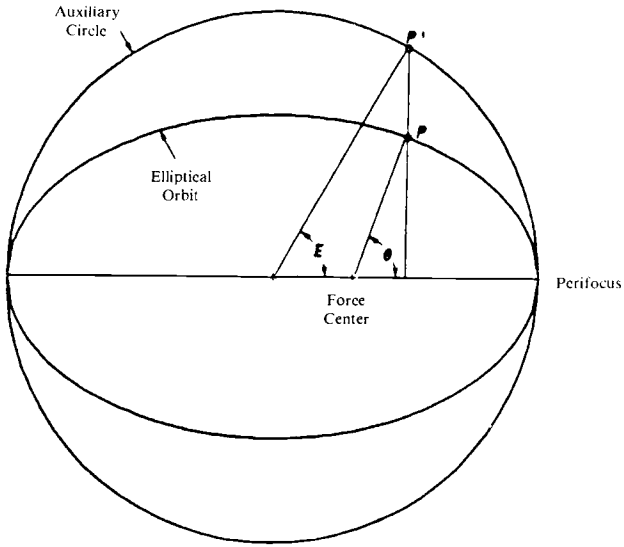


Fig. 4.4 Relationship of true anomaly and eccentric anomaly.

The circle that just fits snugly outside the elliptical orbit is called the "auxiliary circle." It has nothing to do with reality but is a convenient concept introduced to relate position and time. From Kepler's second law, we readily see that, just as the motion of point P in the elliptical orbit is not uniform (i.e., it moves faster when near the perifocus and slower in the region of apofocus), the motion of its image P' in the auxiliary circle is also not uniform.

Now, let us consider another auxiliary circle, as shown in Fig. 4.5. A position P'' on this circle is described by a variable called the mean anomaly M, which is an angular quantity measured at the center of the circle from some reference direction. As we shall see in a moment, the motion of P'' is uniform, i.e., it revolves at a constant speed that is not like the motion of P and P'.

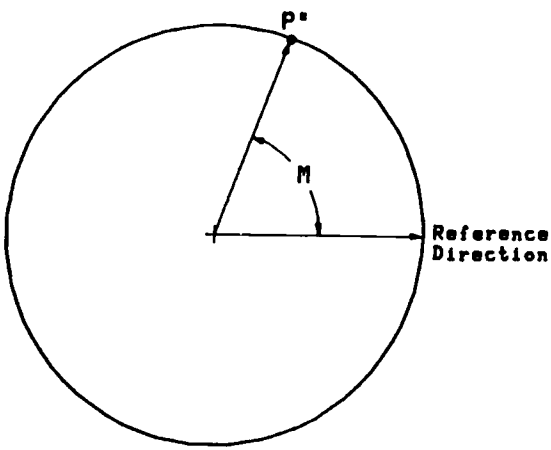


Fig. 4.5 Auxiliary circle of mean anomaly.

Now, consider yet another auxiliary circle. This one is like the clock we are accustomed to seeing; i.e., it shows the time measured from some commonly agreed on reference time such as T. All three auxiliary circles can be viewed as clocks, the time clock, the M clock, and the E clock, each of which is moving at a different rate.

As time moves uniformly in the time clock, P'' in the M clock moves by the relationship of $M = n(t - T)$, where n is a constant determined from Eqs. (4.6) and (4.7). Since n is a constant, we see that the motion of M is also uniform. The M and E clocks are related by Kepler's equation [Eq. (4.4)]. Clearly, the motion of the E clock is not uniform. It is so only if the eccentricity e is zero, making $E = M$. In this special case, the auxiliary circle of E is, in fact, the orbit.

We can synchronize these clocks by noting that when P in the elliptical orbit is at its perifocus, namely, at time T, P' in the E clock is at its zero position. Likewise, P'' in the M clock is also at its zero position.

Now, let us return to our problem and see how easy (or difficult) it is to find a position θ in the orbit that corresponds to some arbitrary time t. First, we must advance the time clock to t. The corresponding point in the M clock is found from

$$M = n(t - T) \tag{4.8}$$

From Eq. (4.5),

$$M_0 = n(t_0 - T)$$

so that

$$M = n(t - t_0) + M_0 \tag{4.9}$$

and we have eliminated the need for T.

Next, we determine E from M through Kepler's equation, $M = E - e \sin E$. Here we encounter our first obstacle. When we used this equation in Eq. (4.4), we were solving for M given E. Now, we are solving for E given M. Since the equation is transcendental, we cannot invert it directly to solve for E. Instead, we must resort to a technique such as the Newton–Raphson successive approximation method to solve for the desired quantity.

Assuming for the moment that E is determined, the final step then is to find θ (true anomaly) from the following equations:

$$\sin\theta = \frac{\sqrt{1 - e^2}\sin E}{1 - e\cos E} \tag{4.10}$$

$$\cos\theta = \frac{\cos E - e}{1 - e\cos E} \tag{4.11}$$

$$\theta = \tan^{-1}\left(\frac{\sin\theta}{\cos\theta}\right) \tag{4.12}$$

Now, let us see how Kepler's equation is solved.

4.2 Solving Kepler's Equation

First, we move mean anomaly M in Eq. (4.4) to the right side of the equation and define a function $f(E)$, of which we seek the roots at $f(E) = 0$.

$$f(E) = E - e \sin E - M \tag{4.13}$$

Recall from the definitions that when E is negative, so is M and, for negative values of E and M, $f(E)$ is an odd function, i.e.,

$$f(-E) = -E + e \sin E + M = -f(E) \tag{4.14}$$

We further note that, for $0 \leq E \leq \pi$,

$$f(2n\pi - E) = (2n\pi - E) - e \sin(2n\pi - E) - (2n\pi - M) \tag{4.15}$$

where $n = 0, 1, 2, \ldots$, and

$$f(2n\pi - E) = -f(E), \qquad 0 \leq E \leq \pi \tag{4.16}$$

Thus, for any value of E, we need only consider the solution in the interval of $0 \leq E \leq \pi$ and adjust the results by a sign and multiples of 2π as appropriate.

Now, consider a Taylor's series expansion about E of

$$f(E + \Delta E) = f(E) + f'(E)\Delta E + \frac{f''(E)}{2!}\Delta E^2 + \cdots \tag{4.17}$$

Discarding second-order $(\Delta E)^2$ terms and higher, we have

$$f(E + \Delta E) = f(E) + f'(E)\Delta E \tag{4.18}$$

which we shall interpret as follows: Assuming that we have E as an approximate root to the function $f(E)$, we wish to find an appropriate ΔE to drive the function at $(E + \Delta E)$ to its root, namely, zero. With this interpretation, Eq. (4.18) becomes

$$0 = f(E) + f'(E)\Delta E \tag{4.19}$$

and

$$\Delta E = -\frac{f(E)}{f'(E)} \tag{4.20}$$

which is the correction term in Newton–Raphson's method of successive approximation. In our case,

$$f(E) = E - e \sin E - M \tag{4.21}$$

and

$$f'(E) = 1 - e \cos E \tag{4.22}$$

and the general procedure can be outlined as follows:

1) Pick a starting value E_0, and let it be $E_i = E_0$. (Typically M is used as the starting value, i.e., $E_0 = M$.)
2) Calculate $f(E_i) = E_i - e \sin E_i - M$.
3) Calculate $f'(E_i) = 1 - e \cos E_i$.
4) Calculate $\Delta E_i = -[f(E_i)/f'(E_i)]$.
5) Determine a new E from $E_{i+1} = E_i + \Delta E_i$.
6) Repeat steps 2–5 until $|\Delta E_i| < \varepsilon$, where ε is some constant appropriately small to correspond to the extent of precision desired in the calculation.

Now that we have outlined a general procedure, we may ask: How efficient is this method of iteration? Before answering this question, we note that the literature on the solution of Kepler's equation is extensive.[1–5] Some of the motivations for seeking alternative methods of computation are

1) Preventing the solution from diverging.
2) Faster convergence (i.e., fewer number of iterations).
3) An algorithm suitable for machine computation.

Basically, the techniques suggested are Newton–Raphson algorithms with suitable starting values for the iteration. These techniques can be found in the references listed at the end of this section.

For expediency, we will not explore these techniques, except to note that all suffer from some computational difficulties or abnormalities, such as the following:

1) Outright divergence.
2) Convergence to fixed points where, after every n iterations, the same set of values of the eccentric anomaly is repeated.
3) Strange attractor behavior, where successive estimates of E remain in a bounded region in an almost random fashion without showing any sign of convergence or divergence.

Now, let us see how efficient the procedure outlined in the six steps is in solving Kepler's equation. In order to do this, Fig. 4.6 is shown giving the number of iterations required for the algorithm to converge to $|\Delta E| < 10^{-12}$ (rad). In every case, the starting value is $E_0 = M$, and the number of iterations required is noted for each e-M pair of starting values, where e is varied from 0.01 to 0.99 in steps of 0.02 and M from 0 to 180 deg in steps of 2 deg. The results are tabulated in an e-M plane, where the number of iterations is printed at the appropriate grid point so that, as an aggregate, they present the form of a contour plot.

Clearly, the number of iterations increases as e approaches 1 and M approaches 0. This is so, for $f'(E)$ in Eq. (4.22) is approaching 0. The uppercase letters of A appearing in the left-hand side of the top row of the figure denote those cases that require 10 iterations or more. For example, for $M = 10$ deg and $e = 0.99$, convergence occurred after 59 iterations! For $M = 18$ deg and $e = 0.99$, the process failed to converge after 100 iterations (denoted by the letter X in the figure).

The situation worsens when the e-M plane is extended at the upper left-hand corner, as shown in Fig. 4.7. Here, e is examined from 0.990 to 0.999 in steps of 0.0005 and M from 0 to 60 deg in steps of 1 deg. The preponderance of A and X suggests the need for an alternative method in solving Kepler's equation, particularly in the region of high e and low M.

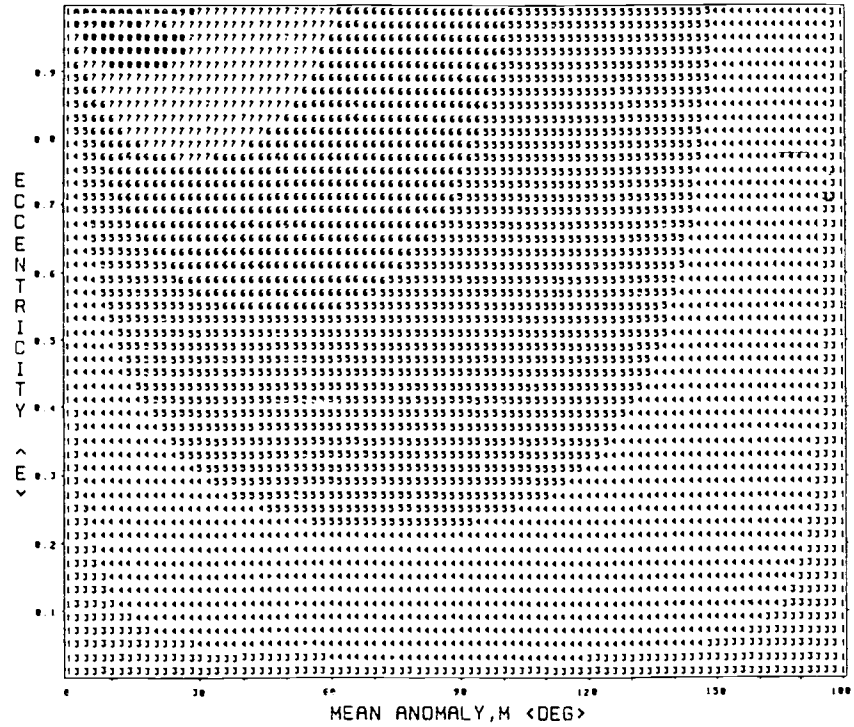


Fig. 4.6 Number of iterations needed to converge by Newton's method using the starting value $E_0 = M$.

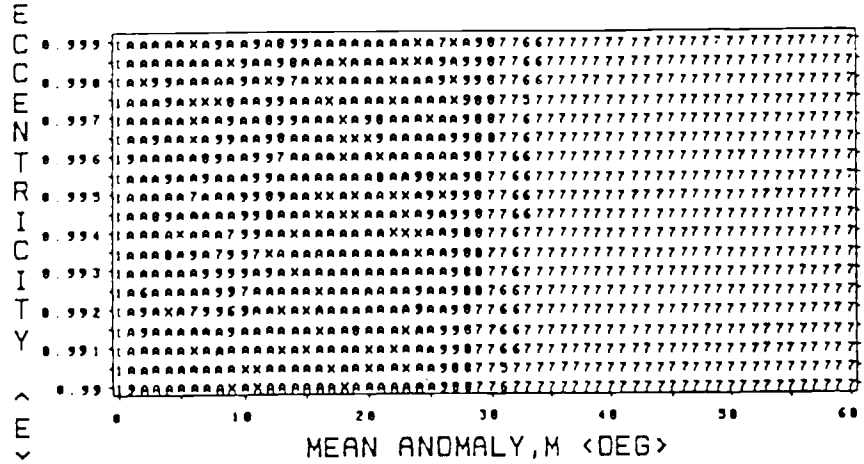


Fig. 4.7 Number of iterations needed to converge by Newton's method in the high-e/low-M region using the starting value of $E_0 = M$.

Recently, Conway,[6] of the University of Illinois, has applied a root-finding method of Laguerre (1834–1886) to the solution of Kepler's equation. Although the method is intended for finding the roots of a polynomial, it works equally as well for Kepler's equation, which is transcendental. Laguerre's iteration method requires the calculation of $f(E)$, $f'(E)$, and $f''(E)$ at each step and has these remarkable properties:

1) It is cubically convergent for simple roots.
2) For algebraic equations with only real roots, it is convergent for every choice of real initial estimate.

In over 500,000 test solutions, Conway has always found the algorithm to converge to the proper value of eccentric anomaly.

Mathematically, the method consists of solving successively the following equation:

$$E_{i+1} = E_i - \frac{nf(E_i)}{f'(E_i) \pm \sqrt{H(E_i)}} \tag{4.23}$$

where

$$H(E_i) = \left|(n-1)\left\{(n-1)\left[f'(E_i)\right]^2 - nf(E_i)f''(E_i)\right\}\right| \tag{4.24}$$

Normally, n is the degree of the polynomial. For our purposes here, we may safely use an arbitrary choice of $n = 5$, making Eqs. (4.23) and (4.24) appear as

$$E_{i+1} = E_i - \frac{5f(E_i)}{f'(E_i) \pm 2\sqrt{\left|4\left[f'(E_i)\right]^2 - 5f(E_i)f''(E_i)\right|}} \tag{4.25}$$

where

$$f(E_i) = E_i - e\sin E_i - M \tag{4.26}$$

$$f'(E_i) = 1 - e\cos E_i \tag{4.27}$$

$$f''(E_i) = e\sin E_i \tag{4.28}$$

Note that, when Eq. (4.25) is calculated, the sign in the denominator should be chosen so that $|E_{i+1} - E_i|$ is small as possible.

Figures 4.8 and 4.9 show the comparable results when Laguerre's method is used in place of Newton's method. Again, $E_0 = M$ is used as the initial starting value for all cases. Note the remarkable improvement in the number of iterations required, especially in the region where e approaches 1 and M approaches 0. Note also the lack of sensitivity to the starting value. Even with the unsophisticated choice of $E_0 = M$, four iterations at the most are sufficient for Laguerre's method to converge to within 10^{-12} rad almost everywhere in the e-M plane. The use of a good starting value would improve the speed of convergence, but this seems unnecessary.

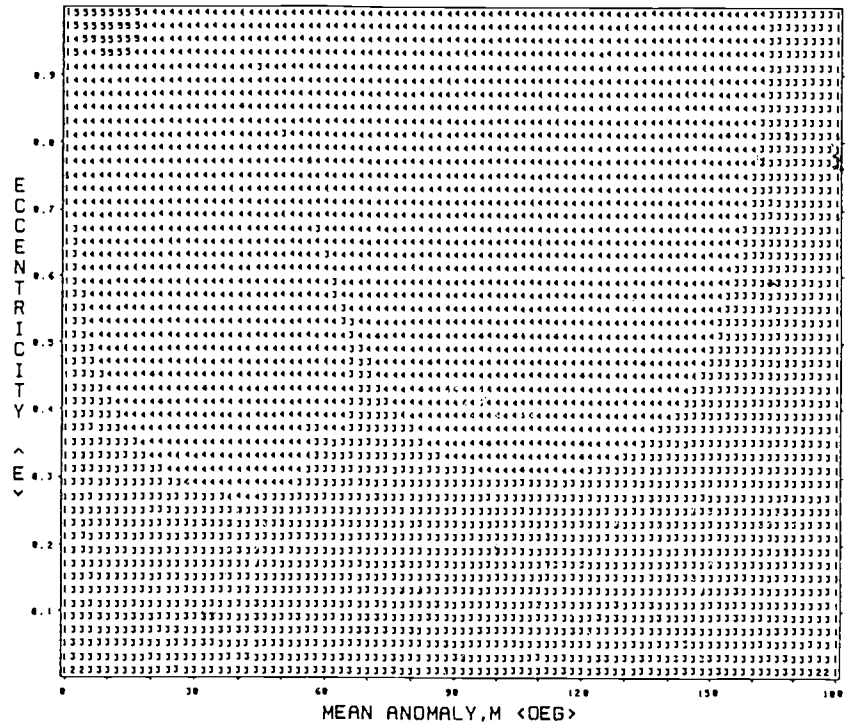


Fig. 4.8 Number of iterations needed to converge by Laguerre's method using the starting value of $E_0 = M$.

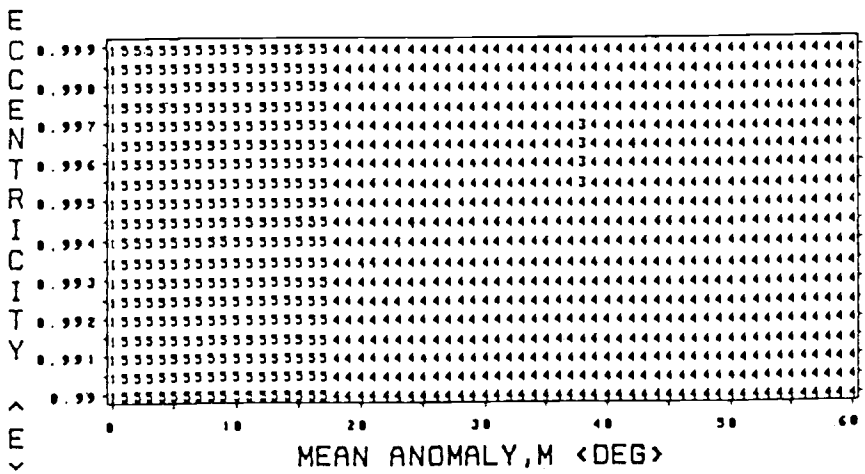


Fig. 4.9 Number of iterations needed to converge by Laguerre's method in the high-e/low-M region using the starting value of $E_0 = M$.

For those who wish to use a starting value other than $E_0 = M$, the following formula is suggested:

$$E_0 = M + \frac{e \sin M}{B + M \sin e} \tag{4.29}$$

where

$$B = \cos e - \left(\frac{\pi}{2} - e\right) \sin e \tag{4.30}$$

The computational requirements of this starting value are modest, and the resulting number of iterations needed to solve Kepler's equation is shown in Figs. 4.10 and 4.11.

As a comparison with Figs. 4.8 and 4.9 shows, one iteration less is needed in most instances to arrive at the same accuracy. In some instances, the number of iterations is reduced by two. Gratifyingly, this reduction extends in part to the high-e/low-M region.

It is interesting to note how close the starting value is to the actual value. This closeness can be seen in Figs. 4.12 and 4.13, where the percent error is noted for each starting value in the e-M plane. Specifically, the following percent error is

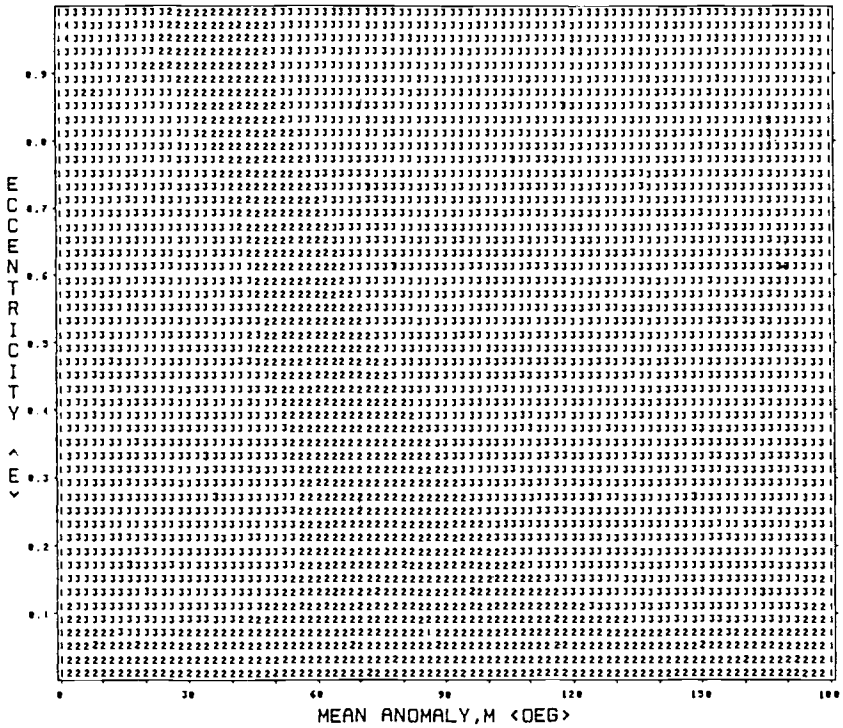


Fig. 4.10 Number of interations needed to converge by Laguerre's method using the starting value of Eqs. (4.29) and (4.30).

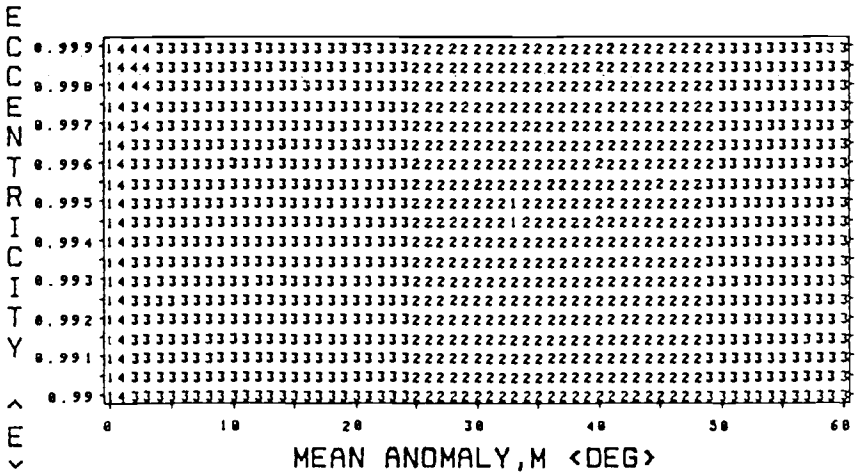


Fig. 4.11 Number of interations needed to converge by Laguerre's method in the high-*e*/low-*M* region using the starting value of Eqs. (4.29) and (4.30)

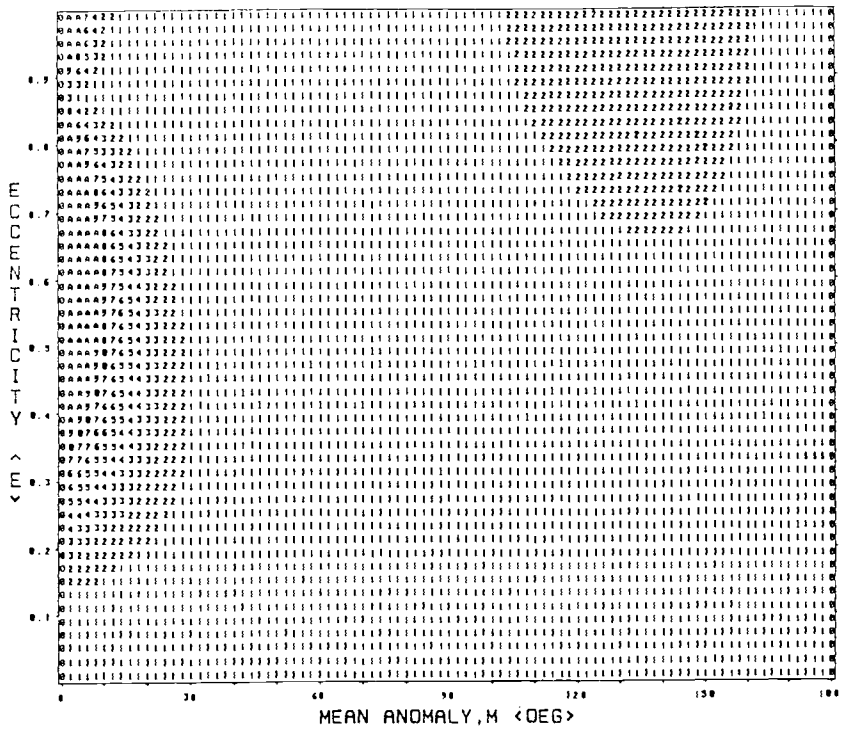


Fig. 4.12 Percent of starting values, as calculated by Eqs. (4.29) and (4.30).

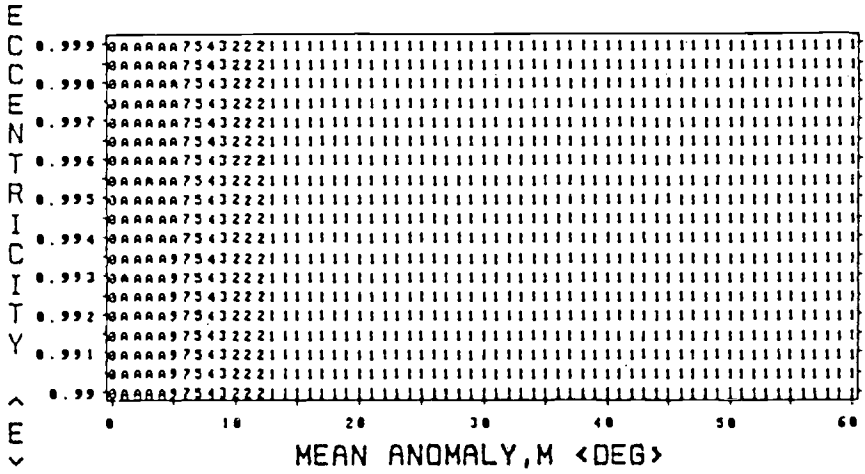


Fig. 4.13 Percent error of starting values in the high-*e*/low-*M* region, as calculated by Eqs. (4.29) and (4.30).

rounded up to the nearest integer and presented so that, as an aggregate, a contour of like percent error is readily visible from the figure.

$$\text{Percent error} = \frac{E_0 - E}{E} \times 100 \tag{4.31}$$

where E_0 is the starting value calculated from Eqs. (4.29) and (4.30) and E is the actual eccentric anomaly correct to within 10^{-12} rad.

With the exception of the high-*e*/low-*M* region, where the percent errors are 10% or greater (denoted by the letters *A* in the figures), the errors are less than 1% in magnitude in most instances.

Given that the starting values of Eqs. (4.29) and (4.30) are very close to the actual eccentric anomaly values, we may ask if their application to Newton's original method will eliminate the divergence problem encountered earlier. The results of this application are shown in Figs. 4.14 and 4.15. The divergence problem has been eliminated, but a fair number of iterations are still required in the extremely high-*e* and low-*M* region.

By way of comparison, the accuracy requirements of the solved-for eccentric anomaly are relaxed to 10^{-5} rad (approx. 0.0006 deg). Figures 4.16a–4.16c show the number of iterations needed to converge to this accuracy. The superiority of the combined Laguerre/initial starting value method is clearly demonstrated in these and the previous figures.

So far, the comparisons between Newton's and Laguerre's methods have been based on the number of iterations required to converge to a specified tolerance. That is, we have been concerned primarily with the issue of convergence. We have viewed a smaller number of iterations as a sign of faster convergence, meaning that, as the desired value is approached more efficiently, we expect diverging to be minimized. A smaller number of iterations, however, does not necessarily imply a smaller (or shorter) computation time, as we shall see in the computing time chart of Fig. 4.17.

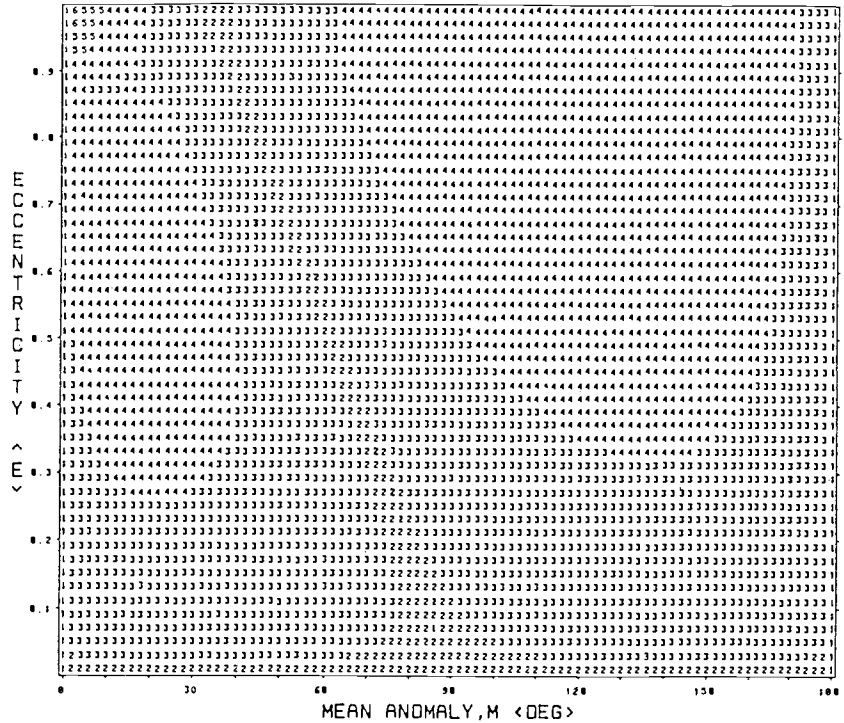


Fig. 4.14 Number of iterations needed to converge by Newton's method using the starting value of Eqs. (4.29) and (4.30).

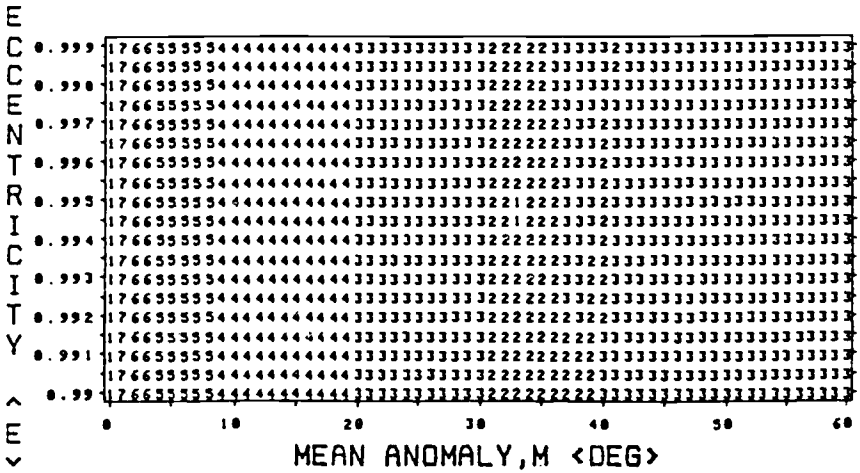


Fig. 4.15 Number of iterations needed to converge by Newton's method in the high-*e*/low-*M* region using the starting value of Eqs. (4.29) and (4.30).

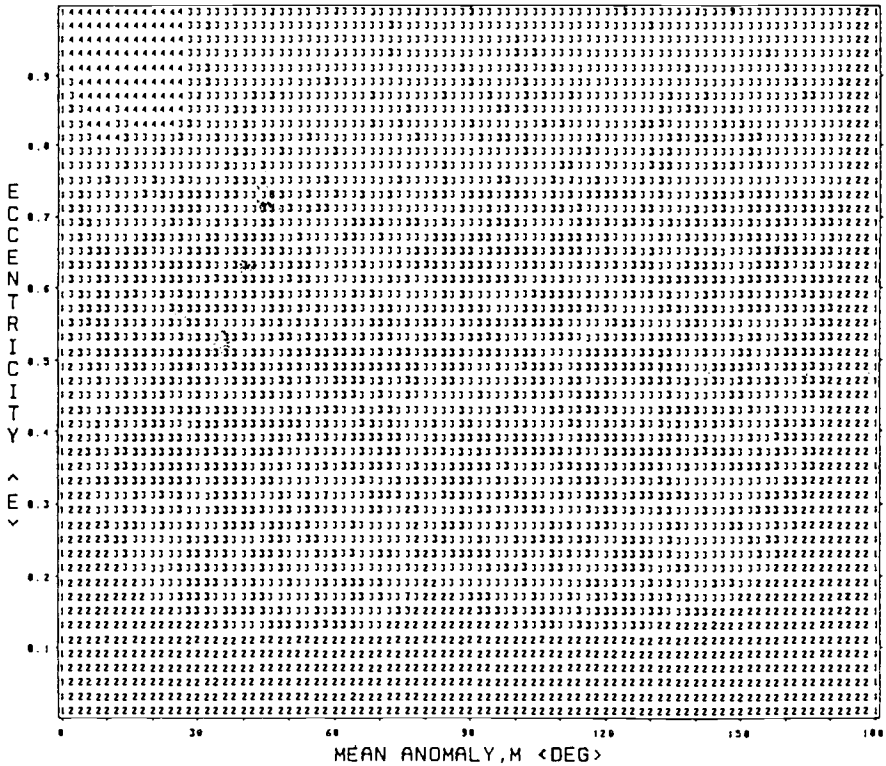


Fig. 4.16a Number of iterations needed to converge to within 0.00001 rad by Laguerre's method using the starting value of $E_0 = M$.

Before we discuss Fig. 4.17 in detail, we must understand the following. First, for any given algorithm, the associated computing time will vary with the computer used to perform the calculation. Second, the computational efficiency will depend also on the manner in which the algorithm is implemented, that is, coded. For these reasons, the traditional approach is to break down the algorithm by its mathematical operations and then compare the number of operations from algorithm to algorithm. For example, an algorithm can be broken down to 5 additions, 3 subtractions, 4 multiplications, 2 divisions, 1 square root, and so on. Needless to say, the number of iterations that are needed must be considered, and it is the total number of operations that is of concern for each of the specified e-M values. This is all well and fine, but the true difference may not be seen until the total number of operations is ultimately translated into a single quantity, namely, computing time.

Figure 4.17 then shows the computing times that are needed if the solutions are sought via the IBM PC-XT system. The coding is done in the APL language in the most straightforward manner possible. Special testing and branching (such as for $M = 0$) have been deliberately avoided. The computing units, where 4660.874753 units are equal to 1 s of time, are unique to the IBM PC-XT, but this should be of no concern to us at the moment. So long as the computations are done on the same computer, the comparisons are all valid. Since, on any computer, multiplications

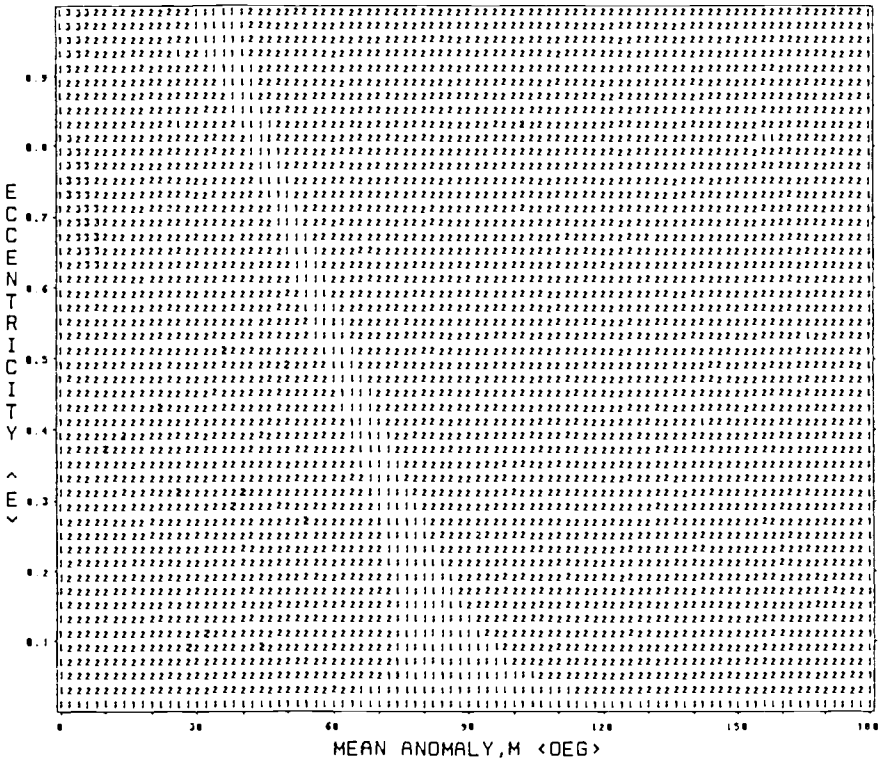


Fig. 4.16b Number of iterations needed to converge to within 0.00001 rad by Laguerre's method using the starting value of Eqs. (4.29) and (4.30).

and divisions take relatively longer to accomplish than additions and subtractions, transcendental functions take longer than multiplications and divisions, and so on, so that, in a sense, the general nature of Fig. 4.17 is applicable to other computing systems as well. Strictly speaking, however, a computing timing study, whose results we will discuss shortly, must be performed for the specific computer on which the algorithm is to be implemented.

In Fig. 4.17, the short dashed line shows the computing time of Laguerre's method where $E_0 = M$ is used as a starting value. The long dashed line is also for Laguerre's method, but with a starting value as calculated by Eqs. (4.29) and (4.30). The difference in time of the two lines is clearly the added time needed to calculate Eqs. (4.29) and (4.30). Interestingly, the steepness of the slope of these lines is such that the value of the upper line at any n iterations is less than the value of the lower line at $n + 1$ iterations. In other words, if the introduction of Eqs. (4.29) and (4.30) reduces the number of iterations by one or more, a saving in computing time is achieved. That a saving is achieved in most instances can be seen by comparing Figs. 4.8 and 4.11 or by glancing at Fig. 4.18, where the e-M region with shorter computing times is shaded with asterisks. Following the style of the previous figures, a similar asterisk plot is shown in Fig. 4.19 for the high-e/low-M region.

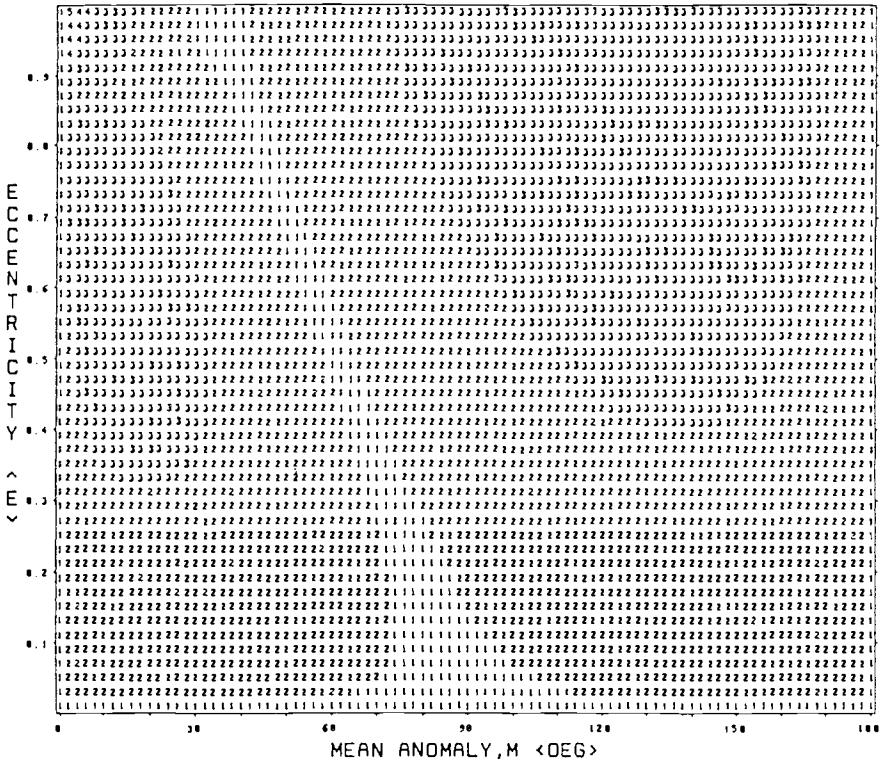


Fig. 4.16c Number of iterations needed to converge to within 0.00001 rad by Newton's method using the starting value of Eqs. (4.29) and (4.30).

We can carry this analysis one step further by plotting the computing time of Newton's method with starting values of Eqs. (4.29) and (4.30). This is shown as the dot-dash line in Fig. 4.17. Newton's method with $E_0 = M$ is also shown (as a solid line) but is not considered in this analysis because of the diverging behavior observed in Figs. 4.6 and 4.7.

In comparing Laguerre's (long-dash line) and Newton's (dot-dash line) methods, we see that 3 iterations by Newton's method require less computing time than 2 iterations by Laguerre's method. For 3 iterations by Laguerre's method, 4 or 5 iterations by Newton's method are computationally faster. For 4 iterations by Laguerre, 5, 6, or 7 iterations by Newton are faster, and so on. A comparison of Figs. 4.10 and 4.14 can identify the *e-M* region where Newton's method, with Eqs. (4.29) and (4.30) as starting values, is computationally faster than Laguerre's method, with similar starting values. Again, for ease of comparison, Figs. 4.20 and 4.21 are provided, where the shaded (asterisk) region identifies faster convergence by Newton's method over Laguerre's method.

After these figures are examined, it seems reasonable to adopt Newton's method, with starting values of Eqs. (4.29) and (4.30) for all *e-M* values except for the region of $e \geq 0.99$, and $M < 4$ deg, where Laguerre's method, with similar starting values, can be applied.

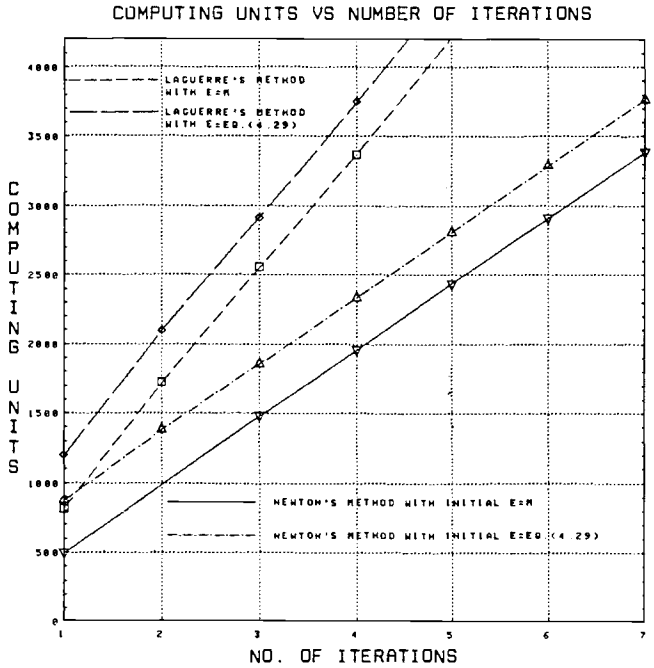


Fig. 4.17 Computing times to solve Kepler's equation.

Now that we know how to solve Kepler's equation analytically (albeit in an iterative fashion), is that all there is to it? Unfortunately, the answer is no. What we have just solved is the situation in which the orbit of concern is elliptical. If the orbit is hyperbolic, we must use Kepler's equation in the form of

$$M = e \sinh F - F \tag{4.32}$$

If the orbit is parabolic, there is yet another form of Kepler's equation, which we will not pursue at this moment.

In the hyperbolic case, it is not enough to change Kepler's equation; the expressions relating the eccentric and true anomalies must also be changed as follows. Equations (4.1–4.3) are changed to

$$\sinh F = \frac{\sqrt{e^2 - 1}\sin\theta}{1 + \cos\theta} \tag{4.33}$$

$$\cosh F = \frac{\cos\theta + e}{1 + \cos\theta} \tag{4.34}$$

$$F = \tanh^{-1}\left(\frac{\sinh F}{\cosh F}\right) \tag{4.35}$$

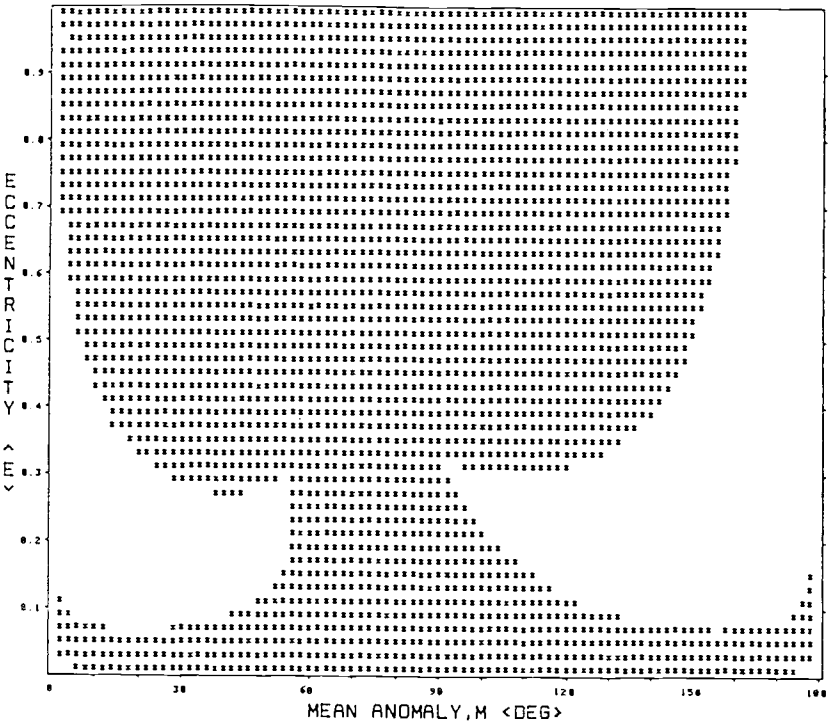


Fig. 4.18 Region (shaded) where Laguerre's method, with starting values of Eqs. (4.29) and (4.30), is computationally faster than the same method with $E_0 = M$.

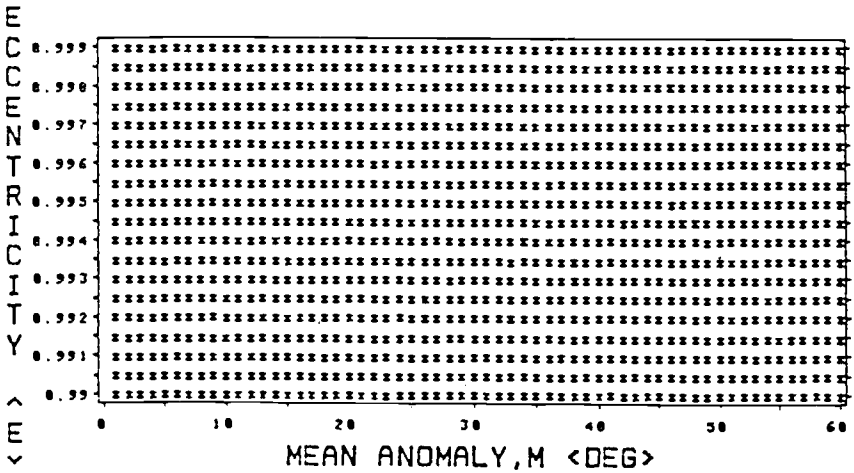


Fig. 4.19 High-e/low-M region (shaded) where Laguerre's method, with starting values of Eqs. (4.29) and (4.30), is computationally faster than the same method with $E_0 = M$.

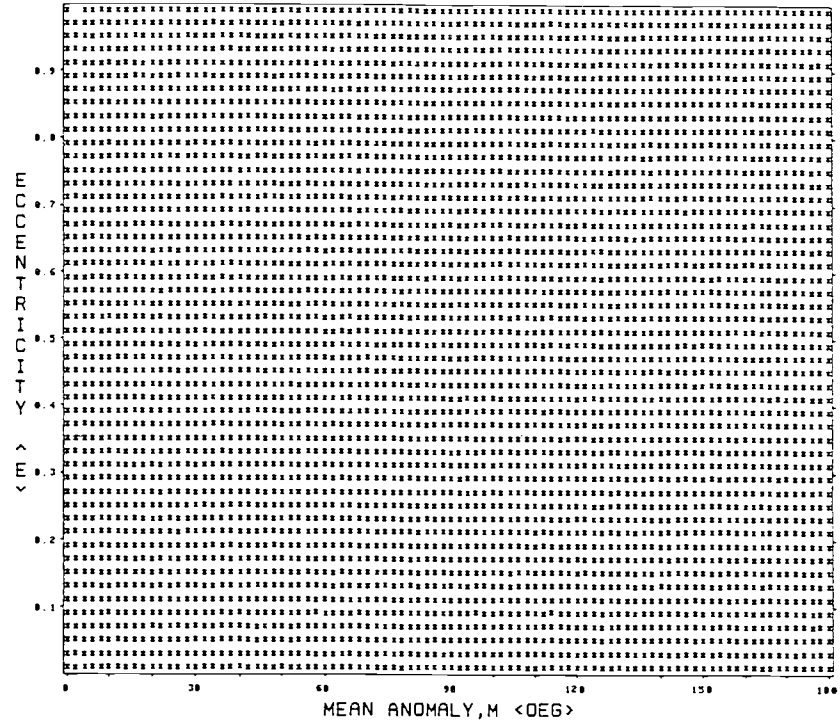


Fig. 4.20 Region (shaded) where Newton's method, with Eqs. (4.29) and (4.30), is computationally faster than Laguerre's method, with Eqs. (4.29) and (4.30).

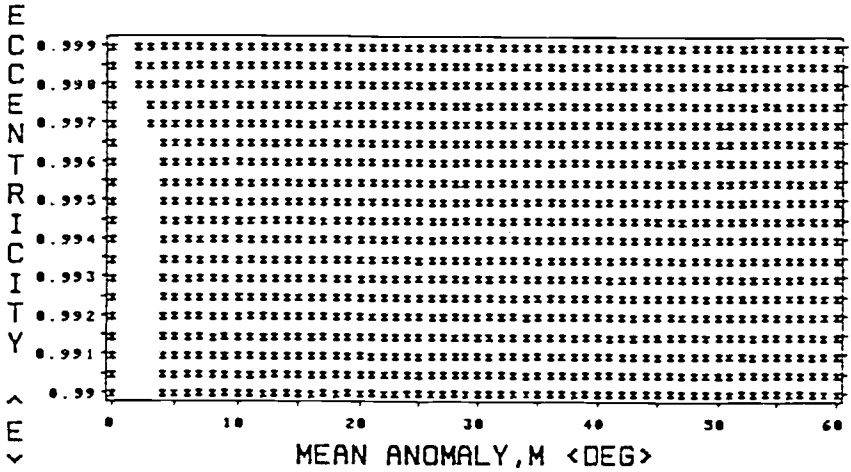


Fig. 4.21 High-*e*/low-*M* region (shaded) where Newton's method, with Eqs. (4.29) and (4.30), is computationally faster than the Laguerre's method with Eqs. (4.29) and (4.30).

and Eqs. (4.10–4.12) are changed to

$$\sin\theta = \frac{\sqrt{e^2 - 1}\sinh F}{e\cosh F - 1} \tag{4.36}$$

$$\cos\theta = \frac{e - \cosh F}{e\cosh F - 1} \tag{4.37}$$

$$\theta = \tan^{-1}\left(\frac{\sin\theta}{\cos\theta}\right) \tag{4.38}$$

Since we do not know ahead of time whether the orbit is elliptic or hyperbolic, we must alter the equations to fit the case at the time of calculation. For a computer program, both sets of equations must be programmed, and a test is made to branch to the appropriate set of equations whenever Kepler's solution is called for. And this is not the only problem. The method we have been following does not work very well for near-parabolic orbits. We have already seen these difficulties when we examined the region where e is nearly 1 and M very small. As we shall see in the next section, these problems are circumvented by the introduction of new auxiliary variables that are different from the eccentric anomaly.

4.3 A Universal Approach

A change of variable known as the "Sundman transformation" was first proposed in 1912. Recently, many authors have used this technique to develop formulas for computing the so-called time of flight via generalized or "universal" variables. In our discussion, we will use the formulation derived by Bate.[7] Since the referenced text goes into considerable details, we will, for practical reasons, confine our discussion to the change of variables introduced, the results of this introduction, and a summarized outline on how the results are used to solve our problem.

We start with the energy equation of

$$\text{Energy} = \frac{1}{2}V^2 - \frac{\mu}{r} = -\frac{\mu}{2a} \tag{4.39}$$

Resolving V into its radial and transverse components gives

$$\frac{1}{2}\dot{r}^2 + \frac{1}{2}(r\dot{\theta})^2 - \frac{\mu}{r} = -\frac{\mu}{2a} \tag{4.40}$$

Noting that

$$h = r^2\dot{\theta} = \sqrt{\mu p} \tag{4.41}$$

we have

$$\dot{r}^2 = -\frac{\mu p}{r^2} + \frac{2\mu}{r} - \frac{\mu}{a} \tag{4.42}$$

Bate now introduces a new independent variable X, which is defined as

$$\dot{X} = \frac{\sqrt{\mu}}{r} \tag{4.43}$$

Note that the X defined here and Z, which will be introduced shortly, are not to be confused with the x, y, and z normally used to describe the rectangular Cartesian coordinate system. Rather than introducing new symbols, we retain, for consistency, the symbols used in Bate's text.

Dividing the square of Eq. (4.43) into the preceding Eq. (4.42) and separating the variables yield

$$\mathrm{d}X = \frac{\mathrm{d}r}{\sqrt{-p + 2r - r^2/a}} \tag{4.44}$$

For $e \neq 1$, we can integrate Eq. (4.44) to obtain

$$X + C_0 = \sqrt{a}\sin^{-1}\left(\frac{r/a - 1}{e}\right) \tag{4.45}$$

where C_0 is the constant of integration.

Solving for r, and recalling that $p = a(1 - e^2)$,

$$r = a\left(1 + e\sin\frac{X + C_0}{\sqrt{a}}\right) \tag{4.46}$$

Substituting Eq. (4.46) into the definition of the universal variable, Eq. (4.43), and integrating give

$$\sqrt{\mu}t = aX - ae\sqrt{a}\left(\cos\frac{X + C_0}{\sqrt{a}} - \cos\frac{C_0}{\sqrt{a}}\right) \tag{4.47}$$

where we assume that $X = 0$ at $t = 0$.

Let us take stock in what we have so far. We have developed two equations for r and t in terms of X. In doing so, a constant of integration was introduced but not as yet evaluated.

We will now restate our problem, namely, Kepler's problem, and apply what we have developed to this problem.

In Kepler's problem, we have at some initial time, say t_0, the state of the orbit in terms of position and velocity, i.e., $\boldsymbol{r}_0$ and $\boldsymbol{V}_0$. Given this, we wish to find the state, i.e., position and velocity, $\boldsymbol{r}$ and $\boldsymbol{V}$, at some other time t.

First, we have assumed that $X = 0$ at $t = 0$. Thus, from Eq. (4.46),

$$e\sin\frac{C_0}{\sqrt{a}} = \frac{r_0}{a} - 1 \tag{4.48}$$

Next, differentiate Eq. (4.46) with respect to time to get

$$\dot{r} = \frac{e}{r}\sqrt{\mu a}\cos\left(\frac{X + C_0}{\sqrt{a}}\right) \tag{4.49}$$

At the initial condition, we have $X = 0$ and, since $r\dot{r} = \boldsymbol{r}_0 \cdot \boldsymbol{V}_0$, Eq. (4.49) becomes

$$e \cos \frac{C_0}{\sqrt{a}} = \frac{\boldsymbol{r}_0 \cdot \boldsymbol{V}_0}{\sqrt{\mu a}} \tag{4.50}$$

Substituting Eqs. (4.48) and (4.50) into Eqs. (4.47) and (4.46) yields

$$\begin{aligned} \sqrt{\mu} t &= a\left(X - \sqrt{a} \sin \frac{X}{\sqrt{a}}\right) \\ &+ \frac{\boldsymbol{r}_0 \cdot \boldsymbol{V}_0}{\sqrt{\mu}} a\left(1 - \cos \frac{X}{\sqrt{a}}\right) + r_0 \sqrt{a} \sin \frac{X}{\sqrt{a}} \end{aligned} \tag{4.51}$$

$$r = a + a\left[\frac{\boldsymbol{r}_0 \cdot \boldsymbol{V}_0}{\sqrt{\mu a}} \sin \frac{X}{\sqrt{a}} - \left(1 - \frac{r_0}{a}\right) \cos \frac{X}{\sqrt{a}}\right] \tag{4.52}$$

And the constant of integration C_0 is eliminated.

Now, Bate[7] introduces another variable, which is defined as

$$Z = \frac{X^2}{a} \tag{4.53}$$

or

$$a = \frac{X^2}{Z} \tag{4.54}$$

Removing a by this new variable from Eqs. (4.51) and (4.52) yields

$$\begin{aligned} \sqrt{\mu} t &= \frac{\sqrt{Z} - \sin \sqrt{Z}}{Z^{3/2}} X^3 \\ &+ \frac{\boldsymbol{r}_0 \cdot \boldsymbol{V}_0}{\sqrt{\mu}} \frac{1 - \cos \sqrt{Z}}{Z} X^2 + r_0 \frac{\sin \sqrt{Z}}{\sqrt{Z}} X \end{aligned} \tag{4.55}$$

$$r = \frac{1 - \cos \sqrt{Z}}{Z} X^2 + \frac{\boldsymbol{r}_0 \cdot \boldsymbol{V}_0}{\sqrt{\mu}} \frac{\sin \sqrt{Z}}{\sqrt{Z}} X + r_0 \cos \sqrt{Z} \tag{4.56}$$

Both these equations are indeterminate when $Z = 0$. To remove this dilemma, two very useful functions are introduced.

$$\begin{aligned} C(Z) &= \frac{1 - \cos \sqrt{Z}}{Z} \\ &= \frac{1}{2!} - \frac{Z}{4!} + \frac{Z^2}{6!} - \frac{Z^3}{8!} + \cdots \\ &= \sum_{k=0}^{\infty} \frac{(-Z)^k}{(2k+2)!} \end{aligned} \tag{4.57}$$

$$S(Z) = \frac{\sqrt{Z} - \sin\sqrt{Z}}{Z^{3/2}}$$
$$= \frac{1}{3!} - \frac{Z}{5!} + \frac{Z^2}{7!} - \frac{Z^3}{9!} + \cdots$$
$$= \sum_{k=0}^{\infty} \frac{(-Z)^k}{(2k+3)!} \tag{4.58}$$

Also, by differentiating Eqs. (4.57), (4.58), and (4.53), we note that

$$\frac{dC}{dZ} = \frac{1}{2Z}(1 - SZ - 2C) \tag{4.59}$$

$$\frac{dS}{dZ} = \frac{1}{2Z}(C - 3S) \tag{4.60}$$

$$\frac{dZ}{dX} = \frac{2X}{a} \tag{4.61}$$

Using the two functions of Eqs. (4.57) and (4.58), we can write Eqs. (4.55) and (4.56) as

$$\sqrt{\mu}t = SX^3 + C\frac{\boldsymbol{r}_0 \cdot \boldsymbol{V}_0}{\sqrt{\mu}}X^2 + r_0(1 - SZ)X \tag{4.62}$$

$$r = \frac{\boldsymbol{r}_0 \cdot \boldsymbol{V}_0}{\sqrt{\mu}}(1 - SZ)X + r_0(1 - CZ) + CX^2 \tag{4.63}$$

Now, the solution to Kepler's problem is sought by first finding X corresponding to time t and then by finding $\boldsymbol{r}$ and $\boldsymbol{V}$ corresponding to X.

In Eqs. (4.62) and (4.63), Z can be removed by $Z = X^2/a$ since a is known from $\boldsymbol{r}_0$ and $\boldsymbol{V}_0$ and the vis-viva equation. But still X cannot be solved for directly because the equations are transcendental in C and S. This suggests the use of a Newton–Raphson iteration method or, better yet,the Laguerre's method that was discussed in solving the original Kepler's equation.

In order to apply Laguerre's method, we must first rewrite Eq. (4.62) as follows:

$$f(X) = \left(1 - \frac{r_0}{a}\right)SX^3 + \frac{\boldsymbol{r}_0 \cdot \boldsymbol{V}_0}{\sqrt{\mu}}CX^2 + r_0X - \sqrt{\mu}t \tag{4.64}$$

Then, by differentiating and making use of Eqs. (4.59), (4.60), and (4.61),

$$f'(X) = CX^2 + \frac{\boldsymbol{r}_0 \cdot \boldsymbol{V}_0}{\sqrt{\mu}}(1 - SZ)X + r_0(1 - CZ) \tag{4.65}$$

$$f''(X) = \left(1 - \frac{r_0}{a}\right)(1 - SZ)X + \frac{\boldsymbol{r}_0 \cdot \boldsymbol{V}_0}{\sqrt{\mu}}(1 - CZ) \tag{4.66}$$

Now we may apply these three equations to

$$X_{i+1} = X_i - \frac{5f(X_i)}{f'(X_i) \pm 2\sqrt{|4[f'(X_i)]^2 - 5f(X_i)f''(X_i)|}} \tag{4.67}$$

where the sign in the denominator is chosen so that $|X_{i+1} - X_i|$ is as small as possible for each iteration.

Note that the t in Eq. (4.64) is, in reality, $\Delta t = t - t_0$. Our choice of $t_0 = 0$ was arbitrary in formulating a solution using the universal variables. Since, in general, $t_0 \neq 0$, a correct Δt must be calculated for the rightmost term of Eq. (4.64). Finally, to start the iterative process, the following initial X value is suggested:

$$X_0 = \sqrt{\mu}\,\frac{t - t_0}{|a|} \tag{4.68}$$

If the orbit is elliptic, the term $t - t_0$ in the numerator should be reduced by the largest integer multiple of P, where $P = 2\pi a^{3/2}/\sqrt{\mu}$. Other initial approximations may speed the convergence of the solution but seem unnecessary with Laguerre's method of iteration.

4.4 Expressions with *f* and *g*

Once X corresponding to t is determined, we must now find $\boldsymbol{r}$ and $\boldsymbol{V}$ in terms of X, r_0 and V_0.

Since Keplerian motion is confined to a plane, the four vectors $\boldsymbol{r}$, $\boldsymbol{V}$, $\boldsymbol{r}_0$, and $\boldsymbol{V}_0$ are all coplanar. Thus, we can write

$$\boldsymbol{r} = f\boldsymbol{r}_0 + g\boldsymbol{V}_0 \tag{4.69}$$

and differentiating

$$\boldsymbol{V} = \dot{f}\boldsymbol{r}_0 + \dot{g}\boldsymbol{V}_0 \tag{4.70}$$

where f, g, $\dot{f}$ and $\dot{g}$ are time-dependent scalars.

One interesting property of the f and g terms is seen by crossing $\boldsymbol{r}$ and $\boldsymbol{V}$ as follows:

$$\boldsymbol{r} \times \boldsymbol{V} = \boldsymbol{h} = (f\dot{g} - \dot{f}g)\boldsymbol{h} \tag{4.71}$$

from which

$$1 = f\dot{g} - \dot{f}g \tag{4.72}$$

This relationship states that f, g $\dot{f}$, and $\dot{g}$ are not independent and, if we know any three, we can determine the fourth from this identity.

Now, the approach taken to develop these f and g expressions is 1) to write Eqs. (4.69) and (4.70) in terms of a perifocal coordinate system, 2) relate Eq. (4.46) to the conic equation and find expressions of the perifocal components in terms of X, 3) substitute the results of step 2 into the results of step 1, and 4) introduce the

definitions of Z, $S(Z)$, and $C(Z)$. We shall omit the details of these steps, but the results of such an endeavor are

$$f = 1 - \frac{X^2}{r_0} C \tag{4.73}$$

$$g = t - \frac{X^3}{\sqrt{\mu}} \tag{4.74}$$

$$\dot{f} = \frac{\sqrt{\mu}}{r r_0} X(ZS - 1) \tag{4.75}$$

$$\dot{g} = 1 - \frac{X^2 C}{r} \tag{4.76}$$

4.5 Summary of the Universal Approach

As a summary of the universal approach discussed so far, we will outline the entire procedure and pertinent equations as follows:

Given: $\boldsymbol{r}_0$ and $\boldsymbol{V}_0$ at t_0 and t

$$r_0 = |\boldsymbol{r}_0| \tag{4.77}$$

$$V_0 = |\boldsymbol{V}_0| \tag{4.78}$$

$$\frac{1}{a} = \frac{2\mu/r_0 - V_0^2}{\mu} \tag{4.79}$$

$$\Delta t = t - t_0 \tag{4.80}$$

If $\frac{1}{a} \leq 0$, skip to Eq. (4.83). Otherwise,

$$P = 2\pi \frac{a^{3/2}}{\sqrt{\mu}} \tag{4.81}$$

$$\Delta t = \Delta t - [\text{sign}(\Delta t)]\text{int}\left[\frac{|\Delta t|}{P}\right] p \tag{4.82}$$

where int $|\xi|$ denotes the integer part of ξ.

$$X_0 = \sqrt{\mu} \frac{\Delta t}{|a|} \tag{4.83}$$

Start with $n = 0$, and calculate

$$Z_n = \frac{X_n^2}{a} \tag{4.84}$$

$$C_n = \frac{1}{2!} - \frac{Z_n}{4!} + \frac{Z_n^2}{6!} - \frac{Z_n^3}{8!} - \cdots \tag{4.85}$$

$$S_n = \frac{1}{3!} - \frac{Z_n}{5!} + \frac{Z_n^2}{7!} - \frac{Z_n^3}{9!} + \cdots \tag{4.86}$$

$$f(X_n) = \left(1 - \frac{r_0}{a}\right) S_n X_n^3 + \frac{\boldsymbol{r}_0 \cdot \boldsymbol{V}_0}{\sqrt{\mu}} C_n X_n^2 + r_0 X_n - \sqrt{\mu} t \tag{4.87}$$

$$f'(X_n) = C_n X_n^2 + \frac{\boldsymbol{r}_0 \cdot \boldsymbol{V}_0}{\sqrt{\mu}}(1 - S_n Z_n) X_n + r_0(1 - C_n Z_n) \tag{4.88}$$

$$f''(X_n) = \left(1 - \frac{r_0}{a}\right)(1 - S_n Z_n) X_n + \frac{\boldsymbol{r}_0 \cdot \boldsymbol{V}_0}{\sqrt{\mu}}(1 - C_n Z_n) \tag{4.89}$$

$$\delta_n = 2\sqrt{4[f'(X_n)]^2 - 5 f(X_n) f''(X_n)} \tag{4.90}$$

$$\Delta X_n = \frac{5 f(X_n)}{f'(X_n) + [\text{sign } f'(X_n)]\delta_n} \tag{4.91}$$

$$X_{n+1} = X_n - \Delta X_n \tag{4.92}$$

Repeat Eqs. (4.84–4.92) for $n = 1, 2, 3 \ldots$ until

$$\left| \frac{(\Delta X_n)^2}{a} \right| < \varepsilon \tag{4.93}$$

where $\varepsilon = 10^{-8}$.

$$f = 1 - \frac{X^2}{r_0} C \tag{4.94}$$

$$g = t - \frac{X^3}{\sqrt{\mu}} S \tag{4.95}$$

$$\dot{f} = \frac{\sqrt{\mu}}{r r_0}(SZ - 1) X \tag{4.96}$$

$$\dot{g} = 1 - \frac{X^2}{r} C \tag{4.97}$$

$$\boldsymbol{r} = f \boldsymbol{r}_0 + g \boldsymbol{V}_0 \tag{4.98}$$

$$\boldsymbol{V} = \dot{f} \boldsymbol{r}_0 + \dot{g} \boldsymbol{V}_0 \tag{4.99}$$

4.6 The Classical Element Set

Let us pause for the moment and examine the coordinate frames we have used so far. When we first described Kepler's problem, we started with an element set that consisted of the following: a, e, i, Ω, ω, and θ. This is a modified form of the classical element set in which the sixth element is θ (true anomaly). Sometimes, M (mean anomaly) is used instead of θ. Rarely is E (eccentric anomaly) used. The true classical element set consists of a, e, i, Ω, ω, and τ, where the sixth term, τ, is the time of perifocal passage or, more exactly, the time of the last previous

perifocal passage as measured from a specified reference time (usually midnight of Greenwich mean time of an epoch date). It is because of the difficulty encountered in handling this sixth term that practical considerations have led to the use of θ and M in its place.

4.7 The Rectangular Coordinate System

When we discussed the universal variables, the initial and final orbit states were expressed by two quantities, position and velocity, in their vector form. Position and velocity as vectors can be resolved into a variety of components. Most common are those expressed in a rectangular Cartesian coordinate frame, in which case the components are x, y, z, $\dot{x}$, $\dot{y}$, and $\dot{z}$. Often, this coordinate frame is referred to as the Earth-centered inertial (ECI) frame.

Since it is apparent that we need to change from one coordinate frame to another, this is a good time to see how we transform from the modified classical element set to the rectangular Cartesian set and vice versa.

4.8 Modified Classical to Cartesian Transformation

This transformation is accomplished in two steps. First, the classical set is expressed in a perifocal coordinate frame. Second, the perifocal coordinate frame is converted to a Cartesian frame through a series of rotations of the axes.

Looking at Fig. 4.22, we can immediately write

$$\boldsymbol{r} = r\cos\theta\hat{\boldsymbol{P}} + r\sin\theta\hat{\boldsymbol{Q}} \tag{4.100}$$

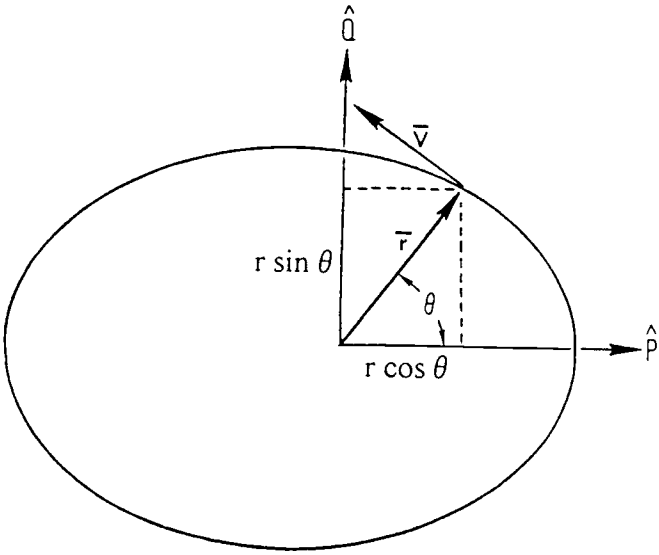


Fig. 4.22 Position and velocity in a perifocal coordinate system.

where the magnitude of r is determined from the equation of the conic

$$r = \frac{p}{1 + e\cos\theta} \tag{4.100a}$$

p being equal to the semilatus rectum of the orbit, which is $a(1 - e^2)$. Differentiating and simplifying the preceding two equations yields

$$\boldsymbol{V} = \sqrt{\frac{\mu}{p}}[(-\sin\theta)\hat{\boldsymbol{P}} + (e + \cos\theta)\hat{\boldsymbol{Q}}] \tag{4.101}$$

In differentiating, note that $\dot{\hat{\boldsymbol{P}}} = \dot{\hat{\boldsymbol{Q}}} = 0$ since the perifocal coordinate frame is "inertial" in space. Also, $\dot{r} = \sqrt{(\mu/p)}e\sin\theta$ and $r\dot{\theta} = \sqrt{(\mu/p)}(1 + e\cos\theta)$ from Eq. (4.41) and Eq. (4.100a) and its derivative.

Now, from Fig. 4.23, we see that the IJK axes can become the PQW axes by three successive rotations as follows: 1) rotation about the $\hat{z}$ axis by $+\Omega$, 2) rotation about the $\hat{x}$ axis by $+i$, and 3) rotation about the $\hat{z}$ axis by $+\omega$. The first transformation is accomplished by

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} \cos\Omega & \sin\Omega & 0 \\ -\sin\Omega & \cos\Omega & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}_{IJK} \tag{4.102}$$

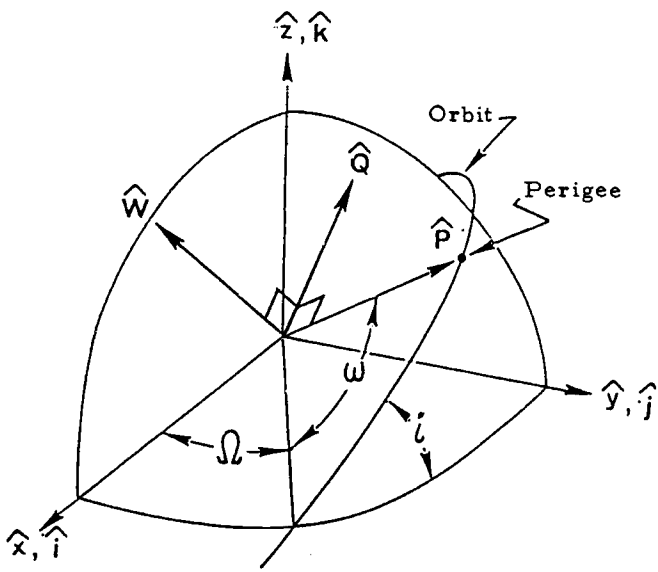


Fig. 4.23 The relationship between the *IJK* and *PQW* systems.

The second transformation is accomplished by

$$\begin{bmatrix} x'' \\ y'' \\ z'' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos i & \sin i \\ 0 & -\sin i & \cos i \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} \tag{4.103}$$

and the third transformation by

$$\begin{bmatrix} x''' \\ y''' \\ z''' \end{bmatrix}_{PQW} = \begin{bmatrix} \cos\omega & \sin\omega & 0 \\ -\sin\omega & \cos\omega & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x'' \\ y'' \\ z'' \end{bmatrix} \tag{4.104}$$

Actually, we need to change the PQW axes to the IJK axes. That is, instead of Eq. (4.102), we need

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{IJK} = [T_1] \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} \tag{4.105}$$

where $[T_1]$ is the inverse of the transformation matrix given in Eq. (4.102). Similarly, we need

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = [T_2] \begin{bmatrix} x'' \\ y'' \\ z'' \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} x'' \\ y'' \\ z'' \end{bmatrix} = [T_3] \begin{bmatrix} x''' \\ y''' \\ z''' \end{bmatrix}_{PQW} \tag{4.106}$$

where $[T_2]$ and $[T_3]$ are the inverses of the transformation matrices given in Eqs. (4.103) and (4.104). The inverses are obtained simply by changing the sign to the sine terms of the matrices. Thus, we have

$$[T_1] = \begin{bmatrix} \cos\Omega & -\sin\Omega & 0 \\ \sin\Omega & \cos\Omega & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4.107}$$

$$[T_2] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos i & -\sin i \\ 0 & \sin i & \cos i \end{bmatrix} \tag{4.108}$$

and

$$[T_3] = \begin{bmatrix} \cos\omega & -\sin\omega & 0 \\ \sin\omega & \cos\omega & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4.109}$$

The complete transformation is then accomplished by three successive transformations

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{IJK} = [T_1][T_2][T_3] \begin{bmatrix} x \\ y \\ z \end{bmatrix}_{PQW} \tag{4.110}$$

Combining the three transformation matrices into one then gives

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = [R] \begin{bmatrix} r\cos\theta \\ r\sin\theta \\ 0 \end{bmatrix} \tag{4.111}$$

and

$$\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{bmatrix} = [R] \begin{bmatrix} -\sqrt{\frac{\mu}{p}}\sin\theta \\ \sqrt{\frac{\mu}{p}}(e+\cos\theta) \\ 0 \end{bmatrix} \tag{4.112}$$

where

$$[R] = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix} \tag{4.113}$$

and

$$\begin{aligned}
R_{11} &= \cos\Omega\cos\omega - \sin\Omega\sin\omega\cos i \\
R_{12} &= -\cos\Omega\sin\omega - \sin\Omega\cos\omega\cos i \\
R_{13} &= \sin\Omega\sin i \\
R_{21} &= \sin\Omega\cos\omega + \cos\Omega\sin\omega\cos i \\
R_{22} &= -\sin\Omega\sin\omega + \cos\Omega\cos\omega\cos i \\
R_{23} &= -\cos\Omega\sin i \\
R_{31} &= \sin\omega\sin i \\
R_{32} &= \cos\omega\sin i \\
R_{33} &= \cos i
\end{aligned} \tag{4.114}$$

4.9 Rectangular to Modified Classical Elements Transformation

In this transformation, we start with $\boldsymbol{r}(x, y, z)$ and $\boldsymbol{V}(\dot{x}, \dot{y}, \dot{z})$ and seek to change these into the set of $(a, e, i, \Omega, \omega$ and $\theta)$. The computational steps for this transformation are as follows:

$$r = \sqrt{x^2 + y^2 + z^2} \tag{4.115}$$

$$V = \sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2} \tag{4.116}$$

$$\frac{V^2}{\mu} = \frac{2}{r} - \frac{1}{a} \rightarrow a \tag{4.117}$$

$$\hat{\boldsymbol{W}} = \frac{\boldsymbol{r} \times \boldsymbol{V}}{|\boldsymbol{r} \times \boldsymbol{V}|} \tag{4.118}$$

$$\cos i = \hat{\boldsymbol{W}} \cdot \hat{\boldsymbol{k}} \rightarrow i \tag{4.119}$$

$$\boldsymbol{e} = \frac{1}{\mu}\left[\left(V^2 - \frac{\mu}{r}\right)\boldsymbol{r} - (\boldsymbol{r} \cdot \boldsymbol{V})\boldsymbol{V}\right] \rightarrow e \tag{4.120}$$

$$\hat{\boldsymbol{N}} = \frac{\hat{\boldsymbol{k}} \times \hat{\boldsymbol{W}}}{|\hat{\boldsymbol{k}} \times \hat{\boldsymbol{W}}|} \tag{4.121}$$

$$\cos \Omega = \hat{\boldsymbol{i}} \cdot \hat{\boldsymbol{N}} \tag{4.122}$$

$$\sin \Omega = (\hat{\boldsymbol{i}} \times \hat{\boldsymbol{N}}) \cdot \hat{\boldsymbol{k}} \tag{4.123}$$

$$\Omega = \tan^{-1}\left(\frac{\sin \Omega}{\cos \Omega}\right) \tag{4.124}$$

$$\cos \omega = \frac{\hat{\boldsymbol{N}} \cdot \boldsymbol{e}}{|\boldsymbol{e}|} \tag{4.125}$$

$$\sin \omega = \frac{\hat{\boldsymbol{N}} \times \boldsymbol{e}}{|\boldsymbol{e}|} \cdot \hat{\boldsymbol{W}} \tag{4.126}$$

$$\omega = \tan^{-1}\left(\frac{\sin \omega}{\cos \omega}\right) \tag{4.127}$$

$$\cos \theta = \frac{\boldsymbol{e} \cdot \boldsymbol{r}}{|\boldsymbol{e}||\boldsymbol{r}|} \tag{4.128}$$

$$\sin \theta = \frac{\boldsymbol{e} \times \boldsymbol{r}}{|\boldsymbol{e}||\boldsymbol{r}|} \cdot \hat{\boldsymbol{W}} \tag{4.129}$$

$$\theta = \tan^{-1}\left(\frac{\sin \theta}{\cos \theta}\right) \tag{4.130}$$

4.10 The Spherical (ADBARV) Coordinate System

Another often used and practical coordinate system is the spherical coordinate system. In this system, the position and velocity are expressed in terms of the following six quantities:

α = right ascension
δ = declination
β = flight-path angle
A = azimuth
r = radius
V = velocity

All of these are scalar quantities and, because of the symbols used, the coordinate system is often referred to as the ADBARV system. The first two quantities reflect the fact that the spherical system is similar to the celestial (right ascension-declination) system used in astronomy.

Figure 4.24 is a pictorial representation of the ADBARV components in the rectangular x, y, z coordinate system.

In the Fig. 4.24, α, δ, and the magnitudes of $\boldsymbol{r}$ and $\boldsymbol{V}$ are self-evident.

The flight-path angle β is the angle between the radius and velocity vectors. At either perigee or apogee, the velocity vector is perpendicular to the radius vector, and $\beta = 90$ deg. On half of the orbit from perigee to apogee, β is less than 90 deg. On the other half, from apogee to perigee, β is greater than 90 deg. Depending

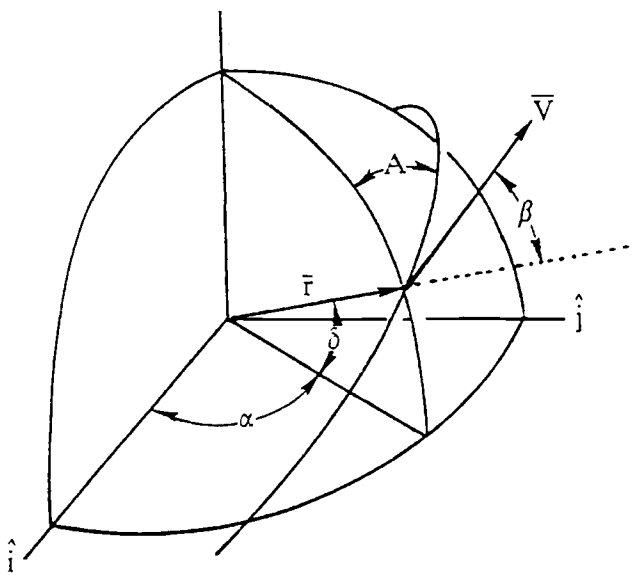


Fig. 4.24 The spherical coordinate system.

on the discipline to which one belongs (control, guidance, etc.), sometimes the flight-path angle is measured relative to the instantaneous geocentric horizontal rather than the vertical. In this case, $\gamma = 90\,\text{deg} - \beta$, and the flight-path angle γ is zero at perigee and apogee. It is positive when approaching apogee from perigee and negative when approaching perigee from apogee. You may also hear the term "pitch angle" used to describe the flight-path angle γ. Here, nose up and nose down are equivalent to the positive and negative sense of the angle, γ.

The azimuth A is measured in the instantaneous geocentric horizontal plane at the point in question. It is the angle between the northerly direction and the projection of the velocity vector onto this plane. Typically, it is positive when measured clockwise from due north when viewed down along the radius vector toward the center of the Earth. Care should be taken in noting the positive/negative sense of this component since some disciplines define the counterclockwise direction as being the positive direction of this angle.

In referring to the instantaneous horizontal, the term "geocentric" has been introduced. This means that the horizontal line or plane is perpendicular to the geocentric radius at that instant. There is also a "geodetic" horizontal, which should not be confused with the geocentric. More on this topic is discussed in Sec. 4.14.

4.11 Rectangular to Spherical Transformation

Now, let us see how we can transform to the spherical (ADBARV) coordinate frame from the rectangular Cartesian coordinate system. The computational steps are

$$r = \sqrt{x^2 + y^2 + z^2} \tag{4.131}$$

$$V = \sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2} \tag{4.132}$$

$$\sin\alpha = \frac{y}{\sqrt{x^2 + y^2}} \tag{4.133}$$

$$\cos\alpha = \frac{x}{\sqrt{x^2 + y^2}} \tag{4.134}$$

$$\alpha = \tan^{-1}\left(\frac{\sin\alpha}{\cos\alpha}\right) \tag{4.135}$$

$$\delta = \sin^{-1}\left(\frac{z}{r}\right) \tag{4.136}$$

$$\beta = \cos^{-1}\left(\frac{\boldsymbol{r}\cdot\boldsymbol{V}}{|\boldsymbol{r}\cdot\boldsymbol{V}|}\right) \tag{4.137}$$

$$\hat{\boldsymbol{W}} = \frac{\boldsymbol{r}\times\boldsymbol{V}}{|\boldsymbol{r}\times\boldsymbol{V}|} \tag{4.138}$$

$$\hat{\boldsymbol{A}} = \frac{\hat{\boldsymbol{W}}\times\boldsymbol{r}}{|\hat{\boldsymbol{W}}\times\boldsymbol{r}|} \tag{4.139}$$

$$\hat{P} = \frac{(r \times \hat{k}) \times r}{|(r \times \hat{k}) \times r|} \tag{4.140}$$

$$\cos A = \hat{A} \cdot \hat{P} \tag{4.141}$$

$$\sin A = \hat{A} \times \hat{P} \cdot \frac{r}{|r|} \tag{4.142}$$

$$A = \tan^{-1}\left(\frac{\sin A}{\cos A}\right) \tag{4.143}$$

4.12 Spherical to Rectangular Transformation

The inverse transformation from spherical to rectangular is very similar to the procedure used in the transformation from the modified classical element set to the rectangular set. Using Fig. 4.24 as a guide, we see that a rotation of the axes first about the $\hat{z}$ axis by $+\alpha$ and then about the $\hat{y}$ axis by $-\delta$ will cause the $\hat{x}$ axis to become colinear with the radius vector. By inspection, then, we can write the inverse transformation of the position vector as follows:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\delta & 0 & -\sin\delta \\ 0 & 1 & 0 \\ \sin\delta & 0 & \cos\delta \end{bmatrix} \begin{bmatrix} r \\ 0 \\ 0 \end{bmatrix} \tag{4.144}$$

By completing the matrix multiplications, we have

$$x = r\cos\delta\cos\alpha \tag{4.145}$$

$$y = r\cos\delta\sin\alpha \tag{4.146}$$

$$z = r\sin\delta \tag{4.147}$$

The inverse transformation of the velocity vector is slightly more complicated but only in that we must go through four rotations instead of two. These rotations are: 1) about the $\hat{z}$ axis by $+\alpha$, 2) about the $\hat{y}$ axis by $-\delta$, 3) about the $\hat{x}$ axis by $-A$, and 4) about the $\hat{y}$ axis by $-\beta$. When these four rotations are completed, we find the $\hat{x}$ axis aligned along the direction of the velocity vector. Again, by inspection, we can write

$$\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{bmatrix} = T_1 T_2 T_3 T_4 \begin{bmatrix} V \\ 0 \\ 0 \end{bmatrix} \tag{4.148}$$

where

$$T_1 = \begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4.149}$$

$$T_2 = \begin{bmatrix} \cos\delta & 0 & -\sin\delta \\ 0 & 1 & 0 \\ \sin\delta & 0 & \cos\delta \end{bmatrix} \tag{4.150}$$

$$T_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos A & \sin A \\ 0 & -\sin A & \cos A \end{bmatrix} \tag{4.151}$$

$$T_4 = \begin{bmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{bmatrix} \tag{4.152}$$

And by completing the matrix multiplications we have

$$\dot{x} = V[\cos\alpha(-\cos A\sin\beta\sin\delta + \cos\beta\cos\delta) - \sin A\sin\beta\sin\alpha] \tag{4.153}$$

$$\dot{y} = V[\sin\alpha(-\cos A\sin\beta\sin\delta + \cos\beta\cos\delta) + \sin A\sin\beta\cos\alpha] \tag{4.154}$$

$$\dot{z} = V(\cos A\cos\delta\sin\beta + \cos\beta\sin\delta) \tag{4.155}$$

4.13 The Earth-Relative Spherical (LDBARV) Coordinate System

Another very useful coordinate system, especially when the orbit must be referenced to various positions on the Earth, is a modified form of the spherical coordinate system. This Earth-relative system is identical to the spherical system except that the first term, right ascension, is replaced by the geographic longitude, usually denoted by the symbol λ. Because of the symbols used in describing this system, it is often referred to as the LDBARV system. In map coordinates, longitude spans 0 to 180 deg both east and west. In astrodynamics, we often consider the longitude to span 0 to 360 deg, with the positive direction coinciding with the easterly direction. Again, care must be exercised since some disciplines consider west as being the positive direction rather than east.

The relationship between longitude and right ascension is

$$\lambda = \alpha - \alpha_g \tag{4.156}$$

where α_g is the right ascension of Greenwich at time t. The right ascension of Greenwich is tabulated for different days in the *American Ephemeris and Nautical Almanac*. It may be more expedient, however, to calculate this value from a set of equations such as the following:

$$\begin{aligned} \alpha_{g@\text{midnight}} = {} & [100.46061838 + d(0.7700537 + 3.88 \times 10^{-4} \times d) \\ & + 360 \times \text{fractional part}\{100d\}]_{\text{mod } 360 \text{ deg}} \end{aligned} \tag{4.157}$$

where

$$d = \frac{\text{MJD} - 51{,}544.5}{36{,}525} \tag{4.158}$$

and

$$\begin{aligned}
\text{YR} &= \text{year (e.g., 1989)} \\
\text{MO} &= \text{month (1 for Jan., 2 for Feb., etc)} \\
\text{DY} &= \text{day (day of the month)} \\
Y &= \text{YR} - D
\end{aligned}$$

where

$$\begin{aligned}
D &= 0, \text{ if MO } \geq 3 \\
&= 1, \text{ if MO } < 3 \\
A &= \text{integer part } \{Y/100\} \\
B &= 2 + \text{integer part } \{A/4\} - A \\
M &= \text{MO} + 12 \times D \\
\text{MJD} &= B + DY + \text{integer part } \{30.6001(M+1)\} \\
&\quad + \text{integer part } \{365.25Y\} - 679{,}006
\end{aligned} \tag{4.159}$$

Note that MJD, which stands for "Modified Julian Date," is introduced to reduce the number of significant digits that must be carried in the computation without loss of precision or the need to invoke double precision operations. MJD is related to Julian date (JD) by the following:

$$\text{MJD} = \text{JD} - 2{,}400{,}000.5 \tag{4.160}$$

Note also that a Julian day is reckoned from noon to noon of the following day. A modified Julian day is from midnight to midnight of the following day.

The right ascension of Greenwich at any time during the day is then found from

$$\alpha_g = \alpha_{g@\text{midnight}} + \omega_e t \tag{4.161}$$

where

$$\begin{aligned}
t &= \text{time from midnight} \\
\omega_e &= \text{rotational rate of the Earth} \\
&= 0.2506844537 \text{ deg/min}
\end{aligned}$$

4.14 Geodetic and Geocentric Altitudes

In Sec. 4.10, the term "geodetic horizontal" was introduced. This term, which is intimately related to the vertical direction, is synonymous with the term "geodetic vertical." Geodetic vertical, in turn, raises the question of what constitutes an altitude of an object above an Earth that is not a sphere but an oblate spheroid.

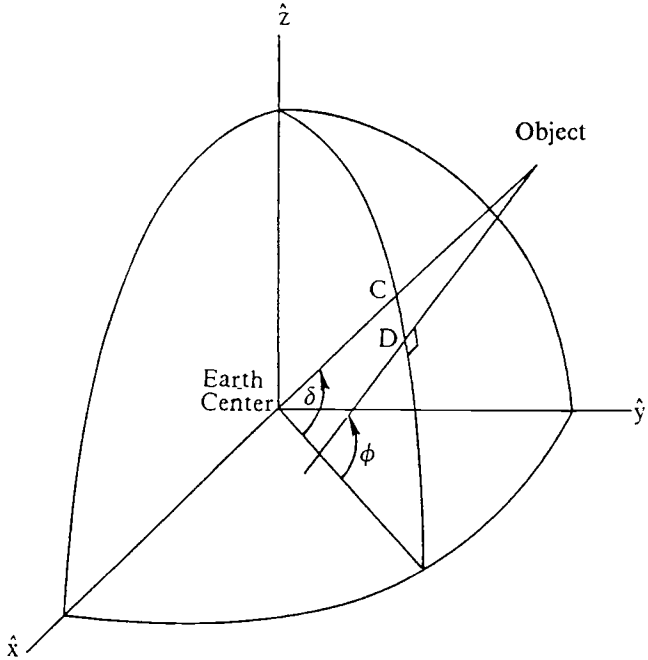


Fig. 4.25 A three-dimensional view of geocentric and geodetic altitudes.

Figure 4.25 gives a three-dimensional view of what we mean by altitude. The distance from C to the object is the geocentric altitude measured along the geocentric radius of the object. The distance from D to the object is the geodetic altitude measured along the vertical line that passes through the object and is also perpendicular to the Earth's surface. Because the Earth is not a sphere, C and D are not colocated. This is more clearly illustrated in the two-dimensional representation of Fig. 4.26, where the meridian containing the two altitudes is viewed from the side.

Ironically, because the coordinate system we use is Earth-centered, it is more direct and easier to calculate the geocentric altitude. Yet, when we say altitude, we really mean geodetic altitude. And this geodetic altitude, if calculated formally, is obtained not directly but by an iterative process. This is not an attractive proposition. When we generate an ephemeris of, say, 10,000 time points and ask the question of altitude at each of these points, this implies introducing an iterative scheme, such as the one we studied in Sec. 4.2 to solve Kepler's equation, at every one of these 10,000 points after we have determined their positions.

Fortunately, we can rely on an approximate but very effective closed-form procedure[8] to determine the geodetic sublatitude and altitude of an object in space. The errors introduced by using this approximate method are only 5 parts in 100 million for latitude and 8 parts in 100 million for altitude.

Referring to Fig. 4.26, we first calculate the radius and declination of the object as follows:

$$r = \sqrt{x^2 + y^2 + z^2} \tag{4.162}$$

$$\delta = \sin^{-1}(z/r) \tag{4.163}$$

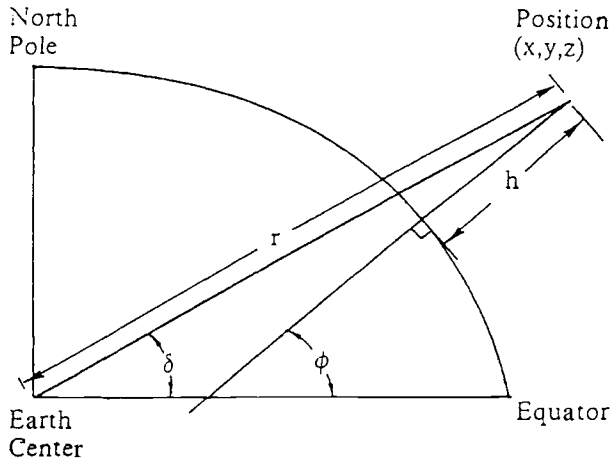


Fig. 4.26 A two-dimensional view of geocentric and geodetic altitudes.

Once these two quantities are determined, the following two equations are used directly.

$$h = r - a_e\left[1 - f\sin^2\delta - \frac{f^2}{2}\sin^2 2\delta\left(\frac{a_e}{r} - \frac{1}{4}\right)\right] \tag{4.164}$$

where

$$h = \text{geodetic altitude}$$

$$a_e = \text{equatorial radius of the Earth}$$

$$f = \text{flattening of the Earth}$$

and

$$\sin(\phi - \delta) = \frac{a_e}{r}\left[f\sin 2\delta + f^2\sin 4\delta\left(\frac{a_e}{r} - \frac{1}{4}\right)\right] \tag{4.165}$$

From Eq. (4.165), then,

$$\phi = \delta + \sin^{-1}[\sin(\phi - \delta)] \tag{4.166}$$

where

$$\phi = \text{geodetic latitude of the sublatitude point}$$

The longitude λ of the object is calculated by first determining the right ascension of the position (x, y, z) and then relating it to an Earth-relative frame whose x axis is in the plane of the Greenwich meridian. Figures 4.27 and 4.28 describe the geometry relating to calculating the longitude.

First, from Fig. 4.27,

$$\alpha = \tan^{-1}\left(\frac{y}{x}\right) \tag{4.167}$$

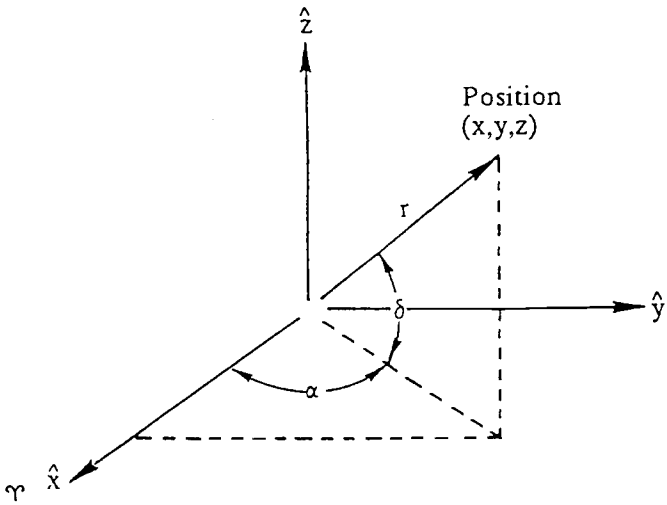


Fig. 4.27 The position in spherical coordinates.

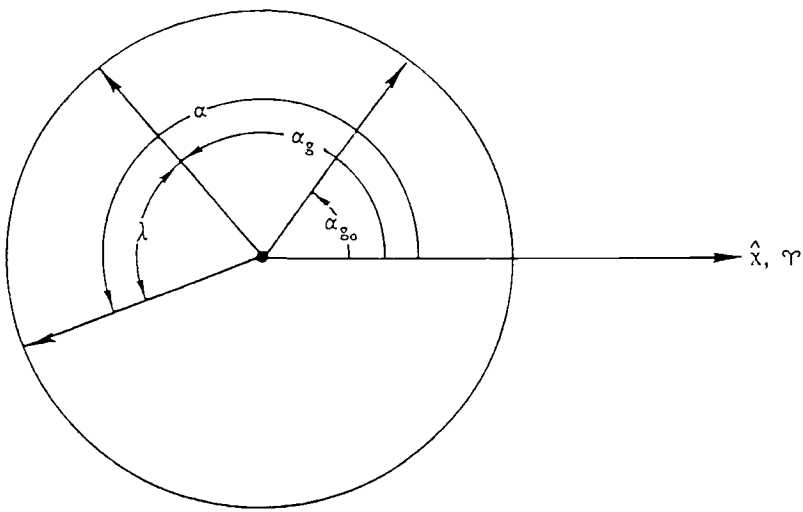


Fig. 4.28 Inertial and Earth-relative angles: α_{g0} = right ascension of Greenwich at epoch; α_g = right ascension of Greenwich at (t – epoch); α = right ascension of the vehicle at (t – epoch).

And, from Fig. 4.28,

$$\lambda = \alpha - \alpha_g \tag{4.168}$$

or

$$\lambda = \{\alpha - [\alpha_{g_0} + \omega_e(t - t_0)]\}_{\text{mod } 360 \text{ deg}}, \qquad 0 \le \lambda < 360 \text{ deg} \tag{4.169}$$

where

$$\omega_e = \text{the rotational rate of the Earth}$$

$$t_0 = \text{the epoch or the reference time}$$

$$\alpha_{g_0} = \text{the right ascension of Greenwich at } t_0$$

By letting t_0 occur at midnight, $\alpha_{g_0} = \alpha_{g@\text{midnight}}$, which can be calculated as shown in Sec. 4.13.

To complete this topic, we will also examine the inverse process, namely, to convert geodetic sublatitude, longitude, and geodetic altitude to the equivalent (x, y, z) position. Unlike earlier conversions, this process is direct and exact and does not use any approximate formulas.

First, as an intermediate step, the geocentric latitude and the geocentric radius of the sublatitude point are determined. Referring to Fig. 4.29, we obtain

$$\phi' = \tan^{-1}[(1 - f)^2 \tan\phi], \qquad -90 \text{ deg} \le \phi \le +90 \text{ deg} \tag{4.170}$$

$$r_E = \frac{a_e(1 - f)}{\sqrt{1 - f(2 - f)\cos^2\phi'}} \tag{4.171}$$

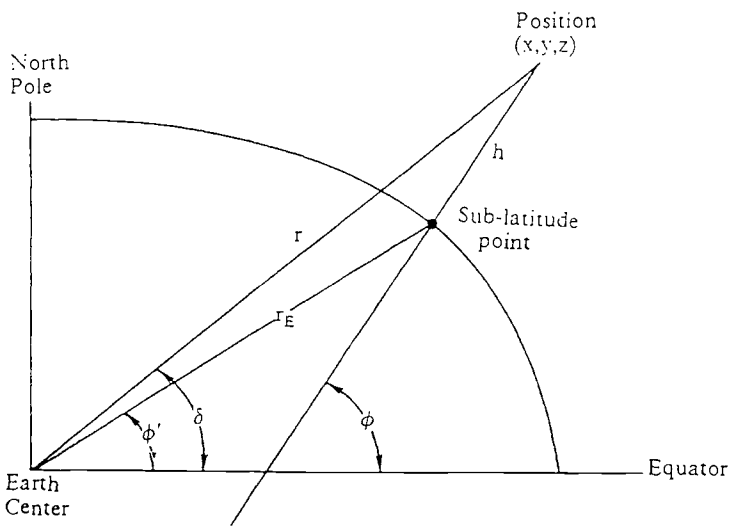


Fig. 4.29 Geometry for converting *h* to *r*.

where ϕ' and r_E are the geocentric latitude and radius of the sublatitude point, respectively.

Then

$$x' = r_E \cos\phi' + h\cos\phi \tag{4.172}$$

$$y' = 0 \tag{4.173}$$

$$z' = r_E \sin\phi' + h\sin\phi \tag{4.174}$$

and

$$r = \sqrt{x'^2 + z'^2} \tag{4.175}$$

$$\delta = \sin^{-1}\left(\frac{z'}{r}\right) \tag{4.176}$$

Now, from Eq. (4.169),

$$\alpha = [\lambda + \alpha_{g_0} + \omega_e(t - t_0)]_{\text{mod 360 deg}}, \qquad 0 \le \alpha \le 360 \text{ deg} \tag{4.177}$$

where, again, by letting t_0 to be at midnight, $\alpha_{g_0} = \alpha_{g@\text{midnight}}$, which can be calculated as shown in Sec. 4.13.

Finally,

$$x = x'\cos\alpha \tag{4.178}$$

$$y = x'\sin\alpha \tag{4.179}$$

$$z = z' \tag{4.180}$$

4.15 Converting from Perigee/Apogee Radii to Perigee/Apogee Altitudes

An interesting variation on what has been described so far is the process to convert a set of perigee/apogee radii to its equivalent set of perigee/apogee altitudes. To be more specific, we have as initial quantities the values for r_p (perigee radius), r_A (apogee radius), and δ_p (declination of the perigee). Inherent in this statement is the assumption that perigee and apogee lie on a straight line that passes through the center of the Earth (i.e., they are 180 deg apart in Earth-centered angle). It follows then that the declination of the apogee is equal in magnitude to the declination of perigee except for its sign, which is opposite. What may not be obvious is that the geodetic sublatitude at perigee is not the same in magnitude as the sublatitude at apogee. Figure 4.30 illustrates this point.

Since r_p and r_A lie on a straight line, $|\delta_p| = |\delta_A|$. But, because $r_p < r_A$, the subapsidal points do not occur at the same latitude (magnitudewise) as shown in the Fig. 4.30. Thus, $|\phi_p| \neq |\phi_A|$. Needless to say, this affects the values of h_p and h_A.

Mathematically, we solve this problem by first using r_p and δ_p in Eqs. (4.164–4.166)

$$h_p = r_p - a_e\left[1 - f\sin^2\delta_p - \frac{f^2}{2}\sin^2 2\delta_p\left(\frac{a_e}{r_p} - \frac{1}{4}\right)\right] \tag{4.181}$$

$$\phi_p = \delta_p + \sin^{-1}\left\{\frac{a_e}{r_p}\left[f\sin 2\delta_p + f^2\sin 4\delta_p\left(\frac{a_e}{r_p} - \frac{1}{4}\right)\right]\right\} \tag{4.182}$$

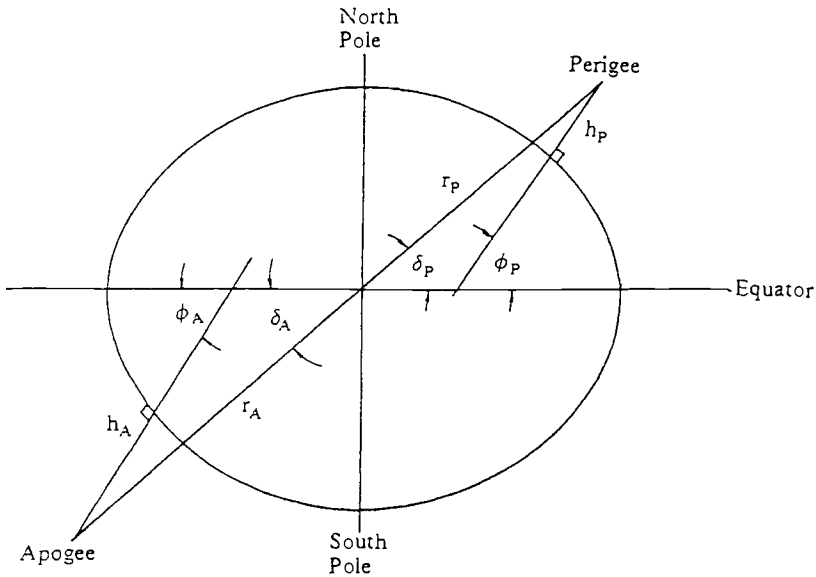


Fig. 4.30 Differences in sublatitudes between perigee and apogee.

This process is then repeated using r_A and δ_A, where $\delta_A = -\delta_p$

$$h_A = r_A - a_e\left[1 - f\sin^2\delta_A - \frac{f^2}{2}\sin^2 2\delta_A\left(\frac{a_e}{r_A} - \frac{1}{4}\right)\right] \tag{4.183}$$

$$\phi_A = \delta_A + \sin^{-1}\left\{\frac{a_e}{r_A}\left[f\sin 2\delta_A + f^2\sin 4\delta_A\left(\frac{a_e}{r_A} - \frac{1}{4}\right)\right]\right\} \tag{4.184}$$

4.16 Converting from Perigee/Apogee Altitudes to Perigee/Apogee Radii

The most interesting and perhaps most misused conversion occurs when perigee and apogee altitudes are converted to perigee and apogee radii. This occurs typically when perigee and apogee altitudes are converted to their ECI (Earth-centered inertial) counterparts. Symbolically, this conversion is represented by

$$\begin{array}{ccc} \text{Orbital elements} & & \text{ECI elements} \\ \begin{bmatrix} h_p \\ h_A \\ i \\ \phi_{PD} \\ \dot{z}_p/|\dot{z}_p| \\ \lambda_P \end{bmatrix} & \Rightarrow & \begin{bmatrix} x \\ y \\ z \\ \dot{x} \\ \dot{y} \\ \dot{z} \end{bmatrix} \end{array} \tag{4.185}$$

where

$$i = \text{inclination}$$

$$\phi_{PD} = \text{geodetic sublatitude of perigee}$$

$$\dot{z}_p/|\dot{z}_p| = \text{direction of motion at perigee}$$

$$(+1 \text{ for northbound; } -1 \text{ for southbound})$$

$$\lambda_P = \text{perigee longitude}$$

What is not explicitly stated is that apogee altitude h_A is measured above a geodetic latitude whose magnitude is not equal to the magnitude of the latitude of the subperigee point. Whatever the difference in the magnitudes of the geodetic latitudes, apogee altitude h_A is placed so that the resulting apogee will lie on a straight line that contains both perigee and the center of the Earth.

Mathematically, the following process is used, provided that $h_A \neq 0$ and $h_P < h_A$. First, the perigee radii r_p and its declination δ_P are determined from

$$\phi'_P = \tan^{-1}\left[(1-f)^2 \tan\phi_{\mathrm{PD}}\right] \tag{4.186}$$

$$r_E = \frac{a_E(1-f)}{\sqrt{1-f(2-f)\cos^2\phi'_P}} \tag{4.187}$$

$$x' = r_E \cos\phi'_p + h_P \cos\phi_{\mathrm{PD}} \tag{4.188}$$

$$z' = r_E \sin\phi'_p + h_P \sin\phi_{\mathrm{PD}} \tag{4.189}$$

$$r_p = \sqrt{x'^2 + z'^2} \tag{4.190}$$

$$\delta_P = \sin^{-1}\left(\frac{z'}{r}\right) \tag{4.191}$$

Then, the apogee radius r_A and declination δ_A are calculated from

$$r_A = a_e \frac{B + \sqrt{B^2 - C}}{2} \tag{4.192}$$

where

$$B = \left(1 - f\sin^2\delta_A + \frac{f^2}{8}\sin^2 2\delta_A\right) + \frac{h_A}{a_e} \tag{4.193}$$

$$C = 2f^2 \sin^2 2\delta_A \tag{4.194}$$

and

$$\delta_A = -\delta_P \tag{4.195}$$

Equation (4.192) is derived from Eq. (4.164) by moving h to the right side of the

equation, multiplying by r, and then rearranging the terms to get

$$r^2 - \left[a_e\left(1 - f\sin^2\delta + \frac{f^2}{8}\sin^2 2\delta\right) + h\right]r + \frac{a_e^2 f^2}{2}\sin^2 2\delta = 0 \tag{4.196}$$

which is a quadratic equation with solutions in the form of

$$r = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{4.197}$$

where

$$a = 1 \tag{4.198}$$

$$b = -\left[a_e\left(1 - f\sin^2\delta + \frac{f^2}{8}\sin^2 2\delta\right) + h\right] \tag{4.199}$$

$$c = \frac{a_e^2 f^2}{2}\sin^2 2\delta \tag{4.200}$$

Now, in seeking a solution, Eq. (4.197) must make sense with any value of f, including the case in which $f = 0$. Take the case in which $f = 0$, and reduce the equation to

$$r = \frac{(a_e + h) \pm \sqrt{(a_e + h)^2}}{2} \tag{4.201}$$

Immediately we see that the sign to the square root term must be + so that $r = a_e + h$, otherwise, we will obtain a solution of $r = 0$ that does not make sense.

Accordingly, we can now write Eq. (4.197) as

$$\frac{r}{a_e} = \frac{B + \sqrt{B^2 - C}}{2} \tag{4.202}$$

where

$$B = \left(1 - f\sin^2\delta + \frac{f^2}{8}\sin^2 2\delta\right) + \frac{h}{a_e} \tag{4.203}$$

and

$$C = 2f^2\sin^2 2\delta \tag{4.204}$$

To continue, the perigee declination determined in Eq. (4.191) must satisfy the following condition; otherwise, there is an inconsistency in the orbital elements, as stated in Eq. (4.185), and the process must be aborted.

$$|\delta_p| \leq i, \qquad \text{if } i \leq 90 \text{ deg} \tag{4.205}$$

or

$$|\delta_P| \leq 180 - i, \qquad \text{if } i > 90 \text{ deg} \tag{4.206}$$

Next, the right ascension of perigee is calculated from

$$\alpha_p = [\alpha_{g_0} + \lambda_P]_{\text{mod } 360 \text{ deg}}, \qquad 0 \le \alpha_P < 360 \text{ deg} \tag{4.207}$$

where

$$\alpha_{g_0} = \text{right ascension of Greenwich at epoch} \tag{4.208}$$

Note that, in Eq. (4.207), perigee is assumed to occur at epoch.

At this point, we have converted the orbital elements of perigee/apogee altitudes and perigee longitude to their counterparts of perigee/apogee radii and perigee right ascension. To complete the process, two more conversions are necessary, as shown symbolically as follows. Their mathematical processes are described in the two steps that follow.

$$\begin{array}{ccccc} \text{Orbital elements} & & \text{Classical elements} & & \text{ECI elements} \\ \begin{bmatrix} r_p \\ r_A \\ i \\ \delta_P \\ \dot{z}_p/|\dot{z}_p| \\ \alpha_P \end{bmatrix} & \Rightarrow & \begin{bmatrix} a \\ e \\ i \\ \Omega \\ \omega \\ \theta \end{bmatrix} & \Rightarrow & \begin{bmatrix} x \\ y \\ z \\ \dot{x} \\ \dot{y} \\ \dot{z} \end{bmatrix} \end{array} \tag{4.209}$$

Step 1: The first step is to convert from the orbital elements of perigee/apogee radii to the modified classical elements used in Sec. 4.1.

The required conditions are

$$r_A \neq 0 \tag{4.210}$$

$$r_A \ge r_P \tag{4.211}$$

$$|\delta_p| \le i, \quad \text{if} \quad i \le 90 \text{ deg} \tag{4.212}$$

$$|\delta_P| \le 180 \text{ deg} - i, \quad \text{if} \quad i > 90 \text{ deg} \tag{4.213}$$

If these are satisfied, then,

$$a = \frac{r_P + r_A}{2} \tag{4.214}$$

$$e = \frac{r_A - r_P}{r_A + r_P} \tag{4.215}$$

If $i \neq 0$ or 180 deg,

$$\omega = \sin^{-1}\left(\frac{\sin \delta_P}{\sin i}\right) \tag{4.216}$$

$$\omega = 180 \text{ deg} - \omega, \qquad \text{if} \left(\frac{\dot{z}_P}{|\dot{z}_P|}\right) < 0 \tag{4.217}$$

If $i = 0$ or 180 deg,

$$\omega = 0 \tag{4.218}$$

$$\omega = [\omega]_{\text{mod } 360 \text{ deg}} \tag{4.219}$$

if $i \neq 90$ deg,

$$\Delta\alpha = \tan^{-1}\left(\frac{\sin\omega\cos i}{\cos\omega}\right) \tag{4.220}$$

If $i = 90$ deg,

$$\Delta\alpha = 0 \text{ for } (\cos\omega > 0) \text{ or } (\cos\omega = 0 \text{ and } \sin\omega > 0) \tag{4.221}$$

$$\Delta\alpha = 180 \text{ deg for } (\cos\omega < 0) \text{ or } (\cos\omega = 0 \text{ and } \sin\omega < 0) \tag{4.222}$$

$$\Omega = [\alpha_P - \Delta\alpha]_{\text{mod } 360 \text{ deg}} \tag{4.223}$$

$$\theta = 0 \tag{4.224}$$

Step 2: The second step is to convert the modified classical elements to their ECI counterparts as shown symbolically here:

$$\begin{array}{ccc} \text{Classical elements} & & \text{ECI elements} \\ \begin{bmatrix} a \\ e \\ i \\ \Omega \\ \omega \\ \theta \end{bmatrix} & \Rightarrow & \begin{bmatrix} x \\ y \\ z \\ \dot{x} \\ \dot{y} \\ \dot{z} \end{bmatrix} \end{array} \tag{4.225}$$

The mathematical steps are as follows:

$$p = a(1 - e^2) \tag{4.226}$$

$$r = \frac{p}{1 + e\cos\theta} \tag{4.227}$$

$$V = \sqrt{\mu\left(\frac{2}{r} - \frac{1}{a}\right)} \tag{4.228}$$

where μ = gravitational parameter

$$\sin\gamma = \sqrt{\frac{\mu}{p}}\frac{e}{V}\sin\theta \tag{4.229}$$

$$\cos\gamma = \sqrt{1 - \sin^2\gamma} \tag{4.230}$$

Then the unit position vector in ECI coordinates is obtained by the following series of rotations:

$$\begin{bmatrix}\hat{x}\\ \hat{y}\\ \hat{z}\end{bmatrix} = \begin{bmatrix}\cos\Omega & -\sin\Omega & 0\\ \sin\Omega & \cos\Omega & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\ 0 & \cos i & -\sin i\\ 0 & \sin i & \cos i\end{bmatrix}\begin{bmatrix}\cos u & -\sin u & 0\\ \sin u & \cos u & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}1\\ 0\\ 0\end{bmatrix} \tag{4.231}$$

where $u = \omega + \theta$.

The unit velocity vector is obtained by the same series of rotations with one additional rotation.

$$\begin{bmatrix}\dot{\hat{x}}\\ \dot{\hat{y}}\\ \dot{\hat{z}}\end{bmatrix} = \begin{bmatrix}\cos\Omega & -\sin\Omega & 0\\ \sin\Omega & \cos\Omega & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\ 0 & \cos i & -\sin i\\ 0 & \sin i & \cos i\end{bmatrix}\begin{bmatrix}\cos u & -\sin u & 0\\ \sin u & \cos u & 0\\ 0 & 0 & 1\end{bmatrix}$$

$$\begin{bmatrix}\cos\left(\frac{\pi}{2}-\gamma\right) & -\sin\left(\frac{\pi}{2}-\gamma\right) & 0\\ \sin\left(\frac{\pi}{2}-\gamma\right) & \cos\left(\frac{\pi}{2}-\gamma\right) & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}1\\ 0\\ 0\end{bmatrix} \tag{4.232}$$

By carrying out these matrix multiplications and multiplying by r and V, respectively, we obtain

$$\begin{bmatrix}\hat{x}\\ \hat{y}\\ \hat{z}\end{bmatrix} = \begin{bmatrix}\cos\Omega\cos u - \sin\Omega\cos i\sin u\\ \sin\Omega\cos u + \cos\Omega\cos i\sin u\\ \sin i\sin u\end{bmatrix} \tag{4.233}$$

$$\begin{bmatrix}\dot{x}\\ \dot{y}\\ \dot{z}\end{bmatrix} = V\begin{bmatrix}\hat{x}\sin\gamma - \cos\gamma(\cos\Omega\sin u + \sin\Omega\cos i\cos u)\\ \hat{y}\sin\gamma - \cos\gamma(\sin\Omega\sin u - \cos\Omega\cos i\cos u)\\ \hat{z}\sin\gamma + \cos\gamma\cos u\sin i\end{bmatrix} \tag{4.234}$$

and

$$\begin{bmatrix}x\\ y\\ z\end{bmatrix} = r\begin{bmatrix}\hat{x}\\ \hat{y}\\ \hat{z}\end{bmatrix} \tag{4.235}$$

References

[1]Prussing, J. E., "Bounds on the Solution to Kepler's Problem," *Journal of the Astronautical Sciences,* Vol. 25, 1977, pp. 123–128.

[2]Ng, E. W., "A General Algorithm for the Solution of Kepler's Equation for Elliptic Orbits," *Celestial Mechanics,* Vol. 20, 1979, pp. 243–249.

[3]Smith, G. R., "A Simple, Efficient Starting Value for the Iterative Solution of Kepler's Equation," *Celestial Mechanics,* Vol. 19, 1979, pp. 163–166.

[4]Danby, J. M. A., and Burkhardt, T. M., "The Solution of Kepler's Equation, I," *Celestial Mechanics,* Vol. 31, 1983, pp. 95–107.

[5]Brouke, R., "On Kepler's Equation and Strange Attractors," *Journal of the Astronautical Sciences,* Vol. 28, 1980, pp. 255–265.

[6]Conway, B. A., "An Improved Algorithm Due to Laguerre for the Solution of Kepler's Equation," *Celestial Mechanics,* Vol. 39, 1986, pp. 199–211.

[7]Bate, R. R., Mueller, D. O., and White, J. E., *Fundamentals of Astrodynamics,* Dover, New York, 1971.

[8]Gersten, R. H., "Geodetic Sub-latitude and Altitude of a Space Vehicle," *Journal of the Astronautical Sciences,* Vol. 8, Technical Notes, 1961, pp. 28–29.

Problems

4.1. Using the Newton–Raphson iteration method, solve Kepler's equation to find the eccentric anomaly E in degrees, where the eccentricity e is 0.1 and the mean anomaly M is 90 deg. Carry out the calculation for two iterations only, and remember to work in radians.

a) Use M for the first value of E.

b) Use $M + e$ for the first value of E.

c) For parts a and b, what are the magnitudes (in degrees) of the errors in E after two iterations?

4.2. Solve Kepler's equation using the Newton–Raphson iteration method, and find E to three significant figures for $M = 5$ deg and $e = 0.9$. Start with $E_0 = M$.

4.3. Repeat Problem 4.2 using the Laguerre–Conway iteration method, where

$$f(E) = E - e \sin E - M$$

and the correction term is

$$\Delta E = \frac{5f(E)}{f'(E) \pm 2\sqrt{|4[f'(E)]^2 - 5f(E)f''(E)|}}$$

so that

$$E_{i+1} = E_i - \Delta E_i$$

Note: In calculating ΔE, the sign before the square root term should be chosen to maximize the absolute value of the denominator.

4.4. A satellite is in a Keplerian orbit with a period $P = 270$ min and eccentricity $e = 0.5$. It has passed its perigee and is now at a point at which the orbit intersects the semilatus rectum of the orbit. How much time (in minutes) has elapsed since perigee passage?

4.5. A rocket carrying a payload is launched into a ballistic trajectory. At apogee, the payload separates and, using its own propulsion, proceeds into orbit. The rocket continues on until it impacts with the surface of the Earth. An observer on Earth notes that, at apogee, the rocket is at an altitude of 150 km and is traveling at a

speed of 2.22 km/s. Once launched, the rocket is in free flight until it impacts the Earth.

a) At how many kilometers downrange (measured on the surface of the Earth from the subapogee point) can the rocket be expected to impact the Earth? Assume the Earth to be a sphere, with a radius of 6378.137 km ($\mu = 398600.5$ km^3/s^2).

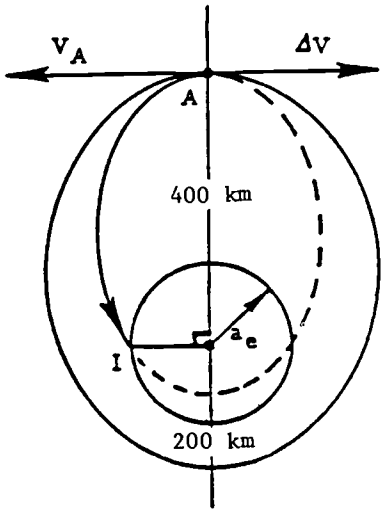
b) How long (in seconds) does it take for the rocket to impact the Earth after the payload is separated at apogee?

c) By how many kilometers will the original impact point be missed if the observer reported the apogee velocity incorrectly as 2.23 km/s?

4.6. A spaceship is moving around the Earth in a 200-×400-km orbit. At apogee, a retrorocket is fired in the direction exactly opposite to the spaceship's velocity vector. The magnitude of the retrorocket ΔV is such that the spaceship will impact the Earth after traveling an Earth-centered angle of 90 deg (i.e., $\Delta\theta = 90$ deg).

a) How much time Δt (in minutes) would have elapsed between the retro firing and impact? Assume the Earth to be a sphere, with a radius of 6378.137 km ($\mu = 398600.5$ km^3/s^2).

b) What is the magnitude of the retrorocket ΔV in kilometers per second?



4.7. A spaceship, in a circular orbit about the Earth at an altitude of 200 km, fires a projectile in a direction opposite to the spaceship's motion. After leaving the spaceship, the projectile impacts the Earth at an Earth-centered angle of 120 deg. Assume the Earth to be a sphere, with $a_e = 6378.137$ km ($\mu = 398600.5$ km^3/s^2).

a) What is the magnitude of the ΔV given the projectile in kilometers per second?

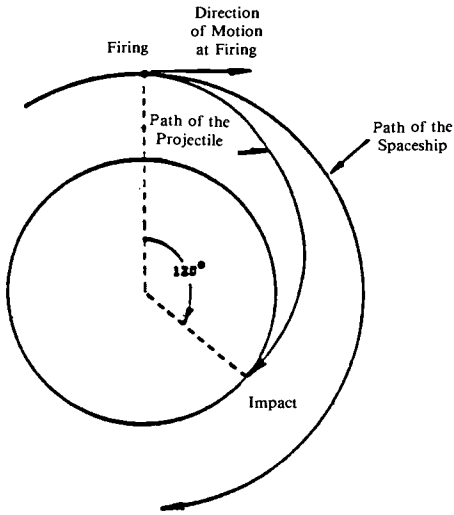
b) How long (in minutes) does it take the projectile to reach the Earth?

c) At impact, where is the spaceship relative to the impact point (e.g., directly overhead, before it, or after it)?

d) What is the time difference (in seconds) between impact and the spaceship's direct flyover of the impact point?

e) Describe briefly and *qualitatively* the relative motion of the projectile with respect to the spaceship from firing to impact.
Remember:

$$r = \frac{a(1-e^2)}{1+e\cos\theta} \text{ and } \frac{V^2}{\mu} = \frac{2}{r} - \frac{1}{a}$$



Selected Solutions

4.1. a) 95.7 deg

4.2. 33.3 deg

4.4. 26.39 min

4.5. a) 409.95 km
b) 186.1 s
c) 2.01 km

4.6. a) 22.828 min
b) 0.172 km/s

4.7. a) 0.08011 km/s
b) 29.095 min
d) 24.15 s

5
Orbital Maneuvers

5.1 Orbital Energy

Any analysis of orbital maneuvers, i.e., the transfer of a satellite from one orbit to another by means of a change in velocity, logically begins with the energy or vis-viva equation derived in Chapter 3 as

$$V^2 = \mu\left[\frac{2}{r} - \frac{1}{a}\right] \tag{5.1}$$

where V is the magnitude of the orbital velocity at some point, r the magnitude of the radius from the focus to that point, a the semimajor axis of the orbit, and μ the gravitational constant of the attracting body. Figure 5.1 illustrates r, V, and a.

Equation (5.1) can be rearranged as

$$\frac{V^2}{2} - \frac{\mu}{r} = -\frac{\mu}{2a} \tag{5.2}$$

where it is evident that

$$\frac{\text{Kinetic energy}}{\text{Satellite mass}} + \frac{\text{Potential energy}}{\text{Satellite mass}} = \frac{\text{Total energy}}{\text{Satellite mass}}$$

Note that total energy/satellite mass is dependent only on a. As a increases, energy increases.

Figure 5.2 illustrates total energy/satellite mass as a function of orbit period,

$$P = \frac{2\pi}{\sqrt{\mu}}a^{3/2} \tag{5.3}$$

so that the figure is really a plot of $-\mu/2a$ vs $(2\pi/\sqrt{\mu})a^{3/2}$. The lowest energy and period point corresponds to a Space Shuttle parking orbit, a circular orbit at an altitude of 280 km. Points corresponding to other interesting orbits are labeled on the curve.

Note that many of the orbits are circular for which $a = r$. When substituted into Eq. (5.2), $V_c = \sqrt{\mu/r}$ is the expression for the circular orbit velocity.

Note also that the total energy values for the orbits in this figure are all negative. For the Shuttle parking orbit, total orbit energy/satellite mass $= -30$ km^2/s^2. For the orbit of the moon, total orbit energy/satellite mass $= -0.52$ km^2/s^2. As Eq. (5.2) shows, the reference level of zero total energy is achieved as $a \to \infty$, and the orbit becomes a parabola. Substituting into Eq. (5.2),

$$\frac{V^2}{2} - \frac{\mu}{r} = 0, \qquad V = \sqrt{\frac{2\mu}{r}} \quad \text{for a parabola}$$

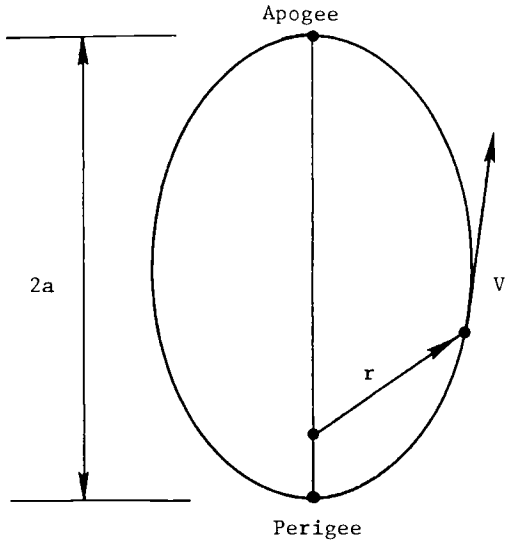


Fig. 5.1 Conservation of energy relates *r*, *V*, and *a*.

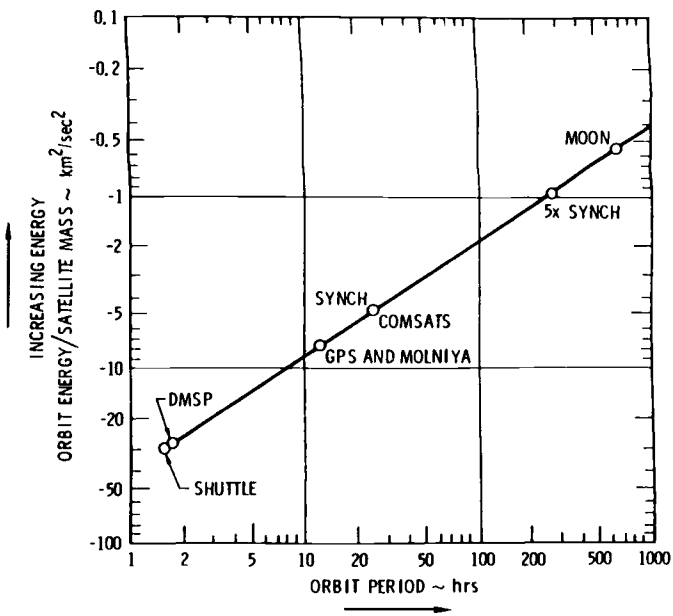


Fig. 5.2 Specific orbital energy as a function of orbit period.

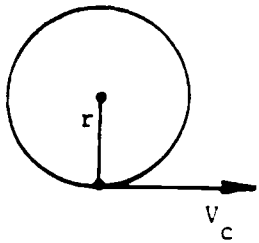


Fig. 5.3 For a circular orbit $V_c = \sqrt{\mu/r}$.

This value is called the "escape velocity" because a satellite possessing this velocity will always increase its distance (after the point of closest approach) from its gravitationally attracting body approaching an infinite distance in an infinite time.

Equation (5.2) also applies to hyperbolic orbits for which semimajor axis a is by definition, negative. Therefore, $(-\mu/2a)$ is positive for hyperbolas. Figure 5.4 illustrates total energy for the three kinds of orbits.

5.2 Single-Impulse Maneuvers

Orbital maneuvers are characterized by a change in orbital velocity. If a velocity increment ΔV, which is a vector, is added to a satellite velocity V_1, also a vector, then a new satellite velocity, V_2, results. If the ΔV is added instantaneously, the maneuver is called an impulsive maneuver. Figure 5.5 illustrates an impulsive maneuver or transfer between intersecting coplanar orbits. The transfer could have been made at point O instead of point P. The selection of transfer points may be made on the basis of $|\Delta V|$, i.e., the smaller the better.

A plane change maneuver rotates the orbit plane in inertial space while holding the size and shape, a and e, fixed. Figure 5.6 illustrates a circular orbit plane change through an angle θ. The ΔV is applied so as to rotate the velocity, V_{c1}, through the angle θ. The resulting velocity triangle is isosceles, as seen in Fig. 5.7. From the right triangle,

$$\frac{\Delta V/2}{V_{c1}} = \sin\theta/2$$

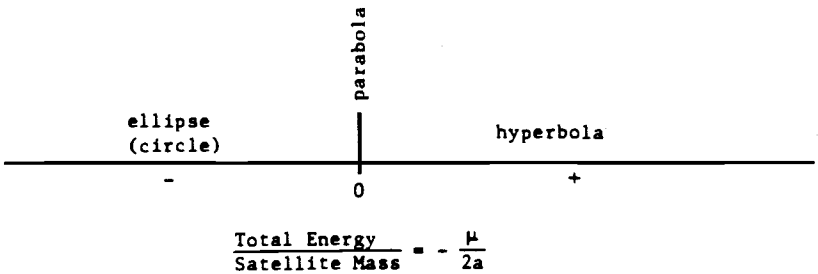


Fig. 5.4 Total energy for conic section orbits.

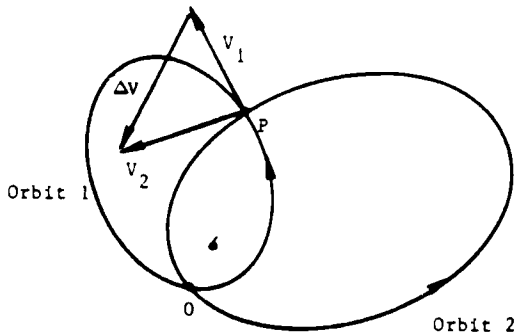


Fig. 5.5 Transfer from orbits 1 to 2; vector difference $\Delta V = V_2 - V_1$.

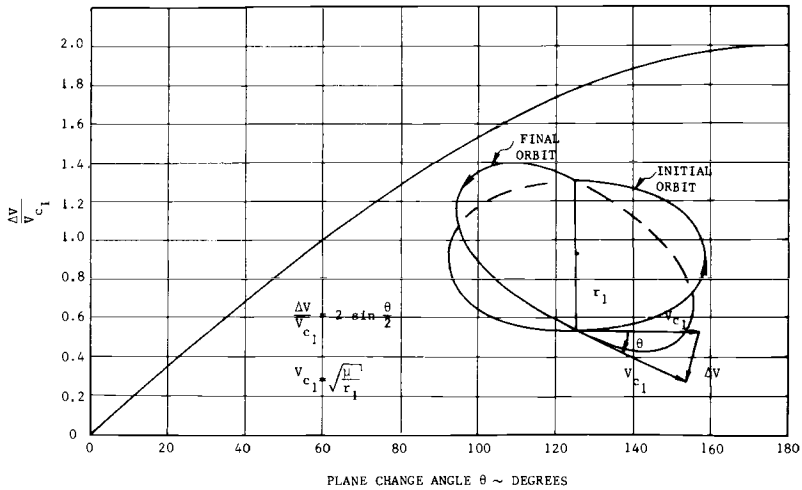


Fig. 5.6 Single impulse plane change for circular orbits.

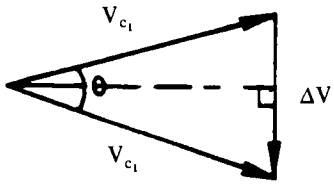


Fig. 5.7 Velocity vector triangle for circular orbit plane change.

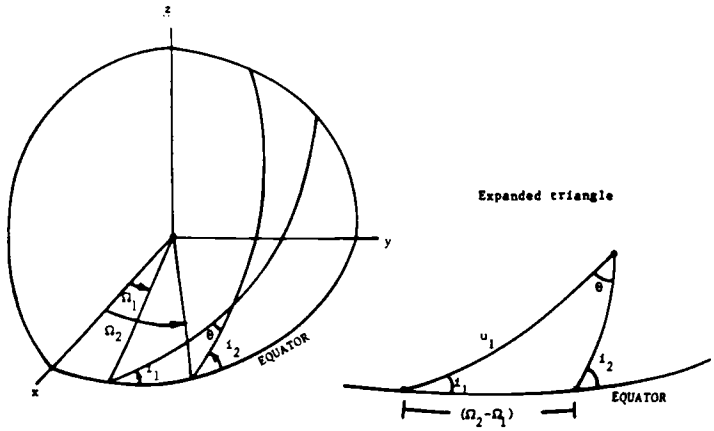


Fig. 5.8 Plane change angle θ in terms of i and Ω

or

$$\frac{\Delta V}{V_{c1}} = 2\sin\theta/2$$

Figure 5.6 plots $\Delta V/V_{c1}$ as a function of θ and shows that plane changes require large ΔV for even modest values of θ. For $\theta = 60$ deg, $\Delta V = V_{c1}$.

Figure 5.8 displays the geometry of a plane change θ, which will, in general, change both the inclination i and the right ascension of ascending node Ω of the original orbit. The plane change will be constrained to an inclination change, i.e., no change in Ω, only if the plane change maneuver is performed at an equatorial crossing. In general, the angle θ between two planes is a function of both inclination i and right ascension of ascending node Ω. Given the initial orbit elements i_1 and Ω_1, the plane change angle θ, and the argument of latitude u_1 of the plane change maneuver, the following two equations from spherical trigonometry can be solved for the final orbit elements, i_2 and Ω_2:

$$\cos i_2 = \cos i_1 \cos\theta - \sin i_1 \sin\theta \cos u_1$$

$$\cos\theta = \cos i_1 \cos i_2 + \sin i_1 \sin i_2 \cos(\Omega_2 - \Omega_1)$$

Another common single-impulse maneuver is a tangential ΔV applied to a circular orbit or to an elliptical orbit at perigee. Figure 5.9 shows a ΔV being added in the direction of motion, i.e., along the velocity vector, to the circular orbit velocity V_c. If the ΔV is relatively small, the resulting orbit is an ellipse, with perigee at the point of ΔV application and apogee located 180 deg away in central angle, i.e., true anomaly, $v = 180$ deg. Now, if the ΔV had been larger or if another relatively small ΔV is added tangentially at the next perigee passage, then the resulting orbit is a larger ellipse with the same perigee but a higher apogee. If the added $\Delta V = V_c(\sqrt{2} - 1)$, then the total velocity $= \sqrt{2}V_c = V_{esc}$ and the resulting orbit is parabolic. And if the added ΔV is greater than this value, then the resulting orbit is hyperbolic. It is interesting to note that, from a circular orbit,

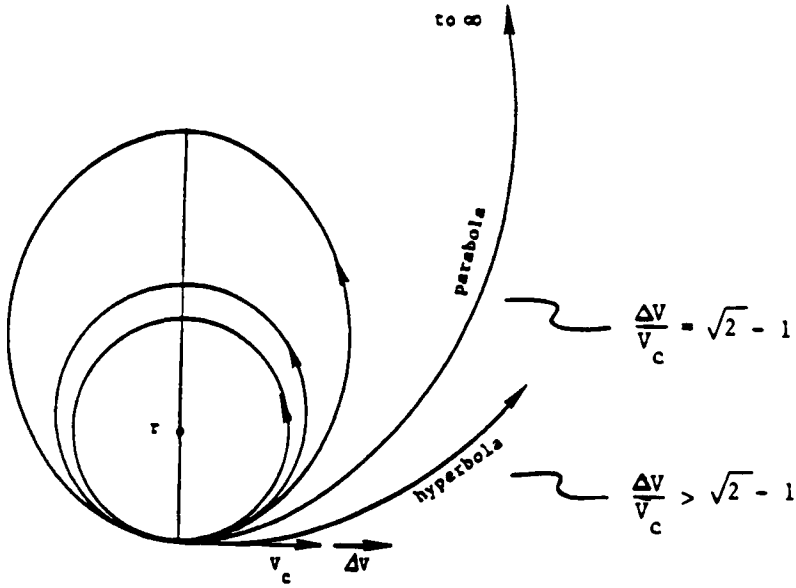


Fig. 5.9 Orbits resulting from a tangential velocity addition.

considerably more ΔV is required to transfer to a rectilinear orbit (drop into the center along a straight line) than to escape.

5.3 Single- and Two-Impulse Transfer Comparison for Coplanar Transfers Between Elliptic Orbits That Differ Only in Their Apsidal Orientation

Single- and Two-Impulse Transfers

For coplanar orbits 1 and 2 in Fig. 5.10, $a_1 = a_2$ and $e_1 = e_2$, but their lines of apsides are rotated by $\Delta\omega$. For single-impulse transfer at either intersection point,

$$\frac{\Delta V/2}{V} = \sin\gamma$$

where γ is the flight-path angle. To solve, substitute the orbit equation $r = p/(1 + e\cos v)$ into the energy equation

$$V^2 = \mu\left[\frac{2}{r} - \frac{1}{a}\right]$$

$$V^2 = \mu\left[\frac{2(1 + e\cos v)}{a(1 - e^2)} - \frac{1}{a}\right] = \mu\left[\frac{2(1 + e\cos v) - (1 - e^2)}{a(1 - e^2)}\right]$$

$$V^2 = \frac{\mu}{p}[2 + 2e\cos v - 1 + e^2] = V^2_{C_{r=p}}[1 + 2e\cos v + e^2]$$

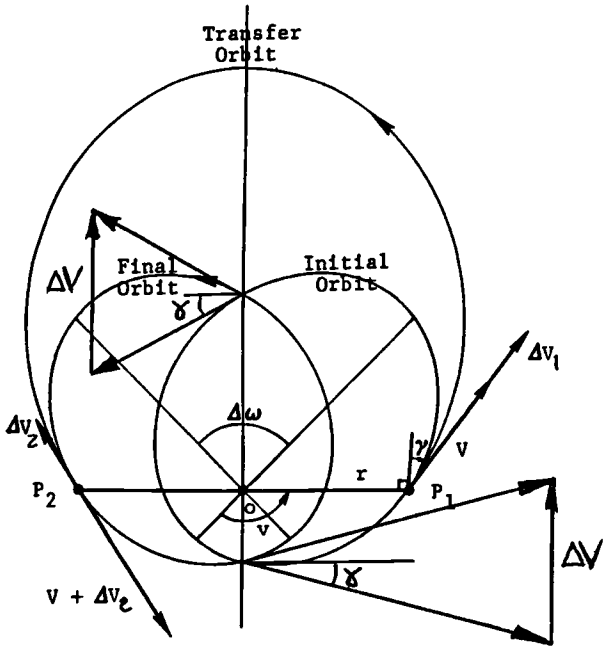


Fig. 5.10 Single- and two-impulse transfer comparison.

and substitute both into the equation for angular momentum, $h = rV\cos\gamma = \sqrt{\mu p}$, squared,

$$\frac{p^2}{(1+e\cos v)^2}\frac{\mu}{p}[1+2e\cos v+e^2]\cos^2\gamma = \mu p$$

so that

$$\cos\gamma = \frac{1+e\cos v}{\sqrt{1+2e\cos v+e^2}}, \qquad \sin\gamma = \frac{e\sin v}{\sqrt{1+2e\cos v+e^2}}$$

Substituting into the ΔV equation,

$$\frac{\Delta V}{V_{c_{r=p}}} = 2e\sin v$$

At the higher intersection,

$$v = 180\text{ deg} \pm \frac{\Delta\omega}{2}$$

and so

$$\frac{\Delta V}{V_{c_{r=p}}} = 2e\sin\frac{\Delta\omega}{2}$$

The solution is exactly the same at the lower intersection.

The solution for the optimal two-impulse cotangential transfer is

$$\frac{\Delta V_{\text{TOTAL}}}{V_{c_{r=p}}} = e \sin \frac{\Delta\omega}{2}$$

when e is assumed small. The sum of the two impulses is half the single-impulse value. The two impulses are equal in magnitude, but one is in the direction of motion while the other is opposite to the direction of motion. The optimal point of application for ΔV_1 is at $v_1 = 90$ deg $+\Delta\omega/2$, and ΔV_2 is applied 180 deg away (see Fig. 5.10). The optimal two-impulse transfer between these orbits is given by Lawden.[1]

5.4 Hohmann Transfer

The Hohmann transfer[2] is the minimum two-impulse transfer between coplanar circular orbits. Derivations of the velocity requirements ΔV_1 and ΔV_2 and the transfer time, as well as a figure of the transfer and plotted results, are presented in the following pages.

Referring to Fig. 5.11, the Hohmann transfer is a relatively simple maneuver. A tangential ΔV_1 is applied to the circular orbit velocity. The magnitude of ΔV_1 is determined by the requirement that the apogee radius of the resulting transfer ellipse must equal the radius of the final circular orbit. When the satellite reaches apogee of the transfer orbit, another ΔV must be added or the satellite will remain in the transfer ellipse. This ΔV is the difference between the apogee velocity in the transfer orbit and the circular orbit velocity in the final orbit. After ΔV_f has been applied, the satellite is in the final orbit, and the transfer has been completed.

Derivation of Velocity Requirements and Transfer Time

Using the vis-viva equation and referring to Fig. 5.11,

$$V_1^2 = \mu\left[\frac{2}{r_1} - \frac{2}{r_1 + r_f}\right] \qquad V_{c1}^2 = \frac{\mu}{r_1}$$

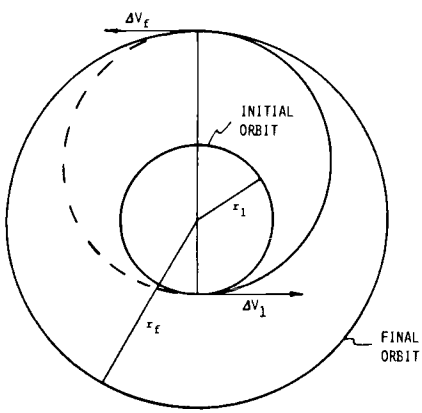


Fig. 5.11 Geometry of the Hohmann transfer.

$$V_1^2 = \frac{\mu}{r_1}\left[2 - \frac{2}{1 + (r_f/r_1)}\right] \qquad 2a = r_1 + r_f = r_1\left(1 + \frac{r_f}{r_1}\right)$$

$$V_1^2 = V_{c1}^2\left[\frac{2 + 2(r_f/r_1) - 2}{1 + (r_f/r_1)}\right] \qquad \frac{V_1}{V_{c1}} = \sqrt{\frac{2(r_f/r_1)}{1 + (r_f/r_1)}}$$

$$\frac{\Delta V_1}{V_{c1}} = \frac{V_1 - V_{c1}}{V_{c1}} = \sqrt{\frac{2(r_f/r_1)}{1 + (r_f/r_1)}} - 1$$

Again, from the vis-viva equation,

$$V_f^2 = \mu\left[\frac{2}{r_f} - \frac{2}{r_f[1 + (1/r_f/r_1)]}\right] \qquad r_1 + r_f = r_f\left(1 + \frac{1}{r_f/r_1}\right)$$

$$V_f^2 = \frac{\mu}{r_f}\left[2 - \frac{2r_f/r_1}{1 + r_f/r_1}\right] \qquad \begin{aligned} V_{cf}^2 &= \tfrac{\mu}{r_f} \\ V_{cf}^2 &= V_{c1}^2 \tfrac{r_1}{r_f} \end{aligned}$$

$$V_f^2 = V_{cf}^2\left[\frac{2 + 2(r_f/r_1) - 2(r_f/r_1)}{1 + (r_f/r_1)}\right]$$

$$V_f^2 = \frac{V_{c1}^2}{r_f/r_1}\left[\frac{2}{1 + (r_f/r_1)}\right]$$

$$\frac{V_f}{V_{c1}} = \sqrt{\frac{2}{r_f/r_1[1 + (r_f/r_1)]}}$$

$$\frac{\Delta V_f}{V_{c1}} = \frac{V_{cf} - V_f}{V_{c1}} = \sqrt{\frac{1}{r_f/r_1}} - \sqrt{\frac{2}{r_f/r_1[1 + (r_f/r_1)]}}$$

The total velocity requirement is

$$\Delta V_{\text{TOTAL}} = \Delta V_1 + \Delta V_f$$

From the equation for the orbital period,

$$\frac{\text{Transfer time}}{P_1} = \frac{\pi/\sqrt{\mu}[(r_f + r_1)/2]^{3/2}}{(2\pi/\sqrt{\mu})r_1^{3/2}} = \frac{1}{2^{5/2}}\left(1 + \frac{r_f}{r_1}\right)^{3/2}$$

where P_1 is the period of the initial orbit.

Referring to the plotted numerical results of Fig. 5.12, one of the interesting features of the Hohmann transfer is that, as the radius ratio r_f/r_1 increases, the total velocity requirements $\Delta V_{\text{TOTAL}}/V_{c1} = (\Delta V_1 + \Delta V_f)/V_{c1}$ increase to a maximum of 0.536 at $r_f/r_1 = 15.58$. For larger values of r_f/r_1, the total velocity requirements decrease and approach $\sqrt{2} - 1$ as $r_f/r_1 \to \infty$. This behavior can be explained somewhat by examining the behaviors of $\Delta V_1/V_{c1}$ and $\Delta V_f/V_{c1}$ in Fig. 5.12. The $\Delta V_1/V_{c1}$ curve monotonically increases with increasing r_f/r_1 and approaches $\sqrt{2} - 1$ as $r_f/r_1 \to \infty$. However, $\Delta V_f/V_{c1}$ increases to a maximum of 0.19 at $r_f/r_1 = 5.88$ and then decreases toward zero as $r_f/r_1 \to \infty$.

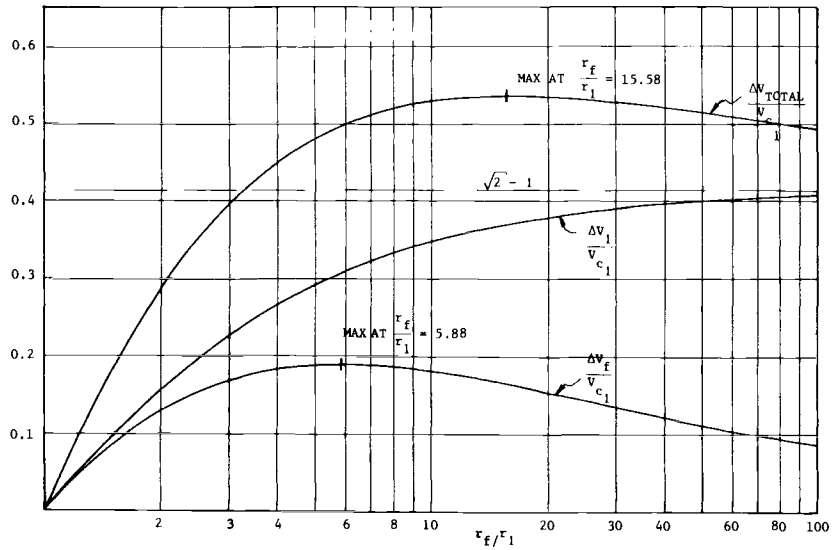


Fig. 5.12 ΔV characteristics of the Hohmann transfer.

5.5 The Bi-elliptic Transfer

The bi-elliptic transfer is completely described in Ref. 3. It is a three-impulse transfer between coplanar circular orbits. Its geometry, ΔV equations, numerical results, and a comparison with the Hohmann transfer are presented in Figs. 5.13–5.16.

The geometry of Fig. 5.13 shows that the transfer begins with a ΔV_1 applied tangentially to the circular orbit velocity. This ΔV_1 is larger than the first impulse of a corresponding Hohmann transfer because the apogee radius r_2 of the resulting transfer ellipse is larger than the final circular orbit radius r_f. At apogee in this first transfer ellipse, ΔV_2 is added tangentially to the existing apogee velocity. The magnitude of ΔV_2 is determined by the requirements to raise the perigee radius of the resulting transfer ellipse from r_1 to r_f. At perigee in this second transfer ellipse, ΔV_f is applied tangentially but opposite to the direction of motion. The

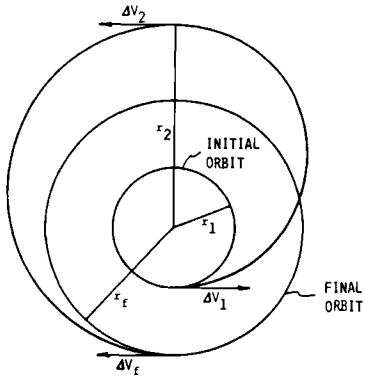


Fig. 5.13 Geometry of the bi-elliptic transfer.

magnitude of ΔV_f is the difference between the perigee velocity of the second transfer ellipse and the final circular orbit velocity.

The equations for the three ΔVs are presented below. Each ΔV is normalized by dividing by V_{c1}. Note that these velocity ratios are functions of r_f/r_1 and r_2/r_1. The final to initial orbit radius ratio is given in any specified transfer, but the ratio r_2/r_1 with the intermediate apogee radius r_2 is open to selection. If r_2/r_1 is selected to be equal to r_f/r_1, a Hohmann transfer will result. Although not practical, the best value of r_2 is infinitely large, i.e., $r_2 \to \infty$

$$\frac{\Delta V_1}{V_{c1}} = \sqrt{\frac{2(r_2/r_1)}{1 + (r_2/r_1)}} - 1$$

where $V_{c1} = \sqrt{\mu/r_1}$, and

$$\frac{\Delta V_2}{V_{c1}} = \sqrt{\frac{2(r_f/r_1)}{(r_2/r_1)[(r_2/r_1) + (r_f/r_1)]}} - \sqrt{\frac{2}{(r_2/r_1)[1 + (r_2/r_1)]}}$$

$$\frac{\Delta V_f}{V_{c1}} = \sqrt{\frac{2(r_2/r_1)}{(r_f/r_1)[(r_f/r_1) + (r_2/r_1)]}} - \frac{1}{r_f/r_1}$$

$$\Delta V_{\text{TOTAL}} = \Delta V_1 + \Delta V_2 + \Delta V_f$$

produces a minimum value of $\Delta V_{\text{TOTAL}} = \Delta V_1 + \Delta V_2 + \Delta V_f$ for all values of r_f/r_1. The bi-elliptic transfer for which $r_2 \to \infty$ is known as the infinite bi-elliptic transfer.

Numerical values of $\Delta V_{\text{TOTAL}}/V_{c1}$ for the infinite bi-elliptic transfer are plotted in Fig. 5.14, along with the Hohmann-transfer results. As r_f/r_1 increases, the

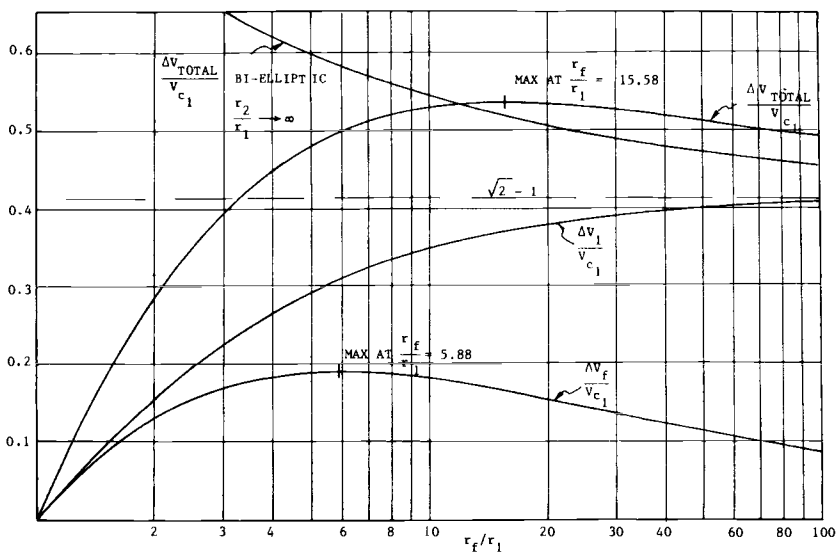


Fig. 5.14 Infinite bi-elliptic and Hohmann transfers.

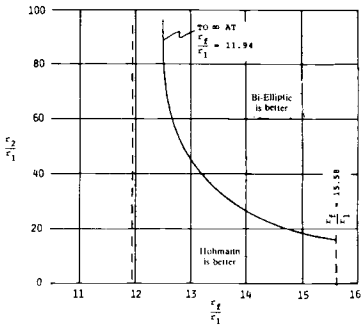


Fig. 5.15 Threshold locus of r_2/r_1 for which bi-elliptic ΔV_{TOTAL} = Hohmann ΔV_{TOTAL}.

Hohmann ΔV_{TOTAL} is less than the infinite bi-elliptic ΔV_{TOTAL} until $r_f/r_1 = 11.94$. At this value, the two curves cross, and the infinite bi-elliptic transfer has a lower ΔV_{TOTAL} than the Hohmann for all greater values of r_f/r_1. Moreover, the authors of Ref. 3 found that as r_f/r_1 increases beyond 11.94, the value of r_2/r_1 that would produce a bi-elliptic ΔV_{TOTAL} equal to the corresponding Hohmann ΔV_{TOTAL} was finite and decreasing in value. Figure 5.15 plots the threshold values of r_2/r_1 for which bi-elliptic ΔV_{TOTAL} = Hohmann ΔV_{TOTAL}. This curve decreases to a threshold value of $r_2/r_1 = 15.58$ at $r_f/r_1 = 15.58$. This means that all bi-elliptic transfers with $r_2/r_1 > r_f/r_1$ for $r_f/r_1 \geq 15.58$ are more economical than the corresponding Hohmann transfer, i.e., bi-elliptic ΔV_{TOTAL} < Hohmann ΔV_{TOTAL}.

It is interesting to determine the maximum savings to be gained from the bi-elliptic transfer relative to the Hohmann transfer. Figure 5.16 presents the

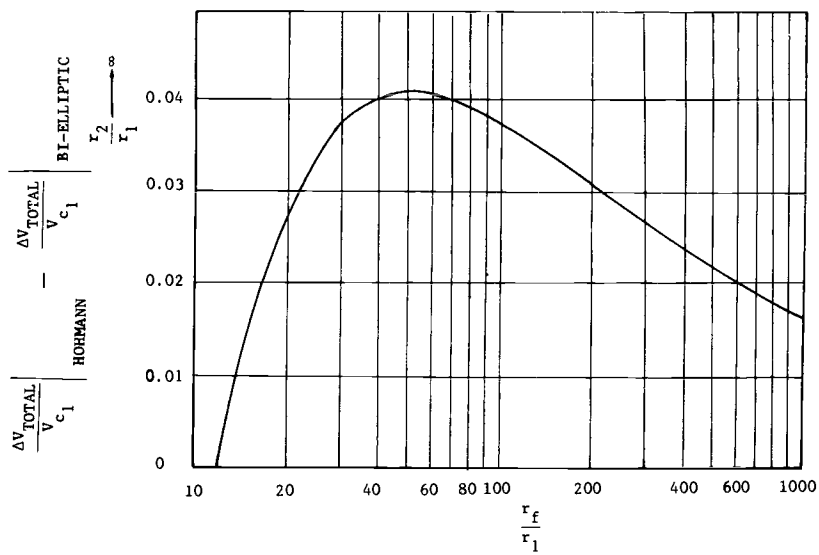


Fig. 5.16 Total velocity requirement difference between infinite bi-elliptic and Hohmann transfers.

difference in $\Delta V_{\text{TOTAL}}/V_{c1}$ between the infinite bi-elliptic and Hohmann transfers. The maximum difference is about 0.0409 at $r_f/r_1 = 50$. Since the Hohmann $\Delta V_{\text{TOTAL}}/V_{c1}$ at this r_f/r_1 is about 0.513, the maximum difference is about 8%. Thus, in terms of ΔV_1, the bi-elliptic transfer is not significantly better than the Hohmann transfer. However, bi-elliptic transfers are very useful when plane changes are necessary.

5.6 Restricted Three-Impulse Plane Change Maneuver for Circular Orbits

For the rotation of circular orbits, the restricted three-impulse plane change maneuver is intended to lower the total ΔV costs relative to the single-impulse plane change maneuver described earlier. The geometry, equations, derivation, and results for this maneuver are presented in Figs. 5.17–5.19 and 5.20–5.22. Figure 5.17 presents the geometry of the maneuver, which proceeds as follows. The first impulse ΔV_1 is added tangentially to the circular orbit velocity in order to achieve a transfer ellipse whose apogee radius is r_2. At apogee, ΔV_2 is used to rotate the apogee velocity through the desired plane change angle θ. The ΔV equations for this maneuver are

$$\Delta V_{\text{TOTAL}} = \Delta V_1 + \Delta V_2 + \Delta V_3$$

$$\frac{\Delta V_1}{V_{c1}} = \frac{\Delta V_3}{V_{c1}} = \sqrt{\frac{2(r_2/r_1)}{1 + (r_2/r_2)}} - 1$$

where $V_{c1} = \sqrt{\mu/r_1}$; and

$$\frac{\Delta V_2}{V_{c1}} = 2\sqrt{\frac{2}{(r_2/r_1)[1 + (r_2/r_1)]}}\sin\frac{\theta}{2}$$

where θ is the angle between the planes.

The equation for ΔV_2 is simply $\Delta V_2 = 2V_A \sin\theta/2$. The whole point of this maneuver is to make the orbit rotation at a point where the orbital velocity is low, i.e., apogee of the transfer ellipse. After the rotation, the satellite returns via

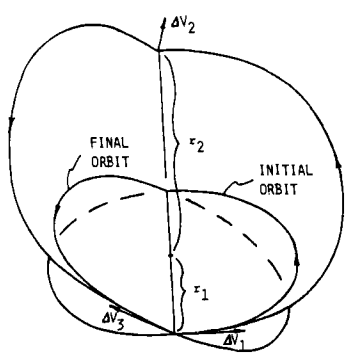


Fig. 5.17 Geometry of restricted three-impulse plane change maneuver for circular orbits.

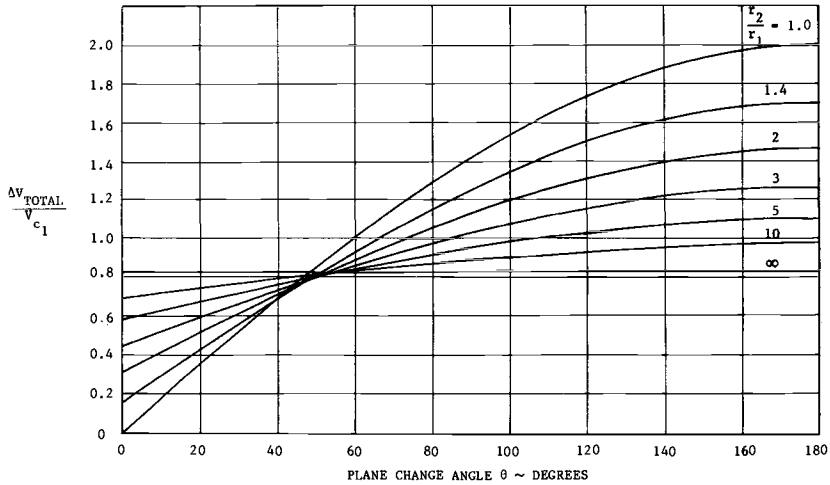


Fig. 5.18 Restricted three-impulse plane change results.

a second transfer ellipse to the original point of departure. At this point, ΔV_3 is applied tangentially in a retro direction to achieve the final circular orbit, which has the same radius as the initial circular orbit but has been rotated through the angle θ. The equations ΔV_1 and ΔV_3 are equal in magnitude. The first description of this maneuver with results was given in Ref. 4.

The ΔVs and their sum are a function of θ and r_2/r_1. Figure 5.18 presents numerical results of $\Delta V_{\text{TOTAL}}/V_{c1}$ vs θ for various values of r_2/r_1. When $r_2/r_1 = 1$, $\Delta V_1 = \Delta V_3 = 0$, and the result is the single-impulse plane change curve presented in Fig. 5.6. For larger values of r_2/r_1, the ΔV_{TOTAL} is larger for small values of θ but smaller for large values of θ. It is apparent that $r_2/r_1 = 1$ is the best curve, i.e., lowest ΔV_{TOTAL}, for values of θ from zero to about 38 or 39 deg and that $r_2/r_1 \to \infty$ is the best curve for values of θ greater than about 60 deg. Figure 5.19 presents an expanded portion of Fig. 5.18 for intermediate values of θ, where the best ΔV curves correspond to many values of r_2/r_1.

In order to determine the envelope of these curves, minimize $\Delta V_T/V_{c1}$ for given plane change angle θ, as illustrated in Fig. 5.20.

Let $\rho = r_2/r_1$. Then,

$$\frac{\Delta V_T}{V_{c1}} = 2\left[\sqrt{\frac{2\rho}{1+\rho}}\left(1 + \frac{\sin\theta/2}{\rho}\right) - 1\right]$$

Now,

$$\left.\frac{\partial(\Delta V_T/V_{c1})}{\partial\rho}\right|_\theta = 0 = 2\left\{\sqrt{\frac{2\rho}{1+\rho}}\left(-\frac{\sin\theta/2}{\rho^2}\right)\right.$$

$$\left. + \left(1 + \frac{\sin\theta/2}{\rho}\left(\frac{1}{2\sqrt{2\rho/(1+\rho)}}\left[\frac{(1+\rho)2 - 2\rho}{(1+\rho)^2}\right]\right)\right)\right\}$$

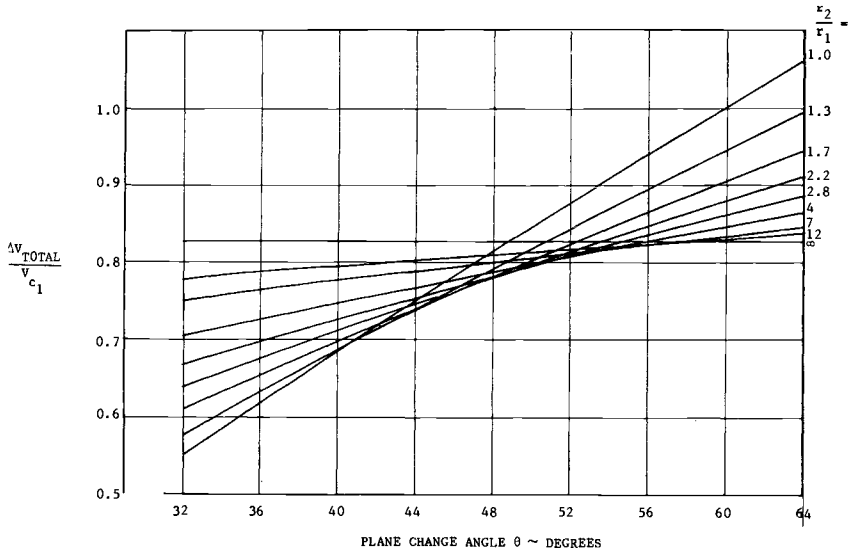


Fig. 5.19 Transition results for restricted three-impulse plane change.

Solving,

$$\rho_{\text{OPT}} = \frac{\sin\theta/2}{1 - 2\sin\theta/2} = \frac{1}{\cos\theta/2 - 2}$$

Upon examination, this equation reveals that, for $0 \leq \theta \leq 38.94$ deg, use

$$\frac{r_2}{r_1} = 1$$

for 38.94 deg $\leq \theta \leq$ 60 deg, use

$$\frac{r_2}{r_1} = \frac{\sin\theta/2}{1 - 2\sin\theta/2}$$

for 60 deg $\leq \theta \leq$ 180 deg, use

$$\frac{r_2}{r_1} \to \infty$$

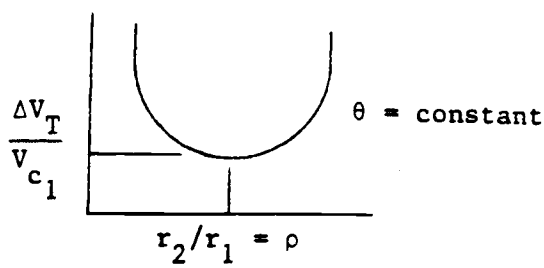


Fig. 5.20 Sketch of minimum $\Delta V_T/V_{c1}$ solution.

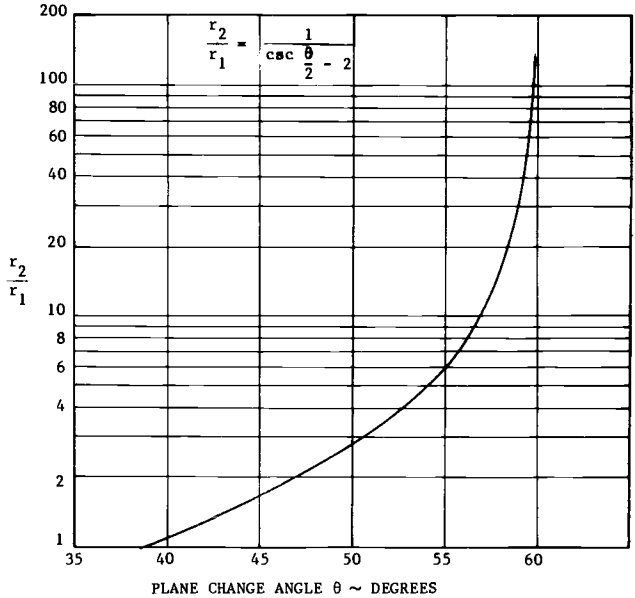


Fig. 5.21 Optimum r_2/r_1 as a function of plane change angle.

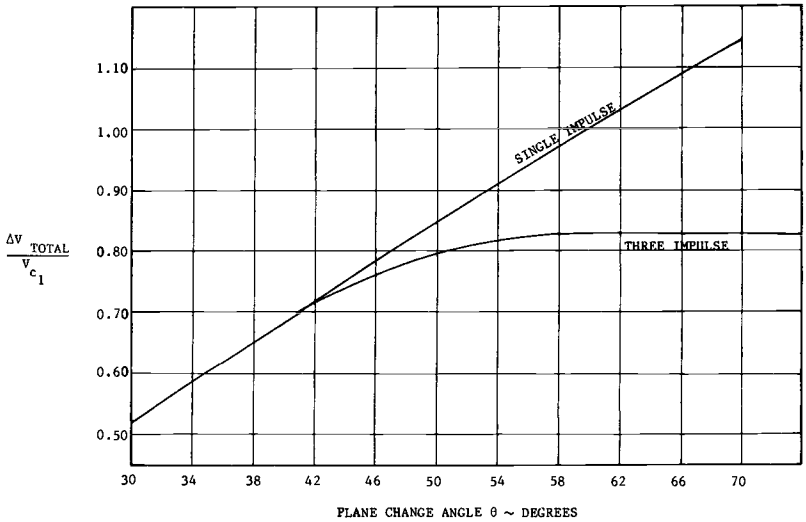


Fig. 5.22 Comparison of single- and restricted three-impulse maneuvers.

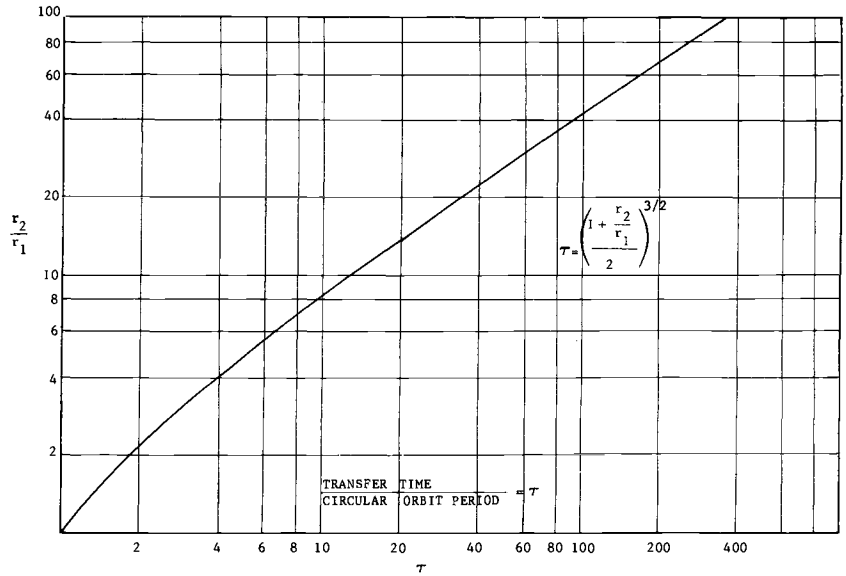


Fig. 5.23 Transfer time for the restricted three-impulse plane change maneuver.

Figure 5.21 presents a plot of optimum r_2/r_1 vs θ for 38.94 deg $\leq \theta \leq$ 60 deg. If these values are then substituted into the ΔV_T equation, they produce the minimum $\Delta V_T/V_{c1}$ as a function of θ. These results are presented in Fig. 5.22 and compared with the single-impulse results. The two curves are the same for $0 \leq \theta \leq$ 38.94 deg. Figure 5.23 presents the transfer time divided by the initial orbit period as a function of r_2/r_1.

As an example, a plane change of 55 deg would prescribe an optimum value of $r_2/r_1 = 6$ from Fig. 5.21. From Fig. 5.22, minimum $\Delta V_T/V_{c1} = 0.820$. From Fig. 5.23, transfer time/circular orbit period = 6.6.

5.7 General Three-Impulse Plane Change Maneuver for Circular Orbits

This maneuver is like the restricted maneuver just described in that the initial and final circular orbits have the same radius but are rotated through an angle and the maneuver utilizes three impulses. The general case, however, makes a plane change at each of the three ΔV applications. The sum of the plane changes equals the total required rotation θ. These are special cases, $r_f = r_1$, in Refs. 5 and 6. However, Ref. 7 presents an analysis that minimizes $\Delta V_{\text{TOTAL}}/V_{c1}$ by determining optimum values of r_2/r_1, i.e., intermediate apogee radius, and of plane change distribution among the three ΔV applications. By symmetry, the plane change performed as part of the first and third ΔV applications is the same.

Results from Ref. 7 are presented in Fig. 5.24 and compared with previous results for the single-impulse and restricted three-impulse maneuvers. The general three-impulse maneuver provides the best results, i.e., lowest $\Delta V_{\text{TOTAL}}/V_{c1}$, for all plane change angles from 0 to 60.185 deg. For small angles, the ΔV results are only slightly smaller than the ΔV results for a single-impulse maneuver. At $\theta = 60.185$

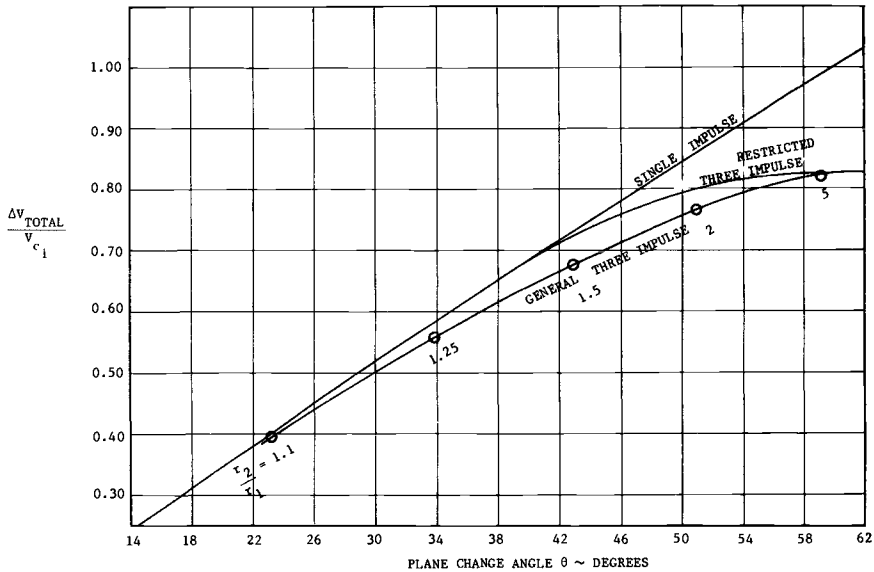


Fig. 5.24 Comparison of circular orbit plane change maneuvers.

deg, the general results merge with the results for a transfer to infinity and return, where $\Delta V_{\text{TOTAL}}/V_{c1} = 2(\sqrt{2} - 1)$. The optimum intermediate apogee radius ratio is noted for the circled points on the general three-impulse curve of Fig. 5.24. These values are generally lower than the corresponding values in Fig. 5.21 for the restricted three-impulse maneuver. This ratio is plotted on Fig. 5.25 as a function of θ. The interesting feature of this curve is that it remains very close to a value of 1 for relatively large values of θ until it increases dramatically as it approaches the value of $\theta = 60.185$ deg.

Figure 5.25 also presents a curve of the first plane change angle α_1 (or the third plane change angle α_3 since $\alpha_3 = \alpha_1$) as a function of θ, which equals $\alpha_1 + \alpha_2 + \alpha_3$. The value of α_1 reaches a maximum of about 4.85 deg at about $\theta = 49$ deg. For larger values of θ, α_1 then decreases, reaching 1.698 deg at $\theta = 60.185$ deg.

5.8 Hohmann Transfer with Split-Plane Change

An important practical circular orbit transfer is one that requires both a plane change and a radius change. The optimal two-impulse transfer to satisfy these requirements is the Hohmann transfer with plane change as shown in Fig. 5.26. The first ΔV not only produces a transfer ellipse whose apogee radius equals the final orbit radius, it also rotates the orbit plane through some angle α_1. At apogee, the second ΔV simultaneously circularizes the orbit and rotates the orbit plane through an angle α_2, where $\alpha_2 = \theta - \alpha_1$.

Figure 5.27 illustrates the velocity vector geometry for Hohmann transfer with plane change θ. Rider in Ref. 5 describes this as the "Mod-2 Hohmann" transfer. Baker in Ref. 6 describes this as the Hohmann transfer with plane change. Sketch 1

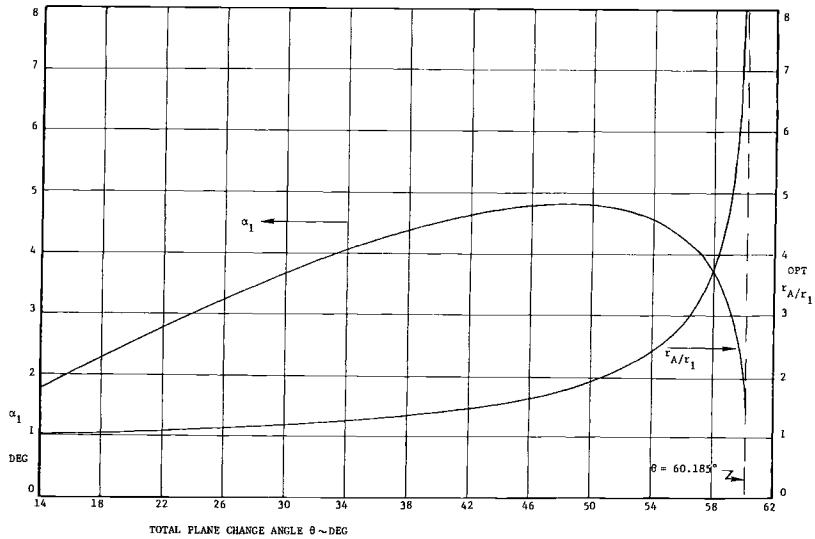
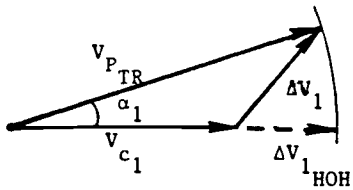


Fig. 5.25 Plane change angle and optimum intermediate apogee radius ratio for the general three-impulse plane change maneuver.

depicts the velocity vector triangle addition of ΔV_1.

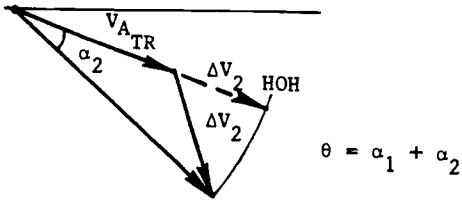


$$\Delta V_1^2 = V_{P_{\mathrm{TR}}}^2 + V_{c1}^2 - 2V_{P_{\mathrm{TR}}} V_{c1} \cos \alpha_1$$

$$V_{P_{\mathrm{TR}}} = V_{c1} + \Delta V_{1_{\mathrm{HOH}}}$$

$$V_{A_{\mathrm{TR}}} = V_{c2} - \Delta V_{2_{\mathrm{HOH}}}$$

Sketch 2 depicts the velocity vector triangle addition of ΔV_2.



$$\Delta V_2^2 = V_{c2}^2 + V_{A_{\mathrm{TR}}}^2 - 2V_{c2} V_{A_{\mathrm{TR}}} \cos(\theta - \alpha_1)$$

$$\theta = \alpha_1 + \alpha_2, \qquad \Delta V_T = \Delta V_1 + \Delta V_2$$

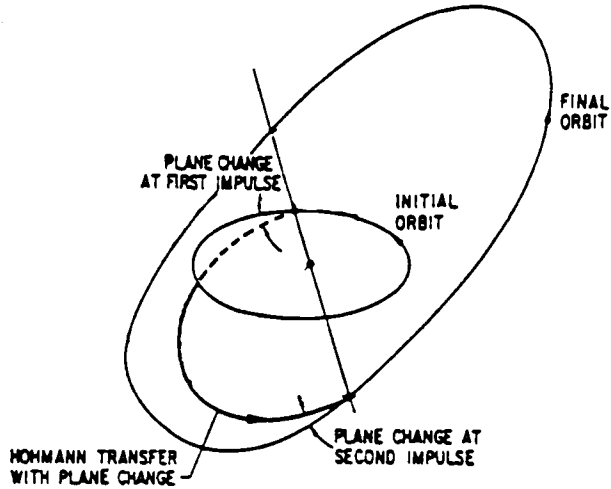


Fig. 5.26 Geometry of the Hohmann transfer with plane change.

To minimize ΔV_T, set $(\partial \Delta V_T / \partial \alpha_1) = 0$

$$\frac{\partial \Delta V_T}{\partial \alpha_1} = \frac{V_{P_{TR}} V_{c1} \sin \alpha_1}{\Delta V_1} - \frac{V_{c2} V_{A_{TR}} \sin(\theta - \alpha_1)}{\Delta V_2} = 0$$

Expanding

$$\frac{V_{P_{TR}} V_{c1} \sin \alpha_1}{\sqrt{V_{P_{TR}}^2 + V_{c1}^2 - 2V_{P_{TR}} V_{c1} \cos \alpha_1}}$$

$$= \frac{V_{c2} V_{A_{TR}} \sin(\theta - \alpha_1)}{\sqrt{V_{c2}^2 + V_{A_{TR}}^2 - 2V_{c2} V_{A_{TR}} \cos(\theta - \alpha_1)}}$$

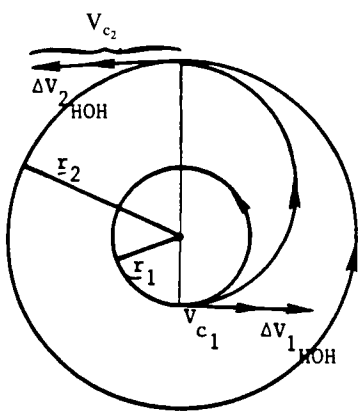


Fig. 5.27 Optimal plane split for Hohmann transfer with plane change.

Solve iteratively for $\alpha_{1_{\text{OPT}}}$. Then $\alpha_{2_{\text{OPT}}} = \theta - \alpha_{1_{\text{OPT}}}$. Substitute into the ΔV equations to get $\Delta V_{1_{\text{OPT}}}$ and $\Delta V_{2_{\text{OPT}}}$. Then,

$$\Delta V_{T_{\text{MIN}}} = \Delta V_{1_{\text{OPT}}} + \Delta V_{2_{\text{OPT}}}$$

5.9 Bi-elliptic Transfer with Split-Plane Change

This transfer is like the bi-elliptic transfer previously described except that plane changes are performed at each of the three ΔV applications. The geometry of the maneuver is presented in Fig. 5.28. Reference 5 describes this as the "Mod-2 bi-elliptic" transfer. Reference 6 describes this as the bi-elliptic transfer with plane change. Reference 7 presents an analysis that optimizes the plane change distribution among the three ΔVs and optimizes the intermediate apogee radius to minimize the total ΔV for a specified ratio of final to initial orbit radii and for a specified total plane change angle θ.

Results from Ref. 7 are presented in Fig. 5.29. In the space of final orbit radius/initial orbit radius and θ, Fig. 5.29 defines the best transfer modes, i.e., Hohmann transfer with plane change, bi-elliptic transfer with plane change, and parabolic transfer to infinity and return.

5.10 Transfer Between Coplanar Elliptic Orbits

Figure 5.30 illustrates two coplanar elliptic orbits, initial and final, and a transfer orbit between them. The transfer begins at a point of departure on the initial orbit. At this point, a velocity increment ΔV_1 is applied in some direction and added vectorially to the orbital velocity at that point. Thus, the satellite achieves a transfer orbit in which it coasts until it reaches an arrival point on the final orbit. At this point, a ΔV_2 that is the vector difference between the velocity in the final orbit and the velocity in the transfer orbit must be applied in order to complete the transfer.

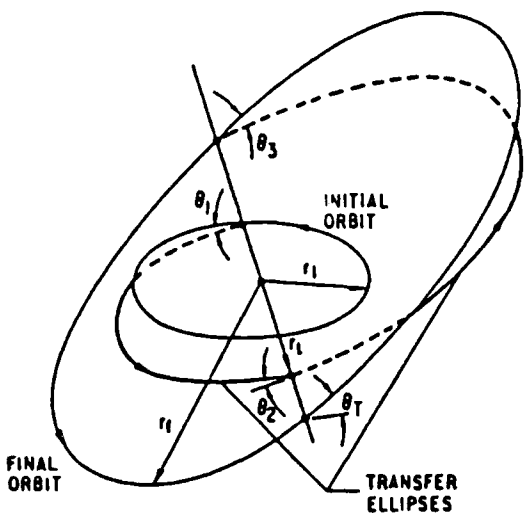


Fig. 5.28 Bi-elliptic transfer with plane change.

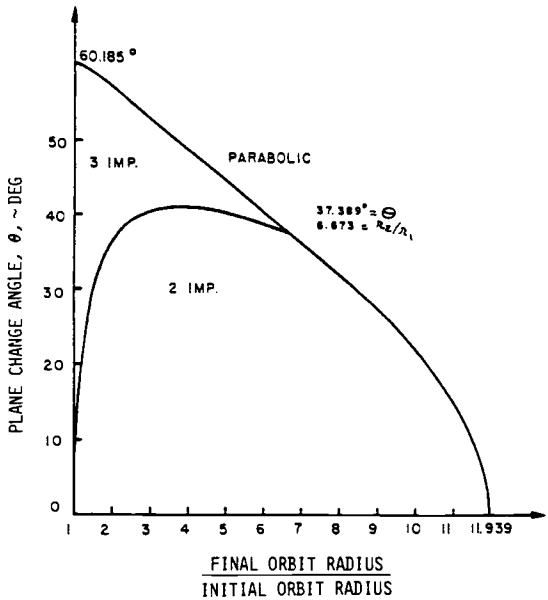


Fig. 5.29 Optimal strategies for transfer between noncoplanar circular orbits of radius r_1 and r_2 with optimal plane change distributions to minimize the total ΔV (from Ref. 7).

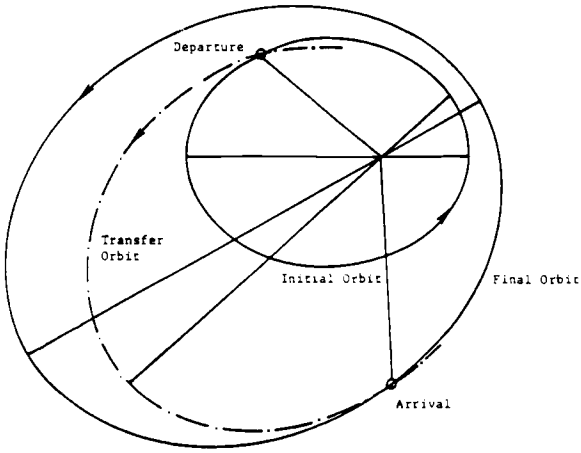


Fig. 5.30 Geometry of transfer between coplanar elliptic orbits.

Lawden, in Ref. 1, derives the equations for the optimal, i.e., minimum total ΔV, two-impulse transfer between coplanar elliptic orbits. These equations cannot be analytically solved in general. However, Lawden solved the special case of transfer between identical elliptic orbits that differ only in their orientation, as described in Sec. 5.3.

Other investigators have tried to solve this transfer problem by specializing the transfer to be a 180-deg transfer or to be cotangential. The 180-deg transfer restricts the departure and arrival points to be separated by 180 deg in central angle. The cotangential transfer requires the impulses to be tangent to the orbits at the points of application. These restrictions tend to simplify the problem of finding a minimum total ΔV solution. However, these solutions may not be as good as the general optimal solution and also are not easy to determine.

Bender, in Ref. 8, describes the theory and equations for the 180-deg transfer and the cotangential transfer. He also formulates a "practically optimum" transfer, which is a 180-deg circumferential transfer. Circumferential means that the ΔVs are applied in the circumferential direction, i.e., perpendicular to the radius vector. This maneuver is simple and easy to evaluate and provides good results. Comparisons of results of several transfer methods are presented in Ref. 8.

A very comprehensive survey of impulsive transfers is given in Ref. 9, which includes a listing of 316 references.

References

[1]Lawden, D. F., "Impulsive Transfer Between Elliptical Orbits," *Optimization Techniques*, edited by G. Leitmann, Academic, New York, 1962, Chap. 11.

[2]Hohmann, W., "Die Ereichbarkeit der Himmelskorper (The Attainability of Heavenly Bodies)," NASA, Technical Translations F-44, 1960.

[3]Hoelker, R. F., and Silber, R., "The Bi-Elliptical Transfer Between Coplanar Circular Orbits," *Proceedings of the 4th Symposium on Ballistic Missiles and Space Technology*, Vol. III, Pergamon, New York, 1961, pp. 164–175.

[4]Rider, L. A., "Characteristic Velocity for Changing the Inclination of a Circular Orbit to the Equator," *ARS Journal*, Vol. 29, Jan. 1959, pp. 48–49.

[5]Rider, L. A., "Characteristic Velocity Requirements for Impulsive Thrust Transfers Between Non-Coplanar Circular Orbits," *ARS Journal*, Vol. 31, March 1961, pp. 345–351.

[6]Baker, J. M., "Orbit Transfer and Rendezvous Maneuvers Between Inclined Circular Orbits," *Journal of Spacecrafts and Rockets*, Vol. 3, 1966, pp. 1216–1220.

[7]Hanson, J. H., "Optimal Maneuvers of Orbital Transfer Vehicles," Ph.D. Dissertation, Univ. of Michigan, Ann Arbor, MI, 1983.

[8]Bender, D. F., "Optimum Coplanar Two-Impulse Transfers Between Elliptic Orbits," *Aerospace Engineering*, Vol. 21, Oct. 1962, pp. 44–52.

[9]Gobetz, F. W., and Doll, J. R., "A Survey of Impulsive Transfers," *AIAA Journal*, Vol. 7, May 1969, pp. 801–834.

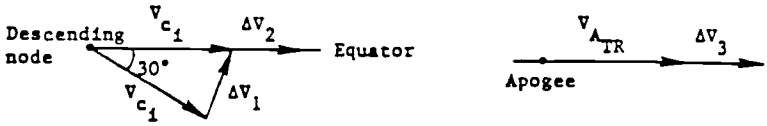
Problems

5.1. Given two circular orbits:

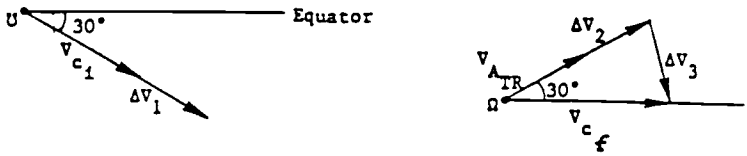
Initial	Final
$r_1 = 6660$ km ($h_1 = 282$ km)	$r_f = 133{,}200$ km
$i = 30$ deg	$i = 0$ (equatorial)

calculate the component and total ΔVs for the following transfer techniques from the initial orbit to the final orbit:

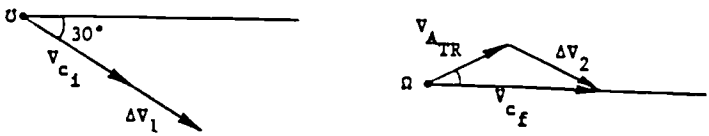
a) Plane change and then Hohmann transfer:



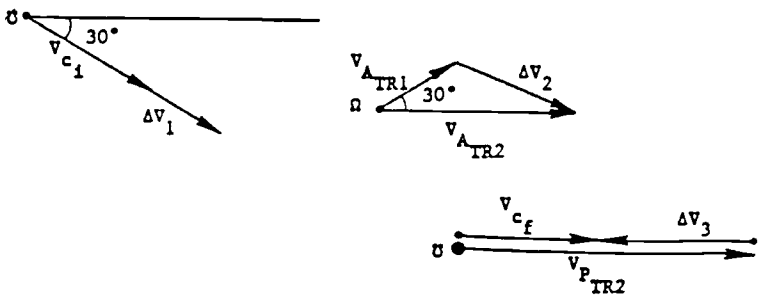
b) Hohmann transfer and then plane change:



c) Hohmann transfer with plane change at apogee in a vectorial combination (two impulses):

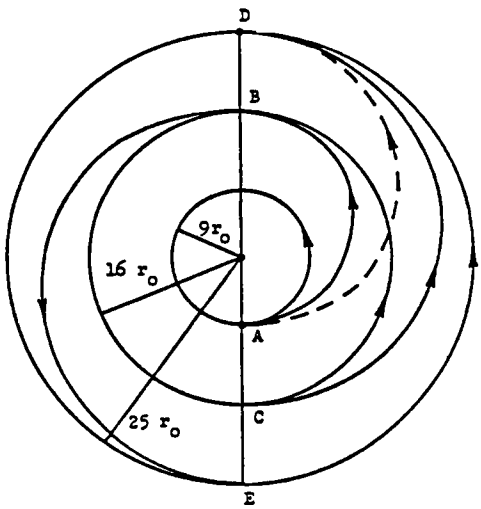


d) Bi-elliptic transfer with vectorial plane change at r_t = 266,400 km (three impulses):



e) Hohmann transfer with optimally split-plane changes (two impulses)

5.2. The sketch illustrates three circular orbits about the Earth. The radii, as shown, are 9, 16, and 25 Earth radii. Determine the characteristic velocity in meters per second. (ΔV_T = sum of ΔV) for a double Hohmann transfer from the inner orbit to the outer orbit ($A - B - C - D$). Calculate ΔV_T in meters per second for a single Hohmann transfer ($A - D$). Finally, determine ΔV_T for an intermediate bi-elliptic transfer ($A - B - E$).



5.3. Given an elliptical orbit whose apogee radius $r_A = 9r_0$ and perigee radius $r_P = 3r_0$ (where r_0 is the radius of the assumed spherical Earth), compute the velocity requirements for two modes of transfer from the surface of the Earth to the ellipse. The first mode is an impulsive launch into a bitangential transfer ellipse that is tangent to the Earth's surface and to the target ellipse at perigee of the target ellipse. At this point, the vehicle impulsively achieves the target orbit. The second mode is via a bitangential transfer ellipse that is tangent to the Earth's surface and the target ellipse at its apogee.

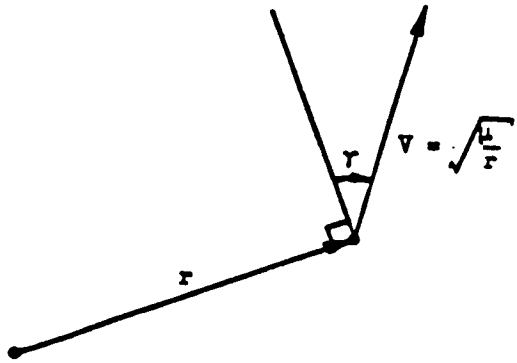
a) Calculate four velocity increments in meters per second.

b) Determine the most economical mode.

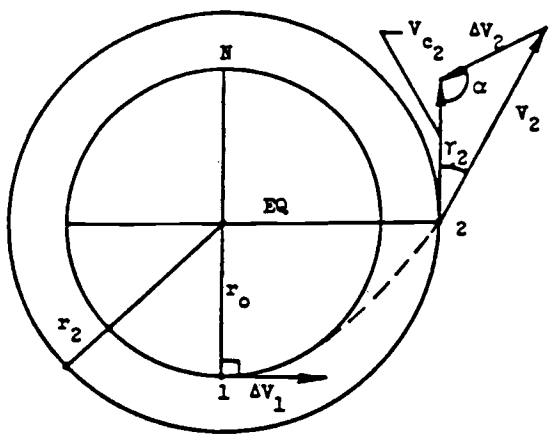
5.4. A satellite is in a circular polar orbit. If, at the ascending node, the velocity vector is rotated counterclockwise 90 deg, what is the new orbit inclination? If the rotation is clockwise 90 deg, what is the new i? If the same rotations occur after the satellite has moved 60 deg and 90 deg from the ascending node, what are the new inclinations?

5.5. Given a set of injection conditions corresponding to the sketch, determine the true anomaly of the injection point as a function only of γ (and perhaps constants), and determine the eccentricity of the resulting orbit as a function only

of γ (and perhaps constants).



5.6. A space vehicle at the South Pole is instantaneously launched, ΔV_1, in a horizontal direction into a parabolic orbit. When the vehicle crosses the equator, point 2, a velocity increment ΔV_2 is applied that instantaneously places the vehicle into a polar, circular orbit. Assuming a spherical Earth of radius $r_0 = 6371$ km, determine the magnitudes of ΔV_1, V_2, V_{c2}, and ΔV_2 in meters per second, and determine the values of γ_2 and α in degrees.

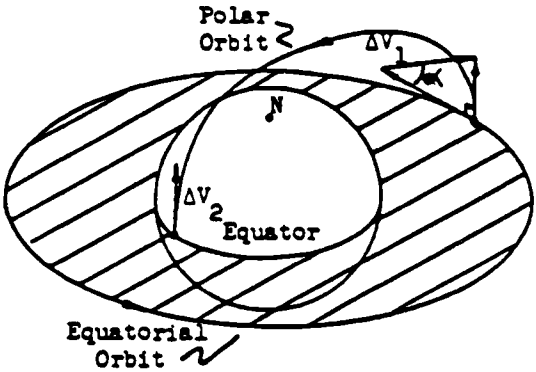


5.7. An astronaut is heading east in a circular equatorial orbit about the Earth at an altitude $h = 3r_0$. At 0° longitude, he applies a velocity increment ΔV_1, which places him in a polar orbit whose perigee grazes the Earth's surface 180 deg away in central angle on the equator.

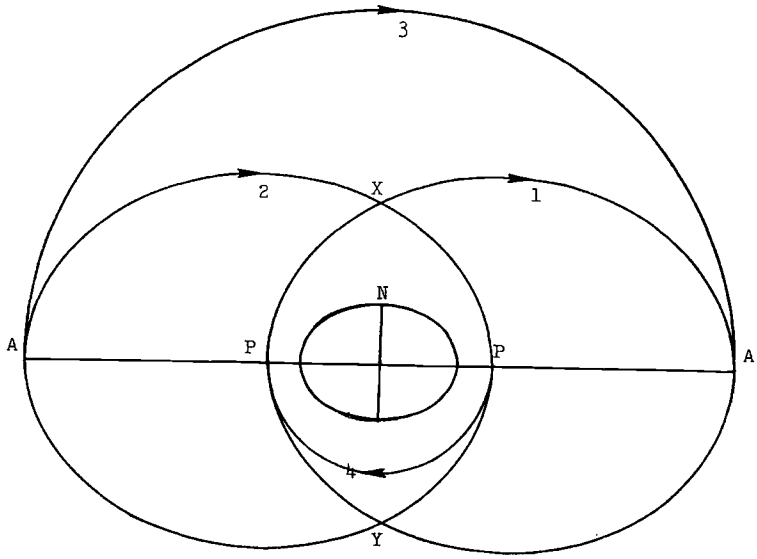
a) What is the magnitude of ΔV_1?

b) What is the angle between ΔV_1 and the original circular orbit velocity?

c) What is the retro velocity increment ΔV_2 at perigee that will reduce his total velocity to zero (soft-land)?

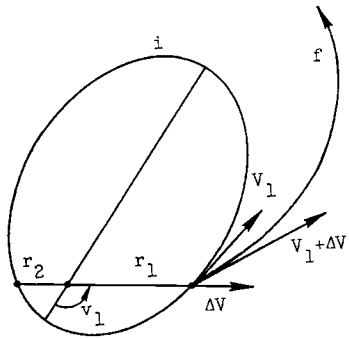


5.8. A satellite is in a polar orbit (orbit 1 on the sketch) about a spherical Earth with no atmosphere. Its perigee and apogee are in the equatorial plane. The perigee altitude is 400 n.mi.; the apogee altitude is 2000 n.mi. Transfer from orbit 2 to orbit 1 can occur in several ways. Determine the total ΔV for transfer via circular orbit 3 from apogee to apogee. Determine the total ΔV for transfer via circular orbit 4 from perigee to perigee. Determine the single ΔV at point X to accomplish the transfer. Would the ΔV at point Y be identical in magnitude? In direction?



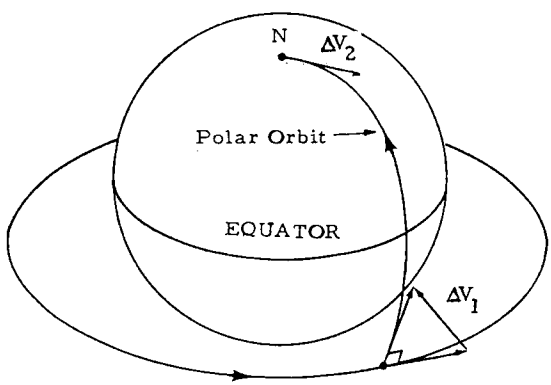
5.9. At an arbitrary point, 1, in an initial orbit i, a velocity increment ΔV is added in the radial direction. A final orbit f is thus achieved. Compare the angular momenta h and the semilatus recta p in the two orbits. Determine the radius in the

final orbit at the point that is 180 deg in central angle away from point 1.



5.10. Boris, a Russian cosmonaut, is in a circular equatorial orbit of radius $r = 1.44r_0$ about the moon (see sketch). He decides to pay a surprise visit to his American friends camped at the North Pole by transferring with ΔV_1 into a polar elliptical orbit whose pericenter is at the camp. When Boris reaches the camp, he retrofires with ΔV_2 to reduce his total velocity to 0. Determine ΔV_1 and ΔV_2 in meters per second. For the moon

$$V_{c0} = \sqrt{\frac{\mu}{r_0}} = 1679 \text{ m/s}$$

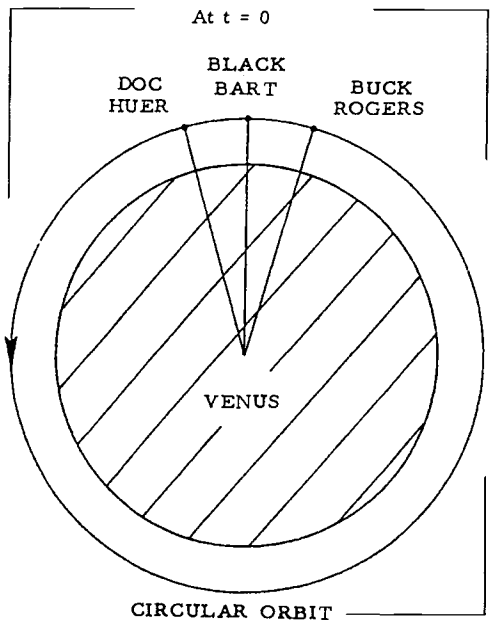


EQUATORIAL ORBIT

5.11. "We've got him surrounded," radios Buck to Doc, referring to Black Bart, intergalactic badman supreme. "We're now all in the same circular orbit that has a period of 90 minutes. I'm 600 km behind him, and you, Doc, are 600 km ahead of him. When I give the word ($t = 0$), we'll both supply circumferential ΔV s to our satellites so that we catch Black Bart after one revolution in our phasing orbits. Since I'll reach Bart before you, I'll have him in handcuffs when you reach us."

Determine the altitude in kilometers of the circular orbit. Calculate the magnitude in meters per second and the direction of ΔV_{BUCK} and ΔV_{DOC}. Determine,

to the nearest kilometer, Buck's minimum altitude during catch-up and Doc's maximum altitude during catch-up. How much time, in minutes, will Buck have to subdue Bart before Doc reaches them? Assume Venus to be spherical and atmosphereless. Assume the radius of Venus to be 6100 km and the circular orbit velocity at the surface of Venus to be 7210 m/s.



Selected Solutions

5.1. a) $\Delta V_1 = 4002$ m/s
$\Delta V_2 = 2939$ m/s
$\Delta V_3 = 1195$ m/s
$\Delta V_T = 8136$ m/s
b) $\Delta V_1 = 2939$ m/s
$\Delta V_2 = 1195$ m/s
$\Delta V_3 = 895$ m/s
$\Delta V_T = 5029$ m/s
c) $\Delta V_1 = 2939$ m/s
$\Delta V_2 = 1295$ m/s
$\Delta V_T = 4234$ m/s
d) $\Delta V_1 = 3068$ m/s
$\Delta V_2 = 776$ m/s
$\Delta V_3 = 267$ m/s
$\Delta V_T = 4112$ m/s

e) $\alpha_{1/\mathrm{OPT}} = 0.7$ deg $\quad \Delta V_1 = 2941$ m/s
$\alpha_{2/\mathrm{OPT}} = 29.3$ deg $\quad \Delta V_2 = 1291$ m/s
$\Delta V_T = 4232$ m/s

5.2. $\Delta V_{ABCD} = 1036$ m/s
$\Delta V_{AD} = 991$ m/s
$\Delta V_{ABE} = 1036$ m/s

5.5. $a = r$
$\cos v = -e$
$e = \sin \gamma$

5.6. $\Delta V_1 = 11{,}180$ m/s
$V_2 = 7905$ m/s
$V_{c2} = 5590$ m/s
$\Delta V_2 = 5590$ m/s
$\gamma_2 = 45$ deg
$\alpha = 90$ deg

5.9. same
same
$r_{2i} = r_{2f}$

5.10. $\Delta V_1 = 2073$ m/s
$\Delta V_2 = 2015$ m/s

5.11. Altitude = 64.2 km
$\Delta V_{\mathrm{BUCK}} = 37.6$ m/s retro
$\Delta V_{\mathrm{DOC}} = 36.5$ m/s forward
Buck's minimum altitude = −63.5 km (underground)
Doc's maximum altitude = 191.2 km
time = 2.8 min if Buck's minimum altitude coincides with a deep canyon.

6
Complications to Impulsive Maneuvers

Having considered one-, two-, and three-impulse ΔV optimal transfers between orbits in Chapter 5, we now turn our attention to the complications of the real world. Are there cases in which four or more impulses would offer significant savings? Do we always want to use the ΔV optimal transfer? How bad is our assumption of impulsive ΔV application, and how does it affect the results? These are the questions that will be investigated in this chapter.

Specifically, the following topics will be considered: 1) N-impulse maneuver, 2) fixed-impulse transfers, 3) finite-duration burns, and 4) very low thrust transfers.

6.1 N-Impulse Maneuvers

In previous chapters, we have seen cases in which a two-impulse maneuver can accomplish the same transfer as a single-impulse but at a considerable ΔV saving (e.g., argument of perigee change). Similarly, cases were studied in which a three-impulse maneuver was best (e.g., bi-elliptic transfers). The obvious next question is: What about four or more impulses?

Edelbaum[1] has answered this question exactly in his paper "How Many Impulses?" The answer, he finds, is that only in rare situations are more than three impulses required to obtain the minimum ΔV transfer.

6.2 Fixed-Impulse Transfers

Introduction

To date, nearly all of the satellites that have been placed in low Earth orbit by the Space Shuttle have achieved their final mission orbit using solid-propellant rocket motors. Whether they are inertial-upper-stage (IUS) or payload-assist-module (PAM) engines, their operation is basically the same. Each solid rocket motor can be viewed simply as a container of solid propellant with a nozzle. An igniter begins the burning of the solid propellant, which continues until all propellant has been consumed. Whereas it is technically feasible to quench the burning of the solid propellant, the typical solid motors in use today are neither stoppable nor restartable.

Since most orbital transfers call for two burns, the typical upper-stage vehicle must consist of two solid motors as depicted in Fig. 6.1. Stage I provides the first burn (ΔV_1) and stage II provides the second burn (ΔV_2).

For a given payload weight and fixed (off-the-shelf) solid rocket motors, the ΔV available from the two motors are fixed by the rocket equation as derived in Chapter 1:

$$\frac{W_i}{W_f} = \exp(\Delta V / g_0 I_{sp}) \tag{6.1}$$

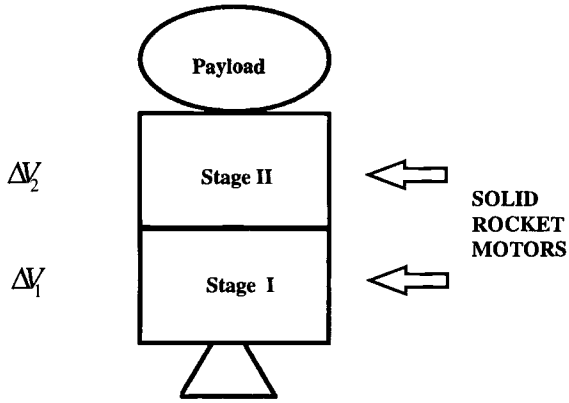


Fig. 6.1 Schematic diagram of a payload atop two solid rocket motors.

where

g_0 = the acceleration of gravity at the Earth's surface
I_{sp} = engine specific impulse
W_i = weight of rocket before burn (initial)
W_f = weight of rocket after burn (final)

Since the solid motors cannot be extinguished or restarted, the ΔV_1 and ΔV_2 supplied by the motors must exactly match the orbital-transfer ΔV requirements. The problem becomes one of finding a transfer between two given orbits that exactly matches the available ΔV_1 and ΔV_2. This problem is often referred to as the "velocity-matching technique" or "fixed-impulse transfers" (referring to the fact that the impulse available from each engine is fixed).

This is clearly a complication to the process of selecting an orbital transfer. In earlier chapters, the orbital transfer was sought that would minimize total ΔV expenditure. The inherent assumption was that the ΔV could be metered out in any size increments by the rocket engine.

Consider the plight of the designer who must assemble a pair of solid motors to form an upper-stage vehicle. In many cases, a range of different payload weights is to be carried, even if it is just to allow for a growth version of a single satellite. In this event, the designer must size the motors for the heaviest payload expected. Having done this, he must find a method to fly the smaller payloads with the same motors. Even the designer who has only a single satellite weight to contend with might be required to select his solid motors from the existing, off-the-shelf array. It is unlikely that he will find motors that will exactly match the ΔV of the optimal transfer. In any event, the designer's problem is the same: how to utilize solid motor stages where either one or both of the stages have excess energy above the optimal-transfer ΔV requirements. The problem is always excess energy because, if the sum of the ΔV provided by the solid motors is less than that of the optimal transfer, no solution is possible.

Several methods have been identified to accommodate this excess energy:

1) Offloading propellant.
2) Adding ballast.
3) Trajectory modification.

The first method, offloading propellants, involves the custom removal of solid propellant from the engine. Propellant is usually removed in the order in which the burn would normally proceed so as to impact the burn characteristics of the motor as little as possible. The disadvantages of this method are that custom propellant offloading costs money and, for large offloads, may require requalifying the motor for flight.

An alternate method consists of adding ballast weights to the appropriate stages to make the ΔV from the motors match the optimal-transfer requirements. A limitation of this approach is that, once enough weight has been added to match the second-stage ΔV to its transfer requirement, the first stage may no longer have enough ΔV. A second drawback is that bolting heavy weights to the payload or motors may invalidate previously performed structural and vibrations tests.

The third method, trajectory modification, is the one we will examine in detail. In trajectory modification, rather than alter the stage ΔV to match the optimal-transfer ΔV, we alter the orbit transfer to match the stage ΔV. The orbit transfer is no longer a ΔV optimal transfer. Physically speaking, the excess energy is being used up via combinations of extra plane change, altitude change, and flight-path-angle change. Usually, there is not a unique solution but rather a range of solutions available to choose from. The following sections will illustrate the trajectory modification method with an example. For more details on this subject, Refs. 2 and 3 are recommended.

Fixed-Impulse Transfer Example

Problem statement. Consider a constellation of satellites that operate in circular orbits with a 12-h period. As shown in Fig. 6.2, the satellite is delivered to low Earth orbit (altitude $h_1 = 278$ km) by the Space Shuttle. From there it is carried to its destination orbit ($h_2 = 20187$ km, inclination $i = 55$ deg) by an upper-stage vehicle consisting of a pair of solid rocket motors.

The arrangement and characteristics of these motors are shown in Fig. 6.3. Using the rocket equation to evaluate the ΔV that are available from each stage

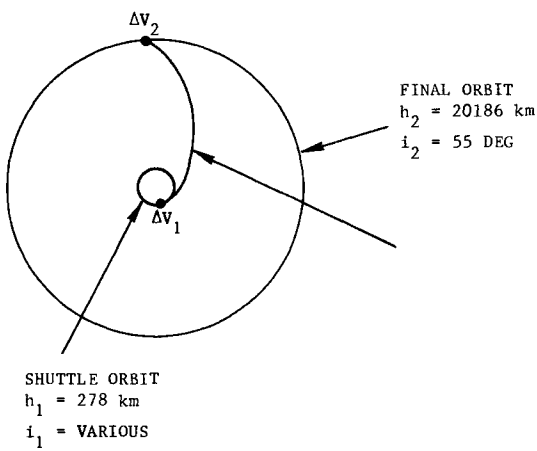


Fig. 6.2 Transfer from Shuttle orbit to final orbit is accomplished by an upper stage (consisting of two motors) attached to the payload.

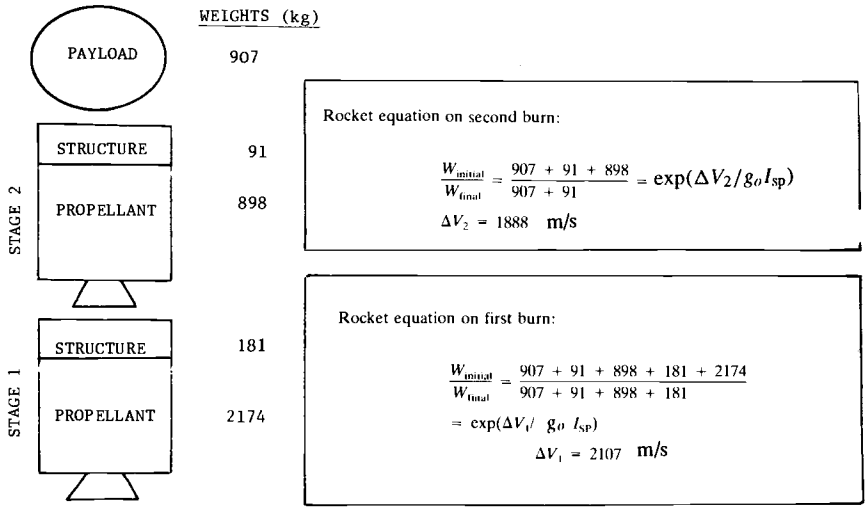


Fig. 6.3 Upper-stage vehicle (I_{SP} for both motors = 300 s; g_o = 9.8066 m/s^2).

yields the results $\Delta V_1 = 2107$ m/s and $\Delta V_2 = 1888$ m/s. Our objective is to find orbit transfers that utilize exactly these amounts of ΔV.

Hohmann transfers. Hohmann-type transfers can still be used if the amount of plane change at both burns is custom-tailored to match the ΔV capabilities of the solid motors. The transfer orbits ΔV_1 and ΔV_2 are derived in the diagram shown in Fig. 6.4 as a function of the amount of plane change (α_1 and α_2) performed on the first and second burns, respectively.

In these equations,

r_1 = radius of Shuttle orbit
r_2 = radius of final orbit
V_{c1} = circular velocity of Shuttle orbit

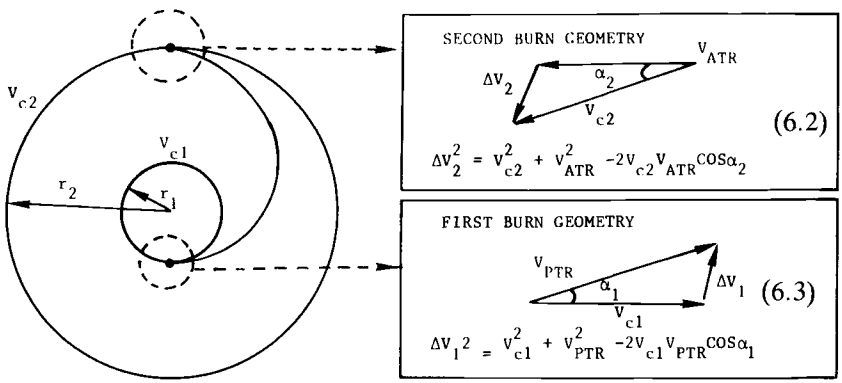


Fig. 6.4 Hohmann transfer with plane change split.

V_{c2} = circular velocity of final orbit
a_{TR} = semimajor axis of transfer orbit
V_{PTR} = transfer-orbit perigee velocity
V_{ATR} = transfer-orbit apogee velocity

For these solid motors, we know

$$\Delta V_1 = 2107 \text{ m/s}$$
$$\Delta V_2 = 1888 \text{ m/s}$$

Solving for the above quantities, we find

$$r_1 = h_1 + r_e = 6656 \text{ km}$$
$$r_2 = h_2 + r_e = 26565 \text{ km}$$
$$r_e = \text{Earth radius}$$
$$V_{c1} = \sqrt{\mu/r_1} = 7739 \text{ m/s}$$
$$V_{c2} = \sqrt{\mu/r_2} = 3874 \text{ m/s}$$
$$a_{TR} = \frac{(r_1 + r_2)}{2} = 16611 \text{ km}$$
$$V_{PTR}^2 = \mu\left(\frac{2}{r_1} - \frac{1}{a_{TR}}\right) \Rightarrow V_{PTR} = 9787 \text{ m/s}$$
$$V_{ATR}^2 = \mu\left(\frac{2}{r_2} - \frac{1}{a_{TR}}\right) \Rightarrow V_{ATR} = 2452 \text{ m/s}$$

Solving the equations in Fig. 6.4 for α_1, α_2, we find

$$\alpha_1 = 3.25 \text{ deg}$$
$$\alpha_2 = 23.25 \text{ deg}$$

Thus, if these specific amounts of plane change are performed as part of a Hohmann ascent, all propellant will be consumed in both motors. What the equations do not specify is the direction in which the plane change is performed. This is free for us to choose. If the two plane changes are performed in the same direction, then the plane change angle θ between the Shuttle orbit and the final orbit must be as follows:

Plane change in same direction:

$$\theta_a = \alpha_1 + \alpha_2 = 3.25 + 23.25 = 26.5 \text{ deg}$$

However, if the two plane changes are performed in opposing directions, then the total plane change angle between the two orbits is only as follows:

Plane changes in opposing directions:

$$\theta_b = |\alpha_1 - \alpha_2| = 23.25 - 3.25 = 20 \text{ deg}$$

If all the plane change is performed as inclination change, then the Shuttle orbit could be placed at any one of four different inclinations i_1:

$$\begin{aligned} i_1 &= i_2 \pm \theta_a = 28.5, 81.5 \\ &= i_2 \pm \theta_b = 35, 75 \end{aligned}$$

In general, however, plane change consists of both inclination change and right ascension of ascending node change ($\Delta\Omega$) according to the relation

$$\cos\theta = \cos i_1 \cos i_2 + \sin i_1 \sin i_2 \cos \Delta\Omega \tag{6.4}$$

Substituting $\theta = \theta_a$ into this equation, varying $28.5 \leq i_1 \leq 81.5$, and solving for $\Delta\Omega_a$ as a function of i_1 will produce all Shuttle orbit orientations ($\Delta\Omega, i$) that are exactly θ_a away from the final orbit in terms of total plane change. This has been done, and the results are plotted in Fig. 6.5. Thus, from each point on the θ_a curve, a Hohmann transfer can be accomplished to the final orbit, and the individual plane changes on each burn ($\alpha_1 = 3.25$ deg, $\alpha_2 = 23.25$ deg) will cause each solid motor to be exhausted exactly. Similarly, substituting $\theta = \theta_b$ in the relation (6.4) yields the second curve in Fig. 6.5. Again, any Shuttle orbit orientation on the θ_b curve, specified as i_1, $\Delta\Omega$ (where $\Delta\Omega = 0$ when the orbit has the same right ascension of ascending node as the final orbit) is a viable orbit in terms of exactly matching the ΔV using a Hohmann transfer. In the case of θ_b, however, the plane changes are in opposing directions. That is, one burn raises the inclination, whereas the other decreases it. Figure 6.5 then contains all possible locations for

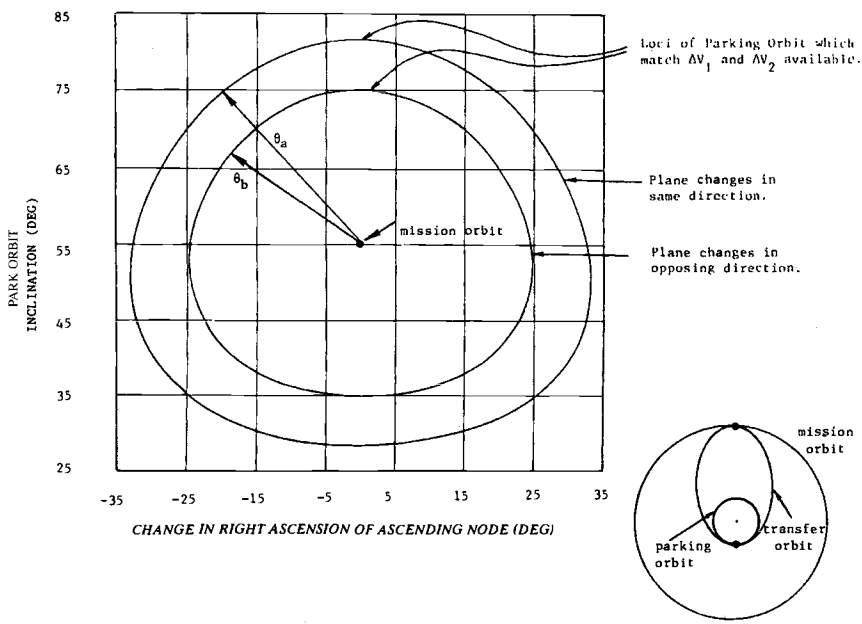


Fig. 6.5 Hohmann flight time transfers.

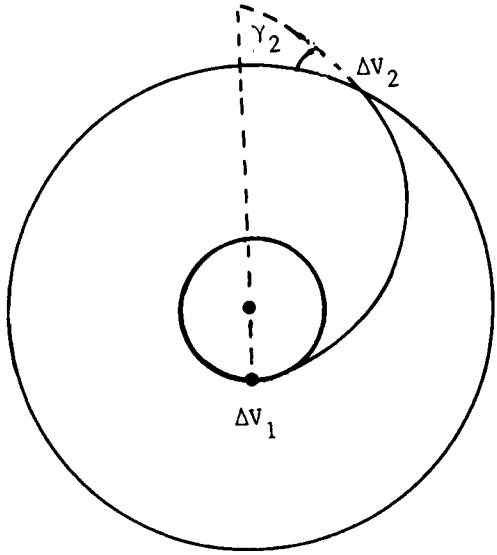


Fig. 6.6 Non-Hohmann transfers.

the Shuttle orbit if the trajectory modification method is to be used to match ΔV and a Hohmann transfer is desired.

Non-Hohmann transfers. But there is nothing sacred about Hohmann transfers. Consider what happens as the transfer-orbit apogee altitude is raised beyond the final orbit altitude. As shown in Fig. 6.6, as the transfer orbit apogee altitude is increased beyond the Hohmann-transfer value, two things happen: 1) α_1 decreases because V_{PTR} increases; and 2) α_2 decreases because an increasing flight-path-angle change Υ_2 must be performed. To reach the higher apogee altitude, the perigee velocity in the transfer orbit (V_{PTR}) must increase. From Eq. (6.3), for a fixed ΔV_1 and V_{cl} as V_{PTR} increases, α_1 must decrease. Similarly, α_2 decreases since an increasing flight-path-angle change Υ_2 must also be performed. The apogee altitude of the transfer orbit can be increased to the point at which either α_1 or α_2 reaches zero. Thus, as transfer-orbit apogee altitude is increased, θ_a decreases, and θ_b approaches θ_a. In the limiting case, θ_b will equal θ_a. (Note that θ_a and θ_b are no longer the simple sum and difference of α_1 and α_2 as a result of different axes of rotation. For a more detailed explanation, see Refs. 2 and 3).

Figure 6.7 shows our example for a non-Hohmann transfer orbit with apogee altitude very near its upper limit. These have been labeled "max/min flight time transfers." The minimum flight time is associated with the ascending crossing of the final orbit, and the maximum flight time is associated with the descending crossing. Again, these curves show, for a specific value of transfer-orbit apogee altitude, all Shuttle orbits from which the final orbit can be reached using $\Delta V_1 =$ 2107 m/s and $\Delta V_2 =$ 1888 m/s. In Fig. 6.5, the transfer-orbit apogee altitude is that of the Hohmann transfer. In Fig. 6.7, the transfer-orbit apogee altitude is very

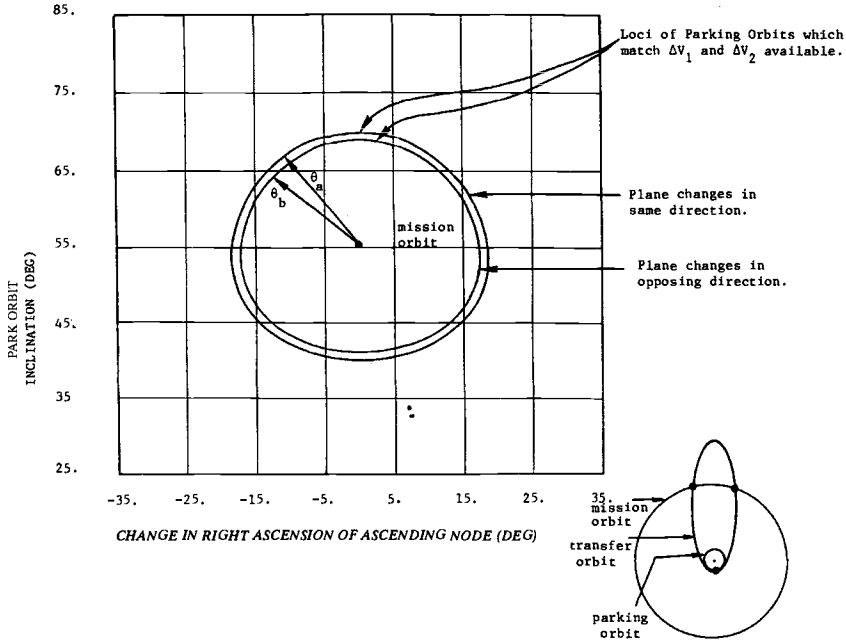


Fig. 6.7 Min/max flight time transfers.

near its maximum allowable value, as evidenced by the closeness of the θ_a and θ_b curves.

Composite trajectory modification results. As the transfer-orbit apogee altitude is raised from the Hohmann value (a lower limit) to the maximum value, the θ_a and θ_b curves traverse the area between Figs. 6.5 and 6.7. The composite result for all possible transfer-orbit apogee altitudes is shown in Fig. 6.8. From a Shuttle orbit located anywhere within the shaded area of Fig. 6.8, at least one transfer exists that satisfies the fixed values of ΔV_1 and ΔV_2 for the final example. Only those points that are on the curves of Fig. 6.5 will have a Hohmann-transfer capability. All others involve non-Hohmann-type transfers. Note that the outer boundary of the region consists of Hohmann transfers, whereas the inner boundary (if any) consists of max/min flight-time solutions. For Shuttle orbits outside the region of Fig. 6.8, there is insufficient ΔV in the solid motors. For Shuttle orbits inside the inner bound of Fig. 6.8, there is too much ΔV in the solid motors for even the trajectory modification method to handle. At a point within the shaded region, usage of the trajectory modification method may yield several choices of transfer orbit that satisfy the ΔV constraints.

Launch Window

It would seem that by supplying excess energy above and beyond the ΔV optimal transfer, we should be receiving some benefit in return. In fact, we do receive a benefit called a launch window, that is, a duration of time within which a ground launch to the Shuttle orbit is possible.

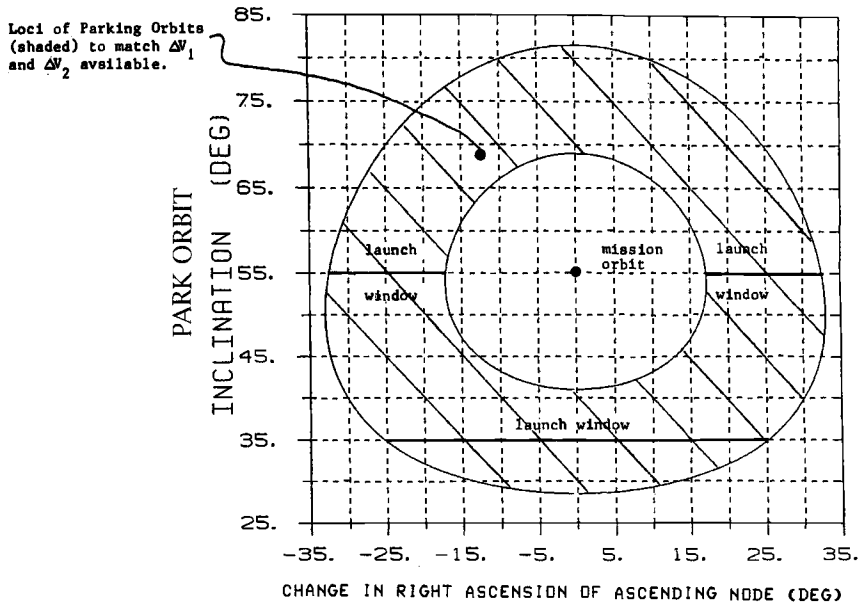


Fig. 6.8 Region of allowable orbits. Mission orbit: h = 20,186 km, i = 55 deg, Ω = 0 deg; parking orbit: h = 278 km, ΔV_1 = 2107 m/s, ΔV_2 = 1888 m/s.

Because the Earth rotates eastward at approximately 15 deg/h (360 deg/23 h, 56 min, 4.09 s), if a launch takes place t hours late, then the right ascension of the ascending node of the orbit will be

$$\Omega = \Omega_0 + 15t$$

where

Ω = actual value of right ascension of ascending node attained
Ω_0 = target value of right ascension of ascending node

The effect of an early or late launch on orbit-transfer geometry is shown in Fig. 6.9. The acceptable amount of time by which a launch may be early or late is called the launch window. For our example, if the Shuttle orbit is located at $i = 55$ deg, the launch window is shown on Fig. 6.8. In terms of $\Delta\Omega$ and time (where $t = 0$ is the time the final orbit plane passes over the launch site), the launch window extends from

$\Delta\Omega$ (deg)	-32 to -17,	17 to 32
t(h)	-2.13 to -1.13	1.13 to 2.13

Note that, because of excess energy constraints, the region around the conodal launch time ($t = 0$) is not part of the launch window. (Conodal launch time is when shuttle and final orbit Ω are the same, $\Delta\Omega = 0$.)

Similarly, for a Shuttle orbit at $i_1 = 35$ deg, the launch window is $t = 0 \pm 1.73$ h. In this case, the conodal launch time is within the launch window.

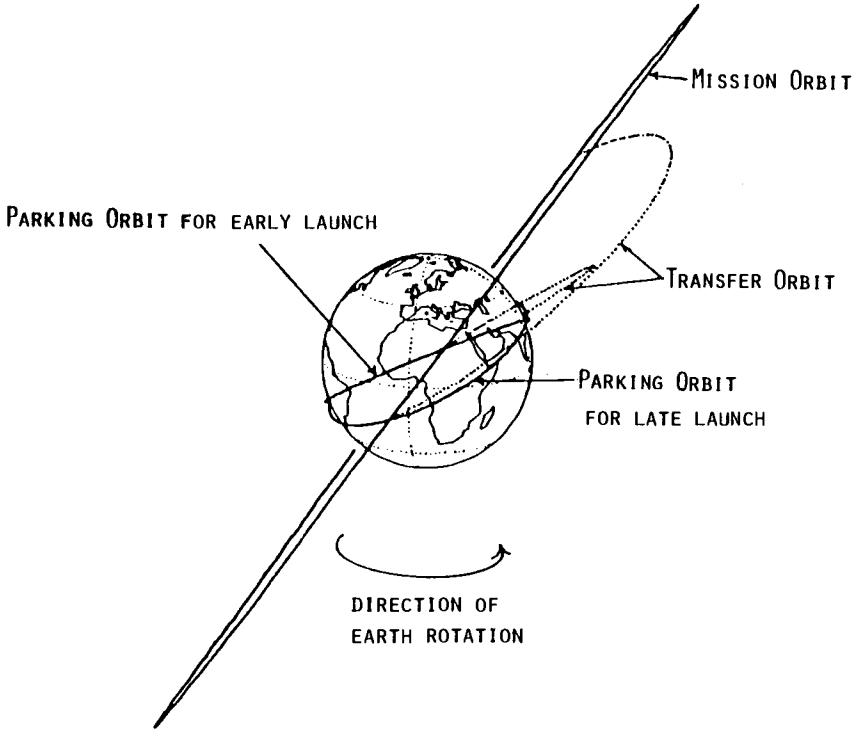


Fig. 6.9 Effect of launch time on orbit geometry.

For a Shuttle orbit at $i_1 = 28.5$ deg, the launch window is simply an instant at $t = 0$ when the final orbit plane is directly above the launch site.

In all cases, as long as the Shuttle launch takes place within the allowable launch window, a transfer trajectory can be found that will exactly match the ΔV available from the solid motors.

6.3 Finite-Duration Burns: Gravity Losses

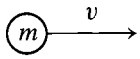
Introduction

All our work to this point has assumed "impulsive maneuvers"; that is, the required ΔV are applied instantaneously. To do this would require a propulsion source of infinite thrust. In reality, the ratio of thrust to weight (T/W) is generally in the range $10^{-5} < T/W < 10$. This finite thrust to weigh ratio complicates our previous study of impulsive maneuvers by introducing a ΔV penalty, the so-called gravity loss.

Gravity Losses

Consider a mass m traveling at velocity v. At some point in time, it ejects a mass dm backward at a relative velocity u.

Before burn:



After burn:

It should be noted that the specific impulse of a rocket engine is a measure of the exhaust velocity of the particles according to the relation

$$u = g_0 I_{sp}$$

The momentum balance equation for this exchange was derived in Eqs. (1.5–1.8) as

$$m\frac{dv}{dt} = -u\frac{dm}{dt} + F$$

where F is the sum of all external forces acting on the masses.

The external forces acting on a launch vehicle are shown in Fig. 6.10. In this simple model, the thrust T and the drag D act along the axis of the vehicle while the weight W of the vehicle acts in the downward direction. The flight-path angle of the vehicle has been labeled Υ. Making the appropriate substitutions in the momentum equation yields

$$m\frac{dv}{dt} = -u\frac{dm}{dt} - D - mg\sin\Upsilon$$

Multiplying by dt/m, we obtain

$$dv = -u\frac{dm}{m} - \frac{D}{m}dt - g\sin\Upsilon\, dt$$

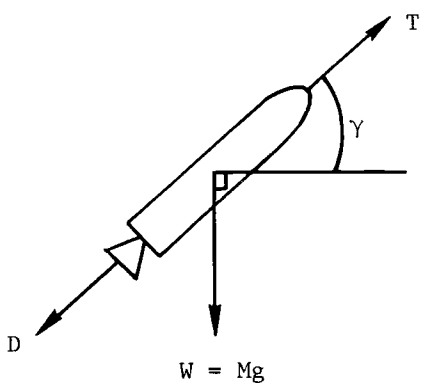


Fig. 6.10 External forces acting on launh vehicle.

Integrating from t_1 to t_2 (duration of the burn) yields*

$$\int_{v_1}^{v_2} \mathrm{d}v \equiv \Delta V = -u \ln m \Big|_{M_1}^{M_2} - \int_{t_1}^{t_2} \frac{D}{m}\,\mathrm{d}t - \int_{t_1}^{t_2} g \sin \Upsilon \,\mathrm{d}t$$

or

$$\Delta V = g_0 I_{\mathrm{sp}} \ln \frac{m_i}{m_f} - \int_{t_1}^{t_2} \frac{D}{m}\,\mathrm{d}t - \int_{t_1}^{t_2} g \sin \Upsilon \,\mathrm{d}t \tag{6.5}$$

where

m_i = mass before burn (initial)
m_f = mass after burn (final)

The result is the familiar rocket equation [Eq. (1.10)], plus two complicating terms. In both cases, these terms detract from the rocket equation and reduce the available characteristic ΔV from the rocket. The first term is related to the drag force D and is not of interest here. In subsequent discussion, we shall assume $D = 0$.

The second term, the gravity-loss term, is the one we are interested in. We cannot easily integrate this term since both g (the gravitational acceleration at the rocket's location) and Υ (the flight-path angle) are unspecified functions of the time (which depend on the trajectory flown). We can, however, analyze this term with the intent of driving it toward zero.

Clearly, as t_1 approaches t_2, this gravity-loss term approaches zero. In the limit with $t_1 = t_2$, the burn is impulsive, and there are no gravity losses. The parameter g, we have no control over, except to note that burns at low altitude (high g) will be inherently more expensive than burns at high altitude (low g). We can control the flight-path angle Υ during the burn and, if it could be kept to zero, there would be no gravity losses. A finite-duration burn is shown in Fig. 6.11 with dashed lines. Typically, the burn surrounds the impulsive-transfer location (shown with solid lines), and the finite-burn direction is very close to the impulsive direction. Recall that, for Hohmann transfers, the flight-path angle is zero for both burns. If the impulsive-burn direction is to be maintained during the finite burn, the initial and final values of flight-path angle Υ cannot be zero. In fact, the initial values of Υ are approximately

$$-\Upsilon_1 \approx \Upsilon_2 \approx \phi_1 \approx \phi_2$$

so that the range of Υ is determined directly by the central angle through which the burn occurs. Near the midpoint of the burn arc, $\Upsilon \approx 0$ as desired but, at the ends, the $\sin \Upsilon$ contribution can be significant.

An alternative would be to vary the thrust direction during the burn so that, at any instant of time, $\Upsilon = 0$. By definition, the gravity-loss term would be zero in

*Note that

$$T = -u\frac{\mathrm{d}m}{\mathrm{d}t} = -g_0 I_{\mathrm{sp}} \frac{\mathrm{d}m}{\mathrm{d}t} = -I_{\mathrm{sp}} \frac{\mathrm{d}W}{\mathrm{d}t}$$

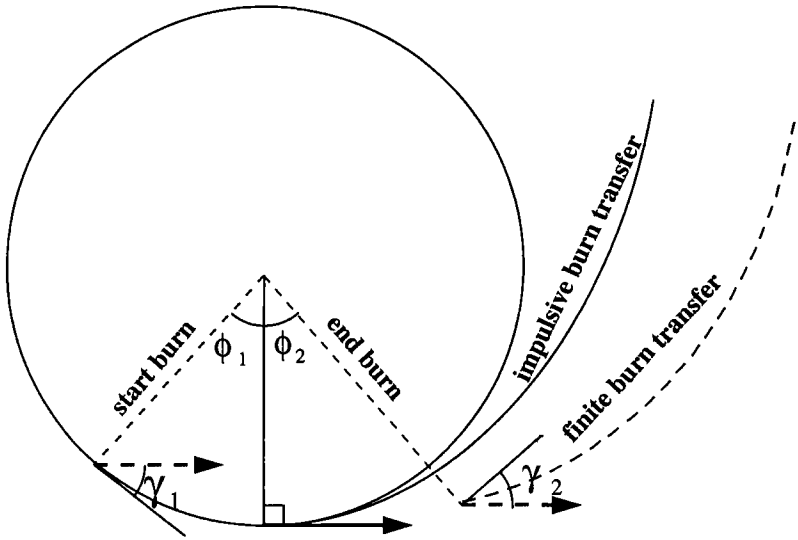


Fig. 6.11 Comparison of impulsive- and finite-burn transfers: solid lines = impulsive-burn characteristics; dashed lines = finite-burn characteristics.

Eq. (6.5). The fallacy in this approach is that the spacecraft would no longer follow the optimal-thrust direction so that the characteristic ΔV in Eq. (6.5) would be increased. One way or another, the penalty must be paid.

The factors that lead to high gravity losses are, then, long burn duration (low thrust/weight), high g (low altitude), and high Υ (large burn central angle). These gravity losses must be added to the impulsive ΔV to arrive at an effective ΔV to accomplish a maneuver. Figure 6.12 shows the total ΔV required to ascend from a low Earth orbit, LEO ($h = 278$ km, $i = 28.5$ deg) to a geosynchronous equatorial orbit, GEO ($h = 35786$ km, $i = 0$ deg) as a function of the initial thrust/weight ratio. For high thrust/weight ratios, the total ΔV approaches the impulsive ΔV (4237 m/s). For low thrust/weight ratios, the gravity loss is obviously significant (over 2200 m/s for $T/W_0 = 10^{-4}$).

Calculating Gravity Losses

For the purpose of determining gravity losses, the problem can be roughly divided into three categories.

High thrust ($T/W_0 \approx (0.5$ *to* $1.0)$). Here the thrust is the dominant force. Gravity losses can be estimated by using the methods of Robbins[4] or even neglected (impulsive assumption).

Low thrust ($T/W_0 \approx 10^{-2}$ *to* 10^{-1}). Here the thrust and gravitational forces are both important, and assumptions or estimates are difficult. The best method of solution is to run an integrated trajectory computer program.

Very low thrust ($T/W_0 \approx 10^{-5}$). Here the thrust may be considered as a perturbation to the trajectory. An orbit transfer consists of many orbit revolutions

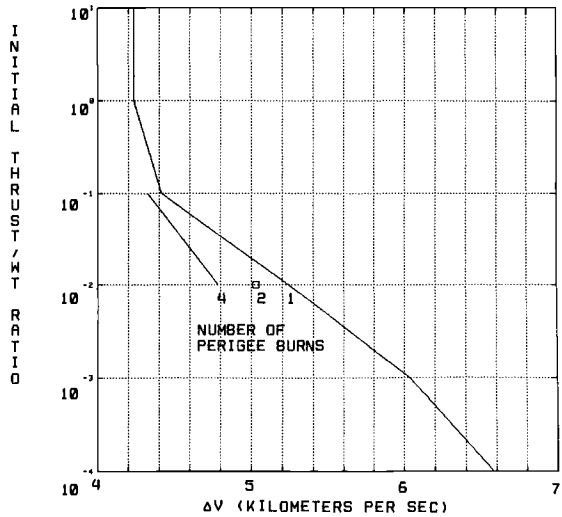


Fig. 6.12 ΔV requirements to transfer from LEO to GEO.

with continuous thrusting. A first-order analytic solution for this case will be described in the next section.

Reducing Gravity Losses

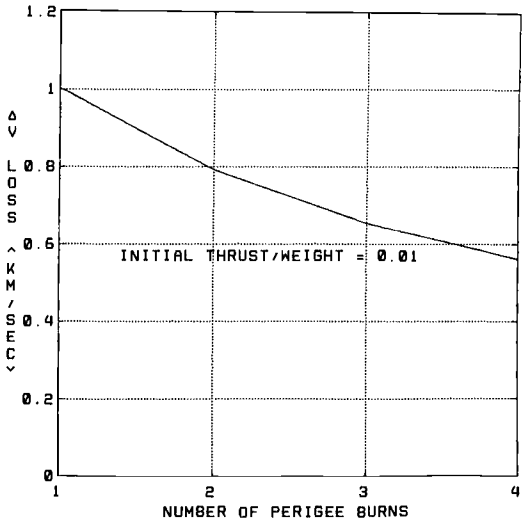
The obvious method for reducing gravity losses is to install an engine with a higher thrust. There is, however, another method. A single large (in terms of central angle traversed) burn can be broken up into a series of smaller burns, separated by one or more revolutions. Consider the transfer of Fig. 6.11. Rather than perform a single large burn over the central angle range $\phi_1 + \phi_2$, the burn could be split into two parts. In the first part, roughly half the total ΔV_1 would be applied, over roughly half the previous central angle range. After a full revolution, the vehicle would return to very nearly the same perigee location, where the remainder of the ΔV_1 would be applied. The net effect is that the central-angle travel (and the corresponding bounds on Υ) has been roughly halved. In the gravity-loss terms of Eq. (6.5), this means lower gravity losses. Breaking the burn into more and more pieces, each of which more closely resembles the impulsive case, leads to lower and lower gravity losses.

For the geosynchronous transfer of Fig. 6.12, some cases are plotted in which the large perigee burn has been broken into two or four pieces. The ΔV saving is substantial. The penalty paid for this is in the form of transfer time. The tradeoff between ΔV savings and transfer time for the case of $T/W_0 = 0.01$ is shown in Fig. 6.13 for the geosynchronous transfer. As the number of perigee burns increases, the gravity loss decreases, but the transfer time increases.

6.4 Very Low Thrust Transfers

If the thrust is small compared to the gravitational force ($T/W_0 \approx 10^{-5}$), then the resulting transfer orbit is a slow spiral outward under continuous thrust. That each of the many revolutions is nearly circular allows certain simplifying assumptions,

a)



b)

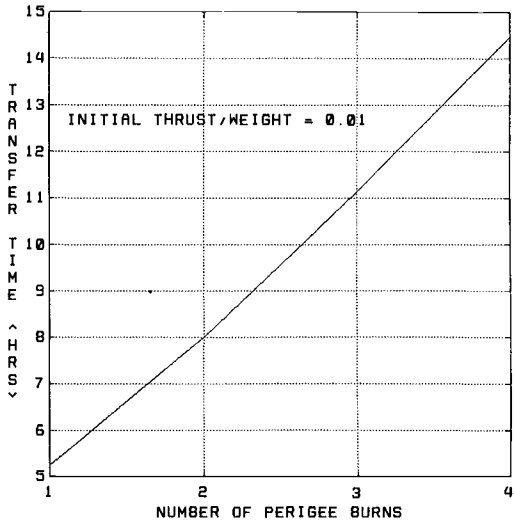


Fig. 6.13 a) ΔV loss as a function of number of perigee burns: b) transfer time as a function of number of perigee burns.

which have been used by Edelbaum[5] to develop a first-order analytic solution to the problem of very low thrust transfers between circular orbits.

The given quantities are

V_0 = initial orbit circular velocity
V_1 = final orbit circular velocity
α = plane change angle between orbits
T = engine thrust

I_{sp} = engine specific impulse
W_1 = payload weight delivered to final orbit

The first step is to compute the ΔV required to make the transfer. This is given by Edelbaum as

$$\Delta V = \sqrt{V_0^2 - 2V_0V_1 \cos\frac{\pi}{2}\alpha + V_1^2}$$

Then, use the rocket equation to find the required weight in the initial orbit (W_0)

$$W_0 = W_1 \exp(\Delta V/g_0 I_{sp})$$

where g_0 is the gravitational acceleration at the Earth's surface. The propellant required for the transfer is simply

$$\Delta W = W_0 - W_1$$

The transfer time is, then,

$$\Delta t = \Delta W/\dot{W}$$

where

$$\dot{W} = \text{thrust}/I_{sp}$$

Using these equations, a first-order estimate of fuel and time requirements for a very low thrust upper-stage vehicle can be obtained. A more complete treatment of this problem is given in Chapter 14.

References

[1]Edelbaum, T. N., "How Many Impulses?," *Aeronautics and Astronautics,* Nov. 1967.

[2]Chu, S. T., Lang, T. J., and Winn, B. E., "A Velocity Matching Technique for Three Dimensional Orbit Transfer in Conceptual Mission Design," *Journal of the Astronautical Sciences,* Vol. 26, Oct.–Dec. 1978, pp. 343–368.

[3]Der, G. J., "Velocity Matching Technique Revisited," AAS/AIAA Astrodynamics Specialist Conference, Lake Tahoe, NV, Aug. 3–5, 1981.

[4]Robbins, H. M., "Analytical Study of the Impulsive Approximation," *AIAA Journal,* Vol. 4, Aug. 1966, pp. 1423–1477.

[5]Edelbaum, T. N., "Propulsion Requirements for Controllable Satellites," *ARS Journal,* Vol. 31, Aug. 1961, pp. 1079–1089.

Problems

6.1. For the fixed-impulse transfer example in the text, it is desired to increase the satellite weight from 907 to 990 kg. If this heavier satellite is placed atop the same upper-stage vehicle

a) Calculate the available ΔV_1 and ΔV_2.

b) Using these ΔV in a Hohmann-type transfer, what plane changes α_1 and α_2 must be performed?

c) What is the locus of Shuttle orbits (inclination and node) from which these fixed-impulse Hohmann-type, transfers can be achieved? Sketch in the locus onto Fig. 6.5.

d) Are non-Hohmann transfers available? Do they increase the allowable locus of Shuttle parking orbits?

e) Briefly describe the launch window available for a Shuttle parking orbit at $i = 55$ deg.

6.2. Instead of using two solid rocket motors to propel the satellite from the Shuttle parking orbit to the final orbit, let us now consider a very low thrust ion-propulsion engine:

$$\text{Thrust} = 4.45 \text{ N}$$
$$I_{sp} = 3000 \text{ s}$$

where 1 N = 1 kg m/s^2 and the weight of 1 kg is $W = mg_0 = 1 \text{ kg} \times 9.8066$ m/s^2 = 9.8066 N

The payload weight to be placed on orbit now weighs 1361 kg and consists of the satellite, the ion engine, and empty propellant tanks.

a) For a Shuttle park orbit at $i = 55$ deg, what is the minimum plane change that must be performed by the ion upper stage?

b) How much ΔV is required from the ion engine?

c) How much gravity loss does this represent? (See example for optimal impulsive ΔV.)

d) How much propellant must the ion upper stage carry?

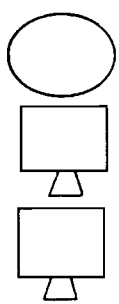
e) How long will the transfer take?

6.3. A payload of space experiments will be delivered to circular low Earth orbit (296-km altitude) by the Space Shuttle. Its final destination is a 741-km altitude circular orbit at as high an inclination as possible (preferably above $i = 72$ deg). Because of range safety constraints, the highest inclination available for the Shuttle orbit is $i = 57$ deg. The transfer from the Shuttle orbit to the destination orbit will be accomplished using two identical solid-propellant motors. These motors are not restartable, so that one motor will be consumed in leaving the Shuttle orbit and the second will be consumed in entering the destination orbit. The characteristics of the motors are as follows:

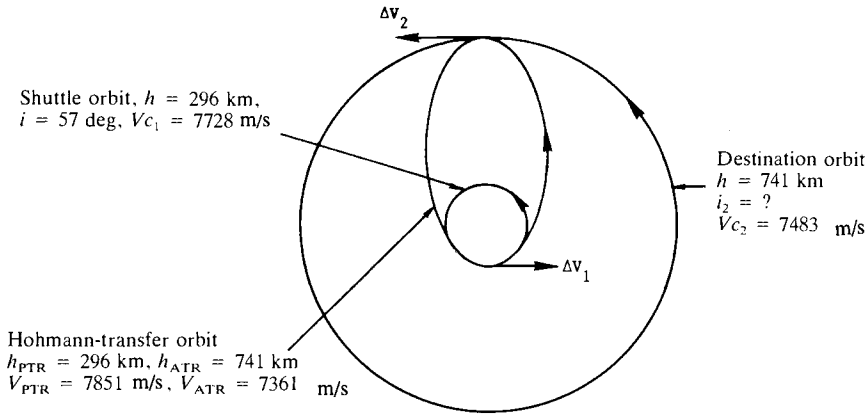
Payload = 1452 kg

Solid motor #2
- propellant wt = 1044 kg
- structure wt = 136 kg
- $I_{sp} = 285$ s

Solid motor #1
- propellant wt = 1044 kg
- structure wt = 136 kg
- $I_{sp} = 285$ s



The required orbital transfer (a Hohmann-type transfer will be used) is summarized in the following sketch:



For this orbital-transfer situation

a) What ΔV (ΔV_1 and ΔV_2) are available from the two solid motors?

b) How much plane change (α_1 and α_2) can be provided by each of the solid motors when flown on this transfer?

c) What is the maximum value of inclination for the destination orbit using these motors? Have the experimenters met their goal of $i \geq 72$ deg?

d) What is the transfer-orbit inclination for this case, and at what latitude must the two burns take place?

Selected Solutions

6.1. a) $\Delta V_1 = 2048$ m/s
$\Delta V_2 = 1779$ m/s
b) $\alpha_1 \approx 0$
$\alpha_2 = 19.97$ deg
c) $\alpha_2 \approx 19.97$ deg

6.2. a) $\Delta i = 0$
b) $\Delta V = 3865$ m/s
c) 395 m/s
d) $m_p = 191$ kg
e) $\Delta t = 14.7$ days

6.3. a) $\Delta V_1 = 894$ m/s
$\Delta V_2 = 1412$ m/s
b) $\alpha_1 = 6.52$ deg
$\alpha_2 = 10.87$ deg
c) $i_{\max} > 72$ deg
d) $i_{tr} = 57 + 6.52 = 63.52$ deg

7
Relative Motion in Orbit

7.1 Space Rendezvous

Rendezvous in space between two satellites is accomplished when both satellites attain the same position and velocity, both vectors, at the same time. However, at the time a rendezvous sequence is initiated, the two satellites may be far apart in significantly different orbits. In fact, one satellite may be starting with a launch from the ground.

This chapter will address the rendezvous sequence in two parts. The first part will be concerned with phasing for rendezvous, i.e., developing the maneuvers and timing sequence that will bring the two satellites into close proximity. The material presented in the sections dealing with Hohmann and bi-elliptic transfer is based on the approach presented in Ref. 1. The second part, terminal rendezvous, will examine the motion of one satellite with respect to the other in a coordinate frame attached to one of the satellites. Relative motion between the satellites and terminal maneuvers required for docking will be examined.

Phasing for Rendezvous

Hohmann transfer. The requirements for rendezvous between two satellites in circular coplanar orbits are both illustrative and operationally useful. Figure 7.1 presents a sketch of two circular orbits with radii r_i and r_f. Assume the satellite in the inner orbit to be the active rendezvous satellite, i.e., the maneuvering satellite. The satellite in the outer orbit is the passive target satellite, i.e., nonmaneuvering. Further, assume that, at some instant in time, the rendezvous satellite is located at the point shown in Fig. 7.1 and the target satellite is located ahead, i.e., in the direction of motion, in its orbit by an amount equal to the central angle θ_H.

Now, assume that the rendezvous satellite initiates a Hohmann transfer in order to rendezvous with the target satellite at the rendezvous point. If travel times for the two satellites are equated,

$$\frac{P_{\text{tr}}}{2} = \frac{\pi - \theta_H}{2\pi} P_f \tag{7.1}$$

where

P_{tr} = the orbital period of the Hohmann-transfer ellipse
P_f = the period of the target satellite orbit

Substituting for both periods,

$$\frac{\pi}{\sqrt{\mu}}\left(\frac{r_i + r_f}{2}\right)^{3/2} = \frac{(\pi - \theta_H)}{2\pi}\left(\frac{2\pi r_f^{3/2}}{\sqrt{\mu}}\right) \tag{7.2}$$

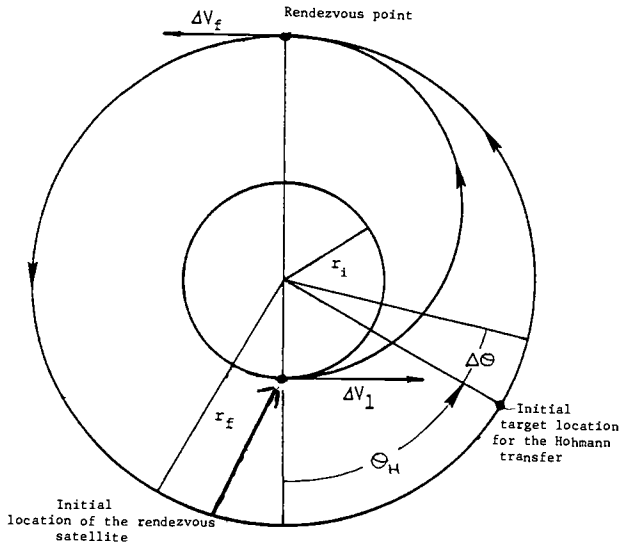


Fig. 7.1 Rendezvous via Hohmann transfer.

Reducing and solving for θ_H,

$$\theta_H = \pi\left[1 - \left(\frac{1 + r_i/r_f}{2}\right)^{3/2}\right] \tag{7.3}$$

The range of θ_H is

$$0 \le \theta_H \le \pi\left[1 - \left(\frac{1}{2}\right)^{3/2}\right] = 0.64645\pi = 116.36 \text{ deg}$$

Figure 7.2 presents θ_H as a function of the final orbit altitude h_f for an initial orbit altitude h_i of 100 n.mi. (185.2 km).

If the initial lead angle of the target satellite with respect to the rendezvous satellite is not θ_H but, instead, is $\theta_H + \Delta\theta$, then the Hohmann transfer cannot be initiated immediately. If it were initiated immediately, then the target satellite would be located an angle $\Delta\theta$ beyond the rendezvous point when the rendezvous satellite reached the rendezvous point. And so the initiation of the Hohmann transfer must wait until the phase angle reduces to θ_H. This will occur naturally because the angular velocity of the inner orbit ω_i is higher than the angular velocity of the outer orbit ω_f.

In time t, angular displacements of the two satellites will be $\theta_i = \omega_i t$ and $\theta_f = \omega_f t$. Therefore, $\Delta\theta = \theta_i - \theta_f = (\omega_i - \omega_f)t_w$, where t_w is the waiting time to achieve a phasing angle change $\Delta\theta$. The maximum value of t_w is the synodic period P_s when $\Delta\theta = 2\pi$

$$P_s = \frac{2\pi}{\omega_i - \omega_f} = \frac{2\pi}{(2\pi/P_i) - (2\pi/P_f)} \tag{7.4}$$

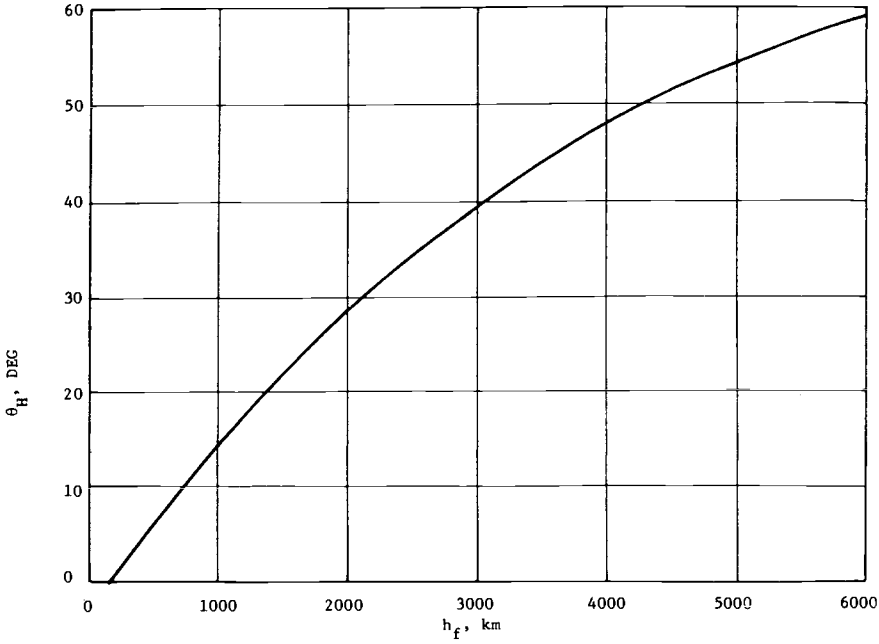


Fig. 7.2 Phase angle for rendezvous via a Hohmann transfer from a 185-km orbit.

or

$$\frac{1}{P_s} = \frac{1}{P_i} - \frac{1}{P_f} \tag{7.5}$$

Figure 7.3 presents P_s vs h_f for $h_i = 100$ n.mi. (185.2 km). Note that, if $h_f = 120$ n.mi. (222.2 km), the synodic period is approximately 10,000 min, or about one week. The possibility of such long waiting times will be circumvented in the next section by using bi-elliptic and semitangential transfers instead of a Hohmann transfer.

For the Hohmann-transfer technique, the total time for rendezvous, t, is the sum of the Hohmann-transfer time t_H and the waiting time t_w,

$$t = t_H + t_W = t_H + \frac{\Delta\theta}{\omega_i - \omega_f} = t_H + \frac{\Delta\theta}{2\pi} P_s \tag{7.6}$$

When the second ΔV of the Hohmann transfer is applied by the rendezvous satellite, both satellites will have the same velocity at the rendezvous point at the same time, and rendezvous will be accomplished.

Bi-elliptic transfer. In Chapter 5, it was concluded that, in terms of ΔV, the bi-elliptic transfer is not significantly better than the Hohmann transfer. However, for rendezvous, the bi-elliptic transfer will be shown to have utility in the case for which the Hohmann transfer is weakest, i.e., for waiting times approaching the synodic period.

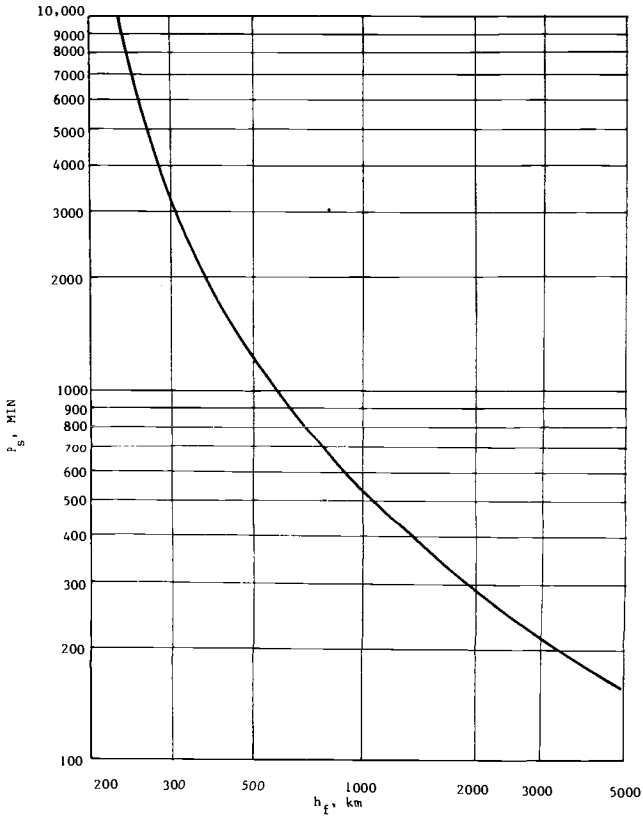


Fig. 7.3 Synodic period for a 185-km inner orbit.

Figure 7.4 presents a sketch of the bi-elliptic transfer previously discussed in Chapter 5. In this case, rendezvous will occur at the rendezvous point after the application of ΔV_3.

Assume the target satellite initially at an angle $(\theta_H + \Delta\theta)$ ahead of the rendezvous satellite. Since the radius r_t is assumed to be greater than r_f, the target satellite must first traverse $2\pi - \Delta\theta - \theta_H$, and then 2π, in order to reach the rendezvous point at the same time as the rendezvous satellite. This total time is

$$t = \frac{2\pi - \Delta\theta}{\omega_f} + \frac{2\pi - \theta_H}{\omega_f} \tag{7.7}$$

But

$$\frac{\pi}{\omega_f} = \frac{P_f}{2}, \qquad \frac{\pi - \theta_H}{2\pi} P_f = t_H$$

from Eq. (7.1), so that

$$t = \frac{2\pi - \Delta\theta}{2\pi} P_f + t_H + \frac{P_f}{2}$$

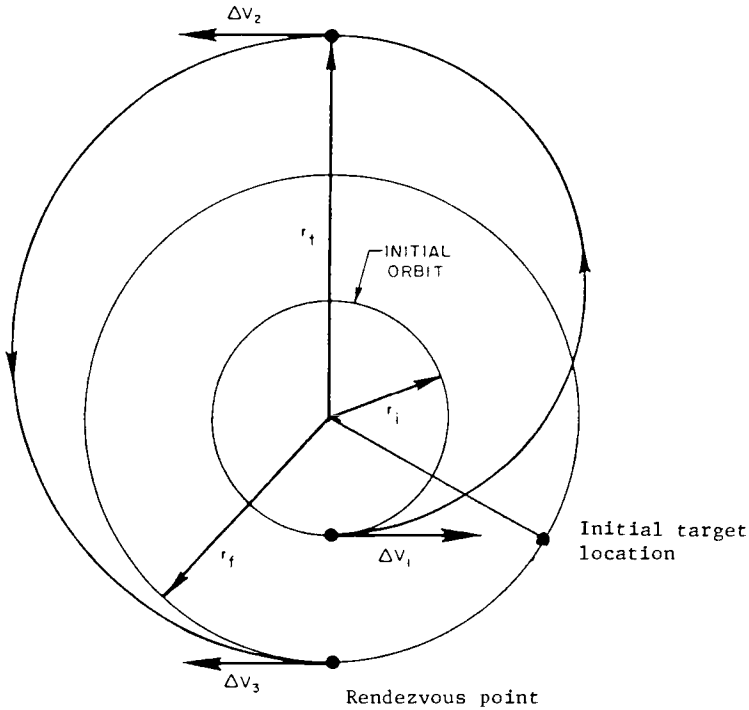


Fig. 7.4 Rendezvous via the bi-elliptic transfer.

and

$$t = t_H + \frac{P_f}{2}\left(3 - \frac{\Delta\theta}{\pi}\right) \tag{7.8}$$

At $\Delta\theta = 0$,

$$t = t_H + \frac{3}{2}P_f$$

And at $\Delta\theta = 2\pi$,

$$t = t_H + \frac{P_f}{2}$$

Thus, because the bi-elliptic transfer occurs mostly beyond the outer circular orbit, it easily accommodates a $\Delta\theta$ that is slightly less than 2π. In this case, r_t will be only slightly larger than r_f.

In all cases, the value of r_t is determined by the value of $\Delta\theta$ because the time spent by the rendezvous satellite in the elliptic-orbit transfer legs is

$$t = \frac{\pi}{\sqrt{\mu}}\left[\left(\frac{r_i + r_t}{2}\right)^{3/2} + \left(\frac{r_t + r_f}{2}\right)^{3/2}\right] \tag{7.9}$$

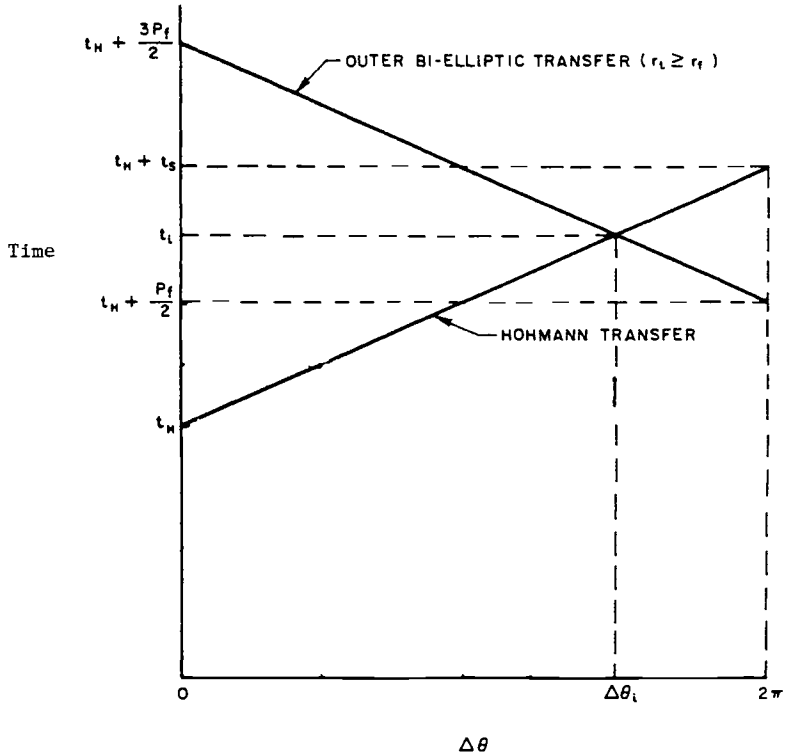


Fig. 7.5 Total time vs phase angle for the Hohmann and bi-elliptic transfers.

Figure 7.5 presents total time vs $\Delta\theta$. The Hohmann line is specified by Eq. (7.6). The bi-elliptic line is specified by Eq. (7.8). Note the difference in slope. The total time for the bi-elliptic transfer is a minimum at $\Delta\theta = 2\pi$. For smaller values of $\Delta\theta$, the total bi-elliptic time becomes longer. Therefore, there is no advantage to be gained by waiting because waiting reduces $\Delta\theta$, which increases the total bi-elliptic time.

Figure 7.5 depicts an intersection of the Hohmann and bi-elliptic lines. An intersection will exist only if $P_s \geq P_f/2$. Substituting for P_s,

$$\frac{P_i P_f}{P_f - P_i} \geq \frac{P_f}{2} \tag{7.10}$$

or

$$3P_i \geq P_f \tag{7.11}$$

Since period $P = 2\pi r^{3/2}/\sqrt{\mu}$, then,

$$r_f \leq 3^{2/3} r_i \tag{7.12}$$

Since $3^{2/3}$ is approximately 2.08, Eq. (7.12) determines that, for an initial orbit altitude of 100 n.mi. (185.2 km), the limit of usefulness for the bi-elliptic phasing technique is a final orbit altitude of 3927.7 n.mi. (7274.1 km). For final altitudes above this value, the Hohmann-transfer technique should be employed for all values of $\Delta\theta$.

When there is an intersection at $\Delta\theta_i$, the Hohmann technique would be used when $0 \leq \Delta\theta < \Delta\theta_i$, and the bi-elliptic technique would be used when $\Delta\theta_i \leq \Delta\theta \leq 2\pi$. To find $\Delta\theta_i$, set the total Hohmann time to the total bi-elliptic time

$$t_H + \frac{\Delta\theta_i}{(2\pi/P_i) - (2\pi/P_f)} = t_H + \frac{P_f}{2}\left(3 - \frac{\Delta\theta_i}{\pi}\right) \tag{7.13}$$

Solving for $\Delta\theta_i$,

$$\Delta\theta_i = 3\pi\left(1 - \frac{P_i}{P_f}\right) \tag{7.14}$$

The corresponding total time is

$$t_i = t_H + \frac{3P_i}{2} \tag{7.15}$$

The break-even phasing angle $\Delta\theta_i$ is presented as a function of h_f for $h_i = 100$ n.mi. (185.2 km) in Fig. 7.6.

Semitangential transfer. One more transfer technique to achieve coplanar rendezvous should be examined. Figure 7.7 illustrates the semitangential transfer. The rendezvous satellite achieves a transfer ellipse by applying a ΔV_1 that is larger than ΔV_1 for a Hohmann transfer. This transfer ellipse intersects the final circular orbit at two points, I_1 and I_2. Rendezvous can be accomplished at either point by the application of a second ΔV to circularize the orbit.

The rendezvous solution proceeds as follows:

1) Apply a specified ΔV_1 to V_{ci} in the direction of motion.

2) Knowing the perigee radius r_p and the perigee velocity V_p of the transfer ellipse, calculate the semimajor axis a, for the transfer ellipse from the energy equation

$$V_p^2 = \mu\left[\frac{2}{r_p} - \frac{1}{a}\right] \tag{7.16}$$

3) Calculate the eccentricity e of the transfer ellipse from

$$r_p = a(1 - e) \tag{7.17}$$

4) Calculate the true anomaly v_1 of the intersection point, I_1, from the orbit equation

$$r_f = \frac{a(1 - e^2)}{1 + e\cos v_1} \tag{7.18}$$

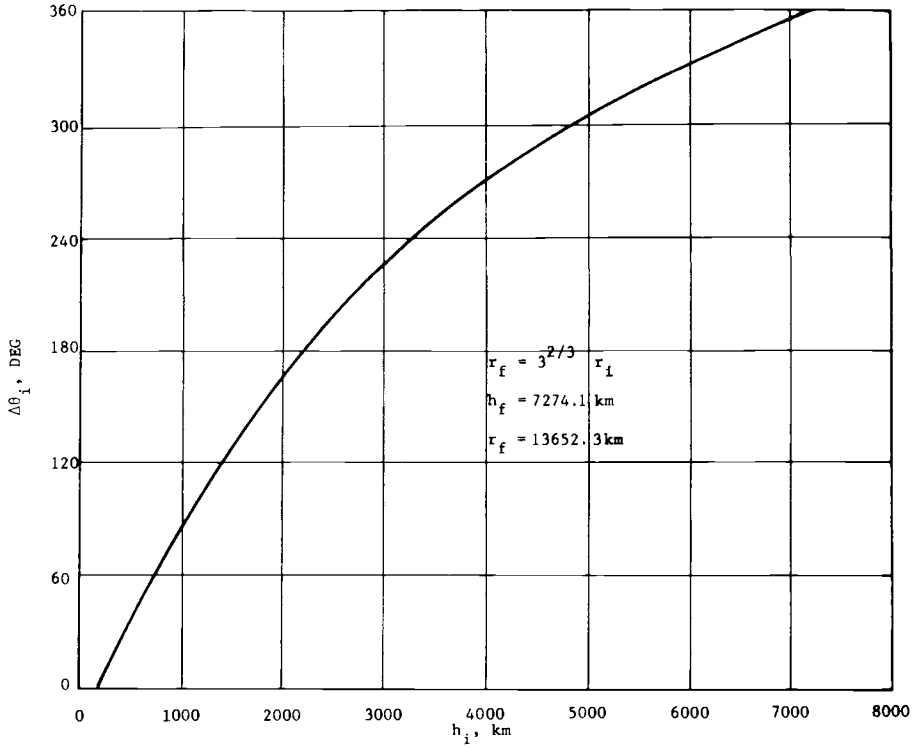


Fig. 7.6 Break-even point phasing angle for h_i = 185 km.

For I_2, $v_2 = 360 \text{ deg} - v_1$,
5) Calculate the flight-path angle γ_1 from

$$\tan \gamma_1 = \frac{e \sin v_1}{1 + e \cos v_1} \tag{7.19}$$

For I_2, $\gamma_2 = -\gamma_1$,
6) Calculate the eccentric anomaly E_1 from

$$\cos E_1 = \frac{e + \cos v_1}{1 + e \cos v_1} \tag{7.20}$$

7) Calculate the period P_t of the transfer orbit from

$$P = \frac{2\pi a^{3/2}}{\sqrt{\mu}} \tag{7.21}$$

8) Calculate the time t_1 from the application of ΔV_1 to the intersection point I_1 from Kepler's equation,

$$t_1 = \frac{P}{2\pi}(E_1 - e \sin E_1) \tag{7.22}$$

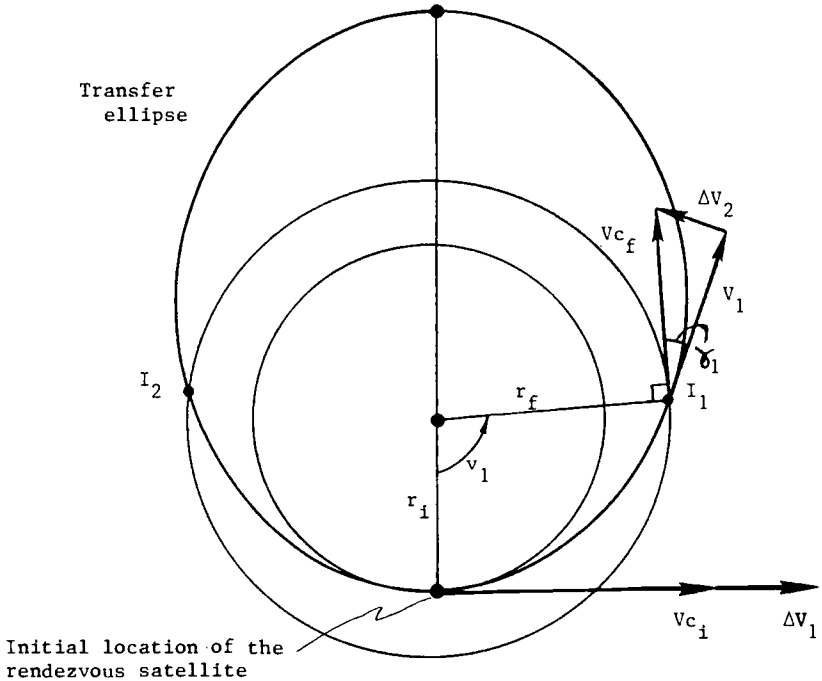


Fig. 7.7 Semitangential transfer geometry.

For I_2, $t_2 = P - t_1$,
9) Calculate $\Delta\theta_1$ for the first intersection point I_1 from

$$\Delta\theta_1 = \left[v_1 - \frac{(360\text{ deg})t_1}{P_f} - \theta_H\right] \text{mod}(2\pi) \tag{7.23}$$

For I_2, calculate $\Delta\theta_2$ from Eq. (7.23) using v_2 and t_2.
10) Calculate ΔV_2 from the vector triangle described in Fig. 7.7

$$\Delta V_2 = \sqrt{V_1^2 + V_{cf}^2 - 2V_1 V_{cf} \cos\gamma_1} \tag{7.24}$$

The magnitude of ΔV_2 is the same for I_1 and I_2 because the vector triangles are congruent.

The previous semitangential transfer was tangent to the initial orbit at the point of departure. A different semitangential transfer is tangent to the final orbit at the point of arrival. The solution for this technique proceeds by selecting a value for the semimajor axis of the transfer orbit. Because the apogee radius equals the radius of the final orbit, the transfer-orbit eccentricity can be calculated from Eq. (7.17). Similarly, Eqs. (7.18) through (7.22) can be used to calculate conditions at the departure point. Then,

$$\Delta\theta_1 = (180\text{ deg} - v_1) - \frac{(360\text{ deg})\left(\dfrac{P_t}{2} - t_1\right)}{P_f} - \theta_H \tag{7.25}$$

Finally, ΔV_1 is calculated from

$$\Delta V_1 = \sqrt{V_1^2 + V_{ci}^2 - 2V_1 V_{ci} \cos \gamma_1} \tag{7.26}$$

where the subscript 1 refers to the point of departure. ΔV_2 is simply the difference between the circular orbit velocity of the final orbit and the apogee velocity of the transfer orbit.

Authors' note. The authors are grateful to A. S. Ganeshan and R. Gupta, of the ISRO Satellite Centre in Bangalore, India, who pointed out a deficiency in the semitangential transfer results presented in the first edition of this book.

Ganeshan and Gupta also pointed out the usefulness of solutions to Lambert's problem. These solutions should, of course, be included and compared. This second edition includes results for the Lambert, Hohmann, bi-elliptic, and semitangential with tangency at departure techniques. Also included are results described earlier in this section for the semitangential with tangency at arrival technique. Comparisons among these techniques are quite interesting.

Transfers based on solutions to Lambert's problem. Lambert's problem, namely, to find the transfer orbit that connects two given position vectors in a specified transfer time, has been the subject of extensive literature. Pitkin,[2] Lancaster and Blanchard,[3] Herrick,[4] Battin,[5] Gooding,[6] and Prussing and Conway,[7] among others, have made important contributions to the solution of the problem.

Lambert's theorem states that the transfer time depends only on the semimajor axis of the transfer orbit a, the sum of the radii to points 1 and 2, $(r_1 + r_2)$, and the chord length c between the points. Figure 7.8 presents the geometry of the transfer orbit. Note that the transfer-orbit eccentricity is not explicitly included.

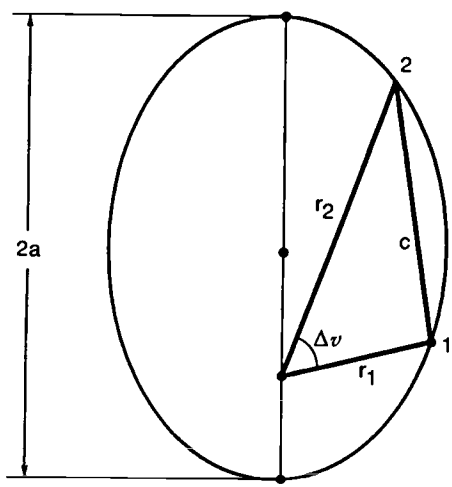


Fig. 7.8 Transfer-orbit geometry.

Lambert's equation, the solution to the problem, is derived very clearly by Prussing and Conway,[7]

$$\sqrt{\mu}(t_2 - t_1) = a^{3/2}[\alpha - \beta - (\sin\alpha - \sin\beta)] \tag{7.27}$$

where

$$r_1 + r_2 + c = 2a(1 - \cos\alpha) = 4a\sin^2\left(\frac{\alpha}{2}\right) \tag{7.28}$$

$$r_1 + r_2 - c = 2a(1 - \cos\beta) = 4a\sin^2\left(\frac{\beta}{2}\right) \tag{7.29}$$

Ambiguities, indeterminancies, and the transcendental nature of Lambert's equation complicate the process of solving for actual values. Herrick[4] developed a universal solution that co-author C. C. Chao developed into a computationally robust and efficient software routine.

Solutions to Lambert's equation are very important to phasing for rendezvous. By letting r_1 and r_2 correspond to the initial- and final-orbit radii and by relating c and ΔV geometrically through

$$c = \sqrt{r_1^2 + r_2^2 - 2r_1r_2\cos\Delta v} \tag{7.30}$$

the following procedure was used to calculate transfer solutions by Lambert's equation.

For this technique, the equation for $\Delta\theta$ is

$$\Delta\theta = \Delta v - \frac{(360\text{ deg})\Delta t}{p_f} - \theta_H \tag{7.31}$$

where $\Delta v = v_2 - v_1 =$ transfer arc and $\Delta t = t_2 - t_1 =$ transfer time. Then, use the following procedure:

1) Select a value for $\Delta\theta$.
2) Select a value for Δv, and calculate Δt from Eq. (7.31).
3) Use the HERRIK or similar solution routine to calculate transfer-orbit characteristics, including the transfer-orbit velocities at points 1 and 2. Because r_1 and r_2 correspond to initial and final circular orbit radii, the routine will also calculate ΔV_1 and ΔV_2 required for the transfer. Finally, $\Delta V_T = \Delta V_1 + \Delta V_2$.

By cycling through many values of Δv for a specified $\Delta\theta$, a minimum value of ΔV_T can be found iteratively. Then, by repeating this process for many values of $\Delta\theta$, a curve of minimum ΔV_T Lambert solutions can be developed. These solutions should be very efficient of ΔV_T because they are unconstrained with respect to the direction of ΔV application.

Comparison of transfer techniques for coplanar rendezvous. Comparisons among rendezvous techniques usually take the form of comparisons of required ΔV_T and elapsed time as a function of initial geometry. A comparison of the Hohmann-, bi-elliptic-, semitangential-, and Lambert-transfer techniques for

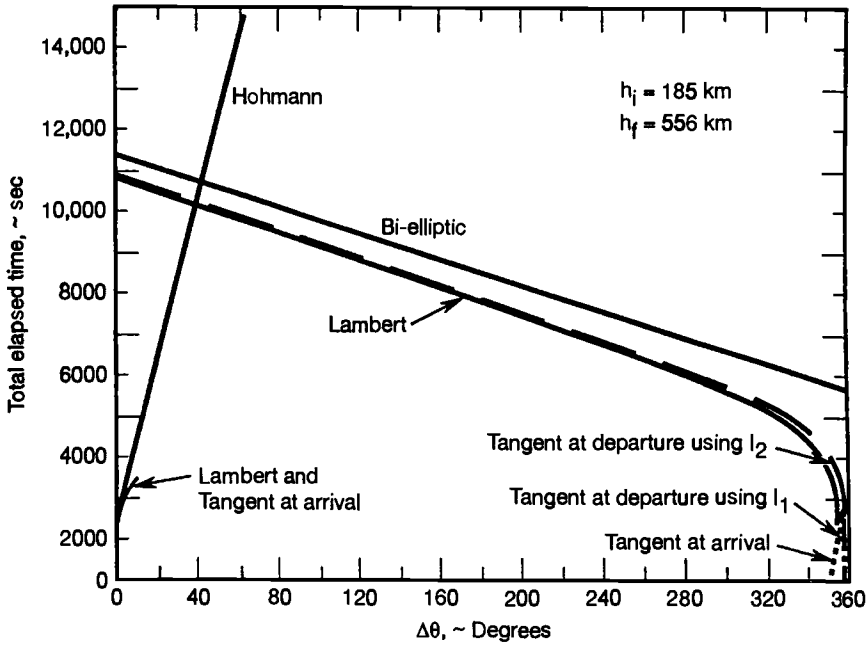


Fig. 7.9 Total elapsed time for rendezvous as a function of initial position of the target satellite for six rendezvous phasing techniques.

coplanar rendezvous will now be made for two specified circular orbits, $h_i = 100$ n.mi. (185.2 km) and $h_f = 300$ n.mi. (555.6 km).

Figure 7.9 presents an actual version of Fig. 7.5 for six transfer techniques. Total elapsed time for rendezvous is presented as a function of $\Delta\theta$. The Hohmann and bi-elliptic curves cross at $\Delta\theta_i = 42.7$ deg, as can be verified by Fig. 7.6. The associated time t_i from Eq. (7.15) is 10,700 s. Only a portion of the Hohmann-transfer curve is shown as it extends off-scale. This curve continues linearly to a value of 69,700 s at $\Delta\theta = 360$ deg. This value is simply the Hohmann-transfer time (2759 s) plus the synodic period (66,941 s) from Eq. (7.6) and Fig. 7.3. The bi-elliptic curve varies linearly from 11,378 s at $\Delta\theta = 0$ to 5632 s at $\Delta\theta = 360$ deg.

For the large middle region of $\Delta\theta$ between approximately $\Delta\theta \approx 10$ deg and $\Delta\theta = 340$ deg, the semitangential and Lambert solutions are represented by a pair of curves whose values are very nearly the same. At $\Delta\theta = 0$, the time for the Lambert transfer is 10,801 s, and the time for the semitangential transfer is 10,807 s. At $\Delta\theta = 340$ deg, the Lambert-transfer time is 4536 s, whereas the semitangential-transfer time is 4633 s. In this large $\Delta\theta$ region, the best semitangential transfers are tangent at departure.

Figure 7.10 presents the total transfer ΔV requirement as a function of $\Delta\theta$ for the six transfer techniques. The Hohmann-transfer curve for ΔV_T is a horizontal line at a value of 211 m/s. The bi-elliptic curve for ΔV_T ranges from 2782 m/s at $\Delta\theta = 0$ to 211 m/s at $\Delta\theta = 360$ deg.

The curves for the Lambert and semitangential transfers are very close. The difference in values for ΔV_T between the two techniques is 5 m/s or less for

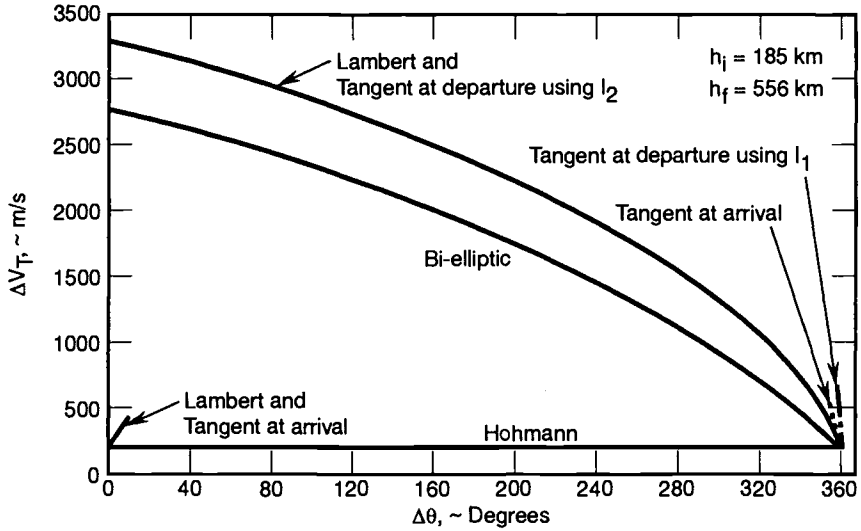


Fig. 7.10 Total ΔV_T required for rendezvous as a function of initial position of the target satellite for six rendezvous phasing techniques.

10 deg $\leq \Delta\theta \leq$ 340 deg, with the Lambert values being the better, that is, the smaller one.

Comparing all six techniques on both Figs. 7.9 and 7.10 in the range 10 deg $\leq \Delta\theta \leq$ 340 deg, the Hohmann technique is best in terms of time and ΔV_T for small values of $\Delta\theta$. As $\Delta\theta$ increases, the Hohmann time increases rapidly while the times for the other techniques decrease gradually. As $\Delta\theta$ increases, the ΔV_T values for the bi-elliptic, semitangential, and Lambert techniques gradually decrease. In this large middle region of $\Delta\theta$, comparative values of time and ΔV_T must be evaluated for each specific application. If ΔV_T is critical and time is available, then the Hohmann technique is attractive. If time is critical and ΔV_T is available, then the Lambert and semitangential techniques are attractive.

Results in two regions, $0 \leq \Delta\theta \leq 10$ deg and 340 deg $\leq \Delta\theta \leq 360$ deg, are very interesting and need to be examined closely. Figures 7.11 and 7.12 present time and ΔV_T vs $\Delta\theta$ for $0 \leq \Delta\theta \approx 9.5$ deg. Curves are presented for the Hohmann, Lambert, and semitangential techniques. This semitangential technique is different from the previously discussed semitangential technique in that this transfer is tangent at arrival. This Lambert solution is very similar to the semitangential transfer and is different from the other Lambert solutions in this $\Delta\Theta$ range shown on Fig. 7.9 and 7.10 in that the times and ΔV are much lower. All of the solutions have common values at $\Delta\theta = 0$, namely, $\Delta t = 2759$ s and $\Delta V_T = 211$ m/s. As $\Delta\theta$ increases, the Lambert and semitangential curves, which are virtually indistinguishable, display shorter travel times and higher ΔV_T values than the Hohmann solutions. The Lambert solutions are slightly shorter in time, 35 s or less, and slightly lower in ΔV_T, 0.5 m/s or less, than the semitangential solutions. In both solutions, the active satellite establishes a transfer trajectory whose perigee altitude is less than 185 km and whose apogee altitude is at or very near 556 km. The transfer arc is slightly more than one-half revolution. An unfortunate

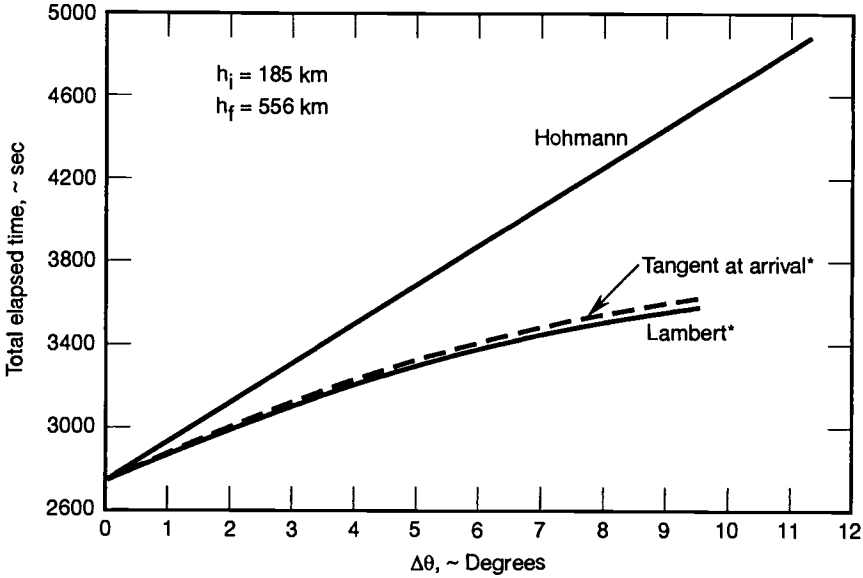


Fig. 7.11 Total elapsed time vs $\Delta\theta$ for $0 \leq \Delta\theta < 9.5$ deg for three rendezvous phasing techniques.

characteristic of these transfer trajectories is that the perigee altitude decreases as $\Delta\theta$ increases. For $\Delta\theta = 6$ deg, the Lambert solution perigee altitude is 120 km and, for $\Delta\theta = 9.5$ deg, the Lambert perigee altitude is 55 km. Therefore, Lambert and semitangential solutions beyond about $\Delta\theta = 6$ deg are impractical because the active satellite would have to negotiate a very low perigee altitude. For $0 \leq \Delta\theta < 6$ deg, the choice between the Lambert and Hohmann techniques is the classic tradeoff between shorter transfer times and higher velocity requirements.

Transfer solutions in the range 340 deg $\leq \Delta\theta \leq$ 360.1 deg are displayed on Figs. 7.13 and 7.14. Figure 7.13 presents travel time for the Lambert solution and both semitangential solutions, that is, tangent at departure and tangent at arrival. In addition, tangent at departure solutions are presented for both intersection points, I_1 and I_2. See Fig. 7.7. Travel-time curves for the Hohmann and bi-elliptic solutions are beyond the scale of this figure.

The curves on Fig. 7.13 display some unusual characteristics. As $\Delta\theta$ decreases from 360 deg, values for Δt increase for tangent at departure using intersection I_2 and decrease for tangent at arrival and for tangent at departure using I_1. As $\Delta\theta$ decreases, values for Δt decrease and then increase for the Lambert technique. Figure 7.14 shows equally surprising results for ΔV_T vs $\Delta\theta$ for these techniques.

In the range 355 deg $\leq \Delta\theta \leq$ 360 deg, the tangent at arrival solutions provide fast transfers for reasonable values of ΔV_T. At $\Delta\theta = 356$ deg, the tangent at arrival solution is $\Delta t = 1940$ s and $\Delta V_T = 366$ m/s. The central angle of travel in the transfer orbit is only 122 deg from departure to arrival at apogee. The perigee altitude is only 90 km and is less as $\Delta\theta$ decreases, but the satellite does not traverse the perigee region in these solutions. The detracting feature of these solutions is that ΔV_T increases rapidly for $\Delta\theta < 355$ deg. However, these solutions are attractive for emergency rendezvous missions where transfer time is critical.

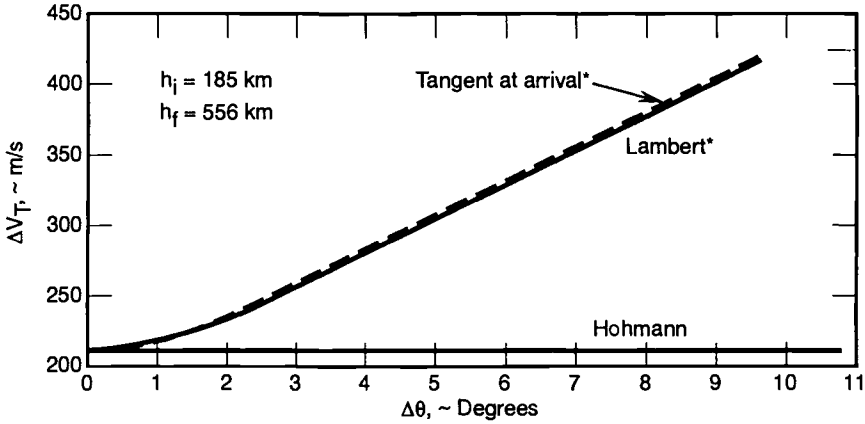


Fig. 7.12 Total ΔV_T vs $\Delta\theta$ for $0 \leq \Delta\theta < 9.5$ deg for three rendezvous phasing techniques.

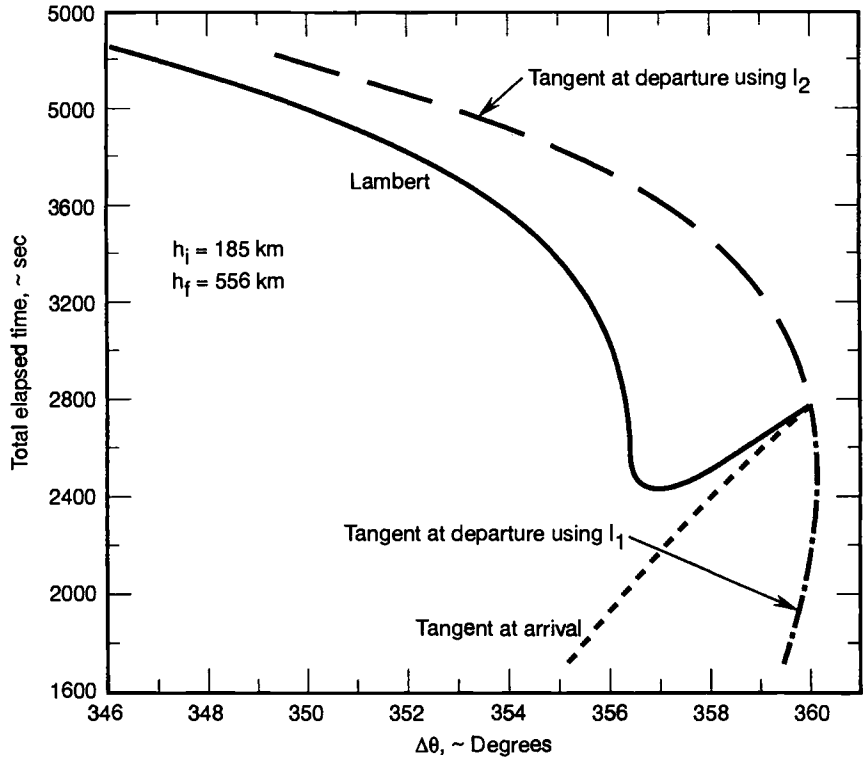


Fig. 7.13 Total elapsed time vs $\Delta\theta$ for $346 \leq \Delta\theta \leq 360.1$ deg for four rendezvous phasing techniques.

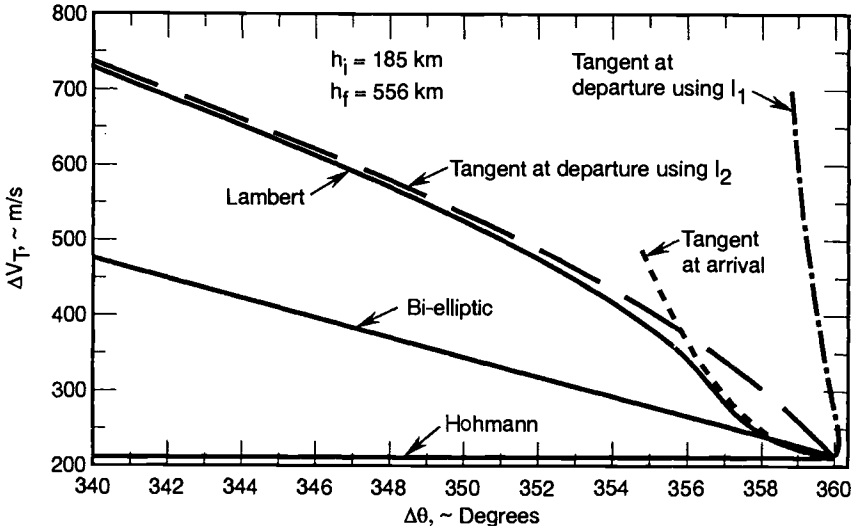


Fig. 7.14 Total ΔV_T vs $340 \leq \Delta\theta \leq 360.1$ deg for six rendezvous phasing techniques.

Tangent at departure solutions using I_1 are available for emergencies in the narrow range 359 deg $< \Delta\theta <$ 360.1 deg. As the figures show, these transfers are very fast, that is, of short duration. However, the ΔV_T rises very quickly as $\Delta\theta$ increases to 360.1 deg and then decreases to 359 deg.

In terms of ΔV_T, the Lambert solutions are always lower, although only slightly so, than the tangent at arrival solutions. In terms of Δt, the Lambert curve decreases as $\Delta\theta$ decreases, but then it levels out at $\Delta\theta \approx 357$ deg and increases thereafter. For $\Delta\theta < 355$ deg, the Lambert solutions more closely resemble the tangent at departure solutions using I_2 than the tangent at arrival solutions. The Lambert solutions are always shorter in time and lower in ΔV_T than the tangent at departure solutions using I_2.

Finally, it is interesting to note that the Lambert and tangent at arrival solutions are lower in ΔV_T than the bi-elliptic solution for 358.5 deg $< \Delta\theta <$ 360 deg. And no solutions are lower in ΔV_T than the Hohmann solution of 211 m/s. All of the solutions coalesce to exactly the same solution at $\Delta\theta = 360$ deg.

Three-Dimensional Space Rendezvous

Modified Hohmann-transfer technique. Figure 7.15 depicts the modified Hohmann-transfer technique for three-dimensional rendezvous. Three-dimensional rendezvous means that the initial and final circular orbits are not coplanar but have a dihedral angle of rotation between the orbit planes α, as shown on the figure. Thus, a Hohmann transfer and a plane change maneuver are required for this technique.

Phasing is accomplished in the same way that was described for the coplanar Hohmann-transfer technique, except that the phasing angles are measured in two different orbit planes. Figure 7.15 shows the line of intersection of the two planes as the line of nodes. Let the in-orbit positions of the satellites be measured from this line, i.e., θ_i describes the position of the rendezvous satellite in the initial orbit,

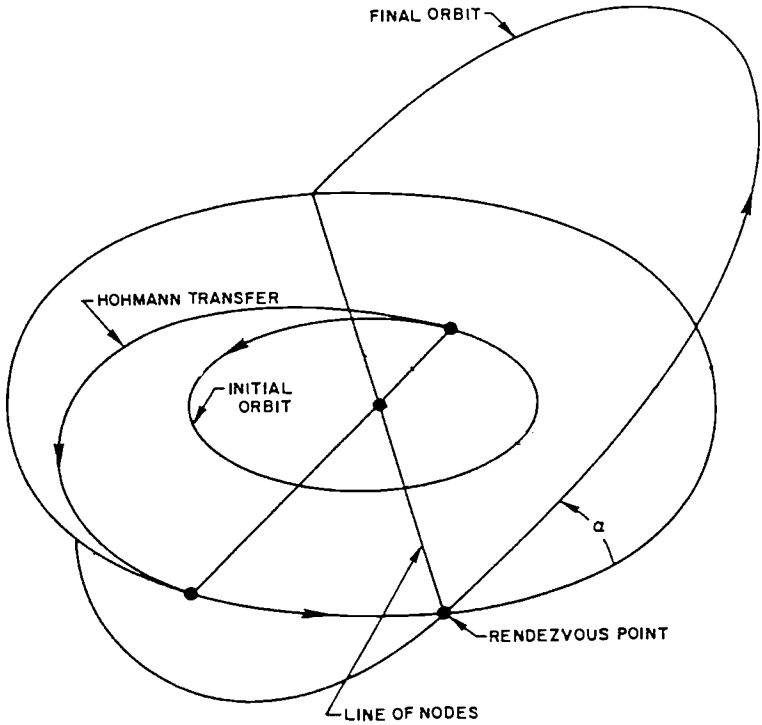


Fig. 7.15 Modified Hohmann-transfer maneuver.

and θ_f describes the position of the target satellite in the final orbit. Let $\theta_i - \theta_f = \theta_H + \Delta\theta$, as in the coplanar case. After a waiting time, $\Delta\theta$ becomes zero, and the Hohmann transfer is initiated. When the rendezvous satellite circularizes into the final orbit, both satellites are equidistant from the rendezvous point. When they simultaneously reach the rendezvous point, the rendezvous satellite performs a single-impulse plane change maneuver to rotate its orbit plane through the angle α, and rendezvous is accomplished. The time required for this technique is the sum of 1) the waiting time to achieve $\Delta\theta = 0$, 2) the Hohmann-transfer time, and 3) the time required for the rendezvous satellite to traverse the final orbit from the circularization point to the line of nodes.

Figure 7.16 describes the total velocity ΔV_T required for the three impulses as a function of the plane change angle α and the final circular orbit altitude h_f when the initial circular orbit altitude $h_i = 100$ n.mi. (185.2 km). When $\alpha = 0$ and $h_f = 300$ n.mi. (555.6 km), $\Delta V_T = 211$ m/s. This corresponds to the value shown on Fig. 7.10 for the coplanar Hohmann transfer. Note that ΔV_T increases very rapidly as α increases.

Bi-elliptic transfer with split plane changes. This transfer technique was previously described in Chapter 5. For three-dimensional rendezvous, the bi-elliptic transfer is initiated when the rendezvous satellite reaches the line of nodes. As in the coplanar case, the value of $\Delta\theta$ determines the altitude h_t of the intermediate transfer point. For the example of $h_i = 100$ n.mi. (185.2 km) and $h_f = 300$

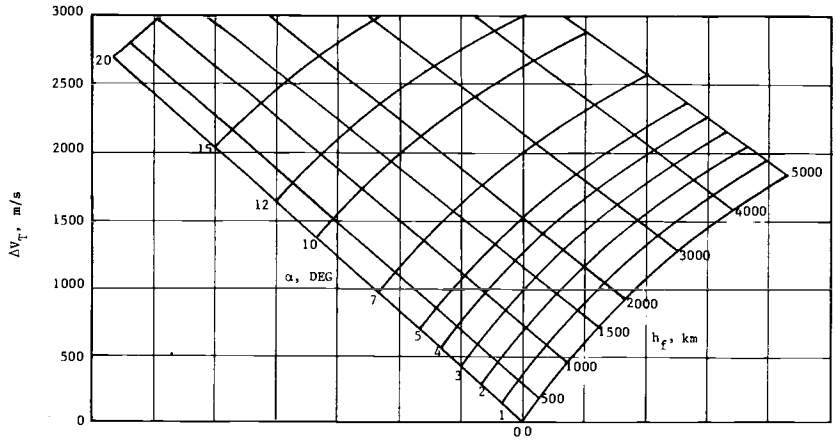


Fig. 7.16 Velocity required for a modified Hohmann transfer from a 185-km parking orbit.

n.mi. (555.6 km), Eqs. (7.8) and (7.9) determine h_t as a function of $\Delta\theta$. Then, given values for h_i, h_t, h_f, and the total plane change angle α_T, Ref. 8 describes the optimal plane change split, i.e., α_1, α_2, α_3, to minimize the total ΔV. Figure 7.17 presents these solutions for ΔV_T as a function of α_T and $\Delta\theta$ for $h_i = 100$ n.mi. (185.2 km) and $h_f = 300$ n.mi. (555.6 km). Because the ΔV_T for the coplanar bi-elliptic transfer is large compared to the Hohmann transfer (see Fig. 7.10), the advantage of the optimal split-plane change for the three-dimensional bi-elliptic transfer produces smaller values of ΔV_T than the modified Hohmann transfer only for large values of $\Delta\theta$.

In-Orbit Repositioning

Maneuvering technique. If a satellite is to be repositioned in its circular orbit, this maneuver can be performed by applying an impulsive velocity along the

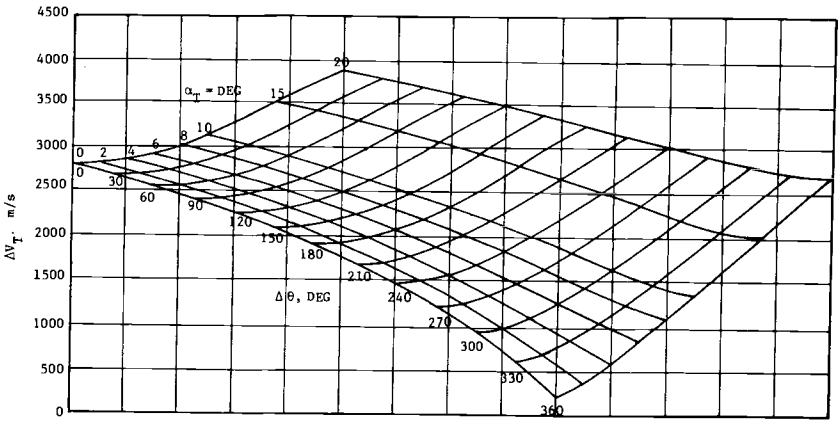


Fig. 7.17 Velocity increment ΔV_T necessary for a bi-elliptic transfer from a 185-km to a 556-km circular orbit.

velocity vector, either forward or retro. With a forward ΔV_1, the satellite will enter a larger phasing orbit. When the satellite returns to the point of ΔV application, it will be behind its original location in the circular orbit. The satellite can re-enter the circular orbit at this point by applying a retro ΔV equal in magnitude to the first ΔV. In a sense, a rendezvous with this point has been performed. Or the satellite can remain in the phasing orbit and re-enter the circular orbit on a future revolution. The satellite will drift farther behind with each additional revolution.

If the first ΔV is in the retro direction, the satellite will enter a smaller phasing orbit and will drift ahead of its original location in the circular orbit. The drift rate, either ahead or behind, is proportional to the magnitude of the ΔV.

Application to geosynchronous circular orbit. A very common application of repositioning is the drifting of a satellite in a geosynchronous circular equatorial orbit from one longitude to another. The change in longitude is given by the equation

$$\Delta L = \dot{L} n P_{\mathrm{PH}} \tag{7.32}$$

where

ΔL = the change in longitude
$\dot{L}$ = the drift rate, positive eastward
n = the number of revolutions spent in the phasing orbit
P_{PH} = the period of the phasing orbit

The drift rate $\dot{L}$ is given by

$$\dot{L} = \omega_E \left(\frac{P_{\mathrm{PH}} - P_o}{P_{\mathrm{PH}}} \right) \tag{7.33}$$

where

ω_E = 360.985647 deg/day is the angular rate of axial rotation of the Earth
P_o = 1436.068 min = 0.9972696 days is the period of the geosynchronous orbit

Substituting Eq. (7.33) into Eq. (7.32),

$$\Delta L = \omega_E n (P_{\mathrm{PH}} - P_o) \tag{7.34}$$

The repositioning problem can now be addressed as follows. Given a desired longitudinal shift, say $\Delta L = +90$ deg, then, from Eq. (7.34),

$$n(P_{\mathrm{PH}} - P_o) = \frac{\Delta L}{\omega_E} = 0.249317 \text{ days} \tag{7.35}$$

Selecting a value of n allows the solution of $(P_{\mathrm{PH}} - P_o)$. Adding P_o solves for P_{PH}. Substitution of P_{PH}, n, and ΔL into Eq. (7.32) allows the solution of $\dot{L}$. Figure 7.18 is a graph of the ΔV required to start and stop a longitudinal drift rate

Table 7.1 Repositioning of geosynchronous satellite solutions for a $\Delta L = +90$ deg

n rev	$P_{PH} - P_o$ days	P_{PH} days	nP_{PH} days	$\dot{L}$ deg/day	ΔV m/s
6	0.04155	1.0388	6.233	14.44	82.0
12	0.02078	1.0181	12.217	7.37	41.8
24	0.01039	1.0077	24.184	3.72	21.3
96	0.00260	0.9999	95.987	0.94	5.49

in a geosynchronous orbit as a function of drift rate. Thus, a ΔV can be associated with a value of $\dot{L}$.

Table 7.1 presents a number of solutions to the $\Delta L = +90$ deg example. Four values of n were assumed. For each value n, the table presents values of P_{PH}, nP_{PH} (the total elapsed time for repositioning), $\dot{L}$, and ΔV. If the repositioning is to be accomplished in 6 rev, then the drift rate is 14.44 deg/day, and the ΔV is 82.0 m/s. However, if the repositioning can be done slowly, i.e., in 96 rev, then the drift rate is only 0.94 deg/day, and the ΔV is only 5.49 m/s. This demonstrates the tradeoff between elapsed time and ΔV.

The curve on Fig. 7.18 was determined by assuming values of ΔV, calculating values of the phasing orbit semimajor axis from the energy equation, calculating

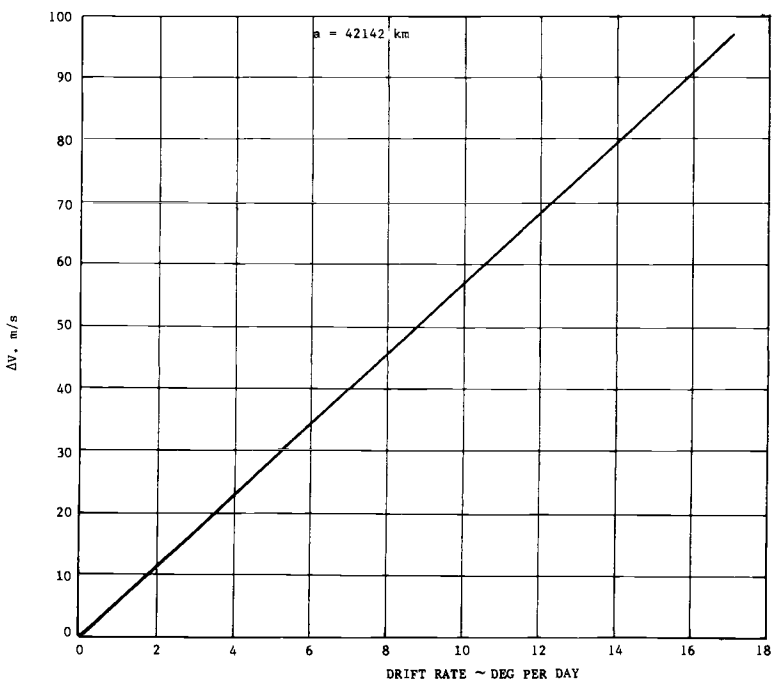


Fig. 7.18 ΔV required to start and stop a longitudinal drift rate in a geosynchronous circular orbit.

values of the phasing orbit period from the period equation, and calculating drift rates from Eq. (7.33). The equations and Fig. 7.18 work equally well for westward drifts.

7.2 Terminal Rendezvous

In the final phase of rendezvous before docking, the satellites are in close proximity, and the relative motion of the satellites is all-important. In this phase, it is common to describe the motion of one satellite with respect to the other. In the following subsections, the relative equations of motion will be derived. A solution to these equations will be obtained for the case in which one of the satellites is in a circular orbit.

Derivation of Relative Equations of Motion

Figure 7.19 presents the vector positions of the rendezvous and target satellites at some time with respect to the center of the Earth, $\boldsymbol{r}$ and $\boldsymbol{r}_T$. The position of the rendezvous satellite with respect to the target satellite is $\boldsymbol{\rho}$. An orthogonal coordinate frame is attached to the target satellite and moves with it. The y axis is radially outward. The z axis is out of the paper. The x axis completes a right-hand triad. The angular velocity, a vector of the target satellite, is given by $\boldsymbol{\omega}$.

The vector positions of the satellites yield

$$\boldsymbol{r} = \boldsymbol{r}_T + \boldsymbol{\rho} \tag{7.36}$$

Differentiating this equation with respect to an inertial coordinate frame results in

$$\ddot{\boldsymbol{r}} = \ddot{\boldsymbol{r}}_T + \ddot{\boldsymbol{\rho}} + 2(\boldsymbol{\omega} \times \dot{\boldsymbol{\rho}}) + \dot{\boldsymbol{\omega}} \times \boldsymbol{\rho} + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \boldsymbol{\rho}) \tag{7.37}$$

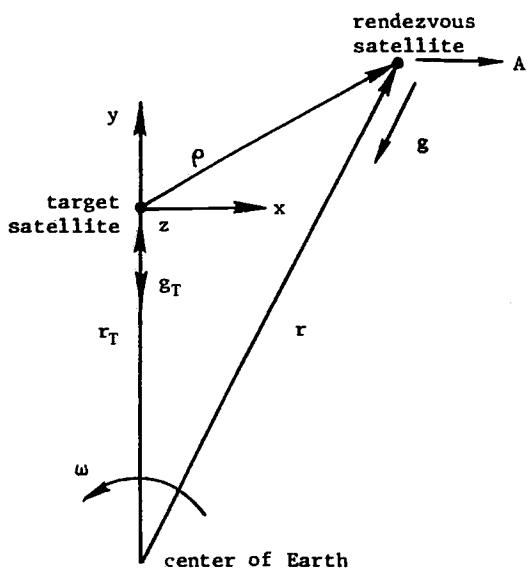


Fig. 7.19 Geometry and coordinate system for terminal rendezvous.

where, as in Eq. (1.2)

$$\begin{aligned}
\ddot{\boldsymbol{r}} &= \text{the inertial acceleration of the rendezvous satellite}\\
\ddot{\boldsymbol{r}}_T &= \text{the inertial acceleration of the target satellite}\\
\ddot{\boldsymbol{\rho}} &= \text{the acceleration of the rendezvous satellite relative to}\\
&\quad\ \text{the target satellite}\\
2(\boldsymbol{\omega} \times \dot{\boldsymbol{\rho}}) &= \text{the Coriolis acceleration}\\
\dot{\boldsymbol{\omega}} \times \boldsymbol{\rho} &= \text{the Euler acceleration}\\
\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \boldsymbol{\rho}) &= \text{the centripetal acceleration}
\end{aligned}$$

Now, let

$$\ddot{\boldsymbol{r}} = \boldsymbol{g} + \boldsymbol{A} \tag{7.38}$$

where $\boldsymbol{g}$ is the gravitational acceleration and $\boldsymbol{A}$ the acceleration applied by external forces (thrust). Resolving Eqs. (7.37) and (7.38) into the x, y, and z components and solving for the relative accelerations produce

$$\begin{aligned}
\ddot{x} &= -g\frac{x}{r} + A_x + 2\omega\dot{y} + \dot{\omega}y + \omega^2 x\\
\ddot{y} &= -g\left(\frac{y + r_T}{r}\right) + A_y + g_T - 2\omega\dot{x} - \dot{\omega}x + \omega^2 y\\
\ddot{z} &= -g\frac{z}{r} + A_z
\end{aligned} \tag{7.39}$$

Assuming that the target-to-satellite distance is much smaller than the orbit radius of the target satellite or that

$$\rho^2 = x^2 + y^2 + z^2 \ll r_T^2 \tag{7.40}$$

the following approximate relations can be written

$$\begin{aligned}
r &= \left[x^2 + (y + r_T)^2 + z^2\right]^{1/2}\\
&\approx r_T\left(1 + \frac{y}{r_T}\right)\\
g &= \frac{g_T r_T^2}{r^2} \approx g_T\left(1 - \frac{2y}{r_T}\right)\\
-g\frac{x}{r} &\approx -g_T\frac{x}{r_T}\\
-g\frac{z}{r} &\approx -g_T\frac{z}{r_T}\\
-g\left(\frac{y + r_T}{r}\right) &\approx -g_T\left(1 - \frac{2y}{r_T}\right)
\end{aligned} \tag{7.41}$$

Therefore, the linearized Eqs. (7.39) become

$$\begin{aligned}
\ddot{x} &= -g_T\frac{x}{r_T} + A_x + 2\omega\dot{y} + \dot{\omega}y + \omega^2 x \\
\ddot{y} &= +2g_T\frac{y}{r_T} + A_y - 2\omega\dot{x} - \dot{\omega}x + \omega^2 y \\
\ddot{z} &= -g_T\frac{z}{r_T} + A_z
\end{aligned} \tag{7.42}$$

When the target is in a circular orbit, $\dot{\omega} = 0$ and $\omega = \sqrt{\frac{\mu}{{r_T}^3}}$, and Eqs. (7.42) become

$$\begin{aligned}
\ddot{x} &= A_x + 2\omega\dot{y} \\
\ddot{y} &= A_y - 2\omega\dot{x} + 3\omega^2 y \\
\ddot{z} &= A_z - \omega^2 z
\end{aligned} \tag{7.43}$$

If there are no external accelerations (e.g., thrust), then,

$$\begin{aligned}
A_x = A_y = A_z &= 0 \\
\ddot{x} - 2\omega\dot{y} &= 0 \\
\ddot{y} + 2\omega\dot{x} - 3\omega^2 y &= 0 \\
\ddot{z} + \omega^2 z &= 0
\end{aligned} \tag{7.44}$$

Solution to the Relative Equations of Motion

The z equation is uncoupled from the x and y equations and can be solved separately. Assume a solution of the form

$$z = A\sin\omega t + B\cos\omega t \tag{7.45}$$

Differentiating

$$\begin{aligned}
\dot{z} &= A\omega\cos\omega t - B\omega\sin\omega t \\
\ddot{z} &= -A\omega^2\sin\omega t - B\omega^2\cos\omega t
\end{aligned}$$

When $t = 0$, $z = z_0$, and $\dot{z} = \dot{z}_0$, and so $z_0 = B$, and $\dot{z}_0 = A\omega$; therefore,

$$\begin{aligned}
z &= \frac{\dot{z}_0}{\omega}\sin\omega t + z_0\cos\omega t \\
\dot{z} &= \dot{z}_0\cos\omega t - z_0\omega\sin\omega t
\end{aligned} \tag{7.46}$$

Substitution into the $\ddot{z}$ equation verifies that these equations are a solution. In mechanics, they correspond to simple harmonic motion.

EQUATIONS OF MOTION

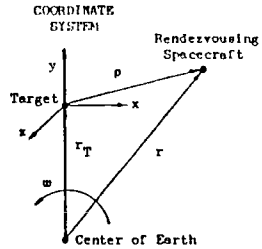
$$\ddot{x} - 2\omega\dot{y} = 0$$

$$\ddot{y} + 2\omega\dot{x} - 3\omega^2 y = 0$$

$$\ddot{z} + \omega^2 z = 0$$

These rendezvous equations apply when the target is in a circular orbit, i.e., $\dot{\omega} = 0$ and $\omega = \sqrt{\frac{\mu}{r_T^3}}$

No external forces are considered, i.e., $A_x = A_y = A_z = 0$. The xy plane is coincident with the orbit plane of the target vehicle.



$$\begin{bmatrix} x \\ y \\ z \\ \dot{x} \\ \dot{y} \\ \dot{z} \end{bmatrix} = \begin{bmatrix} 1 & 6(\omega t - \sin\omega t) & 0 & -3t + \frac{4}{\omega}\sin\omega t & \frac{2}{\omega}(1 - \cos\omega t) & 0 \\ 0 & 4 - 3\cos\omega t & 0 & \frac{2}{\omega}(-1 + \cos\omega t) & \frac{1}{\omega}\sin\omega t & 0 \\ 0 & 0 & \cos\omega t & 0 & 0 & \frac{1}{\omega}\sin\omega t \\ 0 & 6\omega(1 - \cos\omega t) & 0 & -3 + 4\cos\omega t & 2\sin\omega t & 0 \\ 0 & 3\omega\sin\omega t & 0 & -2\sin\omega t & \cos\omega t & 0 \\ 0 & 0 & -\omega\sin\omega t & 0 & 0 & \cos\omega t \end{bmatrix} \begin{bmatrix} x_o \\ y_o \\ z_o \\ \dot{x}_o \\ \dot{y}_o \\ \dot{z}_o \end{bmatrix}$$

Fig. 7.20 Solution to the first-order circular-orbit rendezvous equations.

The x and y equations are coupled but can be solved to produce

$$\begin{aligned} x &= x_0 + 2\frac{\dot{y}_0}{\omega}(1 - \cos\omega t) + \left(4\frac{\dot{x}_0}{\omega} - 6y_0\right)\sin\omega t + (6\omega y_0 - 3\dot{x}_0)t \\ y &= 4y_0 - 2\frac{\dot{x}_0}{\omega} + \left(2\frac{\dot{x}_0}{\omega} - 3y_0\right)\cos\omega t + \frac{\dot{y}_0}{\omega}\sin\omega t \\ \dot{x} &= 2\dot{y}_0\sin\omega t + (4\dot{x}_0 - 6\omega y_0)\cos\omega t + 6\omega y_0 - 3\dot{x}_0 \\ \dot{y} &= (3\omega y_0 - 2\dot{x}_0)\sin\omega t + \dot{y}_0\cos\omega t \end{aligned} \tag{7.47}$$

where x_0, $\dot{x}_0$, y_0, and $\dot{y}_0$ are position and velocity components at $t = 0$.

Figure 7.20 presents these solutions in matrix form. This is a compact, descriptive form. Given the initial position and velocity, the position and velocity at some future time can be determined from these equations.

Two-Impulse Rendezvous Maneuver

Given the initial position $\boldsymbol{\rho}_0$ and velocity $\dot{\boldsymbol{\rho}}_0$ for the rendezvous satellite with respect to the target satellite at the origin of the coordinate system and given the desire to rendezvous at a specified time τ, the problem is to find ΔV_1 at $t = 0$ and ΔV_2 at $t = \tau$ to accomplish rendezvous. Figure 7.21 presents a schematic of this two-impulse rendezvous maneuver.

The solution proceeds as follows.

If at time $t = 0$, the relative position x_0, y_0, z_0 is known (components of $\boldsymbol{\rho}_0$), then the relative velocity components $\dot{x}_{0r}$, $\dot{y}_{0r}$, $\dot{z}_{0r}$ necessary to rendezvous at time

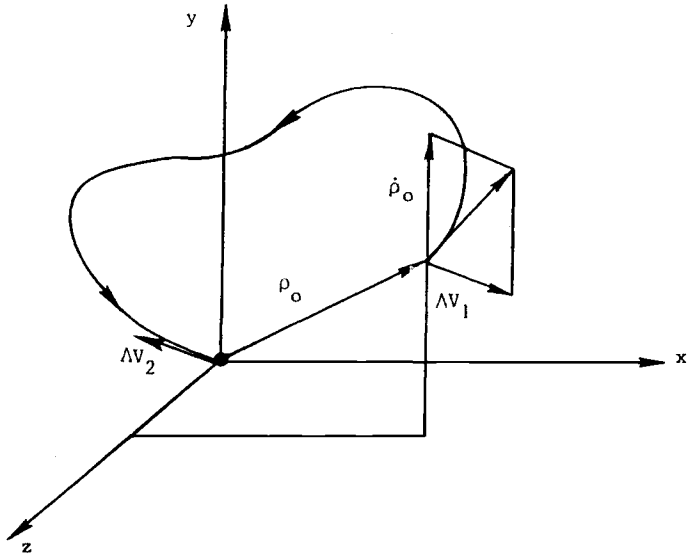


Fig. 7.21 Two-impulse rendezvous maneuver.

$t = \tau$ in the future can be obtained from the x, y, z equations by assuming that $x = y = z = 0$ and solving for $\dot{x}_{0r}, \dot{y}_{0r}, \dot{z}_{0r}$ as follows:

$$\begin{aligned}
\frac{\dot{x}_{0r}}{\omega} &= \frac{x_0 \sin \omega\tau + y_0[6\omega\tau \sin \omega\tau - 14(1 - \cos \omega\tau)]}{\Delta} \\
\frac{\dot{y}_{0r}}{\omega} &= \frac{2x_0(1 - \cos \omega\tau) + y_0(4 \sin \omega\tau - 3\omega\tau \cos \omega\tau)}{\Delta} \\
\frac{\dot{z}_{0r}}{\omega} &= \frac{-z_0}{\tan \omega\tau}
\end{aligned} \tag{7.48}$$

where $\Delta = 3\omega\tau \sin \omega\tau - 8(1 - \cos \omega\tau)$. The first impulse is given by

$$\Delta V_1 = \left[(\dot{x}_{0r} - \dot{x}_0)^2 + (\dot{y}_{0r} - \dot{y}_0)^2 + (\dot{z}_{0r} - \dot{z}_0)^2\right]^{1/2} \tag{7.49}$$

where $\dot{x}_0, \dot{y}_0, \dot{z}_0$ are the actual (initial) velocities of the chaser relative to the target at time $t = 0$.

The components of the second impulse ΔV_2 are the relative velocities $\dot{x}_\tau, \dot{y}_\tau, \dot{z}_\tau$ at time $t = \tau$, with the initial conditions x_0, y_0, z_0, and $\dot{x}_{0r}, \dot{y}_{0r}, \dot{z}_{0r}$. Thus,

$$\Delta V_2 = \left(\dot{x}_\tau^2 + \dot{y}_\tau^2 + \dot{z}_\tau^2\right)^{1/2} \tag{7.50}$$

The ΔV_2 is necessary to stop the chaser vehicle at the target.

Two-Impulse Rendezvous Maneuver Example

Given the ΔV, 21 m/s, for in-track departure from a circular, synchronous (24-h) equatorial orbit, this example will investigate the two-impulse rendezvous

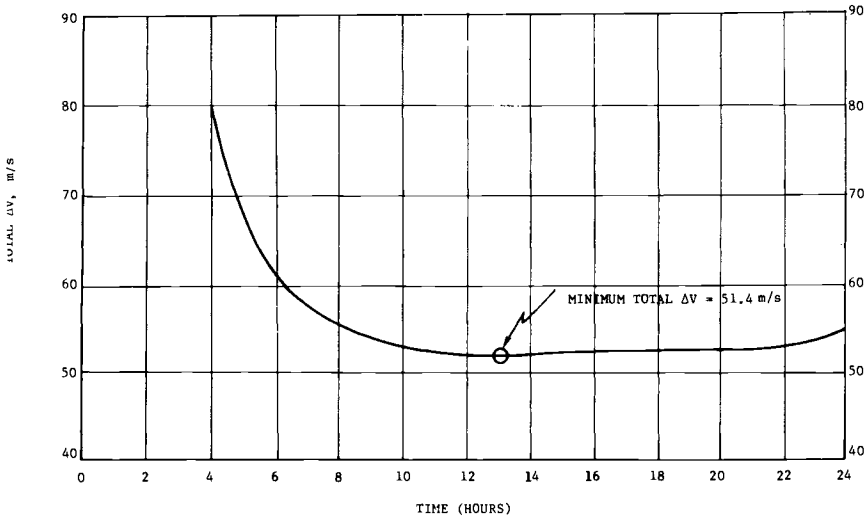


Fig. 7.22 Total ΔV vs time from departure to return for the rendezvous maneuver example.

maneuver to return to the original longitude and orbit in a specified time. This initial ΔV may be applied in order to avoid some debris. The rendezvous maneuver begins 2 h after the application of the 21-m/s ΔV and ends at a specified but variable time, as illustrated in Fig. 7.22.

Figure 7.22 presents the sum of the first ΔV of 21 m/s and the sum of the two impulses required for rendezvous as a function of time from the application of the initial ΔV to the completion of rendezvous. A minimum value in the curve occurs at 13 h. The minimum total ΔV is 51.4 m/s. The two-impulse rendezvous maneuver requires 30.4 m/s.

For the minimum ΔV solution, Fig. 7.23 presents a history of x, in-track, vs y, radial. The position after 2 h is noted. The rendezvous maneuver to return begins at this point and takes 11 h. Figure 7.24 presents $\dot{x}$ vs $\dot{y}$. This figure graphically presents the magnitudes and directions of the ΔV. The first ΔV of 21.0 m/s is applied in-track; i.e., $\dot{x} = -21.0$ m/s, and $\dot{y} = 0$. At $t = 2$ h, the second ΔV of 29.0 m/s is applied. Its components are $\dot{x} = +19.6$ m/s and $\dot{y} = -21.4$ m/s. At $t = 13$ h, the third ΔV of 1.4 m/s is applied. Its components are $x = +1.34$ m/s, and $\dot{y} = -0.40$ m/s. Yaw and pitch angles are measured in the x-y plane. Nose up is (+), and nose down is (−). A yaw angle of 0 means that the nose of the satellite is pointed forward, i.e., in the $-x$ direction. A yaw angle of 180 deg means that the nose of the satellite is pointed in the $+x$ direction.

If the satellite can point its engines in any direction, the total ΔV is the sum of the three ΔV magnitudes, i.e., 51.42 m/s. However, if the satellite's engines point in the x and y directions but cannot be reoriented, then the total ΔV is the sum of all the x and y components, i.e., 63.74 m/s, as tabulated on the figure.

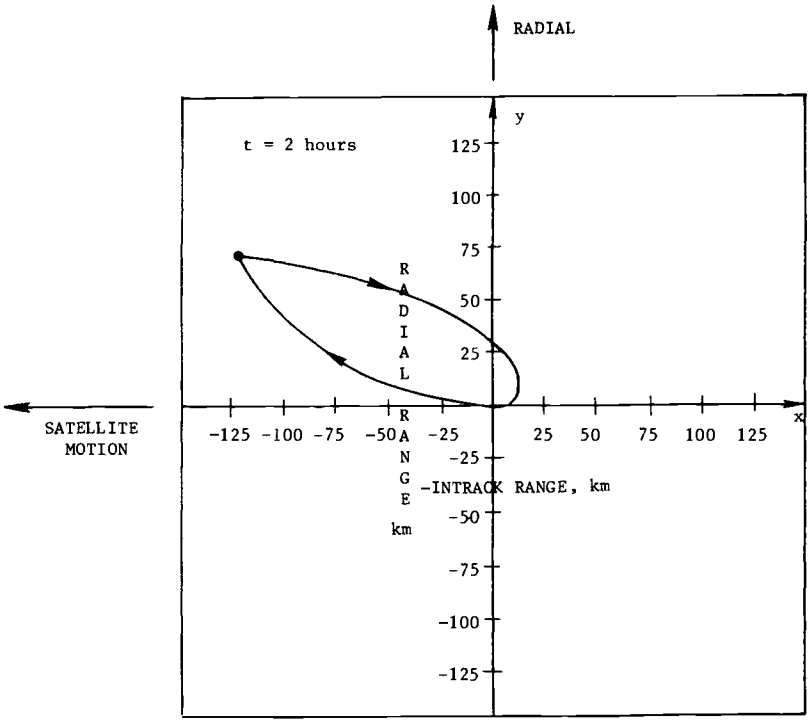


Fig. 7.23 Relative motion for in-track two-impulse solution; starting at t = 2 h and ending at t = 13 h.

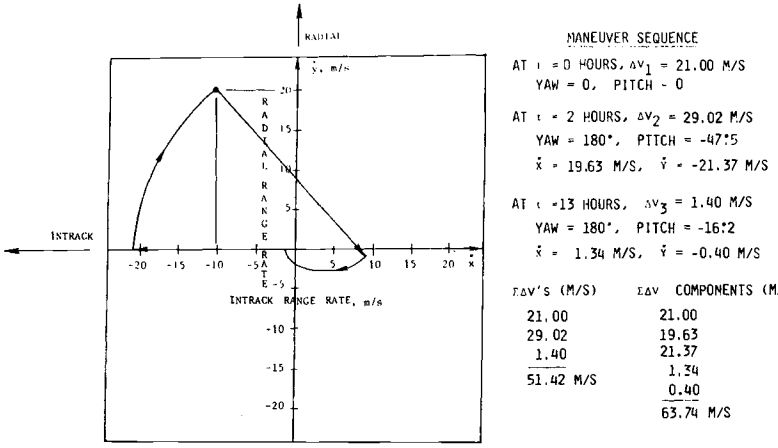


Fig. 7.24 Relative velocity for in-track two-impulse solution, starting at t = 2 h and ending at t = 13.

7.3 Applications of Rendezvous Equations

Co-elliptic Rendezvous

Many space missions require spacecraft rendezvous to dock with another satellite or to perform a rescue or an inspection mission. Typically, a rendezvous mission has a target vehicle in a circular or nearly circular orbit, with the chaser vehicle injected into an orbit of slightly lower altitude. When a specified slant range between the target and chaser vehicle is obtained, the terminal rendezvous phase is initiated. This procedure is often referred to as "co-elliptic rendezvous," for which the conditions

$$\begin{aligned} a_c e_c &= a_t e_t \\ r_{pc} &< r_{pt} \end{aligned} \tag{7.51}$$

are satisfied, where a_c, a_t, are the chaser and target orbit semimajor axes e_c, e_t = eccentricity of each orbit, and r_{pc}, r_{pt} are the respective perigee radii. A typical rendezvous geometry is shown in Fig. 7.25

The basic advantages of co-elliptic rendezvous are: standardized procedure, choice of lighting conditions, star background, and line-of-sight tracking.

Flyaround Maneuvers

After the co-elliptic rendezvous has been performed, a visual inspection of a satellite in orbit may be desired. This can be performed at shorter or longer distances to the satellite with circular, elliptical, or rectilinear trajectories relative to the target vehicle. The elliptical flyaround trajectory can be achieved over an orbit period by a single radial impulse. The ellipse is in the orbit plane whose

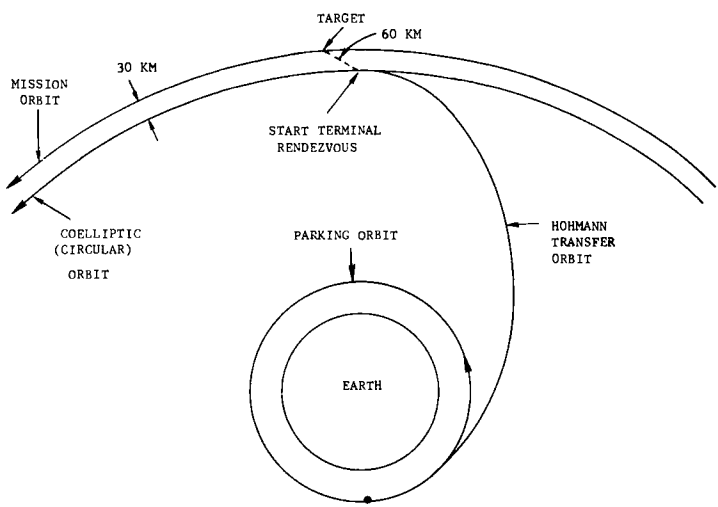


Fig. 7.25 Typical co-elliptic rendezvous geometry.

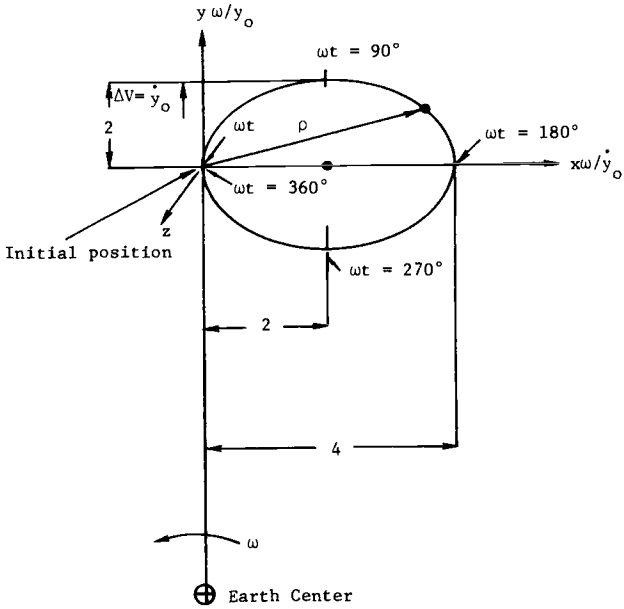


Fig. 7.26 Impulsive flyaround maneuver in orbit plane.

major axis is equal in magnitude to twice its minor axis and is proportional to the magnitude of the applied impulse. The solution is of the form

$$
\begin{aligned}
X &= \frac{x\omega}{\dot{y}_0} = 2(1 - \cos\omega t) \\
Y &= \frac{y\omega}{\dot{y}_0} = \sin\omega t
\end{aligned}
\tag{7.52}
$$

where $\dot{y}_0$ is the radial velocity impulse and ω is the orbital angular velocity. A plot of Eq. (7.52) is shown in Fig. 7.26. The relative displacement with respect to the origin of the coordinate reference frame is

$$
\begin{aligned}
\rho &= (x^2 + y^2)^{1/2} \\
&= \frac{\dot{y}_0}{\omega}[4(1 - \cos\omega t)^2 + \sin^2\omega t]^{1/2}
\end{aligned}
\tag{7.53}
$$

Another approach for close circumnavigation is linear relative translation, which results when the orbit-dependent terms in the rendezvous equations are negligible. The applicable equations are

$$
\ddot{x} = a_x, \qquad \ddot{y} = a_y \tag{7.54}
$$

A circular flyaround at a constant radius can also be performed, which requires a continuous application of thrust to counteract the centrifugal acceleration.

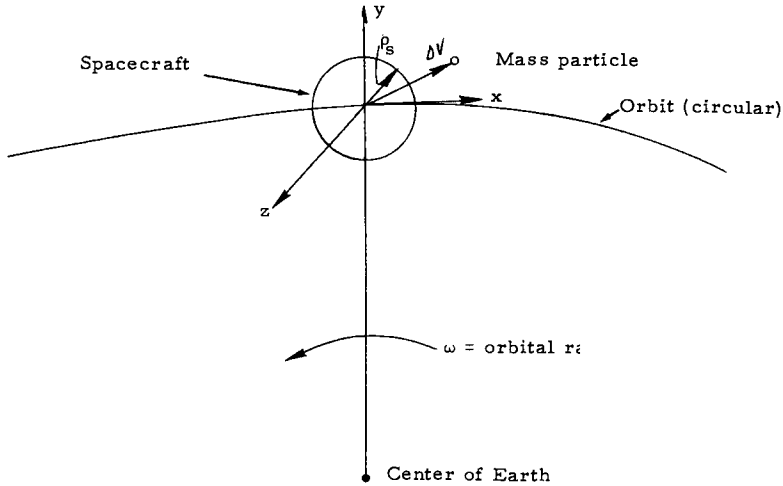


Fig. 7.27 Coordinate system.

Ejected Particle Trajectories

Collision with an ejected particle. The possibility of a spacecraft collision with a particle ejected from a spacecraft is increased greatly if the particle is ejected radially or in a cross-track (out-of-plane) direction. The resulting motion of the particle, to a first-order approximation, is periodic in nature, implying that the particle will return to the ejecting body in an orbit period or a fraction thereof. The in-track (forward or backward) ejection, however, results in a secular increase of the particle distance from the ejecting body, which may or may not be large enough to avoid collision an orbit period later.

Consider the coordinate system attached to a satellite in a circular orbit, as shown in Fig. 7.27, where ρ_s is the radius of the satellite.

For a particle ejected with a relative velocity $\dot{x}_0$, $\dot{y}_0$, or $\dot{z}_0$ along the x, y, or z axes of the reference frame in Fig. 7.27, the trajectory equations from Fig. 7.20 are of the form

$$
\begin{aligned}
x &= \left(-3t + \frac{4}{\omega}\sin\omega t\right)\dot{x}_0 + \frac{2}{\omega}(1-\cos\omega t)\dot{y}_0 \\
y &= \frac{2}{\omega}(-1+\cos\omega t)\dot{x}_0 + \frac{\dot{y}_0}{\omega}\sin\omega t \qquad (7.55)\\
z &= \frac{\dot{z}_0}{\omega}\sin\omega t
\end{aligned}
$$

Solution of Eq. (7.55) leads to the following conclusions:

1) For radial ejection:
   a) Separation is periodic in time.
   b) Maximum separation occurs a half-orbit after ejection.
   c) Separation is reduced to zero upon completion of one orbit.
2) For tangential ejection:
   a) Separation is always finite and variable with time.

b) In-track separation one orbit after ejection is maximum.
c) Separation increases with succeeding orbits.

3) For out-of-plane ejection:
a) Out-of-plane ejection periodic in time (change in orbit inclination).
b) Maximum separation occurs $\frac{1}{4}$ and $\frac{3}{4}$ of an orbit period after ejection.
c) Separation reduced to zero every half-orbit.

For a particle ejected in an arbitrary direction from the satellite, the probability of recontact with the satellite would depend on the magnitudes of the tangential component of velocity $\dot{x}_0$. Thus, for example, one period later, i.e., when $\omega t = 2\pi$, the position of the mass (relative frame x, y, z) is given by the equation

$$x = -\frac{6\pi \dot{x}_0}{\omega} \tag{7.56}$$

This result shows that the mass will be leading (negative x) the spacecraft for a backward ejection at the initial time $t = 0$ and lagging for a forward ejection at $t = 0$.

A sphere of radius $\rho_s = x$ can thus be defined as centered at the coordinate frame (spacecraft origin) that will not be entered by the ejected mass one orbit period later if $|\dot{x}_0| \geq \omega x/6\pi$. Consider now a given ejection velocity $\boldsymbol{\Delta V}$. The x component of $\boldsymbol{\Delta V}$ can be defined as

$$|\dot{x}_0| = \Delta V \cos\beta \tag{7.57}$$

where ΔV is the magnitude of the velocity vector, and β is a half-cone angle measured from the x axis, as shown in Fig. 7.28.

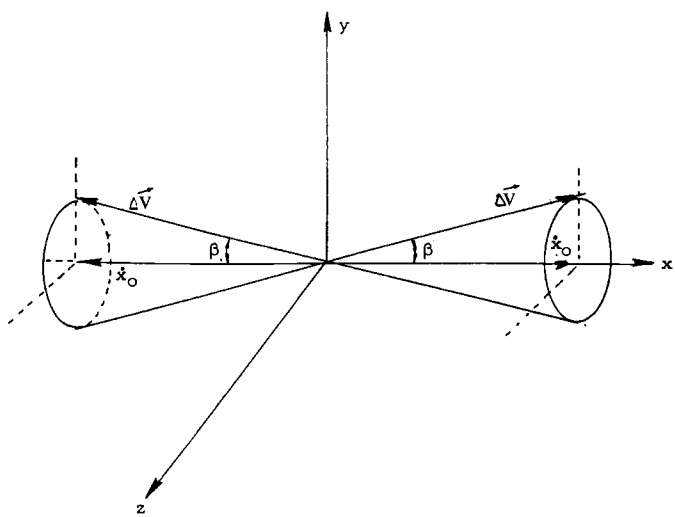


Fig. 7.28 Velocity diagram.

The $|\dot{x}_0| \geq \omega x/6\pi$ condition will be satisfied if and only if $\boldsymbol{\Delta V}$ falls within the cone described by β (either along the positive or negative x axis), and the probability of this occurring can be expressed as

$$P\left(|\dot{x}_0| \geq \frac{\omega x}{6\pi}\right) = \frac{2A_z}{A_s} \tag{7.58}$$

where

$$\begin{aligned} A_z &= \text{an effective area of a spherical zone defined by the cone } \beta \\ &= 2\pi(\Delta V)^2(1 - \cos\beta) \\ A_s &= \text{an effective spherical area} \\ &= 4\pi(\Delta V)^2 \end{aligned}$$

assuming an equal probability of ΔV occurring along any direction. Thus, the probability that a mass initially ejected with a velocity ΔV in an arbitrary direction will be outside a sphere of radius ρ_s one orbital period following the ejection is

$$\begin{aligned} P &= 1 - \cos\beta \\ &= 1 - \frac{|\dot{x}_0|}{\Delta V} \\ &= 1 - \frac{\omega x}{6\pi\,\Delta V} \\ &= 1 - \frac{\omega\rho_s}{6\pi\,\Delta V} \end{aligned} \tag{7.59}$$

The probability that the ejected mass will be within the sphere of radius ρ_s is then

$$\begin{aligned} P_{\rho_s} &= 1 - P \\ &= \frac{\omega\rho_s}{6\pi\,\Delta V} \end{aligned} \tag{7.60}$$

Thus, for example, for a random ejection from a spacecraft with $\Delta V = 10$ m/s and $\rho_s = 100$ m, the probability of recontact (collision) in a 500-km circular orbit within a 100 radius is about 5.8×10^{-4}.

Debris cloud outline. The linearized rendezvous equations presented in Fig. 7.20 can be used to determine the outline of a debris cloud resulting from a breakup of a satellite in orbit. If, for example, it is assumed that the satellite breaks up isotropically; i.e., the individual particles receive a uniform velocity impulse ΔV in all directions, then the position of the particles can be computed as a function of time in an Earth-following coordinate frame attached to the center of mass of the exploding satellite to obtain the outline of the resulting cloud. This can be performed as follows:

Consider an explosion or a collision event in a circular orbit such as the one illustrated in Fig. 7.29. An orbiting orthogonal reference frame xyz is centered at the origin of the event at time $t = 0$ such that x is directed opposite to the

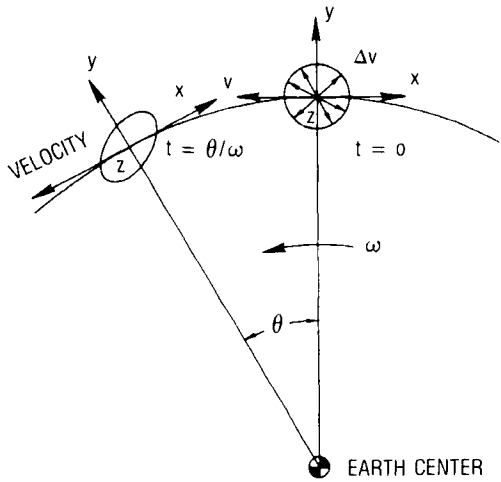


Fig. 7.29 Cloud dynamics.

orbital velocity vector, y is directed along the outward radius, and z completes the triad (along the normal to the orbit plane). The linearized rendezvous equations (7.55) can be used to determine the position of a particle leaving the origin of the coordinate frame with a velocity ΔV; they are of the form

$$
\begin{aligned}
x &= \left(\frac{-3\theta}{\omega} + \frac{4}{\omega}\sin\theta\right)\dot{x}_0 + \frac{2}{\omega}(1-\cos\theta)\dot{y}_0 \\
y &= \frac{2}{\omega}(\cos\theta - 1)\dot{x}_0 + \frac{\dot{y}_0}{\omega}\sin\theta \\
z &= \frac{\dot{z}_0}{\omega}\sin\theta
\end{aligned}
\tag{7.61}
$$

where $\Delta V = (\dot{x}_0^2 + \dot{y}_0^2 + \dot{z}_0^2)^{1/2}$ and $\theta = \omega t$.

The x, y, z coordinates represent particle position at time $t = \theta/\omega$, where θ is the in-orbit plane angle, and ω is the angular rate of the circular orbit. The $\dot{x}_0, \dot{y}_0, \dot{z}_0$ terms are initial velocity components imparted to the particle along the x, y, z axes, respectively. It is assumed that $\Delta V \ll V$, where V is the orbital velocity of the reference frame.

Equations (7.61) can be normalized with respect to $\Delta V/\omega$ as follows:

$$
\begin{aligned}
\frac{x\omega}{\Delta V} &= (-3\theta + 4\sin\theta)h + 2(1-\cos\theta)r \\
&= X \\
\frac{y\omega}{\Delta V} &= 2(\cos\theta - 1)h + (\sin\theta)r \\
&= Y \\
\frac{z\omega}{\Delta V} &= n\sin\theta \\
&= Z
\end{aligned}
\tag{7.62}
$$

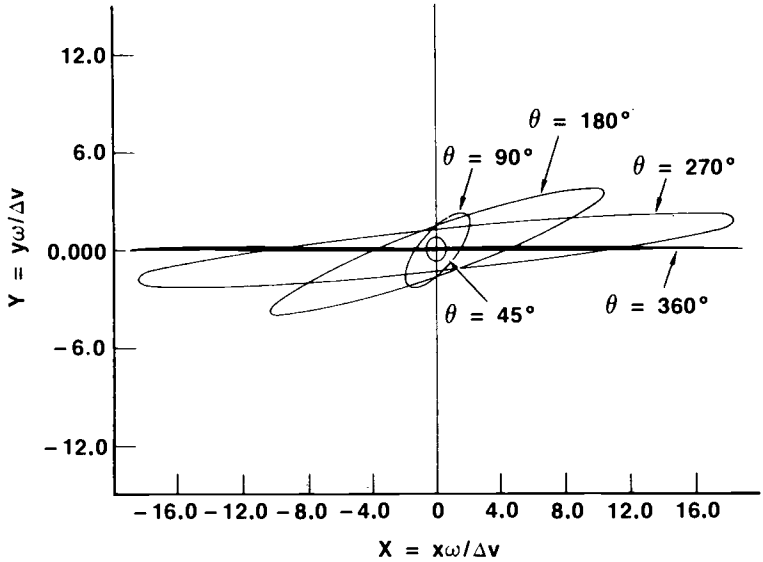


Fig. 7.30 Cloud contours in orbit plane.

where

$$h = \dot{x}_0/\Delta V, \qquad r = \dot{y}_0/\Delta V, \qquad n = \dot{z}_0/\Delta V, \qquad h^2 + r^2 + n^2 = 1$$

Equations (7.62) can be plotted as a function of θ for different values of h, r, and n. If, for example, the initial velocity ΔV distribution for the particles is circular in the x, y plane, then $h = \cos\phi$, $r = \sin\phi$, $0 \leq \phi \leq 360$ deg, and $h^2 + r^2 = 1$ with $n = 0$. The resultant cloud outline is illustrated in Fig. 7.30 for several values of θ.

The accuracy of the results degrades somewhat as time and ΔV increase compared to the orbital velocity V. The results in Fig. 7.30 are representative of the outline of the debris cloud and can be used to compute the volume of the cloud and the resultant collision hazard to orbiting objects in the vicinity of the cloud, as in Ref. 21, for example.

Acceleration and Velocity Impulse Requirements for a Radial Transfer Trajectory

An orbit-transfer maneuver may be required in which a satellite is transferred from one circular orbit to another along the radius vector. For example, it may be of interest to consider the case in which the orbital transfer is radially outward from an initial offset distance d to a smaller offset distance δ measured in an Earth-pointing, rotating coordinate frame, as shown in Fig. 7.31.

An approximate solution of the problem can be obtained using the linearized rendezvous equations for circular orbits. In this case, the satellite m at time $t = 0$ is located at a radial offset distance d and, at a later time, $t = T$ is a radial offset distance δ. The necessary external thrust accelerations a_x and a_y may be derived and integrated to obtain the total velocity impulses required.

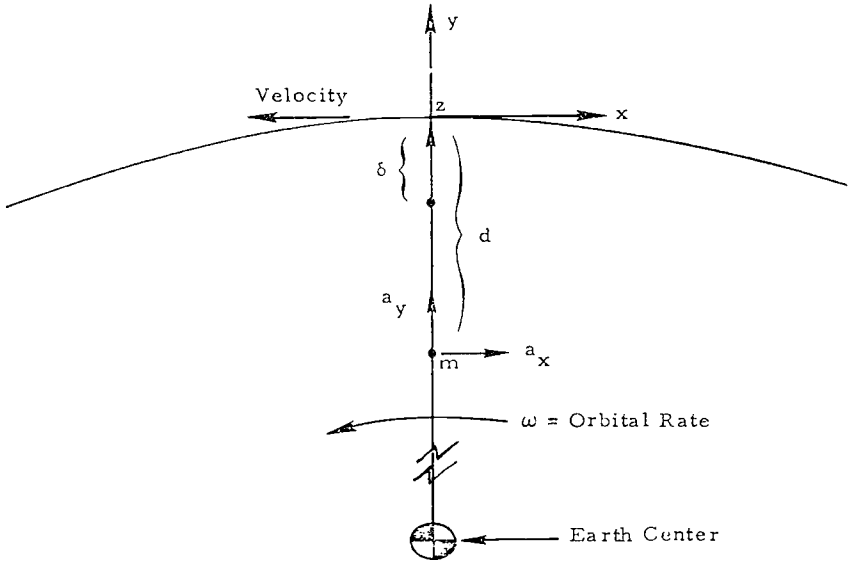


Fig. 7.31 Radial transfer geometry.

The exact nonlinear differential equations governing the motion of a mass m relative to an Earth-oriented reference frame may be expressed as

$$\begin{aligned}
&\ddot{x} - 2\omega\dot{y} + x\omega^2\left[\left(\frac{r_s}{r_f}\right)^3 - 1\right] = \frac{T_x}{m} \\
&\ddot{y} + 2\omega\dot{x} + (y + r_s)\omega^2\left[\left(\frac{r_s}{r_f}\right)^3 - 1\right] = \frac{T_y}{m} \\
&\ddot{z} + \omega^2\left(\frac{r_s}{r_f}\right)^3 = \frac{T_z}{m}
\end{aligned} \tag{7.63}$$

where

T_x, T_y, T_z = external forces
ω = circular orbit rate
r_s = constant radius of the rotating reference frame
$r_f = [x^2 + (y + r_s)^2 + z^2]^{1/2}$
= radial distance to the mass
x, y, z = negative in-track, radial outward, and out-of-plane displacement, respectively

If the $(r_s/r_f)^3$ term is approximated as $(r_s/r_f)^3 \approx 1 - 3y/r_s$ and only planar motion is considered, then Eqs. (7.63) become

$$\begin{aligned}
&\ddot{x} - 2\omega\dot{y} = T_x/m = a_x \\
&\ddot{y} + 2\omega\dot{x} - 3\omega^2 y = T_y/m = a_y
\end{aligned} \tag{7.64}$$

where a_x, a_y are the negative in-track and radially outward accelerations applied to the mass.

For the case of radial transfer, a solution of Eqs. (7.64) can be obtained assuming that $x = \dot{x} = \ddot{x} = 0$ at all times and that

$$a_y = a_{y_0} = \text{const} \qquad \text{for} \quad 0 \le t \le t_1 \tag{7.65}$$

and

$$a_y = 0 \qquad \text{for} \quad t_1 \le t \le T \tag{7.66}$$

where T is the transfer time. Thus, for case (7.65), Eqs. (7.64) become

$$-2\omega\dot{y} = a_x \tag{7.67}$$

$$\ddot{y} - 3\omega^2 y = a_{y_0} \tag{7.68}$$

where Eq. (7.67) represents the Coriolis acceleration resulting from the mass m moving radially outward in a rotating reference frame.

The solution of Eqs. (7.67) and (7.68) is of the form

$$y = Ae^{nt} + Be^{-nt} - \frac{a_{y_0}}{n^2} \tag{7.69}$$

where $n^2 = 3\omega^2$ and A, B are constants to be determined. For this case,

$$\dot{y} = Ane^{nt} - Bne^{-nt} \tag{7.70}$$

$$\ddot{y} = An^2e^{nt} + Bn^2e^{-nt} \tag{7.71}$$

and since, at $t = 0$, $y = -d$, $\dot{y} = 0$,

$$0 = An - Bn$$

$$-d = 2A - \frac{a_{y_0}}{n^2}$$

which yields

$$A = B = \frac{1}{2}\left(\frac{a_{y_0}}{n^2} - d\right) \tag{7.72}$$

Consequently,

$$\begin{aligned} y &= \left(\frac{a_{y_0}}{n^2} - d\right)\cosh nt - \frac{a_{y_0}}{n^2} \\ &= \frac{a_{y_0}}{3\omega^2}(\cosh\sqrt{3}\omega t - 1) - d\cosh\sqrt{3}\omega t \end{aligned} \tag{7.73}$$

$$\begin{aligned} \dot{y} &= n\left(\frac{a_{y_0}}{n^2} - d\right)\sinh nt \\ &= \left(\frac{a_{y_0} - 3\omega^2 d}{\sqrt{3}\omega}\right)\sinh\sqrt{3}\omega t \end{aligned} \tag{7.74}$$

$$\ddot{y} = n^2\left(\frac{a_{y_0}}{n^2} - d\right)\cosh nt$$

$$= (a_{y_0} - 3\omega^2 d)\cosh\sqrt{3}\omega t \tag{7.75}$$

The in-track acceleration a_x is then

$$\begin{aligned} a_x &= -2\omega\dot{y} \\ &= -\frac{2}{\sqrt{3}}(a_{y_0} - 3\omega^2 d)\sinh\sqrt{3}\omega t \end{aligned} \tag{7.76}$$

For the case $t_1 \leq t \leq T$ when outward radial acceleration (thrusting) is zero, the specific solution of Eqs. (7.67) and (7.68) can be obtained from the conditions.

$$\begin{aligned} \dot{y}(T) &= Ane^{nT} - Bne^{-nT} \rightarrow B = Ae^{2nT} \\ &= 0 \\ y(T) &= Ae^{nT} + Ae^{2nT}e^{-nT} \\ &= 2A(e^{nT}) = -\delta \rightarrow A = -\frac{\delta}{2}e^{-nT} \end{aligned}$$

Thus, for $t_1 \leq t \leq T$,

$$\begin{aligned} y &= -\frac{\delta}{2}\{\exp[n(t-T)] + \exp[-n(t-T)]\} \\ &= -\delta\cosh n(t-T) \end{aligned} \tag{7.77}$$

$$\begin{aligned} \dot{y} &= -\delta n\sinh n(t-T) \\ &= -\delta\sqrt{3}\omega\sinh\sqrt{3}\omega(t-T) \end{aligned} \tag{7.78}$$

$$\begin{aligned} \ddot{y} &= -\delta n^2\cosh n(t-T) \\ &= -3\omega^2\delta\cosh\sqrt{3}\omega(t-T) \end{aligned} \tag{7.79}$$

Consequently, for $t_1 \leq t \leq T$, $a_y = 0$ and

$$\begin{aligned} a_x &= -2\omega\dot{y} \\ &= (2n^2\delta/\sqrt{3})\sinh n(t-T) \end{aligned} \tag{7.80}$$

The total (combined) acceleration for $0 \leq t \leq t_1$ is

$$\begin{aligned} a_T &= \left(a_x^2 + a_y^2\right)^{1/2} \\ &= \left[\frac{4}{3}\left(dn^2 - a_{y_0}\right)^2\sinh^2 nt + a_{y_0}^2\right]^{1/2} \end{aligned} \tag{7.81}$$

The velocity impulse (ΔV) is given by

$$\Delta V = \int_0^T a_T \, \mathrm{d}t \tag{7.82}$$

The actual trajectory of the satellite, as obtained by integrating Eq. (7.63), will always contain an in-track ($-x$ direction) component. The magnitude of the in-track component can, however, be controlled by varying slightly the radial acceleration component a_{y_0}.

7.4 An Exact Analytical Solution for Two-Dimensional Relative Motion

Introduction

Interesting and worthwhile solutions for the relative motion of a probe, ejected into an elliptic orbit in the orbital plane of a space station that is in a circular orbit, are derived by Berreen and Crisp in Ref. 9. They have developed an exact analytical solution by coordinate transformation of the known orbital motions to rotating coordinates.

However, there are three restrictions on the solution of Berreen and Crisp that should be noted: 1) The probe is ejected from the space station at time $t = 0$ with relative velocity components x_0' and y_0' but the equations as derived do not permit an initial relative displacement such as $\rho_0 = (x_0^2 + y_0^2)^{1/2}$. Generalized equations that permit an initial relative displacement will be derived here. 2) The motion of the probe is restricted to the orbit plane of the space station and is, therefore, two-dimensional. 3) The space station is assumed to be in a circular orbit. As stated previously, restriction 1 will be relaxed in this section by the derivation of orbit element equations for the probe in terms of arbitrary initial relative velocity and displacement components for the probe with respect to the space station. The relaxation of restrictions 2 and 3 will be the subject of future studies.

Geometry and Coordinate Systems

Using the notation and description of Berreen and Crisp,[9] consider first the coordinate systems of Fig. 7.32. The space station–centered system is X, Y; the X_i, Y_i system is a geocentric inertial system: and the $X_e.Y_e$ system is a geocentric rotating system having its Y_e axis always passing through the space station. Coordinates R_p, θ_P and R_p, α are polar coordinates of the probe in the X_i, Y_i and X_e, Y_e systems, respectively.

Uppercase letters are used henceforth for real distances and velocities: and lowercase letters are used for ratios of distance and velocity, respectively, to the station orbital radius R_s and circular orbit velocity V_s. Thus,

$$x = \frac{X}{R_s}, \quad r_p = \frac{R_p}{R_s}, \quad v_p = \frac{V_p}{V_s} \tag{7.83}$$

where

$$V_s = \sqrt{\frac{\mu}{R_s}} \tag{7.84}$$

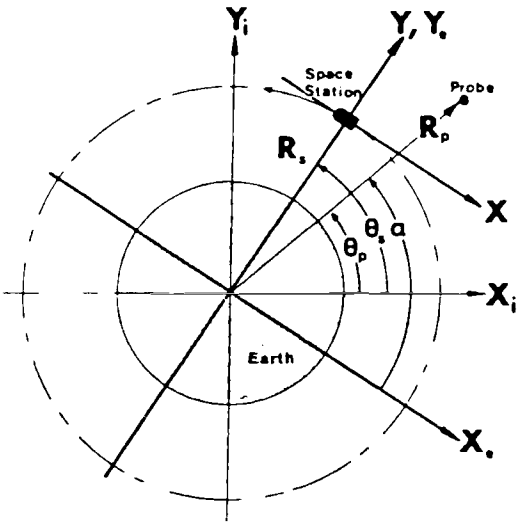


Fig. 7.32 The rectangular coordinate systems $(X_i, Y_i), (X_e, Y_e)$, and (X, Y) and the polar coordinates $(R_s, \theta_s), (R_p, \theta_p)$, and (R_p, α) (from Ref. 9).

and μ is the gravitational constant for the Earth. Subscripts s and p refer to station and probe, respectively.

The mean motion N_s of the station is

$$N_s = \frac{V_s}{R_s} \tag{7.85}$$

The angular coordinate θ_s of the station is then

$$\theta_s = N_s t \tag{7.86}$$

with initial conditions defined at $t = 0$.

Orbital Relations from Berreen and Crisp

In an inertial coordinate system, the probe moves in a Keplerian orbit described by the elements e_p, the eccentricity p_p, the semilatus rectum ratioed to R_s, and the apsidal orientation θ_o^*. Berreen and Crisp[9] describe these elements in terms of the initial relative velocity ratio components x_0' and y_0' by the equations

$$p_p = (1 - x_0')^2 \tag{7.87}$$

$$e_p^2 = 1 + p_p v_p^2 - \frac{2}{r_p} = 1 + (1 - x_0')^2\left[y_0'^2 + (1 - x_0')^2 - 2\right] \tag{7.88}$$

and

$$\theta_p^* = \cos^{-1}[(p_p - 1)/e_p] = -\sin^{-1}[(1 - x_0')y_0'/e_p] \tag{7.89}$$

with $-\pi < \theta_p^* \leq \pi$. The next subsection will generalize these equations to include initial relative displacement components x_0 and y_0, as well as x_0' and y_0'.

Derivation of Generalized Orbital Relations

Figure 7.33 depicts arbitrary, but still coplanar, initial conditions for the probe and space station.

From the geometry of Fig. 7.33,

$$r_p = \left[(1 + y_0)^2 + x_0^2\right]^{1/2} = \left(1 + 2y_0 + x_0^2 + y_0^2\right)^{1/2} \tag{7.90}$$

and

$$v_p = \left[(1 - x_0')^2 = y_0'^2\right]^{1/2} = \left(1 - 2x_0' + x_0'^2 + y_0'^2\right)^{1/2} \tag{7.91}$$

From conservation of angular momentum,

$$p_p = r_p^2 v_{p_{\text{horizontal}}}^2 \tag{7.92}$$

where

$$v_{p_{\text{horizontal}}} = 1 - x_0' \tag{7.93}$$

Substituting Eqs. (7.90) and (7.93) into Eq. (7.92).

$$p_p = \left(1 + 2y_0 + x_0^2 + y_0^2\right)(1 - x_0')^2 \tag{7.94}$$

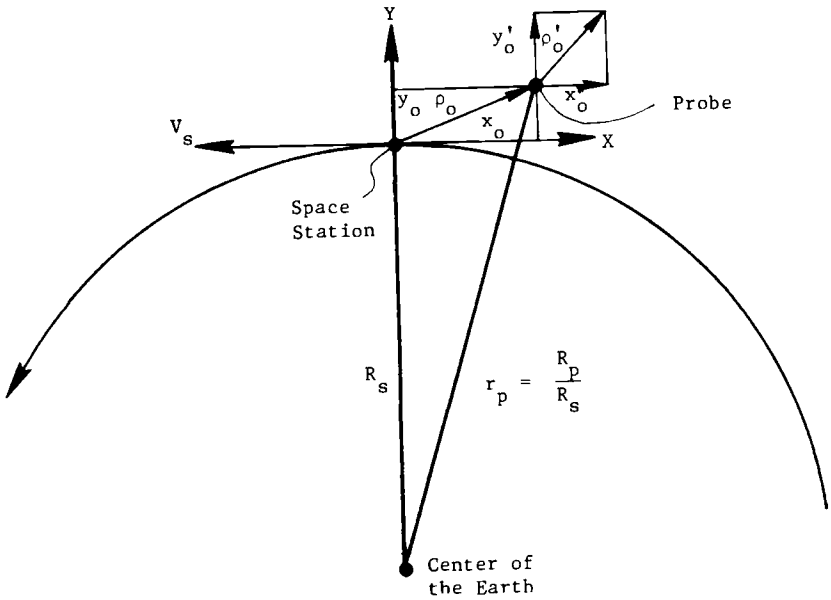


Fig. 7.33 Initial conditions for the probe and space station.

Equation (7.94) is the generalization of Eq. (17) in Berreen and Crisp[9] and reduces to their equation when $x_0 = y_0 = 0$.

From conservation of energy,

$$v_p^2 = \frac{2}{r_p} - \frac{1}{a_p} \tag{7.95}$$

Substituting a_p into the eccentricity equation yields

$$e_p^2 = 1 - \frac{p_p}{a_p} = 1 + p_p v_p^2 - \frac{2}{r_p} \tag{7.96}$$

Substituting Eqs. (7.90) and (7.91) into Eq. (7.96),

$$e_p^2 = 1 + p_p\left[1 - 2x_0' + x_0'^2 + y_0'^2 - \frac{2}{\left(1 + 2y_0 + x_0^2 + y_0^2\right)^{1/2}}\right] \tag{7.97}$$

Equation (7.94) may be used to substitute for p_p. Equation (7.97) is the generalization of Eq. (18) in Berreen and Crisp.[9]

The orbit equation is

$$r_p = \frac{p_p}{1 + e_p \cos \beta_p} \tag{7.98}$$

where β_p is the true anomaly.

The apsidal orientation $\theta_p^* = -\beta_p$, so that

$$\theta_p^* = -\cos^{-1}\left[\frac{(p_p/r_p) - 1}{e_p}\right] \tag{7.99}$$

Equations (7.90), (7.94), and (7.97) may be used to substitute for r_p, p_p, and e_p, respectively. Equation (7.99) is the generalization of Eq. (19) in Berreen and Crisp.[9] Thus, exact equations for p_p, e_p, and θ_p^* have been derived in terms of the initial velocity components x_0' and y_0' and the initial displacement components x_0 and y_0 relative to the space station.

Exact Polar Equations of the Trajectory

Berreen and Crisp[9] derive exact equations for the motion of the probe with respect to the space station in terms of the polar coordinates r_p and α (see Fig. 7.32). The equation for r_p is simply the orbit equation expressed in terms of the true anomaly β_p or the eccentric anomaly E_p,

$$r_p = \frac{p_p}{1 + e_p \cos \beta_p} = a_p(1 + e_p \cos E_p) \tag{7.100}$$

where

$$a_p = \frac{p_p}{1 - e_p^2} \tag{7.101}$$

The variables β_p and E_p are related by

$$\tan\frac{E_p}{2} = \left(\frac{1 - e_p}{1 + e_p}\right)^{1/2} \tan\frac{\beta_p}{2} \tag{7.102}$$

Of course, Kepler's equation must be used to relate time t and E_p:

$$N_p(t - t^*) = E_p - e_p \sin E_p \tag{7.103}$$

where t^* is the time of perigee passage for the probe, and N_p is the mean motion of the probe given by Kepler's third law,

$$\frac{N_s}{N_p} = a_p^{3/2} \tag{7.104}$$

Then, Berreen and Crisp[9] derive an expression for the polar angle γ,

$$\gamma = \beta_p - \frac{N_s}{N_p}(E_p - e_p \sin E_p) \tag{7.105}$$

where $\gamma \equiv \alpha - \alpha^*$ and $\alpha = \alpha^*$ at $t = t^*$. The equation for γ is derived quite directly from

$$\alpha = \frac{\pi}{2} - (\theta_s - \theta_p) \tag{7.106}$$

which is evident from Fig. 7.32.

Thus, Eqs. (7.100) and (7.105) are exact equations in polar coordinates r_p and γ of the probe motion relative to the space station.

Calculation Algorithm for the Exact Solution

An algorithm for calculating r_p and γ is outlined as follows:

1) The radius R_s of the space station orbit is given. The circular orbit velocity V_s is calculated from Eq. (7.84). The initial relative displacement and velocity components of the probe are given. These are normalized by dividing by R_s and V_s to obtain x_0, y_0, x_0', and y_0'.
2) Calculate p_p from Eq. (7.94).
3) Calculate e_p from Eq. (7.97).
4) Calculate a_p from Eq. (7.101).
5) Calculate N_s/N_p from Eq. (7.104).
6) Calculate r_p at $t = 0$ from Eq. (7.90).
7) Calculate E_p at $t = 0$ from Eq. (7.100).

8) Calculate β_p at $t = 0$ from Eq. (7.102).
9) Calculate γ at $t = 0$ from Eq. (7.105).

Then, for other times,

10) Calculate E_p from Eq. (7.03), where N_s is calculated from Eq. (7.85) and N_p from Eq. (7.104).
11) Calculate β_p from Eq. (7.102).
12) Calculate r_p from Eq. (7.100).
13) Calculate γ from Eq. (7.05).

7.5 Optimal Multiple-Impulse Rendezvous

Two-Impulse Time-Fixed Rendezvous

Two-impulse time-fixed rendezvous between satellites in neighboring orbits was investigated in Sec. 7.2. Linearized relative equations of motion were solved for given boundary conditions and a specified time to obtain the two vector impulses required. No optimization was involved; although, in the example headed "Two-Impulse Rendezvous Maneuver" in Sec. 7.2, the time for rendezvous was varied so that a tradeoff of total ΔV vs rendezvous time could be developed. A solution for a minimum total ΔV was identified in the tradeoff. Succeeding sections will identify optimal solutions involving two, three, and four impulses.

Optimal Multiple-Impulse Rendezvous Between Satellites in Neighboring Orbits

A landmark development in optimal multiple-impulse rendezvous was the doctoral thesis of John Prussing in 1967.[10] Prussing considered neighboring nearly circular orbits. He applied Lawden's theory of the primer vector to the equations of motion, linearized about an intermediate circular orbit. Descriptions of the primer locus are used to develop two-, three-, and four-impulse optimal solutions. Figure 7.34 is Fig. 9.4 of Ref. 10. It presents $\Delta V/\delta R$ vs t_F for $\beta/\delta R = 81.4$ deg, where

ΔV = the sum of the impulses
δR = the nondimensional difference between the final and initial circular orbit radii
t_F = time measured in reference orbit periods
β = $\beta_H + \Delta\beta = \theta_H + \Delta\theta$ of Fig. 7.1

Figure 7.34 shows that two impulses are optimal for small values of t_F. As t_F increases, three- and four-impulse solutions become optimal. The indicator 3_F^+ means that a three-impulse solution with a final coast is optimal. As time increases, the Hohmann ΔV cost is finally achieved.

Fig. 7.35, which is Fig. 9.1 of Ref. 10, presents final state variations for circle-to-circle coplanar rendezvous, which are reached optimally using different numbers of impulses, where

$$\delta\theta_F = \beta - \frac{3}{4}\delta R t_F \tag{7.107}$$

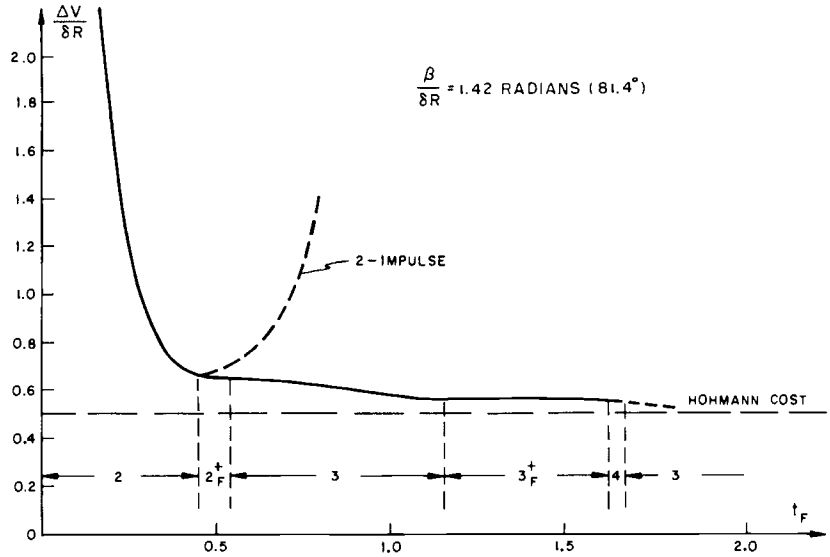


Fig. 7.34 Optimal cost as a function of transfer time for given initial conditions (from Ref. 10).

Figure 7.35 summarizes the results of Ref. 10. These results are also described in Refs. 11 and 12.

Gobetz and Doll[13] provide an excellent survey article of orbit transfers and rendezvous maneuvers. They describe the problems and solutions, and present graphic results and a bibliography of 316 papers, articles, and reports.

Optimal Multiple-Impulse Nonlinear Orbital Rendezvous

Reference 14 extends the circle-to-circle solutions of the preceding subsection to the nonlinear case, i.e., in which the difference in orbit radii may be large. Again, primer vector theory is used to obtain the optimal number of impulses, their times and positions, and the presence of initial or final coasting arcs. Reference 14 is the journal article version of Ref. 15.

Figure 7.36 which is Fig. 4 of Ref. 14, presents total ΔV vs time in reference orbit periods for a final/initial-orbit radius ratio of 1.6 and for $\beta =$ 0, 90, 180, 270 deg. The ΔV curves all decrease with time and eventually reach the Hohmann ΔV cost. The optimal number of impulses are designated at various points on these curves.

An interesting example presented in Refs. 14 and 15 is the optimal trajectory for rendezvous with a target satellite in the same circular orbit at a phase angle β of 180 deg in a specified time of 2.3 reference orbit periods. Figure 7.37, which is Fig. 5.14 of Ref. 15, depicts the trajectory, the locations of the four optimal impulses, and a tabular listing of the ΔV magnitudes and their application times. The total cost is $\Delta V_T/V_c = 0.189$. By comparison, the cost of the best two-impulse rendezvous is $\Delta V_T/V_c = 0.224$. This is the cost that would be obtained by the technique described in the subsection headed "In-Orbit Repositioning" in Sec. 7.1.

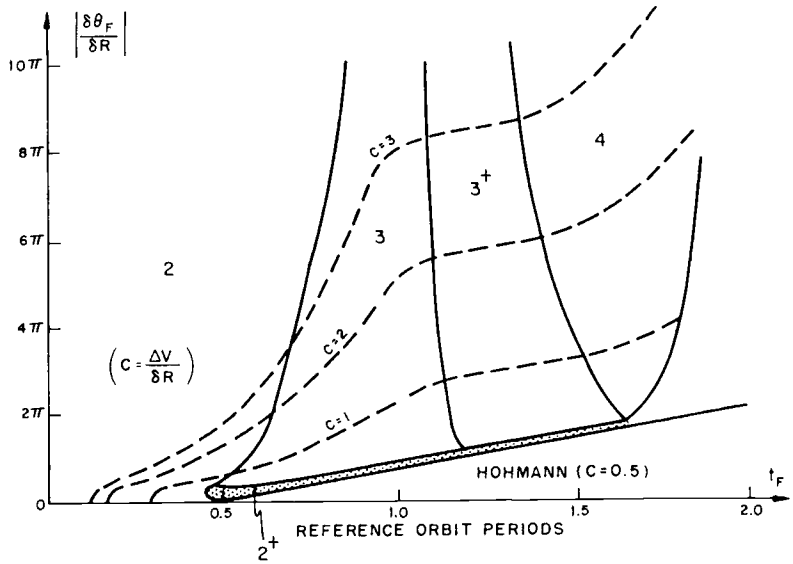


Fig. 7.35 Reachable final state variations; optimal multiple-impulse solutions (from Ref. 10).

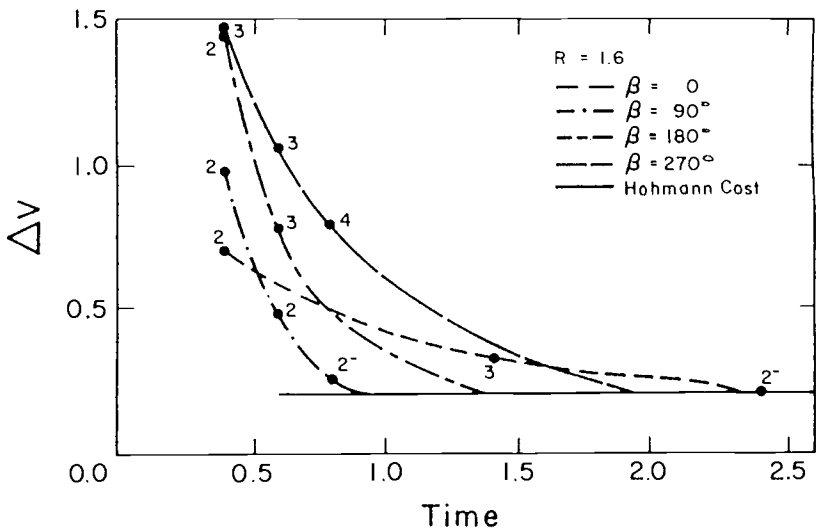


Fig. 7.36 ΔV vs time plot for R = 1.6 (from Ref. 14).

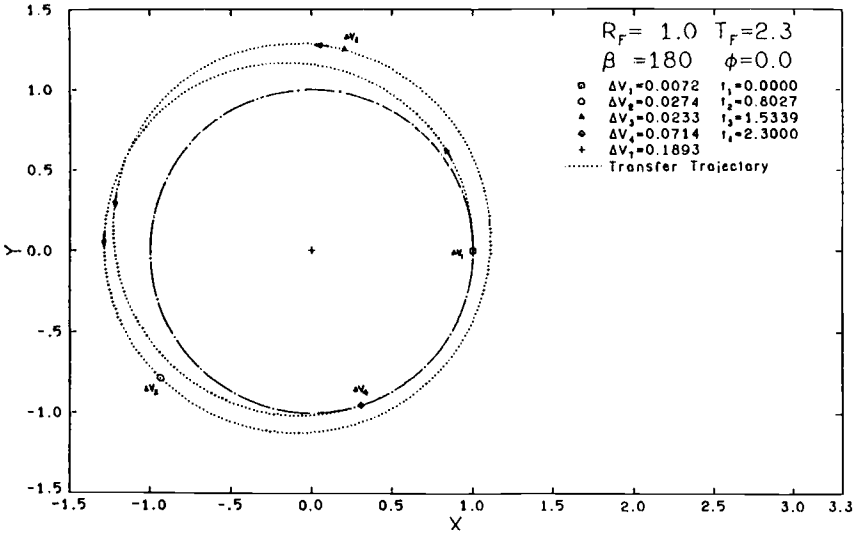


Fig. 7.37 Example rendezvous trajectory with four impulses (from Ref. 15).

Other Optimal Multiple-Burn Rendezvous-Type Maneuvers

Other optimal rendezvous-type maneuvers have been developed for important orbital applications. Some of these are briefly summarized here chronologically.

The paper by Holaday and Swain[16] entitled "Minimum-Time Rescue Trajectories Between Spacecraft in Circular Orbits" uses the nonlinear equations of motion. The multipoint boundary-value problem was solved using the indirect shooting technique with a modified Newton's method for convergence. All of the illustrated minimum-time trajectories describe thrust-coast-thrust histories. If sufficient fuel is available, the optimal solution may utilize continuous maximum thrust. The effects of an altitude constraint are examined. The effects of phase angle, specific impulse, and maximum thrust on rendezvous time are displayed.

Optimal Impulsive Time-Fixed Direct-Ascent Interception

Prussing and Clifton[17] obtained minimum-fuel, impulsive solutions for the problems of attack avoidance on a satellite, followed by a return to the original orbit station. The evasion distance and time are constrained. Both constrained and free final-time cases are considered. Primer vector theory is used to obtain the optimal solutions presented, which include three-impulse solutions for an arbitrarily specified evasion radius vector and two-impulse free-return trajectories for certain evasion radius vectors. Primer vector histories are displayed for a number of examples.

Optimal Trajectories for Time-Constrained Rendezvous Between Arbitrary Conic Orbits

Wellnitz and Prussing[18] generate optimal impulsive trajectories for time-constrained rendezvous between arbitrary conic orbits. Primer vector theory is used to

determine how the cost in ΔV can be minimized by the addition of initial and final costs and by the addition of midcourse impulses.

A universal variable formulation was used. Results are presented for a rendezvous between unaligned coplanar elliptical orbits and for coplanar and inclined elliptical rescue missions.

Optimal Cooperative Time-Fixed Impulsive Rendezvous

In Ref. 19, Mirfakhrale, Conway, and Prussing developed a method for determining minimum-fuel trajectories of two satellites. The method assumes that the satellites perform a total of three impulsive maneuvers, with each satellite being active, i.e., performing at least one maneuver. The method utilizes primer vector theory and analytical expressions for the gradient of the total amount of fuel expended. Results are presented for a number of cases and demonstrate the advantage of performing a rendezvous cooperatively for certain initial geometries and times of flight.

Optimal Orbital Rendezvous Using High and Low Thrust

In Ref. 20, Larson, and Prussing use optimal-control theory to examine a specific class of satellite trajectory problems where high- and low-thrust propulsion systems are used. These problems assume that a satellite is in an established orbit about a planet. An intercept of a predetermined position in space in a specified amount of time using an optimal high-thrust program is then executed. Finally, the satellite returns to the original orbit station using the low-thrust propulsion system in an optimal fashion. Solutions are obtained for problems with a fixed final time. Results for two examples are presented.

References

[1]Billik, B. H., and Roth, H. L., "Studies Relative to Rendezvous Between Circular Orbits," *Astronautica Acta,* Vol. 12, Jan.–Feb. 1967, pp. 23–26.

[2]Pitkin, E. T., "A General Solution of the Lambert Problem," *The Journal of the Astronautical Sciences,* Vol. 15, 1968, pp. 270–271.

[3]Lancaster, E. R., and Blanchard, R. C., "A Unified Form of Lambert's Theorem," NASA Technical Note D-5368, 1969.

[4]Herrick, S., *Astrodynamics,* Vol. 1, Van Nostrand, 1971.

[5]Battin, R. H., *An Introduction to the Mathematics and Methods of Astrodynamics,* AIAA, New York, 1987.

[6]Gooding, R. H., "A Procedure for the Solution of Lambert's Orbital Boundary-Value Problem," *Celestial Mechanics and Dynamical Astronomy,* 48, 1990, pp. 145–165.

[7]Prussing, J. E., and Conway, B. A., *Orbital Mechanics,* Oxford University Press, 1993.

[8]Hanson, J. T., "Optimal Maneuvers of Orbital Transfer Vehicles," Ph.D. Dissertation, Univ. of Michigan, Ann Arbor, MI, 1983.

[9]Berreen, T. F., and Crisp, J. D. C., "An Exact and a New First-Order Solution for the Relative Trajectories of a Probe Ejected from a Space Station," *Celestial Mechanics,* Vol. 13, 1976, pp. 75–88.

[10]Prussing, J. E., "Optimal Multiple-Impulse Orbital Rendezvous," Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, 1967.

[11]Prussing, J. E., "Optimal Four-Impulse Fixed-Time Rendezvous in the Vicinity of a Circular Orbit," *AIAA Journal,* Vol. 7, May 1969, pp. 928–935.

[12]Prussing, J. E., "Optimal Two- and Three-Impulse Fixed-Time Rendezvous in the Vicinity of a Circular Orbit," *AIAA Journal,* Vol. 8, July 1970, pp. 1221–1228.

[13]Gobetz, F. W., and Doll, J. R., "A Survey of Impulsive Trajectories," *AIAA Journal,* Vol. 7, May 1969, pp, 801–834.

[14]Prussing, J. E., and Chiu, J.-H., "Optimal Multiple-Impulse Time-Fixed Rendezvous Between Circular Orbits," *Journal of Guidance, Control, and Dynamics,* Vol. 9, Jan.–Feb. 1986, pp. 17–22.

[15]Chiu, J.-H., "Optimal Multiple-Impulse Nonlinear Orbital Rendezvous," , Ph.D. Thesis, Univ. of Illinois at Urbana-Champaign, IL, 1984.

[16]Holaday, B. H., and Swain, R. L., "Minimum-Time Rescue Trajectories Between Spacecraft in Circular Orbits," *Journal of Spacecraft and Rockets,* Vol. 13, July 1976, pp. 393–399.

[17]Prussing, J. E., and Clifton, R. S., "Optimal Multiple-Impulse Satellite Avoidance Maneuvers," AAS Paper 87-543, Aug. 1987.

[18]Wellnitz, L. J., and Prussing, J. E., "Optimal Trajectories for Time-Constrained Rendezvous Between Arbitrary Conic Orbits," AAS Paper 87-539, Aug. 1987.

[19]Mirfakhrale, K., Conway, B. A., and Prussing, J. E., "Optimal Cooperative Time-Fixed Impulsive Rendezvous," AIAA Paper 88-4279-CP, Aug. 1988.

[20]Larson, C. A., and Prussing, J. E., "Optimal Orbital Rendezvous Using High and Low Thrust," AAS Paper 89-354, Aug. 1989.

[21]Chobotov, V. A., "Dynamics of Orbital Debris Clouds and the Resulting Collision Hazard to Spacecraft," *Journal of the British Interplanetary Society,* Vol. 43, May 1990, pp. 187–194.

Problems

7.1. Assume an Earth satellite in circular orbit, altitude = 278 km and period = 90 min. Assume two-body motion, i.e., no atmospheric drag, no sun, moon perturbations, etc.

At an arbitrary time, a small free-flying experiment package is ejected with a small ΔV from the satellite. Formulate the equations of relative motion for $X/\Delta V$ and $Y/\Delta V$ for two cases: 1) the ΔV is applied in the direction of satellite motion; and 2) the ΔV is applied radially outward. Then, plot $Y/\Delta V$ vs $X/\Delta V$for one period of satellite motion (90 min) for both cases. How far will the free flyer be from the satellite after 90 min?

7.2. A space station is in a 90-min period circular orbit around a spherical, atmosphereless Earth. At $t = 0$, a nearby remote telescope has the following relative position and velocity components in a rendezvous-type coordinate system whose origin is on the space station $x_0 = 0$, $y_0 = 13500/\pi$ m, $z_0 = 0$, $\dot{x}_0 = 10$ mps, $\dot{y}_0 = 0$, $\dot{z}_0 = 0$. How far, in meters, is the telescope from the space station 15 min later? What is the magnitude of the relative velocity, in mps, at this time?

7.3. Buck Rogers in a space bug and Dr. Huer in a space station are in the same circular orbit ($P = 2$ h) about the Earth but are 5486 m apart. Since Buck is ahead of the station and wants to rendezvous with his old friend in 30 min, he decides to

aim his rockets at the station and change his velocity by 3.05 mps (retrofire). After ingeniously determining Buck's trajectory, Dr. Huer exclaims, "Buck's goodfed again!" But has he? To find out, write down the x, y, $\dot{x}$, and $\dot{y}$ equations of Buck's motion relative to Dr. Huer. Determine the distance (in meters) between the two and their relative velocity (in meters/second) after 30 min. At this time, is Buck moving toward or away from the space station?

7.4. In order to avoid some orbiting debris, a geosynchronous satellite in a circular equatorial orbit applies an in-track ΔV. Some time later, the debris has passed, and the satellite has the following relatived position and velocity with respect to its nonmaneuvering location: $x = -120.6$ km, $y = 71.12$ km, $\dot{x} = -10.636$ m/s, and $\dot{y} = 20.221$ m/s. The satellite will now initiate a two-impulse maneuver to return to its original location in 2 h. What are the magnitudes and directions of ΔV_1 and ΔV_2?

7.5. "Phooey!" exclaims Buck Rogers as he throws his sandwich out the airlock of the satellite which he shares with Dr. Huer. "You've goofed again, Buck!" chides Doc. "You know that we're in a circular orbit about the moon. Since you threw that sandwitch radially away from the center of the moon, we'll have tuna salad all over our portholes after one revolution (3 h)." After a lightning calculation, Buck chortles, "You're wrong this time, Doc. You used the first-order rendezvous equations. I used the exact solution of Berreen and Crisp. Assuming a relative ΔV of 0.01 V_{circular}, the tuna salad will miss us by at least 2000 m." What will the distance between the sandwich and the satellite be after one revolution? Use the exact polar equations of Berreen and Crisp and the calculation algorithm presented in Sec. 7.4.

7.6. Use the equations for in-orbit repositioning to shift the longitude of a satellite in a geosynchronous circular equatorial orbit by +12 deg in 3 revolutions of the phasing orbit. Calculate the maneuver ΔV_1, to start the longitude drift. After 3 revolutions the application of $\Delta V_2 = -\Delta V_1$ will stop the drift. Compare the position of the satellite at this time with the position determined from the solution to the relative equations of motion, Eqs. (7.47), when ΔV_1 is applied.

7.7. Solve for relative range and magnitude of velocity between a free-flying space object (e.g. telescope) and a space station in a circular orbit. Use equations in Fig. 7.20 for any specified initial conditions of the space object.

Selected Solutions

7.1. Case 1: $\dfrac{X}{\Delta V} = -\dfrac{4}{\omega}\sin\omega t + 3t$

$$\frac{Y}{\Delta V} = \frac{2}{\omega}(1 - \cos\omega t)$$

After 90 min, $\frac{X}{\Delta V} = 16{,}200$ s

Case 2: $\dfrac{X}{\Delta V} = \dfrac{2}{\omega}(1 - \cos \omega t)$

$\dfrac{Y}{\Delta V} = \dfrac{1}{\omega} \sin \omega t$

After 90 min, 0

7.2. Relative distance = 7746.86 m
Relative velocity = 6.61 m/s

7.3. Relative distance = 10,600 m
Relative velocity = 11.0 m/s
Moving away

7.4. $\Delta V_1 = 39.30$ m/s
$\Delta V_2 = 19.08$ m/s

7.5. Orbit radius = 2.437×10^6 m
Distance = 2297 m (Buck is right)

7.6. $\Delta V_1 = 11.26$ m/s
Position difference = 3390 m or 0.0046 deg

7.7. A solution to this problem is provided in the software that accompanies this book.

8
Introduction to Orbit Perturbations

This chapter provides an introduction to, and an overview of, the orbit perturbations, the perturbing sources, and the physical phenomena associated with orbital motion. Simplified examples and key equations should give readers an easy-to-access reference for understanding and solving typical perturbation problems.

Chapter 9 provides an in-depth description of the mathematical foundations and derivations of the perturbing functions and equations. Chapter 10 gives more detailed descriptions and discussions on various perturbing sources and their special effects on orbital motion. The chapter also includes actual examples of mission analysis and design utilizing the unique orbit-perturbation properties.

8.1 A General Overview of Orbit Perturbations

What are orbit perturbations? By definition, those small deviations from the two-body orbit motion are called orbit perturbations. The two-body orbit motion can be expressed by the conic solutions (ellipse, hyperbola, and parabola) in closed form, which has been explained in detail in Chapters 3 and 4. The equations of two-body motion and its solutions were derived through Newton's law of gravitation and Kepler's laws of orbit motion under the assumption of point mass or mass with spherically symmetrical distribution.

The equations of motion of a two-body problem can be given in the relative form as

$$\frac{\mathrm{d}^2\boldsymbol{r}}{\mathrm{d}t^2} = -\frac{\mu}{r^3}\boldsymbol{r} \tag{8.1}$$

where $\boldsymbol{r}$ is the position vector of the satellite measured from the center of the primary body, and μ is the gravitational constant of $k^2(m_1 + m_2)$.

Because of the presence of various perturbing forces, Eq. (8.1) can be used only as an approximation of the actual motion. The accuracy of the approximation decreases as the time of propagation increases. Those perturbing forces include Earth gravity harmonics (deviations from a perfect sphere), the lunisolar gravitational attractions, atmospheric drag, solar radiation pressure, and Earth tides. For natural satellites with sizable mass such as the inner satellites of Jupiter and Saturn, the mutual gravitational attraction among the satellites is another source of perturbation. The general form of equations of motion, including perturbations, can be expressed as follows:

$$\frac{\mathrm{d}^2\boldsymbol{r}}{\mathrm{d}t^2} = -\frac{\mu}{r^3}\boldsymbol{r} + \boldsymbol{a}_p \tag{8.2}$$

where $\boldsymbol{a}_p$ is the sum of all the perturbing accelerations. In the solar system, the magnitude of all the perturbing accelerations is at least one order of magnitude less than the two-body acceleration. That is why the term "perturbations" is used.

With the perturbing accelerations included, the solutions of Eq. (2) can no longer be expressed by the closed-form conic equations. As an astronomer put it, "In problems of celestial mechanics, the simple solutions are not good and the good solutions are not simple." It will be shown in the next chapter the reasons why the good solutions, including the perturbations, are not so simple.

8.2 Earth Gravity Harmonics

Earth gravity harmonics are derived from the gravity potential through the potential theory, which will be explained in detail in the next chapter. Those harmonics are the terms of a mathematical expansion through which the deviations from a sphere can be represented. The commonly encountered gravity harmonics are J_2 and J_{22}, which are the largest terms of the zonal and tesseral harmonics, respectively. The coefficient of the second harmonic J_2 is related to Earth equatorial oblateness through Earth rotation, and the estimated difference between the polar radius and equatorial radius is about 22 km. J_2 is responsible for the secular rates of the right ascension of ascending node, the argument of perigee, and a small correction to the mean motion of the orbit. These rates can be computed by the following equations[1]:

$$\dot{\Omega} = -\frac{3}{2}\frac{J_2 R^2}{p^2}\bar{n}\cos i \tag{8.3}$$

$$\dot{\omega} = \frac{3}{2}\frac{J_2 R^2}{p^2}\bar{n}\left(2 - \frac{5}{2}\sin^2 i\right) \tag{8.4}$$

$$\bar{n} = \sqrt{\frac{\mu}{a_0^3}}\left[1 + \frac{3}{2}\frac{J_2 R^2}{p^2}\left(1 - \frac{3}{2}\sin^2 i\right)(1 - e^2)^{\frac{1}{2}}\right] \tag{8.5}$$

where

$\dot{\Omega}$ = rate of ascending node
$\dot{\omega}$ = argument of perigee rate
$\bar{n}$ = orbit mean motion with J_2 correction
$J_2 = 0.00108263$
R = Earth equatorial radius
i = orbit inclination
μ = gravitational constant
a_0 = semimajor axis at epoch
e = eccentricity
$p = a_0(1 - e^2)$

The second tesseral harmonic J_{22} is related to the ellipticity of the Earth equatorial plane and is responsible for the long-term (860-day) resonance effects on geosynchronous orbits. The long-term longitude oscillation of the communication satellites, which must be controlled by periodic stationkeeping maneuvers, is caused by the J_{22} effects. The magnitude and frequency of stationkeeping maneuvers depend on satellite longitude and control tolerance. Figure 8.1, from

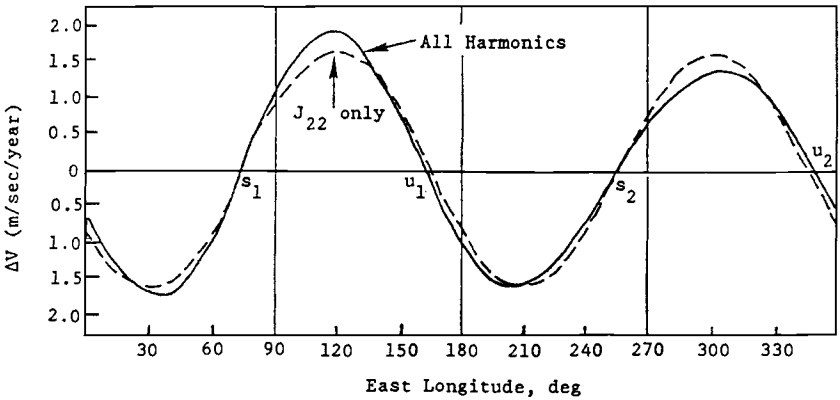


Fig. 8.1 Annual ΔV expenditure for triaxiality correction.

Ref. 2, shows the annual ΔV expenditure for longitude stationkeeping of geosynchronous satellites. The maximum values occur 45 deg from the four equilibrium points (s_1, s_2, u_1, u_2); s_1 and s_2 are stable points, and u_1 and u_2 are unstable points. The typical longitude control tolerance is ± 0.1 deg, and the frequency of maneuver is about once every 14 days, depending on the longitude.

8.3 Lunisolar Gravitational Attractions

To understand the long-term behavior of a satellite orbit under the influence of the sun, imagine both the satellite and the sun smeared out into elliptical rings coinciding with their respective orbits (Fig. 8.2). The mutual gravitational

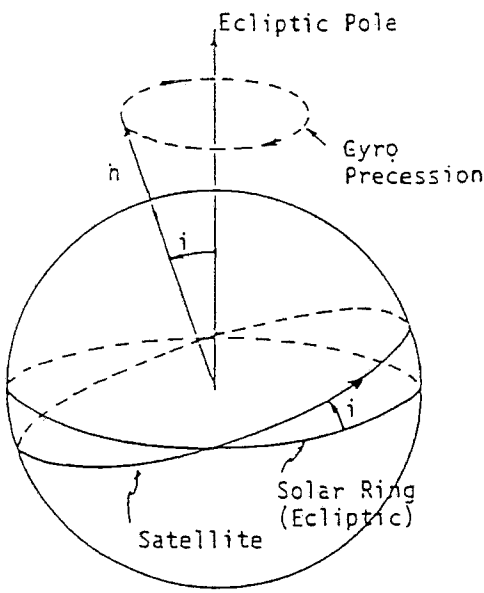


Fig. 8.2 Gyro precession of a satellite orbit.

attractions of the rings will create a torque about the line of nodes tending to turn the satellite ring into the ecliptic. The gyroscopic effect of the torque on the spinning satellite ring will induce a gyro precession of the orbit about the pole of the ecliptic, specifically a regression of the nodes along the ecliptic. Similarly, the moon will cause a regression of the orbit about an axis normal to the moon's orbit plane, which has a 5-deg inclination with respect to the ecliptic plane with a node rate of one rotation in 18.6 yr. The equations of nodal regression and rate of argument of perigee are (from Ref. 3)

$$\dot{\Omega} = -\frac{3}{8}\frac{n_3^2}{n}\frac{\left(1 + (3/2)e^2\right)}{\sqrt{1-e^2}}\cos i\left(3\cos^2 i_3 - 1\right) \tag{8.6}$$

$$\dot{\omega} = \frac{3}{4}\frac{n_3^2}{n}\frac{\left(1 - (3/2)\sin^2 i_3\right)}{\sqrt{1-e^2}}\left(2 - \frac{5}{2}\sin^2 i + \frac{e^2}{2}\right) \tag{8.7}$$

where n_3 and i_3 are the mean motion and inclination with respect to Earth equatorial plane, respectively. For low-altitude orbits, the coefficient (n_3^2/n) is very small compared to J_2 effects, and sun-moon effects can be neglected. For orbits with periods equal to 12 h or longer, the lunisolar effects are significant and should be included.

8.4 Radiation Pressure Effects

The effect of solar radiation on particles moving through interplanetary space has been investigated for many years. The first radiation pressure effect studied was related to very small meteorites or dust particles. It was first developed by Poynting in 1920 and was refined in keeping with the principles of relativity by Robertson in 1937. The net effect of this so-called Poynting–Robertson force is the influence of the impinging photon momentum or radiation pressure on a space vehicle. At one A.U., the solar radiation pressure constant P_0 is 4.7×10^{-5} dyne/cm^2. This value may fluctuate slightly ($<1\%$) as a result of variations in the solar activity index.

The typical radiation pressure effect on satellite orbits is the long-term sinusoidal (yearly for geosynchronous orbits) variations in eccentricity. The magnitude of the variation is proportional to the effective area, surface reflectivity, and inverse of the satellite mass. For a typical communication satellite at geosynchronous altitude, the eccentricity may vary from 0.001 to 0.004 in six months as a result of solar radiation pressure effects. For low-altitude orbits, the period of the long-term variation in eccentricity is governed by the combined rates of the longitude of the mean sun, nodal regress, and argument of perigee. Resonance conditions may occur when one of the combined rates is vanishingly small.

In summary, radiation pressure induces periodic variations in all orbital elements, even exceeding the effects of atmospheric drag at altitudes above 900 km. The induced changes in perigee height can have significant effects on the satellite's lifetime. For example, the 30-m ECHO balloon satellite in its 1852-km-altitude circular orbit displayed the greatest perturbation due to radiation pressure, approximately 3.5-km/day initial decrease in perigee height.

8.5 Atmospheric Drag

When the orbit perigee height is below 1000 km, the atmospheric drag effect becomes increasingly important. Drag, unlike other perturbation forces, is a non-conservative force and will continuously take energy away from the orbit. Thus, the orbit semimajor axis and the period are gradually decreasing because of the effect of drag. The orbit velocity is increasing, however, because Kepler's law ($\mu = n^2a^3 = \text{const}$) must be satisfied. This relationship leads to the "drag paradox": the effect of atmospheric friction is to speed up the motion of the satellite as it spirals inward.

Since the drag is greatest at perigee, where the velocity and atmospheric density are greatest, the energy drain is also greatest at this point. Under this dominant negative impulse at perigee, the orbit will become more circular in each revolution. Figure 8.3 shows the time history of the perigee and apogee altitude of a decaying satellite. The elliptic orbit first becomes circular as the apogee altitude decreases to the same value of perigee and then rapidly spirals into the dense atmosphere.

For near-circular orbits, the orbit semimajor axis decay rate can be computed by the following simple equation:

$$\frac{da}{dt} = -na^2\rho B \tag{8.8}$$

where

n = mean motion
ρ = atmosphere density at that altitude
B = ballistic coefficient = $C_d A/m$

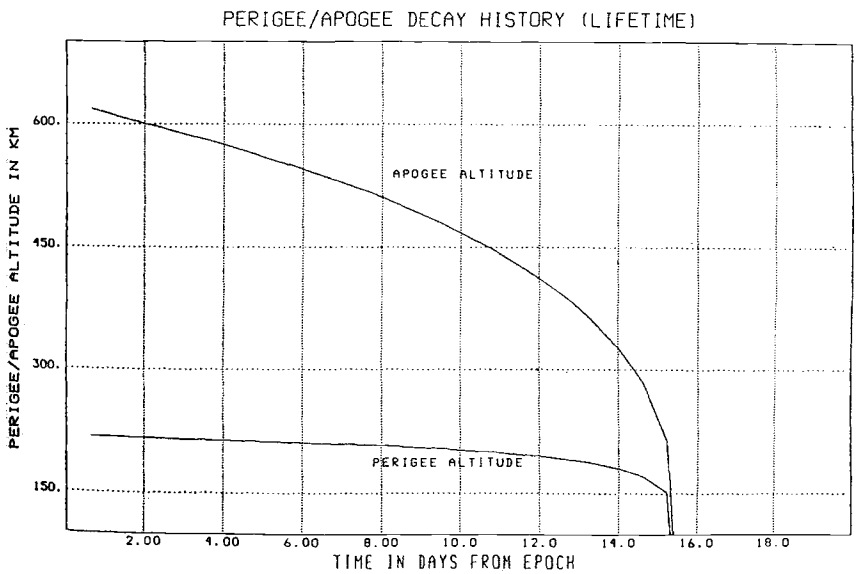


Fig. 8.3 Perigee and apogee decay due to atmospheric drag effects.

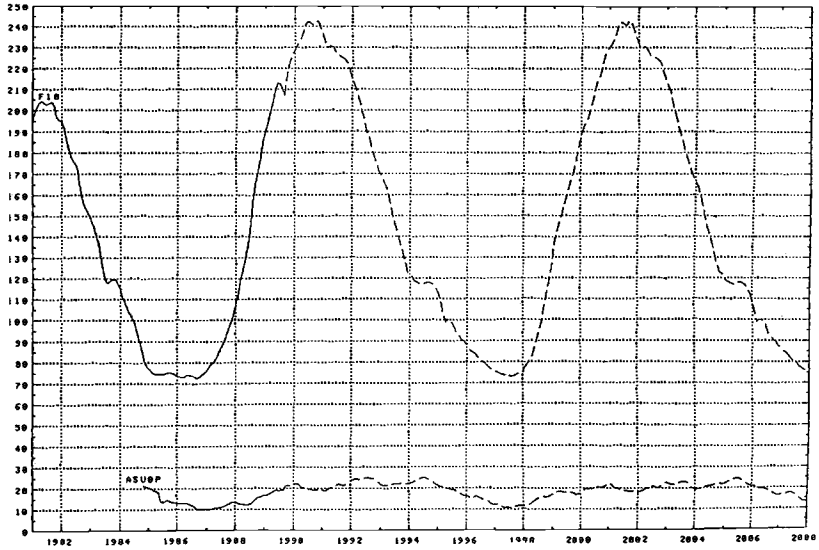


Fig. 8.4 Observed and predicted solar flux index ($F_{10.7}$) and index of magnetic activity (A_p).

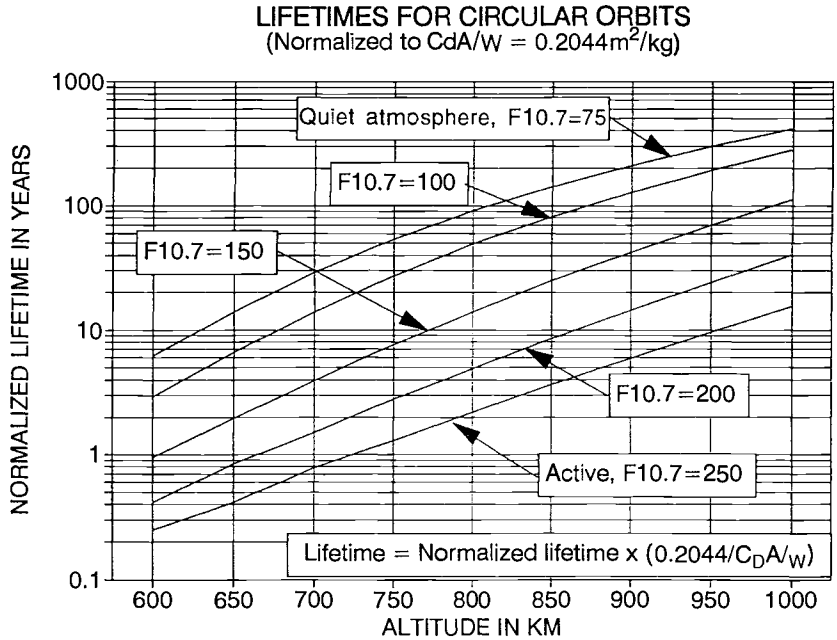
and

m = total mass of spacecraft, kg
A = effective cross-section area, m^2
C_d = drag coefficient = 2.0

An accurate prediction of the satellite motion under the influence of drag requires a good density model of the upper atmosphere. Over the past three decades, various density models have been developed, with varying degrees of complexity and fidelity. The commonly used models are the Jacchia 64 and 71 models. These two models take into account the diurnal variation, the 27-day fluctuation (sun's rotation period), the annual variation, and the 11-year solar cycle. Figure 8.4 shows the predicted 11-year solar cycle variations in $F_{10.7}$ (solar flux index) and A_p (index of magnetic activity) for the period between 1990 and 2008. Based on the Jacchia 1971 model, the estimated orbit lifetimes at various initial orbit altitudes and $F_{10.7}$ values are plotted in Fig. 8.5. Figure 8.6 shows the estimated stationkeeping ΔV per year for maintaining the orbit altitude in an active atmosphere. Both Figs. 8.5 and 8.6 are for orbits with small eccentricity.

8.6 Tidal Friction Effects and Mutual Gravitational Attraction

The magnitude of tidal friction effects on the artificial satellites is very small, and it usually is not included in the perturbation equations. However, tidal friction plays an important role in the evolution of the massive satellites of Jupiter, Saturn, and other outer planets. The coupling effects between tidal friction and mutual gravitational pulls among the massive satellites of the outer planet are responsible for the existence of some of the resonance phenomena discovered



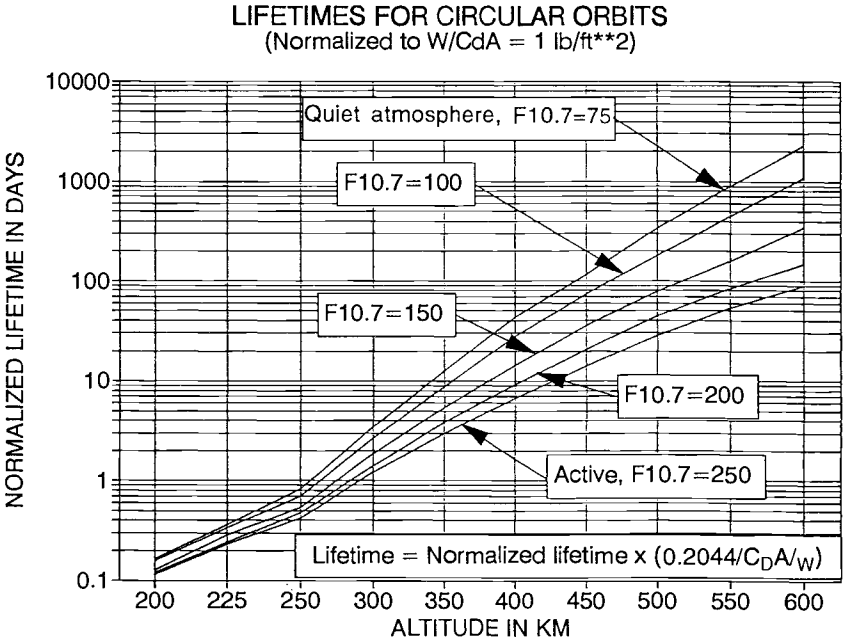


Fig. 8.5 Estimated orbit lifetime for average and active atmosphere.

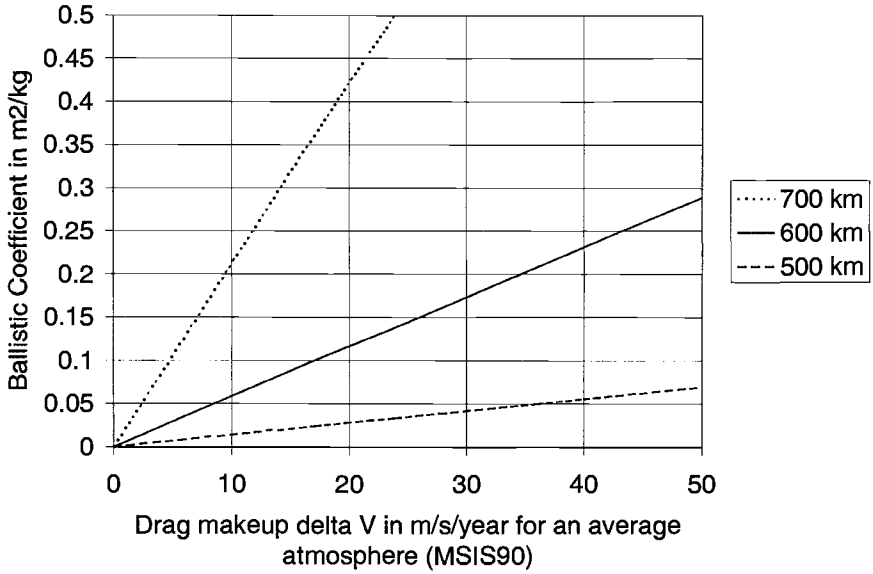


Fig. 8.6 Drag makeup ΔV for an average MSIS90 atmosphere ($F_{10.7}$ = 150, a_p = 15) at various orbit altitudes and C_dA/m.

through the first telescope built by Galileo. These effects will be discussed in detail in Chapter 10.

References

[1]Roy, A. E., *Orbital Motion*, 3rd ed., Adam Hilger, Bristol, UK, 1988.

[2]Michielsen, H. J., and Webb, E. D., "Stationkeeping of Stationary Satellites Made Simple," *Proceedings of the First Western Space Conference*, 1970.

[3]Chao, C. C., "An Analytical Integration of the Averaged Equations of Variation Due to Sun-Moon Perturbations and Its Application," The Aerospace Corp., Tech. Rept. SD-TR-80-12, Oct. 1979.

Problem

8.1. A low altitude Earth satellite moves in near circular orbit with the following elements at time t_0: a = 6800 km, e = 0.002, i = 50 deg., ω = 95 deg., Ω = 120 deg., M = 20 deg.

Determine the secular rates of the last three angular elements (eg., $\dot{\omega}$, $\dot{\Omega}$, $\dot{M}$) of the above set due to J_2 effects.

Selected Solution

8.1. $\dot{\omega} = 4.2446$ deg./day

$\dot{\Omega} = -5.1195$ deg./day

$\dot{M} = 5573.6783 + 0.9537 = 5574.6320$ deg./day

$= 15.4851$ Rev/day

9
Orbit Perturbations: Mathematical Foundations

In Chapter 8, the physical phenomena of orbit perturbations due to various sources have been discussed. This chapter provides an introduction to the mathematical foundations of those perturbations and the various methods of solution.

9.1 Equations of Motion

Two-Body Point Mass

Before going into equations of motion for orbit perturbations, it is important to review the two-body equations of motion in relative form. The equations of motion for a satellite moving under the attraction of a point mass planet without any other perturbations can be given in the planet-centered coordinates as

$$\frac{d^2\boldsymbol{r}}{dt^2} = -\mu\frac{\boldsymbol{r}}{r^3} \tag{9.1}$$

where

$\boldsymbol{r}$ = position vector of the satellite
μ = gravitational constant
t = *time*

Equation (9.1) is a set of three simultaneous second-order nonlinear differential equations. There are six constants of integration. The solution of Eq. (9.1) can be either in terms of initial position and velocity: x_0, y_0, z_0, $\dot{x}_0$, $\dot{y}_0$, $\dot{z}_0$; or in terms of the six orbit elements: $a, e, i, \Omega, \omega, M$.

The closed-form conic solutions of the two-body equations of motion have been given in the earlier chapters, and they may be expressed in a general functional form as

$$\boldsymbol{r}(t) = \boldsymbol{r}(x_0, y_0, z_0, \dot{x}_0, \dot{y}_0, \dot{z}_0, t) \tag{9.2a}$$

or

$$\boldsymbol{r}(t) = \boldsymbol{r}(a, e, i, \Omega, \omega, M) \tag{9.2b}$$

Five of the six orbit elements $(a - \omega)$ in the preceding expression are constants, and M is the mean anomaly defined by

$$M = M_0 + n(t - t_0) \tag{9.3}$$

where

M_0 = mean anomaly at epoch, t_0

n = mean motion = $\sqrt{\dfrac{\mu}{a^3}}$

Figure 9.1 shows the orbit geometry of an orbiting satellite in the inertial Earth-centered equatorial coordinate system (ECI). It is important to know that, without perturbations, the orbit plane and perigee orientation stay fixed in the inertial space.

The orbit elements described earlier are called the classical orbit elements, and they are widely used in celestial mechanics. However, this set of elements becomes poorly defined and ill behaved when the eccentricity and/or the inclination become vanishingly small. To remedy this problem, a particular set of orbit elements was developed,[1] and they are defined as

$$\begin{aligned} a &= a \\ h &= e\sin(\omega + \Omega) \\ k &= e\cos(\omega + \Omega) \\ \lambda &= M + \omega + \Omega \\ p &= \tan(i/2)\sin\Omega \\ q &= \tan(i/2)\cos\Omega \end{aligned} \tag{9.4}$$

Then, the solution may be expressed as

$$\boldsymbol{r} = \boldsymbol{r}(a, h, k, \lambda, p, q) \tag{9.5}$$

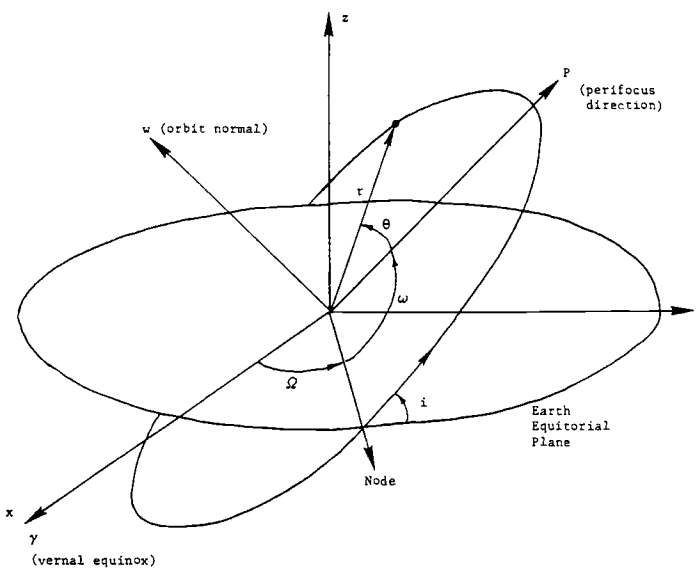


Fig. 9.1 Geometry of a satellite orbit.

where λ, the so-called mean longitude, is the only time-varying parameter, and

$$\lambda = \lambda_0 + n(t - t_0) \tag{9.6}$$

where

$$\lambda_0 = M_0 + \omega + \Omega = \text{mean longitude at epoch } (t_0)$$

For more discussions on equinoctial elements, see Refs. 1, 9, and 10 or Chapter 14. Whether the two-body solutions are given in terms of initial position and velocity or orbit elements, one can always obtain the solutions in closed form. The new position and velocity can be computed at any given time.

In the real world, life is not that simple. Two-body solutions can give approximations for the orbit ephemeris for only a short time before the effect of perturbing accelerations becomes significant.

It is the purpose of this chapter to provide an overview of what the perturbations are and how these effects are computed by various methods.

Equations of Motion with Perturbations

To include the effects of the perturbations, the equations of motion can be written in a general form as

$$\frac{d^2\boldsymbol{r}}{dt^2} = -\mu\frac{\boldsymbol{r}}{r^3} + \boldsymbol{a}_p \tag{9.7}$$

where $\boldsymbol{a}_p$ is the resultant vector of all the perturbing accelerations. In the solar system, the magnitude of the $\boldsymbol{a}_p$ for all the satellite orbits is at least 10 times smaller than the central force or two-body accelerations, or $|\boldsymbol{a}_p \ll |\mu\boldsymbol{r}/r^3|$. $\boldsymbol{a}_p$may consist of the following types of perturbing accelerations.

Gravitational:	Third-body (sun/moon) attractions The nonspherical Earth
Nongravitational:	Atmospheric drag Solar-radiation pressure Outgassing (fuel tank leaks on the spacecraft) Tidal friction effect

These perturbations can also be grouped as conservative and nonconservative. For conservative accelerations, $\boldsymbol{a}_p$ is an explicit function of position only, and there is no net energy transfer taking place. Therefore, the mean semimajor axis of the orbit is constant. For nonconservative accelerations, where $\boldsymbol{a}_p$ is an explicit function of both position and velocity, such as atmospheric drag, outgassing, and tidal friction effect, energy transfer occurs. Consequently, the mean semimajor axis of the orbit changes.

9.2 Methods of Solution

There are two general approaches to solve the equations of motion with perturbations. One approach is through a step-by-step numerical integration, which

is often called "special perturbation." The other approach is through analytical expansion and integration of the equations of variations of orbit parameters. The latter approach is referred as "general perturbation."

Special Perturbation

In special perturbation, there are two basic methods: Cowell's method and Enckel's method. These two methods are explained in several books such as Refs. 2–4, and they are briefly introduced here.

Cowell's method. This method is a straightforward step-by-step integration of the two-body equations of motion with perturbations. The equations of motion may be given,

$$\ddot{\boldsymbol{r}} + \frac{\mu}{r^3}\boldsymbol{r} = \boldsymbol{a}_p \tag{9.8}$$

which, for numerical integration, would be reduced to first-order differential equations

$$\begin{cases} \dot{\boldsymbol{r}} = \boldsymbol{v} \\ \dot{\boldsymbol{v}} = -\dfrac{\mu}{r^3}\boldsymbol{r} + \boldsymbol{a}_p \end{cases} \tag{9.9}$$

where $\boldsymbol{a}_p$ is the vector sum of all the perturbing accelerations to be included in the integration. Cowell's method does not require that the magnitude of $\boldsymbol{a}_p$ should be small. Cowell's method can be further illustrated by the following example of integrating a second-order nonlinear differential equation of the form

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + 3x^2\frac{\mathrm{d}x}{\mathrm{d}t} - 5x = 12 \tag{9.10}$$

Equation (9.10) is first reduced to two first-order differential equations as

$$\begin{cases} \dfrac{\mathrm{d}x}{\mathrm{d}t} = y \\ \dfrac{\mathrm{d}y}{\mathrm{d}t} = -3x^2y + 5x + 12 \end{cases} \tag{9.11}$$

The step-by-step integration may take a simple recursive relation as

$$\begin{aligned} y(t_0 + \Delta t) &= y(t_0) + \left.\frac{\mathrm{d}y}{\mathrm{d}t}\right|_{t_0} \Delta t \\ x(t_0 + \Delta t) &= x(t_0) + \left.\frac{\mathrm{d}x}{\mathrm{d}t}\right|_{t_0} \Delta t \end{aligned} \tag{9.12}$$

where $\mathrm{d}x/\mathrm{d}t|_{t_0}$ and $x(t_0)$ are given initial conditions and $y(t_0)$ and $\mathrm{d}y/\mathrm{d}t|_{t_0}$ can be computed from Eq. (9.11). The step size Δt should be so chosen that the round-off error and the truncation error are smaller than the error tolerance. Three types of

integrators have been used in orbit computations: the Runge–Kutta method (fourth order and higher), the Adams–Moulton multistep predictor/corrector method, and the Gauss–Jackson (second sum) method. The Gauss-Jackson method integrates the second-order equations of motion directly without having to reduce them to first-order equations before integration. It was proved recently by Fox[5] that, for near-circular orbits, the Gauss–Jackson method with Herrick's[2] starting algorithm is the most efficient integrator for Cowell's method. These three methods have been programmed by the author in FORTRAN-77 to integrate satellite orbits with J_2 perturbations, and these programs are available in IBM/PC diskettes.

Encke's method. One disadvantage of Cowell's method is slow computation. Even with the efficient Gauss–Jackson integrator, the computer time required to integrate an orbit to several hundred revolutions is still quite large. Encke's method cuts down the computer time considerably by integrating only the difference from a reference orbit whose ephemeris is known.

Encke's method uses a reference two-body orbit whose initial position and velocity equal that of the orbit with perturbations. Let the equation of motion of the reference orbit be

$$\ddot{\boldsymbol{\rho}} = -\frac{\mu}{\rho^3}\boldsymbol{\rho} \tag{9.13}$$

and define the departure from this reference orbit as

$$\delta\boldsymbol{r} = \boldsymbol{r} - \boldsymbol{\rho} \tag{9.14}$$

where $\boldsymbol{r}$ must satisfy the equation of motion of the true orbit (Eq. 9.8). Then, by differencing Eq. (9.14) twice and with Eq. (9.13), one gets

$$\delta\ddot{\boldsymbol{r}} = \mu\left(\frac{\boldsymbol{\rho}}{\rho^3} - \frac{\boldsymbol{r}}{r^3}\right) + \boldsymbol{a}_p \tag{9.15}$$

The difference $\delta\boldsymbol{r}$ is expected to be small and slowly varying as a result of the small perturbing accelerations $\boldsymbol{a}_p$. Through some binomial expansions and approximations, Eq. (9.15) can be given as[4]

$$\delta\ddot{\boldsymbol{r}} = \frac{\mu}{\rho^3}(fq\boldsymbol{r} - \delta\boldsymbol{r}) + \boldsymbol{a}_p \tag{9.16}$$

The binomial series fq is derived in Ref. 4. To integrate Eq. (9.16), the step size may be much larger than that of Cowell's method because of the slow-varying nature of $\delta\boldsymbol{r}$. However, as the size of $\delta\boldsymbol{r}$ increases to a certain magnitude, the error in the approximation becomes too large, and a new reference orbit should be initialized. This procedure is called rectification, as shown in Fig. 9.2.

The advantages and disadvantages of these two methods in special perturbation may be summarized in Table 9.1.

General Perturbation

The integration of series, analytically, term by term, is the core of general perturbation theory. The integrand series are expansions based on the perturbing ac-

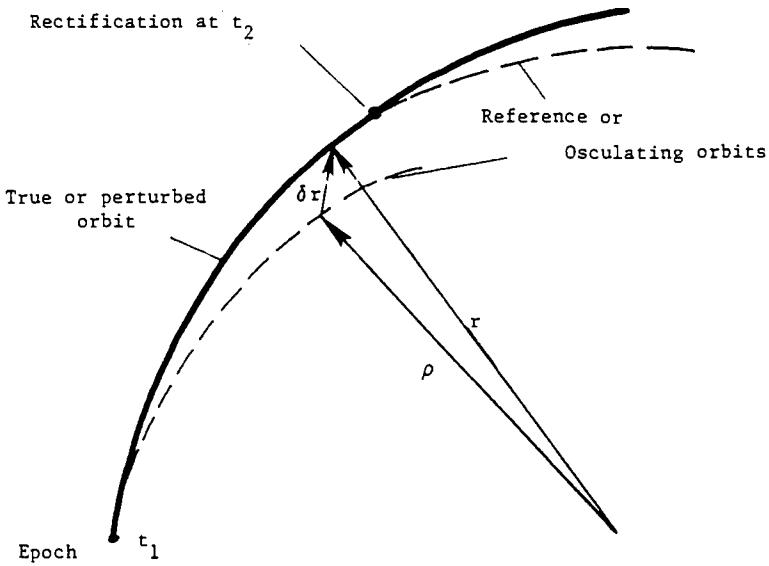


Fig. 9.2 Orbit geometry of Encke's method

celerations or the perturbing functions. These expansions are obtained through the methods of variation of parameters and variation of coordinates (Encke's method). Encke's method has been included in the special perturbation section; therefore, in this section, only the method of variation of parameters will be discussed.

Variation of parameters. From the study of two-body motion, it is known that the coordinates and velocity components at any instant permit the determination of a unique set of six orbit elements. In the problem of two bodies, the elements do not change with time in the inertial reference coordinates. Hence, once the elements are determined at some epoch, the position and velocity of the satellite at any other time can be computed with the same number of significant figures as the basic data at epoch.

In the presence of perturbations, such as drag, third-body attraction, and Earth gravity harmonics, the Keplerian orbit elements are no longer constant. The concept of variation of parameters allows the orbit elements to vary in such a way that, at any instant, the coordinates and velocity components can be computed from a unique set of two-body elements as if there were no perturbations. The equations of the variations can be derived from the concept of perturbed variations. There are two basic approaches to obtain the variational equations in celestial mechanics. They are the force components approach and the perturbing function approach. The former is sometimes called the Gaussian method, and the latter is called the Lagrangian method.

The Force Components Approach

The force components approach directly relates the perturbing force components to the rate of the orbit elements. The general form of equations of variation can be

Table 9.1 Cowell's vs Encke's method

Method	Advantages	Disadvantages
Cowell's	Simplicity in formulation and implementation	Slow in computation
	General purpose and no approximations	Round-off error may affect accuracy where step size is too small
Encke's	Can take much larger step size and thus save computer time	Complex in formulation
	Equations of motion may be linearized for analytical integration	Inadequate number of rectification may cause truncation error

derived through the concept of perturbed variations.[2] Each of the six orbit elements can be expressed in terms of the six states ($\boldsymbol{r}$ and $\dot{\boldsymbol{r}}$) at a given time via two-body relations as

$$\beta = \beta(\boldsymbol{r}, \dot{\boldsymbol{r}}), \qquad \beta \Rightarrow a, e, i, \Omega, \omega, M \tag{9.17}$$

The time derivative of the above-mentioned function leads to the following relation:

$$\frac{\mathrm{d}\beta}{\mathrm{d}t} = \sum_{i=1}^{3} \left(\frac{\partial \beta}{\partial r_i} \frac{\mathrm{d}r_i}{\mathrm{d}t}\bigg|_p + \frac{\partial \beta}{\partial \dot{r}_i} \frac{\mathrm{d}\dot{r}_i}{\mathrm{d}t}\bigg|_p \right) \tag{9.18}$$

where $\mathrm{d}r/\mathrm{d}t|_p$ and $\mathrm{d}\dot{r}/\mathrm{d}t|_p$ are the variations in position and velocity due to the presence of perturbing forces. At any given instant, the position and velocity of the satellite can be represented in terms of a unique set of six orbit elements at that time, as illustrated by the broken osculating orbits in Fig. 9.2. The concept of the osculating orbits requires that the position of the true orbit and that of the two-body osculating orbit coincide at that instant; thus, we require

$$\frac{\mathrm{d}r_i}{\mathrm{d}t}\bigg|_p = 0 \tag{9.19}$$

and the derivative $\mathrm{d}\boldsymbol{r}/\mathrm{d}t|_p$ is nothing but the perturbing accelerations $\boldsymbol{a}_p$. Therefore, Eq. (9.18) reduces to

$$\frac{\mathrm{d}\beta}{\mathrm{d}t} = \sum_{i=1}^{3} \frac{\partial \beta}{\partial \dot{r}_i} a_{pi}(\boldsymbol{r}, \dot{\boldsymbol{r}}) \tag{9.20}$$

The equations of variation of elements represented by relation (9.20) can be developed with accelerations either in Cartesian coordinates or in spherical coordinates. Usually, a simpler set of equations will result if the perturbing accelerations are in spherical coordinates. The rates of the six classical elements are given in Eq. (9.21) in the spherical coordinate system.

$$
\begin{aligned}
\frac{\mathrm{d}a}{\mathrm{d}t} &= \frac{2e\sin\theta}{nx}F_r + \frac{2ax}{nr}F_s \\
\frac{\mathrm{d}e}{\mathrm{d}t} &= \frac{x\sin\theta}{na}F_r + \frac{x}{na^2e}\left(\frac{a^2x^2}{r} - r\right)F_s \\
\frac{\mathrm{d}i}{\mathrm{d}t} &= \frac{r\cos u}{na^2x}F_w \\
\frac{\mathrm{d}\Omega}{\mathrm{d}t} &= \frac{r\sin u}{na^2x\sin i}F_w \\
\frac{\mathrm{d}\omega}{\mathrm{d}t} &= -\frac{x\cos\theta}{nae}F_r + \frac{p}{eh}\left[\sin\theta\left(1 + \frac{1}{1+e\cos\theta}\right)\right]F_s - \frac{r\cot i\sin u}{na^2x}F_w \\
\frac{\mathrm{d}M}{\mathrm{d}t} &= n - \frac{1}{na}\left(\frac{2r}{a} - \frac{x^2}{e}\cos\theta\right)F_r - \frac{x^2}{nae}\left(1 + \frac{r}{ax^2}\right)\sin\theta F_s
\end{aligned}
\tag{9.21}
$$

where

F_r, F_s, F_w = perturbing accelerations along the position vector $\boldsymbol{r}$, 90 deg from $\boldsymbol{r}$ in the velocity-increasing direction and normal to the orbit plane, respectively

θ = true anomaly

n = mean motion

$x = \sqrt{1-e^2}$

$u = \theta + w$, argument of latitude

$p = a(1-e^2)$

$h = \sqrt{\mu p}$

In the force components approach, the accelerations can be either conservative or nonconservative. For example, perturbations due to oblateness (J_2) and third-body (sun/moon) attractions are conservative. Perturbations due to atmosphere drag are nonconservative. By definition, conservative accelerations do not cause secular change in orbit semimajor axis or orbit energy.

The Perturbing Function Approach

All the conservative forces can be derived from potential functions. In this approach, the perturbing accelerations are assumed to be equal to the partial derivative of the perturbing function:

$$a_{pi} = \frac{\partial R}{\partial r_i} \tag{9.22}$$

where R is called the perturbing or disturbing function, and R is a function of satellite position only. Therefore, the equations of motion can be written as

$$\frac{\mathrm{d}^2 r_i}{\mathrm{d}t^2} = -\mu \frac{r_i}{r^3} + \frac{\partial R}{\partial r_i} \tag{9.23}$$

From Hamiltonian mechanics,

$$\begin{aligned} \left.\frac{\mathrm{d}r_i}{\mathrm{d}t}\right|_p &= -\frac{\partial R}{\partial \dot{r}_i} = 0 \\ a_{p_i} = \left.\frac{\mathrm{d}\dot{r}_i}{\mathrm{d}t}\right|_p &= \frac{\partial R}{\partial r_i} \end{aligned} \tag{9.24}$$

Then, after some mathematical derivations, one can obtain the equations of variations of orbit elements represented by the following relation in terms of the perturbing function R:

$$\frac{\mathrm{d}\beta}{\mathrm{d}t} = \sum_{\alpha} (\alpha, \beta) \frac{\partial R}{\partial \alpha}, \quad \alpha \neq \beta \text{ and } \left.\begin{matrix} \alpha \\ \beta \end{matrix}\right\} \Rightarrow a, e, i, \Omega, \omega, M \tag{9.25}$$

where (α, β) is called Poisson's bracket and is defined as

$$(\alpha, \beta) = \sum_{i=1}^{3} \left(\frac{\partial \alpha}{\partial r_i} \frac{\partial \beta}{\partial \dot{r}_i} - \frac{\partial \alpha}{\partial \dot{r}_i} \frac{\partial \beta}{\partial r_i} \right) \tag{9.26}$$

There are 36 brackets with the following properties:

$$\begin{aligned} &(\alpha, \beta) = o \quad \text{when } \alpha = \beta \\ &(\alpha, \beta) = -(\beta, \alpha) \\ &(\alpha, \beta) = \text{const (not a function of time)} \end{aligned} \tag{9.27}$$

These brackets are derived from the two-body relations. The following equations of variations of classical elements are obtained through the disturbing function approach:

$$\begin{aligned} \frac{\mathrm{d}a}{\mathrm{d}t} &= \frac{2}{na} \frac{\partial R}{\partial M} \\ \frac{\mathrm{d}e}{\mathrm{d}t} &= \frac{x^2}{na^2 e} \frac{\partial R}{\partial M} - \frac{x}{na^2 e} \frac{\partial R}{\partial \omega} \\ \frac{\mathrm{d}M}{\mathrm{d}t} &= n - \frac{x^2}{na^2 e} \frac{\partial R}{\partial e} - \frac{2}{na} \frac{\partial R}{\partial a} \\ \frac{\mathrm{d}\Omega}{\mathrm{d}t} &= \frac{1}{na^2 x \sin i} \frac{\partial R}{\partial i} \\ \frac{\mathrm{d}\omega}{\mathrm{d}t} &= \frac{x}{na^2 e} \frac{\partial R}{\partial e} - \frac{\cot i}{na^2 x} \frac{\partial R}{\partial i} \\ \frac{\mathrm{d}i}{\mathrm{d}t} &= \frac{1}{na^2 x} \left[\cot i \frac{\partial R}{\partial \omega} - \csc i \frac{\partial R}{\partial \Omega} \right] \end{aligned} \tag{9.28}$$

For the same perturbing forces, the two approaches, Eq. (9.21) and Eq. (9.28), should lead to the same equations of variations. In actual applications, the force components approach (Gaussian method) is usually preferred because it can handle both conservative and nonconservative forces and because no expansion and partial differentiation of the perturbing function are needed. Equations (9.28) can be solved numerically or analytically, and the methods will be discussed later.

9.3 Potential Theory

In order to have a clear picture of the gravity harmonics of the Earth and other planets, the concept of a gravity potential is introduced. Much as with the potential theory in fluid mechanics, the gravity field of a body with finite mass can be represented by a potential function. If the mass of a celestial body is assumed to be a point mass or uniformly distributed in a sphere, the potential function takes the following simple form:

$$\Phi = \frac{\mu}{r} \tag{9.29}$$

Function (9.29) indicates that the strength of the gravity potential at a point in space is directly proportional to the mass of the body ($\mu = k^2 m$) and inversely proportional to the distance to the center of the body. From potential theory, the gravitational force or acceleration along a given direction is equal to the partial derivative or gradient in that direction. Thus, the equations of motion of a two-body orbit may be given as

$$\begin{aligned} \ddot{x} &= \frac{\partial \Phi}{\partial x} & \ddot{x} &= -\frac{\mu x}{r^3} \\ \ddot{y} &= \frac{\partial \Phi}{\partial y} \Rightarrow & \ddot{y} &= -\frac{\mu y}{r^3} \\ \ddot{z} &= \frac{\partial \Phi}{\partial z} & \ddot{z} &= -\frac{\mu z}{r^3} \end{aligned} \tag{9.30}$$

or, in spherical coordinates,

$$\begin{aligned} \ddot{r} - r\dot{\theta}^2 &= \frac{\partial \Phi}{\partial r} & \ddot{r} - r\dot{\theta}^2 &= -\frac{\mu}{r^2} \\ & \Rightarrow & & \\ r\ddot{\theta} + 2\dot{r}\dot{\theta} &= \frac{1}{r}\frac{\partial \Phi}{\partial \theta} & r\ddot{\theta} + 2\dot{r}\dot{\theta} &= 0 \end{aligned} \tag{9.31}$$

In reality, the point-mass potential cannot accurately represent the gravity field of Earth and other planets in the solar system because of the nonspherical shape of these bodies. Instead, the potential function should be derived from a spheroid that can closely represent the shape and mass distribution of the Earth or other planets.

The basis of the development of the potential of a spheroid is in the integral

$$\Phi = k^2 \int \frac{\mathrm{d}m}{s} \tag{9.32}$$

where $\mathrm{d}m$ is an element of the mass of the attracting body and s is the distance from this point-mass element to the attracted particle, a satellite S whose mass is assumed to be negligible (Fig. 9.3). Figure 9.3 shows two coordinate systems: 1) inertial coordinates (x, y, z) for the satellite motion, and 2) rotating coordinates (ξ, η, ζ) fixed to the attracting body. Through some lengthy mathematical integrations over the entire body, Eq. (9.32) becomes

$$\Phi = \frac{\mu}{r} \sum_{n=0}^{\infty} \sum_{q=0}^{n} \left(\frac{a_e}{r}\right)^n P_n^q(w)(C_{n,q} \cos q\lambda + S_{n,q} \sin q\lambda) \tag{9.33}$$

This infinite series is the potential function of a spheroid with coefficients $C_{n,q}$ and $S_{n,q}$ to be determined from observations, where

a_e = equatorial radius of the body
$P_n^q(w)$ = Legendre polynomials
w = $\sin\delta$, δ = declination of satellite
λ = longitude of the satellite in the body-fixed coordinates

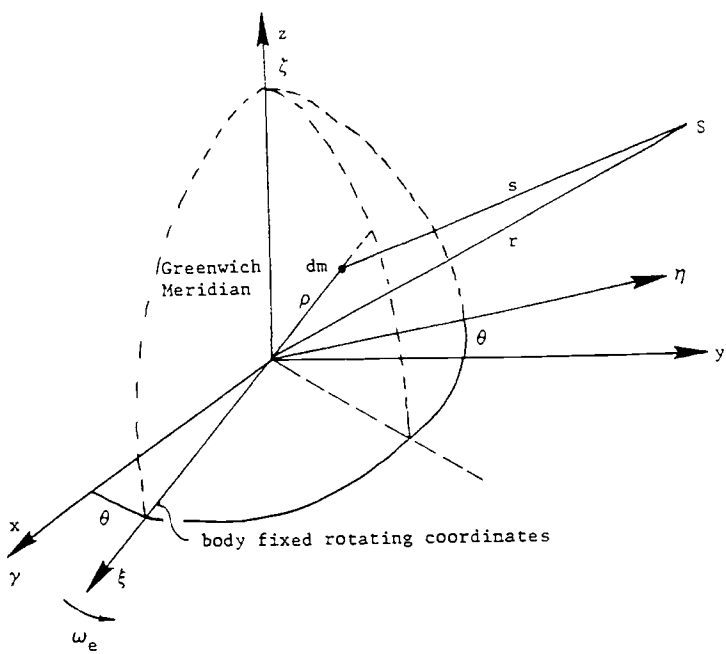


Fig. 9.3 Coordinates for gravity harmonics derivation.

The potential of a spheroid of revolution can be obtained from Eq. (9.33), with the center of mass at the center of the coordinate system as

$$\Phi = \frac{\mu}{r}\left[1 - \sum_{n=2}^{\infty}\left(\frac{a_e}{r}\right)^n J_n P_n(w)\right] \tag{9.34}$$

where

$$J_n = -C_{n,0} = -C_n \tag{9.35}$$

The mathematical definitions of $C_{n,q}$ and $S_{n,q}$ follow:
Zonal harmonics (spheroid of revolution about the spin axis):

$$C_{n,o} = C_n = -J_n = \frac{1}{ma_e^n}\int \rho^n P_n^o(W)\mathrm{d}m \tag{9.36}$$

The lowest-order zonal harmonics J_2 can be defined as

$$C_{2,0} = C_2 = -J_2 = \frac{1}{ma_e^2}\int \frac{\rho^2}{2}(3W^2 - 1)\mathrm{d}m \tag{9.37}$$

where

$\rho = (\xi^2 + \eta^2 + \zeta^2)$
m = mass of the spheroid
$W = \sin\phi'$
ϕ' = latitude of the mass element dm in the body-fixed coordinates

Tesseral and sectorial harmonics:

$$C_{n,q} = \frac{2}{ma_e^n}\frac{(n-q)!}{(n+q)!}\int \rho^n P_n^q(W)\cos q\lambda' \mathrm{d}m \tag{9.38}$$

where λ' is the body-fixed longitude of the mass element, and

$$C_{n,q} \rightarrow S_{n,q} \text{ for } \cos(\ ) \rightarrow \sin(\ ) \tag{9.39}$$

The preceding definitions indicate that the values of the harmonics depend on the shape and mass distribution of the central body. For Earth, various sets of gravity harmonics have been determined in the past two decades. The most widely used set, which is probably the most accurate one, is the WGS 84 model. The values of the low-order harmonics (4 × 4) of that model are listed in Table 9.2 for reference. It is important to know that the value of J_2 is about 400 times larger than the next-largest value J_3. That is why, for most satellite orbits, a reasonably good accuracy can be maintained by simply including the J_2 effect.

9.4 More Definitions of Gravity Harmonics

Sometimes, it is more convenient to represent those harmonic coefficients in terms of J_{nm} and λ_{nm} instead of $C_{n,q}$ and $S_{n,q}$. Let us now review the mathematical

Table 9.2 Value's of low-order (fourth) harmonics of Earth gravitation potential (WGS 84 model)

Zonal harmonics	
$J_2 = 1082.6300E\text{-}6$[a]	
$J_3 = -2.5321531E\text{-}6$	
$J_4 = -1.6109876E\text{-}6$	
Tesseral harmonics	
$C_{22} = 1.5747419E\text{-}6$	$S_{22} = -9.0237594E\text{-}7$
$C_{31} = 29.146736E\text{-}6$	$S_{31} = 2.7095717E\text{-}7$
$C_{32} = 3.0968373E\text{-}7$	$S_{32} = -2.1212017E\text{-}7$
$C_{33} = 1.0007897E\text{-}6$	$S_{33} = 1.9734562E\text{-}7$
$C_{42} = 7.7809618E\text{-}8$	$S_{42} = 1.4663946E\text{-}7$
$C_{44} = -3.9481643E\text{-}9$	$S_{44} = 6.540039E\text{-}9$

[a] $E\text{-}6 = \times 10^{-6}$

expressions of those definitions of the geopotential harmonics in a slightly different representation.

The external geopotential function at any point P specified by the spherical coordinates r, ϕ, λ defined in Fig. 9.4 can be expressed as

$$V = -\frac{\mu}{r}\left[1 - \sum_{n=2}^{\infty} J_n \left(\frac{R_e}{r}\right)^n P_n(\sin\phi) + \sum_{n=2}^{\infty}\sum_{m=1}^{n} J_{nm}\left(\frac{R_e}{r}\right)^n P_{nm}(\sin\phi)\cos m(\lambda - \lambda_{nm})\right] \tag{9.40}$$

where

r = geocentric distance
ϕ = geocentric latitude
λ = geographic longitude
μ = GM_e = Newtonian constant times mass of Earth
R_e = mean equatorial radius of the Earth
P_{nm} = associated Legendre polynomial of degree n and order m
J_{nm} = harmonic coefficients
λ_{nm} = equilibrium longitude for J_{nm}
P_n = Legendre polynomial of degree n and order of zero
J_n = J_{n0}

where the relationships between J_{nm}/λ_{nm} and $C_{n,q}/S_{n,q}$ are

$$J_{nm}^2 = C_{n,q}^2 + S_{n,q}^2$$

$$\lambda_{nm} = \tan^{-1}(S_{n,q}/C_{n,q})/m, \qquad q = m$$

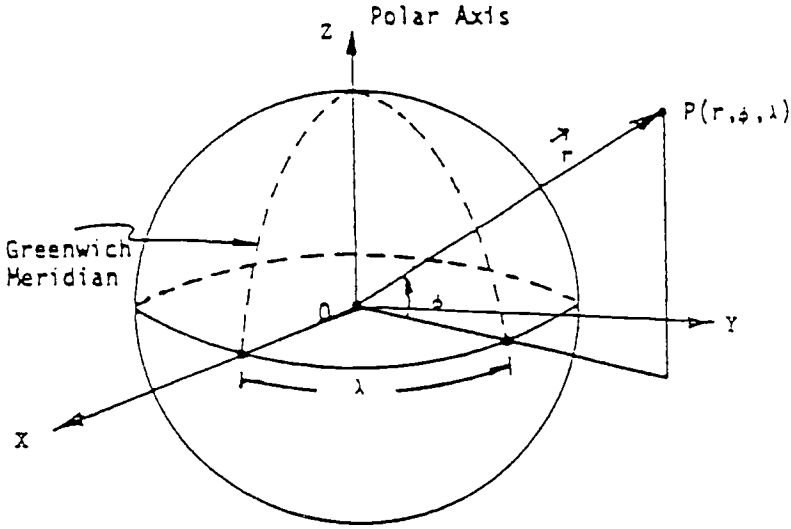


Fig. 9.4 Geopotential function at a given point P in spherical coordinates.

1) *Zonal Harmonics* ($m = 0$): For $m = 0$, the dependence of V on longitude vanishes, and the field is symmetric about the polar axis. For any $P_n(\sin\phi)$, there are n circles of latitude along which P_n is zero and, hence, $(n + 1)$ zones in which the function is alternately $(+)$ and $(-)$.

2) *Sectorial Harmonics* ($n = m$): The polynomial $P_{nm}(\sin\phi)$ are zero only at the poles ($\phi = \pm\pi/2$). On the other hand, the terms

$$\begin{pmatrix}\cos\\ \sin\end{pmatrix} n\lambda$$

are zero for $2n$ different values of λ; hence, the lines along which the function

$$\begin{pmatrix}\cos\\ \sin\end{pmatrix} n\lambda$$

$P_{nn}(\sin\phi)$ vanish are meridians of longitude, which divide the sphere into $2n$ "orange-slice" sectors alternately $(+)$ and $(-)$.

3) *Tesseral Harmonics* ($n \neq m$): For $n \neq m$, the functions

$$\begin{pmatrix}\cos\\ \sin\end{pmatrix} m\lambda$$

$P_{nm}(\sin\phi)$ are referred to as tesseral ("square") harmonics, for the sphere is divided up into a checkerboard array of domains alternately $(+)$ and $(-)$.

In general,

$$P_{nm}(x) = \frac{(1-x^2)^{\frac{m}{2}}}{2^n n!} \cdot \frac{d^{(n+m)}(x^2-1)^n}{dx^{(n+m)}} \tag{9.41}$$

An illustration of the different zonal, tesseral, and sectoral harmonics (with positive and negative values) is shown in Fig. 9.5.

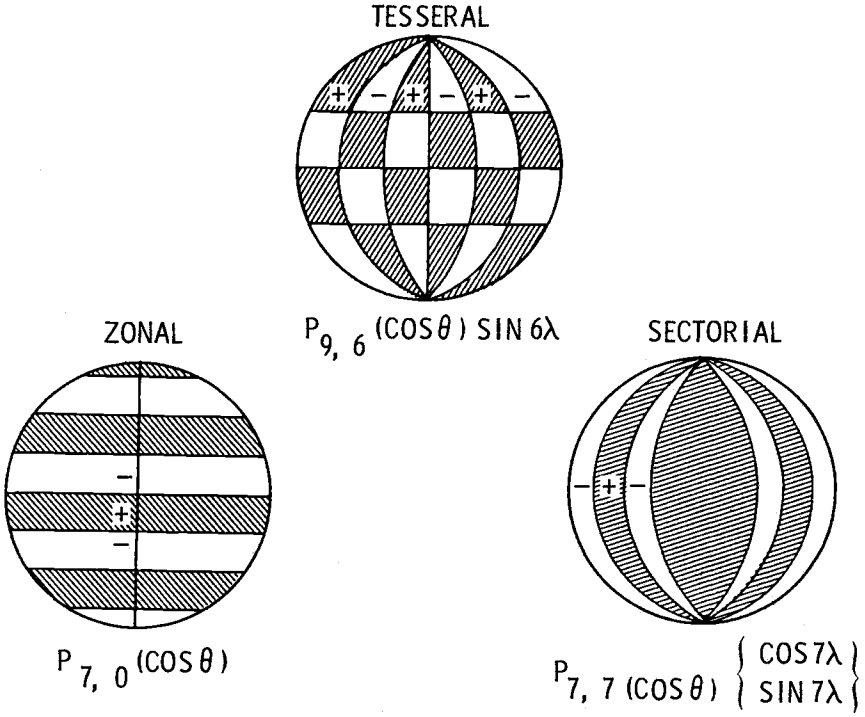


Fig. 9.5 Spherical harmonics.

9.5 Perturbations Due to Oblateness (J_2)

The potential function including the J_2 effect can be given as

$$\Phi = \frac{\mu}{r} + R \tag{9.42}$$

where R is called the perturbing or disturbing function due to J_2, and it may be given either in terms of the orbital parameters as

$$R = \frac{\mu J_2 A_e^2}{2r^3}\left[1 - \frac{3}{2}\sin^2 i(1 - \cos 2u)\right] \tag{9.43}$$

or in terms of the Cartesian coordinates as

$$R = \frac{\mu J_2 A_e^2}{2r^3}\left[1 - 3\left(\frac{z}{r}\right)^2\right] \tag{9.44}$$

In the disturbing function approach, the equations of variation can be obtained by taking partial derivatives of R and using Eq. (9.28); or the perturbing forces along the x, y, z components can be derived by taking partial derivatives of R and using Eq. (9.21). In deriving equations of element variations, the function R must be expanded in terms of the orbit element before the partial differentiation

takes place. After some mathematical manipulations, the time rate of the six orbit elements can be obtained. For illustrative purposes, only the rate of the ascending node Ω is listed next.

$$\begin{aligned}\frac{\mathrm{d}\Omega}{\mathrm{d}t} = -n\frac{3}{2}J_2\left(\frac{a_e}{p}\right)^2 \cos i\Bigg[&1 + 3e\cos M + \frac{9}{2}e^2\cos 2M \\ &-\frac{1}{48}e^3\cos(M-2\omega) + \frac{1}{2}e\left(1-\frac{13}{8}e^2\right)\cos(M+2\omega) \\ &+(1-4e^2)\cos(2M+2\omega) - \frac{7}{2}e\cos(3M+2\omega) \\ &-\frac{17}{2}e^2\cos(4M+2\omega) + \cdots\Bigg]\end{aligned} \tag{9.45}$$

It is interesting to point out that the series equation (9.45) contains a secular term, the first constant term in the bracket, and short-period terms, $\cos M$, $\cos 2M$, etc. The short-period terms are bounded by their coefficients, and the secular term is a linear function of time (to the first order). For near-circular orbits, all the short-period terms are very small except the term $\cos(2M + 2\omega)$. The equations for $\mathrm{d}\omega/\mathrm{d}t$ and $\mathrm{d}M/\mathrm{d}t$ are similar to Eq. (9.45) except that some of the short-period terms have $1/e$ in the coefficient. These terms will become singular when e is approaching zero. The rate equations for $\mathrm{d}a/\mathrm{d}t$, $\mathrm{d}e/\mathrm{d}t$ and $\mathrm{d}i/\mathrm{d}t$ are somewhat different in that they do not contain secular terms and the short-period terms are all sine terms. Furthermore, the coefficients do not contain the $1/e$ factor.

Those rate equations or the variational equations of the classical elements can also be derived via the force components approach. The perturbing accelerations due to J_2 may be given in the spherical coordinate system as

$$\begin{aligned}F_r &= -\frac{3}{2}\frac{\mu J_2 a_e^2}{r^4}\left[1 - \frac{3}{2}\sin^2 i(1-\cos 2u)\right] \\ F_s &= -\frac{3}{2}\frac{\mu J_2 a_e^2}{r^4}\sin^2 i\sin 2u \\ F_w &= -\frac{3}{2}\frac{\mu J_2 a_e^2}{r^4}\sin i\cos i\sin u\end{aligned} \tag{9.46}$$

By substituting the preceding accelerations into Eq. (9.21) and carrying out the expansion, one should obtain the same set of rate equations as the ones derived through the perturbing function approach. When the perturbing force is from the third-body attraction, the disturbing function and force components are as follows:

$$\begin{aligned}R &= k^2 m_3\left(\frac{1}{r_{23}} - \frac{\boldsymbol{r}\cdot\boldsymbol{r}_3}{r_3^3}\right) \\ F_r &= k^2 m_3\left(\frac{r_3\cos\Psi - r}{r_{23}^3} - \frac{\cos\Psi}{r_3^2}\right) \\ F_s &= k^2 m_3 \boldsymbol{s}\cdot\boldsymbol{r}_3\left(\frac{1}{r_{23}^3} - \frac{1}{r_3^3}\right) \\ F_w &= k^2 m_3 \boldsymbol{w}\cdot\boldsymbol{r}_3\left(\frac{1}{r_{23}^3} - \frac{1}{r_3^3}\right)\end{aligned} \tag{9.47}$$

where

m_3 = mass of the third body (sun or moon)
Ψ = angle between $\boldsymbol{r}$ and $\boldsymbol{r}_3$
$r_{23} = |\boldsymbol{r} - \boldsymbol{r}_3|$, distance form the satellite to the third body
r_3 = distance from the central body (Earth) to the third body
$\boldsymbol{s}$ = unit vector along the F_s direction
$\boldsymbol{w}$ = unit vector along the F_w direction

9.6 Integration of the Equations of Variation

Analytical

As an example, the integration of $\mathrm{d}\Omega/\mathrm{d}t$ may be written as

$$\begin{aligned}\Omega(t) &= \Omega(t_0) + \int_{t_0}^{t} \frac{\mathrm{d}\Omega}{\mathrm{d}t}\mathrm{d}t \\ &= \Omega_0 - \frac{3}{2}J_2 \int_{t_0}^{t} n\left(\frac{a_e}{p}\right)^2 \cos i[1 + 3e\cos M + \cdots]\mathrm{d}t\end{aligned} \tag{9.48}$$

where

$$M = M_0 + n(t - t_0)$$

The first-order theory assumes that the parameters n, p, e, and i are constant. Thus, the integration can be carried out analytically term by term as

$$\begin{aligned}\Omega(t) = \Omega_0 - \frac{3}{2}J_2\left(\frac{a_e}{p}\right)^2 \cos i[n(t - t_0) \\ + 3e(\sin M - \sin M_0) + \cdots]\end{aligned} \tag{9.49}$$

In the first-order theory, the elements n, a, e, i, Ω and ω appearing on the right-hand side are assumed to be constant and equal to the epoch values. The validity of the first-order solution is limited to a rather short interval, during which the change in M, Ω and ω due to J_2 is small. To improve the first-order solution, the secular rates of M, Ω and ω determined from the first-order theory are included in evaluating these angles on the right-hand side of the equations of variation.[6,7] This modified method is sometimes called the "first-and-half-order theory." The accuracy of the modified solution has been found to be quite good compared with numerical integration (Figs. 9.6–9.9). When the variations in a, e, and i are significant, as revealed by Figs. 9.8 and 9.9, the second-order solutions should be used. The second-order solutions are generally quite time-consuming in derivation and very complex in the final form. In actual application, the long-term orbit propagation is usually done with a semianalytical averaging method.

Semianalytical (Method of Averaging)

The concept of this method is first to obtain a set of variational equations for the mean rates of the elements and then to integrate numerically these averaged equations with very large step size of one orbit period. After the integration,

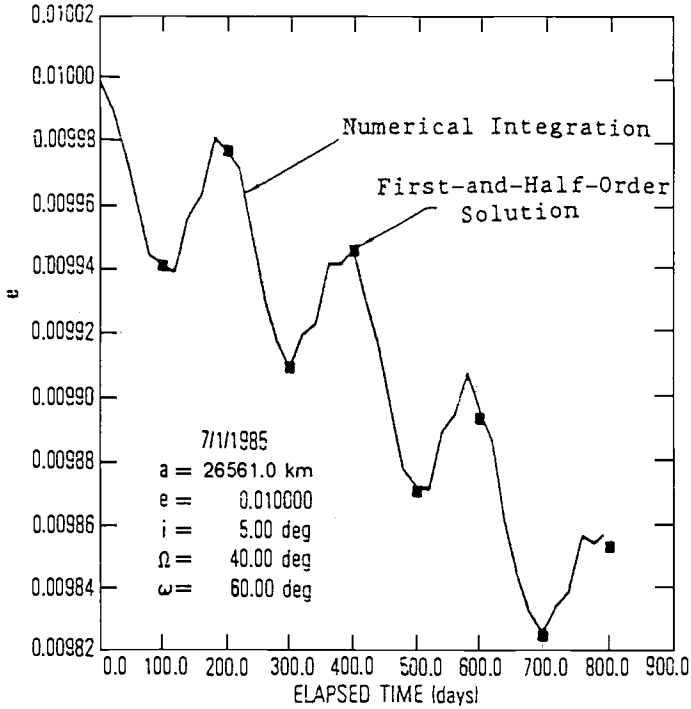


Fig. 9.6 Eccentricity time history for low-inclination orbit (i = 5 deg).

the short-period variations computed from the analytical first-order solution are added to the mean element at a specified time. The procedure of this method can be illustrated as follows:

$$\frac{\mathrm{d}\bar{\Omega}}{\mathrm{d}t} = \frac{1}{2\pi}\int_0^{2\pi} \frac{\mathrm{d}\Omega}{\mathrm{d}t}\mathrm{d}M = -\frac{3}{2}nJ_2\left(\frac{a_e}{p}\right)^2 \cos i \tag{9.50}$$

$$\bar{\Omega}(t) = \Omega_0 + \int_{t_0}^{t} \frac{\mathrm{d}\bar{\Omega}}{\mathrm{d}t}\mathrm{d}t + \delta\Omega_{\text{short period}} \tag{9.51}$$

$$\Omega(t) = \bar{\Omega}_{\text{mean}}(t) + \delta\Omega_{\text{short period}}^{(\bar{a},\bar{e},\bar{i},\bar{\Omega},\bar{\omega},\bar{M})} \tag{9.52}$$

The elements with a bar are the mean (average over one orbit period) orbit elements. The accuracy and efficiency of this semianalytical method have made it a useful tool for long-term orbit propagation and control. The computational speed of the averaging method is generally about 40 or more times faster than the numerical integration of Cowell's method. The short-period variations of J_2 and the mean rates due to J_2, J_3, and J_4 can be found in Ref. 8. The mathematical formulation of semianalytic theory in equinoctial orbit elements is given in Ref. 11. Similar formulation in classical elements can be found in Ref. 12.

A summary of the advantages and disadvantages of the special perturbation method and the general perturbation method, together with their applications, is provided in Table 9.3.

Table 9.3 Summary of special and general perturbation methods

	Special perturbation	General perturbation
Advantages	High precision General purpose Simplicity of formulation	Analytical Fast computation
Disadvantages	Slow computation	Orbit-type-dependent Time-consuming formulation
Applications	Orbit determination	Mission design and analysis
	High-accuracy orbit	Long-term orbit
	Prediction	Prediction
	Navigation	Parametric study

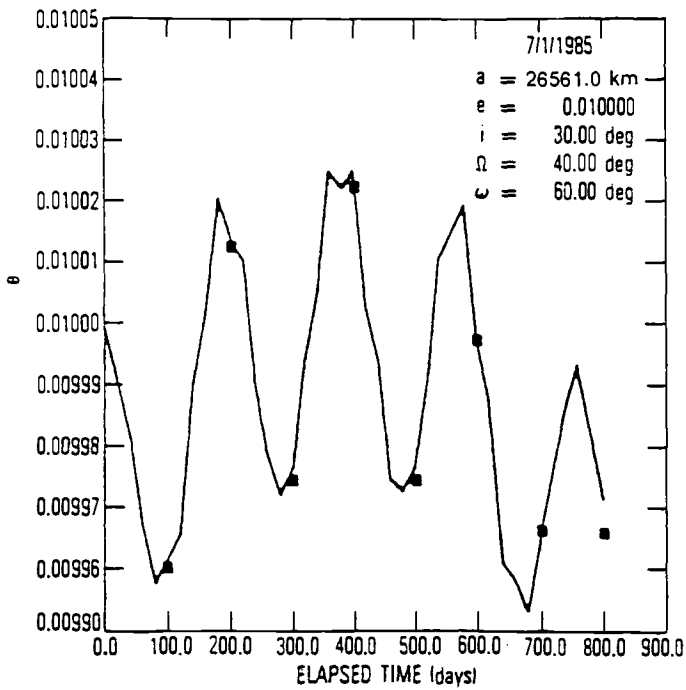


Fig. 9.7 Eccentricity time history for moderate inclination orbit (i = 30 deg).

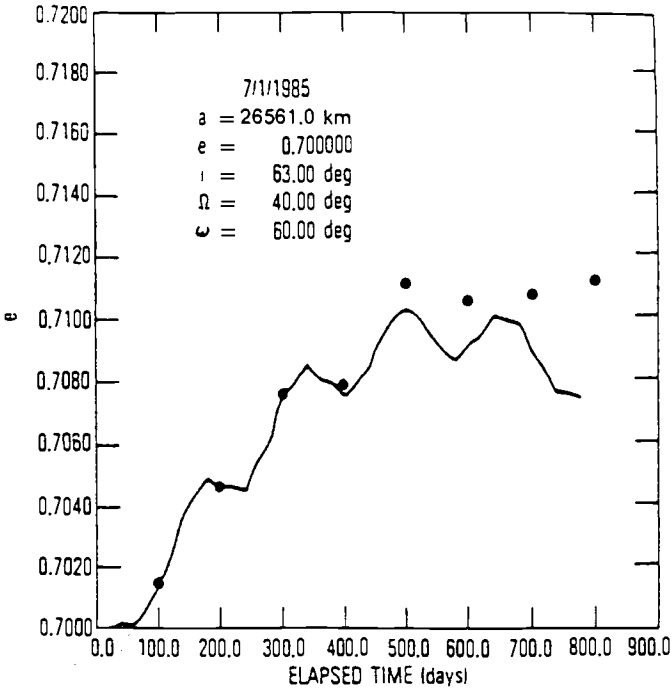


Fig. 9.8 Eccentricity time history for high-eccentricity orbit (i = 63 deg).

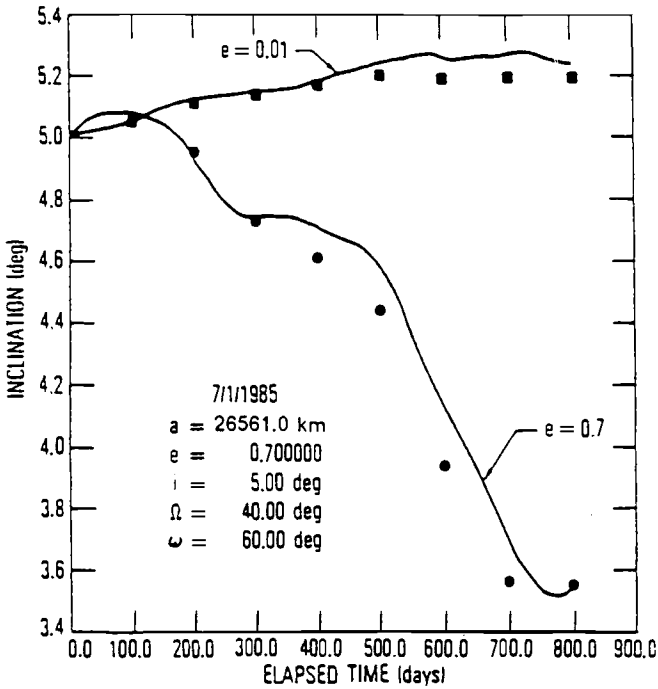


Fig. 9.9 Inclination time history for low-inclination orbit (i = 5 deg).

References

[1]Broucke, R. A., and Cefola, P. J., "On the Equinoctial Orbit Elements," *Celestial Mechanics,* Vol. 5, 1972.

[2]Herrick, S., *Astrodynamics,* Vol. II, Van Nostrand Reinhold, London, 1972.

[3]Brouwer, D., and Clemence, G., *Methods of Celestial Mechanics,* Academic, New York, 1961.

[4]Bate, R. R., Mueller, D. D., and White, J. E., *Fundamentals of Astrodynamics,* Dover, New York, 1971.

[5]Fox, K., "Numerical Integration of the Equations of Motion of Celestial Mechanics," *Celestial Mechanics,* Vol. 33, 1984.

[6]Chao, C. C., "A General Perturbation Method and Its Application to the Motions of the Four Massive Satellites of Jupiter," Ph.D. Dissertation, UCLA, Los Angeles, CA, 1976.

[7]Chao, C. C., "An Analytical Integration of the Averaged Equations of Variation Due to Sun-Moon Perturbations and Its Application," The Aerospace Corp., El Segundo, CA, Rept. SD-TR-80-12, Oct. 1979.

[8]Liu, J.J.F., and Alford, R. L., "Semianalytic Theory for a Close-Earth Artificial Satellite," *Journal of Guidance and Control,* Vol. 3, July–Aug. 1980.

[9]Cefola, P. J., and Broucke, R. A., "On the Formulation of the Gravitational Potential in Terms of Equinoctial Variables," AIAA Paper 75-9, Jan. 1975.

[10]Kaula, W. M., *Theory of Satellite Geodesy,* Blaisdell, Waltham, MA, 1966.

[11]Danielson, D. A., Sagovac, C. P., Neta, B., and Early, L. W., "Semianalytic Satellite Theory," Naval Postgraduate School Report NPS-MA-95-002, Feb. 1995.

[12]Roy, A. E., *Orbital Motion,* 3rd ed., Adam Hilger, Bristol, UK, 1988.

10
Applications of Orbit Perturbations

This chapter provides a description of each of the perturbation forces and considers the actual applications to mission design and analysis of Earth satellite missions. The various perturbation forces are: Earth's oblateness and its triaxiality, the lunisolar gravitational attractions, the solar-radiation pressure, the Earth atmospheric drag, and the tidal friction effects. The interesting properties of orbits with critical inclination (Molniya orbit), sun-synchronous orbits, frozen orbits, and the long-term coupling between sun-moon attractions and Earth oblateness effects will be discussed.

10.1 Earth's Oblateness (J_2) Effects

The principal effects of the J_2 zonal harmonic or Earth's oblateness are secular motions of the node Ω and perigee ω of an orbit. The westward rotation of the orbit plane is illustrated in Fig. 10.1, where S denotes the longitude difference between two successive crossings of the equator.

The motion of the node Ω occurs because of the added attraction of the Earth's equatorial bulge, which introduces a force component toward the equator. The resultant acceleration causes the satellite to reach the equator (node) short of the crossing point for a spherical Earth. The orbit thus regresses a delta node ($\Delta\Omega$) each revolution. This effect may also be regarded as gyrodynamic precession due to a torque acting on the satellite angular momentum vector caused by the additional attraction of the Earth's equatorial bulge. The secular nodal regression rate can be numerically evaluated to first order in the dominant oblateness parameter J_2 from

$$\dot{\Omega} = \frac{-9.9358}{(1-e^2)^2}\left(\frac{R}{R+\bar{h}}\right)^{3.5}\cos i\,\frac{\text{deg}}{\text{mean solar day}} \tag{10.1}$$

where

$$\bar{h} = \frac{h_a + h_p}{2}, \qquad e = \frac{h_a - h_p}{h_a + h_p + 2R}$$

and R is the equatorial radius of the Earth. It should be noted that the node regresses for direct orbits (0 deg $< i <$ 90 deg) and advances for retrograde orbits (90 deg $< i <$ 180 deg). Furthermore, there is no nodal regression to first order for polar orbits. Figure 10.2 illustrates the nodal regression rate vs inclination for various values of average altitude.

The secular motion of the perigee (line of apsides) occurs because the force is no longer proportional to the inverse square radius and the orbit is consequently no longer a closed ellipse. This is illustrated in Fig. 10.3.

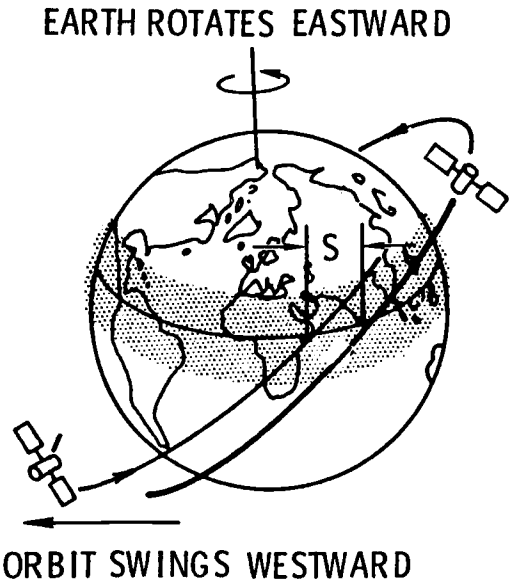


Fig. 10.1 The gravitation pull of the Earth's equatorial bulge causes the orbital plane of an eastbound satellite to swing westward.

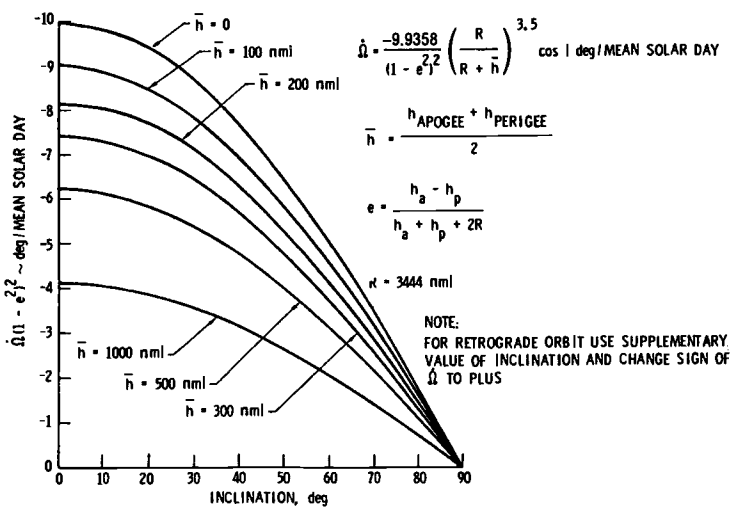


Fig. 10.2 Regression rate due to oblateness vs inclination for various values of average altitude.

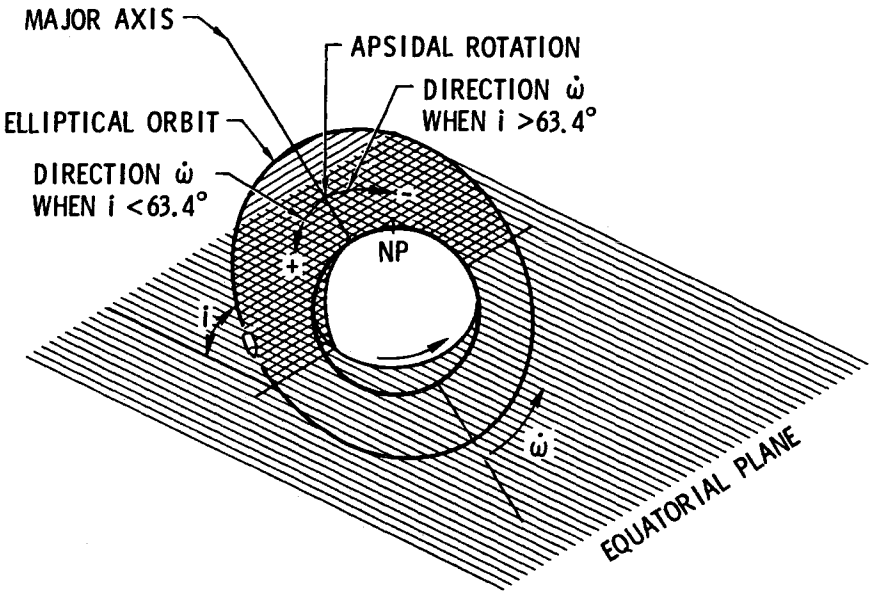


Fig. 10.3 Apsidal rotation.

The apsidal rotation rate $\dot{\omega}$ can be expressed as

$$\dot{\omega} = \frac{9.9358}{(1-e^2)^2}\left(\frac{R}{R+\bar{h}}\right)^{3.5}\left(2-\frac{5}{2}\sin^2 i\right)\frac{\text{deg}}{\text{mean solar day}} \tag{10.2}$$

The apsidal rate is shown in Fig. 10.4 as a function of orbit plane inclination i, with $\bar{h}$ as a parameter.

10.2 Critical Inclination

Equation (10.2) indicates that the apsidal rotation rate of an orbit with J_2 effects is zero when the factor $(2 - \frac{5}{2}\sin^2 i)$ is zero, or

$$\sin i = \frac{2}{\sqrt{5}} \tag{10.3}$$

which yields

$$i = 63.43 \quad \text{or} \quad 116.57 \text{ deg}$$

These two inclination values are called the critical inclination when $\dot{\omega} = 0$. It is interesting to note that the values of the critical inclination do not depend on the value of J_2 and other orbit elements. In other words, the two values of critical inclination are true for any orbits around Earth or any other celestial body with J_2 effect. The application of this orbit property can be found in highly eccentric

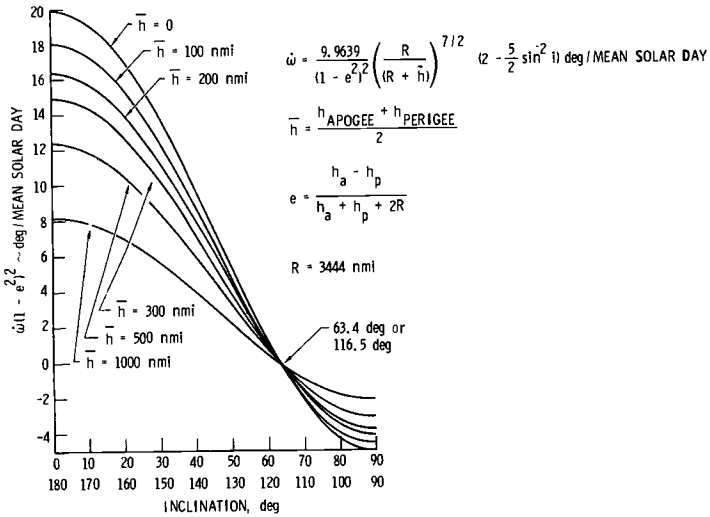


Fig. 10.4 Apsidal rotation rate due to oblateness vs inclination for various values of average altitude.

($e = 0.73$), critically inclined 12-h orbits of the so-called Molniya type. When the argument of perigee is set at 270 deg, the satellite will spend most of the time near apogee and thus stay in view of those desirable ground stations on the northern hemisphere. The nearly 12-h orbit will produce repeating ground tracks, which are required by some missions. Because of the unique property of critical inclination, the argument of perigee remains constant at 270 deg or other preferred values. Consequently, the desirable viewing geometry would not be destroyed.

In actual application, the argument of perigee does not remain at 270 deg or at other initial values indefinitely because of the sun-moon attractions and possible higher-order Earth gravity harmonics. However, these effects are usually very small and can be adjusted with stationkeeping maneuvers.

10.3 Sun-Synchronous Orbits

Sun-synchronous orbits are orbits with the secular rate of the right ascension of the ascending node equal to the right ascension rate of the mean sun. For the satellite orbit to be sun-synchronous, the inclination, semi-major axis, and eccentricity must satisfy the following relation:

$$\left(\frac{d\Omega}{dt}\right)_s = -\frac{3}{2} n J_2 \left(\frac{R}{p}\right)^2 \cos i = \dot{\alpha}_\odot = 0.9856 \text{ deg/day} \tag{10.4}$$

where

$n = \sqrt{\mu/a^3}$ orbit mean motion
R = Earth equatorial radius
$p = a(1 - e^2)$
$\dot{\alpha}_\odot$ = right ascension rate of the mean sun

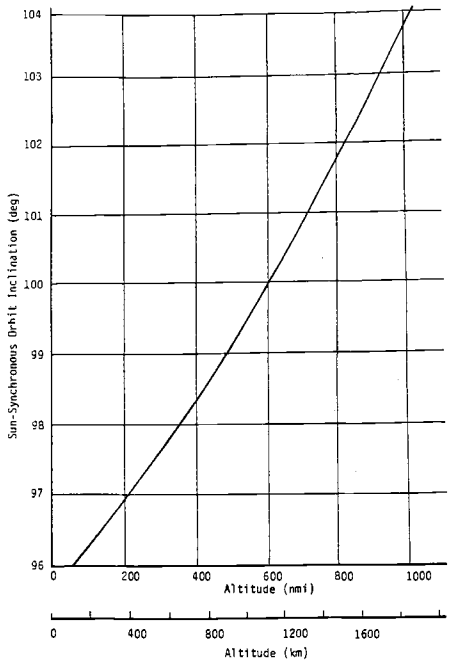


Fig. 10.5 Sun-synchronous inclination for low-altitude circular orbits.

A typical sun-synchronous orbit has a 98.7-deg inclination and 833 km mean orbit altitude. For low-altitude circular orbits, the sun-synchronous inclination and altitude are computed from Eq. (10.4) and plotted in Fig. 10.5. A plot for high altitude orbits is given in Chap. 11. The unique property of the sun-synchronous orbits is that the satellite's ground track has one local time on its ascending half and another local time (12 h away) on its descending half. The two local times remain the same for the entire mission. The sun-synchronous orbits are often referred to by the local time of the ascending node as 6 a.m. orbit, 10 a.m. orbit, etc. The local times at various satellite subpoints of the ground track can be computed by the following relation:

$$t = t_{\text{node}} - \{\tan^{-1}[\tan u \cos(\pi - i)]\} \div \omega_e \tag{10.5}$$

where

t_{node} = local time of the ascending node crossing
u = argument of latitude of the satellite
i = inclination of the orbit
ω_e = Earth rotation rate

There are two kinds of perturbations, the drag and sun's attraction, that will affect the sun-synchronous property and gradually change the local time. The drag perturbs the orbit parameters p and n in Eq. (10.4), and the sun's attraction perturbs the orbit inclination i. The sun-synchronous property induces the deep resonance in the equation of the inclination variation due to solar perturbations.

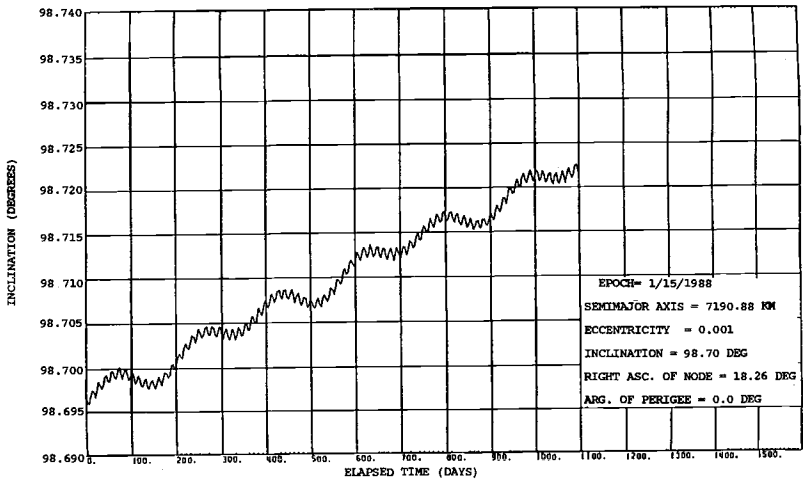


Fig. 10.6 Inclination variation without correction (5:30 orbit).

The particular term of the deep resonance that causes a drift in inclination may be given as[1]

$$\frac{\mathrm{d}i}{\mathrm{d}t} = -0.047(\text{deg/year}) \sin 2(\alpha_{\odot} - \Omega) \tag{10.6}$$

with

$$\alpha_{\odot} = \text{right ascension of the sun}$$

and

$$\dot{\alpha}_{\odot} = \dot{\Omega} \text{ (deep resonance)}$$

The largest inclination drift is 0.047 deg/yr when the angle $2(\alpha_{\odot} - \Omega)$ is either 90 or 270 deg. Figure 10.6 shows a 3-yr inclination variation without correction for the nominal N-ROSS orbit with 5:30 a.m. local time. The corresponding value of $(\alpha_{\odot} - \Omega)$ is 277.5 deg, and the inclination rate as computed from the preceding equation is 0.012 deg/yr, which agrees with the mean slope of the inclination variation in Fig. 10.6. The inclination history shown in Fig. 10.6 was generated from high-precision numerical integration. For a further discussion of sun-synchronous orbits, see Chap. 11.

10.4 J_3 Effects and Frozen Orbits

The third harmonic J_3 is of the order of $10^{-3}J_2$ for the Earth, so that the amplitudes of the short-period perturbations are very small. However, in the equations of long-term variations for eccentricity, the eccentricity appears in the denominator of one term and thus will increase the magnitude of the term to nearly the

magnitude of J_2 if the eccentricity is small enough. As a result, the eccentricity will have a long-term variation of the following form[1]:

$$\Delta e = -\frac{1}{2}\frac{J_3}{J_2}\left(\frac{R}{a}\right)\sin i \sin\omega \tag{10.7}$$

Equation (10.7) indicates that the J_3-induced variation in eccentricity will have a sinusoidal oscillation with a period of $2\pi/\dot{\omega}$ and an amplitude of $(J_3/J_2)\ (R/a)$, where R is the Earth equatorial radius. Similarly, there is a corresponding long-term variation in inclination with the following form[1]:

$$\Delta i = \frac{1}{2}\frac{J_3}{J_2}\left(\frac{R}{a}\right)\frac{e}{1-e^2}\cos i \sin\omega \tag{10.8}$$

It is seen that the induced inclination variation will become significant when the eccentricity is not too small and the inclination is not too large.

Through the coupling effect of J_2 and J_3, the concept of frozen orbit was introduced.[19] The frozen orbit can be achieved by properly selecting a particular combination of initial eccentricity and inclination for a given orbit period such that there will be no variation to the order of J_3 in eccentricity and argument of perigee. In other words, the orbit will be "frozen" in the inertial space, and the orbit altitude history will repeat every revolution. This property may be affected by the solar-radiation pressure and drag perturbations. Thus, in actual application, small orbit maintenance maneuvers will have to be performed periodically to offset those perturbations. For examples of frozen orbits, see Chap. 11.

10.5 Earth's Triaxiality Effects and East-West Stationkeeping

The Earth cross section along the equatorial plane is not a circle but more like an ellipse, as shown in Fig. 10.7. This ellipse-shaped cross section is represented by the so-called tesseral harmonics (C_{22}, S_{22}, C_{32}, S_{32} ...). The primary tesseral harmonic is designated by J_{22}, which combines C_{22} and S_{22}. The longitude of symmetry of the J_{22} harmonic denoted by λ_{22} is determined from observations and has a typical value of -14.7 deg. The equilibrium points are divided into stable (75.3°E and 255.3°E) and unstable (14.7°W and 165.3°E) longitudes. The resonance effect of the 24-h geosynchronous orbits induces a very slow motion,

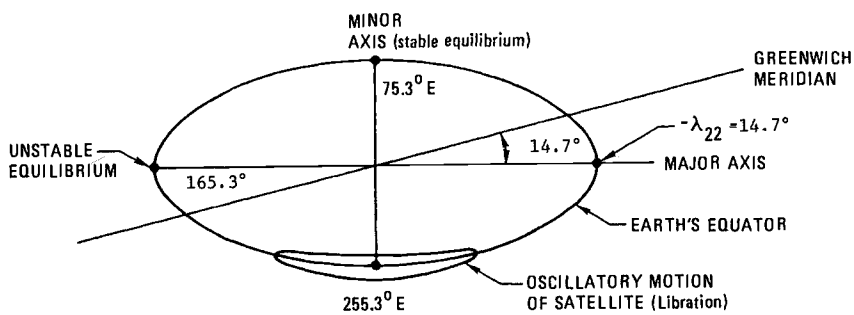


Fig. 10.7 Geosynchronous orbit libration.

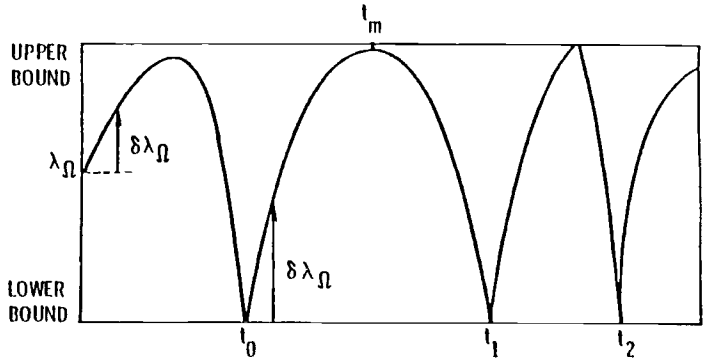


Fig. 10.8 Schematic drawing of a longitude stationkeeping method.

librating around the stable longitudes with a period of about 800 days and an amplitude of nearly 90 deg.

Methods of east-west stationkeeping were developed and discussed in Refs. 8, 9, 11, and 12. The concept is to apply a ΔV maneuver at t_0, where the longitude is about to exceed the tolerance boundary, as shown in Fig. 10.8. The magnitude of the ΔV is determined such that the resonance effect will bring the satellite back before it exceeds the other bound at t_m. The typical longitude tolerance is ± 1 deg, and the average time between two maneuvers is about 100 days. A much tighter tolerance of ± 0.1 deg is required by some missions. The stationkeeping ΔV per year is a function of longitude and can be found in Fig. 8.1.

10.6 Third-Body Perturbations and North/South Stationkeeping

Perturbations due to Earth's oblateness become less important with increasing distance from the Earth. On the other hand, the perturbations from the gravitational force of the moon and sun become more important at higher altitudes.

The combined effects of the solar and lunar gravitational attraction on high-altitude Earth orbits have been studied analytically by several investigators.[2–7] The combined lunisolar gravitational perturbations will induce long-term and/or secular variations in high-altitude orbit eccentricity, inclination, node, and argument of perigee.

The orbit inclination perturbations are caused primarily by the attractions of the sun and moon. For a geosynchronous orbit with initial zero inclination, the sun/moon perturbations increase the orbit inclination at a rate of about 1 deg/yr for the first 10 years, and then the inclination reaches 15 deg in about 17 yr.[10] After that, the inclination decreases to zero in another 27 yr. For most geosynchronous satellite missions, the orbit inclination is not controlled, primarily because of the costly ΔV expenditure in plane change maneuvers. Instead, the initial inclination and right ascension of the ascending node are properly chosen such that the inclination will first decrease to zero and increase to the initial value near the end of the mission. For example, when the initial inclination is selected at 3.5 deg, the inclination will decrease to zero and then increase to 3.5 deg in approximately 7 yr, depending on the epoch. The corresponding initial node is around 270 deg, which must be determined by iteration. The long-term inclination variations will be discussed later in Sec. 10.10.

For missions like DSCS III, north-south stationkeeping is required to keep inclination variation small, less than 0.1 deg. The inclination is controlled by applying ΔV normal to the orbit plane at desirable locations. The nominal locations are at the ascending or descending node. However, for eccentric orbits, the location may be at other points on the orbit to minimize ΔV. The amount of ΔV is determined from the desired inclination correction with the following equation:

$$\Delta V = 2V \sin \frac{\Delta\Theta}{2} \tag{10.9}$$

where V is the circumferential velocity at the point of correction, and $\Delta\Theta$ is the total plane change given by

$$\Delta\Theta = \cos^{-1}[\sin i_1 \sin i_2 \cos(\Omega_2 - \Omega_1) + \cos i_1 \cos i_2] \tag{10.10}$$

In Eq. (10.10), i and Ω are the inclination and right ascension of the ascending node of the orbit plane, respectively. The subscripts 1 and 2 denote conditions before and after the inclination correction.

For a 0.1-deg north-south stationkeeping tolerance, the ΔV required is of the order of 50 m/s, with four to five inclination control maneuvers per year.

10.7 Solar-Radiation-Pressure Effects

The solar-radiation pressure is induced by the light energy (photons) radiated from the sun. At one A.U. (astronomical unit), the solar-radiation-pressure constant P_0 is 4.65×10^{-5} dyne/cm^2. The perturbing acceleration of an Earth satellite due to solar-radiation-pressure effects can be computed by means of the following equation:

$$a_p = 4.65 \times 10^{-6}(1 + \beta)\left(\frac{A}{m}\right)\left(\frac{a_\odot}{r_\odot}\right)^2 \tag{10.11}$$

where a_p is the magnitude of solar-radiation-pressure acceleration, and $a_\odot/r_\odot$ is approximately 1 for near-Earth orbits. Also,

$$\beta = \text{optical reflection constant}$$

where

$$\begin{cases} \beta = 1 \text{ total reflection (mirror)} \\ \beta = 0 \text{ total reception (blackbody)} \\ \beta = -1 \text{ total transmission (transparent)} \end{cases}$$

A = effective satellite projected area, m^2

m = total satellite mass, kg

$r_\odot, a_\odot$ = semimajor axis and radius of the sun's orbit around Earth

The direction of the acceleration a_p is perpendicular to the effective area A, which may or may not be perpendicular to the sun's ray. For a sphere or a flat plate perpendicular to the sun's ray, the normalized components of solar-radiation

acceleration can be expressed by the following equations:

$$
\begin{aligned}
\begin{Bmatrix} F_r \\ F_s \end{Bmatrix} &= \cos^2\frac{i}{2}\cos^2\frac{\varepsilon}{2}\begin{Bmatrix}\cos\\ \sin\end{Bmatrix}(\lambda_\odot - u - \Omega) \\
&\quad - \sin^2\frac{i}{2}\sin^2\frac{\varepsilon}{2}\begin{Bmatrix}\cos\\ \sin\end{Bmatrix}(\lambda_\odot - u + \Omega) \\
&\quad - \frac{1}{2}\sin i\sin\varepsilon\left[\begin{Bmatrix}\cos\\ \sin\end{Bmatrix}(\lambda_\odot - u) - \begin{Bmatrix}\cos\\ \sin\end{Bmatrix}(-\lambda_\odot - u)\right] \qquad (10.12) \\
&\quad - \sin^2\frac{i}{2}\cos^2\frac{\varepsilon}{2}\begin{Bmatrix}\cos\\ \sin\end{Bmatrix}(-\lambda_\odot - u + \Omega) \\
&\quad - \cos^2\frac{i}{2}\sin^2\frac{\varepsilon}{2}\begin{Bmatrix}\cos\\ \sin\end{Bmatrix}(-\lambda_\odot - u - \Omega) \\
F_w &= \sin i\cos^2\frac{\varepsilon}{2}\sin(\lambda_\odot - \Omega) \\
&\quad - \sin i\sin^2\frac{\varepsilon}{2}\sin(\lambda_\odot + \Omega) - \cos i\sin\varepsilon\sin\lambda_\odot
\end{aligned}
$$

where F_r, F_s, and F_w are components of acceleration along the satellite orbit radius vector, perpendicular to F_r in the orbital plane, and along the orbit normal, respectively. The parameters, i, u, and Ω are orbit parameters defined in Chapters 3 and 9.

Here ε denotes the obliquity of the ecliptic, and $\lambda_\odot$, the ecliptic longitude of the sun. The quantities ε, $\lambda_\odot$, and $a_\odot/r_\odot$ and can be computed with sufficient accuracy from the expressions (see *Explanatory Supplement to the Astronomical Ephemeris*, 1961, p. 98).

$$
\begin{aligned}
d &= \text{MJD} - 15019.5 \\
\varepsilon &= 23^\circ.44 \\
M_\odot &= 358^\circ.48 + 0^\circ.98560027d \\
\lambda_\odot &= 279^\circ.70 + 0^\circ.9856473d + 1^\circ.92\sin M_\odot \\
a_\odot/r_\odot &= [1 + 0.01672\cos(M_\odot + 1^\circ 92\sin M_\odot)]/0.99972
\end{aligned}
$$

where MJD is the modified Julian day. The modified Julian day = Julian day − 2400000.5.

By substituting the components F_r, F_s, and F_w into the equation of (9.21) of Chapter 9, one would obtain the variations of orbit elements due to radiation-pressure effects. After examining the dominant terms in the equations of variation, the most significant effect is on orbit eccentricity. For geosynchronous orbits, the dominant term in the equation of variation for eccentricity is $\sin(\lambda_\odot - \omega - \Omega)$. This term suggests that the eccentricity has a long-period (of the year) variation, with the magnitude depending on β, A/W and initial conditions. The same variation appears in the argument of perigee. By properly choosing a unique set of initial conditions, the long-period eccentricity variation will disappear, and a

constant eccentricity, sometimes called the "forced eccentricity" or "resonance eccentricity," will exist as a result of the resonance property discussed in Ref. 11. Therefore, for missions requiring very small eccentricity (or circular orbit), the design of the spacecraft should take into account reducing A/W. Regular eccentricity control maneuvers may be necessary to keep the eccentricity value small in the presence of significant solar-radiation-pressure effects.

An Algorithm to Compute Eccentricity Variations of Geosynchronous Orbits

For a geosynchronous orbit, the eccentricity variations due to solar-radiation pressure can be expressed by a simple relation in closed form[11] as follows:

$$e = \left[\left(\frac{g}{z}\right)^2 - 2\rho\left(\frac{g}{z}\right)\cos\theta + \rho^2\right]^{\frac{1}{2}} \tag{10.13}$$

where

$$\rho = \left[e_0^2 - \left(\frac{2g}{z}\right)e_0\cos\phi_0 + \left(\frac{g}{z}\right)^2\right]^{\frac{1}{2}} \tag{10.14}$$

$$\theta_0 = \sin^{-1}\left(\frac{e_0}{\rho}\sin\phi_0\right) \tag{10.15}$$

$$\phi_0 = \lambda_\theta - \omega_0 - \Omega_0 \tag{10.16}$$

$$g = \frac{3}{2}\frac{(1+\beta)(A/M)P}{V}\left(1 - e_0^2\right)^{\frac{1}{2}}\cos^2\left(\frac{\varepsilon}{2}\right) \tag{10.17}$$

and

e_0 = initial orbit eccentricity
A/M = cross-sectional area/mass ratio of the satellite
P = 4.65×10^{-6} N/m^2
$\lambda_\odot$ = ecliptic longitude of the sun at epoch
ω_0 = argument of perigee at epoch
Ω_0 = right ascension of ascending node at epoch
z = $\dot{\lambda}_\odot - \dot{\Omega}$
Θ = $\Theta_0 + z(t - t_0)$
V = velocity of satellite
ε = obliquity of the ecliptic $\approx$ 23.5 deg
$\dot{\lambda}_\odot$ = rate of the ecliptic longitude of the sun ($\approx$ 1 deg/day)
$\dot{\Omega}$ = nodal regression rate

From the point of view of physical phenomena, the long-term eccentricity variation may be explained by the following analogy with energy addition and subtraction. Consider a circular orbit with counterclockwise motion, as shown in

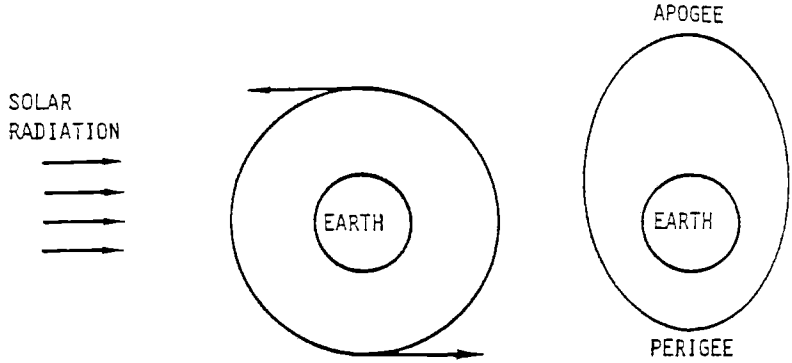


Fig. 10.9 Solar-radiation-pressure effect on an Earth satellite.

the center of Fig. 10.9. The solar radiation coming from the left is equivalent to adding a ΔV at the lower end of the orbit and subtracting the same ΔV at the upper end of the orbit. As a result, the orbit gradually evolves into an elliptical orbit like the one at the right of Fig. 10.9. Six months later, the sun is at the opposite side of the orbit, and the ΔV reverses sign at both ends of the elliptical orbit. Hence, the orbit will be circularized gradually because of the radiation-pressure effect.

For interplanetary missions, the solar-radiation-pressure effect was once considered by engineers at JPL as a means to navigate spacecraft for inner planet encounters. This concept is called "solar sailing." By properly orienting the "sails" of the spacecraft, a desired component along the velocity vector may be achieved. Thus, solar energy can be added or subtracted to the orbit, as shown by Fig. 10.10.

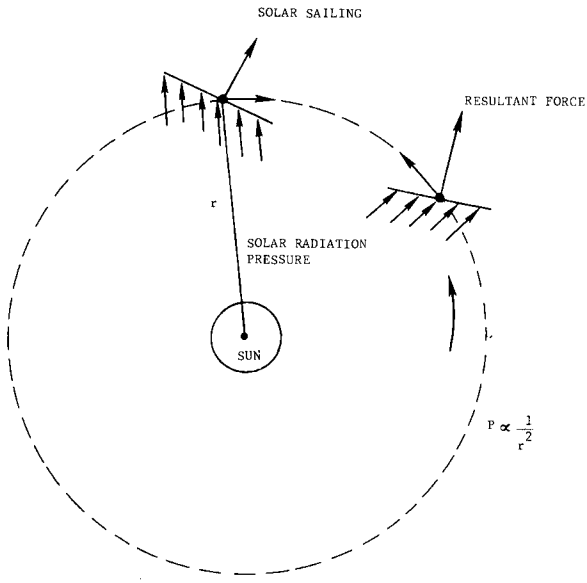


Fig. 10.10 Concept of solar sailing.

Study results have demonstrated that the specially guided energy transfer can, in fact, be used to navigate the spacecraft to its target planet. However, two technical problems hinder the realization of this concept. First, the solar-radiation pressure is inversely proportional to the square of sun-spacecraft distance. Thus, this technique is effective when the spacecraft is closer to the sun. The heat of the radiation poses great difficulty in manufacturing a spacecraft that can resist the severe heat. Second, the large area and light weight make it another very difficult task to design and manufacture such a spacecraft.

10.8 Atmospheric Drag Effects

The Earth's atmosphere produces drag forces that retard a satellite's motion and alter the orbit shape. A low-altitude satellite will eventually be slowed until it spirals in. The rate of orbital decay depends on atmospheric density; this varies with time and geographic position and is not precisely known. Orbital parameters such as height, and the ballistic coefficient, which takes a spacecraft shape and weight into account, are also important factors that must be considered. The effect of atmospheric drag on a satellite orbit is illustrated in Fig. 10.11.

The ballistic drag coefficient $C_D A/M$ is the quotient of a drag quantity divided by the mass. The more "massive" the object and/or the smaller its drag, the smaller will be the value of the ballistic coefficient and, at the same time the less the object will be slowed as a result of its passage through the atmosphere.

The value of the drag coefficient C_D depends on the shape of the vehicle, its attitude with respect to the velocity vector, and whether it is spinning, tumbling, or is stabilized. Above 200-km altitude, the drag coefficient varies from about 2.2 for a sphere to about 3.0 for a cylinder, with other shapes being somewhere in between. Exact values of C_D are best determined by actual flight test. A value of 2.2 will yield a conservative result.

When the orbit altitude is less than 1000 km, atmospheric drag effect should be considered in long-term predictions. The equation for computing drag acceleration is

$$\boldsymbol{a}_D = -\frac{1}{2}\rho V^2 \frac{C_D A}{M} \boldsymbol{i}_v \tag{10.18}$$

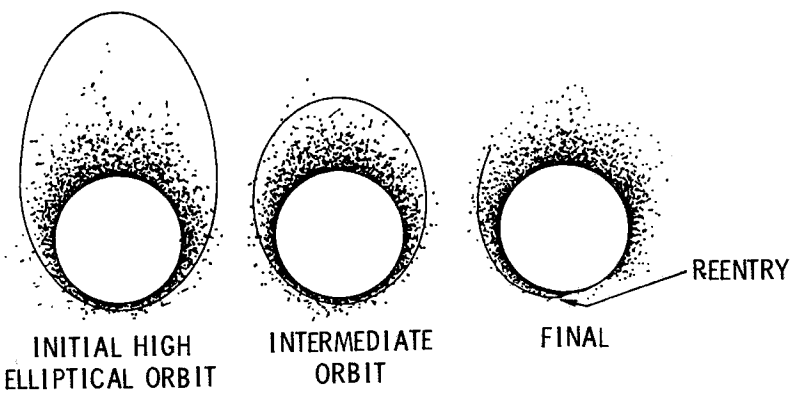


Fig. 10.11 Aerodynamic drag–induced orbit decay.

where

$\boldsymbol{a}_D$ = atmosphere drag acceleration vector
ρ = atmosphere density
V = velocity of the satellite
A = satellite effective (projected) area
M = satellite mass
$\boldsymbol{i}_v$ = unit vector of the satellite velocity

and C_D is the dimensionless drag coefficient of the spacecraft. The difficult part of the drag prediction is the modeling of the density of the atmosphere, which is a function of both altitude and time. The atmosphere density at a given altitude increases with solar flux, which has an 11-yr cycle. The air density also exhibits a day-to-night rhythm, reaching a maximum about 2 h after midday and a minimum between midnight and dawn. According to King-Hele,[13] at a height of 600 km, the maximum daytime density may be as large as eight times greater than the nighttime minimum. Another variation in the air density is the 27-day cycle caused by the extreme ultraviolet (UV) radiation.

Numerous mathematical models have been developed by many researchers since the late 1950s. The most commonly used models are the following:

1) The ARDC 1959 model, which is a static model and is based in part on density data inferred from early satellite observation.

2) The U.S. Standard Atmosphere of 1962, which was designed to represent an idealized, middle-latitude, year-round mean over the range of solar activity between sunspot minima and sunspot maxima. This model is also a static model, and the density values are consistently less than the ARDC 1959 model at various altitudes.

3) The dynamic Jacchia 1964 model, which accounts for the diurnal, 27-day, and 11-yr cycles. The Jacchia 1964 model has been widely used for accurate drag effect predictions. The values of density according to the Jacchia 1964 model for three representative periods of solar activity are shown in Fig. 10.12.

4) The Jacchia 1971 model, widely used for orbit decay predictions, is a dynamic model with improved accuracy.

5) The MSIS90E density model, which is considered the most accurate density model, contains a density profile extending all the way to the Earth's surface.

These three conditions correspond approximately to $F_{10.7}$ (an index used in specifying solar activity) values of 220, 150, and 70 for active, average, and quiet solar periods, respectively. In Fig. 10.12, the density values of the ARDC 1959 and U.S. Standard 1962 models are also shown for comparison.

The drag effect on satellite orbits takes place through energy dissipation. When an orbit loses energy, its semimajor axis decreases. For an orbit with large eccentricity, the drag effect would first circularize the orbit by gradually lowering its apogee and, then, the radius of a circular orbit would continue to decrease until the satellite crashes on the Earth's surface. For circular orbits, the orbit decay rate can be computed by the following equation:

$$\frac{\mathrm{d}a}{\mathrm{d}t} = -\sqrt{\mu a}\,\rho B \tag{10.19}$$

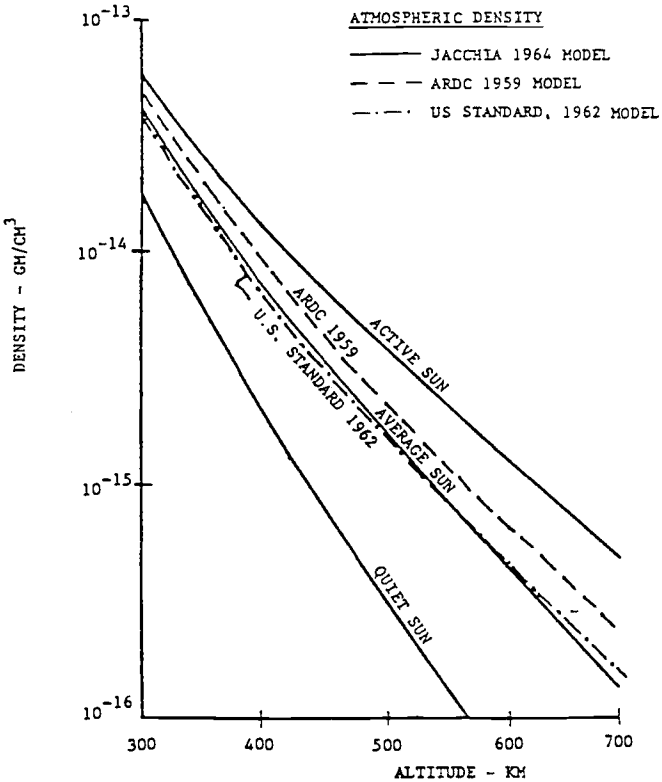


Fig. 10.12 Atmospheric density vs altitudes.

where μ is the Earth gravitational constant, a the semimajor axis, and B the ballistic coefficient $(C_D A/M)$.

Recently, the concept of using drag as a means to save propellant in orbit-transfer maneuvers has received some attention. The transfer from a high-altitude to a low-altitude orbit may be achieved by first lowering the perigee into the atmosphere and then letting the drag gradually reduce the apogee to the desired value. Finally, the perigee may be raised to a higher value if needed. If the spacecraft may be so oriented as to have a lifting surface, the atmospheric pressure can be utilized to make plane changes for highly eccentric orbits. The rotation of the atmosphere would also yield a small component normal to the orbits with high inclination. This component may change the inclination to 0.1 deg in 100 days. This concept is sometimes called the "aeroassisted orbit transfer."

In Fig. 10.13, the approximate lifetime of a circular orbiting satellite is shown in days as a function of its altitude. Several representative values of the ballistic coefficient are given. It should be noted that, for a first approximation, the lifetime is a linear function of the ballistic coefficient $C_D A/M$.

The curves in Fig. 10.14 represent orbits of differing eccentricities. The numbers alongside the curves give approximate period times, in minutes, at that point on the curve.

Perigee altitude is shown on the abscissa, and the ordinate is an odd quantity, the number of revolutions divided by the ballistic coefficient. To find the lifetime

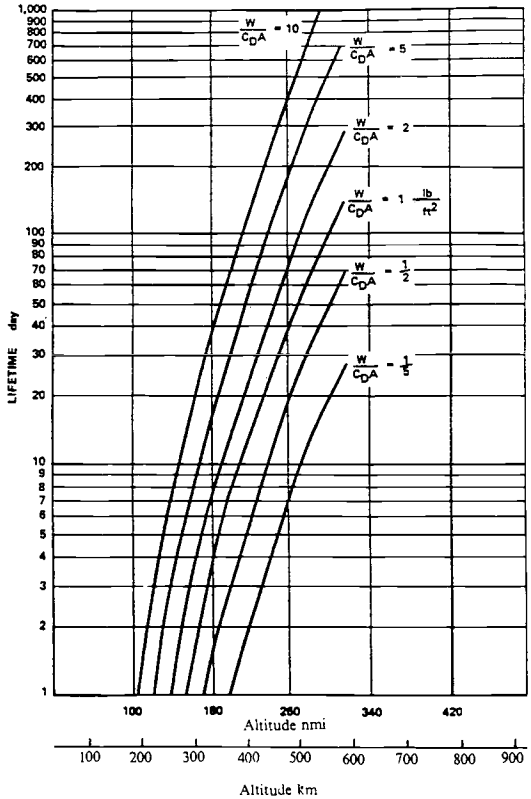


Fig. 10.13 Earth orbit lifetimes: circular orbits.

from this ordinate, the calculated value of the ballistic coefficient is multiplied by the orbital period in minutes. The results in Figs. 10.13 and 10.14 are very approximate as they are based on a static atmosphere.

It should be noted that, because of the atmospheric density changes that result from the 11-yr solar cycle, large variations in orbital lifetime can occur during the course of several years. As solar activity (solar flares and sunspots) increases the effective density, the height of the Earth's upper atmosphere is increased: as a result, lifetime may be less than that shown in Figs. 10.13 and 10.14. The lifetimes determined from the figures should be taken as mean values. Of course, if solar activity decreases, upper-atmosphere densities will be less, and lifetime longer, than shown.

10.9 Tidal Friction Effects

The tidal friction effects can be explained by the Earth-moon system, as illustrated in Fig. 10.15. The bulge is caused by the gravity pull of the moon, and the tidal friction exerts a net torque on Earth. The phase delay ψ occurs because Earth is spinning at a faster rate than the rotation rate of the moon around Earth. As a result of energy dissipation through friction and the conservation of angular

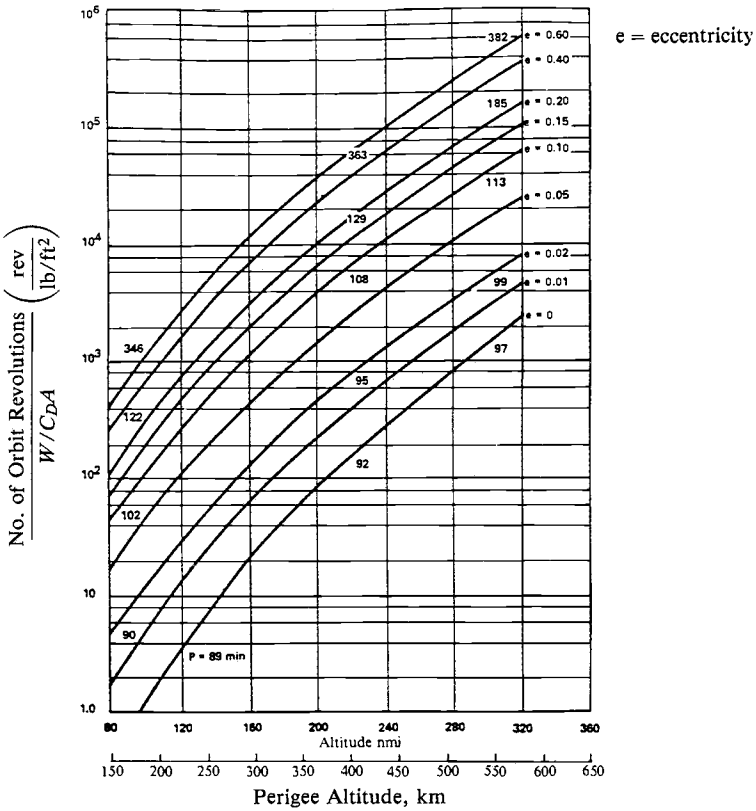


Fig. 10.14 Earth orbit lifetimes: elliptical orbits.

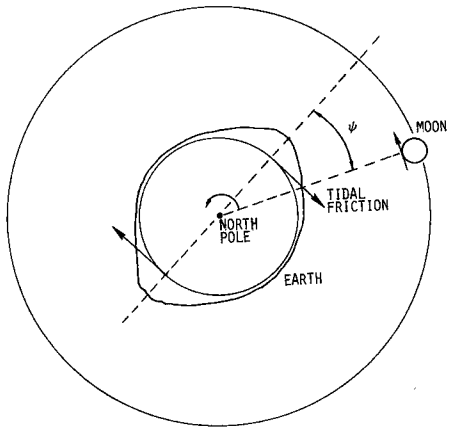


Fig. 10.15 The Earth-moon tidal friction mechanism.

momentum of the Earth-moon system, the Earth spinning rate is slowly decelerating, and the moon is slowly moving away from Earth. This mechanism is true for other outer planets with massive satellites.

The four massive satellites of Jupiter have the common features of most inner satellites of the solar system. These features can be summarized as follows:

1) Small inclination to planet's equatorial plane.
2) Small eccentricity.
3) Orbit-orbit resonance.

Extensive studies have been made in recent years to explain the earlier-listed phenomena. According to Goldreich,[14] dynamical considerations strongly favor low inclinations for close satellites of an oblate precessing planet. The small eccentricity and orbit-orbit resonance are believed to be consequences of tidal dissipation. For the evolution of small eccentricities, a mechanism was proposed by Urey et al.[15] in the form of tidal working in the satellite due to tides raised by the planet. This mechanism can be explained by a simple relation between eccentricity and energy of the satellite orbit.

$$e = \left(1 + \frac{2EL^2}{M_s^3 M_p^2} G\right)^{\frac{1}{2}} \tag{10.20}$$

where E is the energy of the orbit, L is the angular momentum, and M_p and M_s are the planet and satellite masses, respectively. G is the universal gravitational constant. If the satellite is not spinning, the tide raised on it can produce only a radial perturbation force. This means that L is not changed by the tide. Since any energy dissipation in the satellite decreases E and since we have $E < 0$, $0 < e < 1$, and L constant, we find that e is decreasing also. This process will not stop until $e = 0$ because the height of the tide will vary with the oscillation in distance between the satellite and planet. Later, Goldreich and Soter[16] examined such a mechanism, with reasonable values of the tidal energy dissipation function Q, and suggested that tides raised on satellites are of great significance in the evaluation of the eccentricities of these satellites.

A tidal origin of commensurable satellite mean motions was first proposed by Goldreich in 1964,[17] with the following evidence:

1) There are more examples of resonances than can be explained by random distribution of satellite orbital elements.
2) Tides raised on a planet by a satellite tend to vary according to the satellite's orbital period at a rate dependent on the satellite's mass and distance.
3) Mutual gravitational interaction is strong enough to maintain the commensurability of resonant satellites' periods, even against the upsetting influence of the tides.

Later, in 1972, Greenberg[18] gave a realistic model (Titan-Hyperion) involving mutual gravitation and tidal dissipation that provides a detailed explanation for satellite orbit-orbit resonance capture. The conclusion of his study is that once resonant commensurabilities are reached, the angular momentum transferred as a result of tidal friction on the planet should be distributed in such a way that commensurabilities among the four inner satellites are maintained.

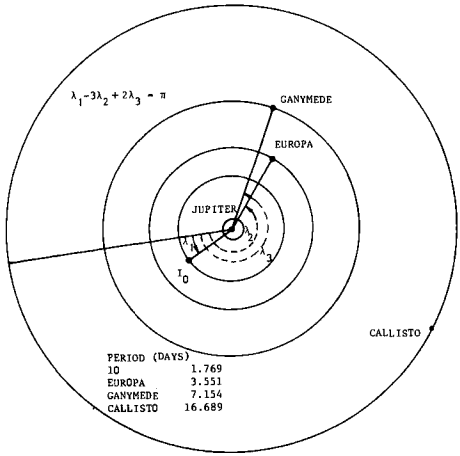


Fig. 10.16 The Galilean satellites of Jupiter.

Although tidal dissipation has been shown to be responsible for small eccentricities and stable commensurabilities, the magnitude of the tidal force exerted on the Galilean satellites is several orders of magnitude smaller[17] than other perturbing forces such as oblateness and mutual and solar attractions. This is true for all the natural and artificial satellites of the solar system.

It is interesting to point out that, within observational accuracy, the mean motions of the inner three Galilean satellites obey the relation $n_1 - 3n_2 + 2n_3 = 0$. The mean longitudes of the satellite satisfy the liberation equation $\lambda_1 - 3\lambda_2 + 2\lambda_3 = 180$, as illustrated in Fig. 10.16. The four massive satellites were discovered by Galileo in 1610 after he invented the telescope.

10.10 Long-Term Inclination Variations

The long-term orbit inclination variation has been one of the important elements in orbit selection and mission design. Inclination control maneuvers require significantly more propellant than the longitude or in-track orbit maintenance maneuvers. Therefore, a good understanding of the inclination perturbations and a proper selection of an initial orbit can often avoid costly ΔV expenditure and reduce the payload.

The long-term inclination perturbations are caused primarily by the sun/moon attractions. In Ref. 7, the averaged equations of variation for third-body perturbations are given in computer-assisted series expansion form, with orbit eccentricity in closed form. The equations for inclination and right ascension of ascending node variations can be reduced for near-circular orbits as

$$\begin{aligned} \frac{\mathrm{d}i}{\mathrm{d}t} &= -\frac{3}{4}\frac{y}{\sin i}\mathrm{IPO} \\ \frac{\mathrm{d}\Omega}{\mathrm{d}t} &= \frac{3}{4}y\left(\frac{\cos i}{\sin i}\mathrm{IPS} - \mathrm{IPC}\right) \end{aligned} \tag{10.21}$$

where

$y = n'^2/n(a'/r')^3 R_m$
n = mean motion of the satellite orbit
n' = mean motion of the perturbing body (sun or moon)
R_m = mass ratio, $R_m = 1$ for solar perturbation, $= 1/82.3$ for lunar perturbation
= mass of the perturbing body/(mass of the Earth + mass of the perturbing body)

and IPO, IPS, and IPC are computer-generated series listed here:

```
IPO = +(-1*SI*CI*SI3*CI3)*SIN(D)
      +(-1/2*SI**2*SI3**2)*SIN(2*D)
      +(+1/2*SI**2 + 1/2*SI**2*CI3 - 1/4*SI**2*SI3**2)*SIN(2*L - 2*D)
      +(-1/2*SI*CI*SI3 - 1/2*SI*CI*SI3*CI3)*SIN(2*L - D)
      +(-1/2*SI*CI*SI3 + 1/2*SI*CI*SI3*CI3)*SIN(2*L + D)
      +(-1/2*SI**2 + 1/2*SI**2*CI3 + 1/4*SI**2*SI3**2)*SIN(2*L + 2*D)
IPS = +(+1*SI*SI3**2)
      +(+1*CI*SI3*CI3)*COS(D)
      +(-1/2*CI*SI3 - 1/2*CI*SI3*CI3)*COS(2*L - D)
      +(-1*SI*SI3**2)*COS(2*L)
      +(+1/2*CI*SI3 - 1/2*CI*SI3*CI3)*COS(2*L + D)
IPC = +(+1*CI - 1/2*CI*SI3**2)
      +(+1*SI*SI3*CI3)*COS(D)
      +(-1/2*CI*SI3**2)*COS(2*D)
      +(-1/2*CI - 1/2*CI*CI3 + 1/4*CI*SI3**2)*COS(2*L - 2*D)
      +(-1/2*SI*SI3 - 1/2*SI*SI3*C13)*COS(2*L - D)
      +(+1/2*CI*SI3**2)*COS(2*L)
      +(+1/2*SI*SI3 - 1/2*SI*SI3*CI3)*COS(2*L + D)
      +(-1/2*CI + 1/2*CI*CI3 + 1/4*CI*SI3**2)*COS(2*L + 2*D)
```

where SI $= \sin i$, CI $= \cos i$, SI3 $= \sin i'$, CI3 $= \cos i'$, $L = U'$, $D = \Omega - \Omega'$
$(\ )'$ = elements of the third body (sun or moon).

The equation set (10.21) can be further reduced by eliminating those intermediate-period terms (terms with U' as argument) and combining the effect due to J_2 to

lead to the following set of useful equations in studying long-term inclination variations:

$$\frac{\mathrm{d}i}{\mathrm{d}t} = \sum_{j=1}^{2} \frac{3}{8} y_j \left[\cos i \sin 2i_j \sin(\Omega - \Omega_j) + \sin i \sin^2 i_j \sin 2(\Omega - \Omega_j) \right] \quad (10.22a)$$

$$\frac{\mathrm{d}\Omega}{\mathrm{d}t} = \sum_{j=1}^{2} \frac{3}{16} \frac{y_j}{\sin i} \left[\sin 2i \left(1 - 3\cos^2 i_j\right) + 2\cos 2i \sin 2i_j \cos(\Omega - \Omega_j) \right.$$
$$\left. + \sin 2i \sin^2 i_j \cos 2(\Omega - \Omega_j) \right] - \frac{3}{2} J_2 \left(\frac{a_e}{a}\right)^2 n \cos i \quad (10.22b)$$

where the summation $\sum$ is summed over the sun and the moon, and

$$y_j = \frac{n_j^2}{n} R_m, \qquad R_m = \begin{cases} 1 \text{ sun} \\ 1/82.3 \text{ moon} \end{cases} \quad (10.23)$$

n_j = mean motion of the sun or moon
i_j = inclination of the orbit of the sun or moon with respect to equatorial plane
J_2 = the second zonal harmonic of Earth
a_e = the equatorial radius of Earth
a = semimajor axis of the orbit
Ω_j = right ascension of the ascending node of the sum or moon with respect to the equatorial plane

The two equations (10.22) show the long-term variations of inclination and node, and the coupling effects between the inclination and right ascension of the ascending node through the sun/moon and J_2 perturbations. Notice that the inclination drift is zero when the value of $\Omega - \Omega_j$ is zero or 180 deg. When $\Omega - \Omega_j$ is less than 180 deg, the drift is positive, or the inclination is gradually increasing. When $\Omega - \Omega_j$ is greater than 180 deg, the inclination is gradually decreasing.

As an example, the upper and lower bounds of the long-term inclination variations of the GPS/NAVSTAR orbits were understood with Eqs. (10.22). Figure 10.17 shows the 10-yr history of inclination variations of the GPS orbits with different initial nodes (Ω_0 = 25, 265, and 145 deg). The solid symbols (squares, triangles, and dots) are the predictions from the reduced equation for inclination (10.22). With this clear understanding, the maximum inclination deviation from a nominal 55 deg was determined to be 1 deg for the mission orbit design specifications.

Another example is the determination of initial inclination and ascending node of a geosynchronous orbit. Figure 10.18 shows the inclination history of two geosynchronous orbits. One curve starts from a zero initial inclination and linearly increases to 9.5 deg after 10 yr. This type of variation is not desired by the mission designers because, after 5 yr, the inclination can become as large as 5 deg. The second curve in Fig. 10.19 starts from an initial inclination of 3.5 deg with an initial node (Ω) of about 270 deg. Based on Eqs. (22), the inclination would

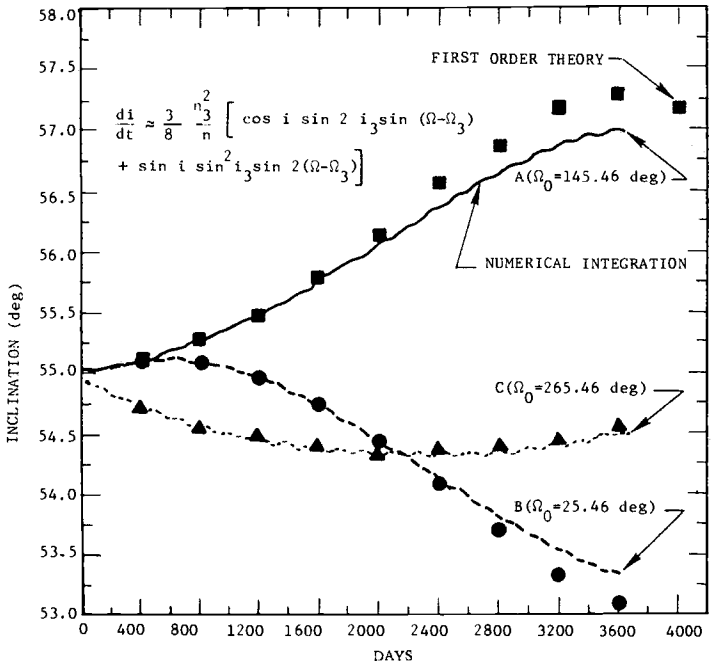


Fig. 10.17 Inclination variation of the phase III GPS satellites.

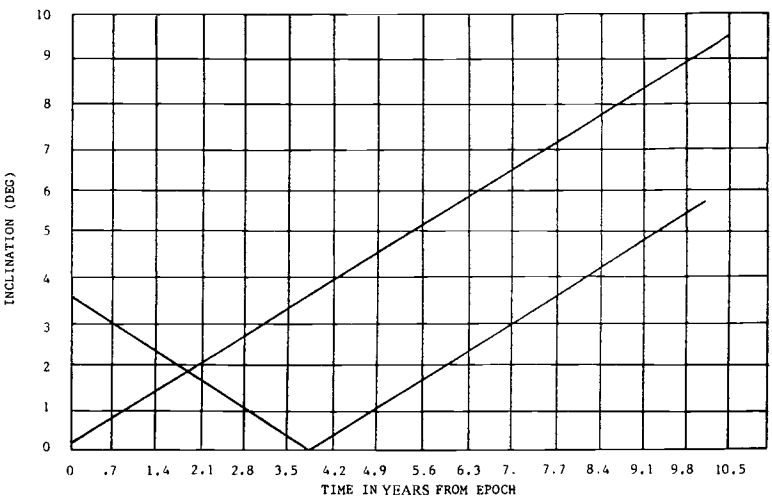


Fig. 10.18 Inclination time history of geosynchronous orbits.

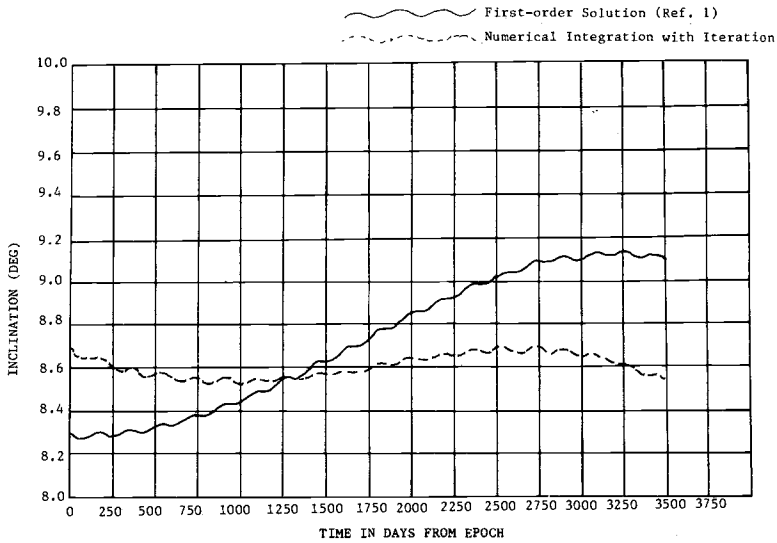


Fig. 10.19 Inclination time history with epoch = 1 January 1987 for a geosynchronous orbit.

follow greatest negative slope to drift. After 3.5 yr, the inclination reaches zero and, in the meantime, the node crosses the zero value and changes the sign of the inclination slope, as shown in Fig. 10.19. As a result, the geosynchronous orbit can stay below 3.5-deg inclination for 7.7 yr, which is usually the length of the mission life. During these 7.7 yr, no inclination controls are required.

Finally, a very interesting example is to achieve a stable inclination value throughout entire mission life by searching for a unique set of initial inclination and node based on Eqs. (10.22). From Eq. (10.22b), an initial inclination may be found such that sun/moon effect and the J_2 effect cancel, thus making the node rate zero. Then, from the inclination equation of Eqs. (10.22), the inclination rate is zero when the node $(\Omega - \Omega_j)$ is zero. Therefore, this coupling effect indicates that, with a proper initial inclination and with the initial right ascension of ascending node equal to zero, the inclination may stay constant for a long time. The solid curve of Fig. 10.19 represents the inclination variation of a geosynchronous orbit with an 8.3-deg initial inclination determined from Eqs. (10.22). The curve starts to deviate from the initial value after 500 days because of the higher-order effects not included in Eqs. (10.22). With some iterations by numerical integration, a new value of 8.7 deg was found, and the deviations from a mean value of 8.6 deg were smaller than 0.1 deg for the entire 10 yr, starting 1 January 1987. The initial value for stable inclination variation depends on the epoch because of the different locations of the sun and moon.

References

[1]Roy, A. E., *Orbital Motion,* 2nd ed., Adam Hilger, Bristol, UK, 1982.

[2]Kozai, Y., "On the Effects of the Sun and the Moon Upon the Motion of a Close Earth Satellite," Smithsonian Institution Astrophysical Observatory, Cambridge, MA, Special Rept. 22, March 1959.

[3]Kaula, W. M., "A Development of the Lunar and Solar Disturbing Function," *Astronomical Journal,* Vol. 67, 1962, p. 300.

[4]Cook, G. E., "Luni-Solar Perturbations of the Orbit of an Earth Satellite," *Geophysics Journal, Royal Astronomical Society,* Vol. 6, 1962, pp. 271–291.

[5]Kaufman, B., "Variation of Parameters and the Long-Term Behavior of Planetary Orbits," AIAA Paper 70-1055, Aug. 1970.

[6]Ash, M. E., "Doubly Averaged Effect of the Moon and Sun on a High Altitude Earth Satellite Orbit," Lincoln Laboratory, Tech. Note ESD-TR-74-45, 1 March 1974.

[7]Chao, C. C., "An Analytical Integration of the Averaged Equations of Variations Due to Sun-Moon Perturbations and Its Application," The Aerospace Corp., El Segundo, CA, Rept. SD-TR-80-12, Oct. 1979.

[8]Michielsen, H. J., and Webb, E. D., "Stationkeeping of Stationary Satellites Made Simple," *Proceedings of the First Western Space Conference,* 1970.

[9]Kamel, A., Kkman, D., and Tibbitts, R., "East-West Stationkeeping Requirements of Nearly Synchronous Satellites Due to Earth's Triaxiality and Luni-Solar Effects," *Celestial Mechanics,* Vol. 8, 1973.

[10]Allan, R. R., and Cook, G. E., "The Long-Period Motion of the Plane of a Distant Circular Orbit," *Proceedings of the Royal Society (London),* Vol. 280, No. 1380, 7 July 1983.

[11]Chao, C. C., and Baker, J. M., "On the Propagation and Control of Geosynchronous Orbits," *Journal of the Astronautical Sciences,* Vol. XXXI, No. 1, Jan.-March 1983, pp. 99–115.

[12]Chao, C. C., "Semiautonomous Stationkeeping of Geosynchronous Satellites," *Journal of Guidance, Control, and Dynamics,* Vol. 7, Jan.-Feb. 1984.

[13]King-Hele, D., *Theory of Satellite Orbits in an Atmosphere,* Butterworths, London, 1964.

[14]Goldreich, P., "Inclination of Satellite Orbits of an Oblate Precessing Planet," *Astronomical Journal,* Vol. 70, 1965.

[15]Urey, H. C., Elsasser, W. M., and Rochester, M. G., *M.G.A.P.J.,* Vol. 129, 1959, p. 842.

[16]Goldreich, P., and Soter, S., "θ in the Solar System," *Icarus,* Vol. 5, 1966, pp. 375–389.

[17]Goldreich, P., "An Explanation of the Frequency Occurrence of Commensurable Mean Motions in the Solar System," *Royal Astronomical Society,* Monthly Notice 130, 1965.

[18]Greenberg, R. J., "Evolution of Satellite Resonances by Tidal Dissipation," *Astronomical Journal,* Vol. 70, No. 4, 1972.

[19]Frautnick, J. C., and Cutting, E., "Flight Path Design Issues for the TOPEX Mission," AIAA Paper 83-0197, Jan. 1983.

Problems

10.1. Calculate nodal regression $\dot{\Omega}$ and apsidal rotation rate $\dot{\omega}$ for a satellite in a 400-km circular orbit at $i = 60$ deg inclination to the equator. What if $i = 90$ deg or $i = 63.4$ deg?

10.2. Determine the semimajor axis and inclination of an orbit of an Earth satellite to satisfy the following constraints:

1) Sun-synchronous.
2) Argument of perigee is constant.
3) Eccentricity = 0.25 If we decrease the semimajor axis by 1000 km, which

one (or more than one) of the three constraints just listed will be violated? What will be the new rates of node and argument of perigee?

Note: Consider only the perturbation due to J_2, $J_2 = 0.0010826$, $\mu = 0.39860047 \times 10^{15}$ m^3/s^2, R_e (Earth equatorial radius) = 6378140 m.

10.3. Determine the semimajor axis of an Earth satellite orbit with eccentricity equal to 0.17 and $d\omega/dt = 0$ (critical inclination, $i = 63.4$ deg or 180 deg − 63.4 deg), and the orbit is sun-synchronous.

10.4. a) Write down the equations of motion in relative form with perturbations $\boldsymbol{a}_p$. If the moon is the perturbing body, write down $\boldsymbol{a}_p = ?$ (see Fig. 12.8).

b) Briefly explain the methods of solution to these equations of motion. If high-accuracy orbit prediction or orbit determination is required, what method should be used and why? If long-term orbit integration is required and the emphasis is on computation speed rather than accuracy, what method should be used?

c) For low-altitude orbits (185–740 km), what perturbations are important? For high-altitude orbits (GPS and geosynchronous altitudes), what perturbations are important and why?

10.5. The potential function for the primary body with J_2 effect is

$$\phi = \frac{u}{r}\left[1 + \frac{1}{2}J_2\left(\frac{a_e}{r}\right)^2(1 - 3w^2)\right]$$

where

$\mu = k^2 m$ gravitational constant
a_e = Earth equatorial radius
r = geocentric distance to satellite $r = \sqrt{x^2 + y^2 + z^2}$
$w = z/r$

Derive the equations of motion from the preceding potential in the Cartesian coordinates. (Hint: $d^2x/dt^2 = \partial\phi/\partial x$, $x \to y, z$.)

10.6. Compute and plot the time history of the right ascension of ascending node Ω of an Earth orbit with the following initial conditions:

$a_0 = 3700$ n.mi.
$e_0 = 0$
$i_0 = 30$ deg
$\Omega_0 = 45$ deg
$\omega_0 = 50$ deg
$M_0 = 100$ deg
($\mu = 62750.1633$ n.mi.^{3}/s^2, Earth equatorial radius = 3444 n.mi.)

Include only the J_2 ($J_2 = 0.0010826$) effect. Plot the Ω variations as a function of time (min) up to 300 min. After you have plotted them, you will notice that the

variations are a combination of secular and short-period variations. Which one of the two is more important?
(Hint: Find the equation $d\Omega/dt = -n(3/2)J_2(a_e/p)^2 \cos i\{1 + 3e\cos M + \cdots\}$ in the class notes, and analytically integrate the equation by ignoring all the terms having e as coefficient. Assume that n, a, i, and ω are constant over the 300-min interval.)

10.7. Compute maximum orbital eccentricity for a 1000-kg synchronous equatorial satellite with a projected area $A = 10\ \text{m}^2$ resulting from solar-radiation pressure. Assume total reflectivity for the surface ($\beta = 1$). What if the area is $10^4\ \text{m}^2$ (balloon)? Assume $e_0 = 0.000007245$, $\omega_0 = 30$ deg, $\Omega_0 = 60$ deg, $\lambda_0 = 270$ deg (sun's ecliptic longitude at epoch), and $d\Omega/dt = 0$. (Hint: Compute g/z and then ρ, and find the maximum eccentricity using Eq. 10.13 for possible values of θ. Read Ref. 11 for more details.)

10.8. What is the lifetime of a 1000-lb (454.5-kg) spherical satellite of 100-ft^2 area in a low-altitude 160 × 360 km (86.5 × 194.6 n.mi.) Earth orbit?

Selected Solutions

10.1. $\dot{\Omega} = -4.0, 0, -3.6$ deg/day
$\dot{\omega} = 1.0, -4.0, 0$ deg/day

10.2. $i = 116.6$ deg
$a = 10{,}187.6$ km

10.3. $a = 9981.25$ km

10.7. $e_{\max} = 0.000438, 0.438$

10.8. 1.42 days

11
Orbital Systems

11.1 Launch Window Considerations

Launch Azimuth

The launch of a satellite into an orbit with inclination i requires launching in an azimuth direction Az defined by the formula

$$\sin Az = \frac{\cos i}{\cos \phi_0} \tag{11.1}$$

where ϕ_0 is the launch site latitude. For a due east launch ($Az = 90$ deg), $i = \phi_0$, or the orbital inclination is equal to the launch site latitude. Differentiating Eq. (11.1) and solving for $\mathrm{d}i$, one obtains

$$(\mathrm{d}i) = \frac{-\cos \phi_0 \cos Az \, \mathrm{d}Az}{\sin i} \tag{11.2}$$

which shows the sensitivity of inclination error $\mathrm{d}i$ due to an error in azimuth $\mathrm{d}Az$. For range safety reasons, the azimuth may be constrained between specified limits, as is shown in Fig. 11.1, for example.

In-Plane or Out-of-Plane Ascent Using a Launch Azimuth Constrained to the Nominal Value

The most efficient launch is an in-plane launch ascent characterized by a wait on the ground until the launch site lies in the mission orbit plane, which is fixed in inertial space. No plane change maneuver is then required unless the orbit inclination is less than the launch site latitude or the launch azimuth constraint is such that a "dog-leg" maneuver is needed. A direct ascent to a parking or mission orbit may be made with or without a *phasing orbit* to place the satellite in a specified geographic location. The phasing problem where the parking and mission orbits are not coplanar may involve excessive waiting times, which can be somewhat reduced by "lofted" or other trajectories.

The out-of-plane launch is a form of direct ascent when the phasing problem can be solved by a wait on the ground. The launch window geometry is shown in Fig. 11.2.

The launch window can be defined as the time,

$$\text{L.W.} = 2\Delta\Omega/\omega_e \tag{11.3}$$

where $\Delta\Omega = \omega_e t$ is the maximum permissible nodal increment consistent with the plane change angle ε required to transfer from the launch orbit plane to the mission orbit plane. The latter can be determined from the equation

$$\varepsilon = \cos^{-1}(\cos^2 i + \sin^2 i \cos \Delta\Omega) \tag{11.4}$$

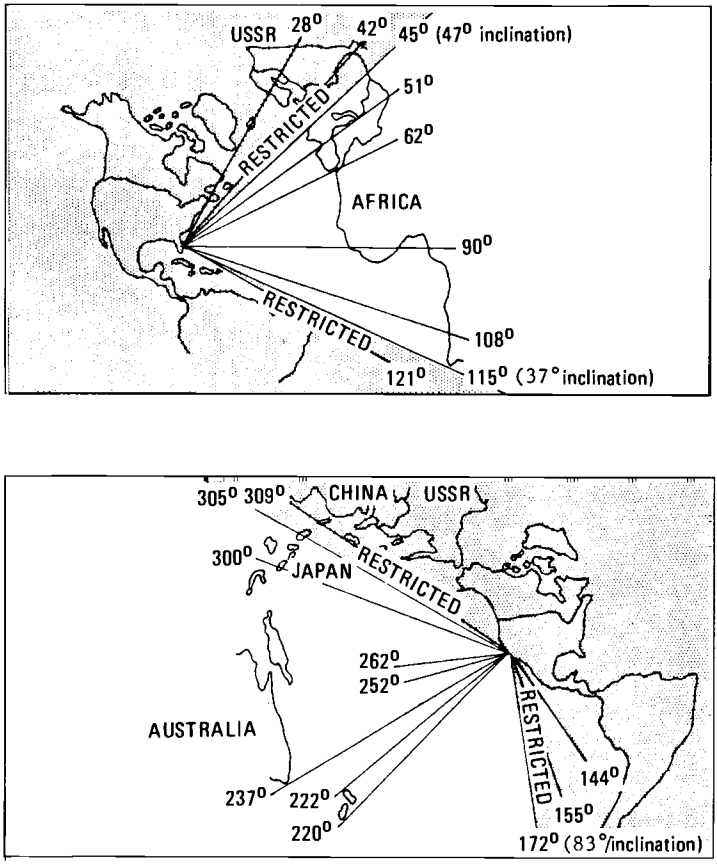


Fig. 11.1 Geographic launch constraints.

where i is the inclination of the orbit and the corresponding launch azimuth is constrained to the nominal value provided by Eq. (11.1). The longitudinal separation $\Delta\Omega$ is the same at all latitudes.

In general, a graph of ε vs time is as shown in Fig. 11.3, where the plane change capability ε of the launch vehicle defines the launch window.

Launch Azimuth Constrained to Off-Nominal Value

If the launch azimuth is constrained to an off-nominal value, then Fig. 11.4 depicts the launch geometry at the time $t = 0$ the launch site lies in the mission orbit plane. The launch geometry at some later time, $t > 0$, is shown in Fig. 11.5. Applying spherical trigonometry to the triangle bounded by the launch orbit, mission orbit, and equator results in the equation

$$\cos\varepsilon = \cos i \cos i' + \sin i \sin i' \cos(\omega_e t + \Delta\Omega_0) \tag{11.5}$$

where $\Delta\Omega_0$ is the difference in right ascension of the ascending node of the launch and mission orbits at $t = 0$.

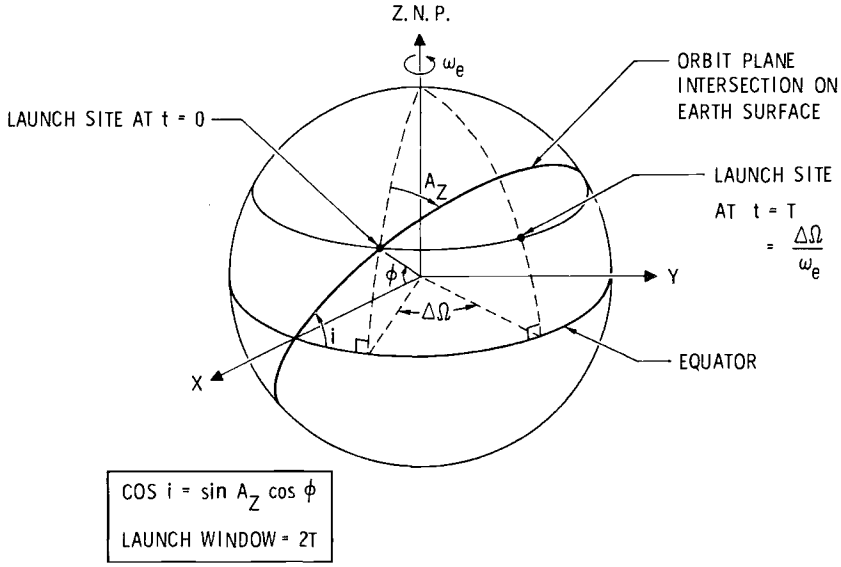


Fig. 11.2 Launch window geometry.

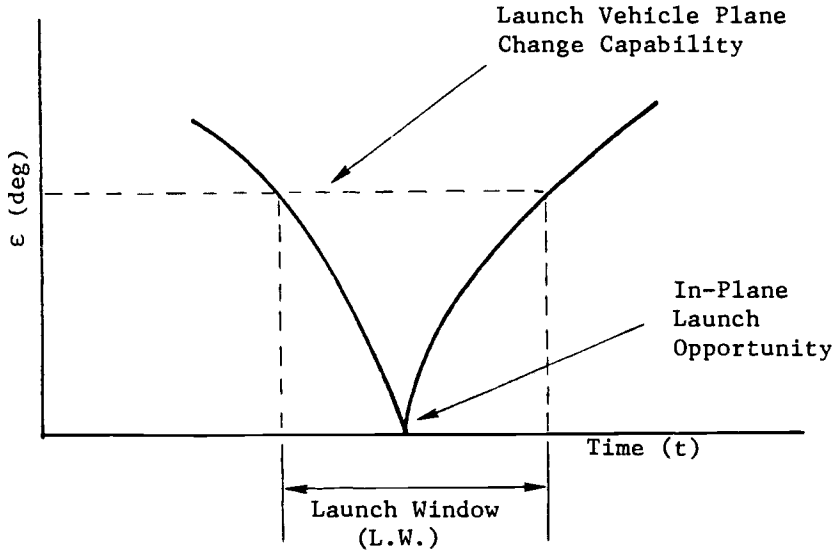


Fig. 11.3 Launch window.

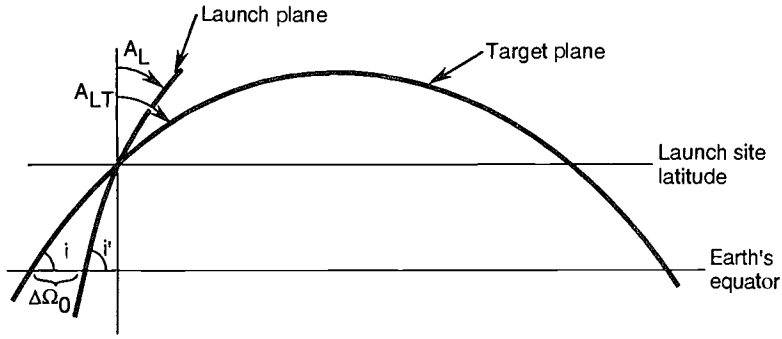


Fig. 11.4 Geometry for fixed launch azimuth at $t = 0$.

Milstead[8] showed that the condition for minimum plane change angle could be found by simply differentiating Eq. (11.5). The resulting condition, $(\omega_e t + \Delta\Omega_0) = 0$, implies that $\varepsilon_{\min} = |i' - i|$ which, in turn, requires that the launch and mission orbit planes intersect on the equator.

Figure 11.6 from Ref. 8 describes plane change angle vs time for several fixed values of launch azimuth and for a launch latitude of 28.34 deg and a mission orbit inclination of 30 deg. For these conditions, a launch azimuth of 79.722 deg satisfies Eq. (11.1), and the satellite can be launched directly into the mission orbit at $t = 0$, with no plane change required. Another opportunity to launch directly into the mission orbit with no plane change required occurs about 168 min later when the launch site rotates into the mission orbit plane again.

For a launch azimuth of 90 deg, the inclination of the launch orbit equals 28.34 deg. The minimum plane change angle $\varepsilon_{\min}$ equals $|28.34 \text{ deg} - 30 \text{ deg}|$, which equals 1.66 deg, as shown on Fig. 11.6.

Launch Azimuth Continuously Optimized to Minimize Plane Change

If the capability exists to vary the launch azimuth with time, then an optimum azimuth solution to minimize the plane change angle can be found. Reference 8 derives this solution, again by differentiation. The intuitively appealing solution is that the minimum plane change angle is achieved by launching the vehicle in such a manner that it will intercept the target plane 90 deg downrange from the

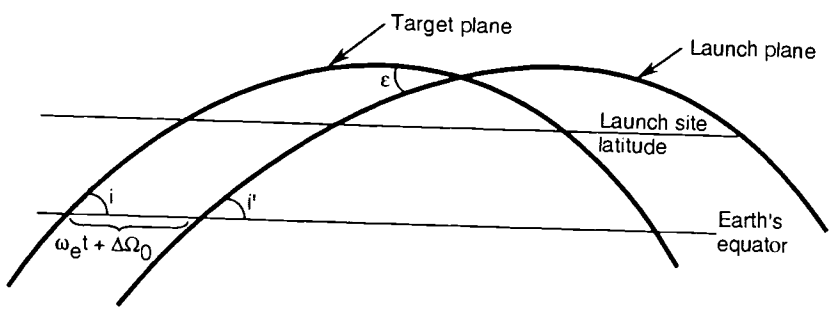


Fig. 11.5 Geometry for fixed launch azimuth at $t > 0$.

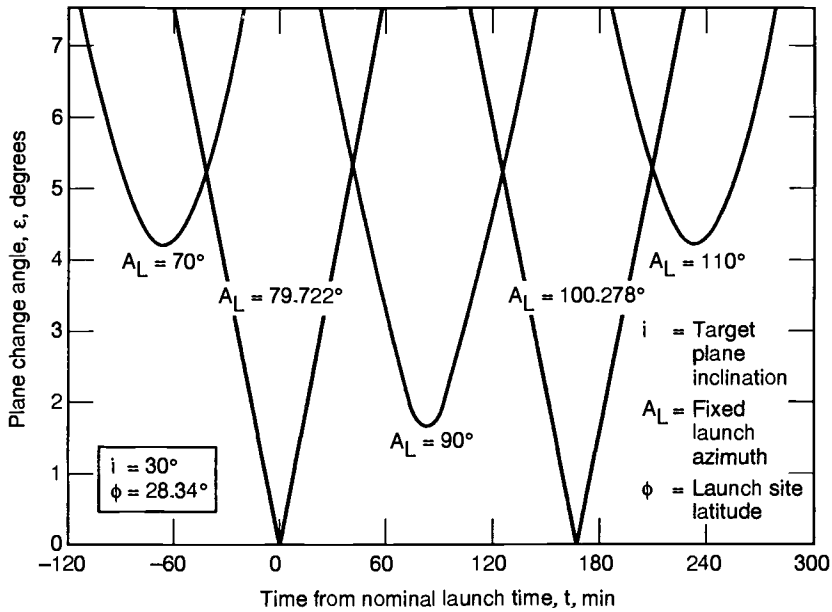


Fig. 11.6 Launch windows for fixed launch azimuth.

position of the launch site at the time of launch. This launch geometry is illustrated in Fig. 11.7. Reference 8 derives the following equation for the minimum plane change angle as a function of time:

$$\sin \varepsilon_{\min} = \sin i \cos \phi \sin(\omega_e t + \Delta L) - \cos i \sin \phi \tag{11.6}$$

where i is the inclination of the target orbit, ϕ is the launch site latitude, and

$$\sin \Delta L = \frac{\tan \phi}{\tan i} \tag{11.7}$$

Figure 11.8 from Ref. 8 illustrates $\varepsilon_{\min}$ as a function of time for several values of target orbit inclination and for a launch site latitude of 28.34 deg. Note the second opportunities to launch directly into the mission orbit with no plane change required when the launch site again rotates into the mission orbit plane.

The optimum launch azimuth to achieve $\varepsilon_{\min}$ is given by[8]

$$\tan A_{\text{opt}} = \frac{\cos i \cos \phi + \sin i \sin(\omega_e t + \Delta L) \sin \phi}{\sin i \cos(\omega_e t + \Delta L)} \tag{11.8}$$

where $0 \leq A_{\text{opt}} \leq 180$ deg, and is illustrated in Fig. 11.9.

Sun-Orbit Orientation

Frequently, thermal or power considerations of spacecraft require that the angle between the sun and the orbit plane be maintained within specified bounds for the

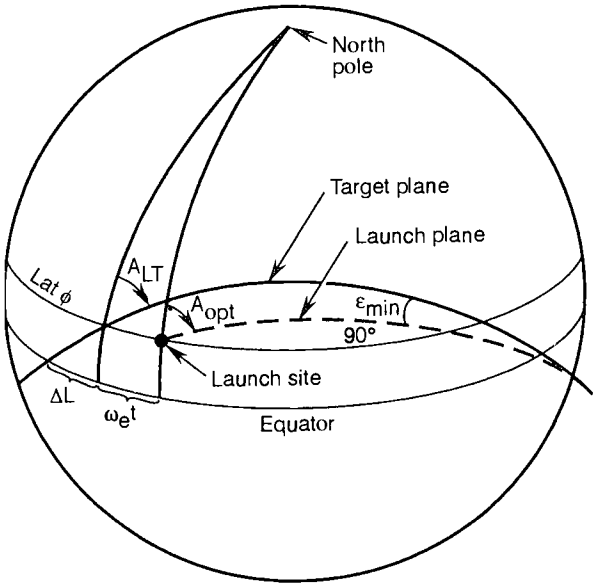


Fig. 11.7 Launch geometry.

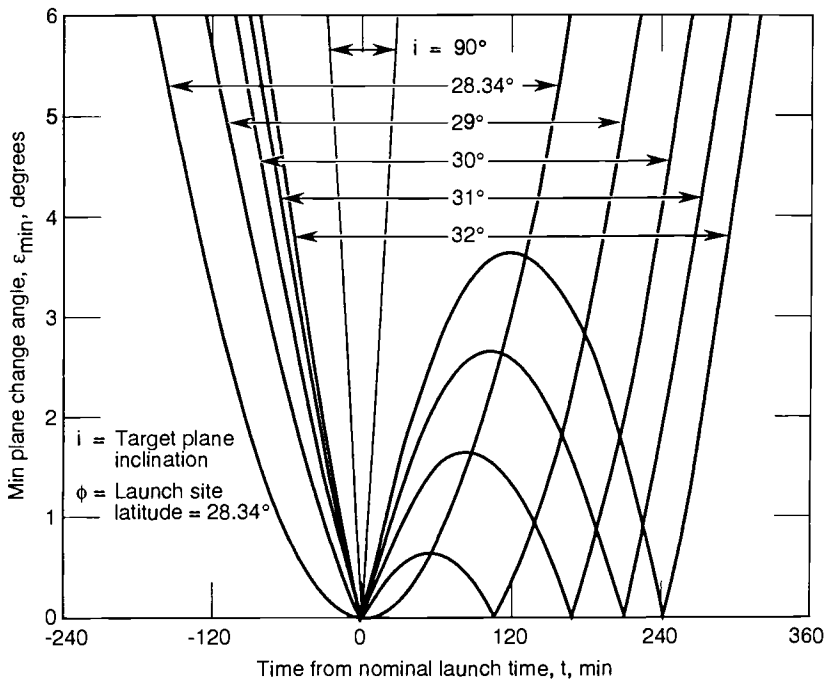


Fig. 11.8 Launch windows for variable optimum launch azimuths.

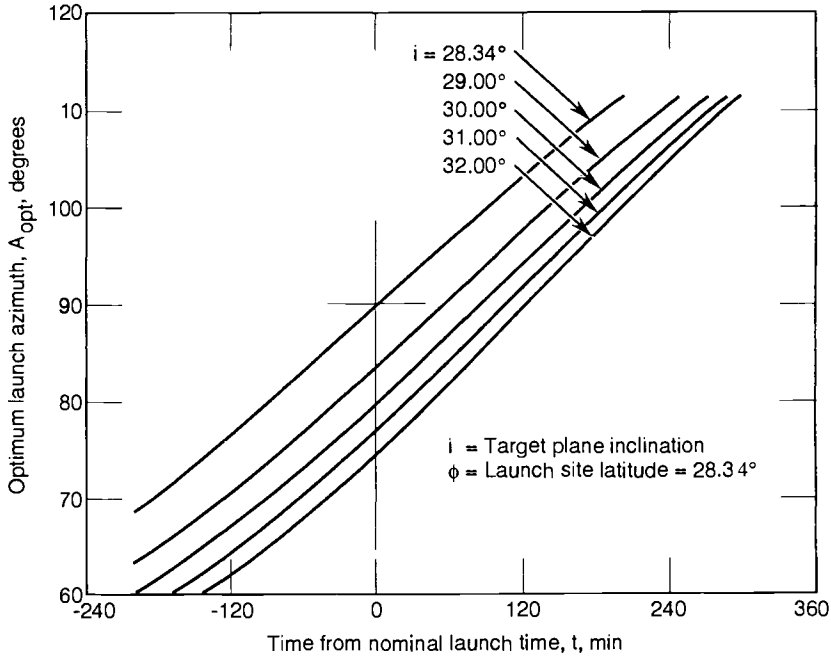


Fig. 11.9 Optimum launch azimuth.

duration of the mission. This angle, conventionally known as the beta angle β, is illustrated in Fig. 11.10. Since β is the complement of the angle between the sun vector $\hat{s}$ and the positive normal to the orbit $\hat{n}$, it follows from their scalar product that

$$\beta = \sin^{-1}(\hat{s} \cdot \hat{n}) \tag{11.9}$$

and ultimately that[1]

$$\beta = \sin^{-1}[\cos\delta_S \sin i \sin(\Omega - \text{RATS}) + \sin\delta_S \cos i] \tag{11.10}$$

where β is defined to lie in the range from -90 to $+90$ deg.

Equation (11.10) reveals that beta angle depends on solar declination δ_S, orbit inclination i, and the difference in right ascensions of the true sun and the ascending node (Ω – RATS). The first of these quantities, δ_S, depends on the date during the mission. The second quantity, i, is essentially constant during the mission. The last quantity (Ω – RATS) changes because of nodal regression (induced by Earth's oblateness perturbations, as described in Chapter 8) and seasonal variation in the right ascension of the true sun.

In light of the variability of the terms on the right side of Eq. (11.10), it is clear that the beta angle cannot be held constant throughout a mission. However, it is generally possible to select conditions at the start of the mission so that the beta angle will stay within some prescribed tolerable range of values for that portion of the mission during which β is essential to performance.

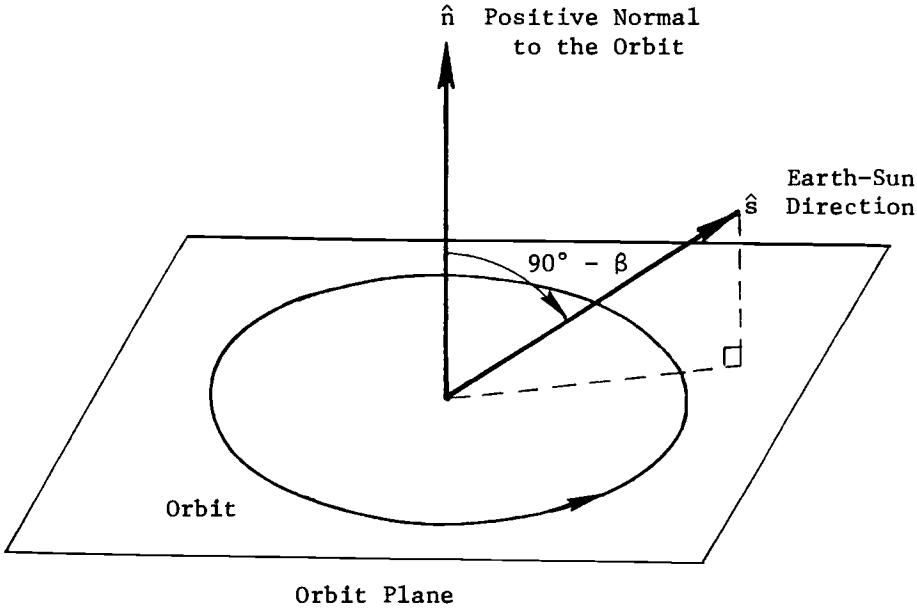


Fig. 11.10 Definition of sun-orbit angle, beta.

Earth Eclipsing of a Circular Orbit

It may be important to determine those occasions during its mission when a satellite is eclipsed by the Earth. Such eclipsing occurs whenever the satellite passes through the Earth's shadow, which is assumed cylindrical in this discussion.* Figure 11.11 shows that the Earth's shadow intersects the orbital sphere of a satellite at altitude h in a minor circle whose Earth-central-angular radius is β^*, where[1]

$$\beta^* = \sin^{-1}[R/(R+h)], \qquad 0 \text{ deg} \leq \beta^* \leq 90 \text{ deg} \tag{11.11}$$

View A–A in Fig. 11.11 reveals that the orbit intersects the perimeter of the shadow circle at points E_1 and E_2. Note that the length of the eclipsed orbital arc E_1E_2 is just twice arc CE_1, where C is the point on the orbit of closest approach to the shadow axis A. That is, the length of arc AC is just the magnitude of β. Hence, it follows from the right spherical triangle ACE_1 that

$$\Delta u = \cos^{-1}(\cos\beta^*/\cos\beta) \tag{11.12}$$

When Eqs. (11.11) and (11.12) are combined, the eclipsed fraction of the circular orbit is found to be[1]

$$f_E = \frac{2\Delta u}{2\pi} = \frac{1}{\pi}\cos^{-1}\left[\frac{\sqrt{h^2+2Rh}}{(R+h)\cos\beta}\right] \tag{11.13}$$

*Such an assumption is valid at low satellite altitudes, where there is no appreciable difference between the umbral and penumbral regions of total and partial eclipsing, respectively.

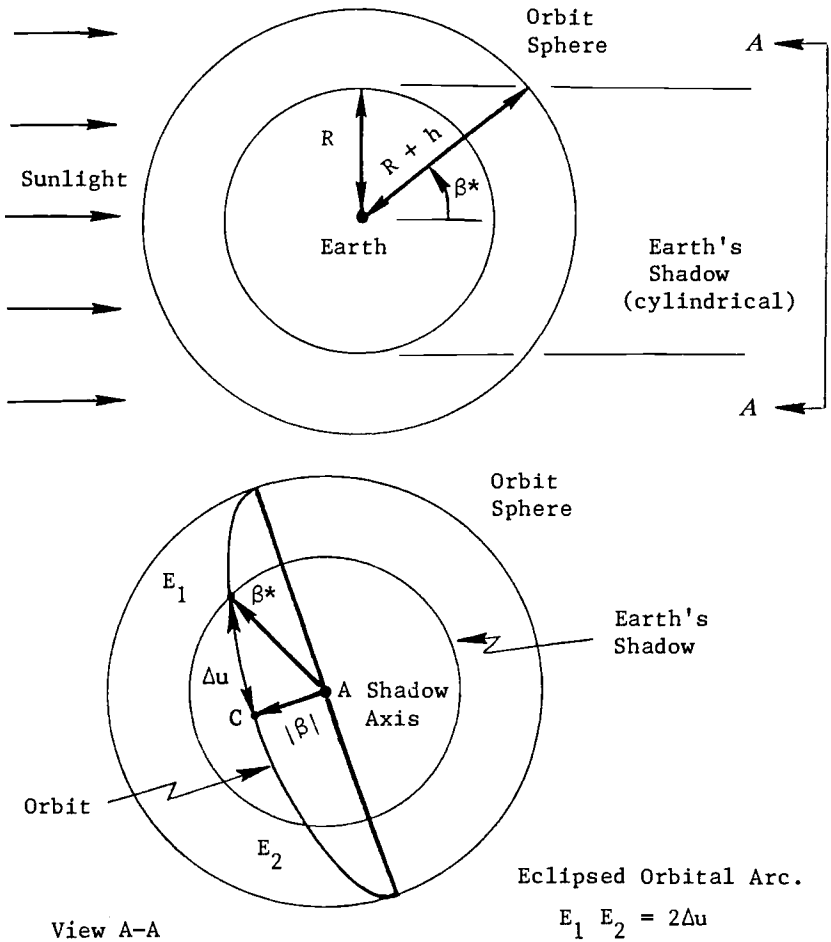


Fig. 11.11 Eclipsing of a circular orbit (from Ref. 1).

Figure 11.12 shows the variation of eclipse fraction with β-angle magnitude for several low-altitude circular orbits. Note that the eclipse fraction diminishes with an increase in either altitude or magnitude of β. One should also observe that eclipsing can occur only if $|\beta|$ is less than the critical angle β^* for the given orbit altitude. For example, a satellite in a 200-n.mi.- altitude circular orbit will be eclipsed through some portion of its orbit if and only if $|\beta| < 70.93$ deg.

For a circular orbit, the time duration of the eclipse interval is directly proportional to its angular extent; in other words, the duration of the eclipse on a given satellite revolution is just the product of the eclipse fraction f_E and the orbital period. However, this is not true for eccentric orbits. The duration and angular extent of eclipsing of a satellite in an eccentric orbit depends on the eccentricity and the relative orientation of sun and perigee, as well as the sun-orbit angle β.

Figure 11.13 presents the Earth, its shadow, and the circular synchronous equatorial orbit at the time of the vernal equinox. The circular synchronous orbit radius

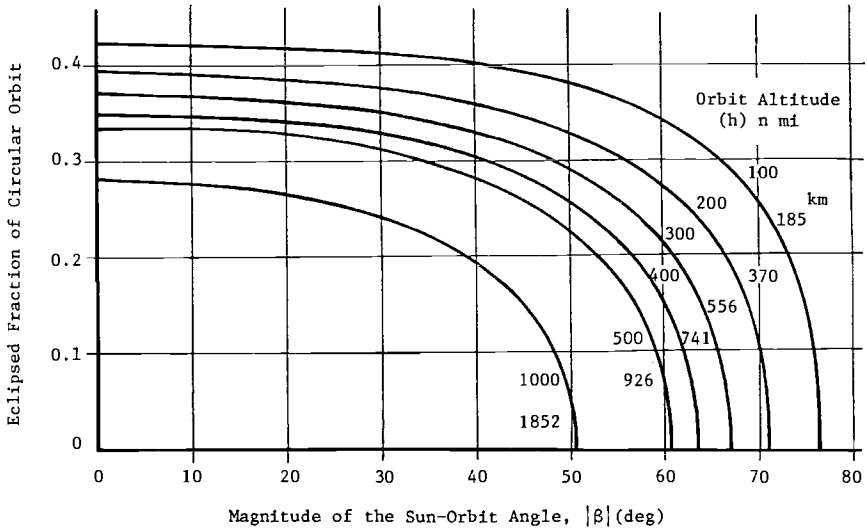


Fig. 11.12 Eclipse fraction vs beta-angle magnitude and circular orbit altitude (from Ref. 1).

of 6.61 Earth radii produces an orbit period of 24 sidereal hours. A satellite in this orbit revolves with an angular rate that exactly equals the angular rate of rotation of the Earth on its axis.

An Earth-centered inertial (ECI) coordinate system whose origin is at the center of the Earth is shown in this figure. The X axis points to the vernal equinox, the Z axis points to the celestial North Pole, and the Y axis forms a right-hand triad with X and Z.

In Fig. 11.13a, the sun is along the X axis, and the shadow axis intersects the synchronous equatorial orbit. The Y axis points 23.45 deg below the ecliptic plane.

This angular difference of 23.45 deg is important at the summer solstice (Fig. 11.13b) because the Earth casts a shadow that passes beneath the circular, synchronous equatorial orbit. And so no eclipsing occurs at this time.

Figure 11.14 shows the geosynchronous eclipse geometry. For geometrical purposes, consider the relative motion of the Earth's shadow as it moves along the ecliptic plane. At the distance of the synchronous orbit, the radius of the disk of the umbra is 8.44 deg. Therefore, 21.6 deg or 22 days before the autumnal equinox, the umbral disk becomes tangent to the equatorial plane. This is the beginning of the eclipse season because this is the first time a satellite in the synchronous equatorial orbit could experience an umbral eclipse, although it would be vanishingly short. At the time of the autumnal equinox, the center of the umbral disk would lie in the equatorial plane. This geometry produces a maximum time of 67.3 min.

Twenty-two days later, the eclipse season ends when the umbral disk is again tangent to the equatorial plane. And so, the eclipse season is 44 days, centered on the autumnal equinox. A similar eclipse season is centered on the vernal equinox.

Sun Synchronism of an Orbit

In Chapter 8, one of the principal perturbations of an orbit caused by Earth's oblateness was identified as nodal regression $\dot{\Omega}$. Recall that Eq. (8.3) showed that

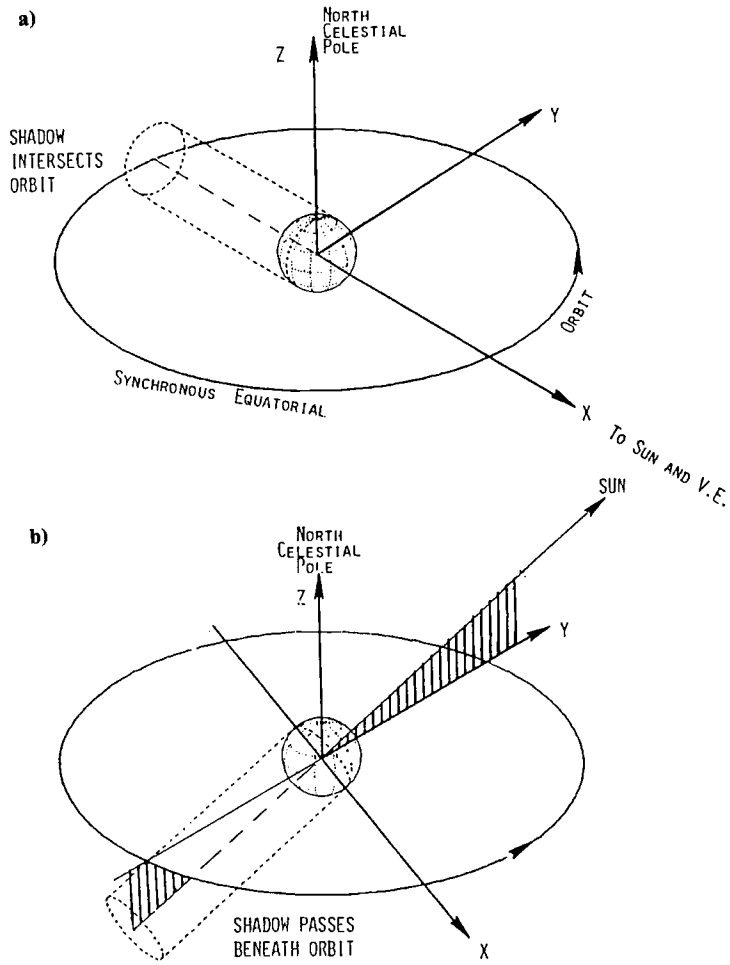


Fig. 11.13 Geosynchronous orbit geometry: a) vernal equinox shadow; b) summer solstice shadow.

$\dot{\Omega}$ depends as follows on orbit inclination i, average altitude h, and eccentricity e:

$$\dot{\Omega} = \frac{-9.9639}{(1-e^2)^2}\left(\frac{R}{R+h}\right)^{3.5} \cos i \frac{\text{deg}}{\text{mean solar day}} \tag{11.14}$$

where

$$h = (h_a + h_p)/2$$
$$e = (h_a - h_p)/(2R + h_a + h_p)$$

and R = the equatorial radius of the Earth.

An orbit is said to be sun-synchronous if its line of nodes rotates eastward at exactly the orbital angular velocity of the mean sun. Since the mean sun

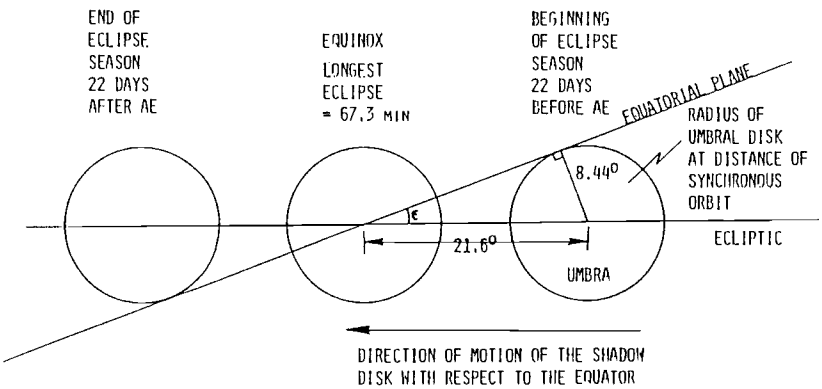


Fig. 11.14 Geosynchronous orbit eclipse at vernal or autumnal equinox.

moves uniformly eastward along the celestial equator through 360 deg in a tropical year (about 365.242 mean solar days), the required rate of nodal regression is 360/365.242, or 0.985647 deg/day. Substituting that numerical value in Eq. (11.14) and solving for the sun-synchronous inclination yields

$$i_{ss} = \cos^{-1}\left\{-0.098922(1-e^2)^2\left(1+\frac{h}{R}\right)^{3.5}\right\} \tag{11.15}$$

Note that, since the cosine of i_{ss} is always negative, the inclination of a sun-synchronous orbit must be greater than 90 deg; that is, sun-synchronous orbits are necessarily retrograde.

For circular orbits, e is zero, and h is the constant altitude above a spherical Earth with radius R. In that case, the variation of sun-synchronous inclination with altitude is shown in Fig. 11.15. Note that sun synchronism is possible for retrograde circular orbits up to an altitude of about 3226 n.mi. or a radius of 6670 n.mi. (for $R = 3444$ n.mi.), at which point the orbit inclination reaches its greatest possible value, 180 deg. Figure 11.16 shows examples of noon-midnight and twilight orbits.

Launch Window to Satisfy Beta-Angle Constraints

The term *launch window*, as used here, will mean the time interval (or intervals) on a given date within which a satellite can be launched into a prescribed orbit with the subsequent satisfaction of sun-orbit angle constraints (upper and lower limits on β) for the specified mission duration.

If the mission duration were zero days, one would simply determine the time interval(s) on the given launch date within which the combination of solar declination δ_S, orbit inclination i, and sun-node orientation $(\alpha_S - \Omega)$ yields acceptable values of β, as shown in Eq. (11.10).

However, if the mission has any extent, the determination of launch window involves more effort. In such a case, one must account for the changes that occur in δ_S and in $(\alpha_S - \Omega)$ during the mission and select the reduced launch window accordingly. As was mentioned earlier, the seasonal change in δ_S is inevitable. However, the initial value of $(\alpha_S - \Omega)$ can easily be controlled by varying the

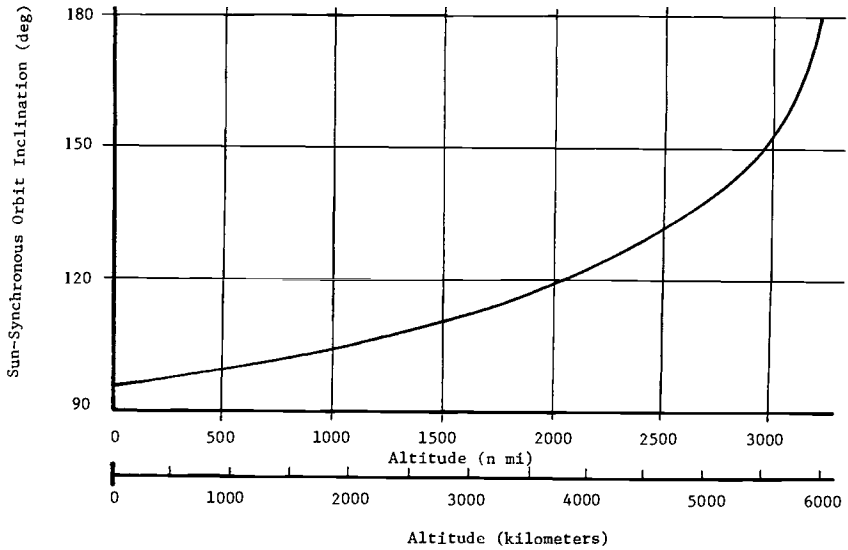


Fig. 11.15 Sun-synchronous inclination vs circular orbit altitude (from Ref. 1).

launch time; a 1-h delay in launch moves the line of nodes 15 deg eastward and thus reduces $(\alpha_S - \Omega)$ by 15 deg on any given launch date. Furthermore, the rate of change of $(\alpha_S - \Omega)$ can be adjusted by varying the orbit inclination or, to a lesser degree, the altitude. In particular, the average rate of change of $(\alpha_S - \Omega)$ can be made to vanish by employing a sun-synchronous combination of altitude and inclination. In this way, a preferred constant value of $(\alpha_S - \Omega)$ and an associated acceptable range of β values can often be maintained throughout the mission.

11.2 Time of Event Occurrence

If an event such as a perigee passage by a satellite occurs over a certain longitude λ, then the time t of the event can be specified unambiguously from the equation

$$\begin{aligned} \alpha_* &= \text{GHA} + \lambda + \omega_e t \\ &= \text{right ascension of the event} \end{aligned} \tag{11.16}$$

where

GHA = Greenwich hour angle (at Greenwich midnight of date)
λ = longitude of the event
ω_e = 15.041067 deg/h = Earth rotation rate
t = time from Greenwich midnight (GMT)

The values of GHA can be found in standard reference works for each date. For example, on January 1, 2000, GHA = 99.97 deg. If the event is a perigee passage by a satellite, then,

$$\alpha_* = \Omega + \mu_{\text{pp}} \tag{11.17}$$

where

$$\mu_{\text{pp}} = \tan^{-1}(\tan\omega \cos i)$$

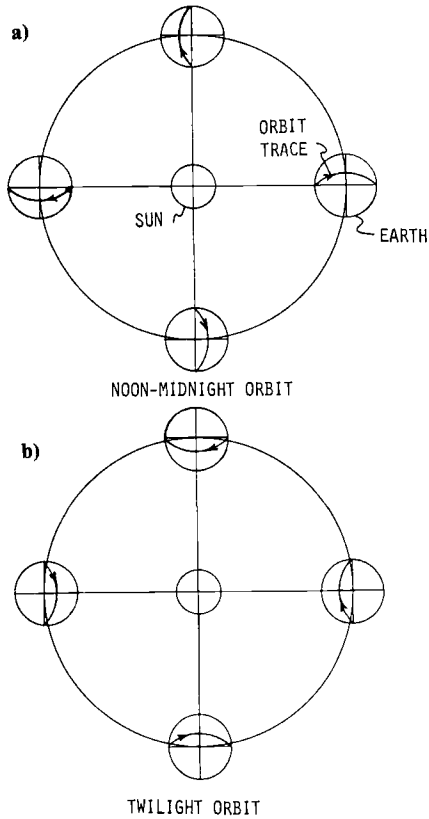


Fig. 11.16 Sun-synchronous orbits: a) noon-midnight orbit ($\Omega - \lambda = 0$); b) twilight orbit ($\Omega - \lambda = 90$ deg).

and

$$\omega = \text{argument of perigee } (<180 \text{ deg})$$

The time of perigee passage t_{pp} is then given by

$$t_{pp} = (\Omega + \mu_{pp} - \text{GHA} - \lambda_{pp})/\omega_e \tag{11.18}$$

where

$$\lambda_{pp} = \text{longitude of perigee passage}$$

11.3 Ground-Trace Considerations

General Characteristics

Circular figure eight, or "eggbeater," types of ground traces can be obtained by using satellites with 12- or 24-h periods at different inclinations to the equator. The use of several orbital planes equally spaced in node can result in several satellites moving in the same ground trace. Two examples of this are shown in

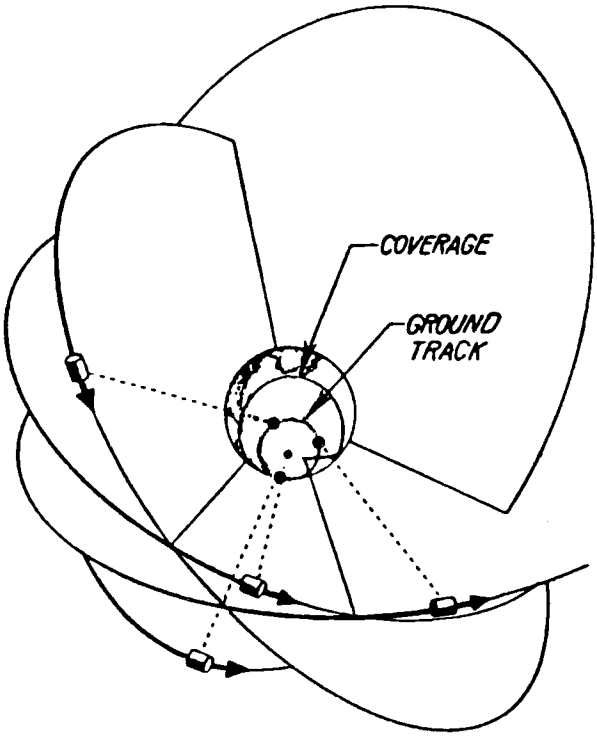


Fig. 11.17 Regional navigation satellite system cencept.

Figs. 11.17 and 11.18, which show the circular and eggbeater types of ground traces with several satellites in each ground trace. Figure 11.17 is for a regional system with four satellites, whereas Fig. 11.18 is for a possible global navigation system employing twenty satellites.

Other orbital systems involving up to four 12-h satellites in each of six equally spaced orbit planes have been found useful for global navigation purposes, as in the global positioning system (GPS), for example.

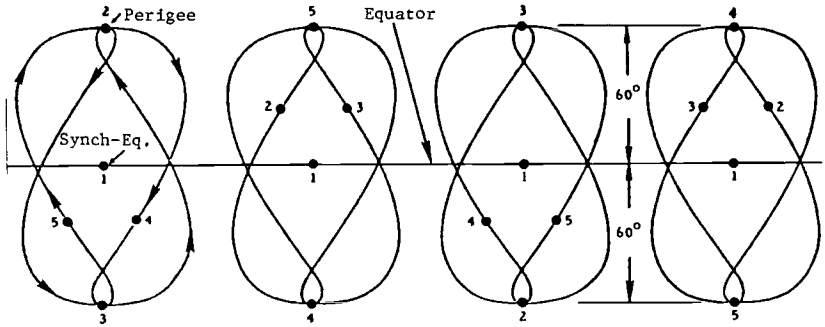
One measure of performance of a navigation satellite system is the performance factor called *geometric dilution of precision* (GDOP), which is a measure of how satellite geometry degrades accuracy. The magnitude of the ranging errors to a minimum of four selected satellites, combined with the geometry of the satellites, determines the magnitude of the user position errors in the GPS navigation fix. The four "best" visible satellites are those with the lowest GDOP.[2] Thus,

$$\text{GDOP} = \sqrt{(\text{PDOP})^2 + (\text{TDOP})^2} \tag{11.19}$$

where

PDOP = ratio of radial error in user position 1 σ in three dimensions to range error 1 σ.

TDOP = ratio of error 1 σ in the range equivalent of the user clock offset to range error 1 σ.



Global System (20 satellites)

4 satellites in each of 2 orbit planes 180° apart in longitude with perigee in southern hemisphere

4 satellites in each of 2 orbit planes 180° apart in longitude with perigee in northern hemisphere

4 satellites in synchronous equatorial orbit, 90° apart in longitude.

All inclined planes (i = 60°) with eccentricity e = .25 are separated 90° in longitude relative to each other.

Fig. 11.18 Global navigation satellite system concept.

Perturbation Effects

Earth's nonsphericity, solar-radiation pressure, and solar/lunar gravitational attraction tend to alter the satellites' orbital elements in time. An example of how the ground trace can change as a result of the Earth's gravitational harmonics is shown in Fig. 11.19, where the initial and a four-year ground trace are illustrated.

The results are for a $Q = 1$ eccentric ($e = 0.27$) orbit, whose period is $23^h55^m59.3^s$, with an initial inclination $i_0 = 28.5$ deg. The perturbations are due to the principal gravitational harmonics and the solar/lunar gravitational accelerations. The satellite trace repetition parameter Q is the number of satellite revolutions that occur during one rotation of the Earth relative to the osculating orbit plane. Q is approximately the number of satellite revolutions per day.

11.4 Highly Eccentric, Critically Inclined $Q = 2$ Orbits (Molniya)

This discussion is devoted to examining the characteristics of a very specialized orbit: the $Q = 2$ orbit, which is highly eccentric and critically inclined, with apogee located over the northern hemisphere. This type of orbit has the useful characteristic of enabling observation of vast areas of the northern hemisphere for extended periods of time each day. Typically, two properly phased spacecraft located in two ideal ground track locations (groundtracks repeat day after day) will view continuously 55–60% of the northern hemisphere, centered at the North Pole, as illustrated in Fig. 11.20.

Orbital Geometry

The following paragraphs examine each of the classical orbital elements, and optimal values are assigned to those elements for which such assignment has meaning.

Semimajor Axis

The semimajor axis a is determined by calculation to be approximately 26,554 km. This value assures that the groundtrack of the orbit will remain fixed relative

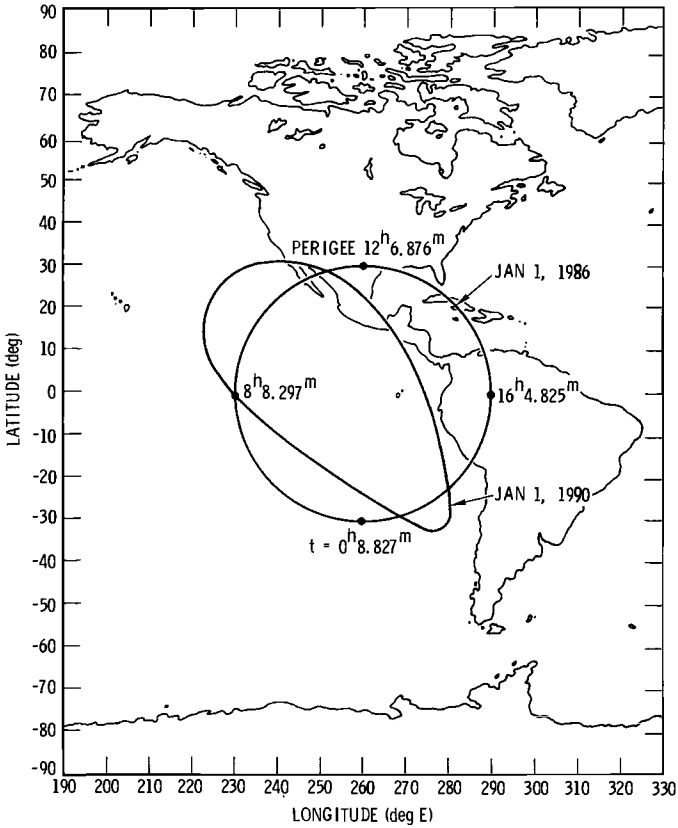


Fig. 11.19 Four-year ground trace change due to Earth's gravitational harmonic effects.

to the Earth and will repeat itself every day ($Q = 2$, an integer, assures this). Operationally, there are forces that tend to alter a (drag, solar pressure, tesseral harmonics, etc.), and fuel must be expended by the spacecraft to make periodic corrections to a.

When $a \approx 26{,}554$ km, the Keplerian period is 11.967 mean solar hours, which is one-half of the quantity 360 deg (one complete Earth rotation on its axis relative to the stars) divided by 15.041067 deg/h (the Earth's rotation rate). And so, to be a repeating groundtrack, $Q = 2$ orbit, the period is not 12 h (or one-half a mean solar day), but 11.967 h, which is one-half of a mean sidereal day.

Eccentricity

The value of eccentricity e will vary over a typical mission lifetime, largely because of solar and lunar gravitational perturbations. Eccentricities ranging from 0.69 to 0.74 are typical of this orbit type. The initial mission value of e is dictated by launch date, mission duration, minimum acceptable value of h_p, groundtrack location, and initial right ascension of ascending node (Ω). An adequately accurate computer program must be utilized to assure that the spacecraft will not re-enter the atmosphere at a premature time.

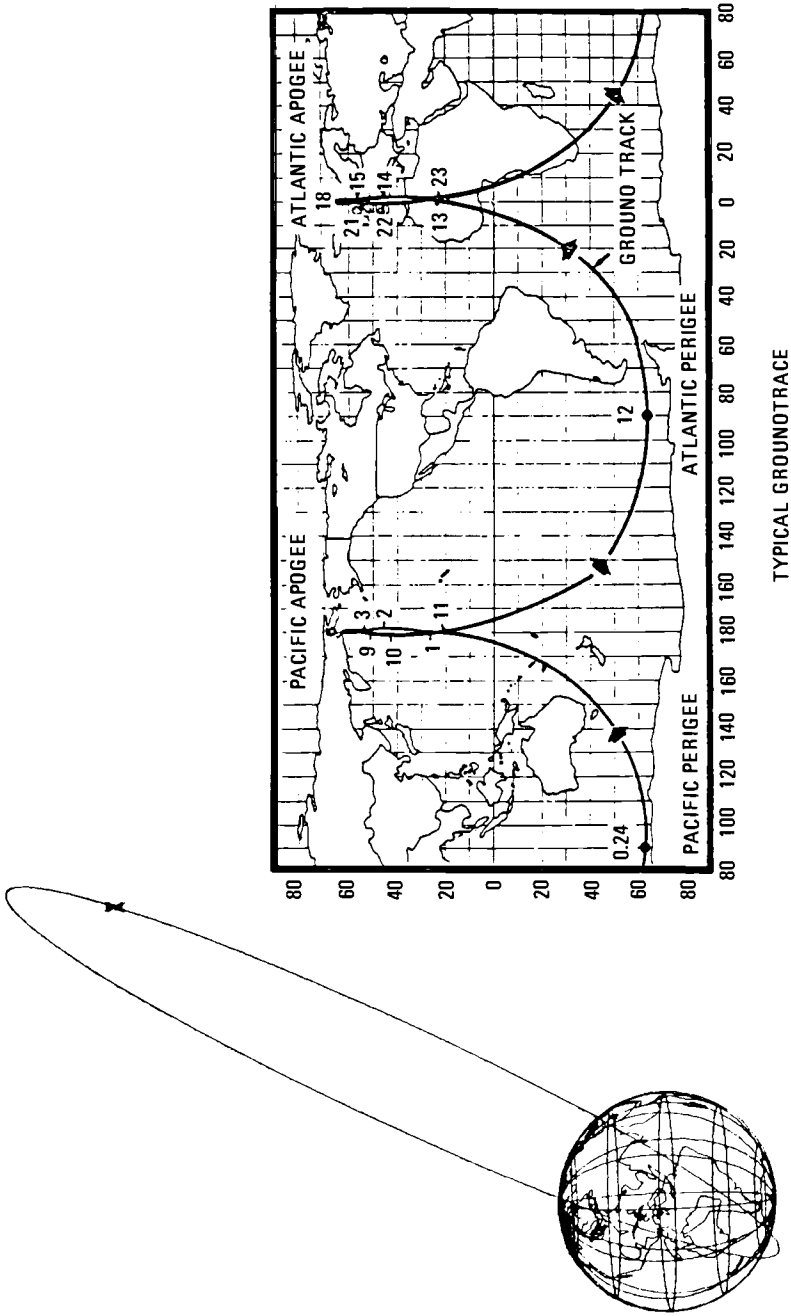


Fig. 11.20 $Q = 2$ (Molniya) type of orbit.

Fig. 11.20 $Q = 2$ (Molniya) type of orbit.

Inclination

For observation of northern hemisphere regions, the inclination should be high. In order that apogee be maintained at the northernmost point for long periods of time, the inclination must be approximately 63.44 deg (the critical inclination). Inclinations higher than 63.44 deg will force the line of apsides to rotate in a direction opposite the direction of satellite motion. Inclinations lower than 63.44 deg will rotate the line of apsides in the same direction as the satellite motion. These statements are valid for direct orbits only. The arguments are reversed for retrograde motion.

Right Ascension of Ascending Node

The node of the orbit can take any value, 0–360 deg. The value chosen essentially pinpoints the time of day of launch. At launch, the node (or time of day) must be carefully chosen so that lunisolar perturbations do not act to reduce h_p below an acceptable level during the nominal mission duration.

Argument of Perigee

In order that optimal visibility of the northern hemisphere be maintained, ω should be held as close to 270 deg as possible. Selecting $i = 63.44$ deg assures that ω will drive away from 270 deg by less than ± 5 deg for mission lifetimes on the order of 5 yr.

11.5 Frozen Orbits

Introduction

The term *frozen orbit* first appeared in the literature in 1978 in a paper by Cutting et al. entitled "Orbit Analysis for SEASAT-A."[3] A frozen orbit is one whose mean elements, specifically, eccentricity e and argument of perigee ω, have been selected to produce constant values, or nearly so, of e and ω with time. Thus, perigee rotation is stopped, and the argument of perigee is frozen at 90 deg. Frozen orbits maintain a constant altitude profile over the oblate Earth from revolution to revolution, with very small altitude variations over the northern hemisphere. This profile appears to decay uniformly because of atmospheric drag. Thus, frozen orbits are particularly useful for low-altitude missions that require eccentricity control. If radiation pressure and, perhaps, drag are not too influential, frozen orbit eccentricity will remain nearly constant for years. Otherwise, active maneuvering to maintain the frozen orbit is feasible and has been demonstrated (see Ref. 4). Frozen orbits are necessary for missions that require tight longitudinal control of the ground trace because there are no longitudinal variations of the ground trace due to perigee rotation.

The orbit will remain frozen in time if only zonal harmonics are assumed in the geopotential, i.e., no tesseral harmonics, and if atmospheric drag, solar-radiation pressure, and third-body gravitational perturbations are ignored.

Geometric interpretations of the first two zonal harmonics, J_2 and J_3, are shown in Fig. 11.21. The Earth's oblateness is caused by a combination of J_2 and the rotation of the Earth. The difference between the polar and equatorial radii of the Earth is 21.4 km. The pear shape of the Earth is caused by J_3. The height of the

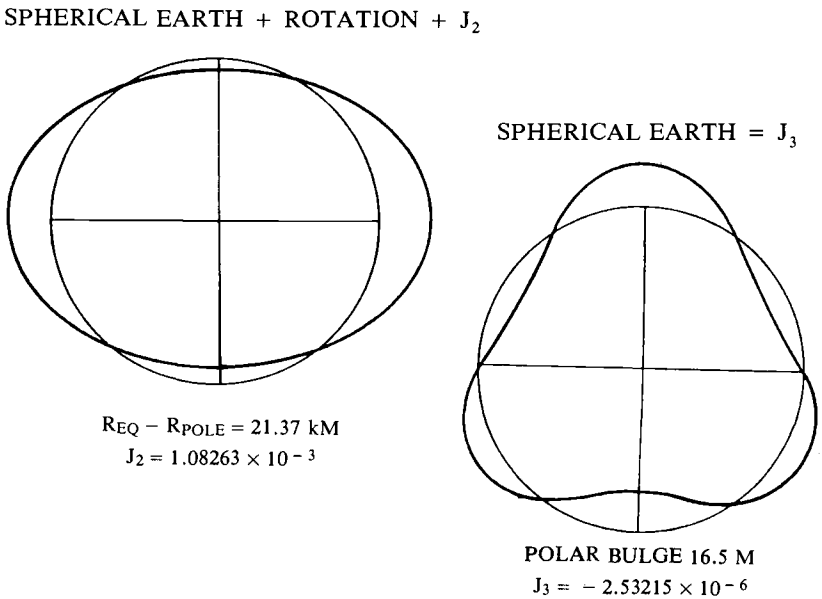


Fig. 11.21 Shape of the Earth.

bulge at the North Pole is about 16.5 m. For the Earth, the value of J_3 is three orders of magnitude less than the value of J_2.

Variational Rate Equations

Equations (11.20) and (11.21) are the averaged variational rate equations for eccentricity and argument of perigee for an Earth model consisting of a rotating Earth plus zonal harmonics J_2 and J_3 (see Chapter 9). All elements are mean elements:

$$\frac{de}{dt} = \frac{3J_3 n}{2(1-e^2)^2}\left(\frac{R}{a}\right)^3 \sin i\left(1 - \frac{5}{4}\sin^2 i\right)\cos\omega \tag{11.20}$$

$$\frac{d\omega}{dt} = \frac{3J_2 n}{(1-e^2)^2}\left(\frac{R}{a}\right)^2\left(1 - \frac{5}{4}\sin^2 i\right)\left[1 + \frac{J_3}{2J_2(1-e^2)} \times\left(\frac{R}{a}\right)\frac{\sin i \sin\omega}{e}\right] \tag{11.21}$$

where

n = the satellite mean motion
R = the Earth's mean equatorial radius

Mean elements reflect secular and long-period perturbative variations. Mean elements can be obtained from osculating elements by averaging out the short-period oscillations.

Note that there is no J_2 term in Eq. (11.20) but that Eq. (11.21) contains both J_2 and J_3. Also, note that both equations contain the well-known expression $(1 - 5/4 \sin^2 i)$. These equations equal zero when this expression equals zero, i.e., when $i = 63.4$ deg or 116.6 deg, the critical inclination. These are frozen-orbit solutions. This solution was probably first implemented in 1964 with the advent of the Molniya satellite orbit.

However, the frozen-orbit solutions of interest are implemented by setting $\omega = 90$ deg, so that $\mathrm{d}e/\mathrm{d}t = 0$, and by setting the bracketed term in Eq. (11.21) equal to zero, so that $\mathrm{d}\omega/\mathrm{d}t = 0$. From this term, the frozen mean eccentricity can be solved for given values of the mean semimajor axis and mean inclination. Note that the eccentricity [obtained as a solution to Eq. (11.22)] is

$$e \approx -\frac{J_3}{2J_2}\left(\frac{R}{a}\right)\sin i \tag{11.22}$$

It is of the order of 10^{-3} because J_3 is three orders of magnitude less than J_2 (see Fig. 11.21).

Frozen-Orbit Solutions

Results from Eq. (11.22) are presented in Fig. 11.22 as the dashed line. Extensive numerical investigations of frozen-eccentricity solutions have revealed the importance of higher-order zonal harmonics, especially for inclinations near the critical inclination. The solid line in Fig. 11.22 describes frozen-eccentricity solutions obtained by using the $J_2 - J_{12}$ zonal harmonic terms in the WGS-84 geopotential model.

There are distinct behaviors of these solutions in three regions of inclination: $i < 63.4$ deg, 63.4 deg $< i <$ 66.9 deg, and 66.9 deg $< i \leq 90$ deg. At $i = 10$ deg, the frozen eccentricity is 0.132×10^{-3}. As inclination increases, the frozen eccentricity increases in an oscillatory manner as shown in Fig. 11.22. The frozen-eccentricity increases only slightly between $i = 35$ deg and $i = 40$ deg. The frozen-eccentricity increases dramatically, however, as the inclination approaches the critical value, i.e., $i = 63.435$ deg. A frozen solution, $e = 10.1 \times 10^{-3}$, was found at $i = 63.0$ deg. Presumably, frozen orbits with larger eccentricities can be found in the inclination region between 63.0 and 63.435 deg.

Inverted frozen orbits were discovered for inclinations slightly greater than 63.435 deg. Inverted frozen-orbit solutions occur at $\omega = 270$ deg rather than $\omega = 90$ deg. Thus, perigee is at the southernmost point in the orbit rather than at the northernmost point. The initial discovery of inverted frozen orbits was published by J. C. Smith in 1986.[6] At $i = 63.6$ deg, the frozen eccentricity was found to be 23.5×10^{-3}, or 0.0235. As inclination increases, the frozen eccentricity decreases dramatically. At $i = 66.6$ deg, the frozen solution is $e = 0.124 \times 10^{-3}$ and, at $i = 66.8$ deg, $e = 0.041 \times 10^{-3}$.

Then, as inclination increases slightly, a transition in frozen solutions from $\omega = 270$ deg to $\omega = 90$ deg occurs, so that, for $i = 67.1$ deg, $e = 0.068 \times 10^{-3}$ at $\omega = 90$ deg. And, for a narrow range of inclinations, no frozen-orbit solutions could be found. As inclination increases from $i = 67.1$ deg, eccentricity increases rapidly to $e = 0.65 \times 10^{-3}$ at $i = 70.0$ deg. Then, as Fig. 11.22 shows, the frozen eccentricity increases less rapidly with increasing inclination until $e = 1.26 \times$

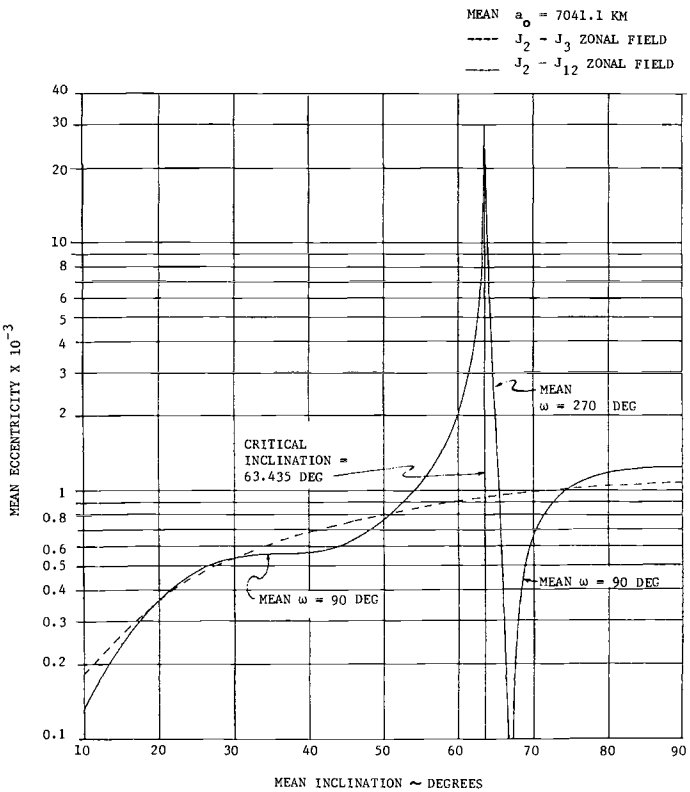


Fig. 11.22 Frozen-orbit solutions.

10^{-3} at $i = 90.0$ deg. Thus, the curve corresponding to the $J_2 - J_{12}$ zonal field is quite different from the curve corresponding to the J_2 and J_3 zonal field, especially for inclinations between 50 and 75 deg. In the vicinity of the critical inclination, the values of de/dt and $d\omega/dt$ from Eqs. (11.20) and (11.21) are near zero. Terms involving higher zonal harmonics than J_2 and J_3 then become relatively more important.

Circulations About the Frozen-Orbit Conditions

It is useful and interesting to examine the behavior of mean e and mean ω for initial values that are near the frozen-orbit values. For nearby initial values, e and ω will circulate about the frozen-orbit conditions as illustrated in Fig. 11.23. This figure is from Ref. 5. For $i = 98.7$ deg and a semimajor axis value of 7198.7 km, the frozen eccentricity is 1.15×10^{-3} at $\omega = 90$ deg. Although the semimajor axis is slightly different, by about 160 km, the value of frozen e is very nearly the value obtained from Fig. 11.22 at an inclination of 81.3 deg, i.e., the supplement of 98.7 deg.

For initial conditions that are near, but not at, the frozen point, e and ω will move counterclockwise in closed contours. For inclinations less than 63.435 deg

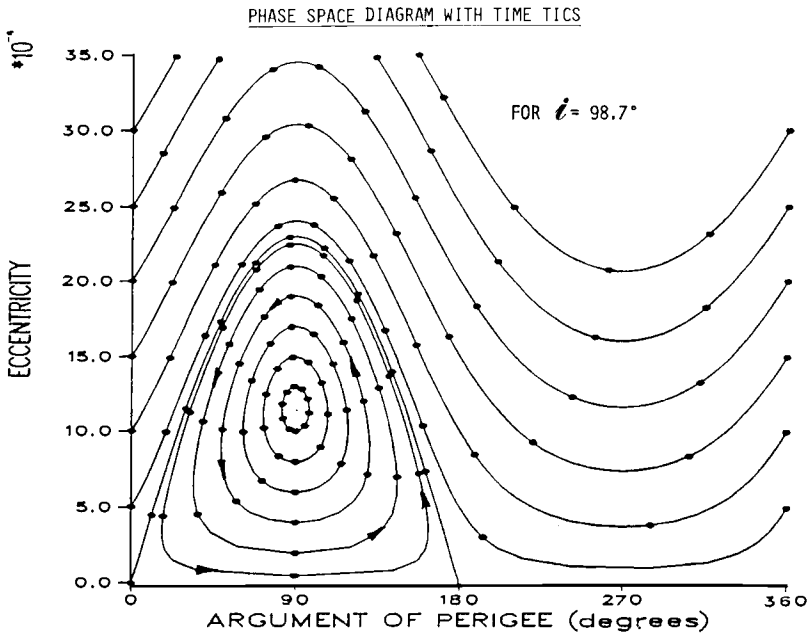


Fig. 11.23 Phase space diagram marked at 13-day interval (from Ref. 5).

or greater than 116.565 deg, the motion is clockwise. The period of circulation is 360 deg/$\dot{\omega}$, where $\dot{\omega}$ is the familiar J_2 secular apsidal rate. A $J_3 - J_{21}$ zonal field was used to generate the circulation curves, and 13-day intervals in time were marked on the curves. The distribution of time tics on the larger closed curves is interesting. More time tics are clustered at the higher than the lower eccentricities. Thus, during one circulation, the eccentricity is larger than the frozen value for a longer period of time than it is smaller than the frozen value. For larger deviations from the frozen point, the contours do not close.

In Ref. 7, Rosborough and Ocampo derive and present frozen-orbit eccentricity for an Earth gravity field complete to degree 50. Their results show that inverted frozen orbits, i.e., $\omega = 270$ deg, occur in a low-altitude ($a < 1.09$ Earth radii), low-inclination ($i < 10$ deg) region as well as in a region of inclination 63.4 deg $< i <$ 66.9 deg.

References

[1]Ginsberg, L. J., and Luders, R. D., *Orbit Planner's Handbook*, The Aerospace Corp., El Segundo, CA, Technical Memorandum, 1976.

[2]Milliken, R. J., and Zoller, C. J., "Principle of Operation of Navstar and System Characteristics," *Navigation: Journal of the Institute of Navigation*, Vol. 25, Summer, 1978, pp. 95–106.

[3]Cutting, E., Born, G. H., and Frautnick, J. C., "Orbit Analysis for SEASAT-A," *Journal of the Astronautical Sciences*, Vol. XXVI, Oct.–Dec. 1978, pp. 315–342.

[4]McClain, W. D., "Eccentricity Control and the Frozen Orbit Concept for the Navy Remote Ocean Sensing System (NROSS) Mission," AAS Paper 87-516, Aug. 1987.

[5]Nickerson, K. G. et al., "Application of Altitude Control Techniques for Low Altitude Earth Satellites," *Journal of the Astronautical Sciences,* Vol. XXVI, April–June 1978, pp. 129–148.

[6]Smith, J. C., "Analysis and Application of Frozen Orbits for the Topex Mission," AIAA Paper 86-2069-CP, Aug. 1986.

[7]Rosborough, G. W., and Ocampo, C. A., "Influence of Higher Degree Zonals on the Frozen Orbit Eccentricity," AAS Paper 91-428, Aug. 1991.

[8]Milstead, A. H., "Launch Windows for Orbital Missions," The Aerospace Corp., El Segundo, CA, Report No. TDR-269(4550-10)-6, April 1, 1964.

Problem

11.1. On 19 June, 1995 the Space Shuttle *Atlantis* was scheduled to launch and rendezvous with the Russian Space Station *Mir.* Cape Canaveral is the launch site, $\delta = 28.34$ deg. Nominally, *Atlantis* would launch directly into *Mir*'s orbit plane whose inclination was 51.6 deg. For performance reasons *Atlantis* has a plane change capability of only 52.4 m/sec. If the plane change occurs at an altitude of 350 km, what was the launch window associated with *Atlantis'* plane change capability? What was *Atlantis'* launch azimuth from the Cape?

Selected Solution

11.1. Launch window = +/− 2 min
Launch azimuth = 44.9 deg

12
Lunar and Interplanetary Trajectories

12.1 Introduction

The solar system consists of a single star (the sun) and nine principal planets that move around the sun in orbital paths that are nearly circular except for that of Pluto, which is highly eccentric. Early man considered the Earth to be the center of the universe and the five planets, Venus, Mercury, Mars, Jupiter, and Saturn, to be divine. The most ancient observations of the planets date back 2000 years B.C. and appear to come from the Babylonian and Minoan civilizations. The term *planet* means "wanderer" in Greek and refers to the celestial objects that move relative to the stars.

The Egyptians, Greeks, and Chinese once thought of Venus, for example, as two stars because it was visible first in the morning and then in the evening sky. The Babylonians called Venus "Istar," the personification of woman and the mother of the gods. In Egypt, the evening star was known as Quaiti and the morning star Tioumoutiri; to the Chinese, Venus was known as Tai-pe, or the Beautiful White One. The Greeks called the morning star Phosphorus and the evening star Hesperos but, by 500 B.C., the Greek philosopher Pythagoras had come to realize that the two were identical. As time passed, the Romans changed the name of the planet to honor their own goddess of love, Venus.

It was not until the Golden Age of Greece that astronomy as a science was placed on a firm foundation and the Earth and the planets were regarded as globes rather than flat surfaces. Had Greek quantitative analysis taken one more step and dethroned the world from its position as the center of the universe, the progress of human thought and logic would have been accelerated. The Greek philosopher and mathematician Aristarchus held a heliocentric view of the solar system, but his ideas were opposed on religious grounds, and the later Greeks reverted to the idea of a central Earth.

Ptolemy, who died in about A.D. 150, left a record of the state of the universe at the end of the classic Greek period. In his Ptolemaic system, the Earth lies in the center of the universe, with the various heavenly bodies revolving around it in perfect circles. First comes the moon, the closest body in the sky; then come Mercury, Venus, and the sun, followed by the three other planets then known (Mars, Jupiter, and Saturn), and finally the stars.[1]

An interesting law of planetary spacing was suggested by Bode in 1771, who stated that the normalized mean distances of the planets from the sun are given by the terms of the series 0, 3, 6, 12, etc., when added to 4 and divided by 10. This law implied that there should also be a planet between Mars and Jupiter, which led to the discovery of the asteroid Ceres and the asteroid belt. Bode's law holds for the planets closest to the sun [e.g., for Earth $(6 + 4)/10 = 1$ A.U. (astronomical unit, mean radius of the Earth orbit)] but fails for Neptune, where it predicts 38.3 vs 30.2 actual distance. It also appears to be valid for the satellites of the planets as, for example, those of Uranus. The planets and their relative size are illustrated in Fig. 12.1, and the planetary and satellite data are given in Tables 12.1 and 12.2, respectively.

This picture represents the relative sizes of the planets in the Solar System, compared with the Sun's disk. The distances from the Sun are given in millions of miles. As a visual reference, the Earth is about 110 Sun diameters away or about 11,700 times Earth's own diameter away. On the scale of this picture, Earth would actually be about 20 meters (65 ft) away from the Sun. Jupiter is 5 times as far, and Saturn nearly 10 times as far. Since Pluto's orbit is highly eccentric, it now is closer to the Sun than Neptune, which would be 30 times farther away than the Earth.

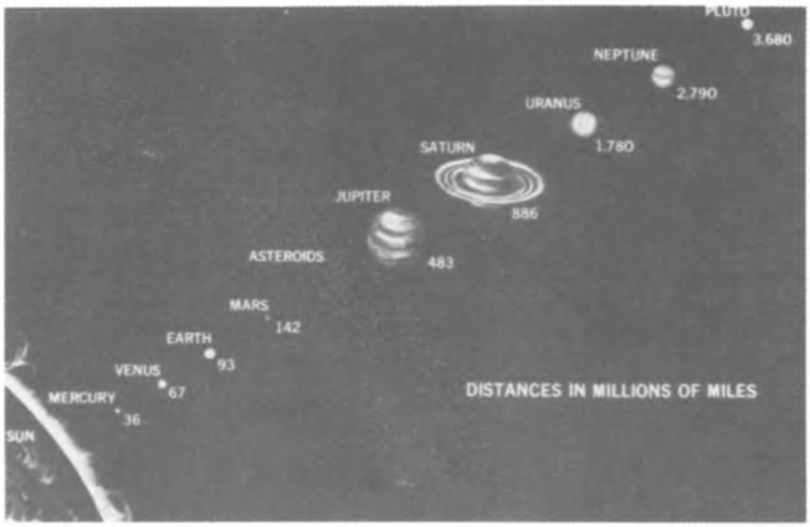


Fig. 12.1 The planets' relative sizes and mean distance from the sun.

Fundamental reasons for study of the solar system are exploration, questions related to the origin of the solar system, and asteroid mining missions, for example. Other potential applications include nuclear waste disposal missions and eventual colonization of the solar system. In this chapter, after a short historical background, the sphere of influence and the gravity-assist concepts are examined. Simple patched-conic approaches are then used to describe missions to the moon, and Mars.

12.2 Historical Background

During the 20-yr period from 1962 to 1981, the exploration of the solar system was initiated by means of unmanned spacecraft. Flybys of the inner planets Venus and Mars were accomplished (by Mariners 2 and 4–7 in the United States). In the 1970s, multiple flybys of the inner planets were performed, including orbiters about Mars and Venus (Mariner 9, Viking, and Pioneer Venus in the United States). Delivery of soft-landing vehicles to Mars and Venus was accomplished and flybys of Jupiter performed.

The 1980s began with the encounters of Pioneer 11 and two Voyager spacecraft with Saturn. During this period, the Soviet Union also sent a number of flyby, orbiting, atmospheric entry, and landing spacecraft to Venus and Mars. Complex terrain across several thousand miles of the surface of Venus was, for example, photographed by the Soviet Venera 16 imaging radar spacecraft, which revealed a

Table 12.1 Planetary data (from Ref. 8)

Planet	μ, km^3/s^2	Equatorial diameter, km	Mean distance from sun, 10^6 km	Sidereal period	Axial rotation (equatorial)	Axial inclination, deg	Mean Synodic period, days
Mercury	2.232×10^4	4670	57.9	88.0 days	58.7 days	?	115.9
Venus	3.257×10^5	12400	108	224.7 days	243 days	?	584.0
Earth	3.986×10^5	12700	149.6	365.3 days	23H 56M 04S	23° 27′	–
Mars	4.305×10^4	6760	227.7	687.0 days	24H 56M 23S	23° 59′	779.9
Jupiter	1.268×10^8	143000	777.8	11.86 yr	9H 50M 30S	3° 04′	398.9
Saturn	3.795×10^7	121000	1486	29.46 yr	10H 14M	26° 44′	378.1
Uranus	5.820×10^6	47100	2869	84.01 yr	10H 49M	97° 53′	369.7
Neptune	6.896×10^6	50700	4475	164.79 yr	About 14H	28° 48′	367.5
Pluto	3.587×10^5	5950	5899	248.43 yr	6D 9H	?	366.7

Planet	Escape velocity, km/s	Density: water=1	Volume: Earth=1	Mass: Earth=1	Surface gravity: Earth=1	Max surface, temperature, °F	Number of Satellites
Mercury	4.2	5.5	0.06	0.06	0.38	+770	0
Venus	10	5.3	0.86	0.82	0.90	+887	0
Earth	11	5.5	1	1	1	+140	1
Mars	6.4	3.9	0.15	0.11	0.38	+80	2
Jupiter	59.7	1.3	1319	318	2.64	−200	13
Saturn	35.4	0.7	744	95	1.16	−240	10
Uranus	22.4	1.7	47	15	1.11	−310	5
Neptune	31	1.8	54	17	1.21	−360	8
Pluto	?	?	0.1?	0.0026	?	?	1

Table 12.1 Satellite data (from Ref. 8)

Planet/Satellite	Mean distance from center of primary[a]	Sidereal period			Diam, km	Density: water = 1	Maximum magnitude	Reciprocal mass primary = 1	Discoverer
Earth									
Moon	385	27D	7H	43M	3470	3.3	12.7	81.3	—
Mars									
Phobos	9.33		7H	39M	23.3	?	11	?	Hall, 1877
Deimos	23.5	1D	6H	18M	11.3	?	12	?	Hall, 1877
Jupiter									
Amalthea (V)	182		11H	57M	241	?	13	?	Barnard, 1892
Io (I)	422	1D	18H	28M	3710	4.1	5.5	26,200	Galileo, 1609
Europa (II)	671	3D	13H	14M	3140	3.7	5.7	40,300	Galileo, 1609
Ganymede (III)	1,070	7D	3H	43M	5150	2.4	5.1	12,200	Galileo, 1609
Callisto (IV)	1,880	16D	16H	32M	4820	2.0	6.3	19,600	Galileo, 1609
Hestia (VI)	11,400	250D	26H		161	?	13.7	?	Perrine, 1904
Hera (VII)	11,700	259D	16H		56.3	?	17	?	Perrine, 1905
Demeter (X)	11,700	260D	12H		24.1	?	18.8	?	Nicolson, 1938
Adrastea (XII)	20,900	631D[b]			28	?	18.9	?	Nicolson, 1951
Pan (XI)	22,500	692D[b]			30.6	?	18.4	?	Nicolson, 1938
Poseidon (VIII)	23,500	744D[b]			56.3	?	18	?	Melotte, 1908
Hades (IX)	23,600	758D[b]			27.4	?	18.4	?	Nicolson, 1914
XIII	22,500		?[b]		8.04	?	?	?	Kowal, 1974

[a]Thousands of kilometers.
[b]Indicates retrograde motion.

Table 12.1 Satellite data (from Ref. 8) (cont.)

Planet/Satellite	Mean distance from center of primary[a]	Sidereal period			Diam, km	Density: water = 1	Maximum magnitude	Reciprocal mass primary = 1	Discoverer
Saturn									
Janus	158		17H	58M	241	?	14	?	Dollfus, 1966
Mimas	182		22H	37M	483	1	12	15,000,000	Herschel, 1789
Enceladus	240	1D	8H	53M	644	1	11	7,000,00	Herschel, 1789
Tethys	295	1D	21H	18M	1130	1.1	10.5	910,000	G. D. Cassini, 1684
Dione	378	2D	17H	41M	1450	3.2	10.5	910,000	G. D. Cassini, 1684
Rhea	528	4D	17H	25M	1770	2	9.3	250,000	G. D. Cassini, 1672
Titan	1,889	15D	22H	41M	5310	2.3	8.3	4,150	Huygens, 1655
Hyperion	1,488	21D	6H	38M	322	3	13	5,000,000	Bond, 1848
Lapetus	3,540	79D	7H	56M	2410	?	9	?	G. D. Cassini, 1671
Phoebe	13,000	550D[b]	10H	50M	241	?	14	?	Pickering, 1898
Uranus									
Miranda	122	1D	9H	50M	322	5	17	1,000,000	Kuiper, 1948
Ariel	192	2D	12H	29M	2410	5	14	67,000	Lassell, 1851
Umbriel	267	4D	3H	28M	1290	4	14.7	170,000	Lassell, 1851
Titania	438	8D	16H	56M	2410	6	14	20,000	Herschel, 1787
Oberon	586	13D	11H	7M	2410	5	14	34,000	Herschel, 1787
Neptune									
Triton	354	550D[b]	10H	50M	4830	?	14	750	Lassell, 1846
Nereid	5,630	550D[b]	10H	50M	322	?	14	?	Kuiper, 1949

[a]Thousands of kilometers.
[b]Indicates retrograde motion.

nearly 11-km-high mountain topped with a 100-km-diam crater, believed to be a massive meteorite impact crater.[2]

Preliminary explorations of comets and the continuing Voyager 2 mission to Uranus and Neptune were performed, with intensive investigation of the outer planets Saturn, Uranus, and Neptune, along with Saturn's satellite Titan.[3] Voyager 2's encounter with Uranus on 24 January 1986 showed, for example, that the planet's magnetic pole is inclined from the pole of rotation by an angle of about 55 deg, which is the largest inclination in the solar system. Images of Miranda and Umbriel, the satellites of Uranus, revealed three terrain types of different age and geology, such as hills, groved valleys, and craters.

One of the major exploration projects in the United States was the Galileo mission to Jupiter originally planned for 1986 but subsequently delayed to October 1989 because of the grounding of the Space Shuttle fleet and the new safety requirements posed by loss of Shuttle Centaur upper stage. A possible backup mission was also planned for July 1991 if the 1989 mission was not possible.

The trajectory to Jupiter employed a two-stage inertial upper stage (IUS) using one gravity-assist maneuver at Venus and two maneuvers at Earth, requiring more than 6 years of travel time.[4]

The new trajectory, which seemed to send Galileo on a cruise through the solar system, soon earned the name "Solar Cruiser."

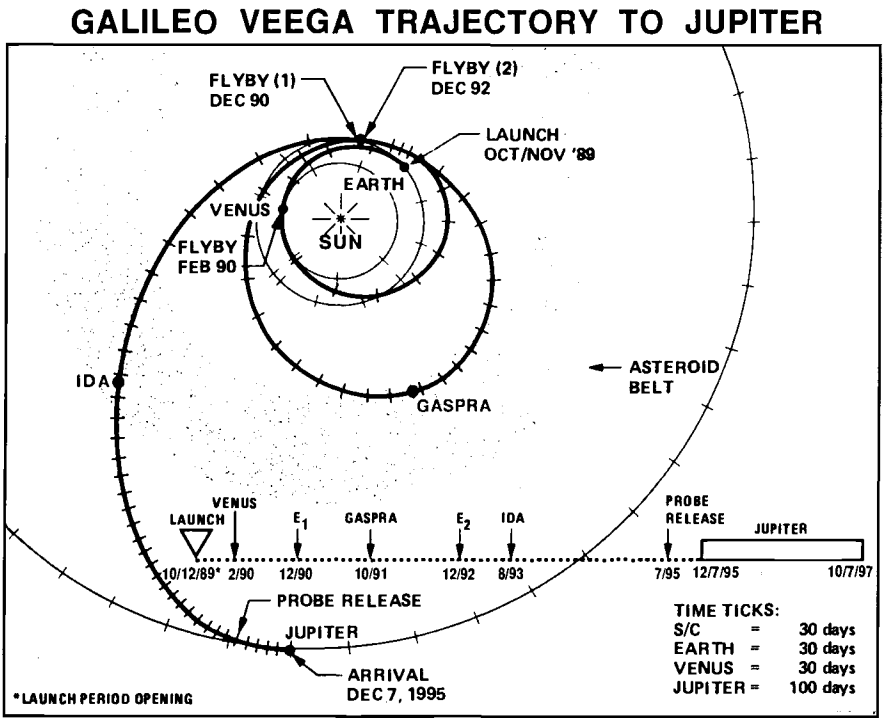


Fig. 12.2 Galileo's route to Jupiter.

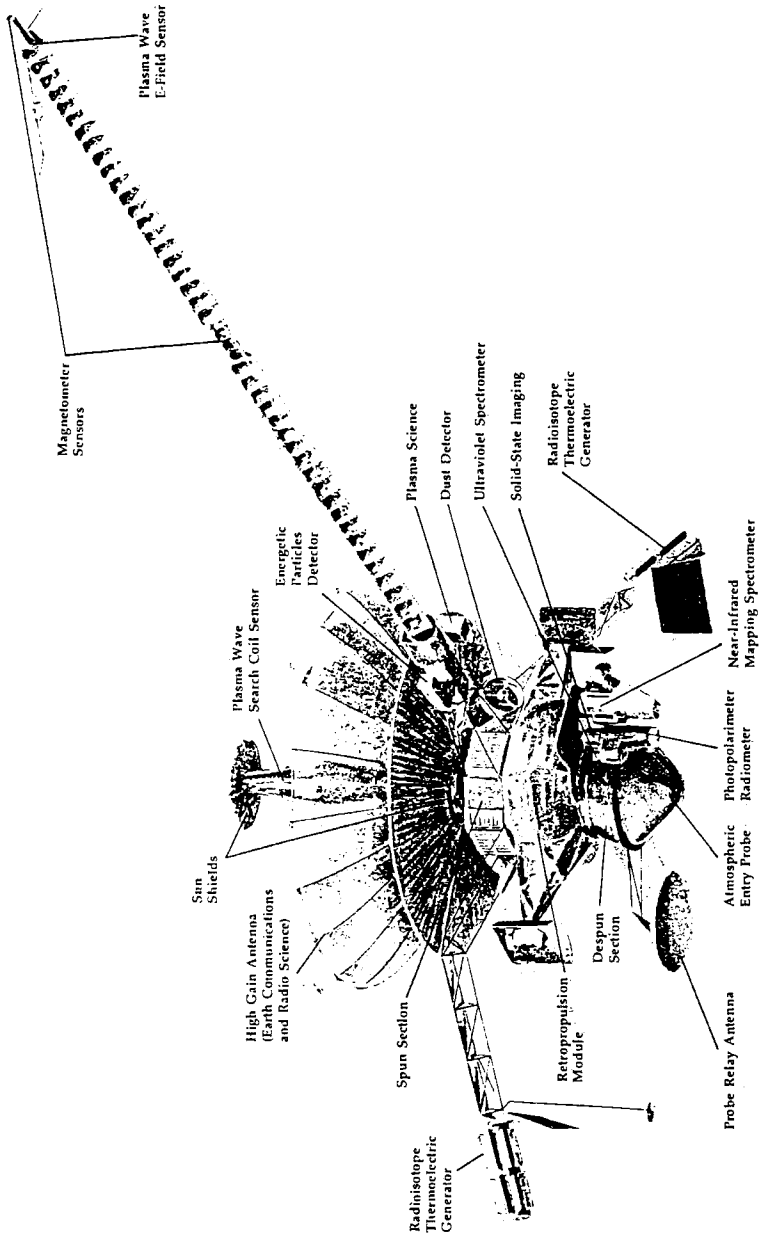


Fig. 12.3 Galileo's spacecraft.

Galileo was launched in October 1989 from Earth on a Space Shuttle and an inertial upper stage, a rocket whose energy is low compared to the Centaur. Instead of heading toward Jupiter or the asteroid belt, Galileo took a flight path that carried it to Venus. Galileo arrived there in February 1990. Venus' gravity accelerated Galileo and sent it on a flight path back toward Earth. When Galileo passed Earth in December 1990, the Earth's gravitational field added energy to send Galileo out to the asteroid belt. A propulsive maneuver, performed in December 1991, brought Galileo past Earth again in December 1992 for a last gravity assist before the spacecraft began its final path to Jupiter. Arrival at Jupiter occurred late in 1995.

A view of the Galileo trajectory and its spacecraft are shown in Figs. 12.2 and 12.3, respectively.

About 150 days before arrival of the Galileo spacecraft, the atmospheric probe separated from the Orbiter. The Orbiter then flew within 1000 km of the satellite Io, whose gravitational field helped to slow the spacecraft. The probe is designed to sample Jupiter's equatorial zone, which consists primarily of ammonia.

According to the plan, the orbiter entered a Jovian orbit ranging from 200,000 km to more than 10 million km for a 20-month study of Jupiter's environment and its satellites. Current knowledge of this environment is illustrated in Figs. 12.4–12.6.

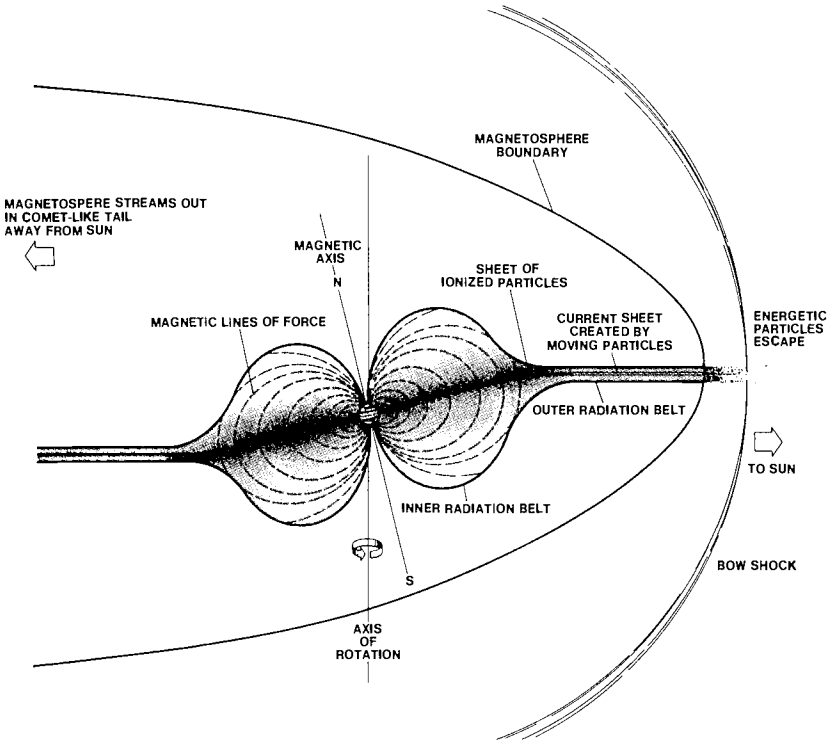


Fig. 12.4 Jupiter's magnetosphere (from Ref. 8).

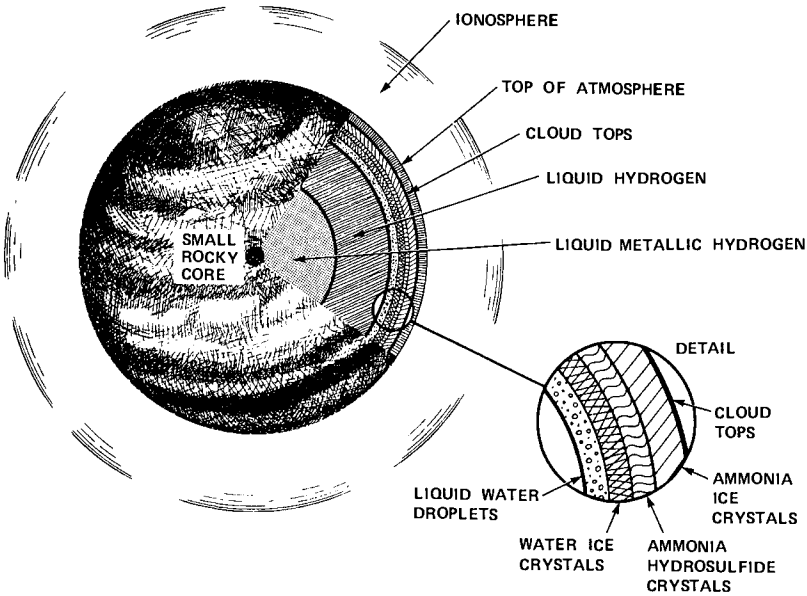


Fig. 12.5 Jupiter atmosphere model (scale exaggerated); atmosphere depth to liquid zone is 1000 km (600 miles) (from Ref. 8).

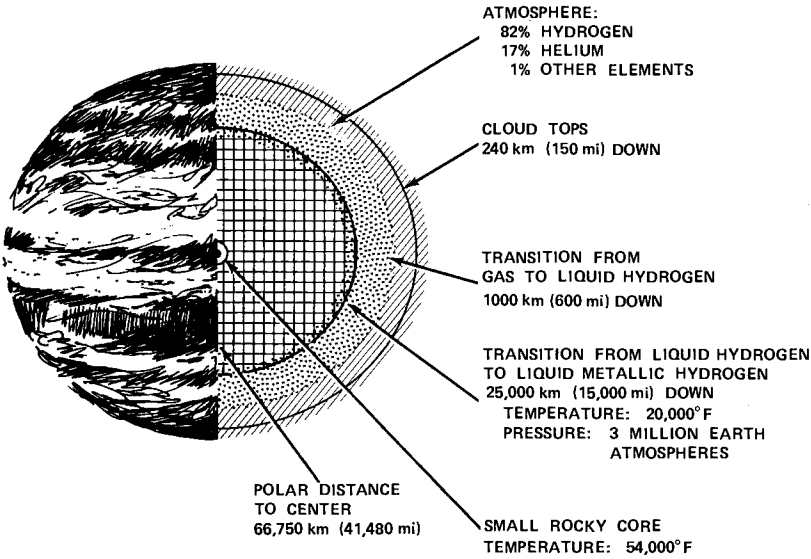


Fig. 12.6 Interior of liquid Jupiter; planet is mainly hydrogen (from Ref. 8).

12.3 Important Concepts

Sphere of Gravitation[5]

Consider the motion of a mass point (P, m) under the influence of two larger masses m_1 and m_2, as shown in Fig. 12.7. The larger masses are termed centers of attraction A_1 and A_2. It is assumed that $m_1 \ll m_2$ and $m \ll m_1$.

The values of the gravitational forces F_1 and F_2 acting on mass m toward A_1 and A_2 are with reference to Fig. 12.7 given by

$$F_1 = \frac{Gmm_1}{|\boldsymbol{A_1P}|^2}, \qquad F_2 = \frac{Gmm_2}{|\boldsymbol{A_2P}|^2} \tag{12.1}$$

where G = universal constant of gravitation.

The locus of points where $F_1 > F_2$ defines the sphere of gravitational attraction of mass m_1 with respect to mass m_2. The location and the radius of the sphere are determined from boundary condition

$$F_1 = F_2$$

or

$$\frac{A_1P}{A_2P} = \sqrt{\frac{m_1}{m_2}} = \text{const} < 1 \tag{12.2}$$

which indicates that the ratio of the distance from P to A_1 and A_2 is constant. From elementary geometry, the locus of points defined by this condition is a sphere, the diameter of which is defined by points C and D in Fig. 12.7.

Thus, letting R_s, R_o be the radius and origin of the sphere of gravitation and using Eq. 12.1 for the collinear points of attractions C and D, the ratio of the distances

$$\frac{A_1C}{A_2C} = \frac{A_1D}{A_2D} = \sqrt{\frac{m_1}{m_2}} \tag{12.3}$$

or

$$\frac{R_s - R_o}{A_1A_2 - (R_s - R_o)} = \frac{R_s + R_o}{A_1A_2 + (R_s + R_o)} = \sqrt{\frac{m_1}{m_2}} \tag{12.4}$$

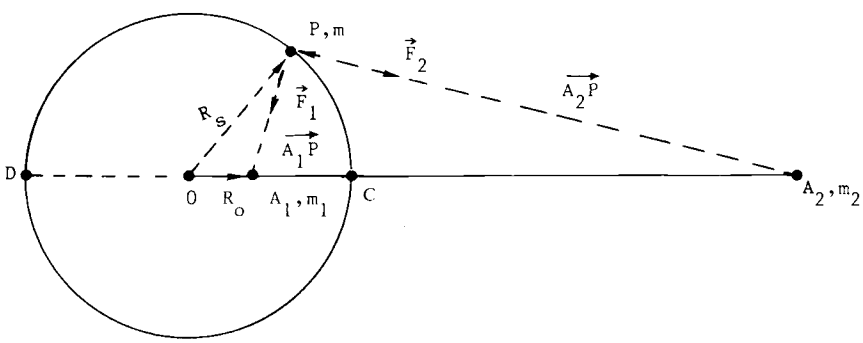


Fig. 12.7 Sphere of gravitation.

which yields

$$R_s = \frac{A_1A_2\sqrt{m_1/m_2}}{1 - m_1/m_2} \tag{12.5}$$

$$R_o = \frac{A_1A_2(m_1/m_2)}{1 - m_1/m_2} \tag{12.6}$$

The notation A_1A_2 denotes the distance from A_1 to A_2.

For example, the lunar sphere of gravitation is found as

$$R_s = \frac{384000\sqrt{1/81}}{1 - (1/81)}$$

$$\approx 43000 \text{ km}$$

and its location

$$R_o = \frac{384000(1/81)}{1 - (1/81)}$$

$$\approx 4500 \text{ km}$$

Sphere of Influence[5]

The sphere of influence is defined as the locus of points measured with respect to the gravitational centers of attraction A_1 or A_2, where the ratios of the perturbative to primary gravitational accelerations of the centers are equal. To find the magnitude of the sphere of influence, consider the notation in Fig. 12.8.

The motion of point P (of mass m) with respect to the gravitational center of attraction A_1 can be expressed as

$$\frac{d^2\boldsymbol{\rho}}{dt^2} = \boldsymbol{a}_1 + \boldsymbol{\Phi}_1 \tag{12.7}$$

where

$$\boldsymbol{a}_1 = \frac{-G(m_1 + m)}{\rho^3}\boldsymbol{\rho} \tag{12.8}$$

$$= \text{primary acceleration of } A_1$$

$$\boldsymbol{\Phi}_1 = Gm_2\left(\frac{\mathbf{r} - \boldsymbol{\rho}}{r_{12}^3} - \frac{\mathbf{r}}{r^3}\right) \tag{12.9}$$

$$= \text{perturbative acceleration of } A_2$$

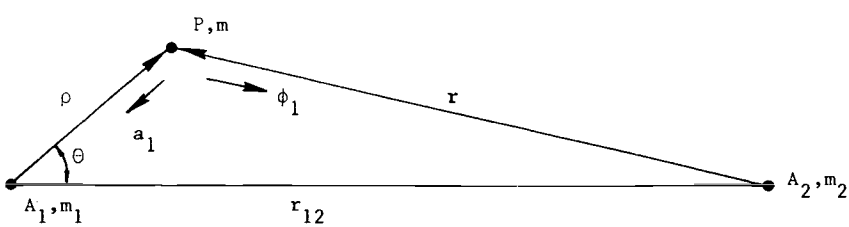


Fig. 12.8 Sphere of influence.

The ratio Φ_1/a_1 determines the magnitude of the deviation from the Keplerian orbit of point P.

With respect to A_2, the equation of motion becomes

$$\frac{d^2\boldsymbol{r}}{dt^2} = \boldsymbol{a}_2 + \boldsymbol{\Phi}_2 \tag{12.10}$$

where

$$\boldsymbol{a}_2 = \frac{-G(m_2 + m)}{r^3}\boldsymbol{r} \tag{12.11}$$

$$\boldsymbol{\Phi}_2 = Gm_1\left(\frac{\boldsymbol{\rho} - \boldsymbol{r}}{r_{12}^3} - \frac{\boldsymbol{\rho}}{\rho^3}\right) \tag{12.12}$$

The sphere of influence with respect to A_1 is defined by the condition

$$\frac{\Phi_1}{a_1} \leq \frac{\Phi_2}{a_2} \tag{12.13}$$

The boundary of this region is defined by the equality sign in Eq. (12.13). Then,

$$(m_1 + m)m_1 r^2 \left|\frac{\boldsymbol{r}_{12}}{r_{12}^3} - \frac{\boldsymbol{\rho}}{\rho^3}\right| = (m_2 + m)m_2\rho^2\left|\frac{\boldsymbol{r}_{12}}{r_{12}^3} - \frac{\boldsymbol{r}}{r^3}\right| \tag{12.14}$$

which, for the case of $m_1 \ll m_2$, approximately yields

$$\left(\frac{\rho}{r_{12}}\right)^{10}(1 + 3\cos^2\theta) = \left(\frac{m_1}{m_2}\right)^4 \tag{12.15}$$

where θ is as defined in Fig. 12.8. Solving for ρ,

$$\rho = r_{12}\left(\frac{m_1}{m_2}\right)^{2/5} \Big/ \sqrt[10]{1 + 3\cos^2\theta} \tag{12.16}$$

Equation (12.16) represents a slightly prolate spheroid that differs little from a sphere when m_1/m_2 is small. Then, the radius of the sphere of influence becomes

$$R_{\text{SOI}} = r_{12}\left(\frac{m_1}{m_2}\right)^{2/5} \tag{12.17}$$

Thus, for the moon,

$$\begin{aligned} R_{\text{SOI}} &= 384{,}000\left(\frac{1}{81}\right)^{2/5} \\ &\approx 66{,}280 \text{ km} \\ &> R_s = 43{,}000 \text{ km} \end{aligned}$$

Velocity at "Infinity"

Velocity at "infinity" V_∞ is the velocity of the vehicle at the sphere of influence of the planet with respect to the coordinates of the planet.

The geometry of the escape hyperbola from Earth is illustrated in Fig. 12.9, where the burnout velocity v_{bo} at the radius r_{bo} and phase angle η required to achieve v_∞ is also known. The value of v_{bo} is found from the fact that the energy ε is constant along the geocentric escape hyperbola and, therefore,

$$\varepsilon = \frac{v_{bo}^2}{2} - \frac{\mu}{r_{bo}}$$

$$= \frac{v_\infty^2}{2} - \frac{\mu}{r_\infty} \approx v_\infty^2/2$$

which yields

$$v_{bo}^2 = v_\infty^2 + 2\mu/r_{bo} \tag{12.18}$$

where r_∞ is the radius of the sphere of the influence of the Earth, and μ is the gravitational constant. The phase angle η is found from

$$\cos\eta = \frac{-1}{e} \tag{12.19}$$

where the eccentricity e of the departure hyperbola is given by

$$e = \sqrt{1 + 2\varepsilon h^2/\mu^2} \tag{12.20}$$

$$h = r_{bo} v_{bo}$$

$$= \text{specific angular momentum} \tag{12.21}$$

for injection at perigee.

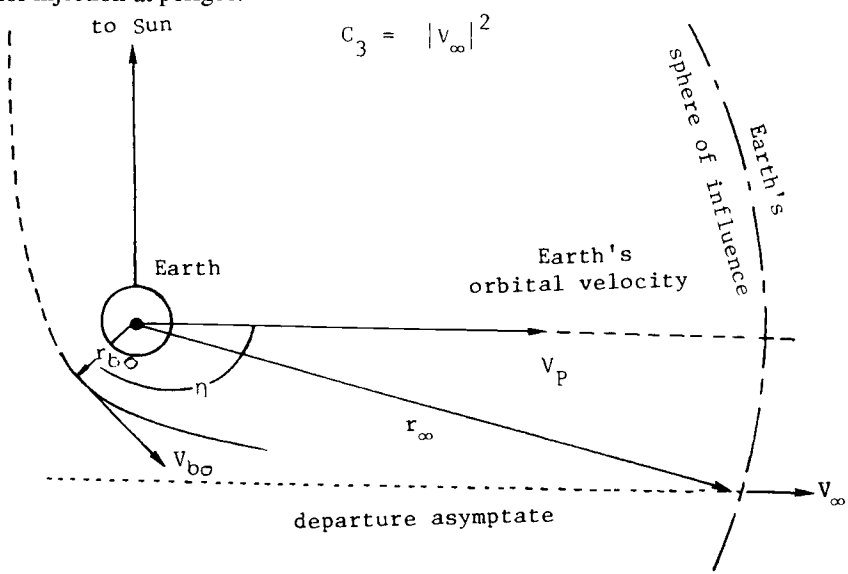


Fig. 12.9 Escape hyperbola (from Ref. 6).

Equation (12.18) may be solved for the velocity at infinity. Thus,

$$v_\infty = \sqrt{v_{bo}^2 - \frac{2\mu}{r_{bo}}} \tag{12.22}$$

The v_∞ velocity vector may then be added to the velocity of the planet in order to obtain the velocity of the vehicle with respect to the sun as in Fig. 12.10, for example.

The characteristic energy $C_3 = v_\infty^2$. It is used to define the energy requirements for departure from the planet's sphere of influence.

Gravity Assist

An important consequence of a spacecraft entering a sphere of influence of a planet (secondary) is the possibility of gaining or losing energy with respect to the sun (primary). The gain or loss of energy is caused by the turning (rotation) of the spacecraft velocity vector under the influence of the secondary's gravitational field. Thus, if the gravitational field of the secondary turns the relative velocity vector of the spacecraft toward that of the secondary, the resultant vector velocity with respect to the primary is greater than that at the entry to the sphere of influence of the secondary. This effect is illustrated in Fig. 12.11, where the notation v_∞ denotes the spacecraft velocity at "infinity" at entry or exit from the sphere of influence of the secondary.

An example of the Mars flyby and gravity-assist vector geometry is shown in Fig. 12.12, where $v_{\infty_{in}}$ and $v_{\infty_{out}}$ are the incoming and outgoing velocity vectors with respect to Mars at the radius of its sphere of influence. The velocity of the satellite with respect to the sun is thus shown to be increased after gravity assist by the equivalent vector ΔV.

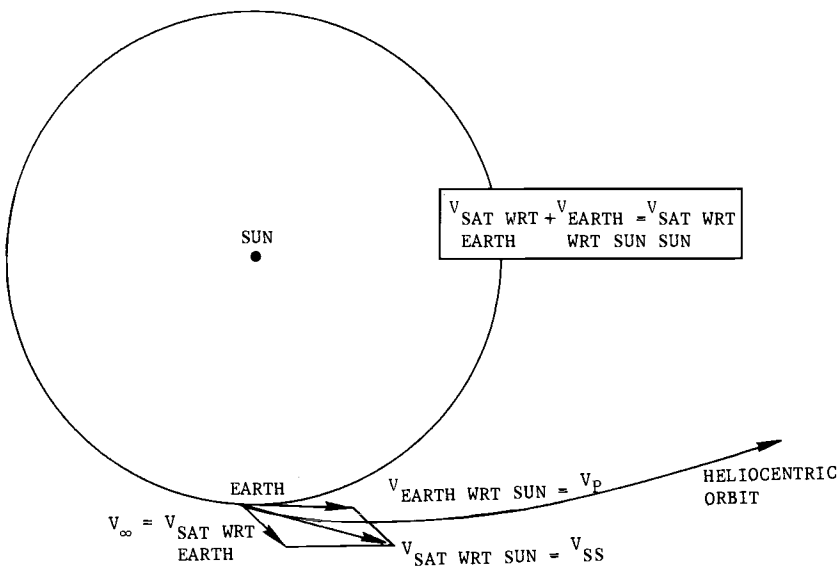


Fig. 12.10 Heliocentric velocity of a vehicle.

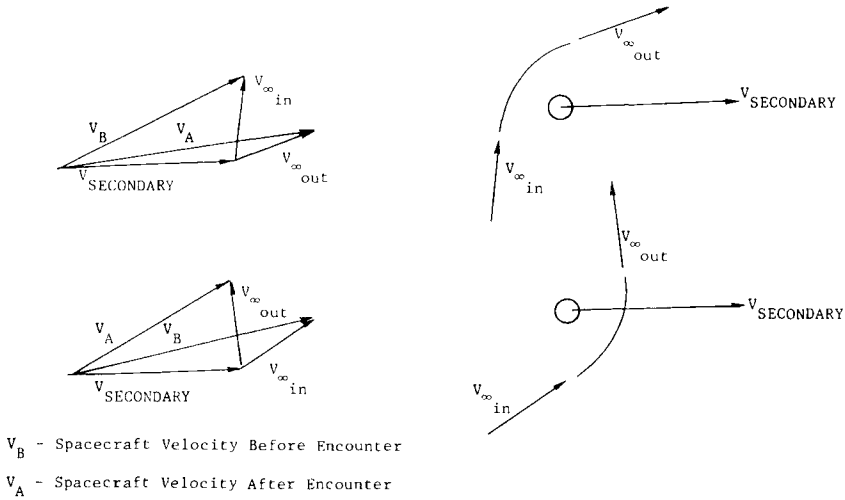


Fig. 12.11 Gravity-assist concept: a) energy increasing; b) energy decreasing V_B = spacecraft velocity before encounter; V_A = spacecraft velocity after encounter.

12.4 Lunar Trajectories

The Earth-Moon System

The motion of the Earth-moon binary system is, in general, complex. The Earth and the moon revolve about their common center of mass, which is 4728 km from the center of the Earth. The mean distance between the Earth and the moon is 384,400 km, and the mass of the moon is 1/81.30 that of the Earth. The period of revolution about the mass center is 27.3 days.

The orbital period of the moon is slowly getting longer because the distance between the Earth and the moon is increasing. This is the result of the tidal bulge in the Earth's oceans, which are carried eastward by the Earth's rotation.

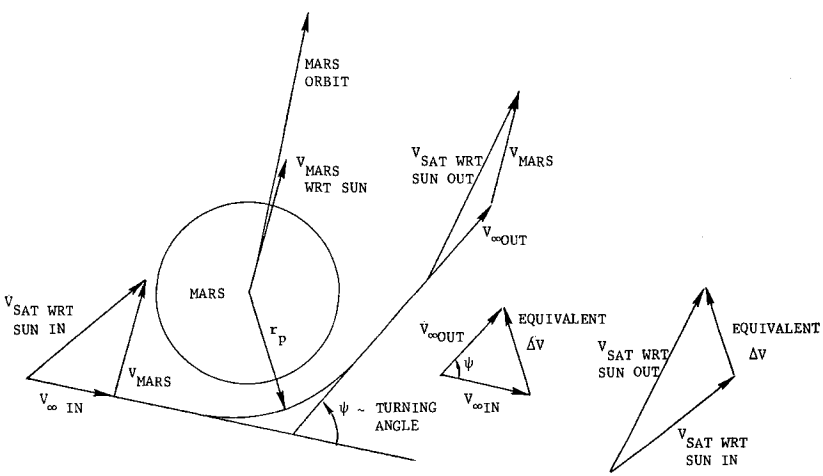


Fig. 12.12 Mars flyby and gravity assist.

Consequently, a shift of the Earth's center of mass relative to the Earth-moon line gives the moon a small acceleration in the direction of its orbital motion. The moon speeds up and moves further away from the Earth. The loss of energy in this process is known as the tidal friction effect.

The mean eccentricity of the moon's orbit is 0.0549. It is inclined to the ecliptic (the plane of Earth's orbit) by about 5°8′. The line of nodes, or the intersection of the orbit planes, moves westward making one complete revolution in 18.6 yr. The angle between the Earth's equatorial plane and the moon's orbital plane varies between 18°19′ and 28°35′ within a period of 18.6 yr. The line of apsides rotates in the direction of the moon's orbital motion within a period of 8.9 yr.

12.5 Analytical Approximations

Two-Body Motion

Several analytical models may be employed for preliminary estimates of the velocity requirements for lunar trajectories. The simplest model assumes the mass of the moon to be negligibly small. In this model, the lunar trajectory is described by an unperturbed two-body conic section relative to the Earth. The inclination of the moon's orbital plane relative to the equator and nonequatorial launch sites can also be included in this model.

The velocity requirements for transfer between the Earth and the moon can be obtained approximately with the use of the vis-viva equation in the form

$$v = \sqrt{\mu\left(\frac{2}{r_{bo}} - \frac{1}{a}\right)} \tag{12.23}$$

where r_{bo} is the radius at burnout and a is the semimajor axis of the lunar trajectory. The flight time can be computed as half of the period of the transfer ellipse or, more accurately, by the use of the Kepler's equation. Typical injection velocities as a function of injection altitude are shown in Fig. 12.13 for 50-, 66-, and 90-h transfer times. The lower and upper values for the shaded regions represent perigee and apogee lunar radii, respectively. These velocities are impulsive inertial (absolute) velocities and may be partly composed of a component of velocity due to Earth's rotation. The range of permissible liftoff times (launch windows) on a given date is determined by the launch azimuth, park orbit coast duration, and liftoff time. Typically, the launch windows may vary from 2 to 5 h, depending on the lunar declination.

Patched-Conic Approach

The patched-conic approach makes use of the spheres of influence of the Earth and the moon. Very near the Earth, the effect of the moon is small and may be neglected regarding the satellite to be on an unperturbed orbit about the Earth. Near the moon, the opposite is assumed, and the trajectory is regarded as an unperturbed orbit about the moon. The "patching" of the trajectories or conics is at the sphere of influence of the moon. This is illustrated in Fig. 12.14, where the change of the reference frames results in a selenocentric (lunar) hyperbola since the velocity of the satellite at the lunar sphere of influence is finite with respect to the moon.

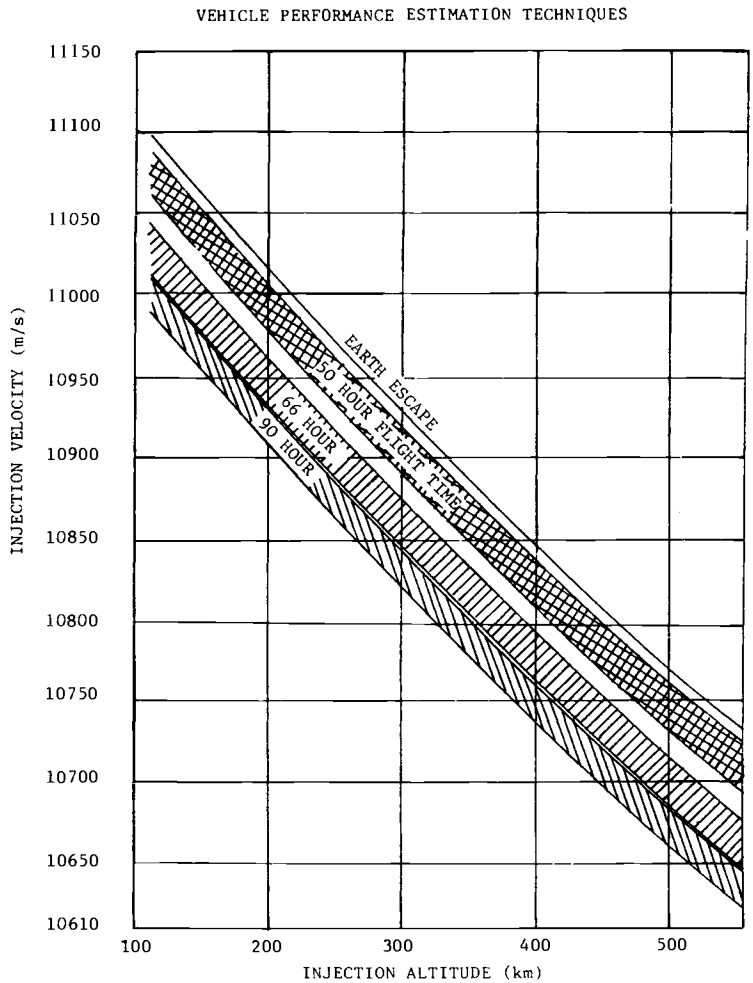


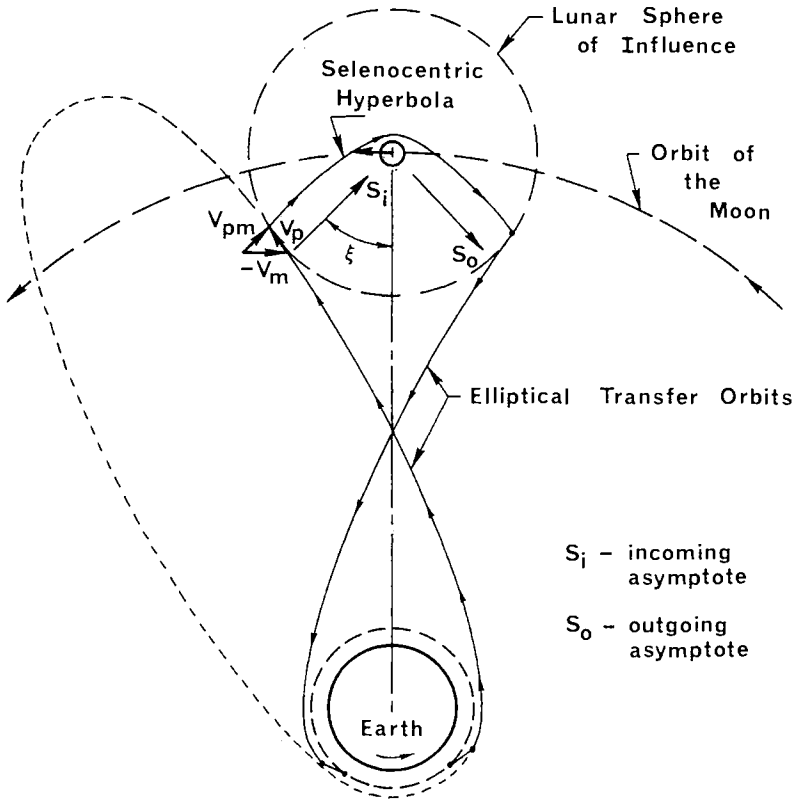
Fig. 12.13 Lunar mission injection velocity vs altitude (from Ref. 9).

Patched-Conic Example Trajectory

The basic patched-conic lunar transfer problem can be illustrated using the approach of Ref. 6. Specifically, it may be desired to arrive at perilune (perigee) at a specified altitude. This can be achieved by selecting the translunar injection (TLI) velocity vector components r_0, v_0, ϕ_0 in Fig. 12.15 such that the desired perilune radius is obtained by iterating on the phase angle λ_1 at arrival to the moon's sphere of influence. An algorithm required to solve this problem and determine the time of flight may be constructed as follows:

With reference to Fig. 12.15, select r_0 (km), v_0 (km/s), ϕ_0 (deg), λ_1 (deg). Then,

$$E_e = \frac{V_0^2}{2} - \frac{\mu_e}{r_0} \tag{12.24}$$



(not to scale)

Fig. 12.14 Lunar trajectory patched-conic concept.

where

$$\mu_e = 3.986 \times 10^5 \text{ km}^3/\text{s}^2 \tag{12.25}$$

$$h = r_0 v_0 \cos \phi_0 \tag{12.26}$$

$$r_1 = \sqrt{D^2 + R_s^2 - 2DR_s \cos \lambda_1}$$

where

$$D = 384400 \text{ km}$$

$$R_s = 66300 \text{ km} \tag{12.27}$$

$$v_1 = \sqrt{2\left(E_e + \frac{\mu_e}{r_1}\right)}$$

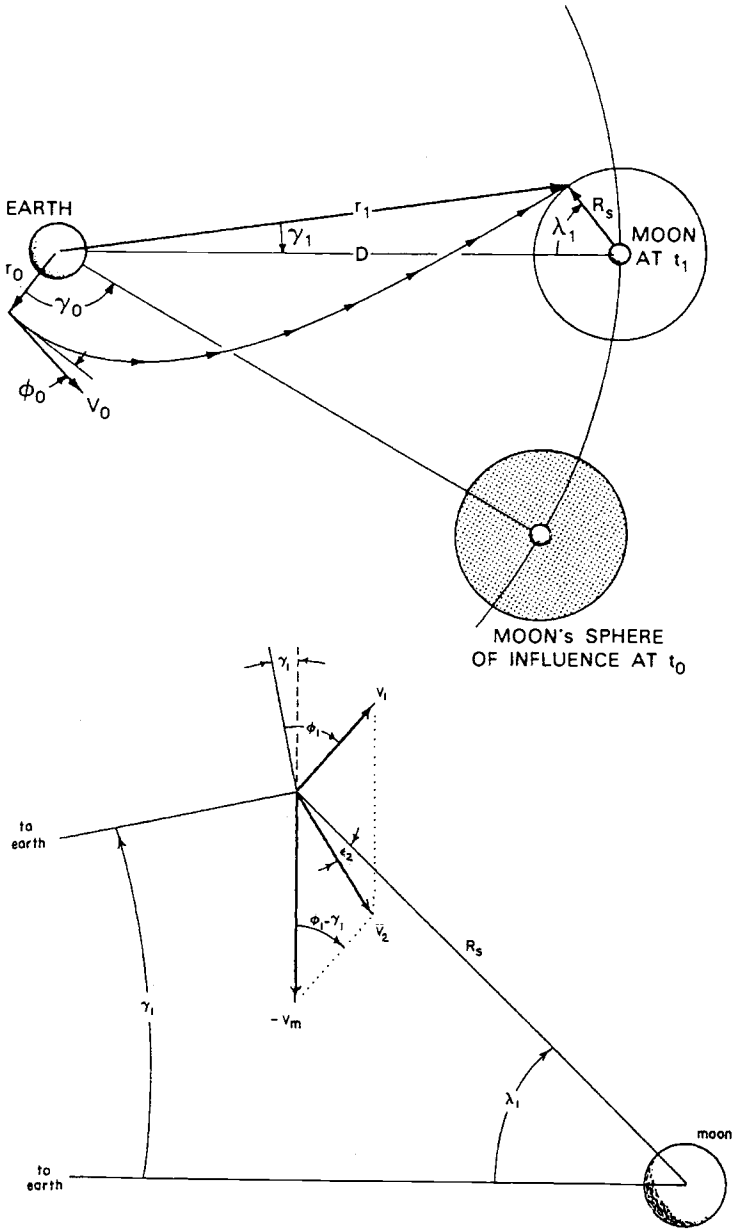


Fig. 12.15 Earth-moon patched-conic geometry (from Ref. 6).

$$\phi_1 = \cos^{-1}\left(\frac{h}{r_1 v_1}\right) \tag{12.28}$$

$$\gamma_1 = \sin^{-1}\left(\frac{R_s}{r_1}\sin\lambda_1\right) \tag{12.29}$$

For the velocity of the moon,

$$v_m = 1.018 \text{ km/s}$$

$$v_2 = \sqrt{v_1^2 + v_m^2 - 2v_1 v_m \cos(\phi_1 - \gamma_1)} \tag{12.30}$$

$$E_m = \frac{v_2^2}{2} - \frac{\mu_m}{R_s} \tag{12.31}$$

where $\mu_m = 4.093 \times 10^3 \text{ km}^3/\text{s}^2$

$$h_m = R_s v_2 \sin\varepsilon_2 \tag{12.32}$$

where

$$\varepsilon_2 = \sin^{-1}\left[\frac{v_m}{v_2}\cos\lambda_1 - \frac{v_1}{v_2}\cos(\lambda_1 - \gamma_1 - \phi_1)\right] \tag{12.33}$$

$$P_m = \frac{h_m^2}{\mu_m} \tag{12.34}$$

$$e_m = \sqrt{1 + 2\frac{E_m h_m^2}{\mu_m^2}} \tag{12.35}$$

$$r_p = \frac{P_m}{1 + e_m} \tag{12.36}$$

$$v_p = \sqrt{2\left(E_m + \frac{\mu_m}{r_p}\right)} \tag{12.37}$$

Here,

$$p = \frac{h^2}{\mu_e} \tag{12.38}$$

$$a = \frac{-\mu_e}{2E_e} \tag{12.39}$$

$$e = \sqrt{1 - \frac{p}{a}} \tag{12.40}$$

For true anomalies, θ_0, θ_1,

$$\cos\theta_0 = \frac{p - r_0}{r_0 e} \rightarrow \theta_0 \tag{12.41}$$

$$\cos\theta_1 = \frac{p - r_1}{r_1 e} \rightarrow \theta_1 \tag{12.42}$$

$$E_0 = \cos^{-1}\left(\frac{e + \cos\theta_0}{1 + e\cos\theta_0}\right) \tag{12.43}$$

$$E_1 = \cos^{-1}\left(\frac{e + \cos\theta_1}{1 + e\cos\theta_1}\right) \tag{12.44}$$

Flight time Δt is given by

$$\Delta t = t_1 - t_0 = \sqrt{\frac{a^3}{\mu_e}}[(E_1 - e\sin E_1) - (E_0 - e\sin E_0)] \tag{12.45}$$

Phase angle at departure is found from

$$\gamma_0 = \theta_1 - \theta_0 - \gamma_1 - \omega_m \Delta t \tag{12.46}$$

where $\omega_m = 2.649 \times 10^{-6}$ rad/s.

Typical flight time as a function of injection velocity at 320 km altitude for the phase angle at arrival $\lambda_1 = 65$ deg is shown in Fig. 12.16. The corresponding perigee radius and velocity at the moon (perilune) are shown in Figs. 12.17 and 12.18, respectively.

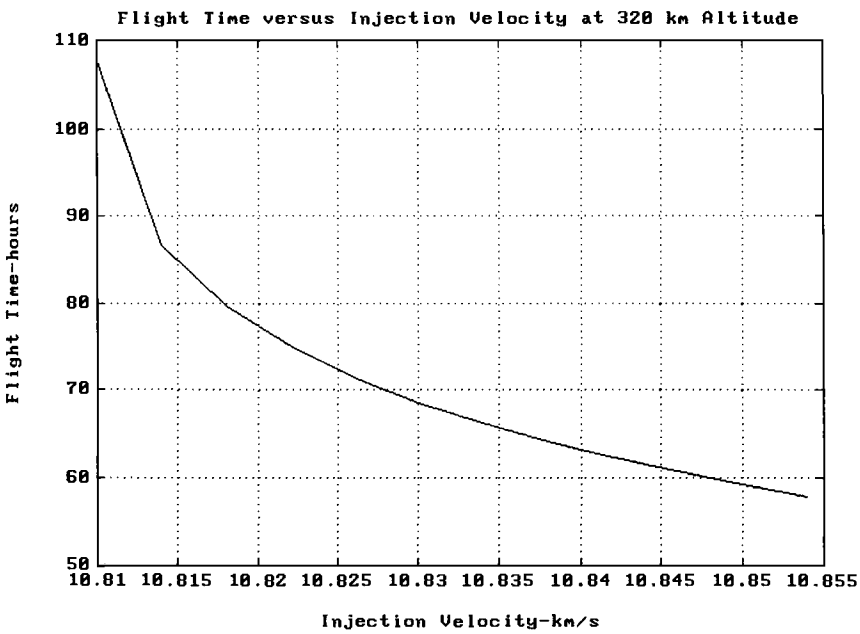


Fig. 12.16 Flight time vs injection velocity at R_0 = 320 km.

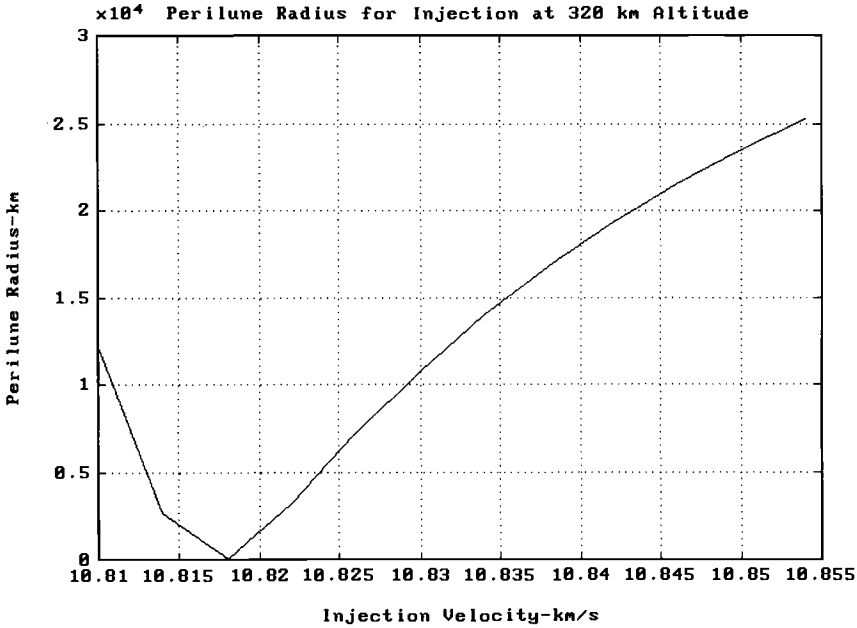


Fig. 12.17 Perilune radius for injection at 320 km altitude.

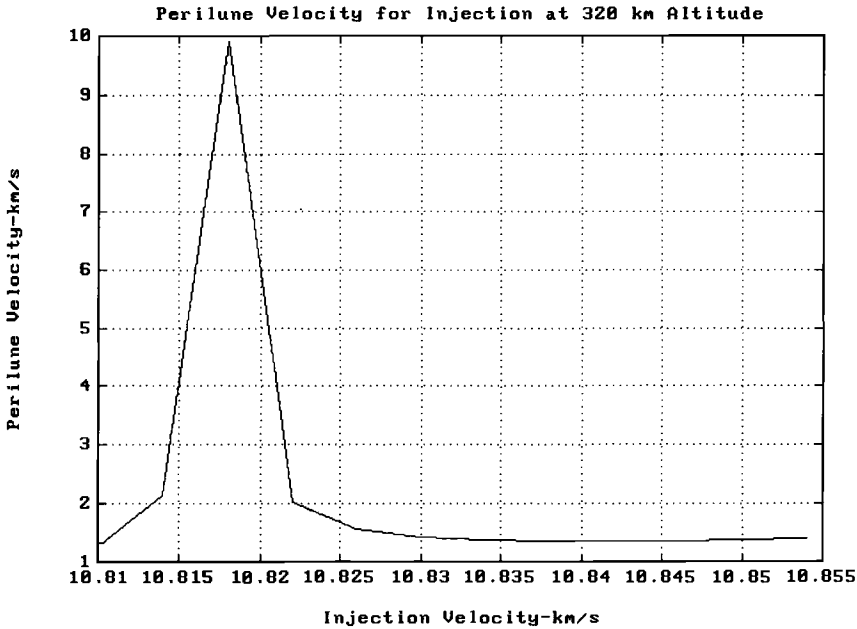


Fig. 12.18 Perilune velocity for injection at 320 km altitude.

12.6 Three-Dimensional Trajectories

The three-dimensional trajectories require the specification of several parameters in addition to those for the two-dimensional case discussed previously. The additional parameters are the initial latitude β, the initial launch azimuth ψ, the longitude difference between the initial point and the moon at impact $\Delta\lambda$, the instantaneous declination of the moon δ, and its maximum value δ_m. The relationships among these parameters determine the launch window at any launch site and the conditions at arrival at the moon. Reference [10] shows, for example, that these families of Earth-moon trajectories departing in either co- or counter-direction with Earth rotation have a common vertex on the far side of the moon (see Fig. 12.19). The corotational direction takes advantage of Earth tangential velocity at the launch site. An example of a Ranger-type (lunar impact) trajectory is illustrated in Fig. 12.20.

The targeting parameter $\boldsymbol{B}$ is defined as a vector originating at the center of the target body (e.g., moon) and directed perpendicular to the incoming asymptote of the target-centered approach hyperbola. The targeting parameter $\boldsymbol{B}$ is resolved into two components that lie in a plane normal to the incoming asymptote $\boldsymbol{Si}$. The orientation of the reference axes in this plane is arbitrary but is usually selected to lie in a fixed plane. For interplanetary trajectories, the unit vector $\hat{T}$ is in the ecliptic plane, and $\hat{R}$ is normal to it, as shown in Fig. 12.21.

12.7 Interplanetary Trajectories

Types of Transfers

Feasibility-type trajectories that are used for preliminary vehicle performance studies involve a simple two-body problem. The classical Hohmann trajectories, for example, are ellipses that are tangent to both the launch and the arrival orbit. The energy for transfer from the launch orbit to the target orbit is, in most cases, a minimum for this trajectory, but the transfer time is usually quite long. The time can be found from Kepler's third law.

There are three main groups of trajectories that can be evolved from the Hohmann transfer: 1) staying tangential to the larger orbit but intersecting the smaller one, 2) intersecting the larger orbit and staying tangential to the smaller one, and 3) intersecting both orbits. Parabolic or hyperbolic transfer trajectories intersect the larger orbit and may, but do not need to be, tangential to the smaller orbit. The flight times for parabolic or hyperbolic trajectories are generally shorter than those for Hohmann transfers. Examples of several types of Earth-Mars trajectories are illustrated in Fig. 12.22.

Mars Landing Mission Example

Taking the approach of Ref. 6, we will examine the example of a spacecraft lander (probe) on Mars. For this case, the spacecraft is subject to the gravitational actions of Earth, sun, and Mars, as shown in Fig. 12.23.

The probe of mass m_p is subject to the gravitational forces F_s, F_e, F_m of sun, earth, and Mars, respectively; thus, the equation of motion is

$$m_p\ddot{\boldsymbol{r}}_p = \boldsymbol{F}_s + \boldsymbol{F}_e + \boldsymbol{F}_m \tag{12.47}$$

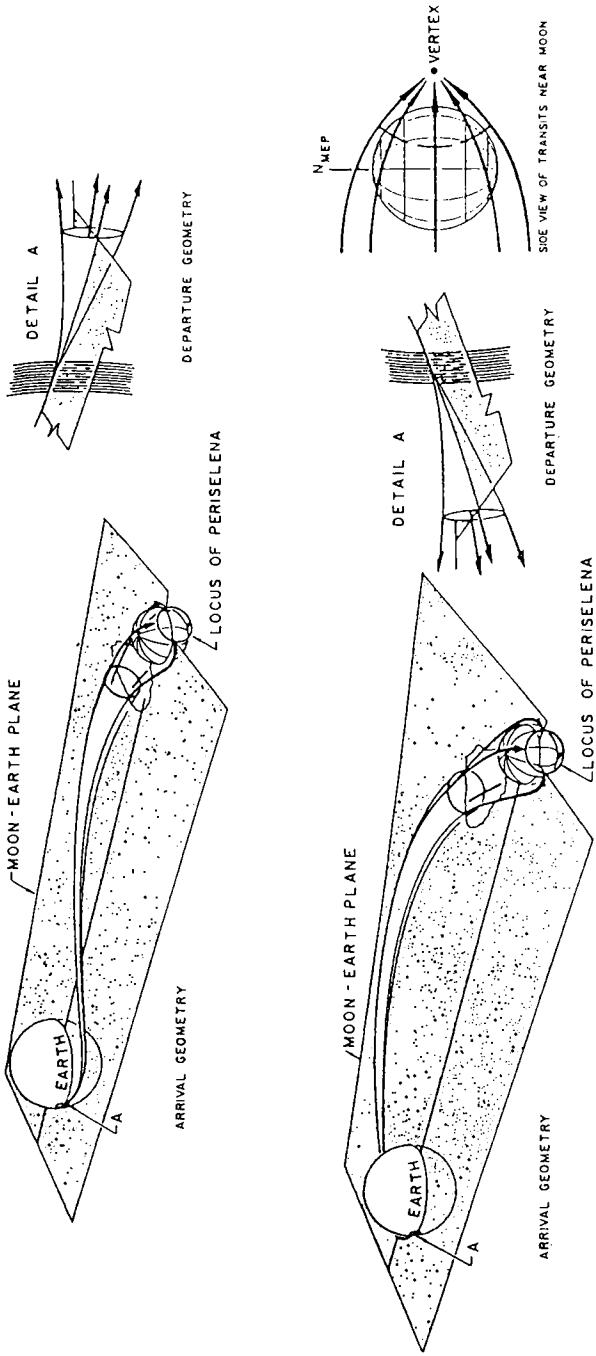


Fig. 12.19 Family of transits departing from moon-Earth plane equatorial perigees in counter-rotational directions (from Ref. 10).

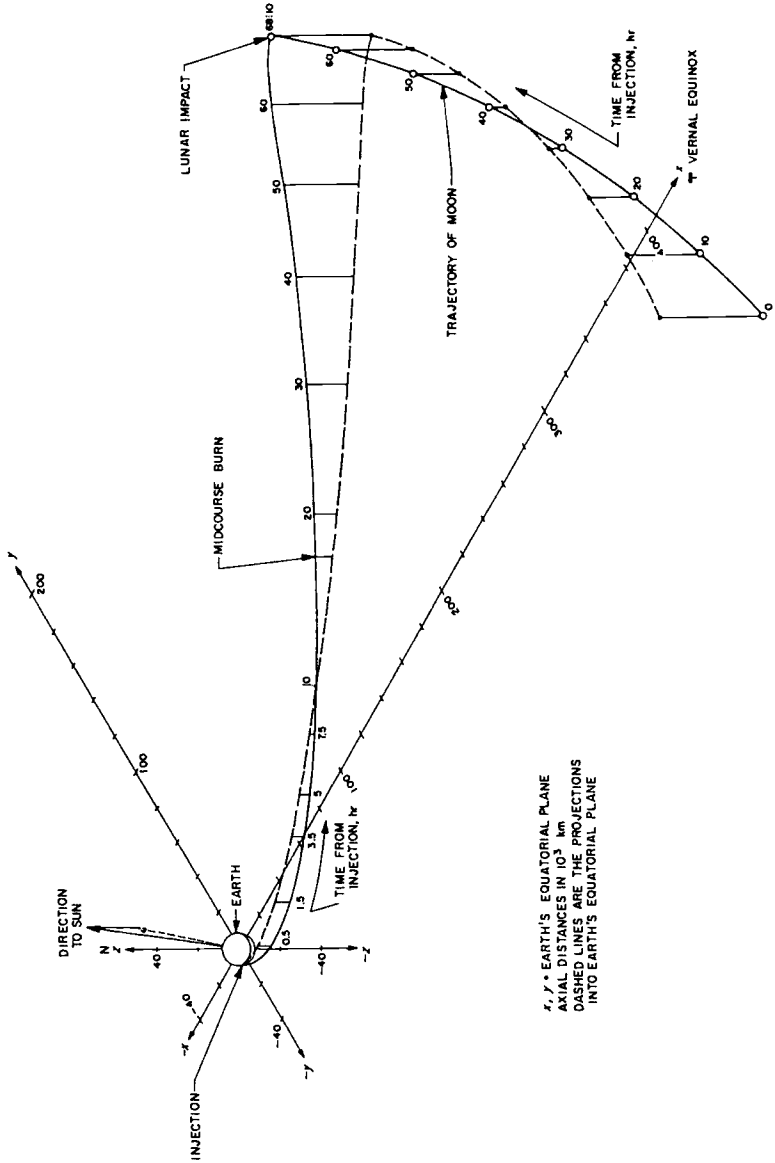


Fig. 12.20 Typical geocentric spatial trajectory trace (Ranger 11).

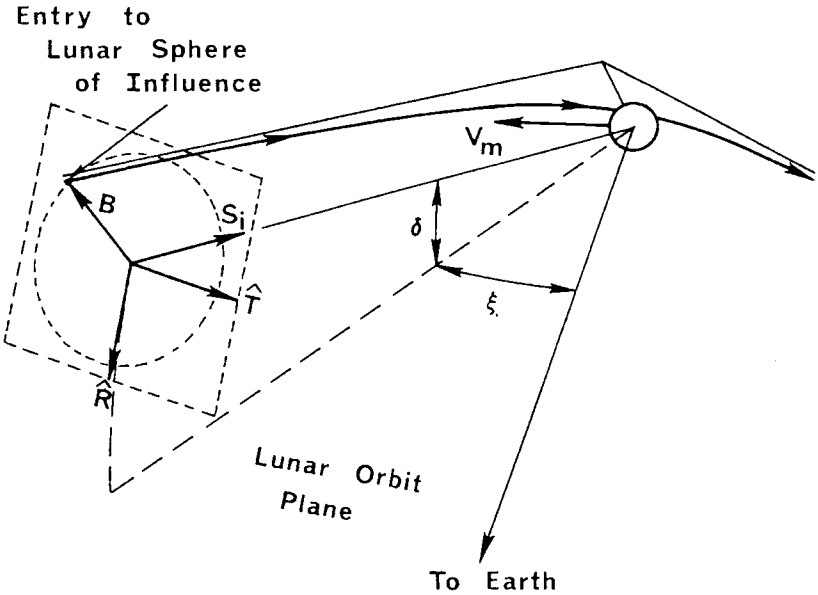


Fig. 12.21 Lunar targeting parameters.

or

$$m_p \ddot{\boldsymbol{r}}_p = \frac{-\mu_s}{r_p^3}\boldsymbol{r}_p - \frac{\mu_e(\boldsymbol{r}_p - \boldsymbol{r}_e)}{|\boldsymbol{r}_p - \boldsymbol{r}_e|^3} - \frac{\mu_m(\boldsymbol{r}_p - \boldsymbol{r}_m)}{|\boldsymbol{r}_p - \boldsymbol{r}_m|^3} \tag{12.48}$$

Here s, e, and m refer to the sun, Earth, and Mars, respectively.

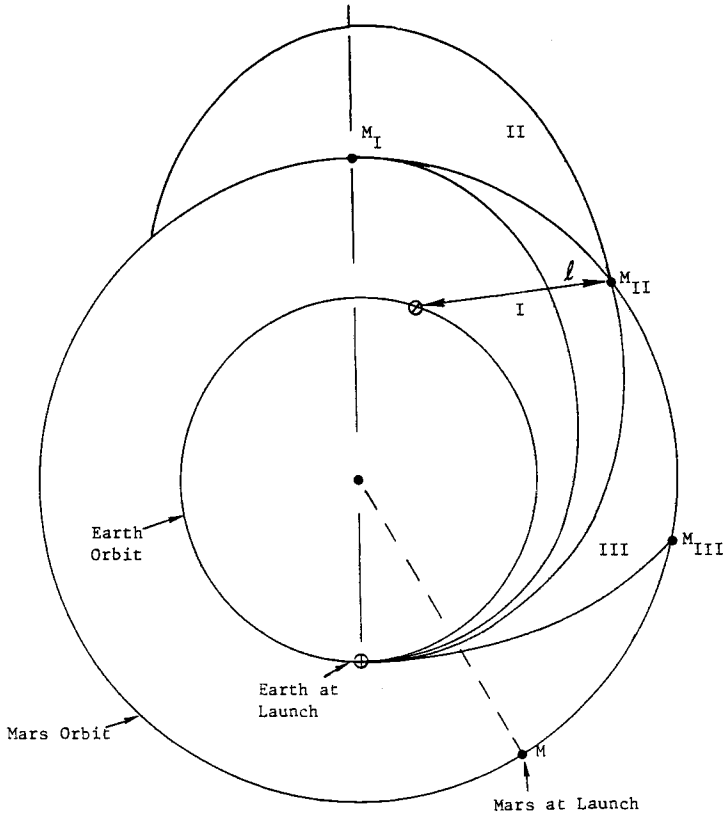
This equation is not integrable in closed form and must be evaluated numerically. Approximate solutions for the trajectory can be obtained using a patched-conic approach. In this approach, the trajectory is divided into three different phases: 1) heliocentric, 2) Earth departure, and 3) Mars arrival. Each phase is a two-body Keplerian orbit, the conic section of which is "patched" with the following phase. Approximate velocity requirements for the mission can in this way be estimated to determine the feasibility of the mission.

Heliocentric phase. Assuming a Hohmann-transfer heliocentric trajectory from Earth to Mars, as illustrated in Fig. 12.24, the required perigee velocity with respect to the sun is given by the equation

$$\begin{aligned} v_{\text{pt}} &= \sqrt{\mu_s\left(\frac{2}{r_e} - \frac{1}{a_t}\right)} \\ &= 32.74 \text{ km/s} \end{aligned} \tag{12.49}$$

where

$$\begin{aligned} r_e &= 1 \text{ A.U.} = 1.49597893 \times 10^8 \text{ km} \\ a_t &= \frac{r_e + r_m}{2} = \frac{2.523}{2} \\ &= 1.262 \text{ A.U.} \\ &= 1.887 \times 10^8 \text{ km} \\ \mu_s &= 1.32712499 \times 10^{11} \text{ km}^3/\text{s}^2 \end{aligned}$$



Trajectory	V_{bo}, km/s	V_∞, km/s	Flight time, days	Arrival Earth-Mars distance l, 10^6 km
I Hohmann	11.6	2.94	259	236
II Elliptic	11.9	3.75	165	133
III Parabolic	16.7	12.33	70	79

V_{bo} = burnout velocity at 500 km (v_c = 7.62 km/s)
V_∞ = velocity at infinity WRT Earth

Fig. 12.22 Earth-Mars trajectory velocity requirements.

The transfer-orbit apogee velocity at Mars is

$$v_{at} = \sqrt{\mu_s\left(\frac{2}{r_m} - \frac{1}{a_t}\right)} \tag{12.50}$$

$$= 21.45 \text{ km/s}$$

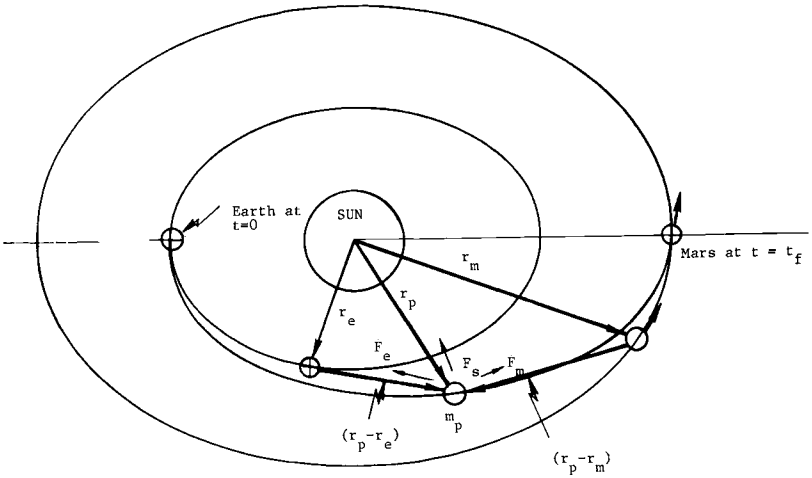


Fig. 12.23 Mars mission trajectory.

The velocity at exit from the Earth sphere of influence is

$$\begin{aligned} v_\infty &= v_{pt} - v_e \\ &= 32.74 - 29.78 \\ &= 2.96 \text{ km/s} \end{aligned} \tag{12.51}$$

Therefore, $C_3 = v_\infty^2 = 8.76$ km^2/s^2, and the energy $E_e = C_3/2 = 4.38$ km^2/s^2.

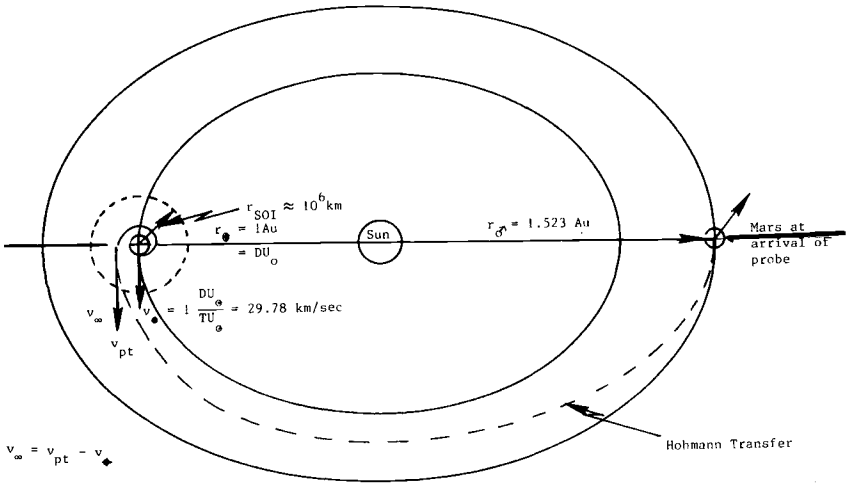


Fig. 12.24 Heliocentric Mars trajectory: $r_{SOI} \approx \ell\ (m_1/m_\odot)^{2/5}$; $\ell = a$ for planet; m_1 = mass of planet; $m_\odot$ mass of sun.

Earth departure phase. Assuming an injection burn at an altitude of 300 km, the burnout velocity required is

$$v_{\text{bo}} = \sqrt{2\left[\frac{\mu_e}{(R_e + h)} + E_e\right]} \tag{12.52}$$
$$= 11.32 \text{ km/s}$$

where

$R_e = 6378$ km
$\quad$ = equatorial radius of Earth
$h = 300$ km
$\quad$ = altitude at injection
$\mu_e = 3.9860064 \times 10^5 \text{ km}^3/\text{s}^2$
$\quad$ = gravitational parameter of earth
$E_e = 4.38 \text{ km}^2/\text{s}^2$
$\quad$ = energy of escape hyperbola

Mars arrival phase. For a soft landing at Mars, the energy of the hyperbolic orbit at Mars is

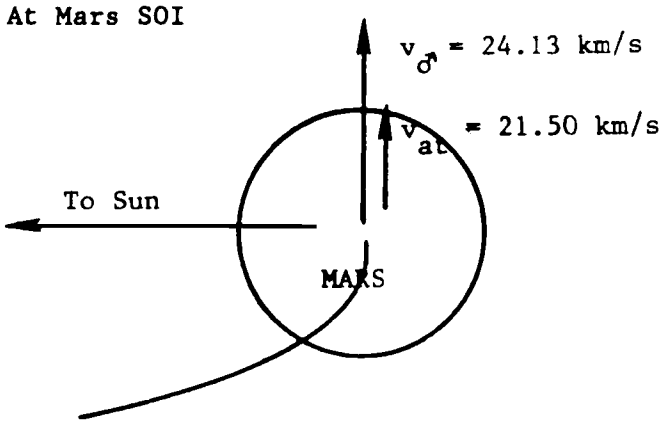
$$\begin{aligned} E_m &= \frac{v_\infty^2}{2} \\ &= \frac{(v_m - v_{at})^2}{2} \\ &= \frac{(24.13 - 21.45)^2}{2} \\ &= 3.59 \text{ km}^2/\text{s}^2 \end{aligned} \tag{12.53}$$

where $v_m = 24.13$ km/s = velocity of Mars. Retro velocity at Mars surface is

$$v_{\text{retro}} = \sqrt{2\left(\frac{\mu_m}{R_m} + E_m\right)}$$
$$= 5.71 \text{ km/s}$$

where

$\mu_m = 43058 \text{ km}^3/\text{s}^2$
$\quad$ = gravitational parameter of Mars
$R_m = 3379$ km
$\quad$ = equatorial radius of Mars
$E_m = 3.59 \text{ km}^2/\text{s}^2$



Total velocity requirements for the mission are

$$\begin{aligned} \text{Burnout velocity at Earth } v_{bo} &= 11.32 \text{ km/s} \\ \text{Retro velocity at Mars } v_{retro} &= \underline{5.71} \text{ km/s} \\ \text{Total} \quad & \quad 17.03 \text{ km/s} \end{aligned}$$

An equal amount of velocity impulse would be required for the return trip to Earth. Small thrust application may also be necessary for midcourse correction of the heliocentric trajectory, and the probe would retrofire first to deboost to a parking orbit before landing. The velocities that were computed are feasibility-type numbers. Precision velocities would be somewhat different as they must be obtained by integration of the equations of motion.

12.8 Galileo Mission

During its six-year journey to Jupiter, the Galileo Orbiter examined the planets Venus and Earth as well as the Moon and made several unique observations. Following each Earth flyby, Galileo made excursions to the asteroid belt and studied two asteroids, Gaspra and Ida, close-up.* The Galileo scientists also observed Comet Shoemaker-Levy 9 fragment impacts on Jupiter.[11]

Objectives and Goals

The primary objectives of the Galileo Orbiter were to

1) investigate the circulation and dynamics of the Jovian atmosphere; the upper Jovian atmosphere and ionosphere; and the composition and distribution of surface minerals on the Galilean satellites;

*Data available online at http://nssdc.gsfc.nasa.gov [cited 2 March 2002].

2) characterize the morphology, the geology, and the physical state of the Galilean satellites;

3) determine the gravitational and magnetic fields and dynamic properties of the Galilean satellites; and

4) study the atmospheres, ionospheres, and extended gas clouds of the Galilean satellites and the interaction of the Jovian magnetosphere with the Galilean satellites.*

The science goals of the Galileo Probe were to "...determine the chemical composition of the Jovian atmosphere; characterize the structure of the atmosphere to a depth of at least 10 bars; investigate the nature of cloud particles and the location and structure of cloud layers; (and) examine the Jovian heat balance."* The Jovian lightning activity and the flux of its energetic charged particles down to the top of the atmosphere were to be determined also.

Results

Several discoveries have been made during the trajectory of the Galileo Orbiter including a satellite, Dactyl, of an asteroid, Ida; an intense interplanetary dust storm; and a new radiation belt, approximately 50,000 km above Jupiter's cloud tops. The existence of a huge ancient impact basin in the southern part of the Moon's far side was confirmed and evidence of more extensive lunar volcanism was found. Jovian winds in excess of 600 km/hr were detected but far less water was found in the Jovian atmosphere than was estimated from earlier Voyager observations. It was also determined that the individual lightning events on Jupiter are about 10 times stronger than on Earth but that the lightning activity is only about 10% of that found in an equal area on Earth. The helium abundance in Jupiter is nearly the same as its abundance in the sun, 24% compared to 25%. Extensive resurfacing of Io's surface due to continuing volcanic activity has been noted and evidence of intrinsic magnetic fields on Io and Ganymede have been seen. Evidence for liquid water ocean under Europa's surface has also been seen. The same evidence of water under Ganymede's ice-reach surface has been obtained by the Orbiter flybys.*

Additional Discoveries

Additional Jovian moons were also discovered.[12] With the use of the University of Hawaii's 2.2 m telescope on Mauna Kea, scientists discovered 11 new satellites of Jupiter. The new satellites are members of Jupiter's outer "irregular" satellite system with large eccentric orbits inclined to its equator. Orbital radii are about 300 times the radius of Jupiter and the inclinations are near 150 to 160 degrees. Nine of these objects and five previous known retrograde satellites of Jupiter constitute a total of 14 such objects. The origin of the retrograde irregular satellites appears to be a possible capture of objects from heliocentric orbit when Jupiter was young. Friction with Jupiter's atmosphere may have provided the energy required for capture of such satellites.†

*Data available online at http://nssdc.gsfc.nasa.gov [cited 2 March 2002].

†Data available online at http://www.ifa.hawaii.edu [cited 2 March 2002].

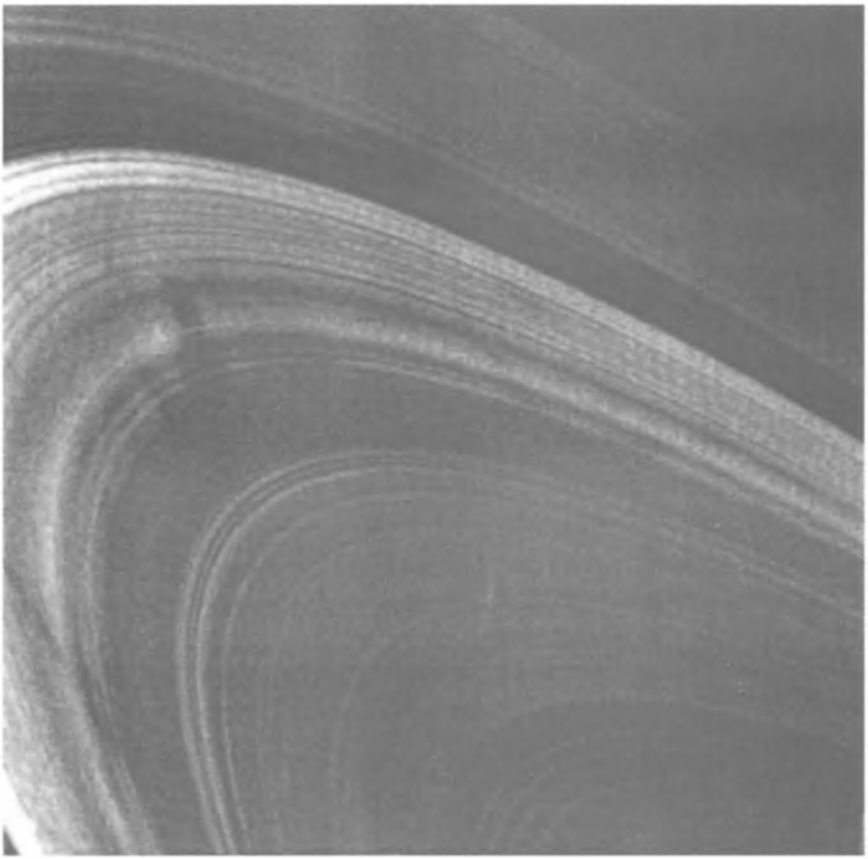


Fig. 12.25 High-resolution picture of Saturn taken 2.5 million miles away. (Courtesy of Jet Propulsion Laboratory, California Institute of Technology/NASA).

12.9 Cassini-Huygens Mission to Saturn and Titan

Saturn, the sixth planet from the sun, is the second largest planet in the Solar System; Jupiter is the largest. With a volume about 760 times that of Earth, Saturn has no solid surface but consists of a gas that compresses into fluid at great depths beneath the clouds. Its atmosphere is composed mostly of hydrogen and helium gas and its rings and icy satellites appear to be made primarily of water ice. A layer of ammonia ice clouds forms at a pressure level that is comparable to sea-level atmospheric pressure on Earth.*

Figure 12.25 is an example of one of the high-resolution pictures of Saturn's rings that Voyager 2 obtained. Numerous offshoots from the rings or "spoke" features can be seen in the middle of this image. Such spoke features "... persist at times for two or three rotations of the ring about the planet" and appear to "... revolve around the planet at the same rate as the rotation of the magnetic field

*Data available online at http://www.jpl.nasa.gov/cassini [cited 2 March 2002].

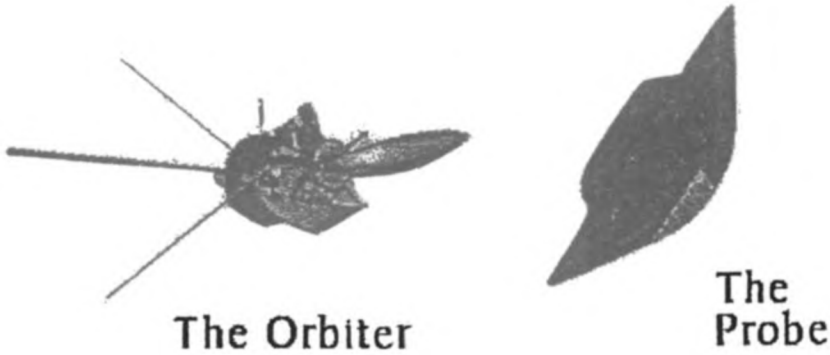


Fig. 12.26 The Cassini Spacecraft. (Courtesy of Jet Propulsion Laboratory, California Institute of Technology/NASA).

and the interior of Saturn, independent of their distance from the center of Saturn." * It is possible that the small dust particles that form the spokes are electrically charged, while older spokes, which may have lost their electrical charge, revolve with the underlying ring particles. Although it is not known exactly what causes these spokes electromagnetic forces could be responsible.*

Since 1997, the Cassini spacecraft consisting of the orbiter and the Huygens Probe (Fig. 12.26) has been traveling to the planet Saturn and its moon Titan. The mission is being managed by NASA's Jet Propulsion Laboratory, the designer and assembler of the Cassini Orbiter. The European Space Technology and Research Center in cooperation with 14 European nations managed the development of the Huygens Titan Probe. The vehicle used to launch the spacecraft was "... made up of a two-stage Titan IV booster rocket, two strap-on solid rocket motors, the Centaur upper stage and a payload enclosure..."† At launch, the spacecraft weighed about 5600 kg (12,346 lbs). Using 26 kg of Plutonium 238 as fuel, radioisotope thermoelectric generators supplied the spacecraft's electrical power. Two 445 N rocket engines provided maneuver capability when needed, and sixteen smaller thrusters were used to control the spacecraft's orientation.‡ The spacecraft must be built to withstand a thermal environment both inside the orbit of Venus (130°C) and at Saturn (−210°C). NASA's Deep Space Network was used for communicating with Cassini.

Cassini's planned interplanetary flight path, illustrated in Fig. 12.27, began with the launch from Earth on 15 October 1997. Next there were two gravity assist flybys of Venus on 26 April 1998 and 24 June 1999; one of Earth on 18 August 1999; and one of Jupiter on 30 December 2000. Saturn arrival is scheduled for 1 July 2004 when a four-year orbital tour of the Saturn system will begin. "With the use of the VVEJGA (Venus-Venus-Earth-Jupiter Gravity Assist) trajectory, it takes 6.7 years for the Cassini spacecraft to arrive at Saturn."*

At arrival, the cone-shaped Huygens Titan Probe will detach from the main spacecraft and conduct its own investigations. The Probe holds six of Cassini's

*Data available online at http://www.jpl.nasa.gov/cassini [cited 2 March 2002].

†Data available online at http://saturn.jpl.nasa.gov [cited 2 March 2002].

‡Data available online at http://www.jpl.nasa.gov [cited 2 March 2002].

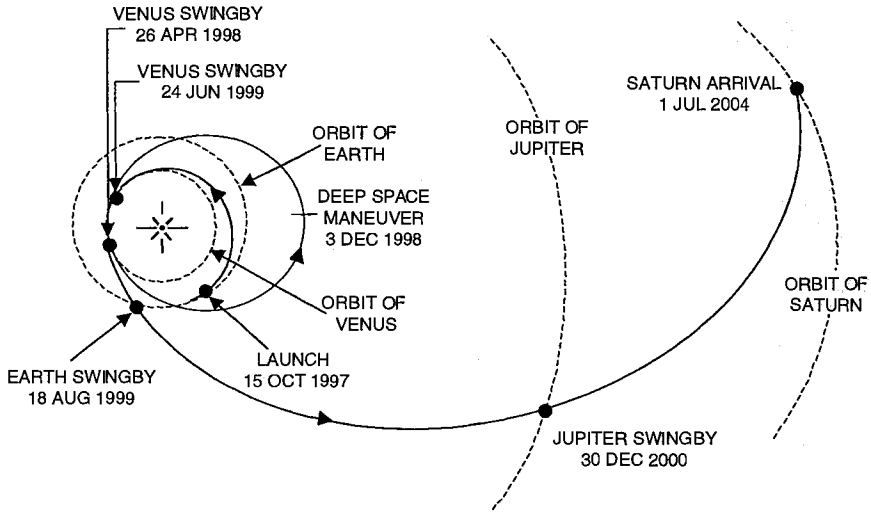


Fig. 12.27 Cassini Interplanetary Trajectory. (Courtesy of Jet Propulsion Laboratory, California Institute of Technology/NASA).

eighteen science instruments including "... a gas chemical analyzer designed to identify various atmospheric elements, a device to collect aerosols for chemical composition analysis, a camera that can take images and make wide-range spectral measurements, and an instrument whose sensors will measure the physical and electrical properties of Titan's atmosphere." *

12.10 Mars Odyssey Mission

Mars Odyssey orbiter lifted off from Kennedy Space Center aboard a Delta II rocket on 6 April 2001. The spacecraft carried scientific instruments to map the chemical and mineralogical composition of Mars and to assess potential radiation hazards to future human explorers. The orbiter reached Mars on 24 October 2001. For the next two and a half years, the orbiter will circle Mars looking for traces of water and evidence of life-sustaining environments.† The mission timeline is shown in Fig. 12.28.

The numbers on the figure are explained in further detail as follows:

1) Launch and Liftoff. Solar array deployed about 36 minutes after launch on 7 April 2001.

2) Cruise phase. Payload health and status check, calibration and data collection. Flight path correction. Duration about 200 days.

3) Mars orbit insertion via aerobraking and main engine firing. Duration about one week after Mars arrival on 24 October 2001.

*Data available online at http://www.jpl.nasa.gov/cassini [cited 2 March 2002].

†Data available online at http://www.mars.jpl.nasa.gov/ [cited 2 March 2002].

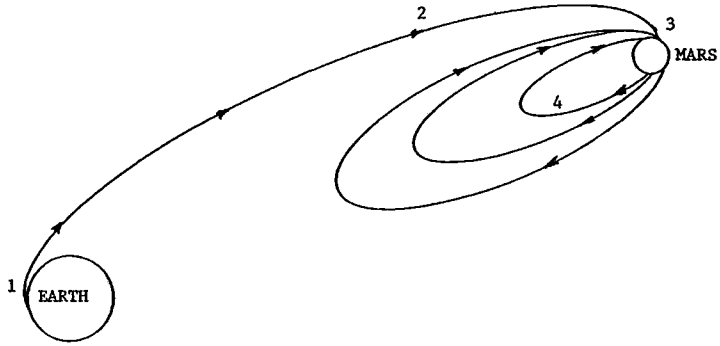


Fig. 12.28 Mars Odyssey Mission. (Courtesy of Jet Propulsion Laboratory, California Institute of Technology/NASA).

4) Mapping orbit. Thermal emission imaging of the Martian surface to determine its chemical and mineralogical composition. Communication support for landers and rovers in the future. Primary science mission duration is 917 days.

References

[1]Mariner-Venus 1962 Final Project Report, NASA SP-59, 1965.

[2]*Aviation Week and Space Technology*, Feb. 1985.

[3]Wood, L. J., and Jordan, J. F., "Interplanetary Navigation Through the Year 2005: The Inner Solar System," *Journal of the Astronomical Sciences*, Vol. 32, Oct.–Dec. 1984.

[4]*Aviation Week and Space Technology*, Dec. 1987.

[5]Balk, M. B., *Elements of Dynamics of Cosmic Flight, Nauka,* 1965 (in Russian).

[6]Bate, R. R., Mueller, D. D., and White, J. E., *Fundamentals of Astrodynamics,* Dover, New York, 1971.

[7]Clarke, T. C., and Fanale, F. P., "Galileo: The Earth Encounters," *Planetary Report,* Jet Propulsion Laboratory, CIT, Sept./Oct. 1989.

[8]"Pioneer to Jupiter," Bendix Field Engineering Corp., Palo Alto, CA, Nov. 1974.

[9]White, J. F., *Flight Performance Handbook for Powered Flight Operations,* Wiley, New York, 1963, pp. 2–54.

[10]Hoolker, R. F., and Braud, N. J., "Mapping the Course for the Moon Trip," *Astronautics & Aeronautics,* Feb. 1964.

[11]Johnson T. V., "The Galilean Mission to Jupiter and its Moons," *Scientific American*, Feb. 2000.

[12]Sheppard, S. S., Jewitt, D. C., Fernandez, Y. R., and Magnier, G., University of Hawaii Report in the International Astronomical Union Circular (IAUC 7555), 15 Feb. 2001.

Problems

12.1. Consider a nuclear waste disposal mission from Earth (1 A.U.) to a heliocentric circular orbit of radius 0.86 A.U.

a) What is the charcteristic energy C_3 required for a Hohmann transfer from Earth to the circular orbit or radius 0.86 A.U.?

b) What is the flight time via the Hohmann-transfer orbit from Earth to the 0.86 A.U. orbit about the sun?

12.2. To accomplish certain measurements of phenomena associated with sunspot activity, it is necessary to establish a heliocentric orbit with a perihelion of 0.85 A.U. The departure from the Earth's orbit will be at apohelion. What must the burnout velocity be at an altitude of 1300 km to accomplish this mission?

12.3. A Venus probe departs from a 6378-km altitude circular parking orbit with a burnout speed of 8.69 km/s. Find the hyperbolic excess speed at infinity.

12.4. Calculate the sphere of influence for the nine planets in the solar system.

12.5. Compute the distance of L_1, the Lagrangian liberation point, from the center of the moon along the Earth-moon line. To a first approximation, L_1 is an equilibrium point between the gravitational and centrifugal accelerations of the attracting bodies.

12.6. Compute the stationkeeping requirements (ΔV) to remain within 10 km of L_1 in Problem 12.5 for a year.

12.7. Compute the velocity impulse (ΔV) for transfer from a 100-km circular orbit at Mars to a hyperbolic orbit Earth return trajectory with an eccentricity $e = 1.5$.

12.8. Write the equation of motion for a solar sail in the solar system. Discuss the type of trajectories possible.

12.9. If the Earth were stopped in its orbit, what would be the elapsed time, in days, until collision with the sun? Assume point masses, and assume the Earth's orbit to be circular, with $r = 1.0$ A.U.

Selected Solutions

12.1. 1.28 (km/s)^2, 163.7 days

12.2. 10.26 km/s

12.3. 3.6 km/s

13
Space Debris

13.1 Introduction

Although the natural meteoroid environment has been considered in the designs of past and existing spacecraft, future satellite designs will have to take account of space debris in addition to the natural environment.

Man-made space debris differs from natural meteoroids because it is in permanent Earth orbit during its lifetime and is not transient through the regions of interest. As a consequence, a given mass of material presents a greater problem in the design and operation of spacecraft because of the extended time period over which there is risk of collision.

Past design practices and deliberate and inadvertent explosions in space have created a significant debris population in operationally important orbits. The debris consists of spent spacecraft and rocket stages, separation devices, and products of explosion. Much of this debris is resident at altitudes of considerable operational interest. Products larger than 10 cm^2 in low orbits can be observed directly. The existence of a substantially larger population of small fragments can be inferred from terrestrial tests in which the particle distributions from explosions have been assayed. From these tests it is reasonable to infer small particle numbers, of the order of 10,000 for each low-intensity explosion and several million for high-intensity explosions.

Two types of space debris are of concern: 1) large objects whose population, while small in absolute terms, is large relative to the population of similar masses in the natural flux (by a factor of about 1000); and 2) a large number of smaller objects whose size distribution approximates natural meteoroids. The interaction of these two classes of objects, combined with their long residual times in orbit, leads to the further concern that inevitably there will be collisions producing additional fragments and causing the total population to grow rapidly.

Some efforts to provide a definitive assessment of the orbiting debris problem have been and are being made by various government agencies and international organizations. Principal areas of concern are the hazards related to the tracked (cataloged), untracked, and future debris populations. Studies are being conducted in the areas of technology, space vehicle design, and operational procedures. Among these are ground- and space-based detection techniques, comprehensive models of Earth-space environment, spacecraft designs to limit accidental explosions, and different collision-hazard assessment methods. Occasional collision avoidance and orbit-transfer maneuvers are being implemented for selected satellites in geosynchronous orbits. The results and experience gained from the activities will, in time, create a better understanding of the problem and all its implications so that appropriate actions can be taken to maintain a relatively low-risk environment for future satellite systems.

A key aspect of the on-orbit debris hazard is that it is self-perpetuating. This arises from three factors: 1) A single spacecraft launch can be responsible for

a multitude of hazardous objects in space; 2) orbital debris tends to disperse randomly, producing high intersection velocities and making avoidance extremely difficult; and 3) objects accumulate in Earth orbit rather than passing through the near-Earth space in the manner of meteoroids. Impact protection may not be feasible in most cases because of the likelihood of very high approach velocities and the fact that certain protuberances, especially those of relatively large areas such as solar arrays and antennas, cannot easily be shielded permanently. Evasive-maneuvering techniques may reduce the present probability of collision for specific satellites in certain circumstances but do not provide a practical long-term solution.

The only natural mechanism opposing debris buildup is removal by atmospheric drag. This process can take a very long time, however, especially from high altitudes, and causes debris to migrate from higher to lower altitudes. Another mechanism, collection by a spacecraft ("orbital garbage truck"), would be extremely difficult and expensive. Prevention of debris formation is the most effective approach.

At the present time, the collision hazard is real but not severe. Continuation of present policies and practices, however, ensures that the probability of collision will eventually reach unacceptable levels, perhaps within a decade. Future problems can be forestalled by initiating studies and implementing their results in five major areas: 1) education, 2) technology, 3) satellite and vehicle design, 4) operational procedures and practices, and 5) national and international space policies and treaties.

13.2 Space Debris Environment: Low Earth Orbit

At any one time, there are about 200 kg of meteoroid mass moving through altitudes below 2000 km at an average speed of about 20 km/s. Most of the mass is found in particles of about 0.1-mm diameter.[1] The meteoroid environment has always been a design consideration for spacecraft. The Apollo and Skylab spacecraft were built to withstand impacts on critical systems from meteoroids having sizes up to 3 mm in diameter. Larger sizes were so few in number as to be of no practical significance for the duration of the mission. Some small spacecraft systems required additional shielding against meteoroids as small as 0.3 mm in diameter in order to maintain an acceptable reliability. The trend in the design of future spacecraft (as for example, the International Space Station) is toward larger structures, lighter construction, and longer times in orbit. These factors increase the concern about damage from particles in the 0.1- to 10-mm size range.

It is no longer sufficient, however, to consider only the natural meteoroid environment in spacecraft design. Since the time of the Apollo and Skylab programs, launch activity has continued and increased. As a result, the population of orbital debris has also increased substantially. The total mass of debris in orbit is now approximately 2300 tons at altitudes below 2000 km. Relative to one another, pieces of debris are moving at an average speed of 10 km/s, or only half the relative speed of meteoroids. The significant difference between the orbital debris population and the meteoroid population is that most of the debris mass is found in objects several meters in diameter rather than 0.1 mm in diameter as for meteoroids. This large reservoir of mass may be thought of as a potential source for particles in the 0.1- to 10-mm range. That is, if only one ten-thousandth of this mass were in this size range, the amount of debris would exceed the natural meteoroid environment. The

potential sources for particles in this size range are many:

1) Explosions: More than 145 spacecraft are known to have exploded in low Earth orbit (LEO) and account for about 42% of the U.S. Space Command (USSPACECOM) Catalog.

2) Hypervelocity collision in space: One known satellite collision (CERISE) and several near misses have been recorded in recent history.

3) Deterioration of spacecraft surfaces: Oxygen erosion, ultraviolet radiation, and thermal stress are known to cause certain types of surfaces to deteriorate, producing small particles. Returned surfaces from orbit (SOLAR MAX, LDEF, SFU, EuReCa, Euro Mir '95, and others) have provided some information on this effect.

4) Solid rocket motor firings: Up to one-third of the exhaust products of a solid rocket motor may be aluminum oxide particles in the size range 0.0001 to 0.01 mm. Slag (up to several centimeters in diameter) may be generated during and after the burn of solid rocket motors.

5) Unknown sources: Other sources are likely to exist. Particulates are commonly observed originating from the Space Shuttle and other objects in space.

13.3 Debris Measurements

What is currently known about the orbital debris flux is from a combination of ground-based and in-space measurements. These measurements have revealed an increasing population with decreasing size. Beginning with the largest sizes, a summary of these measurements follows.

The USSPACECOM tracks and maintains a catalog of "all man-made objects" in space. The catalog, as of January 2001, contained 8978 objects, most in LEO. Figure 13.1 shows the growth of the satellite population from 1958 through 2000. This plot excludes space probes many of which are still in orbit.

The actual ownership of the satellite population is given in Table 13.1 and its composition illustrated in Fig. 13.2. As can be seen from Fig. 13.2, nearly half of the objects in the catalog have resulted from more than 145 satellite breakups, the

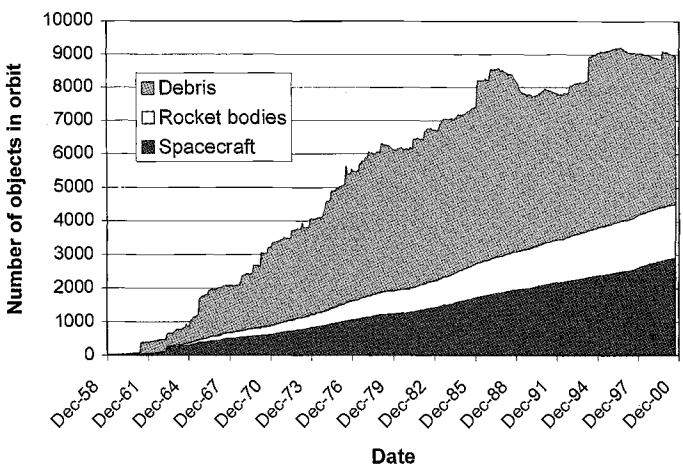


Fig. 13.1 Resident Space object population growth (from Ref. 2).

Table 13.1 Satellite Catalog boxscore 1/24/01 (from Ref. 3)

	Payload			Debris			
Source	On orbit	Decayed	Total	On orbit	Decayed	Total	All
ArabSatCom	7	0	7	0	0	0	7
ASIASAT	3	0	3	0	0	0	3
Argentina	5	2	7	0	0	0	7
Australia	7	2	9	2	0	2	11
Brazil	10	0	10	0	0	0	10
Canada	17	1	18	0	2	2	20
China/Brazil	1	1	1	0	0	0	1
Chile	1	1	1	0	0	0	1
CIS/Russia	1370	1752	3122	2571	10414	12985	16107
Czech Rep.	4	1	5	0	0	0	5
Denmark	1	0	1	0	0	0	1
Egypt	2	0	2	0	0	0	2
ESA	31	6	37	243	524	767	804
ESRO	0	7	7	0	3	3	10
EUTELSAT	18	0	18	0	0	0	18
France/Germ	2	0	2	0	0	0	2
France	31	8	39	16	59	75	114
Germany	20	11	31	1	1	2	33
GLOBALSTAR	52	0	52	0	1	1	53
INMARSAT	9	0	9	0	0	0	9
India	20	7	27	5	10	15	42
Indonesia	9	1	10	0	0	0	10
Iridium	86	2	88	0	12	12	100
Israel	3	3	6	0	3	3	9
IntSpaceStat.	3	0	3	0	3	3	6
Italy	11	6	17	3	0	3	20
INTELSAT	56	1	57	0	0	0	57
Japan	71	12	83	47	103	150	253
South Korea	7	0	7	0	0	0	7
Luxemburg	11	0	11	0	0	0	11
Malaysia	3	0	3	0	0	0	3
Mexico	6	0	6	0	0	0	6
NATO	8	0	8	0	0	0	8
Netherlands	0	1	1	0	0	0	1
Norway	3	0	3	0	0	0	3
ORBCOMM	35	0	35	0	0	0	35
Pakistan	0	1	1	0	0	0	1
Portugal	1	0	1	0	0	0	1
PeopRepChina	33	29	62	342	186	528	590
Taiwan(ROC)	1	0	1	0	0	0	1

continued

Table 13.1 Satellite Catalog boxscore 1/24/01 (from Ref. 3) (continued)

	Payload			Debris			
Source	On orbit	Decayed	Total	On orbit	Decayed	Total	All
Philippines	2	0	2	0	0	0	2
South Africa	1	0	1	0	0	0	1
Sea Launch	1	0	1	4	0	4	5
Spain	6	0	6	0	0	0	6
Sing/Taiwan	1	0	1	0	0	0	1
Sweden	9	0	9	0	0	0	9
Thailand	4	0	4	0	0	0	4
Turkey	4	0	4	0	0	0	4
UnArabEmir.	1	0	1	0	0	0	1
United Kingdom	20	9	29	1	4	5	34
United States	814	695	1509	2922	3827	6749	8258
Total	2821	2557	5378	6157	15152	21309	26687

history of which is illustrated in Fig. 13.3.[4] The ability to catalog small objects is limited by the power and wavelength of individual radar sites, as well as the limitations on data transmission within the network of radar sites. Consequently, objects smaller than about 10 to 20 cm are not usually cataloged.

The most comprehensive effort to track orbiting objects is carried out by the USSPACECOM. An estimate of USSPACECOM's capability to detect objects in Earth orbit is illustrated in Fig. 13.4 (Refs. 5 and 6). Only the region to the right of the heavy line is accessible to operational radar and optical systems.

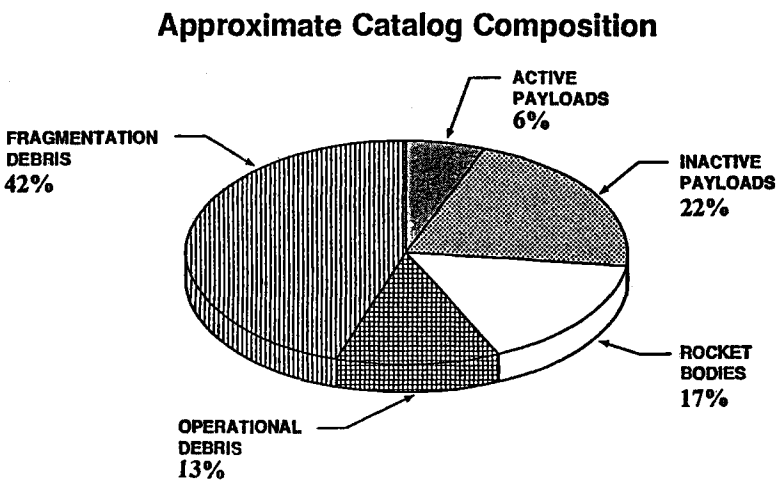


Fig. 13.2 Approximate catalog composition.

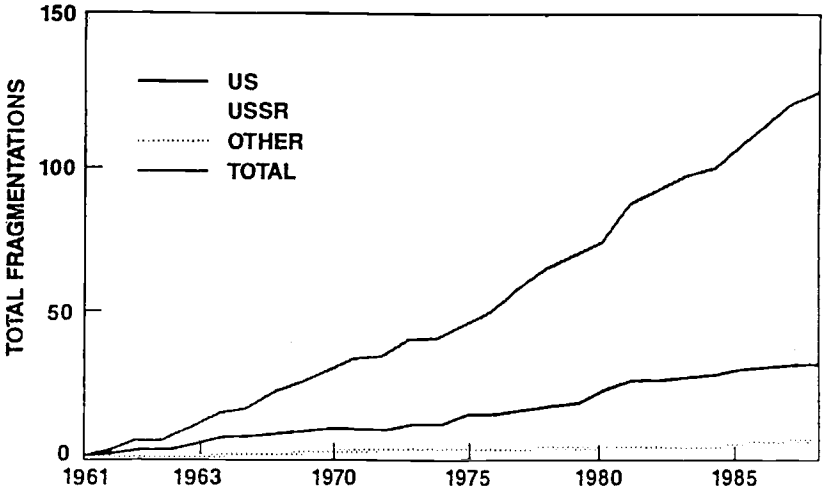


Fig. 13.3 History of fragmentation events (from Ref. 4).

The tracked (>10-cm) object flux is given in Fig. 13.5. The peak object fluxes appear at about 800-, 1000-, and 1500-km altitude and are caused by heavy use of these altitudes. A number of high-intensity explosions or breakups of spacecraft and rocket stages have also contributed to the debris population in this environment. The inclination distribution of the catalog population is shown in Fig. 13.6.

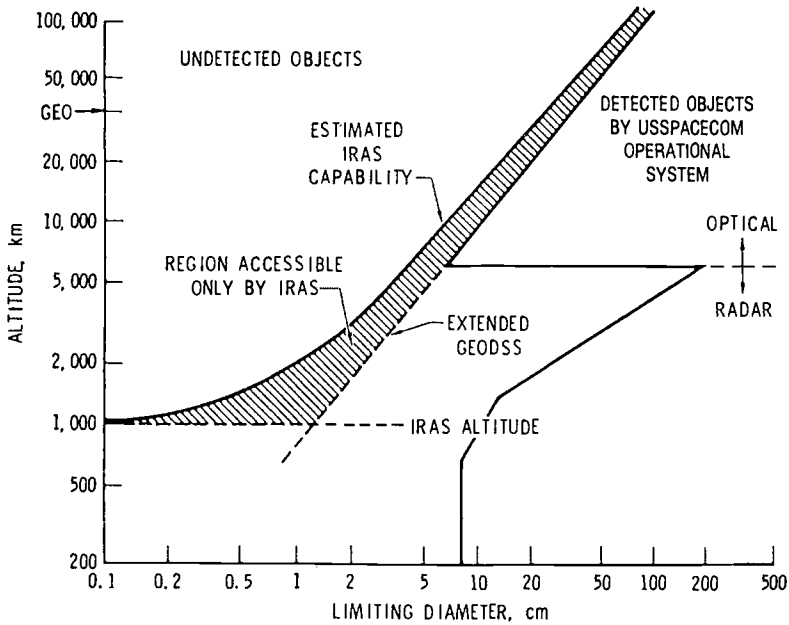


Fig. 13.4 United States capability to detect space objects (from Refs. 5 and 6).

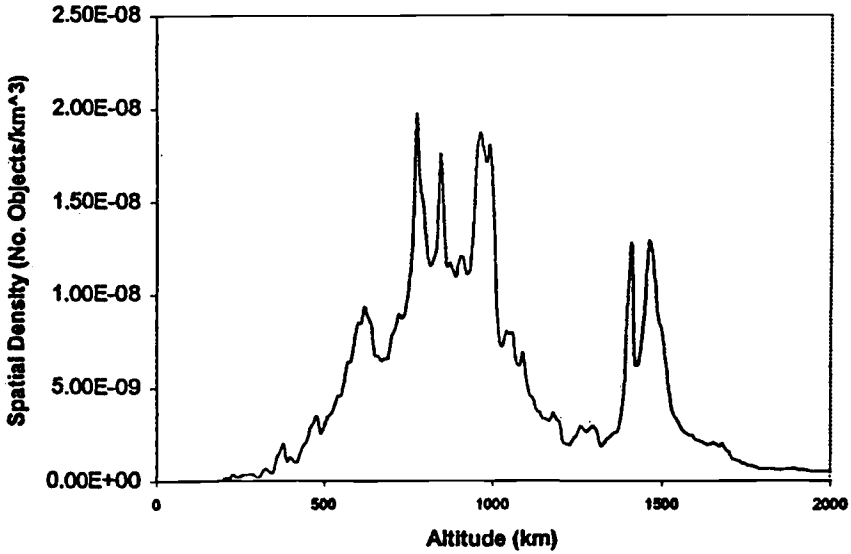


Fig. 13.5 Orbital debris density vs altitude (from Refs. 2 and 3).

Nearly all of the orbital debris measurements to date show an orbital debris flux that exceeds the meteoroid flux. These measurements are summarized and compared with the meteoroid flux in Fig. 13.7.

13.4 Space Debris Environment: Geosynchronous Equatorial Orbit

Orbits in the geosynchronous ring may be classified as 1) low or high inclination, and 2) sub- or supersynchronous-altitude drifting orbits. The ideal geostationary

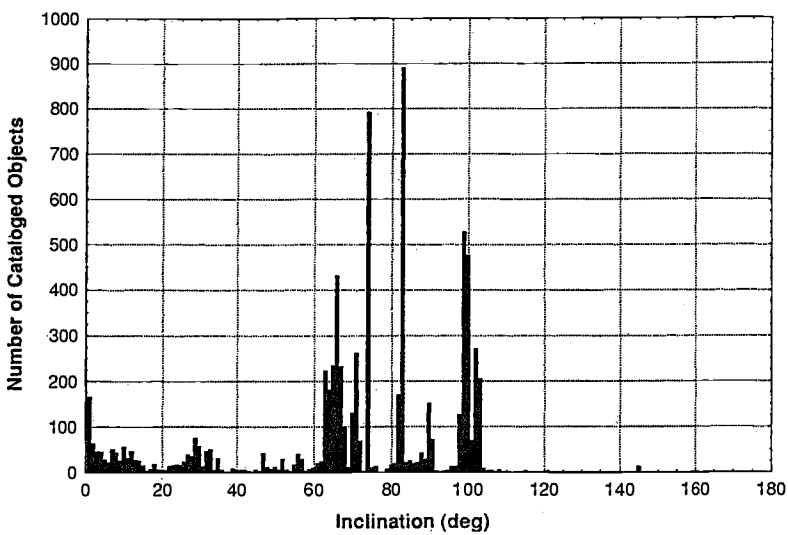


Fig. 13.6 Inclination distribution of catalog population (from Ref. 7).

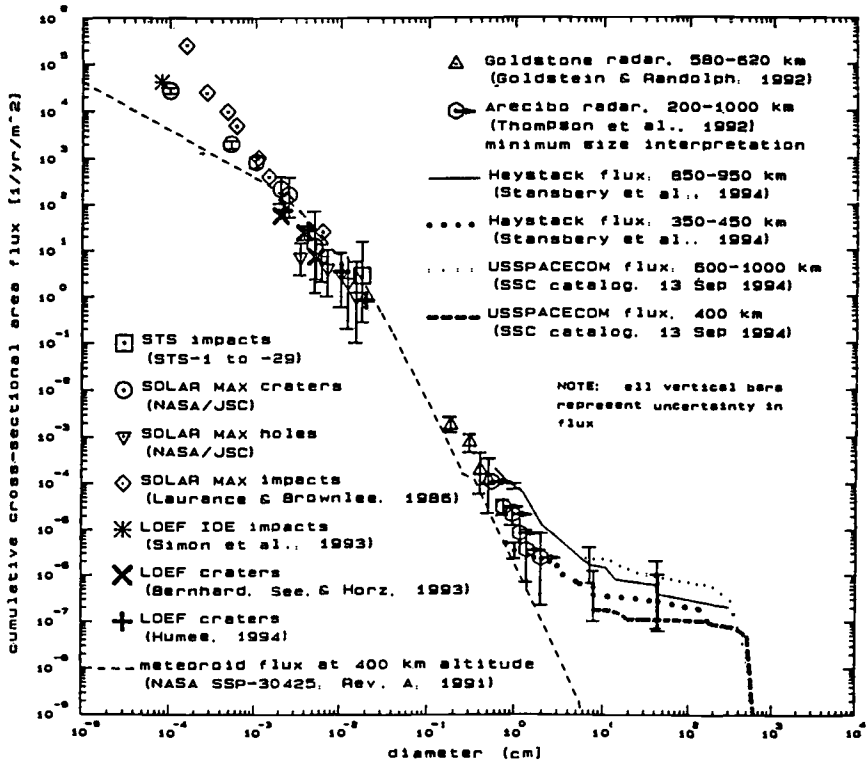


Fig. 13.7 Recent orbital debris measurements compared to meteoroid flux (from Ref. 8).

orbit has an altitude of 35,787 km and a period of 1436.2 min. A satellite whose period of revolution is equal to the period of rotation of the Earth about its axis is called a *geosynchronous satellite*. A geostationary satellite is a geosynchronous satellite in a circular orbit that lies in the plane of the Earth's equator and rotates about the Earth's polar axis in the same direction. Consequently, it remains "fixed" over a point on the equator (Refs. 9–18).

The owners of the 619 GEO spacecraft in the USSPACECOM Catalog of October 2000 are identified in Table 13.2. Fig. 13.8 shows the number of geosynchronous equatorial orbit (GEO) and near GEO objects that spend at least some time at given longitudes. The longitude overlap count of objects with drift range of less than 5° is given in Fig. 13.9. The distribution in right accession of the ascending node (RAAN) is given in Fig. 13.10 illustrating the clustering (bunching) of orbits in the 60- to 100-deg range. A view of the GEO population is also shown in Fig. 13.11, where the apparent sinusoidal distribution in latitude is the result of the orbital bunching of orbit nodes of aging spacecraft and rocket bodies.

Representative distributions of GEO objects in inclination and eccentricity (apogee-perigee difference) are shown in Figs. 13.12 and 13.13, respectively. Corresponding drift rates are shown in Fig. 13.14.

Table 13.2 Geosynchronous Catalog as of 2000–2010 (Ref. 3)

Owner	Spacecraft	Rocket bodies	Debris
Arab Sat. Comm. Org.	7	7	0
Argentina	3	0	0
Asiasat	3	0	0
Australia	5	0	0
Brazil	6	0	0
Canada	11	0	0
Egypt	2	0	0
European Space Agency	19	1	1
Eutelsat	18	0	0
France	9	0	0
France/Germany	3	0	0
Germany	5	0	0
Great Britain	7	0	0
India	12	0	0
Indonesia	10	0	0
Inmarsat	9	0	0
Intelsat	54	0	0
Israel	1	0	0
Italy	3	0	0
Japan	37	1	0
Luxemburg	10	0	0
Malaysia	2	0	0
Mexico	5	0	0
Nato	8	0	0
Norway	3	0	0
People's Rep. of China	19	0	0
Russia (CIS)	132	129	2
Singapore/Taiwan	1	0	0
South Korea	3	0	0
Spain	3	0	0
Sweden	4	0	0
Thailand	3	0	0
Turkey	2	0	0
United States	200	20	3
Total	619	158	6

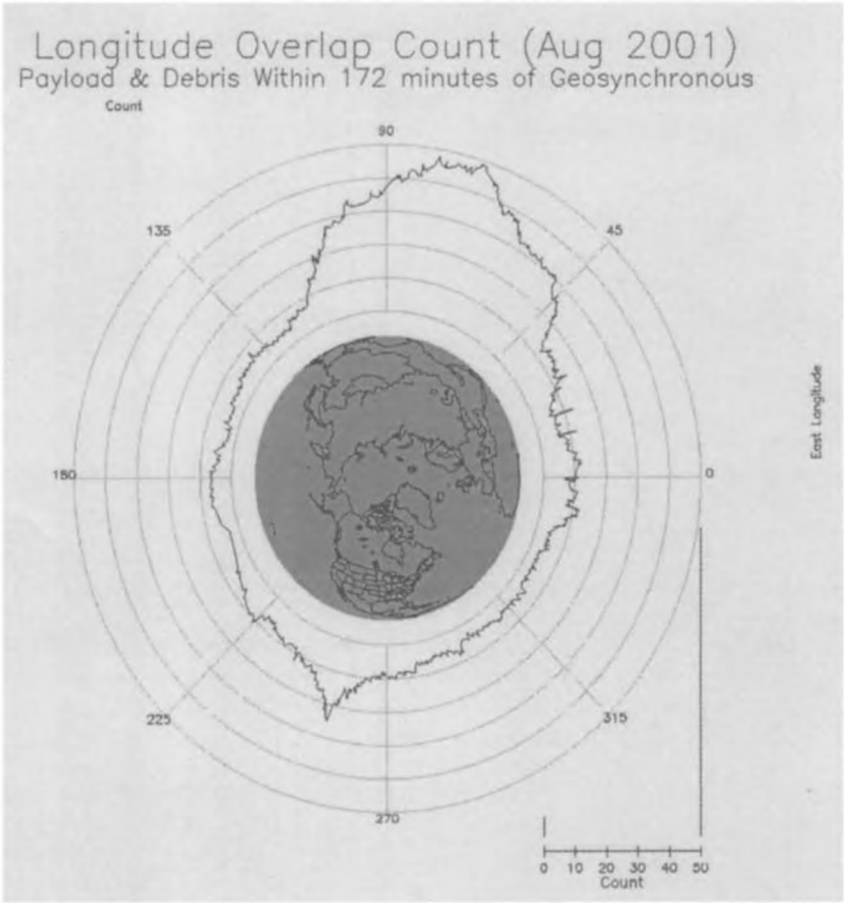


Fig. 13.8 Plot showing the number of GEO and near-GEO objects that spend at least some time at each of the longitudes indicated.*

13.5 Spatial Density

The spatial density as a function of radius r and latitude L of the N objects in the ring, defined by a toroidal volume, can be expressed approximately as

$$\rho(r, L) = \sum_{k=1}^{N} \rho_k \tag{13.1}$$

where

$$\rho_k = \frac{P_k(r_1, r_2) P_k(-L, L)}{\Delta V_k} \tag{13.2}$$

*Private communication from R. G. Gist.

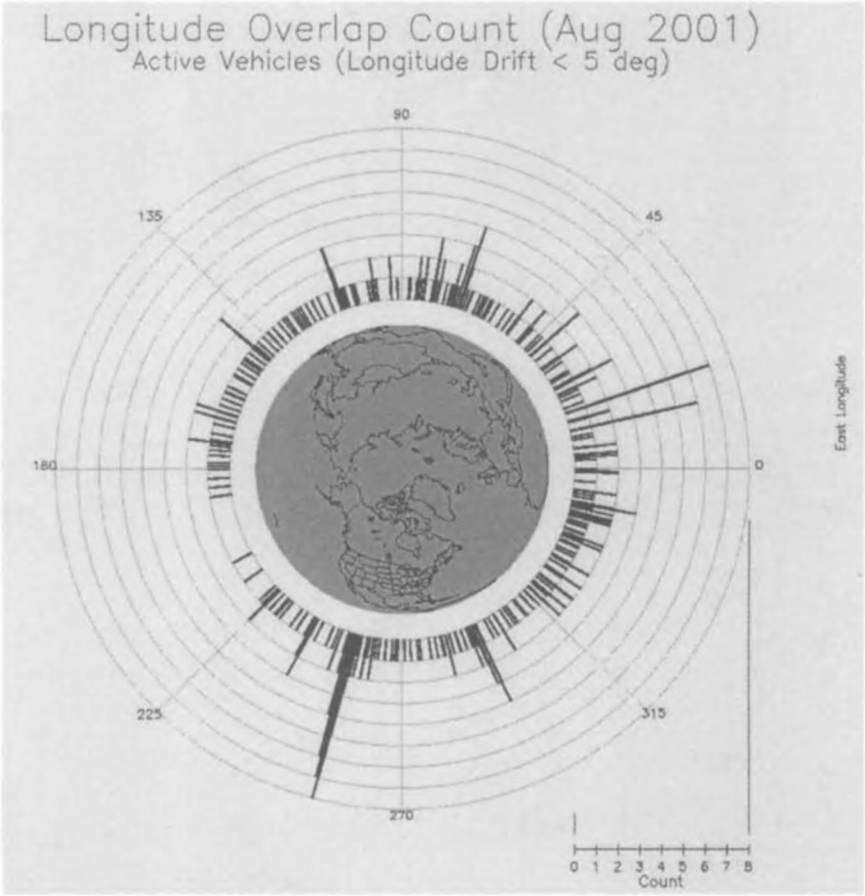


Fig. 13.9 Plot showing the longitude overlap count of objects whose drift range is less than 5 deg.*

with the probability of the kth object being within the toroidal volume,

$$\Delta V_k = \frac{4\pi}{3}\left(r_2^3 - r_1^3\right)\sin L \tag{13.3}$$

defined by the arbitrary orbital radii r_1, r_2, and latitude L as shown in Fig. 13.15. Here,

$P_k(r_1, r_2)$ = probability that the kth object is within the radial range Δr
= $r_2 - r_1$

$P_k(-L, L)$ = probability that the kth object is between the latitudes $-L$ and L

*Private communication from R. G. Gist.

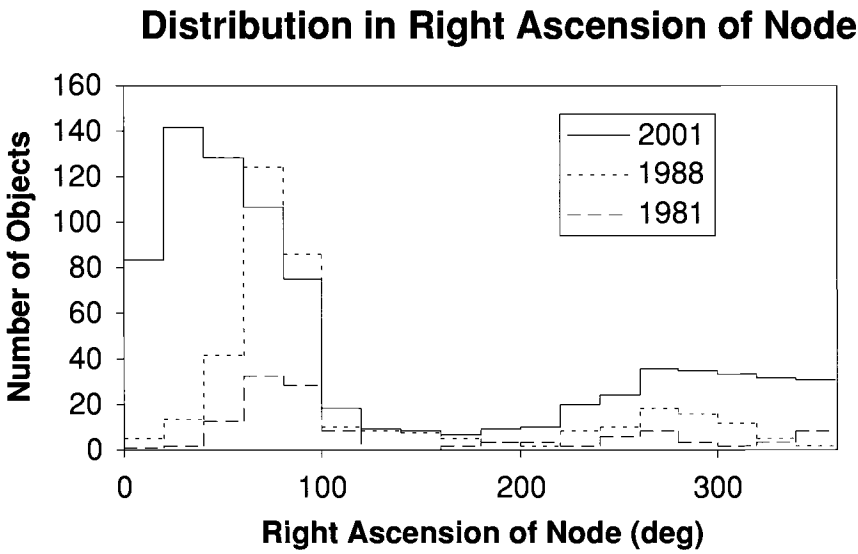


Fig. 13.10 Distribution in right ascension of ascending node (deg).*

For example, if $r_a \geq r_2 > r_1 \geq r_p$, where r_a and r_p are the orbit apogee and perigee, respectively, and, if $i > L$, where i is the orbit plane inclination of the object,

$$P_k(r_1, r_2) = P_k(r_p, r_2) - P_k(r_p, r_1) \tag{13.4}$$

where the probability that the object lies between perigee r_p and some radius r, is derived in Ref. 19 as

$$P(r_p, r) = \frac{1}{2} + \frac{1}{\pi}\sin^{-1}\left[\frac{2(r-a)}{r_a - r_p}\right] - \frac{1}{a\pi}\sqrt{(r_a - r)(r - r_p)} \tag{13.5}$$

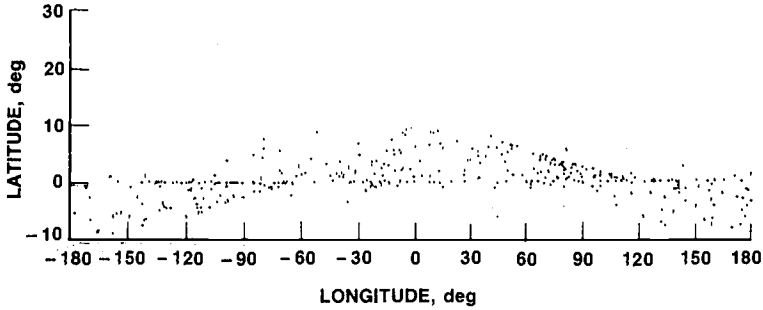


Fig. 13.11 View of geosynchronous objects.

*Private communication from R. G. Gist.

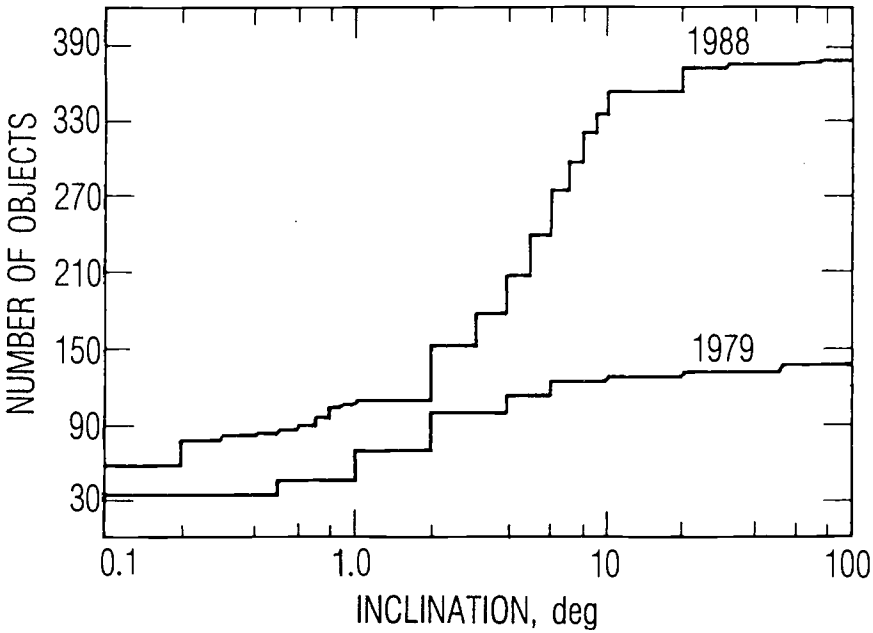


Fig. 13.12 Geosynchronous orbit inclination distribution.

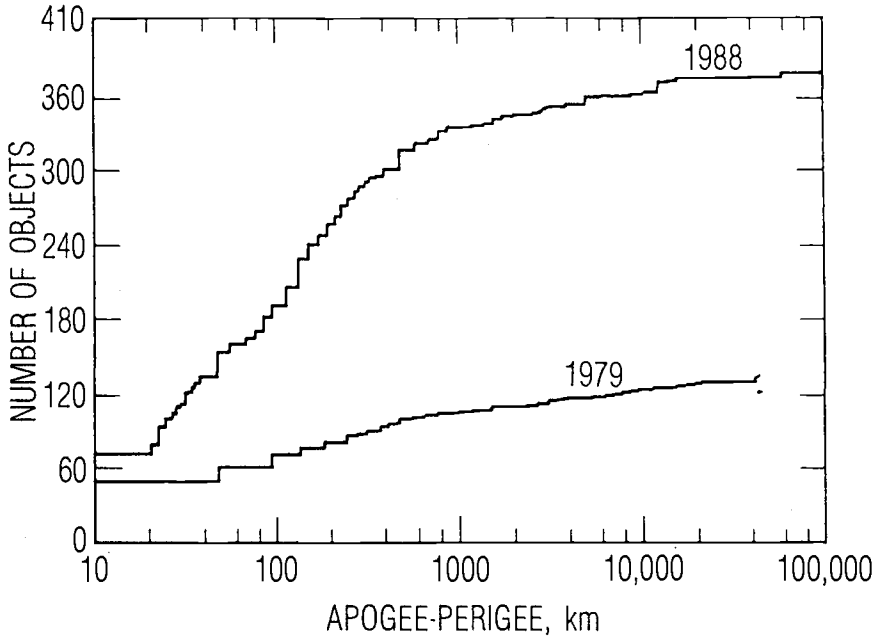


Fig. 13.13 Geosynchronous eccentricity distribution.

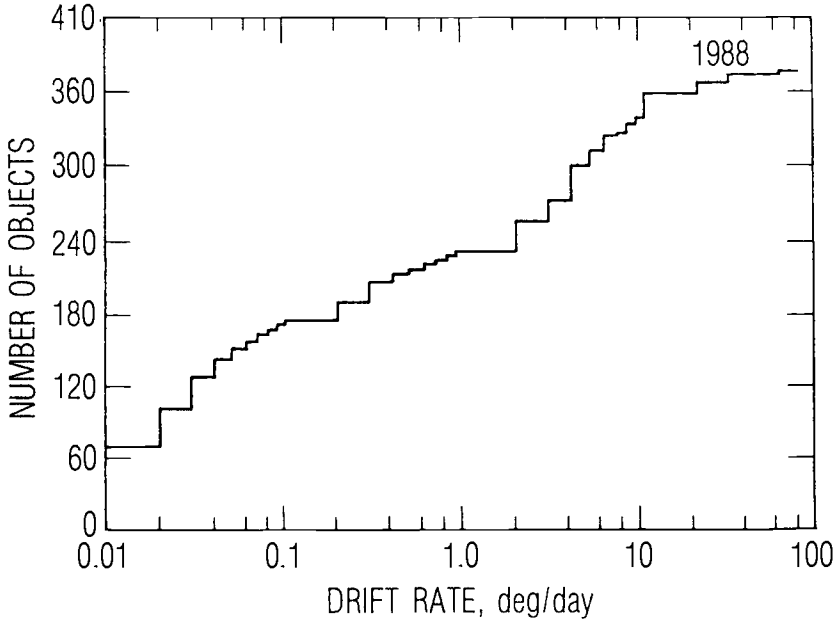


Fig. 13.14 Geosynchronous drift-rate distribution.

and the probability that the object is between the latitudes $-L$ and L is

$$P(-L, L) = \frac{2}{\pi} \sin^{-1}\left(\frac{\sin L}{\sin i}\right) \tag{13.6}$$

where a and i are the object orbit semimajor axis and inclination, respectively.

Equation (13.1) is plotted in Fig. 13.16 for a sample of 379 objects in GEO. The results show that the spatial density is maximum (about 10^{-8} sats/km^3) in a narrow range of the geosynchronous altitude. It decreases by about two orders of magnitude at ± 100 km above or below GEO. The subsequent decrease with

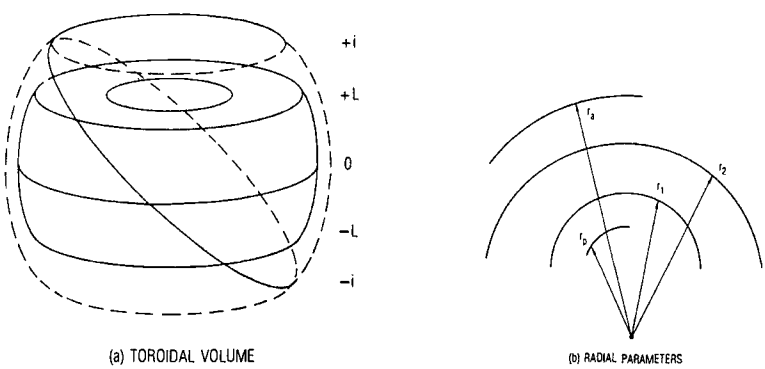


Fig. 13.15 Orbital geometry: a) toroidal volume, b) radial parameters.

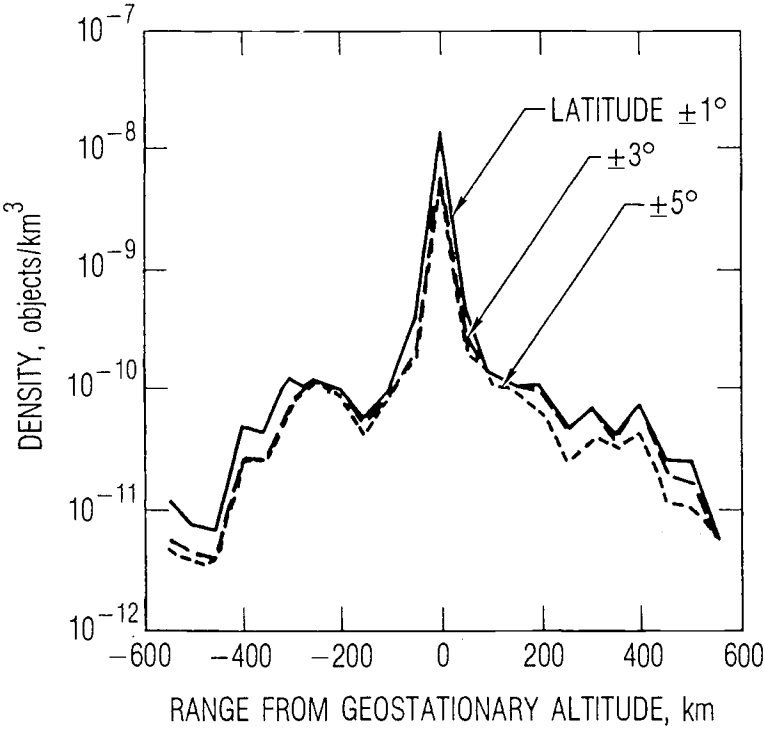


Fig. 13.16 Population density as a function of range and latitude from GEO.

altitude is not as high as in the first 100-km range, being about three orders of magnitude at ±600-km range.

13.6 Collision Hazard Assessment Methods

Uniform Density

Based on the kinetic theory of gases, the Poisson distribution model yields a simple method for computing the probability of collision between a satellite and any one of the other objects in orbit. This approach is accurate if the population of objects can be assumed to resemble random motion of molecules in a gas which is induced by repeated collisions.

The Poisson distribution approach has been used by the NASA engineering models for orbital debris and meteoroids. It has also been used by the NASA environmental predication model EVOLVE, the ESA MASTER model, the Debris Environment and Effects Program (DEEP), and the Russian Space Agency debris contamination model.

By the Poisson distribution method the probability of collision in time Δt between a spacecraft and an object is given by the relation

$$P(\text{col}) = 1 - e^{-N_{\text{enc}}} \tag{13.7}$$

where

$$N_{\text{enc}} = \int_0^{\Delta t} \rho v_r A_c \, \mathrm{d}t \tag{13.8}$$

and

ρ = density of space objects
v_r = relative velocity
A_c = cross-sectional area

Thus, for example a satellite with a projected area A_c, moving with a mean relative velocity $\bar{v}_r$ will sweep out a volume $V = \bar{v}_r A_c \Delta t$ in a time increment Δt. The number of objects encountered is ρV, where ρ is the object density in V. For $\rho V \ll 1$, the probability of collision is approximately

$$p(\text{col}) \approx \rho \bar{v}_r A_c \Delta t \tag{13.9}$$

Distance of Closest Approach

The probability that any two given objects will collide is generally a function of their orbital parameters, object size, and time. However, the collision cannot take place unless the orbits approach each other within an effective collision radius. This may occur even for initially nonintersecting orbits because the Earth's oblateness, air drag, and solar-lunar perturbations tend to alter the orbital parameters in time.

For satellites in mutually inclined circular orbits with equal, or nearly equal, periods of revolution, the distance of closest approach $R_{\min}$ occurs in the vicinity of the nodal axis on the ascending or descending passes. $R_{\min}$ generally depends on the synodic period or the angular increment $\Delta u = u_1|(T_2 - T_1)|$ where u_1 is the mean motion of satellite 1 and T_1, T_2 are the respective periods of revolution for geosynchronous satellites. Δu is typically a fraction of a degree.

If it is assumed that the position uncertainties associated with the three dimensions (coordinates) of the nominal miss distance $R_{\min}$ are Gaussian (normal), with zero biases and equal variance, and are uncorrelated, a bivariate normal density function $f(x, y)$ can be defined in plane xy containing $R_{\min}$ that is oriented normal to the relative velocity vector at encounter. Thus,

$$f(x, y) = (2\pi\sigma^2)^{-1} \exp[-x^2/(2\sigma^2)] \exp[-y^2/(2\sigma^2)] \tag{13.10}$$

A collision can take place only in a region R defined by

$$\begin{aligned} X_{\min} - R_s \leq x \leq X_{\min} + R_s \\ Y_{\min} - R_s \leq y \leq Y_{\min} + R_s \end{aligned} \tag{13.11}$$

where $X_{\min}$, $Y_{\min}$ define the magnitude of $R_{\min}$, i.e., $R_{\min} = (X_{\min}^2 + Y_{\min}^2)^{1/2}$ and R_s is the effective collision radius. The probability of collision for $R_s \ll R_{\min}$ takes the following form:

$$P(\text{col}) = \int\int_R f(x, y)\,\mathrm{d}x\,\mathrm{d}y \tag{13.12}$$
$$= (2/\pi)(R_s/\sigma)^2 \exp\left[-R_{\min}^2/(2\sigma^2)\right]$$

For example, the collision probability in GEO for several close approaches between OPS 6391 and WESTAR-A geosynchronous satellites is shown in Fig. 13.17, plotted as a function of position uncertainty. The values of σ are generally functions of tracking techniques; but, even for the close approaches examined, the collision probabilities were found to be quite small.[18]

A generalization of this approach and a numerical solution to the evaluation of the collision hazard is described in Ref. 20. The evaluation of the object's position and velocity is performed by standard propagation methods and a statistical treatment of the position uncertainty. An evaluation of the effects of orbital parameter errors on the distance of closest approach and the resulting probability of collision between two satellites is given in Ref. 21.

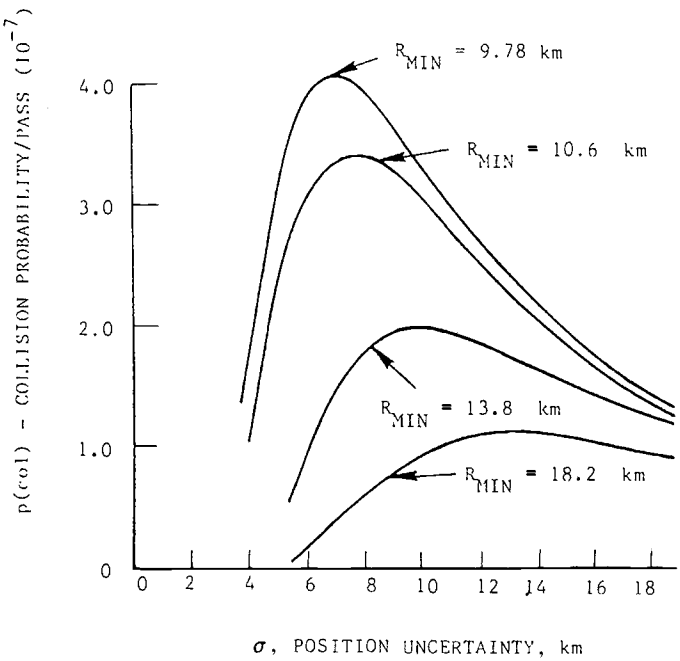


Fig. 13.17 Typical geosynchronous orbit collision probabilities as a function of position at encounter for R_s = 9 m.

Weibull Distribution

A probabilistic assessment of the collision risk can be determined directly by the use of Monte Carlo analysis in conjunction with a Weibull distribution fit to the random sample of the variable of interest. Random samples of a set of variables such as distances of closest approach are generated by numerical means, either by propagating trajectories of the objects of concern or by a Monte Carlo analysis in conjunction with methods of extreme order statistics. An analytical probability density function, i.e., a Weibull function, can be fitted to these data and theorems in extreme order statistics used to justify inference about probability of collision at the small extreme values of the minimum approach distances (Refs. 22–24). The explicit formulation in this approach is as follows:

A Weibull probability distribution function fit to the absolute (global) minimum distances of closest approach is approximated in the form

$$f(x) = \left(\frac{\tau}{\beta}\right)\left(\frac{x}{\beta}\right)^{\tau-1} e^{-(\frac{x}{\beta})^{\tau}} \tag{13.13}$$

Here, x is the random variable (i.e. minimum range); τ and β are the shape and scale parameters, respectively.

The cumulative Weibull collision probability function is of the form

$$\begin{aligned} F(x) &= \int f(x)\mathrm{d}x \\ &= 1 - \exp\left[-\left(\frac{x}{\beta}\right)^{\tau}\right] \\ &\approx \left(\frac{x}{\beta}\right)^{\tau} \quad \text{for} \quad \frac{x}{\beta} \ll 1 \end{aligned} \tag{13.14}$$

The encounter rate N for an object at distance r_c, can be expressed mathematically as:

$$N = \frac{F(r_c)}{E(t_c)} \tag{13.15}$$

where

$$F(r_c) = \int_o^{r_c} f(x)\mathrm{d}x \tag{13.16}$$

and

$$E(t_c) = \frac{E(c)}{v_r} \tag{13.17}$$

Here $E(t_c)$ is the characteristic dwell time by an object within the radius r_c of the given spacecraft. $E(c)$ is the expected (mean) value of the chord of a circle with

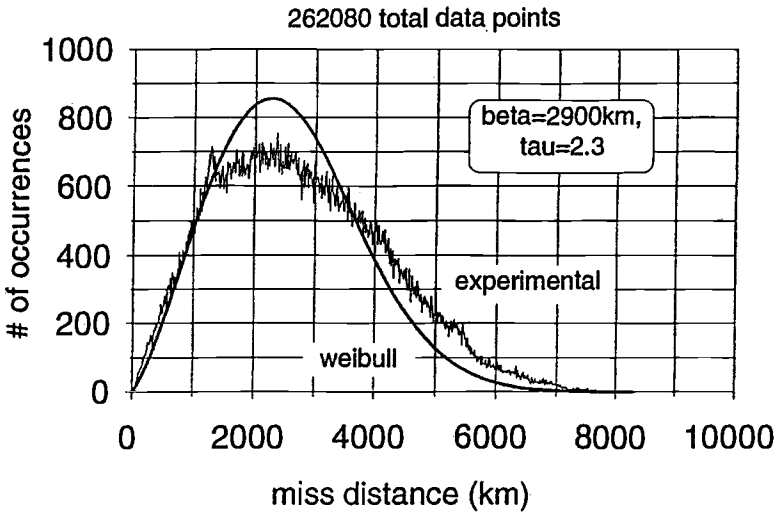


Fig. 13.18 Relative distance statistics (Ref. 24).

radius r_c and v_r is the mean relative velocity of objects. It can be shown that $E(c)$ is numerically equal to one fourth of the circumference of the circle of radius r_c.

The mean (expected) time to collision is then

$$\begin{aligned} T_c &= 1/N \\ &= (\pi r_c/2v_r)(\beta/r_c)^{\tau} \end{aligned} \tag{13.18}$$

An example of a Weibull probability distribution of close encounters between a tether in low earth orbit and a population of debris objects is illustrated in Fig. 13.18. The probability of a specified distance occurrence (Eq. 13.14) for this example is given in Fig. 13.19 for the Weibull distribution function and the calculated (experimental) distribution function respectively[24]. Another method for

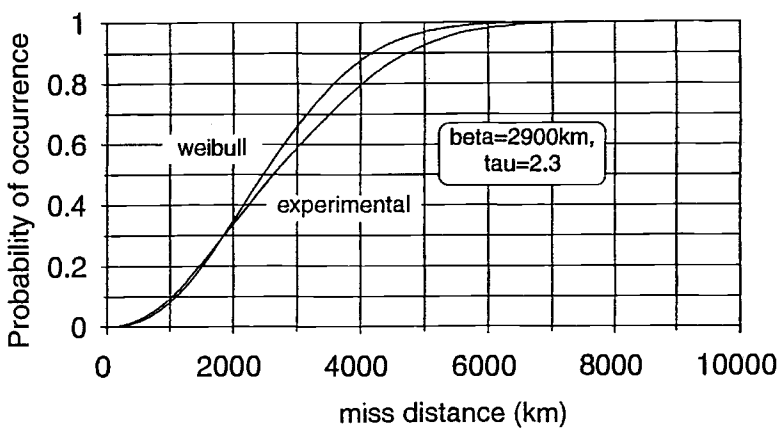


Fig. 13.19 Cumulative relative distance statistics (Ref. 24).

calculating a collision probability between a tether and a satellite is described in Ref. 25.

13.7 Collision Hazards Associated with Orbit Operations

The hazards can be broadly defined in two categories; hazards due to explosion and hazards due to collision. Explosions are a major concern because of the probabilities of collision of other satellites with the debris from the explosion. Explosions can be deliberate or accidental. For example, on 13 November 1986, an Ariane third stage experienced an anomalous explosion in a sun-synchronous orbit and left about 460 trackable fragments.

Collisions are also a major concern because an unplanned collision may damage an active spacecraft and because of the collision probabilities of other satellites with the debris. Like explosions, collisions can be planned and unplanned. A planned collision can be part of an orbital test, such as with the Delta 180 mission, for example conducted by the Department of Defense in 1986.

The collision of the French CERISE spacecraft and a fragment from the SPOT 1 Ariane I upper stage on 24 July 1996 marked the first time that two objects in the U.S. satellite catalog have collided. This collision with CERISE's stabilization boom caused the spacecraft to tumble end-over-end. Collisions with smaller particles that are below the threshold for detection and tracking are more frequent and also can cause serious damage. An example of the damage potential is shown in Fig. 13.20, which shows the 4-mm-diam crater on the Shuttle window

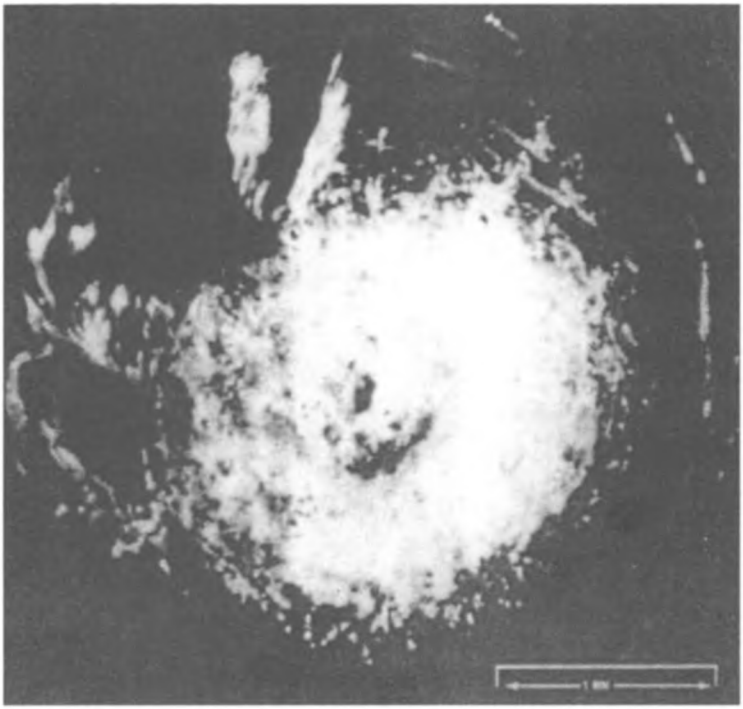


Fig. 13.20 STS-7 window impact.

from the STS-7 mission. Energy-discursive x-ray analysis was used to determine the composition of partially fused material found in the bottom of the pit. Titanium oxide and small amounts of aluminum, carbon, and potassium were found added to the pit glass. Crater morphology places the impacting particle diameter at 0.2 mm, with a velocity between 3 and 6 km/s. From these data, it is concluded that the particle was man-made and probably an orbiting paint fleck.

Low Earth Orbit (LEO)

Risk assessments for LEO satellites have been performed since the 1960s using the Poisson model in cases where there is a large number of independent events and each event has a small probability of occurring. To compute the probability of an impact from space debris requires a meteoroid/space debris (M/OD) environmental model, a satellite configuration and a mission profile. The probability of penetration or a component failure due to space debris requires the geometry of critical subsystems, the ballistic limit equation and the subsystem tolerance to possible debris impacts. The results may be expressed as the probabilities of impact damage to any given subsystem for both the meteoroid and the space debris environment components.

For human space flight, risk assessments are essential in ensuring the safety of the mission. They are also used to design the location and type of space debris shielding which will be necessary to protect the crew and the critical subsystems on the International Space Station. For unmanned spacecraft the risk assessments aid in the placement of shielding materials to protect critical subsystems and components. Typical mean times between impacts on a 10 square meter cross-section area in LEO are given in Table 13.3.

Other techniques such as the distance of closest approach or the statistical (Weibull) distribution methods may be more appropriate when the population of objects is not randomly distributed, but exhibits correlated motion such as in GEO or in debris clouds generated by on-orbit collision or explosion events (Ref. 24).

Geosynchronous Equatorial Orbit (GEO)

The population of space objects in or near GEO is given in Table 13.2 and its longitudinal distribution is illustrated in Figs. 13.8–13.11. Fig. 13.8 illustrates the bunching of GEO objects near 75 degrees east longitude and 255 degrees east longitude. These longitudes correspond to the equatorial stable points caused by the radial asymmetry of the equator. Objects near such points tend to oscillate about them with frequencies of 2–3 years typically depending on the initial conditions, unless acted upon by station keeping impulses.

Table 13.3 Mean time between impacts on a satellite with a cross-section area of 10 square meters (Ref. 26)

Height of circular orbit (km)	Objects 0.1–1.0 cm (yr)	Objects 1–10 cm (yr)	Objects > 10 cm (yr)
500	10–100	3500–7000	150,000
1000	3–30	700–1400	20,000
1500	7–70	1000–2000	30,000

Fig. 13.11 illustrates the apparent sinusoidal motion of satellites in GEO due to their diurnal crossing of the equator. The satellites at higher latitudes traverse the equatorial plane at higher velocities (eg. 500 m/s) and thus pose a higher risk of collision to geostationary satellites in the equatorial plane. Because of the much lower relative velocity in GEO compared to that in LEO (eg. 7 km/s) and the limited number of objects, the overall probability of collision is substantially lower than that in LEO. Special collision possibilities exist in GEO because of the close proximity of operations spacecraft to each other at certain longitude. Such collisions can be eliminated by spacecraft control procedure for actively controlled satellite but may still pose a significant hazard during orbit transfers and for inactive (debris) objects as examined in Ref. 18. The limited number of large objects near GEO also permits the prediction of close approaches between operational spacecraft and tracked debris in sufficient time to perform an evasive maneuver.

Two break-ups are known to have occurred in GEO (one spacecraft and one upper stage). The number of smaller fragments of less than 1 m in diameter is not well known. In such cases the fragments would be widely dispersed and their relative velocities would be higher. Additional space debris measurements in GEO are therefore needed to better define the debris environment so that more accurate risk assessments can be performed. The annual collision probability for average operational spacecraft with other catalogued objects is currently on the order of 10^{-5} (Ref. 26).

13.8 Debris Cloud Modeling

The behavior of debris particles following a breakup in orbit must be modeled in order to determine the collision hazard for resident space objects in the vicinity of the breakup. Short-term behavior (measured in days) of a debris cloud for a hypervelocity collision is, for example, described in Refs. 27–32. The debris cloud dynamics model in Ref. 28 was developed based on the linearized rendezvous equations for relative motion in orbit. Subsequently, the short-term behavior and modeling of debris clouds in eccentric orbits were developed, which described the shape, in-track particle density variation, and volume of a debris cloud resulting from an isotropic breakup.[29] A typical cloud density profile is shown in Fig. 13.21.

Spacecraft Breakup Modeling

There are a number of empirically based fragmentation models that describe the number, mass, and velocity distributions of fragments resulting from hypervelocity collisions or explosions (see, for example, Refs. 27, 30 and 31). There is, however, much uncertainty due to the limited experimental data that must be used to validate such models. Only more testing, coupled with the development of new analytical computational tools, can reduce these uncertainties. The importance of developing a better understanding of the breakup phenomenology for the collisions between LEO objects is also clear. Such collisions are likely to occur with impact velocities in the 8- to 15-km/s range and result in hundreds of trackable fragments and potentially millions of smaller particles. The development of breakup models is essential. These models should include strict energy and momentum conservation laws such as are being implemented in program IMPACT, for example.[31] Other models should be developed by numerical/analytical

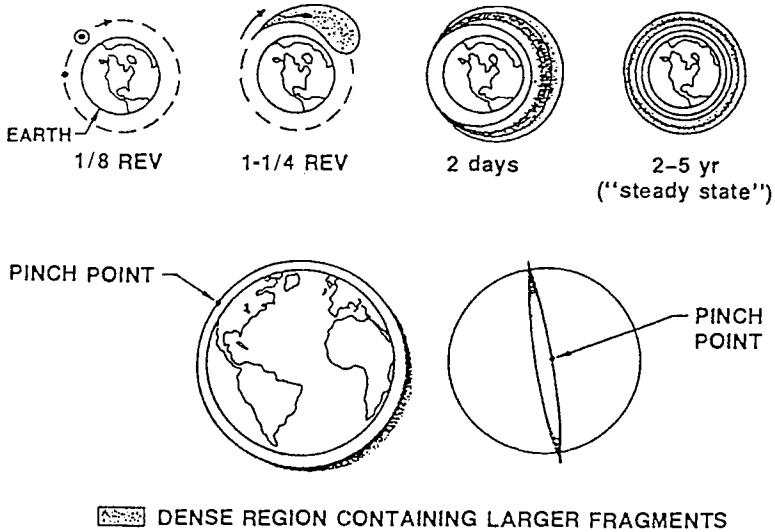


Fig. 13.21 Evolution of a debris cloud.

approaches coupling existing models that describe local area effects due to hypervelocity impact (hydrocodes) with models that describe the spacecraft vibration and deformation due to the impact (structural response codes). Statistical fragmentation models can also be developed using the outputs of other models described previously.

Short- and Long-Term Debris Evolution Modeling

The fragmentation of upper stages and satellites accounts for nearly one-half of all trackable debris objects. Understanding the fragmentation process (due to explosion or hypervelocity collision) is therefore essential in order to estimate the mass, number, velocity, and ballistic coefficients of the resultant fragments. The modeling of the fragmentation debris (as orbiting clouds) can be separated into short- and long-term types. *Short-term* may be defined in terms of days after the breakup, while *long-term* implies subsequent evolution of the fragment clouds and the resulting "steady-state" environment due to the breakup. Primary elements of such modeling include: a) determination of breakup causes, b) orbital lifetime of untrackable debris, c) the breakup process (i.e., breakup modeling), d) debris cloud evolution, e) future traffic projections, and f) collision hazards to resident space objects. Figure 13.22 illustrates the interfaces required between various types of models.

Space debris models may be discrete, representing the debris populations in a detailed format, or engineering approximations. Typically, the initial debris population is represented at an initial epoch and propagated forward in time in a stepwise manner, taking into account various source and sink mechanisms with relevant orbital perturbations. The "steady state" models do not include periodic concentrations of debris following an orbital breakup vehicle.

The models must rely upon historical records of launch activity, in-orbit breakups, and satellite characteristics. There is also limited data on spacecraft material response to impact at high relative velocities. Future traffic scenarios and

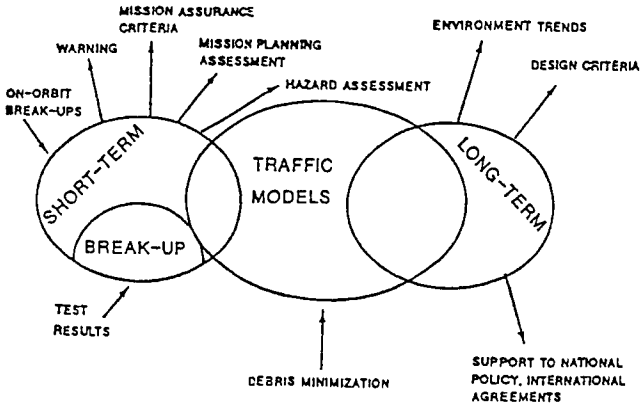


Fig. 13.22 Models.

application of mitigation measures also influence the outcome of model predictions. Continuous updating and validating of space debris is therefore necessary to reflect improvements in the observational and experimental data.

Several of the currently available short- and long-term models are listed in Table 13.4.

The basic characteristics of the debris environment models as given in Ref. 26 are:

1) *CHAIN* and *CHAINEE.* CHAIN was developed by the Technical University of Braunschweig under contract. Since 1993, this model has been maintained and improved by NASA. CHAINEE, the European extension of CHAIN, is used by

Table 13.4 Debris environment models (Ref. 26)

Model name	Source	Evolutionary period	Engineering model available	Minimum size	Orbital regime
CHAIN	NASA	Long term	No	1 cm	LEO
CHAINEE	ESA	Long term	No	1 cm	LEO
EVOLVE	NASA	Short- and long-term	No	1 mm	LEO
IDES	DERA	Short- and long-term	No	0.01 mm	LEO
LUCA	TUBS	Long-term	No	1 mm	LEO/MEO
MASTER	ESA	Short-term	Yes	0.1 mm	LEO/GEO
Nazarenko	RSA	Short- and long-term	No	0.6 mm	LEO
ORDEM96	NASA	Short-term	Yes	1 μm	LEO
SDM/STAT	ESA/CNUCE	Short- and long-term	No		LEO/GEO

ESA. The model, an analytical "particle-in-a-box" model, describes the population and the collision fragments up to an altitude of 2000 km using four altitude bins in LEO and five mass classes.

2) *EVOLVE* was developed by the NASA Johnson Space Center to provide both short-term and long-term forecasts of the LEO environment with excessive source terms and detailed traffic models, based on quasi-deterministic population propagation techniques that are suitable for both LEO and GEO modeling.

3) *ORDEM96* is a semi-empirical engineering model developed by the NASA Johnson Space Center. It is based upon extensive remote and *in situ* observations and is used to support United States Space Shuttle and International Space Station design and operations.*

4) *MASTER* is an ESA semi-deterministic environment model based on three-dimensional discretization of spatial densities and transient velocities. The model is applicable to altitudes from LEO to GEO, providing environment estimates in the short-term. A less detailed version of MASTER is available as an engineering format. Both models were developed by the Technical University of Braunschweig under ESA contract.*

5) *IDES* is a semi-deterministic model of the environment using detailed historical and future traffic models to provide short- and long-term predictions of the space debris environment and the collision flux it presents to specific satellites. The model was developed by the Defence Evaluation and Research Agency (DERA), Farnborough, United Kingdom.

6) *Nazarenko*, a model developed by the Centre for Programme Studies (CPS) of RSA, is a semi-analytic, stochastic model for both short- and long-term predictions of the LEO debris environment, providing spatial density, velocity distributions and particle fluxes. The model takes account, in average form, of debris sources (except for the cascading effect) and of atmospheric drag; it has been adjusted on the basis of Russian and American catalog data and published measurements of somewhat smaller fragments (more than 1 mm), while also taking account of a priori information.

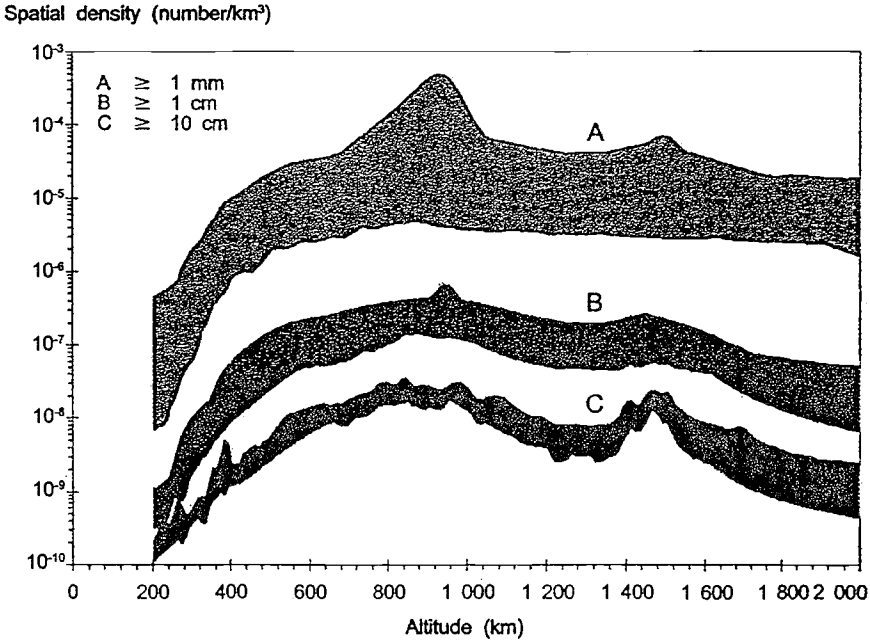
7) *SDM* is a semi-deterministic model to provide both short- and long-term predictions of the space debris environment. The code, developed at CNUCE (Centro Nazionale Universitario di Calcalo Electronico) makes use of a detailed traffic model, including satellite constellations, and considers several source model options for explosions and collisions.

Several different models have been used to develop spatial density contours for different size particles in LEO as is illustrated in Fig. 13.23 (Ref. 26).

Very Long-Term Models

The very long-term models (up to 100 years) predict the number of objects as a function of time, altitude, orbital inclination, and object size. Such models can be used to assess the effectiveness of any mitigation measures and the impact of new space activities. The effects of collisions and or larger objects must also be taken into account in determining future debris environments. The so-called destructive collisions among the population objects is proportional to the square of the number of objects. Therefore, long-term mitigation should focus on the removal

*NASA ORDEM 2000 and ESA MASTER 99 models are also available.



Sources: NASA (ORDEM96); DERA (IDES); ESA (MASTER); CNUCE (SDM); and NAZARENKO

Fig. 13.23 Model values for current spatial density (from Ref. 26).

of mass and cross-sectional areas from orbit. The assessment of the consequences of collisions among larger objects requires reliable breakup models. Such models depend on empirical data of hypervelocity collisions for validation purposes.

Additional assumptions describing future space activities, including debris generation mechanisms, are required. For example: number of launches and related orbits; number and size of payloads per launch; number of mission-related objects (i.e. operational debris); number of expected explosions of spacecraft and upper stages; and new uses of space (e.g. satellite constellations). All of the above parameters are subject to variations with time due to technical, scientific, financial, and political aspects. Consequently, many uncertainties are inherent in such modeling. The long-term models therefore do not agree in their findings. However, some trends can be identified (Ref. 26):

1) The debris population may grow in an accelerated manner in the future if space flight is performed as in the past. This is because of the collisions that will occur given the increased mass on orbit.

2) Currently depending on size, fragments from explosions are the main source of space debris. Without some mitigation, collision fragments may become one of the main debris sources in the future.

3) Collisional fragments may contribute to the number of subsequent collisions. This means that, at some point in the future, which is difficult to predict, the population could grow exponentially. Only by limiting the accumulation of mass in LEO can this be prevented.

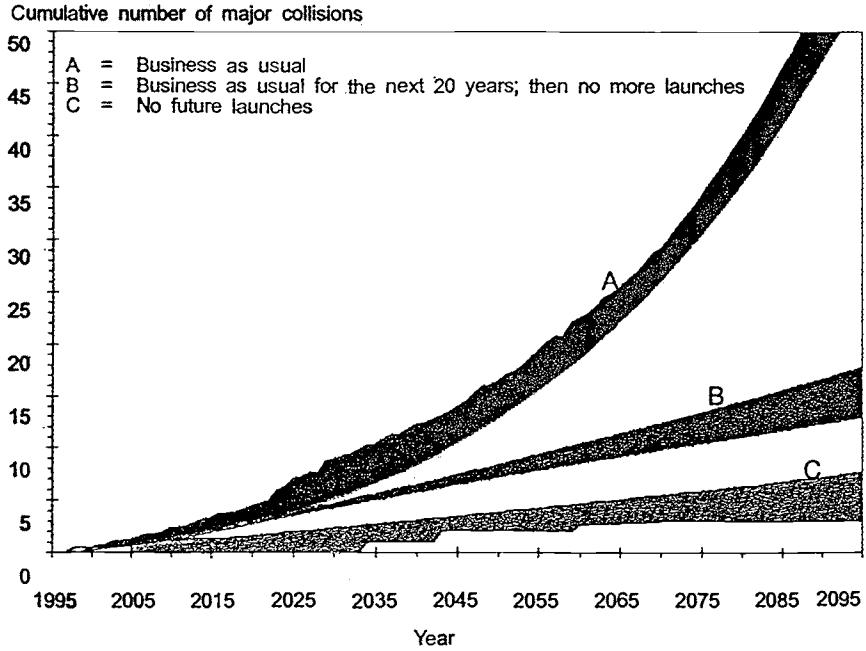


Fig. 13.24 Typical ranges for number of major collisions for three scenarios, 1995–2095 (from Ref. 26).

4) Suppressing explosions can reduce the increase in the number of objects in orbit, but cannot prevent collisions, which are driven by the total mass in orbit and the number of large objects.

Predicted rates of collision are shown in Fig. 13.24 for several future scenarios.

13.9 Lifetime of Nontrackable Debris

The long-term debris environment in Earth orbit may be expressed in terms of the source and sink effects controlling the environment. Sources are all new objects, including fragments resulting from explosions or collisions of objects in orbit. Sink effects govern the removal of objects such as the atmospheric drag that tends to keep the debris population at low altitude nearly constant.

First evidence of the existence of small debris in orbit resulted from examining the Skylab IV/Apollo windows,[33] which showed aluminum-lined pits in about half of the hypervelocity pits found there. Other experiments, such as the Explorer 46 Meteor Bumper Experiment, showed evidence of debris smaller than 0.1 mm.[30] Much good data have also been obtained from the returned Solar-Max insulation blankets and aluminum louvers. These data showed that the man-made debris dominated the micrometeoroid particle flux for sizes smaller than about 0.01 mm. After nearly six years in space, the Long-Duration Exposure Facility (LDEF) results have also shown that there were a total of more than 34,000 impacts on the satellite following its recovery from orbit in January 1990 by the orbiter Columbia. Of these, more than 3000 impacts were in the 0.5- to 5-mm size range. Several

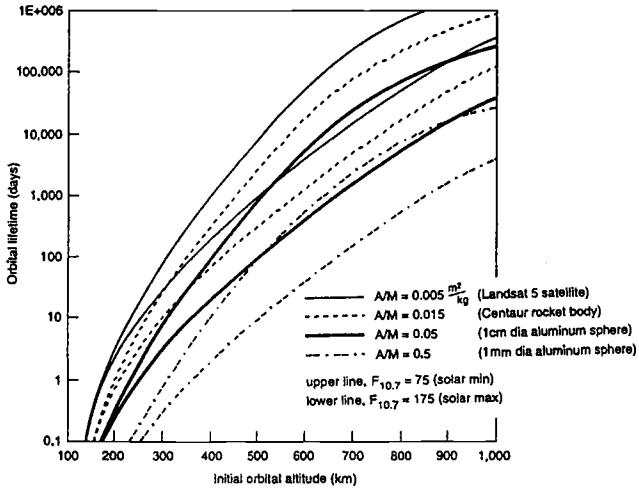


Fig. 13.25 Orbital decay time vs altitude (from Ref. 7).

studies of the spacecraft surfaces returned from LDEF are reported in Refs. 33 and 34. Orbital decay time vs initial altitude is shown in Fig. 13.25.

13.10 Methods of Debris Control

In view of the increasing threat of space debris to operations in orbit, NASA, the Department of Defense, and several space agencies from Russia, Japan, and Europe are participating jointly in the attempt to minimize the threat in the future. Reference 35, for example, outlined a number of debris-control measures which, if implemented, could decrease the quantity of space debris by prevention or debris removal. These control options fall into three categories: those requiring minimal impact on operations, those requiring changes in hardware or operations, and those requiring technology development. Options in the first category recommended for immediate application are:

1) No deliberate breakups of spacecraft, which produce debris in long-lived orbits.

2) Minimization of mission-related debris.

3) Safing procedures for all rocket bodies and spacecraft that remain in orbit after completion of their mission.

4) Selection of transfer orbit parameters to ensure the rapid decay of transfer stages.

5) Reorbiting of geosynchronous equatorial satellites at the end of life (minimum altitude increase, 300–400 km).

6) Upper-stage and separated apogee kick motors used for geostationary satellites should be inserted into a disposal orbit at least 300 km above the geostationary orbit.

Second category options aim for removing upper stages and dead spacecraft from orbit. This can be accomplished with deorbiting maneuvers to ensure

atmospheric entry over ocean areas. Debris-control options in category three require new developments in which, in general, technical feasibility and cost-effectiveness must be demonstrated. Installation of drag-enhancement devices, use of lasers, debris sweeps, and so forth, fall into this category.

In 1995, NASA published end-of-life disposal guidelines for GEO, MEO, and LEO satellites (Ref. 36). Later, these guidelines were reviewed and accepted by the U.S. Air Force, and they became U.S. government guidelines. They are being discussed at the IADC (Inter Agency Space Debris Coordination Committee) meetings, which are being held once to twice a year. The U.S. guidelines may be summarized as follows.

GEO: Maneuver to an orbit with perigee altitude above the geostationary altitude by a distance no less than 235 + 1000 S (A/m) km. S is the solar radiation reflection index, between 1 and 2; and A/m is the area-to-mass ratio of the spacecraft in m^2/Kg.

MEO: Maneuver to a storage orbit between LEO and GEO with apogee altitude below 35,288 km (500 km below GEO altitude) and with perigee altitude above 2500 km. The storage altitude range must exclude the GPS operational zone (GPS reference altitude +/−500 km).

LEO: a) Atmospheric reentry option: Leave the structure in an orbit in which, using conservative projections for solar activity, atmospheric drag will limit the lifetime to no longer than 25 years after completion of mission. If drag enhancement devices are to be used to reduce the orbit lifetime, it should be demonstrated that such devices will significantly reduce the area-time product of the system or will not cause spacecraft or large debris to fragment if a collision occurs while the system is decaying from orbit, b) Maneuver to a storage orbit between LEO and GEO (see MEO), c) Direct retrieval: Retrieve the structure and remove it from orbit within 10 years after completion of mission.

Related GEO orbit studies can be found in Refs. 16–18 and 30–42. Results of long-term (100 to 200 year) stability of GPS disposal orbits revealed a large eccentricity growth, up to 0.5, after 140 years (Refs. 41 and 42). The eccentricity growth is caused by resonance due to sun-moon perturbations interacting with nodal regression resulting from Earth oblateness. Similar effects are found in the inactive GLONASS satellite orbits. Targeting the proper initial argument of perigee and small eccentricity in each GPS orbit plane can significantly reduce eccentricity growth.

13.11 Shielding

Increasing large amounts of space debris in the LEO and GEO environments impacts the design of future space missions. The need to protect spacecraft from debris collisions has prompted a look at necessary design considerations to enhance spacecraft survivability. These design considerations can be categorized into passive and active collision-protection measures.

Current practices in the area of designing passive collision-protection measures have been centered on the concept of shielding entire spacecraft and important spacecraft subsystems. Designs have evolved from the multiple plate shielding concepts of the early 1960s through the present. Some designs, in connection with the space station, use a stand-off aluminum shield approximately 0.16-cm thick. The structural wall is covered with 30 layers of Mylar thermal insulation. This

design has been shown to protect against collision of debris particles up to 0.5 cm in size.

Future work on this concept includes studying the performance of different stand-off shielding materials, additional shield layers, and various shield thicknesses. However, lightweight aluminum alloys and possible suitable composite materials are required to minimize the additional spacecraft weight from the shielding. Research is also needed to determine the effect of shield stand-off distance on shield performance. Theory developed for hypersonic impacts indicates that the protection achieved is relatively insensitive to increases in the shield stand-off distance or thickness and that the size of the debris that can be stopped increases directly with the structural wall thickness. Studies of this nature would be helpful in determining design strategies for optimizing added spacecraft weight and volume required for shielding.

Variations on the shielding methods are being considered for hard-to-protect spacecraft subsystems such as solar arrays and sensitive optics. One concept uses a cascade of stationary louvers to block impacting debris. A disadvantage of this is that the louvers may partially cover solar arrays or block regions of a sensor's field of view. An alternative method that would eliminate these performance problems is a system of active louvers or shutters that could be moved into place to protect delicate equipment in the event of a collision. A similar idea utilizes a "turtle-shell" spacecraft concept. This type of spacecraft would consist of a main protective structure, with ports through which sensors and arrays could be deployed and later retrieved into the protective structure in the event of a collision. These latter two design variations would require the spacecraft to have long-range (>100-km) space debris acquisition and tracking capabilities that could alert the spacecraft of an impending collision in time to take defensive measures.

Still other important considerations are the directionality effects of impacting debris. As has been found in the LDEF experiment, the leading edge of the spacecraft received 10–20 times the number of impacts received by the trailing edge.[34]

13.12 Collision Avoidance

If a large increase in space activity occurs, precautions against collision with existing debris are likely to be required. The most effective precaution would be to have collision avoidance between the structures and the existing large debris. This could reduce, or even eliminate, most fragment-producing collisions and the possibility of catastrophic collisions.

New requirements for collision avoidance would involve the following:

1) Detection of particles.
2) Orbit determination.
3) Preparation of spacecraft for rapid acceleration.
4) Maneuver execution.
5) Return of spacecraft to normal operation.

Spacecraft weight penalty for collision avoidance would be a function of 1) the distance of debris detection, and 2) the specific impulse of the maneuvering rocket system. For objects as small as 0.5 cm in diameter, radar sensors will require millimeter wavelengths, considerable transmitter power, and sizable antenna apertures. Passive sensors could also be employed, including visible and

infrared (IR) types. Another consideration will be the warning time available for maneuver. Detection ranges of as little as 50 km may be possible. The resultant loading on the spacecraft could be quite high and could cause problems for spacecraft with extended lightweight structures such as solar arrays or antennas. Longer warning time may permit reorientation or even reconfiguration of spacecraft in order to reduce the probability of collision or to protect sensitive elements.

For particles larger than 0.5–1 cm, the weight of protective shielding becomes prohibitive, and active protection measures must be considered. Like the two design variations mentioned earlier, all active protection measures require the spacecraft to have knowledge of its environment to at least 100 km. Debris sensing can be accomplished in a number of different ways.

The method nearest to availability is passive sensing using optical or IR sensors. For particles illuminated by the sun (such as those in a high Earth orbit), a visible light sensor with a 0.1-m^2 aperture would be required to sense a 1.0-cm object at 100 km, based on a 0.1 albedo. For particles hidden by Earth's shadow, an infrared sensor with a 0.5-m^2 aperture would provide adequate detection at a range of 100 km.

Onboard radar is a second option. However, the power requirements are beyond the scope of most present-day spacecraft. For instance, for detection of a 0.05-cm debris particle at 100 km, with a coverage of 4π sr, a power-aperture requirement of 100 million W-m^2 would be required. Although such a value is not inconceivable, it would dominate most spacecraft. In both the cases of passive sensing and onboard radar, accurate debris particle range and range-rate determination would need to be made within 1–2 s after acquisition. Based on acquisition at 100 km, this would leave 4–8 s for the spacecraft to complete an avoidance maneuver before impact.

A third concept for sensing debris utilizes a space-based radar system for tracking particles and relaying their position and velocity to other spacecraft. This concept has the advantage over onboard sensing methods of being able to give an early warning to spacecraft. With a space-based radar system, orbit ephemeris for smaller particles could be determined accurately. Spacecraft in danger of collision at some future time could then be identified and repositioned prior to a collision.

With warning available at a range of at least 100 km, two types of avoidance maneuvers can be performed. With extremely accurate sensing capabilities, the spacecraft could merely be reoriented by the reaction control system (RCS) to avoid collision. More realistically, the spacecraft could reposition itself into another orbit, thereby avoiding collision. In the latter case, however, repositioning the spacecraft is no guarantee of safety. In the case of a spacecraft encountering a debris cloud, for example, repositioning from one orbit to another may only move the spacecraft out of one possible collision and into another.

The main preliminary design considerations for active collision-avoidance maneuvers are increased structural needs due to dynamic loading and increased propulsive needs. The acceleration necessary to move the vehicle from a collision in the 5–10 s available after acquisition is proportional to the spacecraft size. It must, however, be restricted to within the limiting load factor of the spacecraft's structure. For example, the g loading due to acceleration of a 20-m spacecraft may be as high as $0.6g$. This acceleration can pose significant problems for structures that have lightweight elements such as solar arrays or larger antenna systems.

The propulsion requirements for each avoidance maneuver are significant. On the average, 50 kg of fuel could be expended per avoidance maneuver based on a generic 10-tonne spacecraft. Both structural loading and propulsion requirements increase with decreasing detection range. For propellant, the amount required increases with the inverse square of the detection range.

Once a better understanding of the debris environment is gained, national and international agreements should be reached to control the space debris problem and ensure safe future use of space for all mankind. For additional information on the subject of space debris, see Refs. 43–55.

References

[1] "Report on Orbital Debris," Interagency Group (Space) for National Security Council, Washington, DC, Feb. 1989.

[2]Chobotov, V. A., and Jenkin, A. B., "Analysis of the Micrometeoroid and Debris Hazard Posed to an Orbiting Parabolic Mirror," *50th International Astronautical Congress,* Amsterdam, 4–8 Oct. 1999.

[3]Chobotov, V. A., "An Overview of Space Debris Research at The Aerospace Corporation 1980–2000," Aerospace Rept. ATR-2001(9637)-1, Dec. 2000.

[4]Johnson, N. L., and McKnight, D. S., *Artificial Space Debris,* Krieger Publishing, Malabar, FL, 1987.

[5]Kessler, D. J., "Orbital Debris-Technical Issues," presentation to the USAF Scientific Advisory Board on Space Debris, Jan. 1987.

[6]Chobotov, V. A., and Wolfe, M. G., "The Dynamics of Orbiting Debris and the Impact on Expanded Operations in Space," *Journal of the Astronautical Sciences,* Vol. 38, Jan.–March 1990.

[7]*Orbital Debris—A Technical Assessment,* National Academy Press, 1995.

[8]Atkinson, D. R., Watts, A. J., and Crowell, L., "Spacecraft Microparticle Impact Flux Definition," Final Report for University of California, Lawrence Livemore National Laboratory, UCRL-RC-108788, 30 Aug. 1991.

[9]Perek, L., "The Scientific and Technical Aspects of the Geostationary Orbit," *38th IAF Congress,* IAA Paper 87-635, 1987.

[10]Hechler, M., and Vanderha, J. C., "Probability of Collisions in the Geostationary Ring," *Journal of Spacecraft and Rockets,* Vol. 18, July–Aug. 1981.

[11]Chobotov, V. A., "The Collision Hazard in Space," *Journal of the Astronautical Sciences,* July–Sept. 1982, pp. 191–212.

[12]Chobotov, V. A., "Classification of Orbits with Regard to Collision Hazard in Space," *Journal of Spacecraft and Rockets,* Vol. 20, Sept.–Oct. 1983, pp. 484–490.

[13]Perek, L., "Safety in the Geostationary Orbit After 1988," *40th IAF Congress,* IAF Paper 89-632, 1989.

[14]Bird, A. G., "Special Considerations for GEO-ESA," AIAA Paper 90-1361, Baltimore, MD, 1990.

[15]Fenoglio, L., and Flury, W., "Long-Term Evolution of Geostationary and Near-Geostationary Orbits," ESA/ESOC, Darmstadt, Germany, MAS Working Paper 260-1987.

[16]Chobotov, V. A., "Disposal of Spacecraft at End of Life in Geosynchronous Orbit," *Journal of Spacecraft and Rockets,* Vol. 27, No. 4, 1990, pp. 433–437.

[17]Yasaka, T., and Oda, S., "Classification of Debris Orbits With Regard to Collision Hazard in Geostationary Region," *41st IAF Congress,* AIAA Paper 90-571, 1990.

[18]Chobotov, V. A., and Johnson, C. G., "Effects of Satellite Bunching on the Probability of Collision in Geosynchronous Orbit," *Journal of Spacecraft and Rockets,* Vol. 31, No. 5, Sept.–Oct. 1994, pp. 895–899.

[19]Dennis, N. G., "Probabilistic Theory and Statistical Distribution of Earth Satellites," *Journal of British Interplanetary Society,* Vol. 25, 1972, pp. 333–376.

[20]Gist, R. G., and Oltrogge, D. L., "Collision Vision: Situational Awareness for Satellite and Reliable Space Operations," *50th International Astronautical Congress,* Amsterdam 4–8 Oct. 1999.

[21]Nazarenko, A. I., and Chobotov, V. A., "The Investigation of Possible Approaches of Cataloged Space Objects to Manned Spacecraft," *Space Debris,* Vol. 1, No. 2, pp. 127–142, 1999 (Published in 2000, Kluwer, Netherlands).

[22]Vedder, J. D., and Tabor, J. L., "New Method for Estimating Low Earth Orbit Collision Probabilities," *Journal of Spacecraft and Rockets,* Vol. 28, No. 2, March–April 1991, pp. 210–215.

[23]Chobotov, V. A., Herman, D. E., and Johnson, C. G., "Collision and Debris Hazard Assessment for a Low-Earth-Orbit Space Constellation," *Journal of Spacecraft and Rockets,* Vol. 34, No. 2, March–April 1997, pp. 233–238.

[24]Chobotov, V. A., and Mains, D. L., "Tether Satellite Systems Collision Study," *Space Debris,* Vol. 1, No. 2, 1999.

[25]Patera, R. P., "A Method for Calculating Collision Probability Between a Satellite and a Space Tether," Paper No. AAS 01-116, Feb. 2001.

[26]Technical Report on Space Debris, United Nations, New York, 1999.

[27]Su, S. Y., and Kessler, D. J., "Contribution of Explosion and Future Collision Fragments to the Orbital Debris Environment," COSPAR, Graz, Austria, June 1984.

[28]Chobotov. V. A., "Dynamics of Orbital Debris Clouds and the Resulting Collision Hazard to Spacecraft," *Journal of the British Interplanetary Society,* Vol. 43, May 1990, pp. 187–195.

[29]Jenkin, A. B., "DEBRIS: A Computer Program for Debris Cloud Modeling," *44th Congress of IAF,* 16–22 Oct. 1993, Graz, Austria (AIAA 6.3-93-746).

[30]Kessler, D. J., Reynolds, R. C., and Anz-Meador, P. D., "Orbital Debris Environment for Spacecraft Designed to Operate in Low Earth Orbit," NASA TM 100-471, April 1988.

[31]Sorge, M. E., and Johnson, C. G., Space Debris Hazard Software: Program Impact Version 3.0 User's Guide, Aerospace Corp. TOR-93(3076)-3, Aug. 1993.

[32]Patera, R. P., and Ailor, W. H., "The Realities of Reentry Disposal," AAS Paper 98-174, Feb. 1998.

[33]*Proceedings of the First European Conference on Space Debris,* Darmstadt, Germany, 5–7 April 1993, ESA SD-01.

[34]McDonnell, J.A.M. (ed.), *Hypervelocity Impacts in Space,* Univ. of Kent, England, 1992.

[35] "Position Paper on Orbital Debris," *International Academy of Astronautics,* 8 March 1993.

[36]*Guidelines and Assessment Procedures for Limiting Orbital Debris,* NASA Safety Standard 1740.14, Office of Safety and Mission Assurance, Aug. 1995.

[37]Newman, L. K., and Folta, D. C., "Evaluation of Spacecraft End-Of-Life Disposal to Meet NASA Management Instruction (NMI) Guidelines," AAS/AIAA Paper No. 95-325, Aug. 1995.

[38]Eichler, P., Reynolds, R., Zhang, J., Bade, A., Jackson, A., Johnson, N., and McNamara, R., "Post Mission Disposal Options for Upper Stages," *Proceedings of Society of Photo-Optical Instrumentation Engineers,* Vol. 3116, San Diego, CA, Nov. 1997, pp. 221–234.

[39]Chao, C. C., "Geosynchronous Disposal Orbit Stability," AIAA Paper 98-4186, Aug. 1998.

[40]Meyer, K. W., and Chao, C. C., "Atmospheric Reentry Disposal for Low-Altitude Spacecraft," *Journal of Spacecraft and Rockets,* Vol. 37, No. 5, Sept.–Oct. 2000.

[41]Chao, C. C., "MEO Disposal Orbit Stability and Direct Reentry Strategy," AAS Paper No. 00-152, Jan. 2000.

[42]Gick, R. A., and Chao, C. C., "GPS Disposal Orbit Stability and Sensitivity Study," AAS Paper 01-244, Feb. 2001.

[43]Portree, D.S.F., and Loftus, J. P., Jr., "Orbital Debris and Near-Earth Environmental Management: A chronology," NASA.

[44]Eichler, P., and Rex, D., "Debris Chain Reactions," AIAA/NASA/DOD Orbital Debris Convergence, AIAA Paper 90-1365, Baltimore, MD, April 1990.

[45]Smirnov, N. N., Lebedev, V. V., and Kiselev, A. B., "Mathematical Modeling of Space Debris Evolution in Low Earth Orbit," *19th ISTS,* Yokohama, Japan, May 1994.

[46]Nazarenko, A. I., "Prediction and Analysis of Orbital Debris Environment Evolution," First European Conference on Space Debris, Darmstadt, Germany, Apr. 1993.

[47]Chernyavskiy, A. G., Chernyavskiy, G. M., Johnson, N., and McKnight, D., "A Simple Case of Space Environmental Effects," *44th International Astronautical Federation Congress,* Graz, Austria, Oct. 1993.

[48]Maclay, T. D., Madler, R. A., McNamara, R., and Culp, R. D., "Orbital Debris Hazard Analysis for Long Term Space Assets," *Proceedings of the Workshop on Hypervelocity Impacts in Space,* Univ. of Kent, Canterbury, U. K., July 1991.

[49]Culp, R. D. et al., "Orbital Debris Studies at the University of Colorado," First European Conference on Space Debris, Darmstadt, Germany, Apr. 1993.

[50]Veniaminov, S. S., "The Methods and Experience of Detecting Small and Weakly Contrasting Space Objects," First European Conference on Space Debris, Darmstadt, Germany, Apr. 1993.

[51]Swinerd, G. G., Barrows, S. P., and Crowther, R., "Short-term debris risk to large satellite constellations," *Journal of Guidance, Control, and Dynamics,* Vol. 22, No. 2, 1999, pp. 291–295.

[52]Swinerd, G. G., Lewis, H. L., Williams, N., and Martin, C., "Self-induced collision hazard in high and moderate inclination satellite constellations," Paper No. IAA-00-IAA.6.6.01, *51st International Astronautical Congress,* Rio de Janeiro, Brazil, 2–6, Oct. 2000.

[53]Rossi, A., Anselmo, L., Pardini, C., Cordelli, A., Farinella, P., and Parrinello, T., "Approaching the Exponential Growth: Parameter Sensitivity of the Debris Evolution," *Proceedings of the First European Conference on Space Debris,* Darmstadt, Germany, 5–7 April 1993, (ESA SD-01), pp. 287–292.

[54]Flury, W., (ed.), "Space Debris," *Advances in Space Research,* Vol. 13, No. 8, 1992.

[55]Toda, S., "Recent Space Debris Activities in Japan," *Earth Space Review,* Vol. 4, No. 3, 1995.

14
Optimal Low-Thrust Orbit Transfer

14.1 Introduction

The theory of optimal low-thrust orbit transfer has received a great deal of attention in the astrodynamics and flight mechanics literature over the past several decades. This chapter begins with a detailed description of some fundamental analytic results obtained by Edelbaum, which are widely in use by the practitioners in the aerospace industry. The reader, after becoming familiar with the simplified transfer analysis, is invited to consider the treatment of the exact transfer problem in the subsequent sections. Drawing on the pioneering work of the Americans Broucke, Cefola, and Edelbaum, who perfected the theory of orbital mechanics in terms of nonsingular orbital elements, examples of optimal orbit transfers are generated and discussed, and all the relevant equations needed to develop unconstrained orbit transfer computer codes are exposed and derived.

14.2 The Edelbaum Low-Thrust Orbit-Transfer Problem

A discussion of the problem of optimal low-thrust transfer between inclined circular orbits was presented by Edelbaum in the early 1960s.[1] Assuming constant acceleration and constant thrust vector yaw angle within each revolution, Edelbaum linearizes the Lagrange planetary equations of orbital motion about a circular orbit and, using the velocity as the independent variable, reduces the transfer optimization problem to a problem in the theory of maxima. The variational integral involves a single constant Lagrange multiplier since it involves a single integral constraint equation for the transfer time or velocity change while maximizing the change in orbital inclination. The control variable being the yaw angle, the necessary condition for a stationary solution is obtained by simply setting the partial derivative of the integrand of the variational integral with respect to the control to zero. This optimum control is then used in the right-hand sides of the original equations of motion, which are integrated analytically to provide expressions for the time and inclination in terms of the independent variable, the orbital velocity. Two expressions for the inclination are provided to cover the case of large inclination change transfers. This complication arises if orbital velocity is adopted as the independent variable. However, a single expression for the inclination change can be obtained that is uniformly valid throughout any desired transfer if the original Edelbaum problem is cast into a minimum-time transfer problem using the more direct formalism of optimal control theory. Following is a discussion of Edelbaum's original analysis, as well as the formulation using optimal control theory.

Edelbaum's Analysis

The full set of the Gaussian form of the Lagrange planetary equations for near-circular orbits is given by

$$\dot{a} = \frac{2af_t}{V} \tag{14.1}$$

$$\dot{e}_x = \frac{2f_t c_\alpha}{V} - \frac{f_n s_\alpha}{V} \tag{14.2}$$

$$\dot{e}_y = \frac{2f_t s_\alpha}{V} + \frac{f_n c_\alpha}{V} \tag{14.3}$$

$$\dot{i} = \frac{f_h c_\alpha}{V} \tag{14.4}$$

$$\dot{\Omega} = \frac{f_h s_\alpha}{V s_i} \tag{14.5}$$

$$\dot{\alpha} = n + \frac{2f_n}{V} - \frac{f_h s_\alpha}{V \tan i} \tag{14.6}$$

where s_α and c_α stand for $\sin\alpha$ and $\cos\alpha$, respectively, and a stands for the orbit semimajor axis, i for inclination, and Ω for the right ascension of the ascending node; $e_x = e\cos\omega$, and $e_y = e\sin\omega$, with e and ω standing for orbital eccentricity and argument of perigee. Finally, $\alpha = \omega + M$ represents the mean angular position, M the mean anomaly, and $n = (\mu/a^3)^{1/2}$ the orbit mean motion, with μ standing for the Earth gravity constant. For near-circular orbits, $V = na = (\mu/a)^{1/2}$. The components of the thrust acceleration vector along the tangent, normal, and out-of-plane directions are depicted by f_t, f_n, and f_h, with the normal direction oriented toward the center of attraction. If we assume only tangential and out-of-plane acceleration, and that the orbit remains circular during the transfer, the Eqs. (14.1–14.6) reduce to

$$\dot{a} = \frac{2af_t}{V} \tag{14.7}$$

$$\dot{i} = \frac{f_h c_\alpha}{V} \tag{14.8}$$

$$\dot{\Omega} = \frac{f_h s_\alpha}{V s_i} \tag{14.9}$$

$$\dot{\alpha} = n - \frac{f_h s_\alpha}{V \tan i} \tag{14.10}$$

If f represents the magnitude of the acceleration vector, and β the out-of-plane or thrust yaw angle, then $f_t = f c_\beta$ and $f_h = f s_\beta$. Furthermore, $\alpha = \omega + M = \omega + \theta^* = \theta$, the angular position when $e = 0$, with $\theta = nt$ and θ^* the true anomaly. If the angle β is held piecewise constant switching sign at the orbital antinodes, then the $f_h s_\alpha$ terms above in Eqs. (14.9) and (14.10) will have a net zero contribution such that the system of differential equations further reduces to

$$\dot{a} = \frac{2af_t}{V} \tag{14.11}$$

$$\dot{i} = \frac{c_\theta f_h}{V} \tag{14.12}$$

$$\dot{\theta} = n \tag{14.13}$$

We can now average out the angular position θ in Eq. (14.12) by integrating with respect to θ and by holding f, β, and V constant

$$\int_0^{2\pi} \left(\frac{\mathrm{d}i}{\mathrm{d}t}\right)\mathrm{d}\theta = \frac{2fs_\beta}{V}\int_{-\pi/2}^{\pi/2} c_\theta \mathrm{d}\theta$$

$$2\pi\frac{\mathrm{d}i}{\mathrm{d}t} = \frac{4fs_\beta}{V}$$

$$\frac{\mathrm{d}i}{\mathrm{d}t} = \frac{2fs_\beta}{\pi V} \tag{14.14}$$

From the energy equation $V^2/2 - \mu/r = -\mu/2a$, with $r = a$, and with Eq. (14.11) used to eliminate the semimajor axis,

$$\mathrm{d}V = -\left[\frac{\mu}{2Va^2}\right]\mathrm{d}a$$

$$= -fc_\beta \mathrm{d}t$$

$$\frac{\mathrm{d}V}{\mathrm{d}t} = -fc_\beta \tag{14.15}$$

Equation (14.14) can also be obtained by dividing Eq. (14.12) by Eq. (14.13),

$$\frac{\mathrm{d}i}{\mathrm{d}\theta} = \frac{c_\theta f_h}{Vn}$$

$$\Delta i = \frac{2f_h}{Vn}\int_{-\pi/2}^{\pi/2} c_\theta \mathrm{d}\theta = \frac{4f_h}{Vn}$$

and, since $\Delta t = 2\pi\, a/V$,

$$\frac{\mathrm{d}i}{\mathrm{d}t} = \frac{\Delta i}{\Delta t} = \frac{2f_h}{V\pi}$$

Equations (14.14) and (14.15) can be replaced by the following set, where V is now the independent variable,

$$\frac{\mathrm{d}i}{\mathrm{d}V} = -\frac{2\tan\beta}{\pi V} \tag{14.16}$$

$$\frac{\mathrm{d}t}{\mathrm{d}V} = -\frac{1}{fc_\beta} \tag{14.17}$$

Let I represent the functional to be maximized,

$$I = \int_{V_0}^{V_f}\left(\frac{\mathrm{d}i}{\mathrm{d}V}\right)\mathrm{d}V = -\int_{V_0}^{V_f}\frac{2}{\pi V}\tan\beta\,\mathrm{d}V \tag{14.18}$$

and let J represent the integral constraint given by

$$J = \int_{V_0}^{V_f} \left(\frac{\mathrm{d}t}{\mathrm{d}V}\right) \mathrm{d}V = \text{const} \tag{14.19}$$

Let us adjoin J to I by way of a constant Lagrange multiplier λ such that the optimization problem is now reduced to a succession of ordinary maximum problems for each value of V between V_0 and V_f, the initial and final velocities, respectively. The necessary condition for a stationary solution of the augmented integral,

$$K = I + \lambda J = \int_{V_0}^{V_f} \left[-\frac{2}{\pi V} \tan\beta - \frac{\lambda}{f c_\beta}\right] \mathrm{d}V \tag{14.20}$$

is then simply given by

$$\frac{\partial}{\partial \beta}\left[\frac{2}{\pi V} \tan\beta + \frac{\lambda}{f c_\beta}\right] = 0 \tag{14.21}$$

The optimization problem consists, therefore, of the maximization of the inclination change subject to the constraint of given total transfer time since

$$\Delta V = f t \tag{14.22}$$

This constraint is equivalent to the fixed ΔV constraint for constant acceleration f. Furthermore, V_0 and V_f being given, the initial and final radii are, therefore, given too since the orbits are assumed circular. With the acceleration being applied continuously, this problem is equivalent to minimizing the total transfer time for a given change in the inclination and velocity. This is also equivalent to minimizing the total ΔV or propellant usage because the thrust is always on and no coasting arcs are allowed. In this latter case, I and J are simply interchanged to yield the optimality condition

$$\frac{\partial}{\partial \beta}\left[\frac{1}{f c_\beta} + \lambda_i \frac{2}{\pi V} \tan\beta\right] = 0 \tag{14.23}$$

From Eq. (14.21), it follows that

$$V s_\beta = -\frac{2f}{\pi \lambda} = \text{const} = V_0 s_{\beta_0} \tag{14.24}$$

$$\lambda = -\frac{2f}{\pi V_0 s_{\beta_0}} \tag{14.25}$$

The optimal β steering law given by Eq. (14.24) can be used in Eq. (14.17) for $\mathrm{d}V/\mathrm{d}t$ in order to obtain the expression for the velocity as a function of time t or $\Delta V = f t$.

$$\frac{\mathrm{d}V}{\mathrm{d}t} = -f c_\beta$$

$$f\,\mathrm{d}t = -\frac{\mathrm{d}V}{c_\beta} = \frac{-\mathrm{d}V}{\pm\left(1 - s_\beta^2\right)^{1/2}}$$

$$\int_0^t f\,\mathrm{d}t = -\int_{V_0}^{V} \frac{V\mathrm{d}V}{\pm\left(V^2 - V_0^2 s_{\beta_0}^2\right)^{1/2}}$$

$$\Delta V = ft = -\frac{1}{\pm}\left[\left(V^2 - V_0^2 s_{\beta_0}^2\right)^{1/2} - (\pm)V_0 c_{\beta_0}\right]$$

$$\Delta V = V_0 c_{\beta_0} \mp \left(V^2 - V^2 s_\beta^2\right)^{1/2} = V_0 c_{\beta_0} \mp (\pm) V c_\beta \tag{14.26}$$

$$\Delta V = V_0 c_{\beta_0} - V c_\beta \tag{14.27}$$

From Eq. (14.26),

$$\Delta V - V_0 c_{\beta_0} = \mp\left(V^2 - V^2 s_\beta^2\right)^{1/2} = \mp\left(V^2 - V_0^2 s_{\beta_0}^2\right)^{1/2}$$

and, after squaring,

$$V^2 = V_0^2 + \Delta V^2 - 2\Delta V V_0 c_{\beta_0} \tag{14.28}$$

This then represents V as a function of time since $\Delta V = ft$. The initial yaw angle β_0 must still be determined. In a similar way, Eq. (14.16) for $\mathrm{d}i/\mathrm{d}V$ can be integrated to provide an expression for the evolution of the inclination in time.

$$\frac{\mathrm{d}i}{\mathrm{d}V} = -\frac{2}{\pi V}\tan\beta = -\frac{2}{\pi V}\frac{s_\beta}{c_\beta} = -\frac{2}{\pi V}\frac{V s_\beta}{\left(V^2 - V^2 s_\beta^2\right)^{1/2}}$$

such that, with the use of $V s_\beta = V_0 s_{\beta_0}$,

$$\int_{i_0}^{i} \mathrm{d}i = -\frac{2}{\pi} V_0 s_{\beta_0} \int_{V_0}^{V} \frac{\mathrm{d}V}{V\left(V^2 - V_0^2 s_{\beta_0}^2\right)^{1/2}}$$

$$\Delta i = -\frac{2}{\pi}\sin^{-1}\left(\frac{-V_0 s_{\beta_0}}{V}\right)\Bigg|_{V_0}^{V}$$

$$\Delta i = -\frac{2}{\pi}\left[\sin^{-1}(s_{\beta_0}) - \sin^{-1}\left(\frac{V_0 s_{\beta_0}}{V}\right)\right]$$

$$\Delta i = -\frac{2}{\pi}\beta_0 + \frac{2}{\pi}\sin^{-1}\left(\frac{V_0 s_{\beta_0}}{V}\right) \tag{14.29}$$

and, since $V_0 s_{\beta_0} = V s_\beta$,

$$\Delta i = -\frac{2}{\pi}(\beta - \beta_0)$$

Now, since the inverse sine function in Eq. (14.29) is double-valued in the interval $(0, 2\pi)$, it is necessary to write this function as

$$\sin^{-1}\left(\frac{V_0 s_{\beta_0}}{V}\right) \text{ if } \sin^{-1}\left(\frac{V_0 s_{\beta_0}}{V}\right) < \frac{\pi}{2}$$

and

$$\frac{\pi}{2} + \left[\frac{\pi}{2} - \sin^{-1}\left(\frac{V_0 s_{\beta_0}}{V}\right)\right] = \pi - \sin^{-1}\left(\frac{V_0 s_{\beta_0}}{V}\right) \text{ if } \sin^{-1}\left(\frac{V_0 s_{\beta_0}}{V}\right) > \frac{\pi}{2}$$

since the function is symmetrical with respect to $\pi/2$. In the second of the preceding conditions, Δi can be written as

$$\Delta i = \frac{2}{\pi}\left[\pi - \sin^{-1}\left(\frac{V_0 s_{\beta_0}}{V}\right)\right] - \frac{2}{\pi}\beta_0 \tag{14.30}$$

or

$$\Delta i = 2 - \frac{2}{\pi}\sin^{-1}\left(\frac{V_0 s_{\beta_0}}{V}\right) - \frac{2}{\pi}\beta_0 \tag{14.31}$$

This is equivalent to writing Eqs. (14.29) and (14.31) as

$$\Delta i = \frac{2}{\pi}(\beta - \beta_0) \quad \text{if } \beta < \frac{\pi}{2} \tag{14.32}$$

$$\Delta i = 2 - \frac{2}{\pi}(\beta + \beta_0) \quad \text{if } \beta > \frac{\pi}{2} \tag{14.33}$$

Of course, the 2 in Eq. (14.33) is given in radians, and it corresponds to 114.6 deg.

Finally, from Eq. (14.26),

$$\Delta V = V_0 c_{\beta_0} - \left(V^2 - V_0^2 s_{\beta_0}^2\right)^{1/2} \quad \text{if } \Delta V - V_0 c_{\beta_0} < 0 \tag{14.34}$$

$$\Delta V = V_0 c_{\beta_0} + \left(V^2 - V_0^2 s_{\beta_0}^2\right)^{1/2} \quad \text{if } \Delta V - V_0 c_{\beta_0} > 0 \tag{14.35}$$

From $\Delta V = V_0 c_{\beta_0} - V c_\beta$, the condition $\Delta V - V_0 c_{\beta_0} < 0$ is identical to $c_\beta > 0$ or $\beta < \pi/2$ or $\sin^{-1}(V_0 s_{\beta_0}/V) < \pi/2$, and the condition $\Delta V - V_0 c_{\beta_0} > 0$ is identical to $\beta > \pi/2$ or $\sin^{-1}(V_0 s_{\beta_0}/V) > \pi/2$ such that the Edelbaum analysis leads to the following set of equations:

1) If $\Delta V - V_0 c_{\beta_0} < 0$, then,

$$\left.\begin{aligned} V &= \left(V_0^2 - 2V_0 \Delta V c_{\beta_0} + \Delta V^2\right)^{1/2} \\ \Delta i &= \frac{2}{\pi}\sin^{-1}\left(\frac{V_0 s_{\beta_0}}{V}\right) - \frac{2}{\pi}\beta_0 = \frac{2}{\pi}(\beta - \beta_0) \end{aligned}\right\} \tag{14.36}$$

2) If $\Delta V - V_0 c_{\beta_0} > 0$, then,

$$\left.\begin{aligned} V &= \left(V_0^2 - 2V_0 \Delta V c_{\beta_0} + \Delta V^2\right)^{1/2} \\ \Delta i &= 2 - \frac{2}{\pi}\sin^{-1}\left(\frac{V_0 s_{\beta_0}}{V}\right) - \frac{2}{\pi}\beta_0 = 2 - \frac{2}{\pi}(\beta + \beta_0) \end{aligned}\right\} \tag{14.37}$$

The preceding equations show that one must monitor the condition $\Delta V - V_0 c_{\beta_0}$ and use Eq. (14.36) to describe the transfer starting from time 0 and later switch to Eqs. (14.37) as soon as $t = \Delta V/f$ exceeds $V_0 c_{\beta_0}/f$, which will take place for large transfers as will be shown later by an example. For large transfers, ΔV as given in Eq. (14.26) could become double-valued in V such that one must use $\Delta V = V_0 c_{\beta_0} - (V^2 - V_0^2 s_{\beta_0}^2)^{1/2}$ from V_0 to $V_0 s_{\beta_0}$ and $\Delta V = V_0 c_{\beta_0} + (V^2 - V_0^2 s_{\beta_0}^2)^{1/2}$ from $V = V_0 s_{\beta_0}$ to V_f, where $V_0 s_{\beta_0} < V_f$. This minimum velocity takes place when $\Delta V = V_0 c_{\beta_0}$, indicating that the orbit will grow to become larger than the final desired orbit and later shrink to that desired orbit. This will happen when larger inclination changes are required since then the orbit plane rotation will be carried out mostly at those higher intermediate altitudes. This, of course, is the result of the trade between inclination and radius or velocity. From

$$\Delta V = V_0 c_{\beta_0} \mp \left(V^2 - V_0^2 s_{\beta_0}^2\right)^{1/2}$$

we have

$$\frac{\partial \Delta V}{\partial V} = \mp \frac{V}{\left(V^2 - V_0^2 s_{\beta_0}^2\right)^{1/2}}$$

which is equal to ∞ for $V = V_0 s_{\beta_0}$, the minimum velocity reached. The initial yaw angle β_0 can be obtained from the terminal conditions at time t_f. At $t = t_f$, $V = V_f$ and $\Delta i = \Delta i_f$. Using Eq. (14.36) for Δi, we get

$$\frac{\pi}{2}\Delta i_f + \beta_0 = \sin^{-1}\left(\frac{V_0 s_{\beta_0}}{V_f}\right)$$

$$\sin\left(\beta_0 + \frac{\pi}{2}\Delta i_f\right) = \frac{V_0 s_{\beta_0}}{V_f}$$

$$\sin\beta_0 \cos\frac{\pi}{2}\Delta i_f + \cos\beta_0 \sin\frac{\pi}{2}\Delta i_f = \frac{V_0 s_{\beta_0}}{V_f}$$

Dividing both sides by c_{β_0} yields

$$\left[\cos\frac{\pi}{2}\Delta i_f - \frac{V_0}{V_f}\right]\tan\beta_0 = -\sin\frac{\pi}{2}\Delta i_f$$

$$\tan\beta_0 = \frac{\sin\frac{\pi}{2}\Delta i_f}{\frac{V_0}{V_f} - \cos\frac{\pi}{2}\Delta i_f} \tag{14.38}$$

Now, carrying out the same manipulations using Eq. (14.37), we get

$$\frac{\pi}{2}\left(\Delta i_f - 2 + \frac{2}{\pi}\beta_0\right) = -\sin^{-1}\left(\frac{V_0 s_{\beta_0}}{V_f}\right)$$

$$\pi - \left(\beta_0 + \frac{\pi}{2}\Delta i_f\right) = \sin^{-1}\left(\frac{V_0 s_{\beta_0}}{V_f}\right)$$

$$\sin\left(\beta_0 + \frac{\pi}{2}\Delta i_f\right) = \frac{V_0 s_{\beta_0}}{V_f}$$

$$\sin\beta_0 \cos\frac{\pi}{2}\Delta i_f + c_{\beta_0}\sin\frac{\pi}{2}\Delta i_f = \frac{V_0 s_{\beta_0}}{V_f}$$

Dividing by c_{β_0}

$$\tan\beta_0 = \frac{\sin\frac{\pi}{2}\Delta i_f}{\frac{V_0}{V_f} - \cos\frac{\pi}{2}\Delta i_f} \tag{14.39}$$

Equations (14.38) and (14.39) indicate that β_0 is given by

$$\beta_0 = \tan^{-1}\left[\frac{\sin\frac{\pi}{2}\Delta i_f}{\frac{V_0}{V_f} - \cos\frac{\pi}{2}\Delta i_f}\right] \tag{14.40}$$

regardless of whether $\Delta V - V_0 c_{\beta_0} < 0$ or $\Delta V - V_0 c_{\beta_0} > 0$ and, from $\Delta V = V_0 c_{\beta_0} - V c_\beta$, the yaw angle β at future times is given by

$$\beta = \cos^{-1}\left[\frac{V_0 c_{\beta_0} - \Delta V}{\left(V_0^2 - 2V_0\Delta V c_{\beta_0} + \Delta V^2\right)^{1/2}}\right] \tag{14.41}$$

where $\Delta V = ft$ and where $0 \le \beta \le \pi$. This expression is better than $V s_\beta = V_0 s_{\beta_0}$, which would yield $\beta = \sin^{-1}(V_0 s_{\beta_0}/V)$ since β could, for large transfers, exceed $\pi/2$.

If the evolution of Δi as a function of time or velocity or ΔV is desired, Eq. (14.36) for Δi is used until $\Delta V = V_0 c_{\beta_0}$. When $\Delta V - V_0 c_{\beta_0} > 0$, Δi as given in Eq. (14.37) is used. However, in Eq. (14.37), the inverse sine function will always return a β angle that is always less than $\pi/2$, and this value for β is the correct value to be used in $\Delta i = 2 - (2/\pi)(\beta + \beta_0)$. This β angle is clearly not the real yaw angle since, in this case, it would be given by $\pi - \beta$ with $\beta < \pi/2$, such that the yaw angle is now larger than $\pi/2$. If the real β angle is used in $\Delta i = 2 - (2/\pi)(\beta + \beta_0)$, we get

$$\Delta i = 2 - \frac{2}{\pi}(\pi - \beta + \beta_0) = \frac{2}{\pi}(\beta - \beta_0) \tag{14.42}$$

Equation (14.42) is universally valid for all yaw angles $0 \le \beta < 180$ deg and should be the only one used. Equation (14.42) will effectively replace Eqs. (14.36) and (14.37), provided that the angle β is computed from Eq. (14.41). Since the sign of $\Delta V - V_0 c_{\beta_0}$ is effectively accounted for in Eq. (14.41), it will return the yaw angle β to be used in Eq. (14.42) for the unambiguous evaluation of Δi. Expressions for β_0 and β can also be obtained by using the identity in Eq. (14.42) since

$$\cos\frac{\pi}{2}\Delta i = c_\beta c_{\beta_0} + s_\beta s_{\beta_0} \tag{14.43}$$

$$\sin\frac{\pi}{2}\Delta i = s_\beta c_{\beta_0} - s_{\beta_0} c_\beta \tag{14.44}$$

If we multiply Eq. (14.43) by VV_0 and replace Vs_β by $V_0 s_{\beta_0}$ and Vc_β by $V_0 c_{\beta_0} - \Delta V$, we get, after regrouping terms,

$$c_{\beta_0} = \frac{V_0 - V\cos\frac{\pi}{2}\Delta i}{\Delta V} \tag{14.45}$$

In a similar manner, from Eq. (14.43), if we replace this time $V_0 s_{\beta_0}$ by Vs_β and $V_0 c_{\beta_0}$ by $\Delta V + Vc_\beta$, we get an expression for c_β:

$$c_\beta = \frac{V_0\cos\frac{\pi}{2}\Delta i - V}{\Delta V} \tag{14.46}$$

Equation (14.44) can also be written as

$$VV_0\sin\frac{\pi}{2}\Delta i = Vs_\beta V_0 c_{\beta_0} - V_0 s_{\beta_0} Vc_\beta$$

If we replace Vs_β by $V_0 s_{\beta_0}$ and $V_0 c_{\beta_0}$ by $\Delta V + Vc_\beta$, then

$$s_{\beta_0} = \frac{V\sin\frac{\pi}{2}\Delta i}{\Delta V} \tag{14.47}$$

If, on the other hand, we replace $V_0 c_{\beta_0}$ by $\Delta V + Vc_\beta$, then the identity will yield

$$s_\beta = \frac{V_0\sin\frac{\pi}{2}\Delta i}{\Delta V} \tag{14.48}$$

Now these expressions will readily yield

$$\tan\beta_0 = \frac{V\sin\frac{\pi}{2}\Delta i}{V_0 - V\cos\frac{\pi}{2}\Delta i} \tag{14.49}$$

$$\tan\beta = \frac{V_0\sin\frac{\pi}{2}\Delta i}{V_0\cos\frac{\pi}{2}\Delta i - V} \tag{14.50}$$

These last two expressions can be used to obtain the initial β_0 and current β provided that the appropriate Δi expression of Eqs. (14.36) or (14.37) is used according to whether $\Delta V - V_0 c_{\beta_0}$ is positive or negative. Although Eqs. (14.49) and (14.50) are valid for any optimal $(V, \Delta i)$ pair during the transfer, there is clearly a singularity at time 0 when $V = V_0$ and $\Delta i = 0$. The angle β_0 is best obtained by setting $V = V_f$ and $\Delta i = \Delta i_f$ in Eq. (14.49). It is better to use Eq. (14.41) for the control time history instead of Eq. (14.50) since we do not have to switch between two Δi expressions to describe that evolution in the first case. Finally, Edelbaum's ΔV equation in terms of the velocities and inclination is obtained from the velocity equation

$$V = \left(V_0^2 - 2V_0\Delta V c_{\beta_0} + \Delta V^2\right)^{1/2}$$

If we square this expression, replace ΔV in the product term by $V_0 c_{\beta_0} - V c_\beta$, and then use the identity

$$c_\beta c_{\beta_0} = \frac{1}{2} c_{\beta-\beta_0} + \frac{1}{2} c_{\beta+\beta_0}$$

with $c_{\beta-\beta_0} = \cos \pi/2\Delta i$ from Eq. (14.42), then we get, with $V s_\beta = V_0 s_{\beta_0}$,

$$V^2 = -V_0^2 + V_0^2 s_{\beta_0}^2 + V V_0 \cos \frac{\pi}{2} \Delta i + V_0 c_{\beta_0} V c_\beta + \Delta V^2$$

However $V_0^2 s_{\beta_0}^2 = V_0 s_{\beta_0} V s_\beta$ and, if this term is combined with $V_0 c_{\beta_0} V c_\beta$, the result will be $V V_0 c_{\beta-\beta_0}$, which can be replaced by $V V_0 \cos \pi/2\Delta i$.

The final result is given by

$$V^2 = -V_0^2 + 2 V V_0 \cos \frac{\pi}{2} \Delta i + \Delta V^2$$

from which

$$\Delta V = \left(V_0^2 - 2 V V_0 \cos \frac{\pi}{2} \Delta i + V^2 \right)^{1/2} \tag{14.51}$$

This is Edelbaum's ΔV equation for constant-acceleration circle to inclined circle transfer. It is valid for any $(V, \Delta i)$ pair along the transfer, provided once again that the appropriate Δi expression is used, i.e., Eqs. (14.36) or (14.37) according to whether $\Delta V - V_0 c_{\beta_0}$ is <0 or >0, respectively. As shown earlier, ΔV is double-valued in the velocity since Δi itself is double-valued in that same variable. However, Eq. (14.51) is mainly used to obtain the total ΔV_{tot} required to achieve a given transfer between V_0 and V_f with a relative inclination change of Δi_f. It is valid for any $0 < \Delta i_f < 114.6$ deg or $0 < \Delta i_f < 2$ rad since this is the limiting Δi in Eq. (14.37). The transfer time t_f is simply obtained from

$$t_f = \frac{\Delta V_{\text{tot}}}{f} \tag{14.52}$$

Formulation Using Optimal Control Theory

Let the system equations be given by Eqs. (14.14) and (14.15), with time as the independent variable and i and V as the state variables. The yaw angle β is the control variable.

$$\frac{\mathrm{d}i}{\mathrm{d}t} = \frac{2}{\pi} \frac{f}{V} s_\beta \tag{14.53}$$

$$\frac{\mathrm{d}V}{\mathrm{d}t} = -f c_\beta \tag{14.54}$$

This problem is now cast as a minimum time transfer problem between initial and final parameters i_0, V_0 and i_f, V_f, respectively. The variational Hamiltonian

is then given by

$$H = 1 + \lambda_i \left(\frac{2}{\pi} \frac{f}{V} s_\beta \right) + \lambda_V (-f c_\beta) \tag{14.55}$$

since the performance index is simply given by

$$J = \int_{t_0}^{t_f} L \mathrm{d}t$$

with $L = 1$. The Euler–Lagrange differential equations are given by

$$\dot{\lambda}_V = -\frac{\partial H}{\partial V} = \frac{2}{\pi} \frac{f s_\beta}{V^2} \lambda_i \tag{14.56}$$

$$\dot{\lambda}_i = -\frac{\partial H}{\partial i} = 0 \tag{14.57}$$

Therefore, λ_i is a constant. The optimality condition is given by

$$\frac{\partial H}{\partial \beta} = \lambda_i \frac{2}{\pi} \frac{f}{V} c_\beta + f \lambda_V s_\beta = 0 \tag{14.58}$$

which yields the optimal control law

$$\tan \beta = -\frac{2}{\pi} \frac{\lambda_i}{V \lambda_V} \tag{14.59}$$

There is no need to integrate $\dot{\lambda}_V$ since we can use the transversality condition $H_f = 0$ at the final time. The Hamiltonian is a constant of the motion since it is not an explicit function of time. Therefore, it is equal to zero all the time. We can therefore solve for λ_V and λ_i from

$$H = 0 = 1 + \frac{2}{\pi} \frac{f}{V} s_\beta \lambda_i - f c_\beta \lambda_V$$

$$\frac{\partial H}{\partial \beta} = 0 = \frac{2}{\pi} \frac{f}{V} c_\beta \lambda_i + f s_\beta \lambda_V$$

This results in

$$\lambda_i = -\frac{\pi s_\beta V}{2f} = \text{const} \tag{14.60}$$

$$\lambda_V = \frac{c_\beta}{f} \tag{14.61}$$

Equation (14.60) reveals that $V s_\beta = V_0 s_{\beta_0}$ since the acceleration f is assumed to be a constant. We can now take advantage of this constancy of $V s_\beta$ in order to

integrate the velocity Eq. (14.54).

$$f\,\mathrm{d}t = -\frac{\mathrm{d}V}{c_\beta} = \frac{-\mathrm{d}V}{\pm\left(1 - s_\beta^2\right)^{1/2}}$$

$$f\int_0^t \mathrm{d}t = -\int_{V_0}^{V} \frac{V\,\mathrm{d}V}{\pm\left(V^2 - V_0^2 s_{\beta_0}^2\right)^{1/2}} = ft = \Delta V$$

This yields, as in the previous section,

$$V^2 = V_0^2 + \Delta V^2 - 2\Delta V V_0 c_{\beta_0}$$

$$V = \left(V_0^2 + f^2t^2 - 2ftV_0c_{\beta_0}\right)^{1/2} \tag{14.62}$$

We can also obtain the preceding equation without integrating $\mathrm{d}V/\mathrm{d}t$ by simply writing

$$V = \frac{V_0 s_{\beta_0}}{s_\beta} = V_0 s_{\beta_0}\frac{(1 + \tan^2\beta)}{\tan\beta} \tag{14.63}$$

However, an expression for $\tan\beta$ is needed first. If we differentiate Eq. (14.59),

$$\frac{\mathrm{d}}{\mathrm{d}t}(\tan\beta) = \frac{2}{\pi}\lambda_i\frac{(\dot{V}\lambda_V + V\dot{\lambda}_V)}{V^2\lambda_V^2} \tag{14.64}$$

Replacing $\dot{V}$ and $\dot{\lambda}_V$ by Eqs. (14.54) and (14.56) and using Eq. (14.60) and (14.61) to eliminate λ_i and λ_V, the above derivative can be written as

$$\frac{\mathrm{d}}{\mathrm{d}\beta}(\tan\beta)\dot{\beta} = \frac{\dot{\beta}}{c_\beta^2} = \frac{fs_\beta}{Vc_\beta^2}$$

which yields

$$\dot{\beta} = \frac{fs_\beta}{V} \tag{14.65}$$

Because $Vs_\beta = V_0 s_{\beta_0}$, this can also be written as

$$\dot{\beta} = \frac{fs_\beta^2}{V_0 s_{\beta_0}}$$

such that

$$\int_{\beta_0}^{\beta}\frac{\mathrm{d}\beta}{s_\beta^2} = \frac{f}{V_0 s_{\beta_0}}\int_0^t \mathrm{d}t$$

$$\cot\beta_0 - \cot\beta = \frac{ft}{V_0 s_{\beta_0}}$$

and, finally, the control law

$$\tan\beta = \frac{V_0 s_{\beta_0}}{V_0 c_{\beta_0} - ft} \tag{14.66}$$

Going back to Eq. (14.63) and replacing $\tan\beta$ with the above expression results in

$$V = \left(V_0^2 + f^2t^2 - 2V_0 c_{\beta_0} ft\right)^{1/2} \tag{14.67}$$

Now the inclination time history can be obtained by direct integration of Eq. (14.53) by using the expression for V in Eq. (14.67),

$$\frac{\mathrm{d}i}{\mathrm{d}t} = \frac{2}{\pi}\frac{f}{V^2} V s_\beta = \frac{2}{\pi}\frac{f}{V^2} V_0 s_{\beta_0}$$

$$\int_0^i \mathrm{d}i = \frac{2}{\pi} V_0 s_{\beta_0} f \int_0^t \frac{\mathrm{d}t}{V_0^2 + f^2t^2 - 2V_0 c_{\beta_0} ft}$$

which yields

$$\Delta i = \frac{2}{\pi}\left[\tan^{-1}\left(\frac{ft - V_0 c_{\beta_0}}{V_0 s_{\beta_0}}\right) - \tan^{-1}(-\cot\beta_0)\right]$$

Since $\tan^{-1}x = -\tan^{-1}(-x)$ and $\tan^{-1}(\cot x) = \pi/2 - \tan^{-1}x$, the final result can be written as

$$\Delta i = \frac{2}{\pi}\left[\tan^{-1}\left(\frac{ft - V_0 c_{\beta_0}}{V_0 s_{\beta_0}}\right) + \frac{\pi}{2} - \beta_0\right] \tag{14.68}$$

This formula is uniformly valid for all t unlike the formulation of the previous section, which resulted in a set of two expressions for Δi because Δi was double-valued in the velocity. This simplification is achieved because time is selected as the independent variable instead of the velocity. If we integrate Eq. (14.56) for $\dot{\lambda}_V$ and use Eq. (14.60) to eliminate the constant λ_i, then,

$$\dot{\lambda}_V = -\frac{V_0^2 s_{\beta_0}^2}{V^3} = -V_0^2 s_{\beta_0}^2 \left(V_0^2 + f^2t^2 - 2V_0 c_{\beta_0} ft\right)^{-3/2}$$

which, upon integration, yields

$$\lambda_V = \frac{V_0 c_{\beta_0} - ft}{fV} \tag{14.69}$$

with $(\lambda_V)_0 = c_{\beta_0}/f$ and, in view of Eq. (14.61),

$$c_\beta = \frac{V_0 c_{\beta_0} - ft}{V} \tag{14.70}$$

From the definition of the influence functions λ_i and λ_V, we have

$$(\lambda_V)_0 = \frac{c_{\beta_0}}{f} = \frac{\partial t_f}{\partial V_0}$$

$$\delta t_f = \frac{c_{\beta_0}}{f}\delta V_0 \tag{14.71}$$

$$(\lambda_i)_0 = -\frac{\pi V_0 s_{\beta_0}}{2f} = \frac{\partial t_f}{\partial(\Delta i)_0}$$

$$\delta t_f = -\frac{\pi V_0 s_{\beta_0}}{2f}\delta(\Delta i)_0 \tag{14.72}$$

Equations (14.71) and (14.72) show how the total transfer time will vary for small variations in the initial velocity and inclination. Let us obtain an expression for ΔV in terms of β_0, V_0, and Δi. From Eq. (14.68), we have

$$\frac{ft - V_0 c_{\beta_0}}{V_0 s_{\beta_0}} = \tan\left(\frac{\pi}{2}\Delta i + \beta_0 - \frac{\pi}{2}\right) = \frac{-1}{\tan\left(\frac{\pi}{2}\Delta i + \beta_0\right)}$$

and since $ft = \Delta V$

$$\Delta V = V_0 c_{\beta_0} - \frac{V_0 s_{\beta_0}}{\tan\left(\frac{\pi}{2}\Delta i + \beta_0\right)} \tag{14.73}$$

This expression can be used to evaluate the total ΔV for the desired transfer once β_0, the initial yaw angle, is known. To obtain β_0, let us first observe that during the integration of $\dot{V}$ in Eq. (14.54), the same intermediate results shown in the previous section and resulting from that particular integration are valid here, too. They are

$$\Delta V = ft = V_0 c_{\beta_0} - V c_\beta$$

$$\Delta V = V_0 c_{\beta_0} \pm \left(V^2 - V_0^2 s_{\beta_0}^2\right)^{1/2} \tag{14.74}$$

From Δi in Eq. (14.68), and using the control law for $\tan\beta$ given in Eq. (14.66), we get

$$\Delta i = \frac{2}{\pi}(\beta - \beta_0) \tag{14.75}$$

We can now obtain β_0 by using the identity in Eq. (14.74) and writing Δi as

$$\Delta i = \frac{2}{\pi}\left\{\tan^{-1}\left[\frac{\pm\left(V^2 - V_0^2 s_{\beta_0}^2\right)^{1/2}}{V_0 s_{\beta_0}}\right] + \frac{\pi}{2} - \beta_0\right\}$$

Since $\tan^{-1} x = \pm\cos^{-1}[1/(x^2+1)^{1/2}]$ according to whether x is >0 or <0, the preceding expression for Δi can be cast as

$$\Delta i = \frac{2}{\pi}\left[\pm\cos^{-1}\left(\frac{V_0 s_{\beta_0}}{V}\right) + \frac{\pi}{2} - \beta_0\right]$$

or

$$\cos\left\{\mp\left[\frac{\pi}{2}-\left(\beta_0+\frac{\pi}{2}\Delta i\right)\right]\right\}=\frac{V_0 s_{\beta_0}}{V}$$

or, since $\cos(-x)=\cos x$,

$$\sin\left(\beta_0+\frac{\pi}{2}\Delta i\right)=\frac{V_0 s_{\beta_0}}{V}$$

which, after expansion and division by c_{β_0}, yields

$$\tan\beta_0=\frac{\sin\frac{\pi}{2}\Delta i}{\frac{V_0}{V}-\cos\frac{\pi}{2}\Delta i} \tag{14.76}$$

For given V_0, V_f, and $(\Delta i)_f$, β_0 can be obtained from Eq. (14.76), which then allows us to describe V and Δi, as well as β as a function of time, from Eqs. (14.67), (14.68), and (14.66), respectively. Equation (14.76) shows that, as Δi approaches the 2-rad value or Δi =114.59 deg, $\sin\pi/2\Delta i$ will approach zero so that $\beta_0\to 0$, indicating that the initial phase of the transfer will be coplanar. The ΔV equation of Edelbaum given by Eq. (14.51) will then approach $\Delta V=V_0+V_f$, which is the sum of the initial and final velocities. Given that V_0 represents the ΔV needed to transfer from V_0 to ∞ or escape, and V_f represents the ΔV needed to transfer from ∞ back to V_f, the transfer is initially coplanar until escape. At infinity, $V_\infty=0$ and the inclination change is achieved at zero cost, after which the return leg to V_f is also coplanar, resulting in $\Delta V=V_0+V_f$. This is shown from

$$\Delta V_1=\left(V_0^2-2V_0V_\infty+V_\infty^2\right)^{1/2}=V_0$$

$$\Delta V_2=\left(V_\infty^2-2V_\infty V_f+V_f^2\right)^{1/2}=V_f$$

$$\Delta V=\Delta V_1+\Delta V_2=V_0+V_f \tag{14.77}$$

For any inclination larger than Δi =114.59 deg, the cost of the orbit rotation is zero and the ΔV remains stationary at V_0+V_f. If the transfer is purely coplanar, i.e., $\Delta i=0$, then

$$\Delta V=|V_0-V_f| \tag{14.78}$$

the difference of the boundary velocities. For given V_0 and V_f, ΔV reaches a maximum if we set

$$\frac{\partial\Delta V}{\partial\Delta i}=0$$

where $\Delta V=(V_0^2-2V_fV_0\cos\pi/2\Delta i+V_f^2)^{1/2}$. This results in $\sin\pi/2\Delta i=0$ or $\Delta i=114.59$ deg as discussed earlier. Therefore, Edelbaum's Eq. (14.51) is to be used for $0\le\Delta i\le 114.59$ deg only. For $\Delta i>114.59$ deg, Eq. (14.77) must be used instead, as the use of Eq. (14.51) in this case will yield the wrong ΔV. Furthermore, from Eq. (14.74), as Δi approaches 114.59 deg from below, ΔV must

approach $V_0 + V_f$, which implies that $\beta_0 \to 0$, as shown earlier, and $\beta_f \to 180$ deg. The ΔV in Eq. (14.73) also approaches $V_0 + V_f$ since, with $\beta_0 = \varepsilon$, $s_{\beta_0} \sim \beta_0$, $c_{\beta_0} \sim 1$, $s_{\beta_f} \sim -s_{\beta_0} \sim -\beta_0$, and $\pi/2\Delta i \sim 180$ deg.

$$\Delta V \cong V_0 - \frac{V_f s_{\beta_f}}{\tan(180 + \varepsilon)} \cong V_0 + \frac{V_f \varepsilon}{\varepsilon} = V_0 + V_f$$

Algorithm of the Edelbaum Transfer Problem

In summary, we have the following algorithm. First, one computes V_0 and V_f from the knowledge of initial and final semimajor axes a_0 and a_f, $V_0 = (\mu/a_0)^{1/2}$, $V_f = (\mu/a_f)^{1/2}$. Given $\mu = 398601.3$ km^3/s^2, Earth's gravity constant; Δi, the total inclination change desired; and f, the low-thrust acceleration, one computes β_0 from Eq. (14.76) and the total ΔV_{tot} from Eq. (14.73) such that the transfer time is known from $t_f = \Delta V_{\text{tot}}/f$. The variation with time of the various variables of interest is obtained from

$$\Delta V = ft$$

$$\beta = \tan^{-1}\left(\frac{V_0 s_{\beta_0}}{V_0 c_{\beta_0} - ft}\right)$$

$$V = \left(V_0^2 - 2V_0 f t c_{\beta_0} + f^2 t^2\right)^{1/2}$$

$$\lambda_V = \frac{c_\beta}{f}$$

$$\Delta i = \frac{2}{\pi}\left[\tan^{-1}\left(\frac{ft - V_0 c_{\beta_0}}{V_0 s_{\beta_0}}\right) + \frac{\pi}{2} - \beta_0\right]$$

λ_i is, of course, constant and given by $-\pi V_0 s_{\beta_0}/(2f)$. The total $(\Delta i)_f$ is obtained from $|i_0 - i_f|$, where i_0 and i_f are the initial and final inclination, respectively. If $i_f > i_0$, the current inclination i is given by $i = i_0 + \Delta i$. If $i_f < i_0$, then $i = i_0 - \Delta i$. This is needed since we assumed $\Delta i > 0$ so that $\beta > \beta_0$, too.

An example of a constant-acceleration LEO to GEO transfer is shown in Figs. 14.1 and 14.2. The transfer is from $a_0 = 7000$ km, $i_0 = 28.5$ deg to $a_f = 42166$ km, $i_f = 0.0$ deg with $f = 3.5 \times 10^{-7}$ km/s^2. The total transfer time $t_f = 191.26259$ days with corresponding $\Delta V_{\text{tot}} = 5.78378$ km/s. The thrust yaw angle increases from its initial value $\beta_0 = 21.98$ deg to $\beta_f = 66.75$ deg as it is more efficient to rotate the orbit plane at higher altitudes. Figures 14.3 and 14.4 show the variation of the thrust yaw angle β, semimajor axis a, velocity V, and inclination i as a function of time for a low-thrust transfer between a_0 =7000 km, $i_0 = 90.0$ deg and $a_f = 42166$ km, $i_f = 0.0$ deg orbit. The acceleration is still constant at $f = 3.5 \times 10^{-7}$km/s^2. Starting from its initial value at $\beta_o = 10.92$ deg, β stays almost stationary for the first 100 days before surging to the final value $\beta_f = 152.29$ deg. It goes through 90 deg at day 245, where it starts to decelerate the vehicle since the orbit radius has exceeded the desired final altitude and, thus, must be shrunk until the final transfer time of $t_f = 335$ days for a total $\Delta V_{\text{tot}} = 10.13$ km/s. This is shown in Fig. 14.4, where the intermediate velocity is much less than the final desired velocity of 3.07 km/s at GEO.

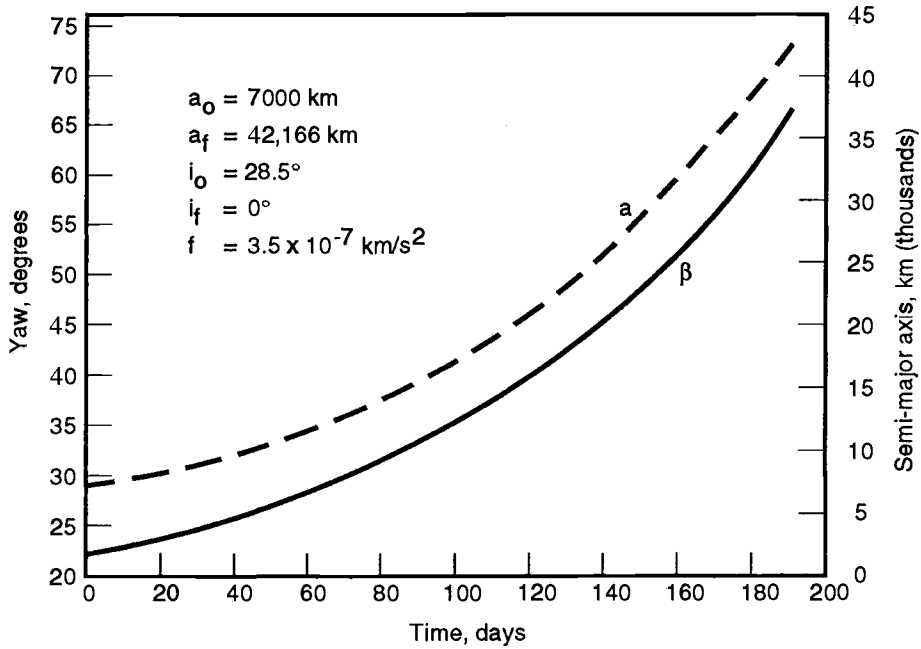


Fig. 14.1 Optimal thrust yaw profile and semimajor axis variation for a low-acceleration LEO to GEO transfer (from Ref. 1).

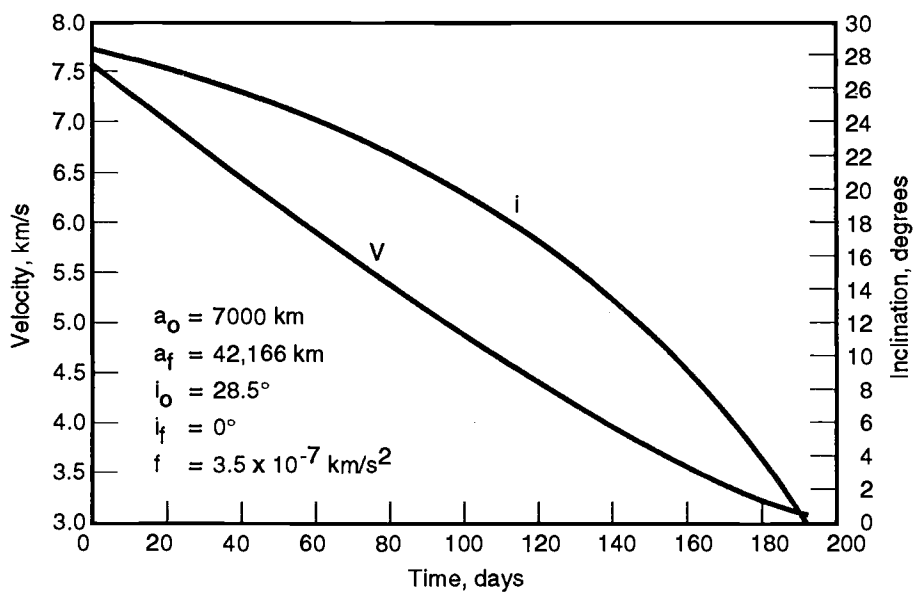


Fig. 14.2 Orbital velocity and inclination variations for a low-acceleration LEO to GEO transfer (from Ref. 1).

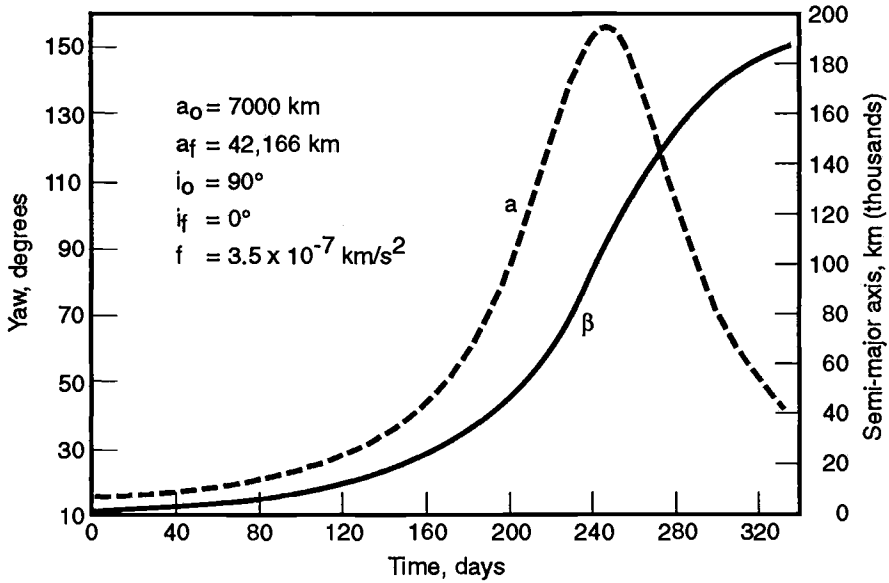


Fig. 14.3 Optimal thrust yaw profile and semimajor axis variation for a large inclination change transfer (from Ref. 1).

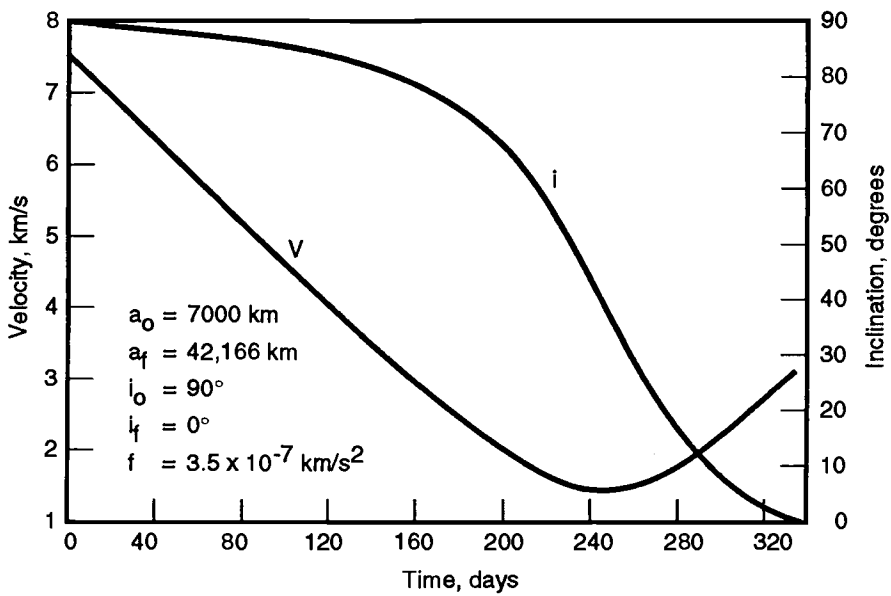


Fig. 14.4 Evolution of velocity and inclination during optimal transfer for a large inclination change case (from Ref. 1).

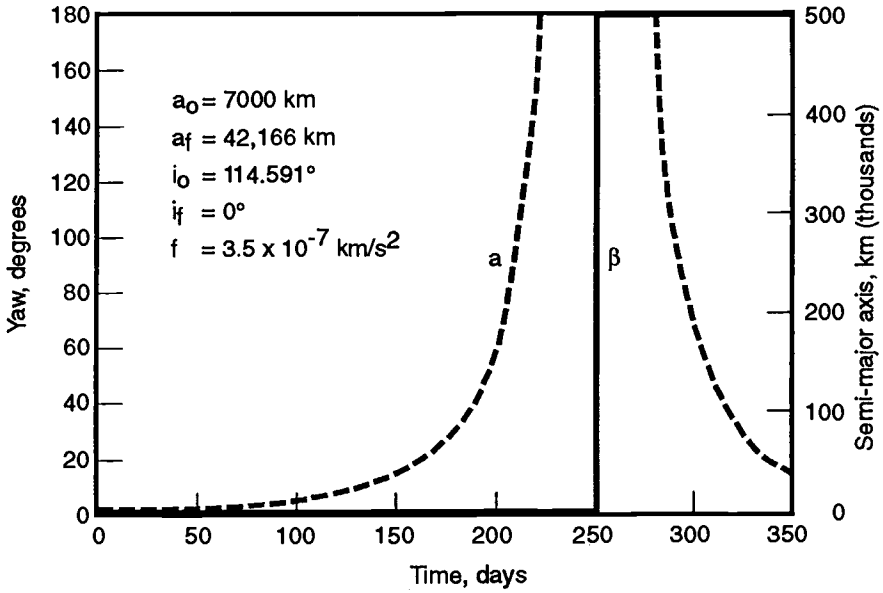


Fig. 14.5 Optimal thrust yaw profile and semimajor axis variation for the limiting case of Edelbaum's theory (Ref. 1).

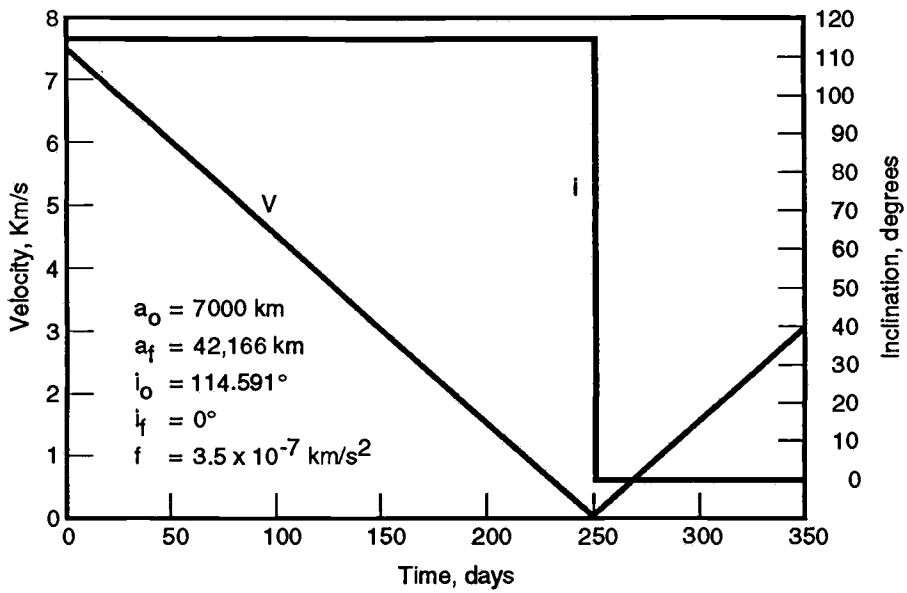


Fig. 14.6 Velocity and inclination profiles for the limiting case of Edelbaum's theory (Ref. 1).

Figures 14.5 and 14.6 are for $i_0 = 114.591$ deg, the uppermost limit of Edelbaum's theory. The β angle stays at zero, indicating a coplanar transfer up to $t = 250$ days, where $V \cong 0$, and the semimajor axis a at infinity, before flipping to 180 deg for the return leg to GEO. The inclination change is carried out instantaneously at infinity with zero cost. The total transfer time is $t_f = 351$ days with a ΔV of 10.61 km/s. This algorithm is valid regardless of whether the transfer is to a higher orbit or a lower orbit, irrespective of the direction of the inclination change.

14.3 The Full Six-State Formulation Using Nonsingular Equinoctial Orbit Elements

The mathematical theory of orbital mechanics in terms of the nonsingular equinoctial orbit elements has benefited from the contribution of Broucke and Cefola in Ref. 2, who took advantage of the well-established results concerning the Lagrange and Poisson brackets of classical elements that are found, for example, in Refs. 3 and 4 in order to transform these brackets in terms of the nonsingular elements. Cefola later developed the single-averaged variation of parameters equations for these elements in Ref. 5, which were applied by Edelbaum, Sackett, and Malchow in Ref. 6 to the problem of optimal low-thrust transfer. Further applications of nonsingular orbit prediction and orbit-transfer optimization problems appeared in Refs. 7 and 8. The variation of parameters perturbation equations based on the nonsingular equinoctial orbit elements are free from singularities for zero eccentricity and 0- and 90-deg inclination orbits. This fact, as well as many additional properties of the equinoctial elements, are derived in a systematic way in Ref. 2. The matrizant, or state transition matrix corresponding to these elements, is based on the partial derivatives of the position and velocity vectors with respect to the equinoctial elements as well as the inverse partial derivatives, meaning the partial derivatives of the equinoctial elements with respect to the position and velocity vectors. These partials were derived in Ref. 2 in terms of the classical elements, and transformed later in Refs. 5 and 7 in terms of the equinoctial elements. The applications of these elements to general and special perturbations are also discussed in the above-mentioned references. The position and velocity components in the direct equinoctial frame $\hat{f}$, $\hat{g}$, $\hat{w}$ of Fig. 14.7, to be described later, are given in terms of the equinoctial elements and the eccentric longitude F, which is related to the mean longitude λ by way of Kepler's equation. The equinoctial elements being defined in terms of the classical elements a, e, i, Ω, ω, M, Refs. 5 and 7 consider the eccentric longitude F to be a function of λ, h, and k, where h and k are the eccentricity vector components along $\hat{f}$ and $\hat{g}$, such that the partials of the position and velocity vectors with respect to the equinoctial elements are derived accordingly; F being defined by $F = E + \omega + \Omega$, where E is the classical eccentric anomaly. Given that the partials of the elements with respect to the velocity are obtained from the partials of the position with respect to the elements as well as the Poisson brackets,

$$\frac{\partial a_\alpha}{\partial \dot{\boldsymbol{r}}} = -\sum_{\beta=1}^{6} (a_\alpha, a_\beta) \frac{\partial \boldsymbol{r}}{\partial a_\beta}$$

where a_α and a_β are the generic elements and (a_α, a_β) are the Poisson brackets of the equinoctial elements, then the partials $\partial a_\alpha / \partial \dot{\boldsymbol{r}}$ are obtained in a form that

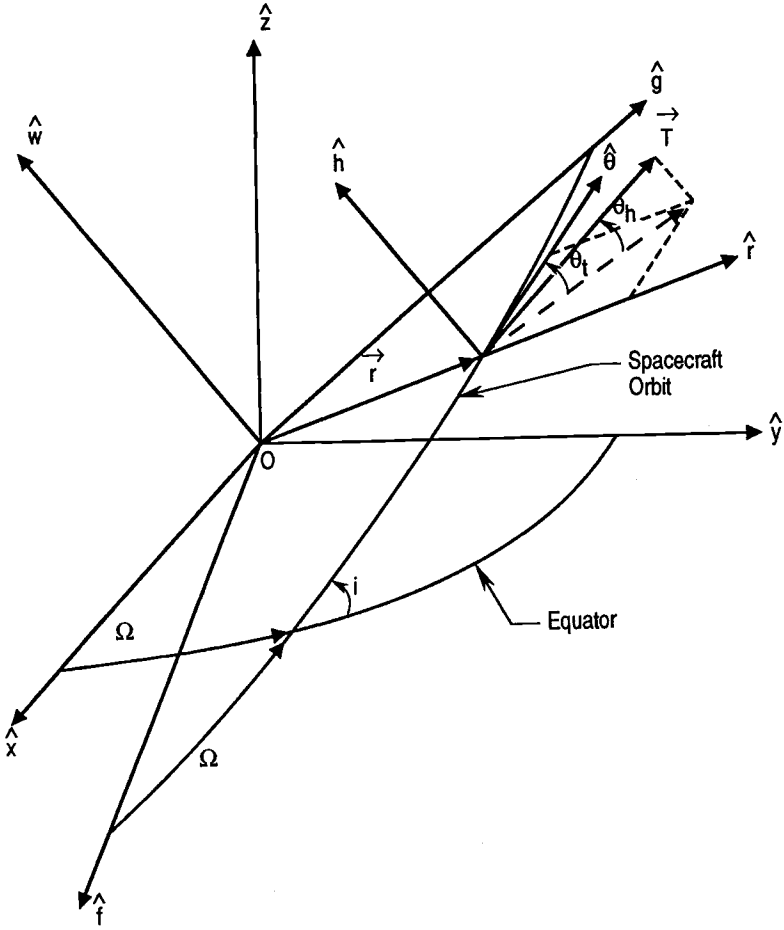


Fig. 14.7 Equinoctial frame and thrust geometry (from Ref. 15).

accounts for the dependence of F on λ, h, and k. It will be shown later how the preceding equation is used to integrate the state differential equations forward in time. In problems of optimal transfer and rendezvous, the Euler–Lagrange differential equations, which must be integrated simultaneously with the state differential equations for the solution of the two-point boundary-value problem, involve the partial derivatives of $\partial a_\alpha / \partial \dot{\boldsymbol{r}}$ with respect to the equinoctial elements. The velocity partials for the first five elements were derived for the case of the optimal transfer problem in Ref. 6, by assuming that F is an independent orbit element. However, we will assume in what follows that F is dependent on λ, h, and k. This assumption is consistent with Broucke and Cefola's such that we can provide the full set of the governing equations used in the solution of space-flight optimization problems that are consistent with the original nonsingular state differential equations used in orbit prediction.

Orbital Mechanics in Terms of Equinoctial Elements

The equinoctial orbit elements were introduced in order to avoid the singularities associated with the use of the classical elements a, e, i, Ω, ω, and M. They are defined with respect to these elements by

$$a = a \tag{14.79}$$

$$h = e\sin(\omega + \Omega) \tag{14.80}$$

$$k = e\cos(\omega + \Omega) \tag{14.81}$$

$$p = \tan\left(\frac{i}{2}\right)\sin\Omega \tag{14.82}$$

$$q = \tan\left(\frac{i}{2}\right)\cos\Omega \tag{14.83}$$

$$\lambda = M + \omega + \Omega \tag{14.84}$$

The mean longitude λ defined in terms of the mean anomaly M can be replaced by the eccentric longitude F or the true longitude L defined respectively in terms of the eccentric anomaly E and the true anomaly θ^* such that

$$F = E + \omega + \Omega \tag{14.85}$$

$$L = \theta^* + \omega + \Omega \tag{14.86}$$

The inverse transformation is readily obtained from the preceding definitions:

$$a = a \tag{14.87}$$

$$e = (h^2 + k^2)^{1/2} \tag{14.88}$$

$$i = 2\tan^{-1}(p^2 + q^2)^{1/2} \tag{14.89}$$

$$\Omega = \tan^{-1}\left(\frac{p}{q}\right) \tag{14.90}$$

$$\omega = \tan^{-1}\left(\frac{h}{k}\right) - \tan^{-1}\left(\frac{p}{q}\right) \tag{14.91}$$

$$M = \lambda - \tan^{-1}\left(\frac{h}{k}\right) \tag{14.92}$$

The last expression can be replaced by either of the two following ones:

$$E = F - \tan^{-1}\left(\frac{h}{k}\right) \tag{14.93}$$

$$\theta^* = L - \tan^{-1}\left(\frac{h}{k}\right) \tag{14.94}$$

This set of equinoctial elements is called the direct equinoctial elements. They do not exhibit any singularity for $e = 0$ and $i = 0$ deg, 90 deg. However, they are singular for i =180 deg. The equinoctial frame $\hat{f}$, $\hat{g}$, $\hat{w}$ is defined in Fig.14.7 respective to the inertial frame $\hat{x}$, $\hat{y}$, $\hat{z}$. The direction of the unit vector $\hat{f}$ is such that both $\hat{f}$ and $\hat{g}$ are in the orbit plane and $\hat{f}$ is obtained through a clockwise rotation of an angle Ω from the direction of the ascending node. The three rotations Ω, $-i$, $-\Omega$ transform the $\hat{f}$, $\hat{g}$, $\hat{w}$ system into the $\hat{x}$, $\hat{y}$, $\hat{z}$ system such that

$$\begin{pmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{pmatrix} = \begin{pmatrix} c_\Omega c_{-\Omega} - s_\Omega c_i s_{-\Omega} & -c_\Omega s_{-\Omega} - s_\Omega c_i c_{-\Omega} & s_\Omega s_i \\ s_\Omega c_{-\Omega} + c_\Omega c_i s_{-\Omega} & -s_\Omega s_{-\Omega} + c_\Omega c_i c_{-\Omega} & -c_\Omega s_i \\ s_i s_{-\Omega} & s_i c_{-\Omega} & c_i \end{pmatrix} \begin{pmatrix} \hat{f} \\ \hat{g} \\ \hat{w} \end{pmatrix} \tag{14.95}$$

which reduces to

$$\begin{pmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{pmatrix} = \begin{pmatrix} c_\Omega^2 + c_i s_\Omega^2 & c_\Omega s_\Omega - s_\Omega c_i c_\Omega & s_\Omega s_i \\ s_\Omega c_\Omega - s_\Omega c_\Omega c_i & s_\Omega^2 + c_i c_\Omega^2 & -c_\Omega s_i \\ -s_\Omega s_i & s_i c_\Omega & c_i \end{pmatrix} \begin{pmatrix} \hat{f} \\ \hat{g} \\ \hat{w} \end{pmatrix} \tag{14.96}$$

The inverse transformation is given by

$$\begin{pmatrix} \hat{f} \\ \hat{g} \\ \hat{w} \end{pmatrix} = \begin{pmatrix} c_\Omega^2 + c_i s_\Omega^2 & s_\Omega c_\Omega - s_\Omega c_\Omega c_i & -s_\Omega s_i \\ s_\Omega c_\Omega - s_\Omega c_\Omega c_i & s_\Omega^2 + c_i c_\Omega^2 & s_i c_\Omega \\ s_\Omega s_i & -c_\Omega s_i & c_i \end{pmatrix} \begin{pmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{pmatrix} \tag{14.97}$$

From the definitions of the equinoctial elements, the following relationships can be established:

$$c_{2\Omega} = \frac{q^2 - p^2}{p^2 + q^2}$$

$$s_\Omega = \frac{p}{(p^2 + q^2)^{1/2}}$$

$$c_\Omega = \frac{q}{(p^2 + q^2)^{1/2}}$$

$$s_{i/2}^2 = \frac{p^2 + q^2}{1 + p^2 + q^2}$$

$$c_{i/2}^2 = \frac{1}{1 + p^2 + q^2}$$

$$s_i = \frac{2(p^2 + q^2)^{1/2}}{1 + p^2 + q^2}$$

$$c_i = \frac{1 - p^2 - q^2}{1 + p^2 + q^2}$$

$$(1 + c_i)^2 = \frac{4}{(1 + p^2 + q^2)^2}$$

It is then possible to obtain expressions relating the various elements of the transformation matrix in Eq. (14.97) in terms of the equinoctial elements p and q such that

$$c_\Omega^2 + c_i s_\Omega^2 = \frac{1 - p^2 + q^2}{1 + p^2 + q^2}$$

$$s_\Omega c_\Omega - s_\Omega c_\Omega c_i = \frac{2pq}{1 + p^2 + q^2}$$

$$s_\Omega s_i = \frac{2p}{1 + p^2 + q^2}$$

$$s_\Omega^2 + c_i c_\Omega^2 = \frac{1 + p^2 - q^2}{1 + p^2 + q^2}$$

$$s_i c_\Omega = \frac{2q}{1 + p^2 + q^2}$$

From Eq. (14.97), the components of the $\hat{f}$, $\hat{g}$, $\hat{w}$ unit vectors in the inertial $\hat{x}$, $\hat{y}$, $\hat{z}$ frame are now readily obtained in terms of the direct equinoctial elements

$$\hat{f} = \frac{1}{(1 + p^2 + q^2)} \begin{pmatrix} 1 - p^2 + q^2 \\ 2pq \\ -2p \end{pmatrix} \tag{14.98}$$

$$\hat{g} = \frac{1}{(1 + p^2 + q^2)} \begin{pmatrix} 2pq \\ 1 + p^2 - q^2 \\ 2q \end{pmatrix} \tag{14.99}$$

$$\hat{w} = \frac{1}{(1 + p^2 + q^2)} \begin{pmatrix} 2p \\ -2q \\ 1 - p^2 - q^2 \end{pmatrix} \tag{14.100}$$

The derivation of Kepler's equation in terms of the equinoctial elements is carried out by using the definitions of e as well as h and k in the classical Kepler equation,

$$M = E - e \sin E$$

$$\lambda - (\omega + \Omega) = F - (\omega + \Omega) - e \sin[F - (\omega + \Omega)]$$

$$\lambda = F - e\left(s_F \frac{k}{e} - c_F \frac{h}{e}\right)$$

$$\lambda = F - k \sin F + h \cos F \tag{14.101}$$

The orbit equation $r = a(1 - e \cos E)$ can be written in terms of the equinoctial elements by using $F = E + \omega + \Omega$ and the definitions in Eqs. (14.80), (14.81), and (14.88), which relate e to h and k:

$$r = a\{1 - e \cos[F - (\omega + \Omega)]\}$$

$$r = a(1 - k \cos F - h \sin F) \tag{14.102}$$

With the radius vector $\boldsymbol{r}$ being contained in the orbit plane, its components X_1 and Y_1 along $\hat{f}$ and $\hat{g}$ must be derived. Then, $\dot{\boldsymbol{r}} = \dot{X}_1 \hat{f} + \dot{Y}_1 \hat{g}$ can also be evaluated. To do this, let us observe that $X_1 = r \cos(\theta^* + \omega + \Omega)$ and $Y_1 = r \sin(\theta^* + \omega + \Omega)$. From classical orbital mechanics, the relations $r c_{\theta^*} = a(\cos E - e)$ and $r s_{\theta^*} = a(1 - e^2)^{1/2} \sin E$ can now be used before the definitions in Eqs. (14.80), (14.81), (14.85), and (14.88) are introduced. Then,

$$\begin{aligned} X_1 &= r c_{\theta^*} c_{\omega+\Omega} - r s_{\theta^*} s_{\omega+\Omega} \\ &= a\left\{\frac{k^2 + h^2(1 - h^2 - k^2)^{1/2}}{(h^2 + k^2)} c_F + \frac{hk\left[1 - (1 - h^2 - k^2)^{1/2}\right]}{(h^2 + k^2)} s_F - k\right\} \end{aligned}$$

This intermediary result is written in order to introduce the quantity β, which is not to be confused with the thrust yaw angle β of the Edelbaum theory described earlier.

$$\beta = \frac{1 - (1 - h^2 - k^2)^{1/2}}{(h^2 + k^2)}$$

$$\beta = \frac{1}{1 + (1 - h^2 - k^2)^{1/2}} \tag{14.103}$$

We observe that

$$\frac{k^2 + h^2(1 - h^2 - k^2)^{1/2}}{(h^2 + k^2)} = 1 - h^2 \beta$$

and find that X_1 takes the final form

$$X_1 = a[(1 - h^2 \beta) \cos F + hk\beta \sin F - k] \tag{14.104}$$

The same manipulations are carried out for the definition of Y_1, which reduces to

$$Y_1 = a[hk\beta \cos F + (1 - k^2 \beta) \sin F - h] \tag{14.105}$$

The velocity components $\dot{X}_1$ and $\dot{Y}_1$ can be obtained directly from Eqs. (14.104) and (14.105) by holding a, h, and k, therefore, also β constant and varying only F.

$$\dot{X}_1 = a\left[-(1-h^2\beta)s_F\dot{F} + hk\beta c_F\dot{F}\right]$$
$$\dot{Y}_1 = a\left[-hk\beta s_F\dot{F} + (1-k^2\beta)c_F\dot{F}\right]$$

From Eq. (14.101) and with t_0 designating time at epoch,

$$\lambda = n(t-t_0) + \lambda_0 = F - ks_F + hc_F$$
$$\dot{\lambda} = n = \dot{F} - kc_F\dot{F} - hs_F\dot{F}$$
$$\dot{F} = \frac{n}{(1-kc_F - hs_F)}$$

and, in view of Eq. (14.102), the orbit equation

$$\dot{F} = \frac{na}{r} \tag{14.106}$$

The velocity components can then be cast into the following form:

$$\dot{X}_1 = \frac{a^2n}{r}[hk\beta c_F - (1-h^2\beta)s_F] \tag{14.107}$$

$$\dot{Y}_1 = \frac{na^2}{r}[(1-k^2\beta)c_F - hk\beta s_F] \tag{14.108}$$

Given $\boldsymbol{r}$ and $\dot{\boldsymbol{r}}$ in the inertial $\hat{x}, \hat{y}, \hat{z}$ system, the equinoctial elements can be obtained by first evaluating a and $\boldsymbol{e}$ with the semimajor axis obtained from the energy equation

$$a = \left(\frac{2}{|\boldsymbol{r}|} - \frac{|\dot{\boldsymbol{r}}|^2}{\mu}\right)^{-1} \tag{14.109}$$

$$\boldsymbol{e} = -\frac{\boldsymbol{r}}{|\boldsymbol{r}|} - \frac{(\boldsymbol{r}\times\dot{\boldsymbol{r}})\times\dot{\boldsymbol{r}}}{\mu} \tag{14.110}$$

Since $\hat{w}$ is along the angular momentum vector $\boldsymbol{h} = \boldsymbol{r}\times\dot{\boldsymbol{r}}$, we have

$$\hat{w} = \frac{\boldsymbol{r}\times\dot{\boldsymbol{r}}}{|\boldsymbol{r}\times\dot{\boldsymbol{r}}|} = \begin{pmatrix} w_x \\ w_y \\ w_z \end{pmatrix} \tag{14.111}$$

From the definition of $\hat{w}$ given by Eq. (14.100), it follows that p and q can be determined from Eq. (14.111) such that

$$p = \frac{w_x}{(1 + w_z)} \tag{14.112}$$

$$q = \frac{-w_y}{(1 + w_z)} \tag{14.113}$$

Since k and h are the components of the eccentricity vector $\boldsymbol{e}$ along $\hat{f}$ and $\hat{g}$, we have

$$k = \boldsymbol{e} \cdot \hat{f} \tag{14.114}$$

$$h = \boldsymbol{e} \cdot \hat{g} \tag{14.115}$$

Finally, the components of $\boldsymbol{r}$ along $\hat{f}$ and $\hat{g}$ are obtained from

$$X_1 = \boldsymbol{r} \cdot \hat{f} \tag{14.116}$$

$$Y_1 = \boldsymbol{r} \cdot \hat{g} \tag{14.117}$$

These quantities are needed to evaluate the mean longitude λ, which itself requires the evaluation of the eccentric longitude F since $\lambda = F - ks_F + hc_F$. The quantities s_F and c_F are obtained by solving the system of Eqs. (14.104) and (14.105), which express X_1 and Y_1 in terms of h, k, and F. This results in

$$s_F = h + \frac{Y_1(1 - h^2\beta) - hk\beta X_1}{a(1 - h^2 - k^2)^{1/2}}$$

$$c_F = \frac{(1 - k^2\beta)X_1 - hk\beta Y_1}{a(1 - h^2 - k^2)^{1/2}} + k$$

X_1 and Y_1 are first obtained from Eqs. (14.116) and (14.117). Finally, the partials of X_1, Y_1, $\dot{X}_1$, and $\dot{Y}_1$ with respect to h and k, which will be needed later, are evaluated from Eqs. (14.104), (14.105), (14.107), and (14.108). The following partials must, however, first be derived from

$$\beta = \frac{1}{1 + (1 - h^2 - k^2)^{1/2}}$$

with the observation that $(1 - h^2 - k^2)^{1/2} = (1 - \beta)/\beta$. They are

$$\frac{\partial \beta}{\partial h} = \frac{h\beta^3}{(1 - \beta)}$$

$$\frac{\partial \beta}{\partial k} = \frac{k\beta^3}{(1 - \beta)}$$

From $\lambda = F - ks_F + hc_F = M + \omega + \Omega$, it follows that

$$\frac{\partial \lambda}{\partial h} = 0 = \frac{\partial F}{\partial h} - kc_F \frac{\partial F}{\partial h} + c_F - hs_F \frac{\partial F}{\partial h}$$

$$= \frac{\partial F}{\partial h}[1 - kc_F - hs_F] + c_F = \frac{\partial F}{\partial h}\left(\frac{r}{a}\right) + c_F$$

$$\frac{\partial F}{\partial h} = -\frac{a}{r} c_F \tag{14.118}$$

Similarly, from $\partial \lambda / \partial k = 0$,

$$\frac{\partial F}{\partial k} = \frac{a}{r} s_F \tag{14.119}$$

which is then used in Eq. (14.104), together with $\partial \beta / \partial h$ developed earlier, to yield, after some manipulations,

$$\frac{\partial X_1}{\partial h} = a\left[-(hc_F - ks_F)\left(\beta + \frac{h^2\beta^3}{(1-\beta)}\right) - \frac{a}{r} c_F(h\beta - s_F)\right] \tag{14.120}$$

In a similar way, and as in Ref. 5,

$$\frac{\partial X_1}{\partial k} = -a\left[(hc_F - ks_F)\frac{hk\beta^3}{(1-\beta)} + 1 + \frac{a}{r} s_F(s_F - h\beta)\right] \tag{14.121}$$

$$\frac{\partial Y_1}{\partial h} = a\left[(hc_F - ks_F)\frac{hk\beta^3}{(1-\beta)} - 1 + \frac{a}{r} c_F(k\beta - c_F)\right] \tag{14.122}$$

$$\frac{\partial Y_1}{\partial k} = a\left[(hc_F - ks_F)\left(\beta + \frac{k^2\beta^3}{(1-\beta)}\right) + \frac{a}{r} s_F(c_F - k\beta)\right] \tag{14.123}$$

Letting $z = (a\ h\ k\ p\ q\ \lambda_0)^T$ represent the orbit state vector at time t, with λ_0 the mean longitude at epoch corresponding to the mean anomaly M_0, and given that $z = f(\boldsymbol{r}, \dot{\boldsymbol{r}})$, we have

$$\dot{z} = \frac{\partial z}{\partial \boldsymbol{r}} \dot{\boldsymbol{r}} + \frac{\partial z}{\partial \dot{\boldsymbol{r}}} \ddot{\boldsymbol{r}} \tag{14.124}$$

where

$$\ddot{\boldsymbol{r}} = \frac{\boldsymbol{T}}{m} + \boldsymbol{g} = \frac{\boldsymbol{T}}{m} - \frac{\mu}{r^3}\boldsymbol{r} \tag{14.125}$$

and where $\boldsymbol{T}$ and m stand for the thrust vector and spacecraft mass, respectively. The partial derivatives of the equinoctial elements with respect to the position and velocity vectors, the so-called inverse partials, are related to the partials of the

position and velocity vectors with respect to the equinoctial elements using the Poisson brackets (a_α, a_β) of equinoctial elements.

$$\frac{\partial a_\alpha}{\partial \boldsymbol{r}} = \sum_{\beta=1}^{6} (a_\alpha, a_\beta) \frac{\partial \dot{\boldsymbol{r}}}{\partial a_\beta} \tag{14.126}$$

$$\frac{\partial a_\alpha}{\partial \dot{\boldsymbol{r}}} = -\sum_{\beta=1}^{6} (a_\alpha, a_\beta) \frac{\partial \boldsymbol{r}}{\partial a_\beta} \tag{14.127}$$

If we write Eq. (14.124) one component at a time, then, using Eq. (14.125),

$$\begin{aligned}
\dot{z} &= \frac{\partial z}{\partial \boldsymbol{r}} \dot{\boldsymbol{r}} + \frac{\partial z}{\partial \dot{\boldsymbol{r}}} \ddot{\boldsymbol{r}} \\
&= \sum_\beta (a_\alpha, a_\beta) \frac{\partial \dot{\boldsymbol{r}}}{\partial a_\beta} \dot{\boldsymbol{r}} - \sum_\beta (a_\alpha, a_\beta) \frac{\partial \boldsymbol{r}}{\partial a_\beta} \ddot{\boldsymbol{r}} \\
&= \sum_\beta (a_\alpha, a_\beta) \frac{\partial \dot{\boldsymbol{r}}}{\partial a_\beta} \dot{\boldsymbol{r}} + \sum_\beta (a_\alpha, a_\beta) \frac{\partial \boldsymbol{r}}{\partial a_\beta} \left(\frac{\mu}{r^3} \right) \boldsymbol{r} + \frac{\partial z}{\partial \dot{\boldsymbol{r}}} \frac{\boldsymbol{T}}{m} \\
&= \sum_\beta (a_\alpha, a_\beta) \left(\frac{\partial \dot{\boldsymbol{r}}}{\partial a_\beta} \dot{\boldsymbol{r}} + \frac{\mu}{r^3} \frac{\partial \boldsymbol{r}}{\partial a_\beta} \boldsymbol{r} \right) + \frac{\partial z}{\partial \dot{\boldsymbol{r}}} \frac{\boldsymbol{T}}{m}
\end{aligned} \tag{14.128}$$

The term in brackets can be written as

$$\frac{1}{2} \frac{\partial (\dot{\boldsymbol{r}}^2)}{\partial z} + \frac{\mu}{r^3} \frac{1}{2} \frac{\partial (\boldsymbol{r}^2)}{\partial z} = \frac{1}{2} \left(\frac{\partial v^2}{\partial z} + \frac{\mu}{r^3} \frac{\partial r^2}{\partial z} \right) \tag{14.129}$$

Since, $|\dot{\boldsymbol{r}}| = v$ and, from the energy equation,

$$\frac{1}{2} v^2 - \frac{\mu}{r} = C$$

we have

$$\frac{1}{2} \frac{\partial v^2}{\partial z} = -\frac{\mu}{r^2} \frac{\partial r}{\partial z}$$

and, in view of $\partial r^2 / \partial z = 2r \, \partial r / \partial z$, the bracket in Eq. (14.129) cancels out such that

$$\dot{z} = \frac{\partial z}{\partial \dot{\boldsymbol{r}}} \frac{\boldsymbol{T}}{m}$$

If we let $\hat{u}$ represent a unit vector in the direction of the thrust,

$$\boldsymbol{T} = T \hat{u} \tag{14.130}$$

and write z as a vector, we obtain the variation of parameters equations, where both $\partial z/\partial \boldsymbol{r}$ and $\hat{u}$ are expressed in the direct equinoctial frame,

$$\dot{z} = \frac{\partial z}{\partial \dot{\boldsymbol{r}}}\frac{T}{m}\hat{u} = \frac{\partial z}{\partial \dot{\boldsymbol{r}}} f\hat{u} \tag{14.131}$$

Partials of the Equinoctial Orbit Elements with Respect to Velocity

The first-order differential equations in the equinoctial orbit elements involve the partials of these elements with respect to the velocity vector $\dot{\boldsymbol{r}}$. These partials are computed using the Poisson brackets of the equinoctial elements (a_α, a_β). However, the partials of the position vector $\boldsymbol{r}$ with respect to the equinoctial elements must be computed first. Furthermore, the Poisson brackets of the equinoctial elements can be obtained from the Poisson brackets of the classical elements, which are found in texts such as Refs. 3 and 4. These brackets have been derived in Refs. 2 and 5. These "velocity partials" are given by Eq. (14.127). First, let us derive the $\partial \boldsymbol{r}/\partial a_\beta$ partials. From the definition of $\boldsymbol{r} = X_1\hat{f} + Y_1\hat{g}$, it follows that

$$\frac{\partial \boldsymbol{r}}{\partial a} = \frac{\partial X_1}{\partial a}\hat{f} + \frac{\partial Y_1}{\partial a}\hat{g}$$

Refer back to the definition of λ to arrive at

$$\frac{\partial \lambda}{\partial a} = \frac{\partial \lambda}{\partial F}\frac{\partial F}{\partial a} = -\frac{3}{2a}nt$$

$$\frac{\partial F}{\partial a} = -\frac{3}{2}\frac{n}{r}t \tag{14.132}$$

Then, use the definitions of $\dot{X}_1$ and $\dot{Y}_1$ in Eqs. (14.107) and (14.108) to obtain

$$\begin{aligned}\frac{\partial X_1}{\partial a} &= [(1-h^2\beta)c_F + hk\beta s_F - k] + a[-(1-h^2\beta)s_F + hk\beta c_F]\frac{\partial F}{\partial a}\\ &= \frac{X_1}{a} + \frac{r\dot{X}_1}{na}\frac{\partial F}{\partial a} = \frac{X_1}{a} - \frac{3}{2}\frac{t}{a}\dot{X}_1\end{aligned}$$

In a similar way,

$$\frac{\partial Y_1}{\partial a} = \frac{Y_1}{a} - \frac{3}{2}\frac{t}{a}\dot{Y}_1$$

and, therefore,

$$\frac{\partial \boldsymbol{r}}{\partial a} = \left(\frac{X_1}{a} - \frac{3}{2}\frac{t}{a}\dot{X}_1\right)\hat{f} + \left(\frac{Y_1}{a} - \frac{3}{2}\frac{t}{a}\dot{Y}_1\right)\hat{g}$$

The partials of $\partial \boldsymbol{r}/\partial h$ and $\partial \boldsymbol{r}/\partial k$ are written in a straightforward manner:

$$\frac{\partial \boldsymbol{r}}{\partial h} = \frac{\partial X_1}{\partial h}\hat{f} + \frac{\partial Y_1}{\partial h}\hat{g} \tag{14.133}$$

$$\frac{\partial \boldsymbol{r}}{\partial k} = \frac{\partial X_1}{\partial k}\hat{f} + \frac{\partial Y_1}{\partial k}\hat{g} \tag{14.134}$$

Equations (14.120–14.123) must, of course, be used in Eqs. (14.133) and (14.134). Now, by way of some manipulations,

$$\frac{\partial \boldsymbol{r}}{\partial \lambda} = \frac{\partial \boldsymbol{r}}{\partial F}\frac{\partial F}{\partial \lambda} = \left(\frac{\partial X_1}{\partial F}\hat{f} + \frac{\partial Y_1}{\partial F}\hat{g}\right)\frac{\partial F}{\partial \lambda}$$

$$\frac{\partial \boldsymbol{r}}{\partial \lambda} = \frac{\dot{X}_1}{n}\hat{f} + \frac{\dot{Y}_1}{n}\hat{g} = \frac{\dot{\boldsymbol{r}}}{n} \tag{14.135}$$

The remaining partials $\partial \boldsymbol{r}/\partial p$ and $\partial \boldsymbol{r}/\partial q$ are derived by observing first that X_1 and Y_1 are not functions of either p or q but that the components of the unit vectors $\hat{f}$ and $\hat{g}$ are such functions; therefore,

$$\frac{\partial \boldsymbol{r}}{\partial p} = X_1\frac{\partial \hat{f}}{\partial p} + Y_1\frac{\partial \hat{g}}{\partial p}$$

From Eqs. (14.98–14.100), it follows that

$$\frac{\partial \boldsymbol{r}}{\partial p} = \frac{2X_1}{(1+p^2+q^2)^2}\begin{pmatrix} -2p-2pq^2 \\ 2q-q(1+p^2-q^2) \\ -2q^2-(1-p^2-q^2) \end{pmatrix} + \frac{2Y_1}{(1+p^2+q^2)^2}\begin{pmatrix} q(1+q^2-p^2) \\ 2pq^2 \\ -2pq \end{pmatrix}$$

$$\frac{\partial \boldsymbol{r}}{\partial p} = \frac{2}{(1+p^2+q^2)}[q(Y_1\hat{f} - X_1\hat{g}) - X_1\hat{w}] \tag{14.136}$$

In a similar way,

$$\frac{\partial \boldsymbol{r}}{\partial q} = X_1\frac{\partial \hat{f}}{\partial q} + Y_1\frac{\partial \hat{g}}{\partial q} = \frac{2}{(1+p^2+q^2)}\left[p(X_1\hat{g} - Y_1\hat{f}) + Y_1\hat{w}\right] \tag{14.137}$$

The Poisson brackets of equinoctial elements are very tedious to derive unless one uses the transformation equation that converts the Poisson brackets of classical elements into the brackets of equinoctial elements. The transformation equation is given in Ref. 2 and, using the nomenclature introduced by Broucke and Cefola, we write

$$[(p_\alpha, p_\beta)] = \left[\frac{\partial p_\alpha}{\partial a_\lambda}\right][(a_\lambda, a_\mu)]\left[\frac{\partial p_\beta}{\partial a_\mu}\right]^T \tag{14.138}$$

The various matrices appearing in Eq. (14.138) are given by

$$[(p_\alpha, p_\beta)] = \begin{bmatrix} (a,a) & (a,h) & (a,k) & (a,\lambda_0) & (a,p) & (a,q) \\ (h,a) & (h,h) & (h,k) & (h,\lambda_0) & (h,p) & (h,q) \\ (k,a) & (k,h) & (k,k) & (k,\lambda_0) & (k,p) & (k,q) \\ (\lambda_0,a) & (\lambda_0,h) & (\lambda_0,k) & (\lambda_0,\lambda_0) & (\lambda_0,p) & (\lambda_0,q) \\ (p,a) & (p,h) & (p,k) & (p,\lambda_0) & (p,p) & (p,q) \\ (q,a) & (q,h) & (q,k) & (q,\lambda_0) & (q,p) & (q,q) \end{bmatrix}$$

$$[(a_\lambda, a_\mu)] = \begin{bmatrix} (a,a) & (a,e) & (a,i) & (a,\Omega) & (a,\omega) & (a,M_0) \\ (e,a) & (e,e) & (e,i) & (e,\Omega) & (e,\omega) & (e,M_0) \\ (i,a) & (i,e) & (i,i) & (i,\Omega) & (i,\omega) & (i,M_0) \\ (\Omega,a) & (\Omega,e) & (\Omega,i) & (\Omega,\Omega) & (\Omega,\omega) & (\Omega,M_0) \\ (\omega,a) & (\omega,e) & (\omega,i) & (\omega,\Omega) & (\omega,\omega) & (\omega,M_0) \\ (M_0,a) & (M_0,e) & (M_0,i) & (M_0,\Omega) & (M_0,\omega) & (M_0,M_0) \end{bmatrix}$$

$$\left[\frac{\partial p_\alpha}{\partial a_\lambda}\right] = \begin{bmatrix} \frac{\partial a}{\partial a} & \frac{\partial a}{\partial e} & \frac{\partial a}{\partial i} & \frac{\partial a}{\partial \Omega} & \frac{\partial a}{\partial \omega} & \frac{\partial a}{\partial M_0} \\ \frac{\partial h}{\partial a} & \frac{\partial h}{\partial e} & \frac{\partial h}{\partial i} & \frac{\partial h}{\partial \Omega} & \frac{\partial h}{\partial \omega} & \frac{\partial h}{\partial M_0} \\ \frac{\partial k}{\partial a} & \frac{\partial k}{\partial e} & \frac{\partial k}{\partial i} & \frac{\partial k}{\partial \Omega} & \frac{\partial k}{\partial \omega} & \frac{\partial k}{\partial M_0} \\ \frac{\partial \lambda_0}{\partial a} & \frac{\partial \lambda_0}{\partial e} & \frac{\partial \lambda_0}{\partial i} & \frac{\partial \lambda_0}{\partial \Omega} & \frac{\partial \lambda_0}{\partial \omega} & \frac{\partial \lambda_0}{\partial M_0} \\ \frac{\partial p}{\partial a} & \frac{\partial p}{\partial e} & \frac{\partial p}{\partial i} & \frac{\partial p}{\partial \Omega} & \frac{\partial p}{\partial \omega} & \frac{\partial p}{\partial M_0} \\ \frac{\partial q}{\partial a} & \frac{\partial q}{\partial e} & \frac{\partial q}{\partial i} & \frac{\partial q}{\partial \Omega} & \frac{\partial q}{\partial \omega} & \frac{\partial q}{\partial M_0} \end{bmatrix}$$

$$\left[\frac{\partial p_\beta}{\partial a_\mu}\right]^T = \begin{bmatrix} \frac{\partial a}{\partial a} & \frac{\partial h}{\partial a} & \frac{\partial k}{\partial a} & \frac{\partial \lambda_0}{\partial a} & \frac{\partial p}{\partial a} & \frac{\partial q}{\partial a} \\ \frac{\partial a}{\partial e} & \frac{\partial h}{\partial e} & \frac{\partial k}{\partial e} & \frac{\partial \lambda_0}{\partial e} & \frac{\partial p}{\partial e} & \frac{\partial q}{\partial e} \\ \frac{\partial a}{\partial i} & \frac{\partial h}{\partial i} & \frac{\partial k}{\partial i} & \frac{\partial \lambda_0}{\partial i} & \frac{\partial p}{\partial i} & \frac{\partial q}{\partial i} \\ \frac{\partial a}{\partial \Omega} & \frac{\partial h}{\partial \Omega} & \frac{\partial k}{\partial \Omega} & \frac{\partial \lambda_0}{\partial \Omega} & \frac{\partial p}{\partial \Omega} & \frac{\partial q}{\partial \Omega} \\ \frac{\partial a}{\partial \omega} & \frac{\partial h}{\partial \omega} & \frac{\partial k}{\partial \omega} & \frac{\partial \lambda_0}{\partial \omega} & \frac{\partial p}{\partial \omega} & \frac{\partial q}{\partial \omega} \\ \frac{\partial a}{\partial M_0} & \frac{\partial h}{\partial M_0} & \frac{\partial k}{\partial M_0} & \frac{\partial \lambda_0}{\partial M_0} & \frac{\partial p}{\partial M_0} & \frac{\partial q}{\partial M_0} \end{bmatrix}$$

The elements of the matrix $[\partial p_\alpha / \partial a_\lambda]$ can be calculated from the definitions of the equinoctial elements that appear in Eqs. (14.79–14.84).

$$\left[\frac{\partial p_\alpha}{\partial a_\lambda}\right] = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & s_{\omega+\Omega} & 0 & ec_{\omega+\Omega} & ec_{\omega+\Omega} & 0 \\ 0 & c_{\omega+\Omega} & 0 & -es_{\omega+\Omega} & -es_{\omega+\Omega} & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & \dfrac{s_\Omega}{2\cos^2\dfrac{i}{2}} & \tan\dfrac{i}{2}c_\Omega & 0 & 0 \\ 0 & 0 & \dfrac{c_\Omega}{2\cos^2\dfrac{i}{2}} & -\tan\dfrac{i}{2}s_\Omega & 0 & 0 \end{bmatrix}$$

which converts to

$$\left[\frac{\partial p_\alpha}{\partial a_\lambda}\right] = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & h/(h^2+k^2)^{1/2} & 0 & k & k & 0 \\ 0 & k/(h^2+k^2)^{1/2} & 0 & -h & -h & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & \dfrac{p(1+p^2+q^2)}{2(p^2+q^2)^{1/2}} & q & 0 & 0 \\ 0 & 0 & \dfrac{q(1+p^2+q^2)}{2(p^2+q^2)^{1/2}} & -p & 0 & 0 \end{bmatrix}$$

The Poisson brackets in classical elements, though difficult to derive, are readily available in advanced books on astrodynamics such as Refs. 3 and 4. The matrix of the Poisson brackets in the classical elements is found to be, with $p' = a(1-e^2)$, the orbit parameter

$$[(a_\lambda, a_\mu)] = P_{c1} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & -2\left(\dfrac{a}{\mu}\right)^{1/2} \\ & 0 & 0 & 0 & \dfrac{(1-e^2)^{1/2}}{e(\mu a)^{1/2}} & \dfrac{-(1-e^2)}{e(\mu a)^{1/2}} \\ & & 0 & \dfrac{1}{(\mu p')^{1/2}s_i} & \dfrac{-c_i}{(\mu p')^{1/2}s_i} & 0 \\ & & & 0 & 0 & 0 \\ -\text{Sym} & & & & 0 & 0 \\ & & & & & 0 \end{bmatrix} \tag{14.139}$$

Carrying out the algebra, and using the identities introduced earlier, as well as the following identities, we obtain

$$s_i c_\Omega = \frac{2q}{(1+p^2+q^2)}$$
$$s_\Omega s_i = \frac{2p}{(1+p^2+q^2)}$$
$$-c_\Omega s_i = \frac{-2q}{(1+p^2+q^2)}$$

The Poisson brackets of equinoctial elements are derived to result in

$$(a, \lambda_0) = \frac{-2}{na}$$
$$(\lambda_0, h) = \frac{-h(1-h^2-k^2)^{1/2}}{na^2\left[1+(1-h^2-k^2)^{1/2}\right]}$$
$$(\lambda_0, k) = \frac{-k(1-h^2-k^2)^{1/2}}{na^2\left[1+(1-h^2-k^2)^{1/2}\right]}$$
$$(\lambda_0, p) = \frac{-p(1+p^2+q^2)}{2na^2(1-h^2-k^2)^{1/2}}$$
$$(\lambda_0, q) = \frac{-q(1+p^2+q^2)}{2na^2(1-h^2-k^2)^{1/2}}$$
$$(h, k) = \frac{-(1-h^2-k^2)^{1/2}}{na^2}$$
$$(h, p) = \frac{-kp(1+p^2+q^2)}{2na^2(1-h^2-k^2)^{1/2}}$$
$$(h, q) = \frac{-kq(1+p^2+q^2)}{2na^2(1-h^2-k^2)^{1/2}}$$
$$(k, p) = \frac{hp(1+p^2+q^2)}{2na^2(1-h^2-k^2)^{1/2}}$$
$$(k, q) = \frac{hq(1+p^2+q^2)}{2na^2(1-h^2-k^2)^{1/2}}$$
$$(p, q) = \frac{-(1+p^2+q^2)^2}{4na^2(1-h^2-k^2)^{1/2}}$$

The property that emanates from the definition of the brackets, namely, $(\alpha, \alpha) = 0$ and $(\alpha, \beta) = -(\beta, \alpha)$, is responsible for the antisymmetric nature of the foregoing matrices. Going back to our original nomenclature for (a_α, a_β) and having

developed all the Poisson brackets (a_α, a_β) and all the partials $\partial \boldsymbol{r}/\partial a_\beta$, we can now derive the partials $\partial z/\partial \dot{\boldsymbol{r}}$. These partials are

$$\frac{\partial a}{\partial \dot{\boldsymbol{r}}} = \frac{2}{n^2 a}(\dot{X}_1 \hat{f} + \dot{Y}_1 \hat{g}) = M_{11}\hat{f} + M_{12}\hat{g} + M_{13}\hat{w} \tag{14.140}$$

$$\frac{\partial h}{\partial \dot{\boldsymbol{r}}} = \frac{(1-h^2-k^2)^{1/2}}{na^2}\left[\left(\frac{\partial X_1}{\partial k} - \frac{h\beta}{n}\dot{X}_1\right)\hat{f} + \left(\frac{\partial Y_1}{\partial k} - \frac{h\beta}{n}\dot{Y}_1\right)\hat{g}\right]$$
$$+ \frac{k(qY_1 - pX_1)}{na^2(1-h^2-k^2)^{1/2}}\hat{w} = M_{21}\hat{f} + M_{22}\hat{g} + M_{23}\hat{w} \tag{14.141}$$

$$\frac{\partial k}{\partial \dot{\boldsymbol{r}}} = -\frac{(1-h^2-k^2)^{1/2}}{na^2}\left[\left(\frac{\partial X_1}{\partial h} + k\beta\frac{\dot{X}_1}{n}\right)\hat{f} + \left(\frac{\partial Y_1}{\partial h} + k\beta\frac{\dot{Y}_1}{n}\right)\hat{g}\right]$$
$$- \frac{h(qY_1 - pX_1)}{na^2(1-h^2-k^2)^{1/2}}\hat{w} = M_{31}\hat{f} + M_{32}\hat{g} + M_{33}\hat{w} \tag{14.142}$$

$$\frac{\partial p}{\partial \dot{\boldsymbol{r}}} = \frac{(1+p^2+q^2)Y_1}{2na^2(1-h^2-k^2)^{1/2}}\hat{w} = M_{41}\hat{f} + M_{42}\hat{g} + M_{43}\hat{w} \tag{14.143}$$

$$\frac{\partial q}{\partial \dot{\boldsymbol{r}}} = \frac{(1+p^2+q^2)X_1}{2na^2(1-h^2-k^2)^{1/2}}\hat{w} = M_{51}\hat{f} + M_{52}\hat{g} + M_{53}\hat{w} \tag{14.144}$$

$$\frac{\partial \lambda_0}{\partial \dot{\boldsymbol{r}}} = \frac{1}{na^2}\left[-2X_1 + 3\dot{X}_1 t + (1-h^2-k^2)^{1/2}\left(h\beta\frac{\partial X_1}{\partial h} + k\beta\frac{\partial X_1}{\partial k}\right)\right]\hat{f}$$
$$+ \frac{1}{na^2}\left[-2Y_1 + 3\dot{Y}_1 t + (1-h^2-k^2)^{1/2}\left(h\beta\frac{\partial Y_1}{\partial h} + k\beta\frac{\partial Y_1}{\partial k}\right)\right]\hat{g}$$
$$+ \frac{(qY_1 - pX_1)}{na^2(1-h^2-k^2)^{1/2}}\hat{w} = M_{61}\hat{f} + M_{62}\hat{g} + M_{63}\hat{w} \tag{14.145}$$

Let us now use $\partial\lambda/\partial\dot{\boldsymbol{r}}$ instead of $\partial\lambda_0/\partial\dot{\boldsymbol{r}}$. There is a simple relationship that links the two partials since $\lambda = \lambda_0 + n(t - t_0)$ such that

$$\frac{\partial \lambda}{\partial \dot{\boldsymbol{r}}} = \frac{\partial \lambda_0}{\partial \dot{\boldsymbol{r}}} + t\frac{\partial n}{\partial a}\frac{\partial a}{\partial \dot{\boldsymbol{r}}}$$

and, since $\partial n/\partial a = -3n/(2a)$ and $\partial a/\partial \dot{\boldsymbol{r}} = 2\dot{\boldsymbol{r}}/(n^2 a)$, then,

$$\frac{\partial \lambda}{\partial \dot{\boldsymbol{r}}} = \frac{\partial \lambda_0}{\partial \dot{\boldsymbol{r}}} - \frac{3}{na^2}t\dot{\boldsymbol{r}} \tag{14.146}$$

The matrix M is then a 6×3 matrix such that

$$M = \begin{bmatrix} (\partial a/\partial \dot{\boldsymbol{r}})^T \\ (\partial h/\partial \dot{\boldsymbol{r}})^T \\ (\partial k/\partial \dot{\boldsymbol{r}})^T \\ (\partial p/\partial \dot{\boldsymbol{r}})^T \\ (\partial q/\partial \dot{\boldsymbol{r}})^T \\ (\partial \lambda/\partial \dot{\boldsymbol{r}})^T \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \\ M_{41} & M_{42} & M_{43} \\ M_{51} & M_{52} & M_{53} \\ M_{61} & M_{62} & M_{63} \end{bmatrix} \tag{14.147}$$

The equations of motion are therefore given by

$$\dot{a} = \left(\frac{\partial a}{\partial \dot{\boldsymbol{r}}}\right)^T \cdot \hat{u} f \tag{14.148}$$

$$\dot{h} = \left(\frac{\partial h}{\partial \dot{\boldsymbol{r}}}\right)^T \cdot \hat{u} f \tag{14.149}$$

$$\dot{k} = \left(\frac{\partial k}{\partial \dot{\boldsymbol{r}}}\right)^T \cdot \hat{u} f \tag{14.150}$$

$$\dot{p} = \left(\frac{\partial p}{\partial \dot{\boldsymbol{r}}}\right)^T \cdot \hat{u} f \tag{14.151}$$

$$\dot{q} = \left(\frac{\partial q}{\partial \dot{\boldsymbol{r}}}\right)^T \cdot \hat{u} f \tag{14.152}$$

$$\dot{\lambda} = n + \left(\frac{\partial \lambda}{\partial \dot{\boldsymbol{r}}}\right)^T \cdot \hat{u} f \tag{14.153}$$

The Euler–Lagrange Differential Equations

The Euler–Lagrange or adjoint differential equations for the set a, h, k, p, q, and λ will now be derived. The necessary condition for optimality consists of maximizing the Hamiltonian of the system given by

$$H = \boldsymbol{\lambda}_z^T \dot{z} = \boldsymbol{\lambda}_z^T M(z, F) f \hat{u} + \lambda_\lambda n \tag{14.154}$$

where the seventh differential equation for the mass flow rate has been neglected. The Euler–Lagrange equations are then given by

$$\dot{\boldsymbol{\lambda}}_z = -\frac{\partial H}{\partial z} = -\boldsymbol{\lambda}_z^T \frac{\partial M}{\partial z} f \hat{u} - \lambda_\lambda \frac{\partial n}{\partial z} \tag{14.155}$$

The optimal thrust direction $\hat{u}$ is chosen such that it is at all times parallel to $\boldsymbol{\lambda}_z^T M(z, F)$ in order to maximize the Hamiltonian. The matrix M for the set (a, h, k, p, q, λ) is as given by Eq. (14.147). The complete set of the partials $\partial M/\partial z$ is shown in the Appendix. These analytic partials are used in Eq. (14.155)

for the numerical integration of the $\dot{\boldsymbol{\lambda}}_z$ rates. The only nonzero $\partial n/\partial z$ partial is given by

$$\frac{\partial n}{\partial a} = -\frac{3n}{2a} = -\frac{3}{2}\mu^{1/2}a^{-5/2}$$

The following partials are used to generate the $\partial M/\partial z$ partials of the Appendix. Here F must be considered to be independent of a but not of h, k, and λ because λ is being integrated and is therefore independent of a, h, k, p, and q such that $\partial\lambda/\partial a$ is now equal to zero. From Kepler's equation, this implies that $\partial F/\partial a = 0$. Thus,

$$\frac{\partial r}{\partial a} = \frac{r}{a}$$

$$\frac{\partial r}{\partial h} = \frac{a^2}{r}(h - s_F)$$

$$\frac{\partial r}{\partial k} = \frac{a^2}{r}(k - c_F)$$

$$\frac{\partial r}{\partial F} = a(ks_F - hc_F)$$

$$\frac{\partial F}{\partial a} = 0$$

$$\frac{\partial F}{\partial h} = -\frac{a}{r}c_F$$

$$\frac{\partial F}{\partial k} = \frac{a}{r}s_F$$

$$\frac{\partial F}{\partial \lambda} = \frac{a}{r}$$

It is also true that $\partial\lambda/\partial F = r/a$. The components u_f, u_g, u_w of the unit vector $\hat{u}$ in the equinoctial frame are obtained from

$$\hat{u} = \left(\boldsymbol{\lambda}_z^T M\right)^T / \left|\boldsymbol{\lambda}_z^T M\right| \tag{14.156}$$

From Fig. 14.7, the thrust vector $\boldsymbol{T} = T\hat{u}$ is defined by the thrust pitch and thrust yaw angles θ_t and θ_h, respectively, in the rotating $\hat{r}, \hat{\theta}, \hat{h}$ frame; $\hat{r}$ is a unit vector along the instantaneous position vector $\boldsymbol{r}$, with $\hat{\theta}$ in the orbit plane and along the direction of motion, and $\hat{h}$ along the angular momentum vector. Therefore,

$$\theta_t = \tan^{-1}(u_r/u_\theta) \tag{14.157}$$

$$\theta_h = \tan^{-1}(u_w/u_\theta) \tag{14.158}$$

with u_r, u_θ, and u_h, representing the components of the unit vector $\hat{u}$ along the $\hat{r}$, $\hat{\theta}$, and $\hat{h}$ directions. We have

$$\hat{r} = \frac{X_1}{|\boldsymbol{r}|}\hat{f} + \frac{Y_1}{|\boldsymbol{r}|}\hat{g} \tag{14.159}$$

$$\hat{\theta} = \hat{h} \times \hat{r} = -\frac{Y_1}{|\boldsymbol{r}|}\hat{f} + \frac{X_1}{|\boldsymbol{r}|}\hat{g} \tag{14.160}$$

such that

$$u_r = \frac{X_1}{r}u_f + \frac{Y_1}{r}u_g \tag{14.161}$$

$$u_\theta = -\frac{Y_1}{r}u_f + \frac{X_1}{r}u_g \tag{14.162}$$

$$u_h = u_w \tag{14.163}$$

where

$$r = |\boldsymbol{r}| = \left(X_1^2 + Y_1^2\right)^{1/2}$$

14.4 Orbit Transfer with Continuous Constant Acceleration

This section presents examples of precision-integrated, optimized low earth orbit (LEO) to geostationary earth orbit (GEO) minimum-time transfer and compares them to the solutions obtained by way of the averaging technique. A 10^{-2} g acceleration applied in a constant and continuous manner is taken as an example in order to generate fast subday transfers that could be flown with nuclear thermal propulsion upper stages. The six-state formulation used here allows the user to generate optimal transfers that first start from a given fixed location on the initial orbit while optimizing the arrival point on the target or final orbit. The analysis is further extended to optimize both departure and arrival points in order to obtain the overall minimum-time free-free solution. This requires the vanishing of the Lagrange multiplier adjoint to the mean longitude at both initial and final times with fixed initial time and optimized final time. These fast, few-revolution, five-state transfers are sensitive to initial and final orbital position, thereby necessitating the use of the full six-state dynamics. These exact results are then compared to the approximate solutions obtained using averaged dynamics with robust and fast convergence characteristics. These examples determine that the ΔVs or transfer time solutions compare rather well, even for these short-duration transfers, but that the element time histories, and especially the eccentricity, are poorly simulated by the approximate solutions. Furthermore, due to the nature of the averaging technique, the sensitivity of the solution to orbital position is totally removed such that the precision-integrated solution must be used instead for accurate guidance.

Minimum-Time Transfer from Fixed Initial State with Continuous Constant Acceleration

We minimize total transfer time by maximizing the performance index $J = \int_{t_0}^{t_f} L\mathrm{d}t = -\int_{t_0}^{t_f} \mathrm{d}t = -(t_f - t_0)$. For fixed t_0, the minimization of t_f or maximization of $-t_f$ gives rise to the transversality condition $H_f = 1$ for $H = \boldsymbol{\lambda}_z^T \dot{z}$ since $H_f = 0$ for the augmented Hamiltonian $H = -1 + \boldsymbol{\lambda}_z^T \dot{z}$. We now use the full six-element formulation to solve five-state orbit-transfer problems by first starting from a fixed initial state and optimizing the final arrival point on the terminal orbit. Given $(a)_0$, $(e)_0$, $(i)_0$, $(\Omega)_0$, $(\omega)_0$, and $(M)_0$ at time $t_0 = 0$ or, equivalently, $(a)_0$, $(h)_0$, $(k)_0$, $(p)_0$, $(q)_0$, and $(\lambda)_0$, the initial values of the Lagrange multipliers, namely, $(\lambda_a)_0$, $(\lambda_h)_0$, $(\lambda_k)_0$, $(\lambda_p)_0$, $(\lambda_q)_0$, and $(\lambda_\lambda)_0$ are guessed, and the dynamic equations $\dot{z} = (\partial z/\partial \dot{\boldsymbol{r}}) \cdot \hat{u} f$ as well as the adjoint Eq. (14.155) are integrated forward to the guessed, transfer time t_f by using the optimal control $\hat{u} = [\boldsymbol{\lambda}_z^T M(z, F)]^T / |\boldsymbol{\lambda}_z^T M(z, f)|$. An iterative scheme next is used in order to adjust the initial values of the six multipliers as well as t_f such that the five terminal state parameters a_f, h_f, k_f, p_f, and q_f are matched and $(\lambda_\lambda)_f = 0$, $H_f = 1$ are satisfied. This is done by minimizing the following objective function:

$$F' = w_1(a - a_f)^2 + w_2(h - h_f)^2 + w_3(k - k_f)^2 + w_4(p - p_f)^2 + w_5(q - q_f)^2 + w_6(\lambda_\lambda - \lambda_{\lambda_f})^2 + w_7(H - H_f)^2$$

or

$$F' = \sum_{i=1}^{5} w_i(z_i - z_{i_f})^2 + w_6(\lambda_\lambda - 0)^2 + w_7(H - 1)^2$$

with w_i standing for certain weights that can be adjusted in order to favor the rapid convergence of some elements relative to others and alleviate, to some extent, certain sensitivity and scaling problems associated with the use of a given optimizer. For this purpose, we make use of the minimization algorithm UNCMIN of Ref. 14, which is designed for the unconstrained minimization of a real-valued function $F'(x)$ of n variables denoted by the vector x. This subroutine is based on a general descent method and uses a quasi-Newton algorithm. In the Newton method, the step p'' is computed from the solution of a set of n linear equations known as the Newton equations:

$$\nabla^2 F'(x) p'' = -\nabla F'(x) \tag{14.164}$$

Therefore, the solution is updated by using

$$x_{k+1} = x_k + p'' = x_k - \left[\nabla^2 F'(x_k)\right]^{-1} \nabla F'(x_k) \tag{14.165}$$

Here $\nabla F'(x)$ denotes the gradient of F' at x, whereas $\nabla^2 F'(x)$, the constant matrix of the second partial derivatives of F' at x, represents the Hessian matrix. The Newton direction given by p'' is guaranteed to be a descent direction only

if $[\nabla^2 F']^{-1}$ is positive definite, i.e., $z^T[\nabla^2 F']^{-1}z > 0$ for all $z \neq 0$ since, in that case, for small ε,

$$\begin{aligned} F'(x + \varepsilon p'') &= F'(x) + \varepsilon \nabla F'^T p'' + \mathcal{O}(\varepsilon^2) \\ &= F'(x) - \varepsilon \nabla F'^T [\nabla^2 F']^{-1} \nabla F' + \mathcal{O}(\varepsilon^2) \end{aligned} \tag{14.166}$$

such that $F'(x + \varepsilon p'') < F'(x)$. This is equivalent to requiring the linear term in Newton's quadratic approximation for the function F' to be negative. This is, of course, the second term in the Taylor series expansion for F' at $x + p''$. The algorithm UNCMIN builds a secant approximation B_k to the Hessian as the function is being minimized such that, at x_k,

$$B_k p'' = -\nabla F'_k \tag{14.167}$$

with the matrix B_k positive definite since there is no guarantee that $[\nabla^2 F']^{-1}$ will always be positive definite for each x. Once the descent direction is established, a line search on α' is established such that $F'(x_k + \alpha' p'') < F'_k$ and the solution updated via $x_{k+1} = x_k + \alpha' p''$. The approximate Hessian B_k is updated next using x_{k+1} and the gradient $\nabla F'(x_{k+1})$. For example, if F' is quadratic,[14]

$$B_{k+1}(x_{k+1} - x_k) = \nabla F'_{k+1} - \nabla F'_k \tag{14.168}$$

Only the gradient values are needed for the update of the approximate Hessian, which is achieved by finite differencing such that the user must provide only the function F' itself. Let the initial orbit be given by

$$a_0 = 7000 \text{ km}, \qquad e_0 = 0, \qquad i_0 = 28.5 \text{ deg}$$

$$\Omega_0 = 0 \text{ deg}, \qquad \omega_0 = 0 \text{ deg}, \qquad M_0 = -220 \text{ deg}$$

and let $f = 9.8 \times 10^{-5}$ km/s^2 or, roughly, $10^{-2}g$. The final orbit is given by

$$a_f = 42{,}000 \text{ km}, \qquad e_f = 10^{-3}, \qquad i_f = 1 \text{ deg}$$

$$\Omega_f = 0 \text{ deg}, \qquad \omega_f = 0 \text{ deg}$$

with M_f free. The following solution is obtained:

$$(\lambda_a)_0 = 1.260484756 \text{ s/km}$$

$$(\lambda_h)_0 = 3.865626962 \times 10^2 \text{ s}$$

$$(\lambda_k)_0 = -9.388262635 \times 10^3 \text{ s}$$

$$(\lambda_p)_0 = -2.277132367 \times 10^3 \text{ s}$$

$$(\lambda_q)_0 = -1.743027218 \times 10^4 \text{ s}$$

$$(\lambda_\lambda)_0 = 5.155487187 \times 10^2 \text{ s/rad}$$

$$t_f = 58{,}624.094 \text{ s}$$

The achieved orbit parameters at $t = t_f$ are:

$$a_f = 42{,}000.007 \text{ km}, \qquad e_f = 1.00022 \times 10^{-3}$$

$$i_f = 1.000012 \text{ deg}, \qquad \Omega_f = 359.999963 \text{ deg}$$

$$\omega_f = 1.966524 \times 10^{-2} \text{ deg}, \qquad M_f = 43.779715 \text{ deg}$$

with

$$H_f = 1.002694, \qquad (\lambda_\lambda)_f = -7.662506 \times 10^{-3} \text{ s/rad}$$

The value of $\lambda_f = 19.613998$ rad is such that, with $\lambda_0 = -3.839724$ rad, a total travel of $\Delta\lambda/(2\pi) = (\lambda_f - \lambda_0)/(2\pi) = 23.453722/(2\pi) = 3.732$ revolutions around the Earth is accomplished. The initial values of the equinoctial elements are given by

$$a_0 = 7000 \text{ km}, \qquad h_0 = 0, \qquad k_0 = 0, \qquad p_0 = 0$$

$$q_0 = 0.2539676465, \qquad \lambda_0 = -3.839724354 \text{ rad}$$

Figure 14.8 shows the variations of the classical elements a and e as functions of time during the 16.284470-h transfer. The eccentricity reaches a peak of around 0.4, with most of the orbit rotation taking place in the final 2-h of the transfer. Figure 14.9 depicts e and i as functions of the semimajor axis, which is monotonically increasing as shown in Fig. 14.8. The final position given by λ_f or M_f with $M_f = 43.779715$ deg is the optimal arrival point on the target orbit that results

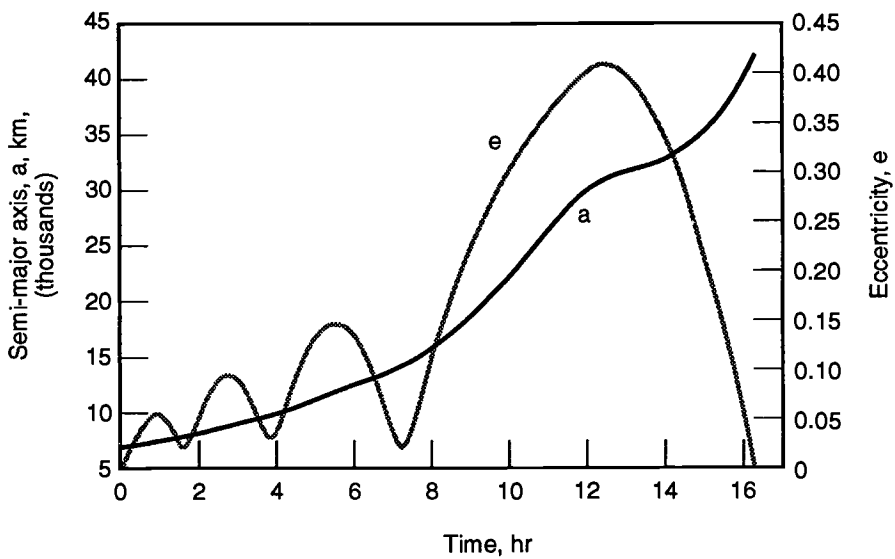


Fig. 14.8 Variation of semimajor axis and eccentricity for a LEO to a near-GEO transfer with initial M_0 = –220 deg (from Ref. 20).

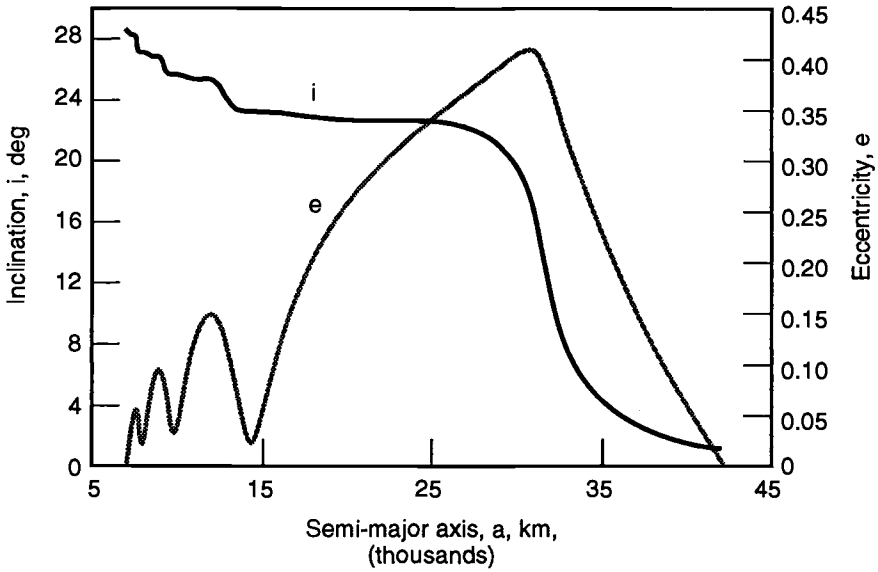


Fig. 14.9 Eccentricity and inclination vs semimajor axis for a LEO to a near-GEO transfer with initial $M_0 = -220$ deg (from Ref. 20).

in the minimum-time transfer for the given or fixed initial $M_0 = -220$ deg. The total ΔV is given by

$$\Delta V = f \cdot t_f = 5.745161 \text{ km/s}$$

We now solve the same transfer problem except that $M_0 = -300$ deg instead of -220 deg on the initial orbit at LEO. This will show whether any sensitivity to transfer time exists related to the departure location on the initial orbit. The solution now is given by

$$(\lambda_a)_0 = 7.131919334 \text{ s/km}$$

$$(\lambda_h)_0 = -6.586531878 \times 10^2 \text{ s}$$

$$(\lambda_k)_0 = 9.135791613 \times 10^3 \text{ s}$$

$$(\lambda_p)_0 = 1.439828925 \times 10^3 \text{ s}$$

$$(\lambda_q)_0 = -2.414079986 \times 10^4 \text{ s}$$

$$(\lambda_\lambda)_0 = -3.736240460 \times 10^2 \text{ s/rad}$$

$$t_f = 58{,}158.832 \text{ s}$$

This transfer requires a flight time of some 465.262 s less than the previous transfer, or some 7.75 min. The achieved parameters are

$$a_f = 42{,}000.012 \text{ km}, \qquad e_f = 9.9965 \times 10^{-4}$$

$$i_f = 1.000030 \text{ deg}, \qquad \Omega_f = 359.998924 \text{ deg}$$

$$\omega_f = 359.975378 \text{ deg}, \qquad M_f = 227.155823 \text{ deg}$$

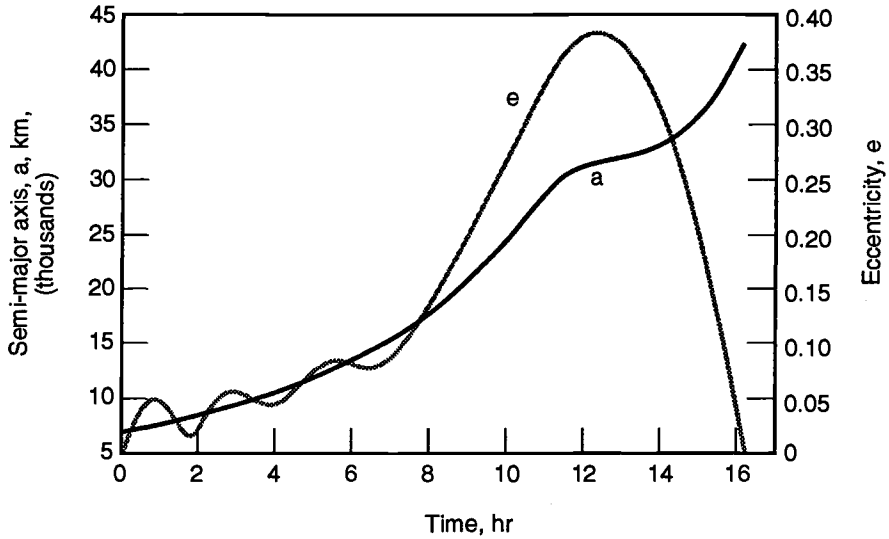


Fig. 14.10 Time history of semimajor axis and eccentricity for a LEO to a near-GEO transfer with initial M_0 = –300 deg (from Ref. 20).

with

$$H_f = 1.000000000$$

$$(\lambda_\lambda)_f = -8.872380 \times 10^{-3} \text{ s/rad}$$

and

$$\lambda_0 = -5.235987 \text{ rad}$$

$$\lambda_f = 16.530539 \text{ rad}$$

such that the total angular travel consists of $(\lambda_f - \lambda_0)/(2\pi) = 21.766526/(2\pi) = 3.464$ revolutions around the Earth. Figure 14.10 shows the variations of the classical orbit elements a and e during the optimal transfer with smaller oscillation amplitudes in the eccentricity during the first revolutions. Finally, Fig. 14.11 shows e and i as functions of the semimajor axis with essentially linear buildup and decay for e. The total ΔV is given by

$$\Delta V = f \cdot t_f = 5.699565 \text{ km/s}.$$

Minimum-Time Transfer with Optimized Departure and Arrival Locations

In this problem, we are given $(a)_0$, $(e)_0$, $(i)_0$, $(\Omega)_0$, and $(\omega)_0$ or, equivalently, $(a)_0$, $(h)_0$, $(k)_0$, $(p)_0$, and $(q)_0$. The guessed parameters are $(\lambda_a)_0$, $(\lambda_h)_0$, $(\lambda_k)_0$, $(\lambda_p)_0$, $(\lambda_q)_0$, and the initial mean longitude $(\lambda)_0$ as well as the transfer time t_f. The precision integration of the dynamic and adjoint systems in Eqs. (14.148–14.153)

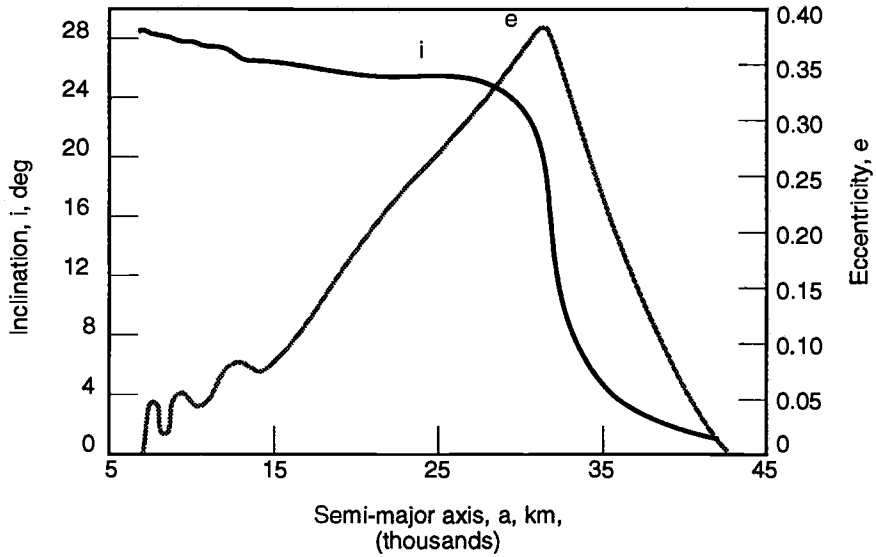


Fig. 14.11 Eccentricity and inclination vs semimajor axis for a LEO to a near-GEO transfer with initial M_0 = –300 deg (from Ref. 20).

and Eq. (14.155) is carried out using $(\lambda_\lambda)_0 = 0$. The boundary conditions at the unknown final time are given by $a_f, h_f, k_f, p_f, q_f, (\lambda_\lambda)_f = 0$, and $H_f = 1$ for a minimum-time solution. The optimal thrust direction still is given by

$$\hat{u} = \left[\lambda_z^T M(z, F)\right]^T / \left|\lambda_z^T M(z, F)\right|$$

The objective function to minimize still is given by F' above. We have essentially traded $(\lambda)_0$ given, with $(\lambda_\lambda)_0 = 0$, since small variations in $(\lambda)_0$ must be such that the performance index $J = -t_f$ is stationary, indicating zero sensitivity to initial location. The same argument still holds true for the arrival point on the target orbit. The solution of the optimal free-free transfer is given by

$$(\lambda_a)_0 = 4.8548563957 \text{ s/km}$$

$$(\lambda_h)_0 = 5.52370740318 \times 10^2 \text{ s}$$

$$(\lambda_k)_0 = -9.51431194293 \times 10^3 \text{ s}$$

$$(\lambda_p)_0 = -1.0373235843 \times 10^2 \text{ s}$$

$$(\lambda_q)_0 = -2.33561012603 \times 10^4 \text{ s}$$

$$(\lambda)_0 = -2.272581909 \text{ rad}$$

corresponding to $M_0 = -130.209352$ deg, and the overall minimum-time $t_f =$ 58,090.031 s, with corresponding $\Delta V = 5.692823$ km/s. The total angular travel

is obtained from $(\lambda)_0$ and $(\lambda)_f = 19.65399449$ rad or 3.4897 revolutions around the Earth. The final achieved parameters are:

$$a_f = 42{,}000.001 \text{ km}$$

$$e_f = 9.78045 \times 10^{-4}$$

$$i_f = 0.999359 \text{ deg}$$

$$\Omega_f = 358.777222 \text{ deg}$$

$$\omega_f = 350.922884 \text{ deg}$$

$$M_f = 56.390827 \text{ deg}$$

with

$$H_f = 1.038422077$$

and

$$(\lambda_\lambda)_f = -8.6830309 \times 10^{-1} \text{ s/rad}$$

The Hamiltonian is constant because the dynamic equations are not explicit functions of time. Furthermore, we can multiply all the initial values of the multipliers by a common constant factor in order to get $H = 1$ exactly, if we so desire. The solution is, of course, unchanged because of the common scaling used. This also means that we can arbitrarily select one of the λ and reduce the order of the integration by 1 since H, a positive constant, is in effect a first integral of the motion. The initial values of the multipliers that scale the Hamiltonian to the unit value are:

$$(\lambda_a)_0 = 4.675224557 \text{ s/km}$$

$$(\lambda_h)_0 = 5.319327780 \times 10^2 \text{ s}$$

$$(\lambda_k)_0 = -9.162278183 \times 10^3 \text{ s}$$

$$(\lambda_p)_0 = -9.989421518 \times 10^1 \text{ s}$$

$$(\lambda_q)_0 = -2.249191516 \times 10^4 \text{ s}$$

Figure 14.12 shows the variations of e and i as functions of the semimajor axis. Most of the inclination change is taking place at or near maximum eccentricity. The thrust pitch and yaw programs are shown in Fig. 14.13. The yaw profile, as expected, changes sign every one-half revolution during the first three orbits, rotating the orbit slowly, but the pitch angle stays near zero for maximum energy buildup. The sharp buildup in the yaw angle is responsible for most of the orbit rotation that takes place as soon as eccentricity reaches its maximum value as we have observed. Figure 14.14 shows r vs θ, the angular position with $\theta = \omega + \theta^*$. As soon as the orbit reaches the proper energy level, it is able to transfer directly to the GEO altitude on a highly eccentric orbit. The last portion of the trajectory

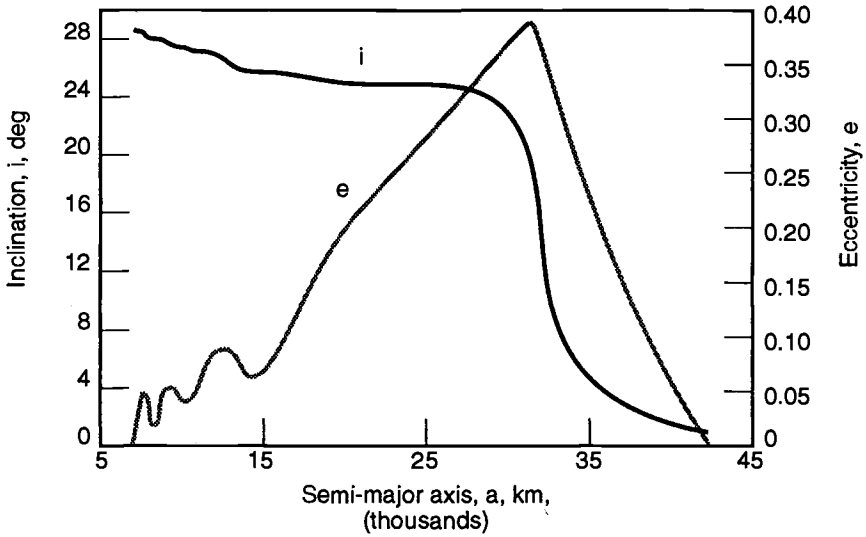


Fig. 14.12 Eccentricity and inclination vs semimajor axis for absolute minimum-time LEO to near-GEO transfer (from Ref. 20).

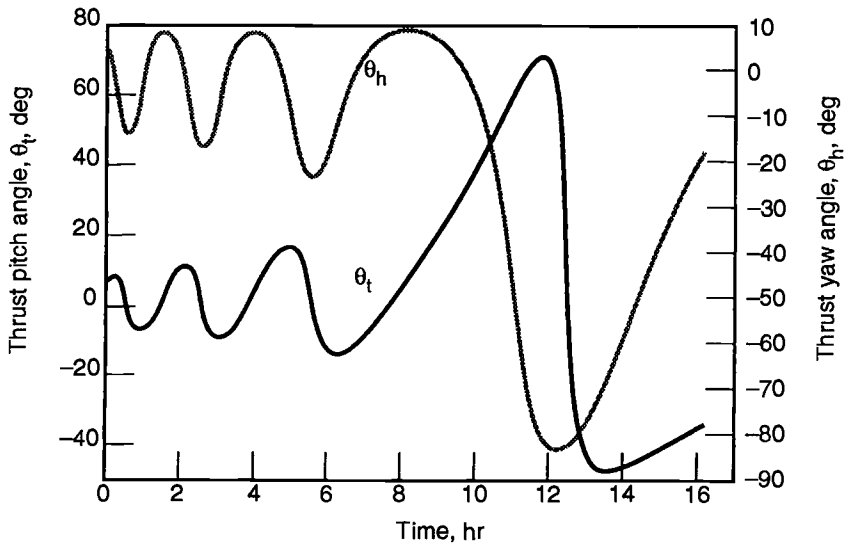


Fig. 14.13 Optimal thrust pitch and yaw programs for absolute minimum-time LEO to near-GEO transfer (from Ref. 20).

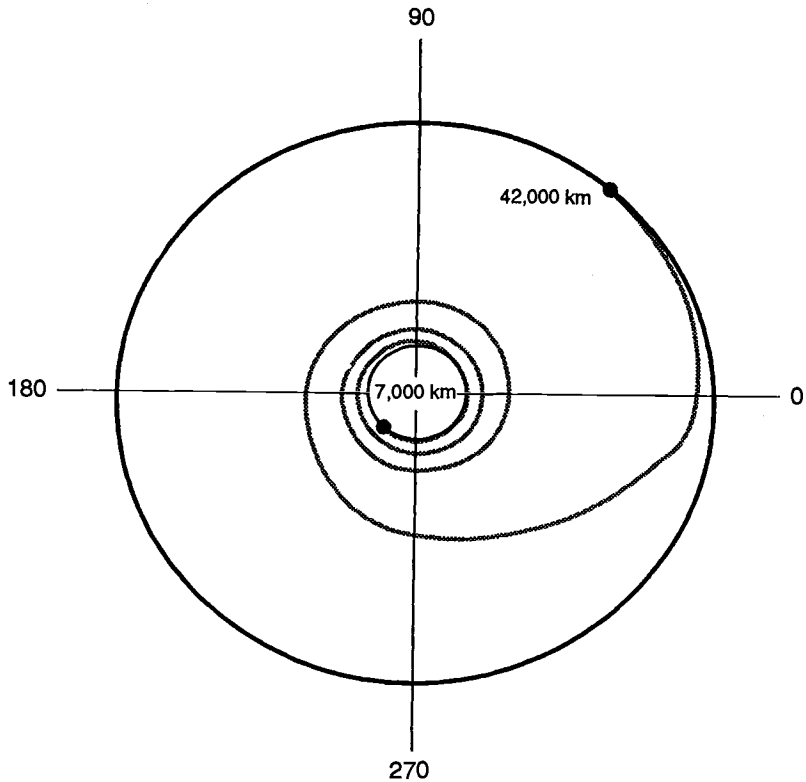


Fig. 14.14 Polar plot of absolute minimum-time LEO to near-GEO transfer (from Ref. 20).

consists of a circularization maneuver in order to cancel the eccentricity and enter the GEO orbit. The pitch angle θ_t becomes negative during the last portion of this phase, as shown in Fig. 14.13 for final eccentricity control. The θ_t is such that -90 deg $< \theta_t <$ 90 deg at all times, indicating that the thrust is never applied in a breaking mode. This is not always the case, especially for larger inclination changes or larger orbit transfers. As another example, the solution for the optimal transfer from

$$a_0 = 7000 \text{ km}, \qquad e_0 = 0, \qquad i_0 = 28.5 \text{ deg}$$
$$\Omega_0 = 0 \text{ deg}, \qquad \omega_0 = 0 \text{ deg}$$

to a circular equatorial orbit given by

$$a_f = 42{,}000 \text{ km}, \qquad e_f = 0, \qquad i_f = 0 \text{ deg}$$
$$\Omega_f = 0 \text{ deg}, \qquad \omega_f = 0 \text{ deg}$$

with $f = 9.8 \times 10^{-5}$ km/s^2 is given by:

$$(\lambda_a)_0 = 4.613252551 \text{ s/km}$$

$$(\lambda_h)_0 = 5.650656277 \times 10^2 \text{ s}$$

$$(\lambda_k)_0 = -9.232366500 \times 10^3 \text{ s}$$

$$(\lambda_p)_0 = -1.043515373 \times 10^2 \text{ s}$$

$$(\lambda_q)_0 = -2.391075386 \times 10^4 \text{ s}$$

$$(\lambda)_0 = -2.331552763 \text{ rad corresponding to } M_0 = -133.588133 \text{ deg,}$$

with

$$t_f = 58{,}325.693 \text{ s} \quad \text{and} \quad \Delta V = 5.715918 \text{ km/s.}$$

The first example used at the end of Sec. 14.2 with Edelbaum's theory consisted of a transfer from $a_0 = 7000$ km, $i_0 = 28.5$ deg to $a_f = 42166$ km, $i_f = 0$ deg with a low acceleration of $f = 3.5 \times 10^{-7}$ km/s^2 The optimal transfer for that same example with

$$e_0 = 0, \qquad \Omega_0 = 0 \text{ deg}, \qquad \omega_0 = 0 \text{ deg}, \qquad e_f = 0$$

$$\Omega_f = 0 \text{ deg}, \qquad \omega_f = 0 \text{ deg}, \qquad \text{and} \qquad f = 9.8 \times 10^{-5} \text{ km/s}^2$$

is given by:

$$(\lambda_a)_0 = 4.632642932 \text{ s/km}$$

$$(\lambda_h)_0 = 5.521541143 \times 10^2 \text{ s}$$

$$(\lambda_k)_0 = -9.247467017 \times 10^3 \text{ s}$$

$$(\lambda_p)_0 = -1.005059692 \times 10^2 \text{ s}$$

$$(\lambda_q)_0 = -2.368567314 \times 10^4 \text{ s}$$

$$(\lambda)_0 = -2.328897129 \text{ rad corresponding to } M_0 = -133.435976 \text{ deg}$$

$$t_f = 58{,}413.667 \text{ s with corresponding } \Delta V = 5.724539 \text{ km/s.}$$

The ΔV_{tot} from Edelbaum's formula, being independent of the acceleration level, yielded $\Delta V_{\text{tot}} = 5.77584$ km/s, or some 50 m/s more than the exact solution, which is quite accurate even for this relatively high-acceleration example.

Continuous Constant Acceleration Orbit Transfer Using Averaging

Following Ref. 6, an averaged Hamiltonian $\tilde{H}$ is formed first, from which a first-order approximation to the state and costate is derived by holding these quantities

constant over the averaging interval and varying only F, the eccentric longitude on a conic orbit.

$$\tilde{H} = \frac{1}{T_0}\int_0^{T_0} H\,\mathrm{d}t = \frac{1}{T_0}\int_{-\pi}^{\pi} \frac{H\,\mathrm{d}F}{\dot{F}(\tilde{z}, F)} \tag{14.169}$$

where T_0, the orbital period at time t, is given by $T_0 = 2\pi/\tilde{n}$, with $\tilde{n} = \mu^{1/2}\tilde{a}^{-3/2}$ and $\tilde{a}$ indicating the averaged value of the semimajor axis at time t, and so forth. From Kepler's equation, $\tilde{\lambda} = F - \tilde{k}s_F + \tilde{h}c_F = \tilde{n}t$, we have

$$\dot{F} = \frac{2\pi}{T_0(1 - \tilde{k}c_F - \tilde{h}s_F)} \tag{14.170}$$

such that, with

$$s = \frac{1/T_0}{\dot{F}} = \frac{1}{2\pi}(1 - \tilde{k}c_F - \tilde{h}s_F)$$

the equations for the approximation for the state and costate variables are given by[6]

$$\dot{\tilde{z}} = \left(\frac{\partial \tilde{H}}{\partial \tilde{\lambda}_z}\right)^T = \int_{-\pi}^{\pi} \left(\frac{\partial H}{\partial \tilde{\lambda}_z}\right)^T s(\tilde{z}, F)\,\mathrm{d}F \tag{14.171}$$

$$\dot{\tilde{\lambda}}_z = -\left(\frac{\partial \tilde{H}}{\partial \tilde{z}}\right)^T = -\int_{-\pi}^{\pi} \left[\left(\frac{\partial H}{\partial \tilde{z}}\right)^T s(\tilde{z}, F) + H\left(\frac{\partial s}{\partial \tilde{z}}\right)^T\right]\mathrm{d}F \tag{14.172}$$

The partials $\partial s/\partial \tilde{z}$ are given by

$$\frac{\partial s}{\partial \tilde{a}} = 0, \qquad \frac{\partial s}{\partial \tilde{h}} = -\frac{s_F}{2\pi}, \qquad \frac{\partial s}{\partial \tilde{k}} = -\frac{c_F}{2\pi}, \qquad \frac{\partial s}{\partial \tilde{p}} = 0, \qquad \frac{\partial s}{\partial \tilde{q}} = 0$$

and

$$\frac{\partial s}{\partial \tilde{\lambda}} = \frac{(\tilde{k}s_F - \tilde{h}c_F)}{[(1 - \tilde{k}c_F - \tilde{h}s_F)2\pi]}$$

since

$$\frac{\partial s}{\partial \tilde{\lambda}} = \left(\frac{\partial s}{\partial F}\right)\cdot\left(\frac{\partial F}{\partial \tilde{\lambda}}\right)$$

and

$$\tilde{\lambda} = F - \tilde{k}s_F + \tilde{h}c_F$$

The optimal trajectory is such that H is constant throughout. This means that, when thrust is applied continuously during every complete revolution, the contribution

from the $H(\partial s/\partial \tilde{z})^T$ term is equal to zero because the $\partial s/\partial \tilde{z}$ terms will give a zero net contribution. For a continuous constant acceleration f, the averaged equations are given by

$$\dot{\tilde{a}} = \frac{1}{2\pi} f \int_{-\pi}^{\pi} \left(\frac{\partial a}{\partial \dot{\boldsymbol{r}}}\right)^T \cdot \hat{u}(1 - \tilde{k}c_F - \tilde{h}s_F)\mathrm{d}F \tag{14.173}$$

$$\dot{\tilde{h}} = \frac{1}{2\pi} f \int_{-\pi}^{\pi} \left(\frac{\partial h}{\partial \dot{\boldsymbol{r}}}\right)^T \cdot \hat{u}(1 - \tilde{k}c_F - \tilde{h}s_F)\mathrm{d}F \tag{14.174}$$

$$\dot{\tilde{k}} = \frac{1}{2\pi} f \int_{-\pi}^{\pi} \left(\frac{\partial k}{\partial \dot{\boldsymbol{r}}}\right)^T \cdot \hat{u}(1 - \tilde{k}c_F - \tilde{h}s_F)\mathrm{d}F \tag{14.175}$$

$$\dot{\tilde{p}} = \frac{1}{2\pi} f \int_{-\pi}^{\pi} \left(\frac{\partial p}{\partial \dot{\boldsymbol{r}}}\right)^T \cdot \hat{u}(1 - \tilde{k}c_F - \tilde{h}s_F)\mathrm{d}F \tag{14.176}$$

$$\dot{\tilde{q}} = \frac{1}{2\pi} f \int_{-\pi}^{\pi} \left(\frac{\partial q}{\partial \dot{\boldsymbol{r}}}\right)^T \cdot \hat{u}(1 - \tilde{k}c_F - \tilde{h}s_F)\mathrm{d}F \tag{14.177}$$

$$\begin{aligned}\dot{\tilde{\lambda}} = {} & \frac{1}{2\pi} f \int_{-\pi}^{\pi} \left(\frac{\partial \lambda}{\partial \dot{\boldsymbol{r}}}\right)^T \cdot \hat{u}(1 - \tilde{k}c_F - \tilde{h}s_F)\mathrm{d}F \\ & + \frac{1}{2\pi} \int_{-\pi}^{\pi} \tilde{n}(1 - \tilde{k}c_F - \tilde{h}s_F)\mathrm{d}F\end{aligned} \tag{14.178}$$

$$\begin{aligned}\dot{\tilde{\lambda}}_a = {} & \frac{1}{2\pi} f \int_{-\pi}^{\pi} -\tilde{\boldsymbol{\lambda}}_z^T \frac{\partial M}{\partial \tilde{a}} \cdot \hat{u}(1 - \tilde{k}c_F - \tilde{h}s_F)\mathrm{d}F \\ & + \frac{1}{2\pi} \int_{-\pi}^{\pi} -\tilde{\lambda}_\lambda \frac{\partial \tilde{n}}{\partial \tilde{a}}(1 - \tilde{k}c_F - \tilde{h}s_F)\mathrm{d}F\end{aligned} \tag{14.179}$$

$$\dot{\tilde{\lambda}}_h = \frac{1}{2\pi} f \int_{-\pi}^{\pi} -\tilde{\boldsymbol{\lambda}}_z^T \frac{\partial M}{\partial \tilde{h}} \cdot \hat{u}(1 - \tilde{k}c_F - \tilde{h}s_F)\mathrm{d}F \tag{14.180}$$

$$\dot{\tilde{\lambda}}_k = \frac{1}{2\pi} f \int_{-\pi}^{\pi} -\tilde{\boldsymbol{\lambda}}_z^T \frac{\partial M}{\partial \tilde{k}} \cdot \hat{u}(1 - \tilde{k}c_F - \tilde{h}s_F)\mathrm{d}F \tag{14.181}$$

$$\dot{\tilde{\lambda}}_p = \frac{1}{2\pi} f \int_{-\pi}^{\pi} -\tilde{\boldsymbol{\lambda}}_z^T \frac{\partial M}{\partial \tilde{p}} \cdot \hat{u}(1 - \tilde{k}c_F - \tilde{h}s_F)\mathrm{d}F \tag{14.182}$$

$$\dot{\tilde{\lambda}}_q = \frac{1}{2\pi} f \int_{-\pi}^{\pi} -\tilde{\boldsymbol{\lambda}}_z^T \frac{\partial M}{\partial \tilde{q}} \cdot \hat{u}(1 - \tilde{k}c_F - \tilde{h}s_F)\mathrm{d}F \tag{14.183}$$

$$\dot{\tilde{\lambda}}_\lambda = \frac{1}{2\pi} f \int_{-\pi}^{\pi} -\tilde{\boldsymbol{\lambda}}_z^T \frac{\partial M}{\partial \tilde{\lambda}} \cdot \hat{u}(1 - \tilde{k}c_F - \tilde{h}s_F)\mathrm{d}F \tag{14.184}$$

The thrust direction $\hat{u}$ is a function of $\tilde{z}$ and F as well as $\tilde{\boldsymbol{\lambda}}_z$, and the second term in Eq. (14.178) is equal to zero. An eight-order Gauss–Legendre quadrature is used to evaluate the averaged rates given by Eqs. (14.173–14.184), which then are used by a variable step size integrator to integrate trajectories from given initial conditions, with step sizes spanning several orbital revolutions. The two-point boundary-value problem is identical to the one used in the preceding subsection, except that all variables as well as the Hamiltonian have averaged values. The averaged Hamiltonian is obtained from the integrated variables

$$\tilde{H} = \tilde{\lambda}_a \dot{\tilde{a}} + \tilde{\lambda}_h \dot{\tilde{h}} + \tilde{\lambda}_k \dot{\tilde{k}} + \tilde{\lambda}_p \dot{\tilde{p}} + \tilde{\lambda}_q \dot{\tilde{q}} + \tilde{\lambda}_\lambda \dot{\tilde{\lambda}}$$

The same example is used with the optimized $(\lambda)_0$ or, equivalently,

$$(M)_0 = -130.209352 \text{ deg}$$

yielding the following solution given by

$$\left(\tilde{\lambda}_a\right)_0 = 0.597196701 \times 10^{-4} \text{ day/km}$$

$$\left(\tilde{\lambda}_h\right)_0 = -0.167705951 \times 10^{-11} \text{ day}$$

$$\left(\tilde{\lambda}_k\right)_0 = 0.502638000 \times 10^{-3} \text{ day}$$

$$\left(\tilde{\lambda}_p\right)_0 = -0.161888685 \times 10^{-12} \text{ day}$$

$$\left(\tilde{\lambda}_q\right)_0 = -0.967633609 \text{ day}$$

$$\left(\tilde{\lambda}_\lambda\right)_0 = 0 \text{ day/rad}$$

$$t_f = 0.6566499459 \text{ days (or 56,734.555 s)}$$

The achieved final orbit is given by

$$\tilde{a}_f = 41{,}999.999 \text{ km}, \qquad \tilde{e}_f = 9.999999 \times 10^{-4}$$

$$\tilde{i}_f = 1.000000 \text{ deg}, \qquad \tilde{\Omega}_f = 360.000000 \text{ deg}$$

$$\tilde{\omega}_f = 359.999999 \text{ deg}, \qquad \tilde{M}_f = 127.049617 \text{ deg}$$

with $\tilde{\lambda}_f = 21.066990$ rad, $(\tilde{\lambda}_\lambda)_f = 0$ day/rad, and $\tilde{H}_f = 1.000000007$ indicating a converged solution. Since $(\tilde{\lambda})_0 = -2.272581$ rad, the transfer requires $(\tilde{\lambda}_f - \tilde{\lambda}_0)/(2\pi) = 3.7146$ revolutions of the Earth. This information is available since we also have integrated the $\dot{\tilde{\lambda}}$ equation. The multiplier $\tilde{\lambda}_\lambda$ stays constant at zero throughout the transfer as it should because of the averaging procedure, which eliminates all sensitivity to the position along the orbit. This is verified by starting from $(\tilde{\lambda})_0 = 0$ rad or $\tilde{M}_0 = 0$ deg and generating the exact duplicate trajectory except that $(\tilde{\lambda})_f = 23.339571$ rad or $\tilde{M}_f = 257.258969$ deg such that the transfer still requires 3.7146 revolutions of the Earth. This averaged transfer requires some 1355.476 s less time than the precision-integrated exact transfer

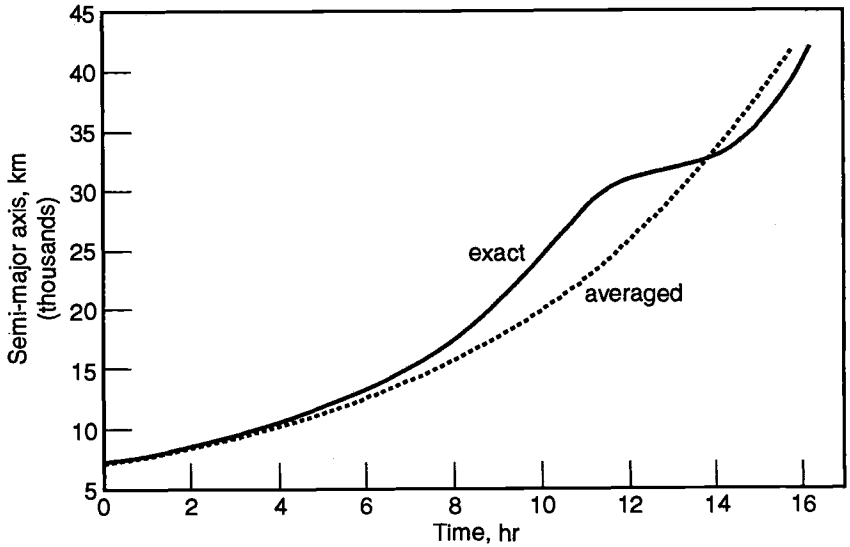


Fig. 14.15 Evolution of exact and averaged solutions for the semimajor axis for a LEO to near-GEO transfer (from Ref. 20).

of the preceding subsection, which required $t_f = 58{,}090.031$ s. Therefore, the averaged transfer solution is optimistic by some 22.591 min, requiring a ΔV of 5.559986 km/s, which is rather accurate. In order to compare the various λ, we first must convert the units from days to seconds such that

$$(\tilde{\lambda}_a)_0 = 5.159779497 \text{ s/km}$$

$$(\tilde{\lambda}_h)_0 = -1.448979417 \times 10^{-7} \text{ s}$$

$$(\tilde{\lambda}_k)_0 = 4.342792320 \times 10^{1} \text{ s}$$

$$(\tilde{\lambda}_p)_0 = -1.398718238 \times 10^{-8} \text{ s}$$

$$(\tilde{\lambda}_q)_0 = -8.360354382 \times 10^{4} \text{ s}$$

$$(\tilde{\lambda}_\lambda)_0 = 0 \text{ s/rad}$$

The exact and averaged plots for the semimajor axis and inclination are shown in Figs. 14.15 and 14.16. On the other hand, the averaged classical elements $\tilde{\Omega}$ and $\tilde{\omega}$ stay at zero while the eccentricity $\tilde{e}$ slowly increases from 0 to 10^{-3} at the final time. The $\tilde{\lambda}_a$ multiplier is shown in Fig. 14.17 with its exact counterpart. The effect of averaging is to smooth out all the fluctuations seen in the exact curves. The main shortcoming of the averaging of the dynamics is the inability to describe the important eccentricity buildup seen in the exact solution. However, the ΔV evaluation or, equivalently, transfer time is quite accurate even for this difficult, short-duration, large transfer problem. If we now use an acceleration f of 3.5×10^{-7}km/s^2, this averaged transfer will be achieved in the minimum time

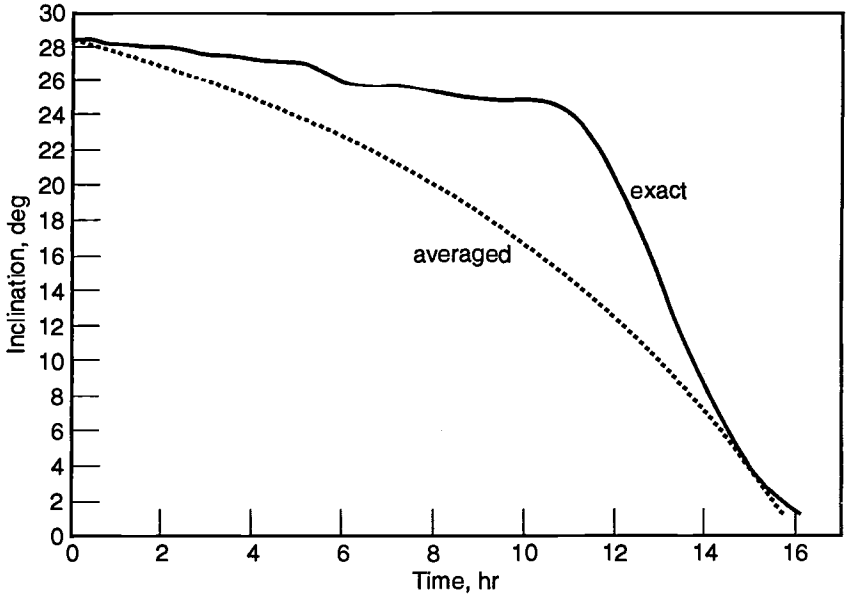


Fig. 14.16 Evolution of exact and averaged solutions for the inclination for a LEO to near-GEO transfer (from Ref. 20).

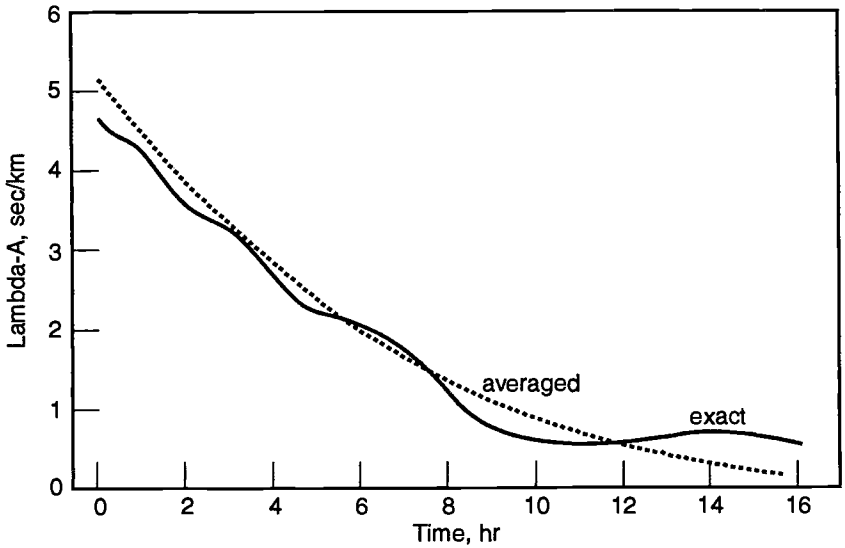


Fig. 14.17 Comparison between exact and averaged solutions for λ_a for a LEO to near-GEO transfer (from Ref. 20).

of 183.861985 days, or $t_f = 1.588567550 \times 10^7$ s. This corresponds to a ΔV of 5.559986 km/s, which is identical to the ΔV required by the fast transfer using averaging. This shows that the averaged transfer time is exactly inversely proportional to the acceleration such that $\Delta V = f \cdot t_f$ remains constant, regardless of the duration of the transfer. Going back to the example of Fig. 14.1, consisting of a LEO to GEO transfer with

$$\tilde{a}_0 = 7000 \text{ km}, \qquad \tilde{i}_0 = 28.5 \text{ deg}, \qquad \tilde{e}_0 = 0, \qquad \tilde{\Omega}_0 = 0 \text{ deg}$$

$$\tilde{\omega}_0 = 0 \text{ deg}, \qquad \tilde{a}_f = 42{,}166 \text{ km}, \qquad \tilde{e}_f = 0$$

$$\tilde{i}_f = 0 \text{ deg}, \qquad \tilde{\Omega}_f = 0 \text{ deg}, \qquad \tilde{\omega}_f = 0 \text{ deg}$$

and using $f = 9.8 \times 10^{-5}$ km/s^2, the optimal solution using averaging is given by

$$(\tilde{\lambda}_a)_0 = 0.595400784 \times 10^{-4} \text{ day/km}$$

$$(\tilde{\lambda}_h)_0 = 0.455846926 \times 10^{-11} \text{ day}$$

$$(\tilde{\lambda}_k)_0 = -0.708506385 \times 10^{-11} \text{ day}$$

$$(\tilde{\lambda}_p)_0 = 0.706094060 \times 10^{-11} \text{ day}$$

$$(\tilde{\lambda}_q)_0 = -0.986509235 \text{ day}$$

$$(\tilde{\lambda}_\lambda)_0 = 0 \text{ day/rad}$$

$$t_f = 57{,}505.1722 \text{ s}$$

with corresponding $\Delta V = 5.635506$ km/s. The exact solution requires a ΔV of 5.724539 km/s and the Edelbaum's $\Delta V_{\text{tot}} = 5.78378$ km/s. The solution of this problem for $f = 3.5 \times 10^{-7}$ km/s^2 with the present averaging method is given by

$$(\tilde{\lambda}_a)_0 = 0.166712215 \times 10^{-1} \text{ day/km}$$

$$(\tilde{\lambda}_h)_0 = 0.148064422 \times 10^{-4} \text{ day}$$

$$(\tilde{\lambda}_k)_0 = 0.184348796 \times 10^{-5} \text{ day}$$

$$(\tilde{\lambda}_p)_0 = 0.349863862 \times 10^{-5} \text{ day}$$

$$(\tilde{\lambda}_q)_0 = -2.762226204 \times 10^{2} \text{ day}$$

$$(\tilde{\lambda}_\lambda)_0 = 0 \text{ day/rad}$$

$$t_f = 1.610144786 \times 10^{7} \text{ s}$$

with corresponding $\Delta V = 5.635506$ km/s. This is only some 148 m/s less than Edelbaum's ΔV. The exact solution for this case would require very long integration times, and it has not been attempted. However, it is expected that the

exact ΔV will be very close to the value of 5.635506 km/s since the transfer orbit will remain nearly circular throughout the exact transfer; this is the case with the averaged solution, too, which also uses a variable thrust yaw and pitch profile within each revolution during the ascent. Gravity losses are inherently more important when the orbit departs further from the circular shape as is the case with the higher acceleration of $f = 9.8 \times 10^{-5}$ km/s^2. The averaged solution fails to account effectively for the gravity losses since it provides an optimistic ΔV with respect to the exact solution, whereas the Edelbaum's ΔV_{tot} equation is overall more pessimistic, reflecting the losses associated with the use of a nonoptimal thrust profile with fixed yaw angle within each revolution. This more than compensates its inherent optimistic characteristics since it assumes a perfectly circular orbit with no gravity losses.

14.5 Orbit Transfer with Variable Specific Impulse

The problem of minimum-fuel time-fixed orbit transfer and rendezvous using continuous low thrust bounded from above and below is analyzed next. Specific impulse, or I_{sp}, is no longer considered to be a fixed quantity during the transfer and is now allowed to vary between well-defined minimum and maximum bounds such that both thrust magnitude and direction are optimized to yield the overall minimum-fuel solution. For example, for the case of the electric engine, we can assume that the power remains constant such that the thrust magnitude is inversely proportional to the specific impulse, which is continuously adjusted by varying the beam voltage. We first revisit the fundamentals of flight mechanics and low-thrust propulsion developed in Refs. 12, 17, and 18 and derive the equivalent expressions for the optimal controls for the thrust-magnitude unconstrained case using equinoctial elements instead of the usual Cartesian coordinates. The necessary conditions for optimality for the thrust-bounded case are derived following Refs. 11–13 and the problem of low-thrust transfer and rendezvous of Refs. 6, 15, and 19 extended to the case of continuously varying I_{sp}.

The Optimization of the Thrust Magnitude

From rocket propulsion fundamentals and using Newton's law for a variable mass body, the equation of motion of a rocket-powered vehicle is given by

$$m\ddot{\boldsymbol{r}} = \dot{m}\boldsymbol{c} + m\boldsymbol{g}$$

where $\boldsymbol{g}$ is the acceleration of gravity, $\boldsymbol{c}$ is the exhaust velocity, and $\dot{m} < 0$ is the rate at which mass is expelled from the engine. In this chapter, we use c for the exhaust velocity since $\hat{\boldsymbol{u}}$ is used to define the thrust direction. The thrust vector $\boldsymbol{T} = \dot{m}\boldsymbol{c}$ is directed opposite the exhaust velocity vector such that the acceleration can be written as

$$\boldsymbol{f} = \frac{\boldsymbol{T}}{m} = \ddot{\boldsymbol{r}} - \boldsymbol{g}$$

For a low-thrust electric rocket, the exhaust stream or beam power can be written as

$$P_B = \frac{Tc}{2} \tag{14.185}$$

and, since $T = -\dot{m}c$, it can be written as

$$P_B = -\frac{T^2}{2\dot{m}} \tag{14.186}$$

It can also be expressed in terms of the beam voltage V_B and the beam current I_B as

$$P_B = V_B I_B$$

Equation (14.185) can be derived from the following two expressions based on electrostatics considerations:

$$V_B e' = \frac{m_i}{2} c^2 \tag{14.187}$$

$$I_B = \left(\frac{e'}{m_i}\right)(-\dot{m})$$

The first expression states that if a particle of mass m_i and charge e' and negligible initial velocity passes through a potential difference V_B, it will acquire a kinetic energy of $1/2 m_i c^2$. The second expression is the definition of the current such that the beam power P_B is written as

$$P_B = I_B V_B = \frac{1}{2}(-\dot{m})c^2$$

and, since $T = -\dot{m}c$, then,

$$P_B = \frac{1}{2} T c$$

Alternately, from Eq. (14.187), an expression relating the I_{sp} to the beam voltage can be obtained since, with $c = I_{sp} g$,

$$c = \left[2\left(\frac{e'}{m_i}\right) V_B\right]^{1/2}$$

$$I_{sp} = \frac{1}{g}\left[2\left(\frac{e'}{m_i}\right) V_B\right]^{1/2}$$

For a given beam power P_B, the thrust vs mass flow rate curve of an electric rocket is parabolic since, from Eq. (14.186),

$$T = \sqrt{-2\dot{m} P_B} \tag{14.188}$$

This behavior is very different from that of a constant exhaust velocity rocket since, in the latter case, the curve is linear,

$$T = -\dot{m}c$$

Conversely, the mass flow rate expressions for both types of vehicles are

$$\dot{m} = -\frac{T}{c}$$

$$\dot{m} = \frac{-T^2}{2P_B} \tag{14.189}$$

It is therefore advantageous from a propellant consumption point of view to have high exhaust velocity or high power. Furthermore, Eq. (14.188) shows that the same level of thrust can be achieved by different combinations of $\dot{m}$ and P_B since, if a lower power level is selected, an appropriate increase in $\dot{m}$ will maintain T constant. However, from Eq. (14.189) and for a given T, it is seen that $\dot{m}$ is at a minimum if P_B is chosen at its maximum level. This means that, at each instant of time, the selection of P_B at its maximum value, namely, $P_{B_{\max}}$, will achieve the required thrust for minimum $\dot{m}$ or minimum fuel consumption. In other words, from all the possible ways of flying a required trajectory, meaning an acceleration time history $f(t) = \ddot{r}(t) - g(t)$, the selection of $P_{B_{\max}}$ is the only one that results in minimum fuel expenditure.[12] Electric rockets must therefore always operate at $P_{B_{\max}}$. Then, from Eq. (14.185) with $P_B = P_{B_{\max}}$, as the thrust is decreased, I_{sp} will increase and vice versa. Since the I_{sp} is dependent on the beam voltage V_B, it can be obtained by adjusting V_B, provided that I_B is also adjusted to obtain the required $P_{B_{\max}}$ power since $P_B = I_B V_B$. All these fundamental ideas are encapsulated in Figs. 14.18–14.20. In Fig. 14.18, the thrust vs mass flow rate curve for a low-thrust rocket with constant exhaust velocity is depicted. Since c cannot be varied in this type of rocket, the thruster operates at $T_{\max}$ only.[12]

Figure 14.19 corresponds to the variable c case with $T = (-2\dot{m}P)^{1/2}$ for given power P_B or for short, P. The power levels below $P_{\max}$ result in Region II, with Region I completely inaccessible since it corresponds to $P > P_{\max}$, which is impossible. As is pointed out by Marec in Ref. 12, it is not optimal to operate

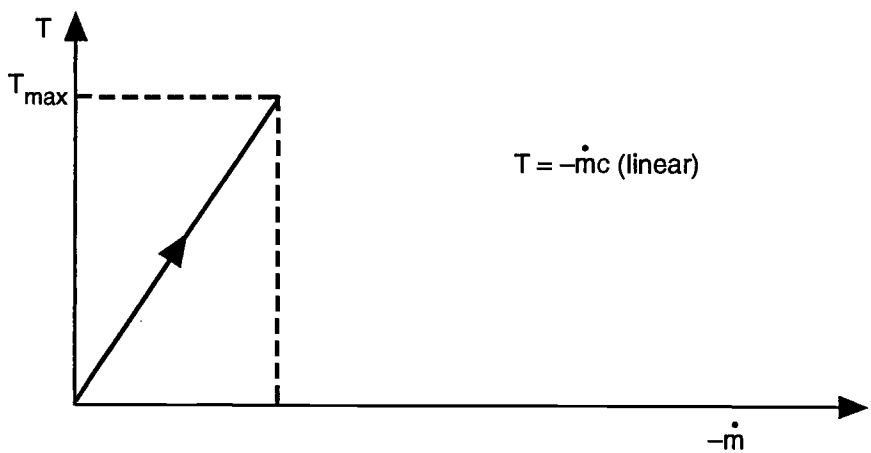


Fig. 14.18 Thrust vs mass flow rate for constant–exhaust velocity low-thrust rocket (from Ref. 12).

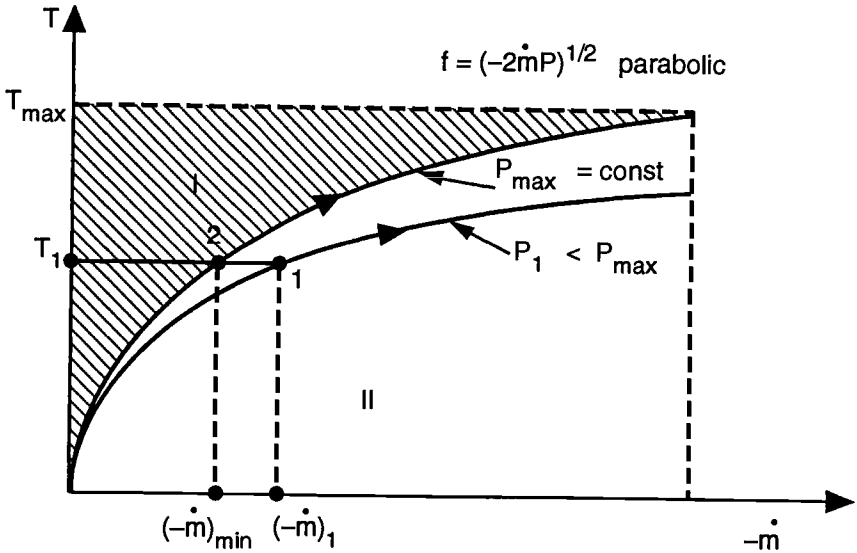


Fig. 14.19 Selection of optimal thrust for variable–exhaust velocity low-thrust rocket (from Ref. 12).

these types of thrusters at a power level $P_1 < P_{max}$ since, for the same thrust T_1, the operation at P_{max} depicted by point 2 results in the minimum mass flow rate or propellant expenditure. In Fig. 14.20, and for given P_{max}, any operating point A corresponds to a unique combination of beam voltage V_B and beam current I_B. If V_B is increased, then I_{sp} or c will also increase according to the I_{sp} equation and, since power is held at $P_{max} =$ constant, the thrust will decrease accordingly since $P = 1/2Tc$. Conversely, decreasing I_{sp} or V_B will result in increasing thrust and I_B. In Fig. 14.21, the unreachable Region I is extended further by the inclusion of the boundary 0O′, which corresponds to the equation $T = -\dot{m}c_{max}$, where c_{max} is the maximum exhaust velocity achieved by the rocket. This boundary is necessary to prevent the exhaust velocity or the I_{sp} to grow to very large values as the thrust is decreased toward its minimum value. This minimum is conveniently defined at point O such that the operating arc is the arc O3 on the P_{max} parabola. Following Edelbaum and letting $\boldsymbol{r}$ and $\boldsymbol{v}$ stand for the spacecraft position and velocity vectors, the second-order differential equation of motion is reduced to the following first-order form:

$$\dot{\boldsymbol{v}} = \frac{\boldsymbol{T}}{m} + \boldsymbol{g}(\boldsymbol{r}, t)$$

$$\dot{\boldsymbol{r}} = \boldsymbol{v}$$

The mass flow rate obeys the general form

$$\dot{m} = \dot{m}(\boldsymbol{r}, t, \boldsymbol{T})$$

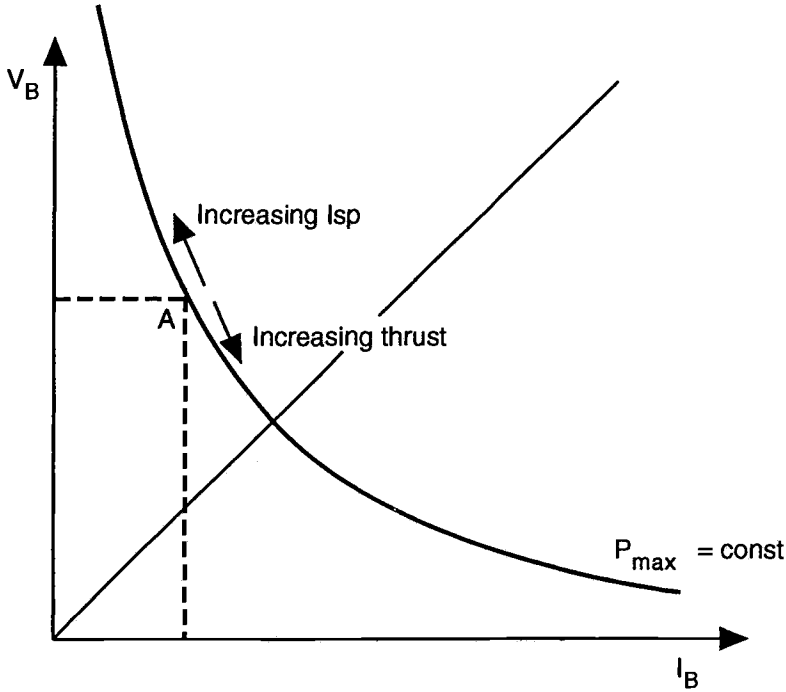


Fig. 14.20 Beam voltage and current variation for given power P_{max} (from Ref. 16).

The Hamiltonian of such a system is given by

$$H = \boldsymbol{\lambda}_v \cdot \frac{\boldsymbol{T}}{m}(\boldsymbol{r}, t) + \boldsymbol{\lambda}_v \cdot \boldsymbol{g}(\boldsymbol{r}, t) + \boldsymbol{\lambda}_r \cdot \boldsymbol{v} + \lambda_m \dot{m}(\boldsymbol{r}, t, \boldsymbol{T})$$

The Euler–Lagrange equations are, therefore,

$$\dot{\boldsymbol{\lambda}}_v = -\frac{\partial H}{\partial \boldsymbol{v}} = -\boldsymbol{\lambda}_r$$

$$\dot{\boldsymbol{\lambda}}_r = -\frac{\partial H}{\partial \boldsymbol{r}} = -\frac{\boldsymbol{\lambda}_v}{m} \cdot \frac{\partial \boldsymbol{T}}{\partial \boldsymbol{r}} - \boldsymbol{\lambda}_v \cdot \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{r}} - \lambda_m \frac{\partial \dot{m}}{\partial \boldsymbol{r}}$$

$$\dot{\lambda}_m = -\frac{\partial H}{\partial m} = \frac{\boldsymbol{\lambda}_v \cdot \boldsymbol{T}}{m^2}$$

The application of Pontryagin's maximum principle requires that the control vector, namely, the thrust vector $\boldsymbol{T}$, must be selected in such a manner as to maximize H at each instant of time. Since $\boldsymbol{T}$ appears in $\dot{m}$, this necessary condition translates into

$$\frac{\boldsymbol{\lambda}_v \cdot \boldsymbol{T}(\boldsymbol{r}, t)}{m} + \lambda_m \dot{m}(\boldsymbol{r}, t, \boldsymbol{T}) = \max$$

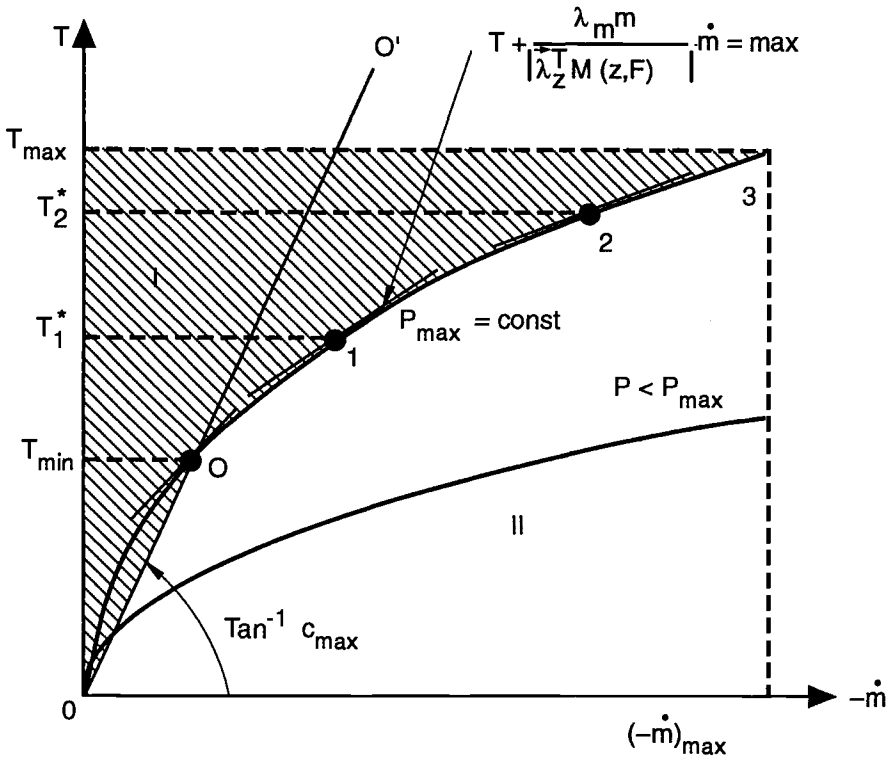


Fig. 14.21 Locus of optimal thrust magnitude for bounded thrust rocket (from Ref. 12).

This can be simplified further, if $\dot{m}$ is a function of thrust magnitude but not direction. This is the case of the unconstrained transfer with free yaw, pitch, and roll. In these cases, $\dot{m} = \dot{m}(\boldsymbol{r}, t)$ such that the preceding condition reduces to

$$\frac{\boldsymbol{\lambda}_v \cdot \boldsymbol{T}(\boldsymbol{r}, t)}{m} + \lambda_m \dot{m}(\boldsymbol{r}, t) = \max$$

This requires that the thrust vector $\boldsymbol{T}$ remain aligned with $\boldsymbol{\lambda}_v$ at all times since $\boldsymbol{\lambda}_v \cdot \boldsymbol{T}$ is then maximum. This results in

$$\frac{\lambda_v T}{m} + \lambda_m \dot{m} = \max \tag{14.190}$$

From Eq. (14.189), $\dot{m}$ can be replaced by $-T^2/(2P)$, which reduces the expression in Eq. (14.190) to

$$T - \frac{\lambda_m m}{2\lambda_v P} T^2 = \max$$

Maximizing this expression with respect to the thrust magnitude T results in

$$T^* = \frac{\lambda_v P}{\lambda_m m} \tag{14.191}$$

and, from Eq. (14.185), the optimal I_{sp}^* is then

$$I_{\text{sp}}^* = \frac{2P}{gT^*} = \frac{2m\lambda_m}{g\lambda_v}$$

Let us now carry out the optimization of T for the equinoctial formulation, using the equinoctial orbit elements represented by the vector $z = (a\ h\ k\ p\ q\ \lambda)^T$ instead of the $\boldsymbol{r}$ and $\boldsymbol{v}$ formulation used above. The equation for the mass is, as before, given by

$$\dot{m} = -\frac{T^2}{2P} = -\frac{2P}{c^2} = -\frac{T}{c}$$

The system equations can now be written as

$$\dot{z} = \frac{T}{m} M(z, F)\hat{u} \tag{14.192}$$

$$\dot{m} = -\frac{T^2}{2P(z, F, \hat{u})} \tag{14.193}$$

Here P is assumed to be a function of the vehicle orientation. Let us further assume that the thrust magnitude is not constrained such that the Hamiltonian of the system above is written as

$$\begin{aligned} H &= \boldsymbol{\lambda}_z^T \cdot \dot{z} + \lambda_m \dot{m} \\ &= \boldsymbol{\lambda}_z^T \cdot \frac{T}{m} M(z, F)\hat{u} - \lambda_m \frac{T^2}{2P(z, F, \hat{u})} + \lambda_\lambda n \end{aligned} \tag{14.194}$$

If P is not a function of $\hat{u}$, then,

$$H = \boldsymbol{\lambda}_z^T \cdot \frac{T}{m} M(z, F)\hat{u} - \lambda_m \frac{T^2}{2P(z, F)} + \lambda_\lambda n$$

H is maximized if $\hat{u}$ is chosen parallel to $[\boldsymbol{\lambda}_z^T (T/m) M(z, F)]^T$ or

$$\hat{u} = \frac{\left[\boldsymbol{\lambda}_z^T \frac{T}{m} M(z, F)\right]^T}{\left|\boldsymbol{\lambda}_z^T \frac{T}{m} M(z, F)\right|}$$

Then, H is reduced to

$$H = \left|\boldsymbol{\lambda}_z^T \frac{T}{m} M(z, F)\right| - \lambda_m \frac{T^2}{2P(z, F)} + \lambda_\lambda n \tag{14.195}$$

and the optimal thrust magnitude is obtained from the optimality condition $\partial H/\partial T = 0$, which results in

$$T^* = \frac{|\boldsymbol{\lambda}_z^T M(z, F)| P(z, F)}{\lambda_m m} \tag{14.196}$$

This is equivalent to the maximization of

$$T + \frac{m\lambda_m}{|\boldsymbol{\lambda}_z^T M(z, F)|}\dot{m} = \max \tag{14.197}$$

since $\dot{m}$ is equal to $-T^2/(2P)$ in Eq. (14.195). Equation (14.196) is identical to Eq. (14.191), where the $\boldsymbol{r}, \boldsymbol{v}$ formulation was used. Here $|\boldsymbol{\lambda}_z^T M(z, F)|$ replaces λ_v since the state variables are the equinoctial elements. Equation (14.197) is the equation of a straight line in the $(T, -\dot{m})$ system, and the maximum takes place when the line is tangent to the $P_{\max}$ curve (Fig. 14.21) such as at point 1. At a later time, the optimal operating point is at 2 with T_2^* representing the optimal time-varying thrust magnitude.[12,17] The corresponding mass flow rate at each instant of time is, of course, obtained from $\dot{m} = -T^2/(2P)$, and the optimal I_{sp}^* from

$$I_{\text{sp}}^* = \frac{2m\lambda_m}{g|\boldsymbol{\lambda}_z^T M(z, F)|}$$

The Euler–Lagrange differential equations are given by

$$\dot{\boldsymbol{\lambda}}_z = -\frac{\partial H}{\partial z} = -\boldsymbol{\lambda}_z^T \frac{\partial M}{\partial z}\frac{T}{m}\hat{u} - \lambda_\lambda \frac{\partial n}{\partial z} \tag{14.198}$$

$$\dot{\lambda}_m = -\frac{\partial H}{\partial m} = \boldsymbol{\lambda}_z^T M \frac{T}{m^2}\hat{u} \tag{14.199}$$

where it is assumed that the power P is independent of the state vector z, being considered constant throughout the transfer. For the optimization of the thrust orientation, the maximum principle is used to get, as before,

$$\hat{u} = \frac{(\boldsymbol{\lambda}_z^T M)^T}{|\boldsymbol{\lambda}_z^T M|} \tag{14.200}$$

and thereby the three components u_f, u_g, and u_w which, in turn, provide the values of u_r, u_θ, and u_h in the rotating $\hat{r}, \hat{\theta}, \hat{h}$ frame. If the thrust magnitude is unconstrained, its optimal value is obtained from the optimality condition

$$\frac{\partial H}{\partial T} = 0 = \boldsymbol{\lambda}_z^T \frac{M}{m}\hat{u} - \frac{T}{P}\lambda_m$$

which yields the optimal thrust T^*

$$T^* = \frac{P(\boldsymbol{\lambda}_z^T M\hat{u})}{(m\lambda_m)} \tag{14.201}$$

The optimal value of the specific impulse is then given by

$$I_{\text{sp}}^* = \frac{2P}{T^* g}$$

since $c = 2P/T = I_{\text{sp}} g$. Now the optimal acceleration program is obtained from

$$f^* = \frac{T^*}{m} = \frac{\boldsymbol{\lambda}_z^T M \hat{u}}{m^2 \lambda_m} P$$

If the thrust magnitude is bounded from above and below, then,

$$T_{\min} < T < T_{\max}$$

such that we have the following inequality constraints on the control variable T:

$$C_1 = T - T_{\max} \leq 0$$
$$C_2 = -T + T_{\min} \leq 0$$

These constraints can be adjoined to the original Hamiltonian by way of Lagrange multipliers μ_1 and μ_2 such that

$$H = \frac{T}{m} \boldsymbol{\lambda}_z^T M \hat{u} - \frac{T^2}{2P} \lambda_m + \lambda_\lambda n + \mu_1 (T - T_{\max}) + \mu_2 (-T + T_{\min})$$

The necessary condition on H is

$$H_T = \frac{\partial H}{\partial T} = \boldsymbol{\lambda}_z^T \frac{M}{m} \hat{u} - \frac{T}{P} \lambda_m + \mu_1 - \mu_2 = 0$$

The multipliers μ_1 and μ_2 are such that $\mu_1 > 0$ when $C_1 = 0$ or $T = T_{\max}$, and $\mu_1 = 0$ when $C_1 < 0$ or $T < T_{\max}$ and, similarly, $\mu_2 > 0$ when $C_2 = 0$ or $T = T_{\min}$, and $\mu_2 = 0$ when $C_2 < 0$ or $T > T_{\min}$. When $T_{\min} < T < T_{\max}$ assumes an intermediate value, $\mu_1 = \mu_2 = 0$ and $H_T = 0$ reduces to $\boldsymbol{\lambda}_z^T (M/m) \hat{u} - (T/P) \lambda_m = 0$, yielding the optimal control given by Eq. (14.201). The values of μ_1 and μ_2 are obtained from

$$\mu_1 = \frac{T_{\max}}{P} \lambda_m - \frac{\boldsymbol{\lambda}_z^T M \hat{u}}{m}$$
$$\mu_2 = -\frac{T_{\min}}{P} \lambda_m + \frac{\boldsymbol{\lambda}_z^T M \hat{u}}{m}$$

The Lagrange multipliers are still given by Eqs. (14.198) and (14.199), and the optimal T^* is selected by monitoring the value of $\boldsymbol{\lambda}_z^T M \hat{u} P / (m \lambda_m)$. If it is less than $T_{\min}$, then $T^* = T_{\min}$ and, if it is larger than $T_{\max}$, then $T^* = T_{\max}$ and, finally, if it is intermediate between $T_{\min}$ and $T_{\max}$, then T^* is given by Eq. (14.201), which

is the presently calculated value of $\boldsymbol{\lambda}_z^T M\hat{u}P/(m\lambda_m)$. We can also use the simpler Hamiltonian H^* without adjoining the constraints, namely,

$$H^* = \boldsymbol{\lambda}_z^T \dot{z} + \lambda_m \dot{m} = \frac{T}{m}\left(\boldsymbol{\lambda}_z^T M\hat{u} - \frac{m\lambda_m}{c}\right) + \lambda_\lambda n$$

This is equivalent to Eq. (14.202) since $c = 2P/T$ is a function of the control T,

$$H^* = \frac{T}{m}\boldsymbol{\lambda}_z^T M\hat{u} - \frac{T^2}{2P}\lambda_m + \lambda_\lambda n \tag{14.202}$$

The optimality condition yields with

$$\frac{\partial H^*}{\partial T} = H_T^* = \boldsymbol{\lambda}_z^T \frac{M}{m}\hat{u} - \frac{T}{P}\lambda_m \tag{14.203}$$

the following condition: $\delta H^* = H_T^* \delta T \leq 0$ since this is equivalent to $\delta J = \int_{t_0}^{t_f} H_T^* \delta T \mathrm{d}t = \int_{t_0}^{t_f} \delta H^* \mathrm{d}t \leq 0$ for the control T to be maximizing for all admissible values of δT. δJ is the variation in J, the performance index, due to variations in T for fixed $z(t_0)$. The optimal control is selected by monitoring the value of H_T^*. If H_T^*, as calculated by Eq. (14.203), is positive, then $T = T_{\max}$ and, if H_T^* is negative, then $T = T_{\min}$ and, finally, if $H_T^* = 0$ then T^* is given by Eq. (14.203):

$$H_T^* > 0 \Rightarrow \delta T < 0 \Rightarrow T = T_{\max} \tag{14.204}$$

$$H_T^* < 0 \Rightarrow \delta T > 0 \Rightarrow T = T_{\min} \tag{14.205}$$

$$H_T^* = 0 \Rightarrow T = T^* = \boldsymbol{\lambda}_z^T \frac{M\hat{u}}{m\lambda_m} P \tag{14.206}$$

In practice, the last condition for $H_T^* = 0$ is replaced by $\left|H_T^*\right| < \varepsilon$, where ε is a small number, say 10^{-10}. The Euler–Lagrange equations are still given by Eqs. (14.198) and (14.199). We now maximize the value of the mass at the fixed final time t_f such that the performance index $J = \phi = m_f$ and the optimal thrust direction given by Eq. (14.200) is obtained directly from the maximum principle. Since the equations of motion given by Eqs. (14.192) and (14.193) are not explicit functions of time, the Hamiltonian H in Eq. (14.194), or H^* in Eq. (14.202), is constant throughout the transfer. Given initial state parameters $(a)_0$, $(h)_0$, $(k)_0$, $(p)_0$, $(q)_0$, $(\lambda)_0$, and $(m)_0$, the initial values of the seven Lagrange multipliers are guessed, namely, $(\lambda_a)_0$, $(\lambda_h)_0$, $(\lambda_k)_0$, $(\lambda_p)_0$, $(\lambda_q)_0$, $(\lambda_\lambda)_0$, $(\lambda_m)_0$, and the state and adjoint equations given by Eqs. (14.192), (14.193) and (14.198), (14.199) are integrated forward from t_0 to t_f by using the optimal thrust direction $\hat{u}$ in Eq. (14.200) and the thrust magnitude from Eqs. (14.204–14.206). The initial values of the multipliers are iterated until the desired terminal state given by $(a)_f$, $(h)_f$, $(k)_f$, $(p)_f$, $(q)_f$, $(\lambda)_f$, and $(\lambda_m)_f = (\partial\phi/\partial m)_{t_f} = 1$ is satisfied. This is achieved by minimizing the following objective function:

$$\begin{aligned} F' = {} & w_1[a - (a)_f]^2 + w_2[h - (h)_f]^2 + w_3[k - (k)_f]^2 \\ & + w_4[p - (p)_f]^2 + w_5[q - (q)_f]^2 + w_6[\lambda - (\lambda)_f]^2 + w_7[\lambda_m - 1]^2 \end{aligned}$$

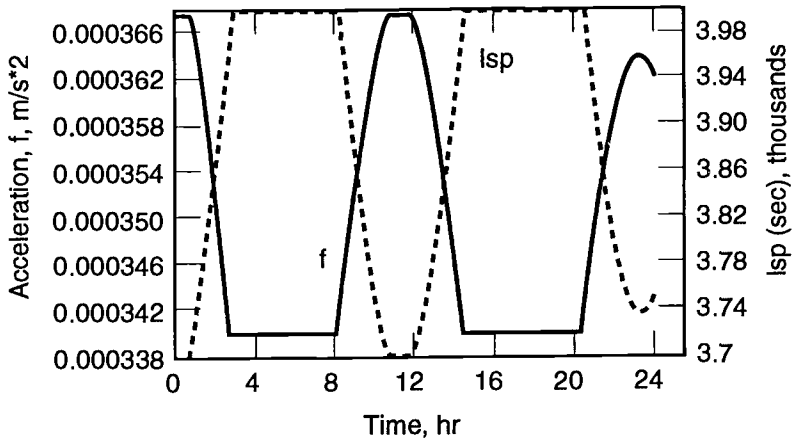


Fig. 14.22 Optimal acceleration and I_{sp} profiles for a short-duration orbital rendezvous with $(I_{sp})_{max}$ = 4000 s and $(I_{sp})_{min}$ = 3700 s (from Ref. 16).

where the w_j are certain weights and where the values of a, h, k, p, q, λ, and λ_m are evaluated at the fixed terminal time t_f. Since the power is assumed to remain constant, let us select $(I_{sp})_{min}$ and $(I_{sp})_{max}$ such that $T_{min} = 2P/[(I_{sp})_{max}\, g]$, $T_{max} = 2P/[(I_{sp})_{min}\, g]$. The values of T_{max} and T_{min} are used as input to the optimizer, and they correspond to initial accelerations of $(f)_{max} = T_{max}/m_0$ and $(f)_{min} = T_{min}/m_0$. Furthermore, the acceleration, being now variable, will be optimized to achieve minimum fuel consumption. The various λ have units that are the inverse of the corresponding element rates, namely, seconds per meter for λ_a, seconds per radian for λ_λ, and seconds per kilogram for λ_m, with the remaining λ units being the second since h, k, p, and q are unitless. This is needed because the Hamiltonian is unitless. Figure 14.22 shows a rendezvous example borrowed from Ref. 16, where both minimum-I_{sp} and maximum-I_{sp} thrust arcs are used. Substantial savings in fuel can be achieved, especially with higher values of the specific power or power-to-mass ratio, for highly elliptic orbit transfers and for larger changes in the energy and angular momentum vector. Furthermore, since the thrust is continuously on, or the control is otherwise continuous, it is much easier to generate a series of transfers with increasing I_{sp} bandwidths such that, in the limit, the discontinuous thrust case can be produced with full thrust cutoff arcs that correspond to infinite I_{sp} or, in practice, negligible thrust. These coast arcs are beneficial for certain transfer geometries, especially if the fixed transfer time is further relaxed from its minimum-time value obtained with continuous thrust. Finally, when a given engine design is used with specified I_{sp} bounds, the thrust magnitude unconstrained solution may not be capable of flying because of the engine limitations which will, in turn, necessitate the generation of the optimal I_{sp}-constrained solution.

Appendix: The Partials of the *M* Matrix

The following complete set of the $\partial M/\partial z$ partials are needed to numerically integrate the differential equations for the adjoints such as those given in Eq. (14.155).

The Partial Derivatives of M with Respect to h

$$\frac{\partial M_{11}}{\partial h} = \frac{2}{n^2 a}\frac{\partial \dot{X}_1}{\partial h} \tag{A.1}$$

$$\frac{\partial M_{12}}{\partial h} = \frac{2}{n^2 a}\frac{\partial \dot{Y}_1}{\partial h} \tag{A.2}$$

$$\frac{\partial M_{13}}{\partial h} = 0 \tag{A.3}$$

$$\begin{aligned}\frac{\partial M_{21}}{\partial h} &= \frac{-h}{na^2(1-h^2-k^2)^{1/2}}\left(\frac{\partial X_1}{\partial k} - \frac{h\beta}{n}\dot{X}_1\right)\\ &+ \frac{(1-h^2-k^2)^{1/2}}{na^2}\left[\frac{\partial^2 X_1}{\partial h \partial k} - \frac{\dot{X}_1}{n}\left(\beta + \frac{h^2\beta^3}{1-\beta}\right) - \frac{h\beta}{n}\frac{\partial \dot{X}_1}{\partial h}\right]\end{aligned} \tag{A.4}$$

$$\begin{aligned}\frac{\partial M_{22}}{\partial h} &= \frac{-h}{na^2(1-h^2-k^2)^{1/2}}\left(\frac{\partial Y_1}{\partial k} - \frac{h\beta}{n}\dot{Y}_1\right)\\ &+ \frac{(1-h^2-k^2)^{1/2}}{na^2}\left[\frac{\partial^2 Y_1}{\partial h \partial k} - \frac{\dot{Y}_1}{n}\left(\beta + \frac{h^2\beta^3}{1-\beta}\right) - \frac{h\beta}{n}\frac{\partial \dot{Y}_1}{\partial h}\right]\end{aligned} \tag{A.5}$$

$$\frac{\partial M_{23}}{\partial h} = \frac{hk(1-h^2-k^2)^{-3/2}}{na^2}(qY_1 - pX_1) + \frac{k\left(q\frac{\partial Y_1}{\partial h} - p\frac{\partial X_1}{\partial h}\right)}{na^2(1-h^2-k^2)^{1/2}} \tag{A.6}$$

$$\begin{aligned}\frac{\partial M_{31}}{\partial h} &= \frac{h}{na^2(1-h^2-k^2)^{1/2}}\left(\frac{\partial X_1}{\partial h} + k\beta\frac{\dot{X}_1}{n}\right)\\ &- \frac{(1-h^2-k^2)^{1/2}}{na^2}\left(\frac{\partial^2 X_1}{\partial h^2} + \frac{hk\beta^3}{1-\beta}\frac{\dot{X}_1}{n} + \frac{k\beta}{n}\frac{\partial \dot{X}_1}{\partial h}\right)\end{aligned} \tag{A.7}$$

$$\begin{aligned}\frac{\partial M_{32}}{\partial h} &= \frac{h}{na^2(1-h^2-k^2)^{1/2}}\left(\frac{\partial Y_1}{\partial h} + k\beta\frac{\dot{Y}_1}{n}\right)\\ &- \frac{(1-h^2-k^2)^{1/2}}{na^2}\left(\frac{\partial^2 Y_1}{\partial h^2} + \frac{hk\beta^3}{1-\beta}\frac{\dot{Y}_1}{n} + \frac{k\beta}{n}\frac{\partial \dot{Y}_1}{\partial h}\right)\end{aligned} \tag{A.8}$$

$$\begin{aligned}\frac{\partial M_{33}}{\partial h} &= \frac{-1}{na^2(1-h^2-k^2)^{1/2}}\left[(qY_1 - pX_1) + h\left(q\frac{\partial Y_1}{\partial h} - p\frac{\partial X_1}{\partial h}\right)\right]\\ &- \frac{h^2(qY_1 - pX_1)}{na^2(1-h^2-k^2)^{3/2}}\end{aligned} \tag{A.9}$$

$$\frac{\partial M_{43}}{\partial h} = \frac{(1+p^2+q^2)}{2na^2(1-h^2-k^2)^{1/2}}\left[\frac{\partial Y_1}{\partial h} + \frac{hY_1}{(1-h^2-k^2)}\right] \tag{A.10}$$

$$\frac{\partial M_{53}}{\partial h} = \frac{(1+p^2+q^2)}{2na^2(1-h^2-k^2)^{1/2}}\left[\frac{\partial X_1}{\partial h} + \frac{hX_1}{(1-h^2-k^2)}\right] \tag{A.11}$$

The partials

$$\frac{\partial M_{41}}{\partial h} = \frac{\partial M_{42}}{\partial h} = \frac{\partial M_{51}}{\partial h} = \frac{\partial M_{52}}{\partial h} = 0$$

are all identically zero.

$$\frac{\partial M_{61}}{\partial h} = \frac{1}{na^2}\left\{-2\frac{\partial X_1}{\partial h} - h\beta(1-h^2-k^2)^{-1/2}\left(h\frac{\partial X_1}{\partial h} + k\frac{\partial X_1}{\partial k}\right)\right.$$
$$+(1-h^2-k^2)^{1/2}\left[\left(\beta + \frac{h^2\beta^3}{1-\beta}\right)\frac{\partial X_1}{\partial h} + \frac{hk\beta^3}{1-\beta}\frac{\partial X_1}{\partial k}\right.$$
$$\left.\left. + \beta\left(h\frac{\partial^2 X_1}{\partial h^2} + k\frac{\partial^2 X_1}{\partial h \partial k}\right)\right]\right\} \tag{A.12}$$

$$\frac{\partial M_{62}}{\partial h} = \frac{1}{na^2}\left\{-2\frac{\partial Y_1}{\partial h} - h\beta(1-h^2-k^2)^{-1/2}\left(h\frac{\partial Y_1}{\partial h} + k\frac{\partial Y_1}{\partial k}\right)\right.$$
$$+(1-h^2-k^2)^{1/2}\left[\left(\beta + \frac{h^2\beta^3}{1-\beta}\right)\frac{\partial Y_1}{\partial h} + \frac{hk\beta^3}{1-\beta}\frac{\partial Y_1}{\partial k}\right.$$
$$\left.\left. + \beta\left(h\frac{\partial^2 Y_1}{\partial h^2} + k\frac{\partial^2 Y_1}{\partial h \partial k}\right)\right]\right\} \tag{A.13}$$

$$\frac{\partial M_{63}}{\partial h} = \frac{(1-h^2-k^2)^{-1/2}}{na^2}$$
$$\times\left[\left(q\frac{\partial Y_1}{\partial h} - p\frac{\partial X_1}{\partial h}\right) + h(1-h^2-k^2)^{-1}(qY_1 - pX_1)\right] \tag{A.14}$$

The Partial Derivatives of M with Respect to k

$$\frac{\partial M_{11}}{\partial k} = \frac{2}{n^2 a}\frac{\partial \dot{X}_1}{\partial k} \tag{A.15}$$

$$\frac{\partial M_{12}}{\partial k} = \frac{2}{n^2 a}\frac{\partial \dot{Y}_1}{\partial k} \tag{A.16}$$

$$\frac{\partial M_{13}}{\partial k} = 0 \tag{A.17}$$

$$\frac{\partial M_{21}}{\partial k} = \frac{-k}{na^2(1-h^2-k^2)^{1/2}}\left(\frac{\partial X_1}{\partial k} - \frac{h\beta}{n}\dot{X}_1\right)$$
$$+\frac{(1-h^2-k^2)^{1/2}}{na^2}\left[\frac{\partial^2 X_1}{\partial k^2} - \frac{hk\beta^3}{n(1-\beta)}\dot{X}_1 - \frac{h\beta}{n}\frac{\partial \dot{X}_1}{\partial k}\right] \tag{A.18}$$

$$\frac{\partial M_{22}}{\partial k} = \frac{-k}{na^2(1-h^2-k^2)^{1/2}}\left(\frac{\partial Y_1}{\partial k} - \frac{h\beta}{n}\dot{Y}_1\right)$$

$$+ \frac{(1-h^2-k^2)^{1/2}}{na^2}\left[\frac{\partial^2 Y_1}{\partial k^2} - \frac{hk\beta^3}{n(1-\beta)}\dot{Y}_1 - \frac{h\beta}{n}\frac{\partial \dot{Y}_1}{\partial k}\right] \tag{A.19}$$

$$\frac{\partial M_{23}}{\partial k} = \frac{(qY_1 - pX_1)}{na^2(1-h^2-k^2)^{1/2}} + \frac{1}{na^2(1-h^2-k^2)^{1/2}}$$

$$\times\left[k\left(q\frac{\partial Y_1}{\partial k} - p\frac{\partial X_1}{\partial k}\right) + \frac{k^2(qY_1 - pX_1)}{(1-h^2-k^2)}\right] \tag{A.20}$$

$$\frac{\partial M_{31}}{\partial k} = \frac{k}{na^2(1-h^2-k^2)^{1/2}}\left(\frac{\partial X_1}{\partial h} + k\beta\frac{\dot{X}_1}{n}\right) - \frac{(1-h^2-k^2)^{1/2}}{na^2}$$

$$\times\left[\frac{\partial^2 X_1}{\partial k \partial h} + \left(\beta + \frac{k^2\beta^3}{1-\beta}\right)\frac{\dot{X}_1}{n} + \frac{k\beta}{n}\frac{\partial \dot{X}_1}{\partial k}\right] \tag{A.21}$$

$$\frac{\partial M_{32}}{\partial k} = \frac{k}{na^2(1-h^2-k^2)^{1/2}}\left(\frac{\partial Y_1}{\partial h} + k\beta\frac{\dot{Y}_1}{n}\right) - \frac{(1-h^2-k^2)^{1/2}}{na^2}$$

$$\times\left[\frac{\partial^2 Y_1}{\partial k \partial h} + \left(\beta + \frac{k^2\beta^3}{1-\beta}\right)\frac{\dot{Y}_1}{n} + \frac{k\beta}{n}\frac{\partial \dot{Y}_1}{\partial k}\right] \tag{A.22}$$

$$\frac{\partial M_{33}}{\partial k} = \frac{-h}{na^2(1-h^2-k^2)^{1/2}}\left(q\frac{\partial Y_1}{\partial k} - p\frac{\partial X_1}{\partial k}\right)$$

$$- \frac{hk}{na^2(1-h^2-k^2)^{3/2}}(qY_1 - pX_1) \tag{A.23}$$

$$\frac{\partial M_{43}}{\partial k} = \frac{(1+p^2+q^2)}{2na^2(1-h^2-k^2)^{1/2}}\frac{\partial Y_1}{\partial k} + \frac{k(1+p^2+q^2)}{2na^2(1-h^2-k^2)^{3/2}}Y_1 \tag{A.24}$$

$$\frac{\partial M_{53}}{\partial k} = \frac{(1+p^2+q^2)}{2na^2(1-h^2-k^2)^{1/2}}\frac{\partial X_1}{\partial k} + \frac{k(1+p^2+q^2)}{2na^2(1-h^2-k^2)^{3/2}}X_1 \tag{A.25}$$

$$\frac{\partial M_{41}}{\partial k} = \frac{\partial M_{42}}{\partial k} = \frac{\partial M_{51}}{\partial k} = \frac{\partial M_{52}}{\partial k} = 0$$

$$\frac{\partial M_{61}}{\partial k} = \frac{1}{na^2}\left\{-2\frac{\partial X_1}{\partial k} - k\beta(1-h^2-k^2)^{-1/2}\left(h\frac{\partial X_1}{\partial h} + k\frac{\partial X_1}{\partial k}\right)\right.$$

$$+ (1-h^2-k^2)^{1/2}\left[\left(\beta + \frac{k^2\beta^3}{1-\beta}\right)\frac{\partial X_1}{\partial k} + \frac{hk\beta^3}{1-\beta}\frac{\partial X_1}{\partial h}\right.$$

$$\left.\left. + \beta\left(h\frac{\partial^2 X_1}{\partial k \partial h} + k\frac{\partial^2 X_1}{\partial k^2}\right)\right]\right\} \tag{A.26}$$

$$\frac{\partial M_{62}}{\partial k} = \frac{1}{na^2}\left\{ -2\frac{\partial Y_1}{\partial k} - k\beta(1-h^2-k^2)^{-1/2}\left(h\frac{\partial Y_1}{\partial h} + k\frac{\partial Y_1}{\partial k}\right) \right.$$
$$\left. +(1-h^2-k^2)^{1/2}\left[\left(\beta + \frac{k^2\beta^3}{1-\beta}\right)\frac{\partial Y_1}{\partial k} + \frac{hk\beta^3}{1-\beta}\frac{\partial Y_1}{\partial h} + \beta\left(h\frac{\partial^2 Y_1}{\partial k \partial h} + k\frac{\partial^2 Y_1}{\partial k^2}\right)\right]\right\} \tag{A.27}$$

$$\frac{\partial M_{63}}{\partial k} = \frac{(1-h^2-k^2)^{-1/2}}{na^2}\left[\left(q\frac{\partial Y_1}{\partial k} - p\frac{\partial X_1}{\partial k}\right) + k(1-h^2-k^2)^{-1}(qY_1 - pX_1)\right] \tag{A.28}$$

The Partial Derivatives of M with Respect to p

The nonzero partials are

$$\frac{\partial M_{23}}{\partial p} = \frac{-kX_1}{na^2(1-h^2-k^2)^{1/2}} \tag{A.29}$$

$$\frac{\partial M_{33}}{\partial p} = \frac{hX_1}{na^2(1-h^2-k^2)^{1/2}} \tag{A.30}$$

$$\frac{\partial M_{43}}{\partial p} = \frac{pY_1}{na^2(1-h^2-k^2)^{1/2}} \tag{A.31}$$

$$\frac{\partial M_{53}}{\partial p} = \frac{pX_1}{na^2(1-h^2-k^2)^{1/2}} \tag{A.32}$$

$$\frac{\partial M_{63}}{\partial p} = \frac{-X_1}{na^2(1-h^2-k^2)^{1/2}} \tag{A.33}$$

The Partial Derivatives of M with Respect to q

The nonzero partials are

$$\frac{\partial M_{23}}{\partial q} = \frac{kY_1}{na^2(1-h^2-k^2)^{1/2}} \tag{A.34}$$

$$\frac{\partial M_{33}}{\partial q} = \frac{-hY_1}{na^2(1-h^2-k^2)^{1/2}} \tag{A.35}$$

$$\frac{\partial M_{43}}{\partial q} = \frac{qY_1}{na^2(1-h^2-k^2)^{1/2}} \tag{A.36}$$

$$\frac{\partial M_{53}}{\partial q} = \frac{qX_1}{na^2(1-h^2-k^2)^{1/2}} \tag{A.37}$$

$$\frac{\partial M_{63}}{\partial q} = \frac{Y_1}{na^2(1-h^2-k^2)^{1/2}} \tag{A.38}$$

The partial derivatives of $\dot{X}_1$ with respect to h and k are

$$\frac{\partial \dot{X}_1}{\partial h} = \frac{a}{r}\dot{X}_1\left[s_F + \frac{a}{r}c_F(ks_F - hc_F)\right] + \frac{na^2}{r}\left\{h\beta s_F + (kc_F + hs_F)\left(\beta + \frac{h^2\beta^3}{1-\beta}\right) + \frac{a}{r}c_F\left[hk\beta s_F + (1 - h^2\beta)c_F\right]\right\} \tag{A.39}$$

$$\frac{\partial \dot{X}_1}{\partial k} = -\frac{\dot{X}_1}{r}a\left[-c_F + \frac{a}{r}s_F(ks_F - hc_F)\right] + \frac{na^2}{r}\left\{\frac{hk\beta^3}{1-\beta}(kc_F + hs_F) + h\beta c_F - \frac{a}{r}s_F\left[hk\beta s_F + (1 - h^2\beta)c_F\right]\right\} \tag{A.40}$$

The partials of $\dot{Y}_1$ with respect to h and k are

$$\frac{\partial \dot{Y}_1}{\partial h} = -\frac{\dot{Y}_1}{r}a\left[-s_F - \frac{a}{r}c_F(ks_F - hc_F)\right] + \frac{na^2}{r}\left\{-\frac{hk\beta^3}{1-\beta}(kc_F + hs_F) - k\beta s_F + \left[hk\beta c_F + (1 - k^2\beta)s_F\right]\frac{a}{r}c_F\right\} \tag{A.41}$$

$$\frac{\partial \dot{Y}_1}{\partial k} = \frac{-\dot{Y}_1}{r}a\left[-c_F + \frac{a}{r}s_F(ks_F - hc_F)\right] + \frac{na^2}{r} \times \left\{-\left(\beta + \frac{k^2\beta^3}{1-\beta}\right)(kc_F + hs_F) - k\beta c_F - \frac{a}{r}s_F\left[hk\beta c_F + (1 - k^2\beta)s_F\right]\right\} \tag{A.42}$$

The second partials of X_1 and Y_1 with respect to h and k are

$$\frac{\partial^2 X_1}{\partial h^2} = a\left\{-\frac{2a}{r}c_F\left(\beta + \frac{h^2\beta^3}{1-\beta}\right) - \frac{h\beta^3}{1-\beta}(hc_F - ks_F)\left[3 + \frac{h^2\beta^2(3-2\beta)}{(1-\beta)^2}\right] + \frac{a^2}{r^2}c_F(h\beta - s_F)\left[-s_F + \frac{a}{r}(h - s_F)\right] - \frac{a^2}{r^2}c_F^3\right\} \tag{A.43}$$

$$\frac{\partial^2 X_1}{\partial k^2} = -a\left\{-\frac{2a}{r}s_F\frac{hk\beta^3}{(1-\beta)} + (hc_F - ks_F)\left[1 + \frac{k^2\beta^2(3-2\beta)}{(1-\beta)^2}\right]\frac{h\beta^3}{(1-\beta)} + \frac{a^2}{r^2}s_F(h\beta - s_F)\left[-c_F + \frac{a}{r}(k - c_F)\right] + \frac{a^2}{r^2}c_F s_F^2\right\} \tag{A.44}$$

$$\frac{\partial^2 X_1}{\partial h \partial k} = -a\left\{\frac{a}{r}c_F\frac{hk\beta^3}{(1-\beta)} + (hc_F - ks_F)\left[1 + \frac{h^2\beta^2(3-2\beta)}{(1-\beta)^2}\right]\frac{k\beta^3}{(1-\beta)}\right.$$
$$\left. + (s_F - h\beta)\left[\frac{a}{r}(s_F^2 - hs_F) - c_F^2\right]\frac{a^2}{r^2} - \frac{a^2}{r^2}s_F c_F^2 - \frac{a}{r}s_F\left(\beta + \frac{h^2\beta^3}{1-\beta}\right)\right\} \tag{A.45}$$

$$\frac{\partial^2 X_1}{\partial k \partial h} = a\left\{\frac{a}{r}s_F\left(\beta + \frac{h^2\beta^3}{(1-\beta)}\right) - (hc_F - ks_F)\left[1 + \frac{h^2\beta^2(3-2\beta)}{(1-\beta)^2}\right]\frac{k\beta^3}{(1-\beta)}\right.$$
$$\left. + \frac{a^2}{r^2}\left[\frac{a}{r}\left(kc_F - c_F^2\right) + s_F^2\right](h\beta - s_F) - \frac{a}{r}c_F\frac{hk\beta^3}{(1-\beta)} + \frac{a^2}{r^2}c_F^2 s_F\right\} \tag{A.46}$$

$$\frac{\partial^2 Y_1}{\partial h^2} = a\left\{\frac{2a}{r}c_F\frac{hk\beta^3}{(1-\beta)} + (hc_F - ks_F)\frac{k\beta^3}{(1-\beta)}\left[1 + \frac{h^2\beta^2(3-2\beta)}{(1-\beta)^2}\right]\right.$$
$$\left. + \frac{a^2}{r^2}c_F\left[-\frac{a}{r}(h - s_F) + s_F\right](k\beta - c_F) - \frac{a^2}{r^2}s_F c_F^2\right\} \tag{A.47}$$

$$\frac{\partial^2 Y_1}{\partial k^2} = a\left\{\frac{-2a}{r}s_F\left(\beta + \frac{k^2\beta^3}{1-\beta}\right) + (hc_F - ks_F)\left[3 + \frac{k^2\beta^2(3-2\beta)}{(1-\beta)^2}\right]\right.$$
$$\left. \times\frac{k\beta^3}{(1-\beta)} + \frac{a^2}{r^2}s_F\left[-\frac{a}{r}(k - c_F) + c_F\right](c_F - k\beta) - \frac{a^2}{r^2}s_F^3\right\} \tag{A.48}$$

$$\frac{\partial^2 Y_1}{\partial h \partial k} = a\left\{\frac{a}{r}c_F\left(\beta + \frac{k^2\beta^3}{1-\beta}\right) + (hc_F - ks_F)\right.$$
$$\times\frac{h\beta^3}{(1-\beta)}\left[1 + \frac{k^2\beta^2(3-2\beta)}{(1-\beta)^2}\right] - \frac{a^2}{r^2}\left[\frac{a}{r}s_F(h - s_F) + c_F^2\right]$$
$$\left. \times(c_F - k\beta) + \frac{a^2}{r^2}c_F s_F^2 - \frac{a}{r}s_F\frac{hk\beta^3}{(1-\beta)}\right\} \tag{A.49}$$

$$\frac{\partial^2 Y_1}{\partial k \partial h} = a\left\{-\frac{a}{r}s_F\frac{hk\beta^3}{(1-\beta)} + (hc_F - ks_F)\frac{h\beta^3}{(1-\beta)}\right.$$
$$\times\left[1 + \frac{k^2\beta^2(3-2\beta)}{(1-\beta)^2}\right] - \frac{a^2}{r^2}\left[\frac{a}{r}c_F(k - c_F) + s_F^2\right]$$
$$\left. \times(k\beta - c_F) + \frac{a}{r}c_F\left(\beta + \frac{k^2\beta^3}{1-\beta}\right) + \frac{a^2}{r^2}c_F s_F^2\right\} \tag{A.50}$$

The accessory partials $\partial^2 X_1/\partial a \partial k$, $\partial^2 X_1/\partial a \partial h$, $\partial^2 Y_1/\partial a \partial k$, and $\partial^2 Y_1/\partial a \partial h$ are generated with $\partial F/\partial a = 0$. In all the following partials, $\partial X_1/\partial a = X_1/a$ and $\partial Y_1/\partial a = Y_1/a$:

$$\frac{\partial^2 X_1}{\partial a \partial k} = \frac{1}{a}\frac{\partial X_1}{\partial k} \tag{A.51}$$

$$\frac{\partial^2 X_1}{\partial a \partial h} = \frac{1}{a}\frac{\partial X_1}{\partial h} \tag{A.52}$$

$$\frac{\partial^2 Y_1}{\partial a \partial k} = \frac{1}{a}\frac{\partial Y_1}{\partial k} \tag{A.53}$$

$$\frac{\partial^2 Y_1}{\partial a \partial h} = \frac{1}{a}\frac{\partial Y_1}{\partial h} \tag{A.54}$$

The Partials Derivatives of M with Respect to a

$$\frac{\partial M_{11}}{\partial a} = \frac{4}{n^2 a^2}\dot{X}_1 + \frac{2}{n^2 a}\frac{\partial \dot{X}_1}{\partial a} \tag{A.55}$$

$$\frac{\partial M_{12}}{\partial a} = \frac{4}{n^2 a^2}\dot{Y}_1 + \frac{2}{n^2 a}\frac{\partial \dot{Y}_1}{\partial a} \tag{A.56}$$

$$\frac{\partial M_{13}}{\partial a} = 0 \tag{A.57}$$

$$\frac{\partial M_{21}}{\partial a} = \frac{(1-h^2-k^2)^{1/2}}{na^2}\left[-\frac{1}{2a}\frac{\partial X_1}{\partial k} + \frac{\partial^2 X_1}{\partial a \partial k} - \frac{h\beta}{na}\dot{X}_1 - \frac{h\beta}{n}\frac{\partial \dot{X}_1}{\partial a}\right] \tag{A.58}$$

$$\frac{\partial M_{22}}{\partial a} = \frac{(1-h^2-k^2)^{1/2}}{na^2}\left[-\frac{1}{2a}\frac{\partial Y_1}{\partial k} + \frac{\partial^2 Y_1}{\partial a \partial k} - \frac{h\beta}{na}\dot{Y}_1 - \frac{h\beta}{n}\frac{\partial \dot{Y}_1}{\partial a}\right] \tag{A.59}$$

$$\frac{\partial M_{23}}{\partial a} = \frac{k}{na^2(1-h^2-k^2)^{1/2}}\left[-\frac{1}{2a}(qY_1 - pX_1) + q\frac{\partial Y_1}{\partial a} - p\frac{\partial X_1}{\partial a}\right] \tag{A.60}$$

$$\frac{\partial M_{31}}{\partial a} = -\frac{(1-h^2-k^2)^{1/2}}{na^2}\left[-\frac{1}{2a}\frac{\partial X_1}{\partial h} + \frac{\partial^2 X_1}{\partial a \partial h} + \frac{k\beta}{na}\dot{X}_1 + \frac{k\beta}{n}\frac{\partial \dot{X}_1}{\partial a}\right] \tag{A.61}$$

$$\frac{\partial M_{32}}{\partial a} = -\frac{(1-h^2-k^2)^{1/2}}{na^2}\left[-\frac{1}{2a}\frac{\partial Y_1}{\partial h} + \frac{\partial^2 Y_1}{\partial a \partial h} + \frac{k\beta}{na}\dot{Y}_1 + \frac{k\beta}{n}\frac{\partial \dot{Y}_1}{\partial a}\right] \tag{A.62}$$

$$\frac{\partial M_{33}}{\partial a} = \frac{-h}{na^2(1-h^2-k^2)^{1/2}}\left[-\frac{1}{2a}(qY_1 - pX_1) + q\frac{\partial Y_1}{\partial a} - p\frac{\partial X_1}{\partial a}\right] \tag{A.63}$$

$$\frac{\partial M_{41}}{\partial a} = 0 \tag{A.64}$$

$$\frac{\partial M_{42}}{\partial a} = 0 \tag{A.65}$$

$$\frac{\partial M_{43}}{\partial a} = \frac{(1+p^2+q^2)}{2na^2(1-h^2-k^2)^{1/2}}\left(-\frac{1}{2a}Y_1 + \frac{\partial Y_1}{\partial a}\right) \tag{A.66}$$

$$\frac{\partial M_{51}}{\partial a} = 0 \tag{A.67}$$

$$\frac{\partial M_{52}}{\partial a} = 0 \tag{A.68}$$

$$\frac{\partial M_{53}}{\partial a} = \frac{(1+p^2+q^2)}{2na^2(1-h^2-k^2)^{1/2}}\left(-\frac{1}{2a}X_1 + \frac{\partial X_1}{\partial a}\right) \tag{A.69}$$

$$\frac{\partial M_{61}}{\partial a} = \frac{-M_{61}}{2a} + \frac{1}{na^2} \times \left[-2\frac{\partial X_1}{\partial a} + (1-h^2-k^2)^{1/2}\left(h\beta\frac{\partial^2 X_1}{\partial a \partial h} + k\beta\frac{\partial^2 X_1}{\partial a \partial k}\right)\right] \tag{A.70}$$

$$\frac{\partial M_{62}}{\partial a} = -\frac{M_{62}}{2a} + \frac{1}{na^2} \times \left[-2\frac{\partial Y_1}{\partial a} + (1-h^2-k^2)^{1/2}\left(h\beta\frac{\partial^2 Y_1}{\partial a \partial h} + k\beta\frac{\partial^2 Y_1}{\partial a \partial k}\right)\right] \tag{A.71}$$

$$\frac{\partial M_{63}}{\partial a} = -\frac{M_{63}}{2a} + \frac{1}{na^2}\left[\left(q\frac{\partial Y_1}{\partial a} - p\frac{\partial X_1}{\partial a}\right)(1-h^2-k^2)^{-1/2}\right] \tag{A.72}$$

with

$$\frac{\partial \dot{X}_1}{\partial a} = -\frac{1}{2}\frac{na}{r}\left[hk\beta c_F - (1-h^2\beta)s_F\right] \tag{A.73}$$

$$\frac{\partial \dot{Y}_1}{\partial a} = \frac{1}{2}\frac{na}{r}\left[hk\beta s_F - (1-k^2\beta)c_F\right] \tag{A.74}$$

The Partial Derivatives of M with Respect to λ

$$\frac{\partial M_{11}}{\partial \lambda} = \frac{2}{n^2 r}\frac{\partial \dot{X}_1}{\partial F} \tag{A.75}$$

$$\frac{\partial M_{12}}{\partial \lambda} = \frac{2}{n^2 r}\frac{\partial \dot{Y}_1}{\partial F} \tag{A.76}$$

$$\frac{\partial M_{13}}{\partial \lambda} = 0 \tag{A.77}$$

$$\frac{\partial M_{21}}{\partial \lambda} = \frac{(1-h^2-k^2)^{1/2}}{nar}\left(\frac{\partial^2 X_1}{\partial F \partial k} - \frac{h\beta}{n}\frac{\partial \dot{X}_1}{\partial F}\right) \tag{A.78}$$

$$\frac{\partial M_{22}}{\partial \lambda} = \frac{(1-h^2-k^2)^{1/2}}{nar}\left(\frac{\partial^2 Y_1}{\partial F \partial k} - \frac{h\beta}{n}\frac{\partial \dot{Y}_1}{\partial F}\right) \tag{A.79}$$

$$\frac{\partial M_{23}}{\partial \lambda} = \frac{k\left(q\frac{\partial Y_1}{\partial F} - p\frac{\partial X_1}{\partial F}\right)}{nar(1-h^2-k^2)^{1/2}} \tag{A.80}$$

$$\frac{\partial M_{31}}{\partial \lambda} = -\frac{(1-h^2-k^2)^{1/2}}{nar}\left(\frac{\partial^2 X_1}{\partial F \partial h} + \frac{k\beta}{n}\frac{\partial \dot{X}_1}{\partial F}\right) \tag{A.81}$$

$$\frac{\partial M_{32}}{\partial \lambda} = -\frac{(1-h^2-k^2)^{1/2}}{nar}\left(\frac{\partial^2 Y_1}{\partial F \partial h} + \frac{k\beta}{n}\frac{\partial \dot{Y}_1}{\partial F}\right) \tag{A.82}$$

$$\frac{\partial M_{33}}{\partial \lambda} = \frac{-h\left(q\frac{\partial Y_1}{\partial F} - p\frac{\partial X_1}{\partial F}\right)}{nar(1-h^2-k^2)^{1/2}} \tag{A.83}$$

$$\frac{\partial M_{41}}{\partial \lambda} = \frac{\partial M_{42}}{\partial \lambda} = 0 \tag{A.84}$$

$$\frac{\partial M_{43}}{\partial \lambda} = \frac{(1+p^2+q^2)}{2nar(1-h^2-k^2)^{1/2}}\frac{\partial Y_1}{\partial F} \tag{A.85}$$

$$\frac{\partial M_{51}}{\partial \lambda} = \frac{\partial M_{52}}{\partial \lambda} = 0 \tag{A.86}$$

$$\frac{\partial M_{53}}{\partial \lambda} = \frac{(1+p^2+q^2)}{2nar(1-h^2-k^2)^{1/2}}\frac{\partial X_1}{\partial F} \tag{A.87}$$

$$\frac{\partial M_{61}}{\partial \lambda} = \frac{1}{nar}\left[-2\frac{\partial X_1}{\partial F} + (1-h^2-k^2)^{1/2}\left(h\beta\frac{\partial^2 X_1}{\partial F \partial h} + k\beta\frac{\partial^2 X_1}{\partial F \partial k}\right)\right] \tag{A.88}$$

$$\frac{\partial M_{62}}{\partial \lambda} = \frac{1}{nar}\left[-2\frac{\partial Y_1}{\partial F} + (1-h^2-k^2)^{1/2}\left(h\beta\frac{\partial^2 Y_1}{\partial F \partial h} + k\beta\frac{\partial^2 Y_1}{\partial F \partial k}\right)\right] \tag{A.89}$$

$$\frac{\partial M_{63}}{\partial \lambda} = \frac{\left(q\frac{\partial Y_1}{\partial F} - p\frac{\partial X_1}{\partial F}\right)}{nar(1-h^2-k^2)^{1/2}} \tag{A.90}$$

The auxiliary partials are

$$\frac{\partial X_1}{\partial F} = a\left[hk\beta c_F - (1-h^2\beta)s_F\right] \tag{A.91}$$

$$\frac{\partial Y_1}{\partial F} = a\left[-hk\beta s_F + (1-k^2\beta)c_F\right] \tag{A.92}$$

$$\frac{\partial \dot{X}_1}{\partial F} = -\frac{a}{r}(ks_F - hc_F)\dot{X}_1 + \frac{a^2 n}{r}\left[-hk\beta s_F - (1-h^2\beta)c_F\right] \tag{A.93}$$

$$\frac{\partial \dot{Y}_1}{\partial F} = -\frac{a}{r}(ks_F - hc_F)\dot{Y}_1 + \frac{a^2 n}{r}\left[-hk\beta c_F - (1-k^2\beta)s_F\right] \tag{A.94}$$

$$\frac{\partial^2 X_1}{\partial F \partial h} = a\left[(hs_F + kc_F)\left(\beta + \frac{h^2\beta^3}{1-\beta}\right) + \frac{a^2}{r^2}(h\beta - s_F)(s_F - h) + \frac{a}{r}c_F^2\right] \tag{A.95}$$

$$\frac{\partial^2 X_1}{\partial F \partial k} = -a\left[-(hs_F + kc_F)\frac{hk\beta^3}{1-\beta} + \frac{a^2}{r^2}(s_F - h\beta)(c_F - k) + \frac{a}{r}s_F c_F\right] \tag{A.96}$$

$$\frac{\partial^2 Y_1}{\partial F \partial h} = a\left[-(hs_F + kc_F)\frac{hk\beta^3}{1-\beta} - \frac{a^2}{r^2}(k\beta - c_F)(s_F - h) + \frac{a}{r}s_F c_F\right] \tag{A.97}$$

$$\frac{\partial^2 Y_1}{\partial F \partial k} = a\left[-(hs_F + kc_F)\left(\beta + \frac{k^2\beta^3}{1-\beta}\right) + \frac{a^2}{r^2}(c_F - k\beta)(c_F - k) - \frac{a}{r}s_F^2\right] \tag{A.98}$$

References

[1]Edelbaum, T. N., "Propulsion Requirements for Controllable Satellites," *ARSJ,* Aug. 1961, pp. 1079–89.

[2]Broucke, R. A., and Cefola, P. J., "On the Equinoctial Orbit Elements," *Celestial Mechanics* 5, pp. 303–310, 1972.

[3]Herrick, S., *Astrodynamics,* Vol. II, Van Nostrand Reinhold, London, 1972.

[4]Plummer, H. C., *An Introductory Treatise on Dynamical Astronomy,* Dover, New York, 1960.

[5]Cefola, P. J., "Equinoctial Orbit Elements: Application to Artificial Satellite Orbits," AIAA Paper 72-937, AIAA/AAS Astrodynamics Conference, Palo Alto, CA, Sept. 11–12, 1972.

[6]Edelbaum, T. N., Sackett, L. L., and Malchow, H. L., "Optimal Low Thrust Geocentric Transfer," AIAA Paper 73-1074, AIAA 10th Electric Propulsion Conference, Lake Tahoe, NV, Oct. 31–Nov. 2, 1973.

[7]Cefola, P. J., Long, A. C., and Holloway, G., Jr., "The Long-Term Prediction of Artificial Satellite Orbits," AIAA Paper 74-170, AIAA 12th Aerospace Sciences Meeting, Washington, DC, Jan. 30–Feb. 1, 1974.

[8]Sackett, L. L., and Edelbaum, T. N., "Effect of Attitude Constraints on Solar-Electric Geocentric Transfers," AIAA Paper 75-350, AIAA 11th Electric Propulsion Conference, New Orleans, LA, March 19–21, 1975.

[9]Kechichian, J. A., "Equinoctial Orbit Elements: Application to Optimal Transfer Problems," AIAA Paper 90-2976, AIAA/AAS Astrodynamics Conference, Portland, OR, Aug. 20–22, 1990.

[10]Kechichian, J. A., "Trajectory Optimization with a Modified Set of Equinoctial Orbit Elements," AAS/AIAA Paper 91-524, Astrodynamics Specialist Conference, Durango, CO, Aug. 19–22, 1991.

[11]Bryson, A. E., and Ho, Y-C., *Applied Optimal Control,* Ginn, Waltham, MA, 1969.

[12]Marec, J-P., *Optimal Space Trajectories,* Elsevier, Amsterdam, 1979.

[13]Vinh, N. X., *Optimal Trajectories in Atmospheric Flight,* Elsevier, Amsterdam, 1981.

[14]Kahaner, D., Moler, C., and Nash, S., *Numerical Methods and Software,* Prentice-Hall, Englewood Cliffs, NJ, 1989.

[15]Kechichian, J. A., "Optimal Low-Thrust Rendezvous Using Equinoctial Orbit Elements," IAF Paper 92-0014, 43rd Congress of the International Astronautical Federation, Washington, DC, Aug. 28–Sept. 5, 1992.

[16]Kechichian, J. A., "Optimal Low-Thrust Transfer Using Variable Bounded Thrust," IAF Paper 93A.2.10, 44th Congress of the International Astronautical Federation, Graz, Austria, Oct. 16–22, 1993.

[17]Edelbaum, T. N., "Optimal Space Trajectories," Analytic Mechanics Associates, Inc., Jericho, NY, Dec. 1969.

[18]Hill, P. G., and Peterson, C. R., *Mechanics and Thermodynamics of Propulsion,* Addison-Wesley, Reading, MA, 1970.

[19]Kechichian, J. A., "Minimum-Fuel Time-Fixed Rendezvous Using Constant Low Thrust," AAS Paper 93-130, AAS/AIAA Spaceflight Mechanics Meeting, Pasadena, CA, Feb. 22–24, 1993.

[20]Kechichian, J. A., "Optimal LEO-GEO Intermediate Acceleration Orbit Transfer," AAS Paper 94-125, AAS/AIAA Spaceflight Mechanics Meeting, Cocoa Beach, FL, Feb. 14–16, 1994.

15
Orbital Coverage

For many classes of satellites, the primary function involves coverage of the Earth's surface. Satellites used for communications, weather, navigation, Earth resources, or surveillance clearly fall into this category. In addition, satellites that monitor activity near the Earth (e.g., air-breathing vehicle or launch detection), up to low Earth orbit (LEO), can be placed in this category as well.

Earth coverage requirements are usually specified in terms of area of interest, frequency, and duration. The most common areas of interest for satellite coverage are global, zonal, or regional coverage. Examples of these coverage areas are shown in Fig. 15.1. Global coverage involves coverage of the entire Earth. Such coverage would be important for worldwide communications, weather, or navigation functions. Zonal coverage requires viewing the area between two values of latitude. This type of coverage could be employed to bring telephone communications to the United States and Europe, which areas are bounded roughly by the same latitudes. Regional coverage could be used to provide satellite TV to a single region such as the United States.

The required frequency and duration of coverage is also important. Communication and navigation systems usually must be available 24 h/day. Other functions may require coverage of the area of interest for only part of the day (e.g., 8 h/day). Still others, such as weather, resources, or surveillance satellites, may require only occasional views of the area of interest. A resources satellite, for instance, may require only weekly viewing of points within its area of interest. A surveillance satellite may require viewing certain seaports, for instance, every few hours. This time interval is referred to as the *revisit time*.

To achieve these area and time requirements for coverage, one or more satellites may be required. A single satellite is capable of providing global coverage, but it will view any point on Earth for only a short time on given days. Similarly, a given region of the Earth can be covered continuously (24 h/day) by a single satellite, but only if the region is sufficiently small. More demanding coverage requirements cannot be met by a single satellite, and so groups or constellations of satellites working together are often necessary. As we shall see, continuous global coverage requires at least four high-altitude satellites or even hundreds of low-altitude satellites. The study of optimal satellite constellations serves to reduce program costs by finding the smallest number of satellites required to perform a coverage task.

In studying orbital coverage, we shall first investigate the coverage offered by a single satellite and then look at methods of generating optimal constellations of satellites.

15.1 Coverage from a Single Satellite

Single-Satellite Coverage Geometry

The viewing geometry for a single satellite is shown in Fig. 15.2, where

h = satellite altitude

r_e = Earth's equatorial radius (6378.135 km)

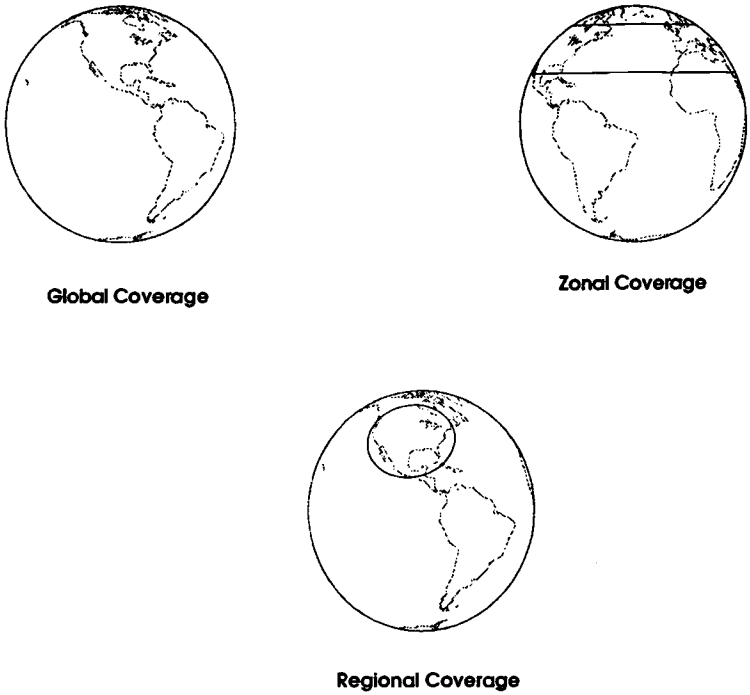


Fig. 15.1 Types of Earth coverage.

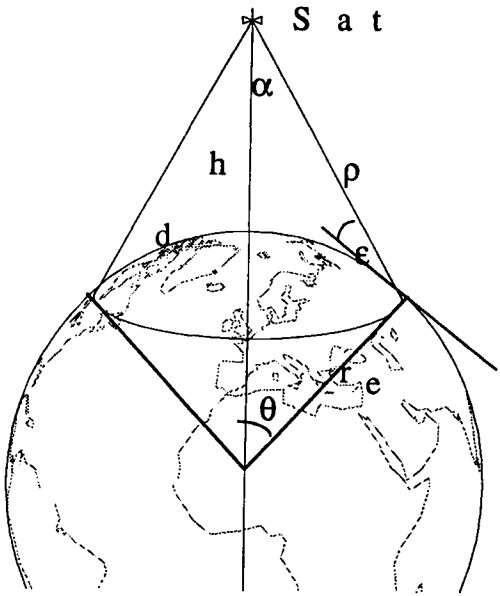


Fig. 15.2 Satellite coverage geometry.

ε = ground elevation angle
θ = Earth central angle of coverage
α = satellite field-of-view angle
d = radius of coverage circle on Earth's surface
ρ = slant range distance

The instantaneous coverage of the satellite is most often defined by a conical field of view that intersects the Earth's surface to form a circular footprint centered on the subsatellite point. The field of view of the satellite is limited by the horizon, $\varepsilon = 0$. Higher minimum values of ground elevation angle (ε) are often used to allow for atmospheric losses or obscuring terrain. Sometimes the satellite sensor further limits visibility because of its particular angular field of view (α) or slant range (ρ) limitations. The parameters that describe the satellite's instantaneous coverage are related by the following equations:

$$\cos(\theta + \varepsilon) = \frac{\cos\varepsilon}{1 + (h/r_e)} = \sin\alpha \tag{15.1}$$

$$\tan\varepsilon = \frac{\cos\theta - [r_e/(r_e + h)]}{\sin\theta} \tag{15.2}$$

$$\theta = \frac{360d}{2\pi r_e} \tag{15.3}$$

$$\rho^2 = r_e^2 + (r_e + h)^2 - 2r_e(r_e + h)\cos\theta \tag{15.4}$$

Equation (15.1) can be used to find the Earth central angle of the coverage circle (θ) when the ground elevation angle (ε) and either the satellite altitude (h) or field-of-view angle (α) are known. Figure 15.3 shows the radius of the coverage circle (θ) as a function of satellite altitude (h) and ground elevation angle (ε). Notice that the size of the coverage circle falls off rapidly at lower altitudes and, thus, is quite sensitive to ground elevation angle (ε). Equation (15.2) allows computation of the ground elevation angle (ε) if the satellite altitude (h) and Earth central angle of coverage (θ) are known. Equations (15.3) and (15.4) relate the Earth's surface coverage radius (d) and maximum slant range (ρ), respectively, to the other parameters.

For a satellite in a circular orbit, the coverage circle on the Earth will not change in size with time but will simply move along the groundtrack of the satellite. The coverage circle for a satellite in an elliptical orbit will grow as the satellite ascends and shrink as it descends.

Coverage from a Single Low-Earth Orbit Satellite

The Earth coverage provided from a low-Earth orbit (LEO) satellite after two revolutions is shown in Fig. 15.4. The dashed line is the satellite groundtrack, and the solid lines surrounding it represent the region around the groundtrack, swept out by the satellite's coverage circle. This region is called the *coverage swath*. The instantaneous coverage circle for a single point in time is also shown. As discussed in Chapter 10, the groundtrack of the LEO satellite has shifted westward on the second revolution by an amount S in longitude as a result of two factors: the first that the Earth has rotated eastward during one revolution of the satellite, and the

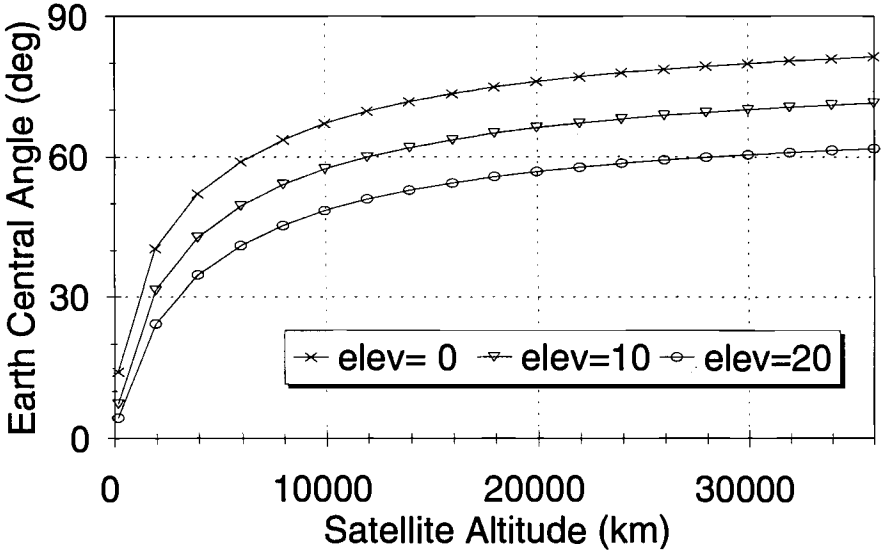


Fig. 15.3 Earth central angle of coverage for a single satellite.

second that the satellite's orbital plane has rotated westward because of Earth's oblateness, J_2, as described in Eq. (10.1). As a result of these two factors, the satellite groundtrack shifts to the west in longitude on each pass by an amount

$$S = P_n(\omega_e - \dot{\Omega}) \tag{15.5}$$

where ω_e is the rotation rate of the Earth ($\omega_e = 0.250684454$ deg/min) and $\dot{\Omega}$ is the rate of nodal regression of the orbit as a result of Earth's oblateness ($J_2 = 0.0010826$). Note that this longitudinal shift of the groundtrack is the same at all latitudes. The nodal regression rate [see Eq. (10.1)] is given by

$$\dot{\Omega} = -\frac{3}{2}\frac{J_2 r_e^2}{a_0^2\left(1-e_0^2\right)^2}\left(\frac{\mu}{a_0^3}\right)^{0.5}\cos i_0 \quad \text{(rad/s)}$$

or (15.6)

$$\dot{\Omega} = \frac{-9.9639}{\left(1-e_0^2\right)^2}\left(\frac{r_e}{a_0}\right)^{3.5}\cos i_0 \quad \text{(deg/mean solar day)}$$

P_n in Eq. (15.5) is the nodal period of the orbit, that is, the time required to go from one ascending node to the next, and is given by

$$P_n = P_0\left\{1 - \frac{3}{2}J_2\frac{(1+e_0\cos\omega_0)^3}{(a_0/r_e)^2\left(1-e_0^2\right)^3} - \frac{3}{4}J_2\frac{\left(4-5\sin^2 i_0\right)}{(a_0/r_e)^2\left(1-e_0^2\right)^{0.5}(1+e_0\cos\omega_0)^2}\right\} \tag{15.7}$$

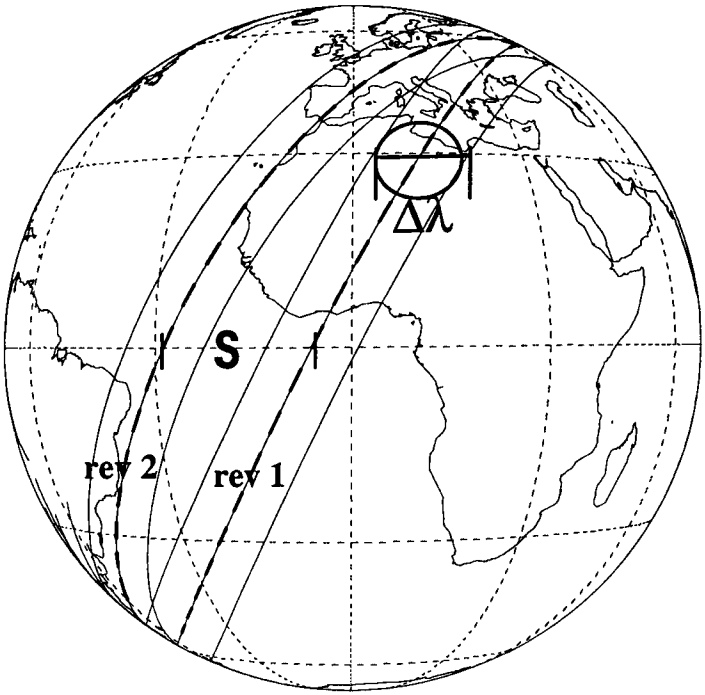


Fig. 15.4 Coverage from a single LEO satellite.

where

a_0 = initial semimajor axis
e_0 = initial eccentricity
i_0 = initial inclination
ω_0 = initial argument of perigee

are the osculating elements at the ascending node and

$$P_0 = \text{Keplerian period} = 2\pi\left(a_0^3/\mu\right)^{0.5}$$

It is useful in examining subsequent revolutions of a satellite to define the dimensionless satellite trace repetition parameter Q:

$$Q = 360/S \tag{15.8}$$

Note that S is measured positively westward. From Fig. 15.4 it can be seen that, if the westward shift S of each revolution is such that it divides into 360 deg an integral number of times, then the $Q + 1$ pass will have the same ascending node longitude as the first pass. Therefore, if Q is an integer, the orbit is called a *repeating groundtrack orbit*, whose groundtrack repeats itself each day after Q revolutions. A geosynchronous orbit (see Sec. 11.3) is a $Q = 1$ orbit since

its groundtrack repeats daily after a single revolution. A Molniya orbit (see Sec. 11.4) is an example of a $Q = 2$ orbit. Its groundtrack repeats daily after two revolutions.

In general, Q need not be an integer but can be expressed as a ratio of integers:

$$Q = \frac{N}{D} \tag{15.9}$$

In this form, D is the number of days until the groundtrack repeats itself, and N is the number of orbit revolutions until repeat. As an example, a $Q = 12\frac{1}{3} = 37/3$ orbit will repeat the same groundtrack every three days, after 37 revolutions. A $Q = 12.3 = 123/10$ orbit will repeat the same groundtrack every 10 days, after 123 revolutions. Even though the groundtrack of this latter orbit does repeat, it would probably not be considered a repeating groundtrack orbit since the repeat interval is so long.

The satellite trace repetition parameter Q can be used to analyze how successive passes and their coverage swaths overlay on the Earth's surface. For satellites at the same altitude, it can be used to interleave their passes to achieve the desired coverage.

At any given latitude (ϕ), the longitude region ($\Delta\lambda$), as shown in Fig. 15.4 swept out by the swath, can be found from

$$\Delta\lambda = \sin^{-1}\left(\frac{\sin\theta + \sin\phi\cos i}{\sin i\cos\phi}\right) + \sin^{-1}\left(\frac{\sin\theta - \sin\phi\cos i}{\sin i\cos\phi}\right) \tag{15.10}$$

for values of latitude ϕ such that $-i \mp \theta \leq \phi \leq i \pm \theta$ (where the upper sign is for posigrade inclinations and the lower sign is for retrograde inclinations). When the swath width θ is small, this equation becomes

$$\Delta\lambda \approx \frac{2\theta}{(\cos^2\phi - \cos^2 i)^{0.5}} \tag{15.11}$$

For the latitude value in which the longitude swept out is equal to the distance between groundtracks (i.e., $\Delta\lambda = S$), there will be no gaps between swaths of successive revolutions. The entire region above this latitude and up to a latitude near $(i + \theta)$ for posigrade orbits will be covered by the satellite in a single day. In fact, it will be covered at least twice, once by northbound passes and once by southbound passes.

Clearly, the equator is the most difficult latitude to cover using highly inclined orbits. Complete coverage of the equator requires a coverage circle size given by

$$\theta = \sin^{-1}\left(\sin\frac{S}{2}\sin i\right) \tag{15.12}$$

Coverage of the poles is assured so long as $i + \theta \geq 90$ (or $i - \theta \leq 90$ for retrograde orbits). In this case, global coverage would be achieved. The revisit time at the equator would be about 12 h, with shorter revisit times toward the poles. A given point on Earth would see the satellite two or more times per day, but each viewing opportunity would be only minutes in duration.

Reducing the revisit time could be accomplished by using additional LEO satellites that come into view of the region of interest at other times of day so as to break up large coverage gap intervals. Hanson et al.[1] have studied methods for selecting the initial values of longitude of ascending nodes and mean anomaly to minimize revisit time for a constellation of satellites to a point on the Earth. To achieve continuous global coverage using LEO satellites obviously requires constellations of many satellites. This subject will be addressed later in the chapter.

The LEO orbit is a good choice for Earth resources, weather, or surveillance satellites, which do not require continuous or even quick revisit time coverage. In addition, sensors with limited slant range capability find the LEO orbit a necessity. The LEO orbit also has the advantage of requiring the least energy to achieve (see Fig. 5.2). Hence, it is favored for extremely heavy satellites (e.g., Space Shuttle, Space Station, Hubble Telescope) and those seeking to launch on small launch vehicles (e.g., Pegasus).

Coverage from a Single Geosynchronous Equatorial Orbit Satellite

By definition, a geosynchronous satellite revolves in its orbit at the same rate at which the Earth rotates about its polar axis. If the orbital plane of the satellite is equatorial ($i = 0$), the satellite remains over the same point on the equator. Its groundtrack is simply a point on the Earth.

Figure 15.5 shows the Earth coverage from a single geosynchronous equatorial orbit (GEO) satellite. Because the satellite does not move relative to the Earth, it has the same view continuously. Contours of different ground elevation angle (ε) are shown on the figure. Because the satellite is at high altitude ($h = 35{,}786$ km), it can see a large region of the Earth continuously. The GEO orbit is an excellent choice for continuous coverage of nearly a hemisphere of the Earth. A single satellite can provide 24-h communication coverage for the North and South American continents (excluding the extreme polar regions). A single satellite could also continuously link most of Europe, Africa, South America, and North America as shown in Fig. 15.5. A set of three GEO satellites would continuously cover all but the polar regions as shown in Fig. 15.6. The drawbacks of the GEO orbit are the large amount of energy required to achieve it (see Fig. 5.2), which translates into high launch costs, and the large sensor range required.

A geosynchronous satellite that has a nonzero inclination will trace out a figure eight on the Earth. The maximum latitude excursion of the subsatellite point will equal the orbital inclination. Figure 15.7 shows the groundtrack and coverage for such a satellite with an inclination of 60 deg. The contours show the regions that are covered 24, 18, 12, 6, and 0 h/day. Because the satellite is moving relative to the Earth's surface, it has visibility to a greater region than the motionless equatorial GEO satellite did. The only region that this satellite never sees is the small football-shaped region 180 deg away from its groundtrack on the Earth. On the other hand, the region that is always in view, namely, the small football-shaped region centered on the groundtrack, is much smaller for the inclined GEO. So, although, the inclined GEO covers more of the Earth, it covers less of it continuously (24 h/day). Unlike the equatorial GEO, an inclined geosynchronous orbit would allow access to the polar regions for significant time periods each day. The North Pole, which could not be seen by the equatorial GEO, can be covered about 10 h/day by the inclined GEO.

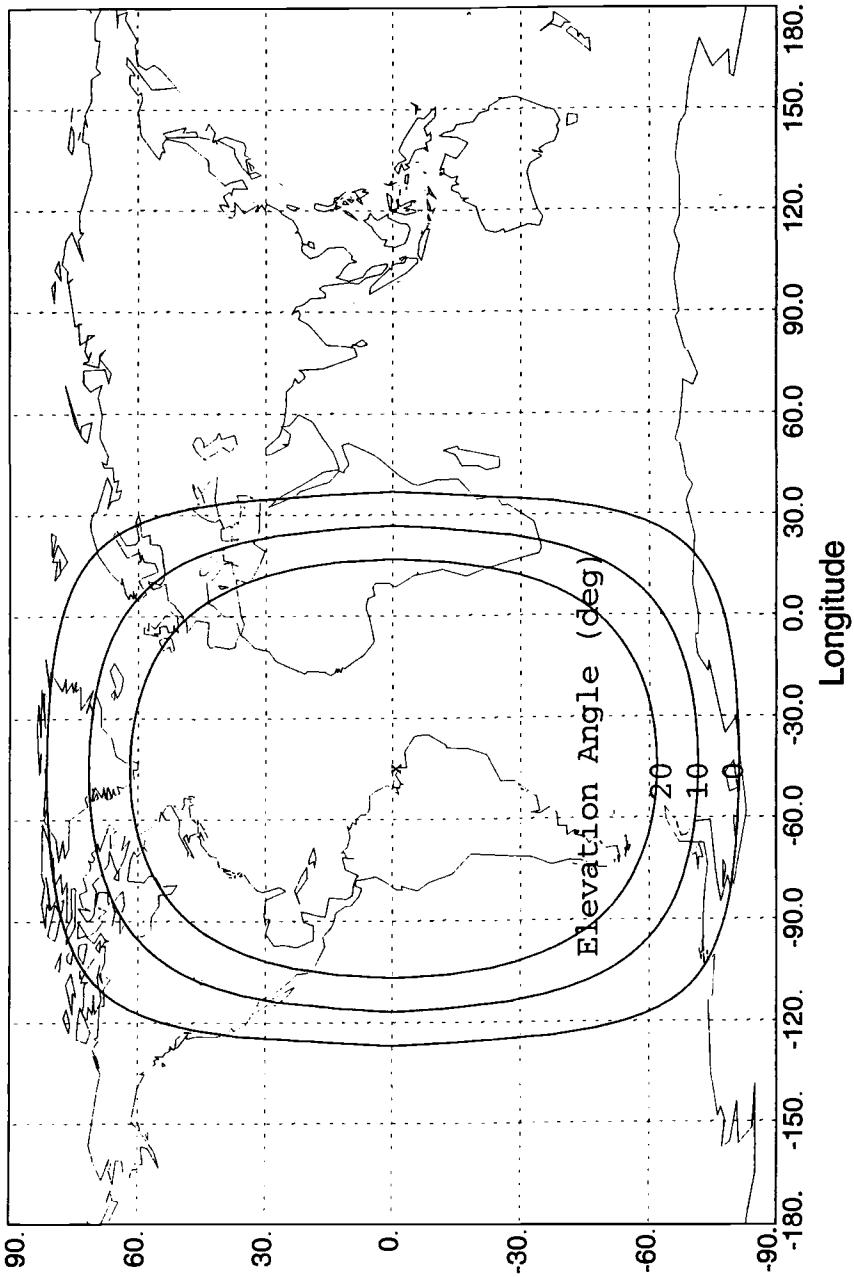


Fig. 15.5 Coverage from a single GEO satellite.

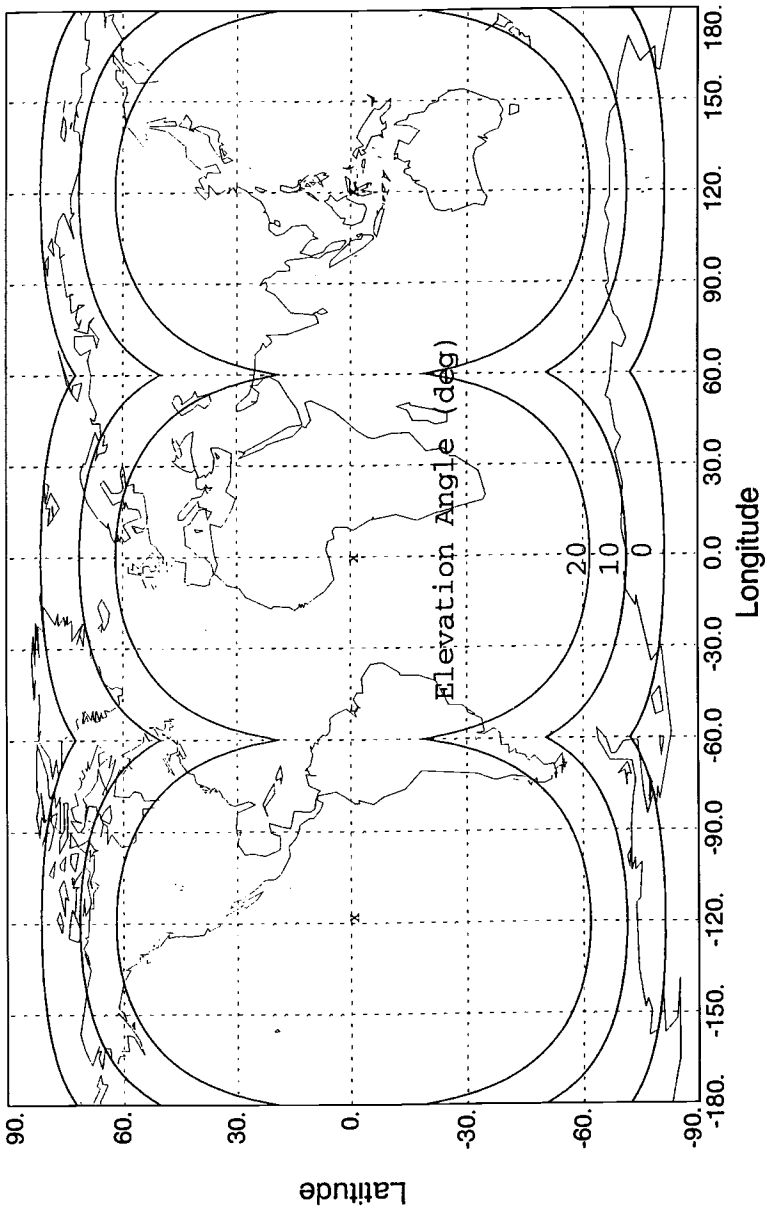


Fig. 15.6 Coverage from three GEO satellites.

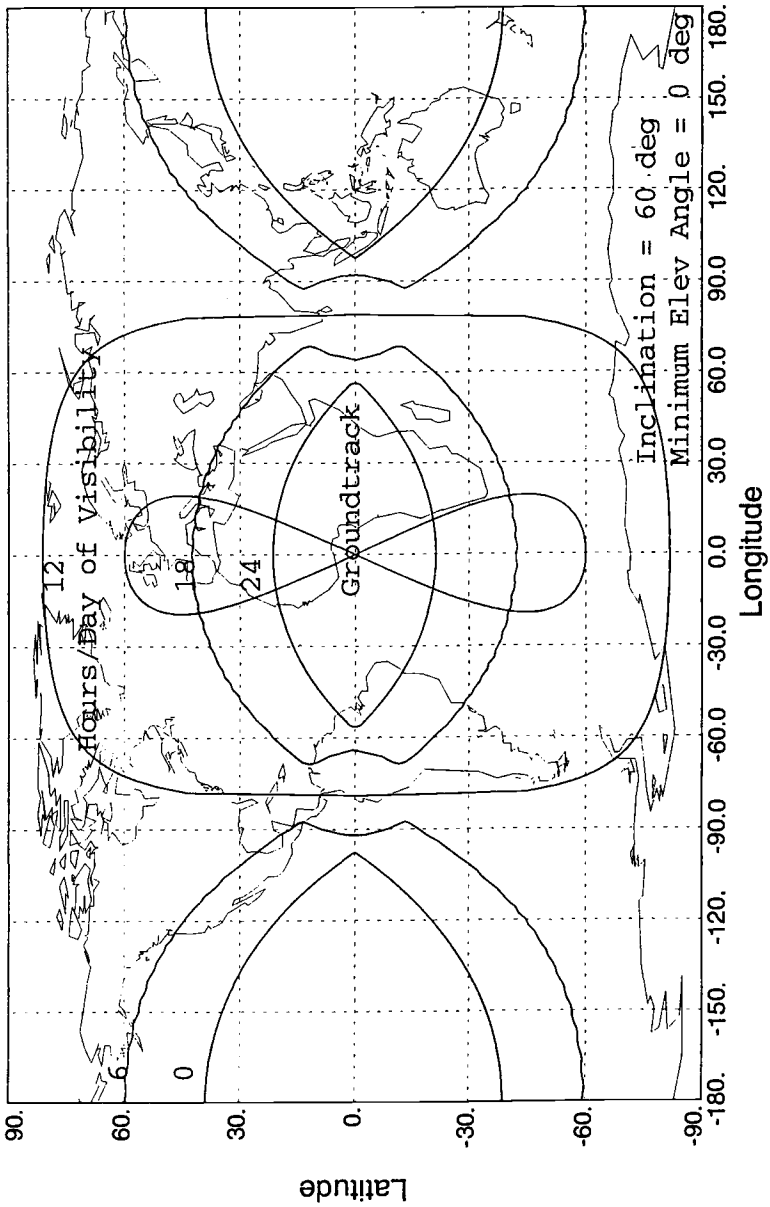


Fig. 15.7 Coverage from one inclined GEO satellite.

Coverage of the polar regions can be further enhanced by using inclined, eccentric GEO orbits. An example is shown in Fig. 15.8, for a geosynchronous satellite with an inclination of 30 deg, an eccentricity of 0.3, and an argument of perigee of 270 deg. (Note that, because the inclination is not at the critical value of 63.435 deg, the argument of perigee will increase slowly with time as a result of the effects of J_2.) This satellite traverses a nearly circular groundtrack in a counterclockwise direction, with apogee at the northernmost point and perigee at the southernmost point. As a result, the satellite spends more time north of the equator than south of it. Consequently, the coverage is shifted toward the northern hemisphere. The North Pole is now covered about 14 h/day compared to about 5 h/day for the South Pole. Again, the region covered continuously (24 h) is smaller than for the circular, equatorial GEO.

Combinations of GEO satellites with and without inclination and eccentricity can be used effectively to provide regional, global, or even polar coverage. Hanson and Higgins[2] examined such combinations of GEO satellites to maximize coverage of six different geographic areas. Their results show that, for global or near-global coverage, constellations of elliptical or circular GEO satellites perform about equally well. For coverage of the northern hemisphere or of a region such as the United States, North Atlantic, and Western Europe, the elliptical GEO constellations offer the better coverage.

Coverage from a Single Highly Eccentric Orbit Satellite

In the previous section, it was noted that eccentric orbits with apogee located at the northernmost point in the orbit could be used to shift coverage to favor the northern hemisphere. In the current section, this concept is taken nearly to extreme in the study of the highly eccentric orbit (HEO).

In a highly eccentric orbit, the satellite spends most of its time in the region of apogee at a high-altitude vantage point, where it sees the largest surface area of the Earth. Relatively little time is spent at low altitudes because the satellite speeds through perigee on its way back out toward apogee. By far the most common HEO is the highly eccentric, critically inclined $Q = 2$ (Molniya) orbit that was examined in Sec. 11.4. The typical Molniya orbit has an apogee altitude higher than GEO and a perigee altitude in the 900- to 1800-km range. Its orbital period of 11.967 h is one-half of a mean sidereal day ($Q = 2$), so that the groundtrack of the satellite repeats itself every two revolutions. The critical inclination of 63.435 deg is employed to prevent rotation of the line of apsides so as to maintain apogee at the northernmost point in the orbit.

The instantaneous view of the Earth from a Molniya satellite at apogee is shown in Fig. 15.9. The groundtrack and coverage provided by a single Molniya satellite are shown in Fig. 15.10. The groundtrack is labeled with two apogees, 180° apart in longitude, and two perigees similarly spaced. In a single day, the satellite traverses the groundtrack from west to east and returns to repeat the same groundtrack the next day. Since the satellite spends 11 h of its 12-h period north of the equator (most of that time in the vicinity of apogee), the coverage is concentrated in the northernmost regions. Notice that a significant amount of the northern region is covered 6, 9, and even 12 h/day. In cases where extended coverage of a polar region (typically northern) is desired, the Molniya orbit is an excellent choice. It

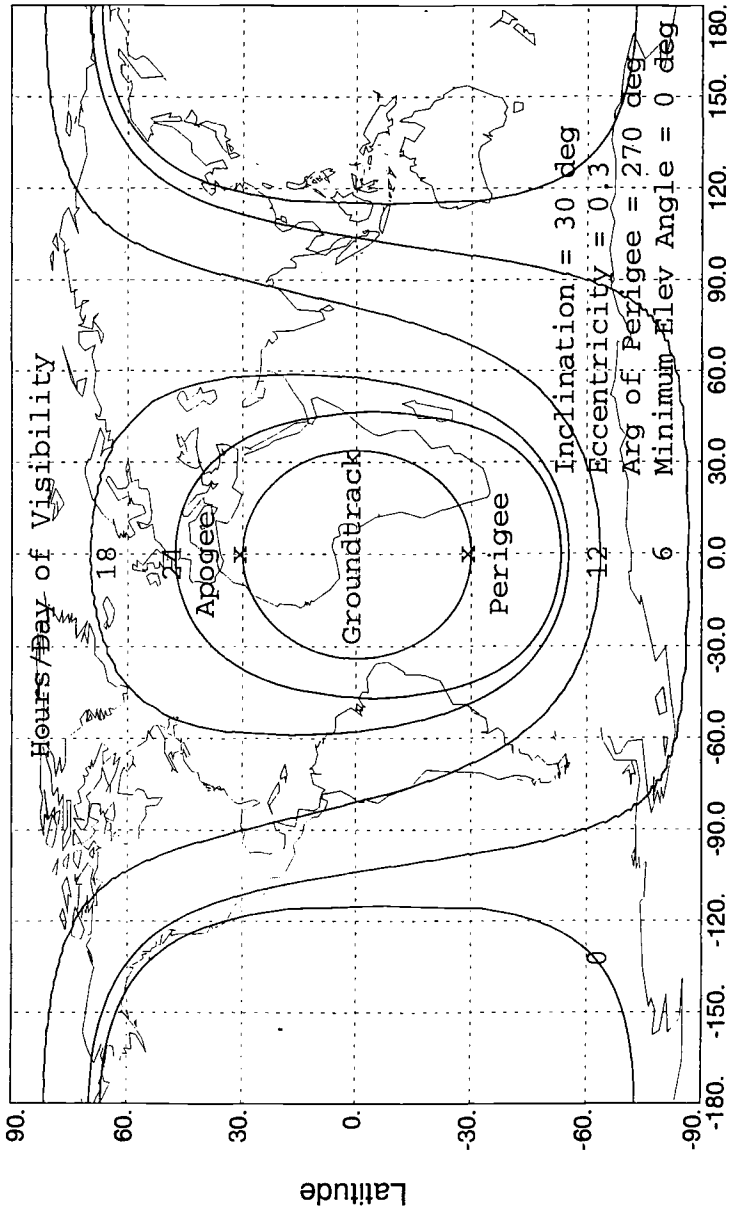


Fig. 15.8 Coverage from one inclined eccentric GEO.

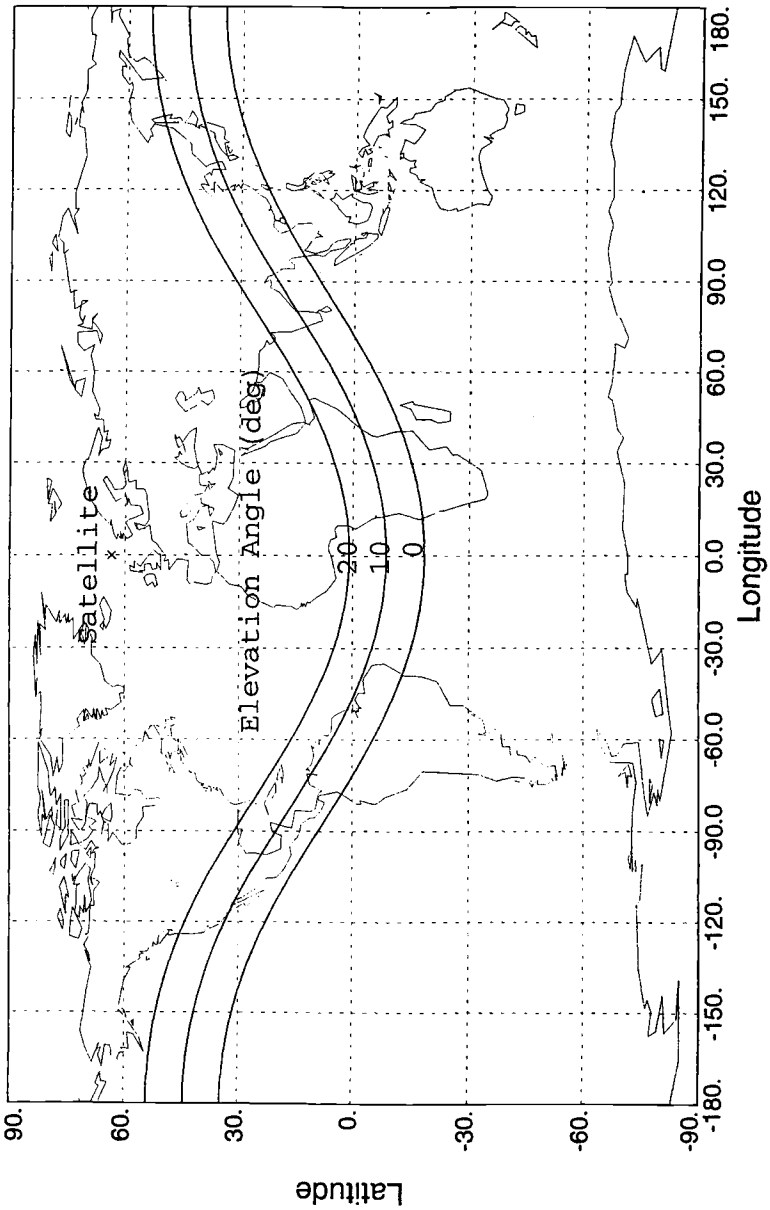


Fig. 15.9 Instantaneous view by HEO at apogee.

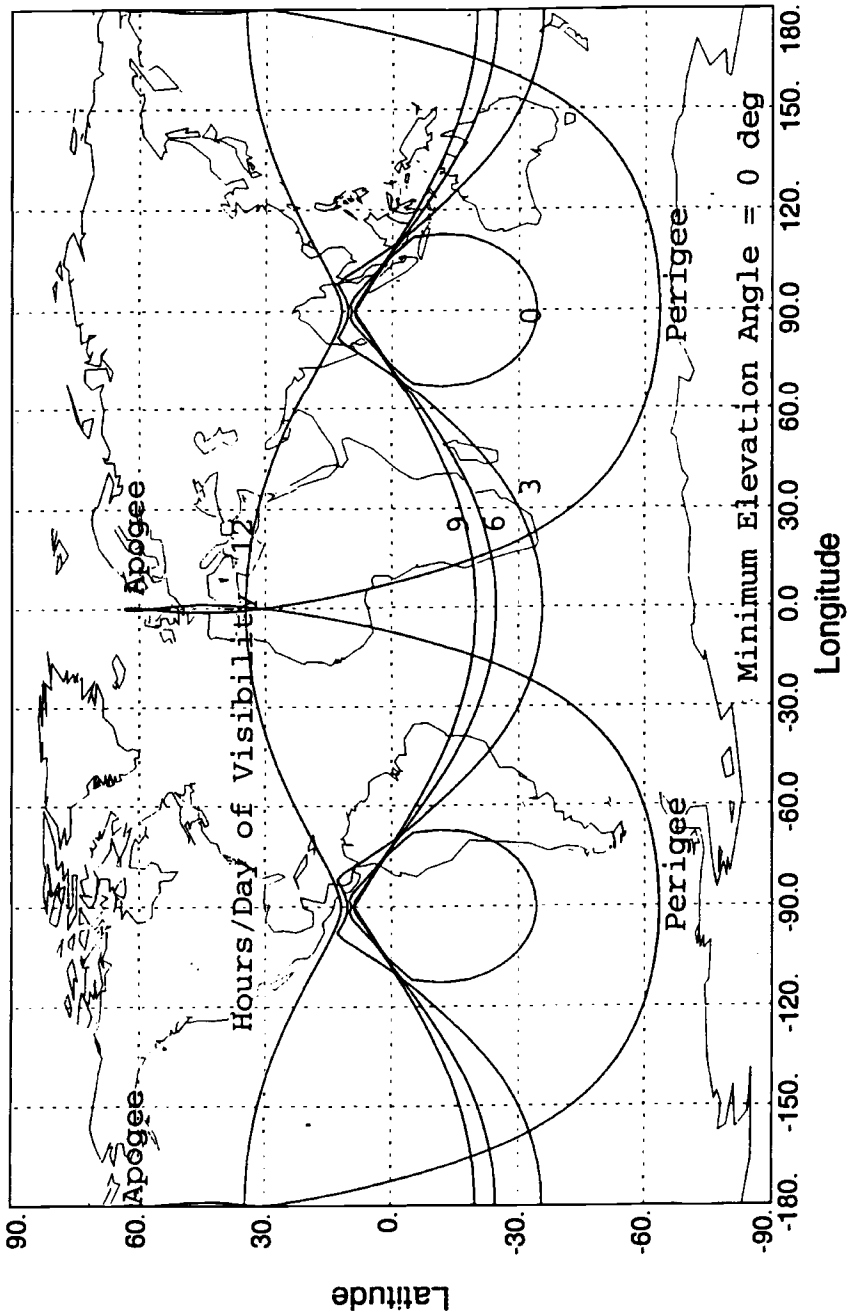


Fig. 15.10 Coverage from a single HEO satellite.

requires less energy to achieve than does the GEO orbit (see Fig 5.2), because perigee remains at low altitude.

For continuous viewing of northern regions, constellations of two or three Molniya satellites are often used. For a two-satellite arrangement, the second satellite is placed in the same groundtrack as the first but phased so that one satellite is at perigee while the other is at apogee. To do this requires using orbital planes 90 deg apart in right ascension of ascending node and phasings (mean anomalies) that differ by 180 deg. For a three-satellite arrangement, all three satellites are placed in the same groundtrack but phased 8 h (240 deg in mean anomaly) apart. The right ascensions of ascending node in this case are spaced 120 deg apart. The continuous and partial coverages from these two- and three-satellite constellations are shown in Figs. 15.11 and 15.12, respectively. The two-satellite HEO constellation can continuously cover most of the Earth above 30° north latitude. The three-satellite HEO constellation can continuously cover nearly all of the northern hemisphere. Clearly, these constellations of HEO satellites are quite efficient at concentrating coverage in the northern (or southern if apogee is placed at the southernmost point in the orbit) regions of the Earth.

Coverage from a Single Medium Earth Orbit Satellite

Although there is no strict definition of what constitutes a medium earth orbit (MEO), it is safe to say that any orbit too high to be labeled LEO, too low for GEO, and not specifically an HEO is an MEO. Typically, the MEO label is applied to orbits whose periods range from about 2 to 18 h and whose eccentricity is small. This range includes orbit altitudes in the 2000- to 30,000-km region. If the LEO, with its small coverage circle, is at one extreme of the coverage realm and the GEO, with its nearly hemispheric coverage circle, is at the other, then the MEO falls in between.

In Fig. 15.3, the LEO orbits are near the left end of the plot, and the GEO orbits are near the right end of the plot. The MEO orbits constitute the rest. In Fig. 15.3, note how quickly the size of the coverage circle increases as the orbital altitude is increased from 2000 to 10,000 km. Beyond this point, the coverage circle size increases only moderately with altitude. The MEO can offer the orbit planner a middle ground between the LEO and GEO alternatives. The MEO offers a coverage circle considerably larger than LEO and not much smaller than GEO. Its orbit requires more energy to attain than LEO but less than GEO. Required sensor range, while greater than for a LEO, is considerably less than for a GEO.

The MEO range is currently inhabited primarily by navigation satellites. The U.S. Global Positioning System (GPS or NAVSTAR) consists of 24 satellites in six orbit planes (four satellites per plane) inclined at 55 deg. The satellites are at an altitude of about 20,000 km, in a $Q = 2$ circular orbit, with a groundtrack that repeats each day. Figure 15.13 shows the groundtrack of a single GPS satellite and the coverage circle projected on the Earth at an instant in time. Note that the groundtrack repeats itself after two revolutions (12-h orbit period) and that the size of the coverage circle is not much smaller than for a GEO satellite, even though the altitude is nearly half that of GEO. The purpose of this satellite system is to provide at least fourfold continuous Earth coverage to allow a user to determine his position accurately. The Russian GLONASS (Global Navigation

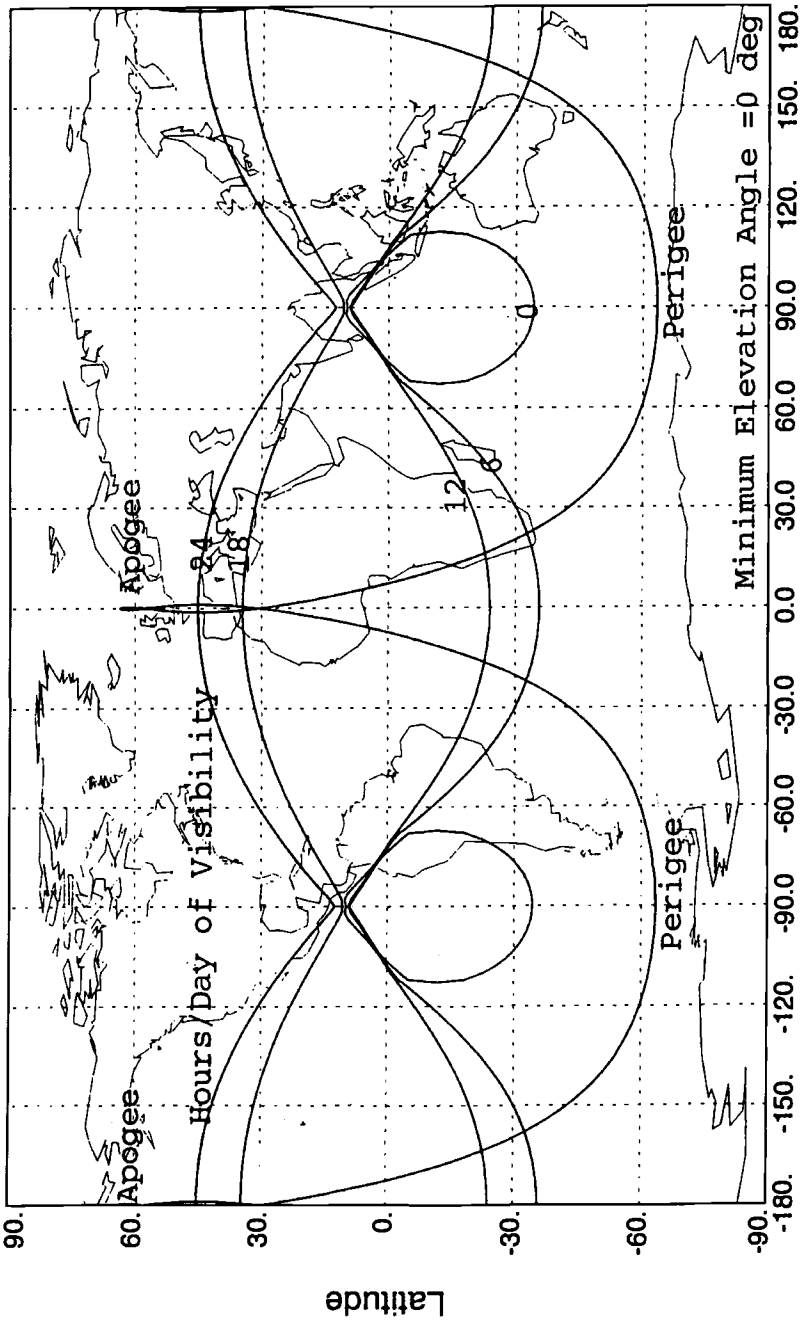


Fig. 15.11 Coverage from two HEO satellites.

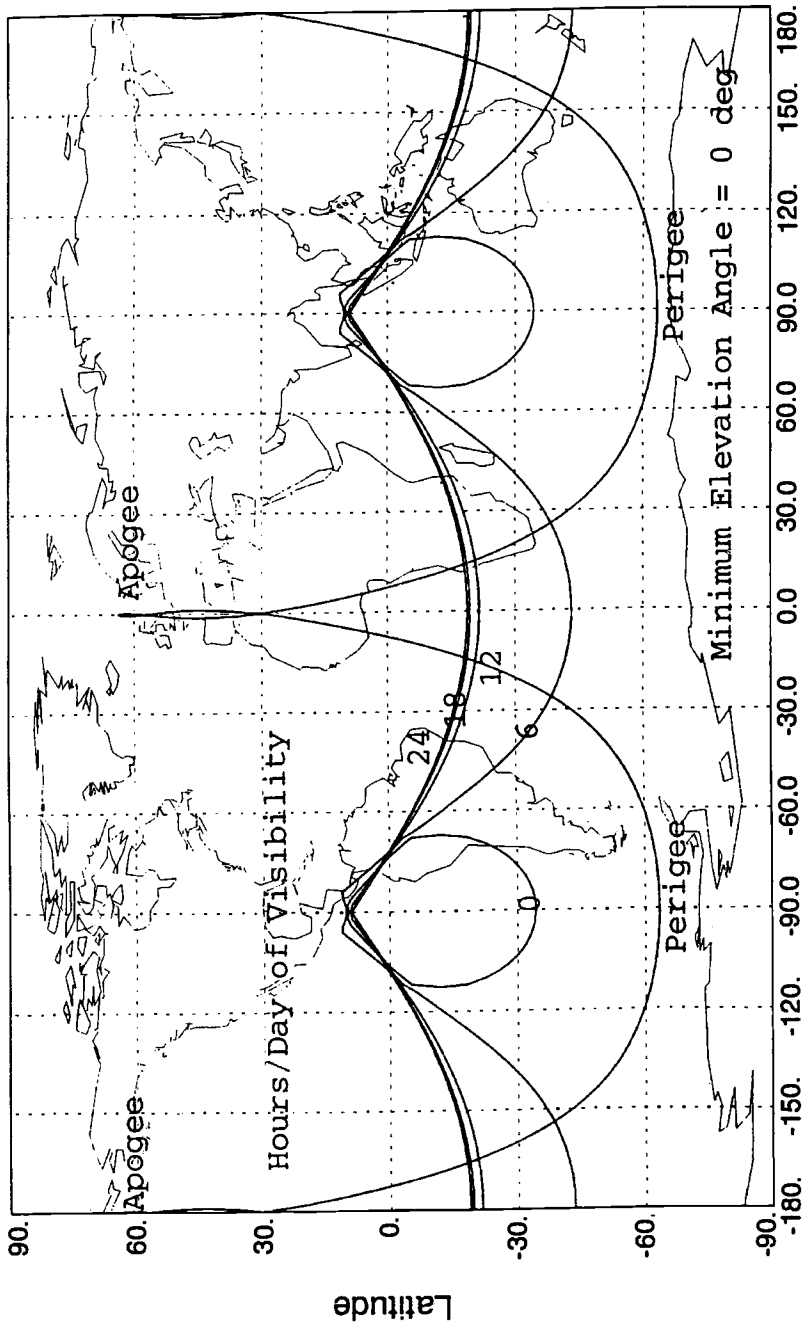


Fig. 15.12 Coverage from three HEO satellites.

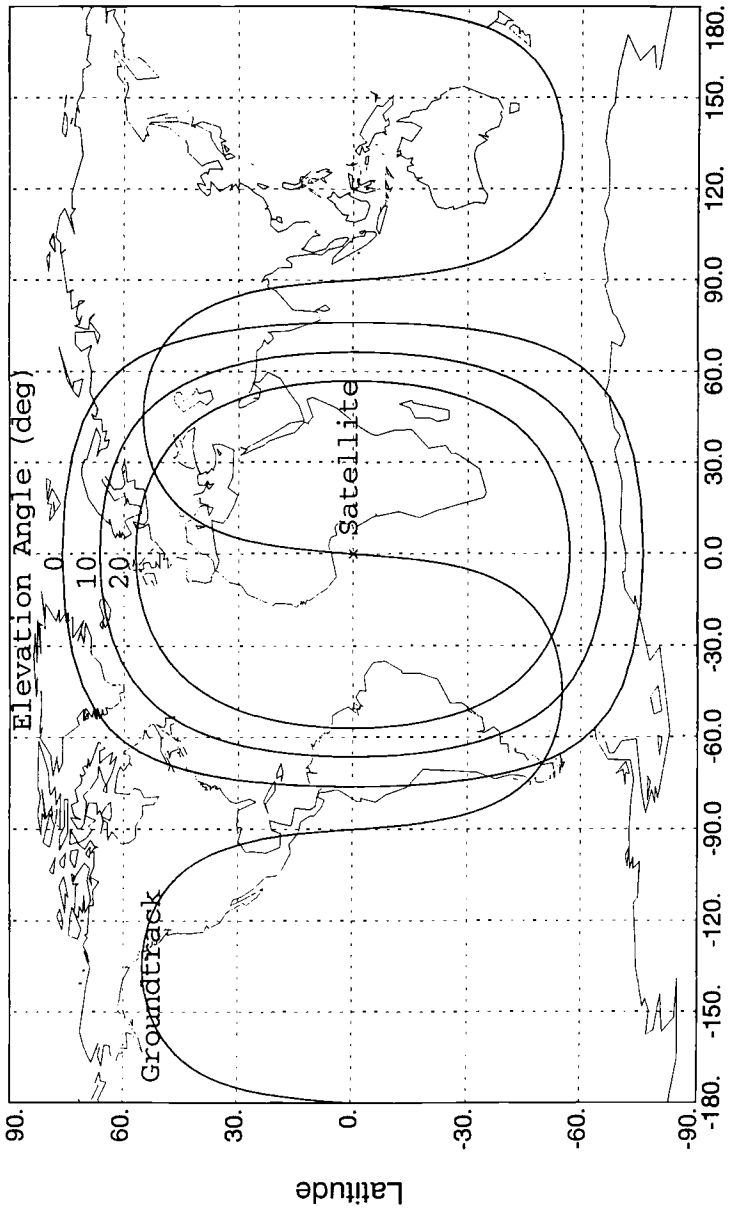


Fig. 15.13 Groundtrack and coverage for a GPS satellite.

Satellite System), whose purpose is similar, consists of 24 satellites in three orbit planes at an inclination of 64.8 deg. The satellites are at an altitude of 19,100 km in an eight-day repeating groundtrack.

Whereas these navigation satellite systems populate the higher altitudes of the MEO region, the lower MEO altitudes are relatively unpopulated. This may be due in part to the Van Allen radiation belts that inhabit this area. The fear of radiation-induced degradation to satellite solar arrays and electronic components may have deterred mission planners from using these orbits in the past. Progress in electronics technologies, however, has made available more radiation-tolerant electronic components and solar arrays. In the future, we may see constellations of MEO satellites proposed as the outcome of LEO vs GEO tradeoffs.

Along these lines, Draim and Kacena[3] have published an elegant paper advocating the advantages of MEO orbits. They propose a nondimensional coverage parameter that indicates the efficiency of an orbit. The parameter includes the effect of the total ΔV needed to achieve the orbit (launchability), the mean surface area covered by the satellite (Earth coverage), and the maximum slant range required by the sensor (satellite weight). The parameter is intended as a measure of what you get divided by what you pay for it. This efficiency parameter tends to be optimized in the 2000- to 20,000-km altitude range and appears to peak in the vicinity of 3500 km. This MEO regime, which currently is sparsely inhabited, seems to have sparked recent interest for the purpose of global communications.

15.2 Design of Optimal Satellite Constellations for Continuous Zonal and Global Coverage

While single satellites are useful for certain coverage functions, missions that involve the coverage of large regions of the Earth for long periods of time usually require constellations of satellites. We have already noted how multiple LEO satellites could be used to reduce the revisit time (the time during which an Earth point is not seen). We observed (Fig. 15.6) that three GEO satellites would provide nearly continuous global coverage (the polar regions are not covered). Finally, we examined how two or three Molniya satellites (Figs. 15.11 and 15.12) could work together to concentrate their coverage in the northern hemisphere. These are simple examples of how constellations of satellites working together can offer improved coverage.

In this section, the goal of designing optimal satellite constellations for the purpose of continuous zonal and global coverage will be examined. Sometimes, the nature of the mission requires visibility from more than a single satellite. A good example of this is the navigation function. The GPS/NAVSTAR requires simultaneous visibility to at least three satellites (i.e., at least continuous threefold coverage) to give a user on the ground a good position fix. Consequently, the subject of multiple folds of coverage will be explored.

The objective of designing optimal satellite constellations is to reduce the number of satellites required at a given altitude to provide the required level or fold of continuous zonal or global coverage. If optimal constellations are available for a number of different altitudes, designers can select the best altitude for their needs based on launch costs and satellite production costs.

As an example, consider the designer of a communication satellite constellation tasked to provide continuous global coverage. He could select a constellation of 5 GEO satellites or a constellation of perhaps 25 LEO satellites to do the same job. There would be far fewer GEO satellites, but his antenna and power systems would have to be sized for slant range operation at perhaps 25 times that of the LEO satellite (note that communication losses increase as r^2). A heavier satellite would be required in GEO, and the energy requirements to attain orbit are much larger for GEO than LEO. Clearly, a larger satellite and launch system would be involved. In contrast, the LEO communication satellite would be much lighter and easier to launch, and several of them might be placed in orbit on the same launch vehicle. Determining the better alternative in this case requires optimal constellations for both options (and perhaps some in between) and detailed cost figures. In this manner, tradeoffs can be performed and the most cost-effective system determined.

A number of methods have been developed by researchers to design optimal satellite constellations for continuous zonal or global coverage. In one method, multiple circular orbit satellites at the same altitude are placed in a single plane so as to create a *street of coverage* that is continuously viewed. The objective is then to determine analytically how many such streets (i.e., planes of satellites at the same inclination) are required to cover the zone of interest or the globe. In another method, satellites in common altitude and inclination orbits are distributed symmetrically and propagated ahead in time. Based on satellite positions at each time interval, the largest required coverage circle size over time is recorded. The orbital inclination and arrangement are then varied numerically to achieve the optimal constellation. These arrangements of symmetric, circular orbits are often referred to as *Walker constellations*, based on the contributions by J. G. Walker. A third class of optimal constellations involves the use of eccentric orbits with a common period and inclination to achieve single or multiple continuous global coverage using fewer satellites than are required with circular orbits. These constellations of symmetrical, elliptical orbits are commonly called *Draim constellations* after their developer, J. E. Draim. The next three sections will deal with these constellation optimization methods individually and will compare their results.

For circular orbits, the constellation optimization problem can be uncoupled from satellite altitude (h) and ground elevation angle (ε) considerations by using the Earth central angle radius of coverage (θ) as the primary independent variable. For constellations of T circular orbit satellites, the goal is to find the arrangement that requires the smallest value of θ and still achieves continuous zonal or global coverage. The constellation with the lowest required value of θ will allow the lowest operating altitude for a fixed value of ε. Conversely, if satellite altitude is fixed, the lower operating limits on ground elevation angle (ε) will be maximized. The value of the Earth central angle radius of coverage (θ), which is required for the constellation to achieve continuous zonal or global coverage, is regarded as a measure of efficiency of a constellation. The lower the value of θ for fixed T, the more efficient the constellation.

For elliptical orbits, the use of θ as the measure of efficiency of a constellation is not feasible, becausee it does not remain constant as the satellite moves around its orbit. Instead, the semimajor axis (a), eccentricity (e), and minimum required ground elevation angle (ε) serve to determine the efficiency of a constellation.

Optimal Satellite Constellations Using the Street of Coverage Method

The goal here is continuous, multiple, zonal, or global Earth coverage. It is assumed in this method that equal numbers of satellites are symmetrically distributed in equally inclined circular orbit planes at the same altitude. At least three satellites (depending on altitude and ground elevation angle) are placed per plane in order to form a continuous street of coverage to associate with each orbit plane. The orbit planes are located with ascending nodes symmetrically distributed around the Earth's equator. The nodal separation between adjacent planes is $2\pi/p$, where p is the number of orbit planes. An exception to this rule is for polar orbits, where it was found best to distribute the orbit planes over approximately 180 deg. Optimal constellations resulting from this method are described by the Earth central angle radius of coverage (θ), inclination, and $T = p \times s$, where T, the total number of satellites, is equal to the number of planes (p) times the number of satellites per plane (s).

A single street of coverage is depicted in Fig. 15.14. The coverage circles of the satellites in this orbit plane overlap so as to form a band or street of coverage

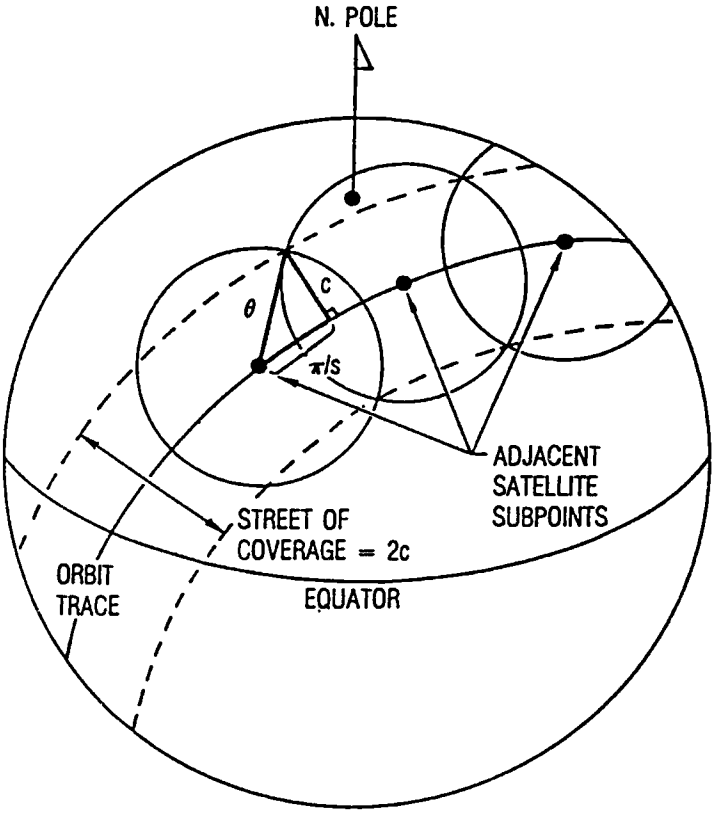


Fig. 15.14 Continuous street of coverage from a single orbit plane.

about the Earth that is continuously covered. The relation between the half-street width (c), the coverage circle size (θ), and the number of satellites in the plane (s) is given by

$$\cos\theta = (\cos c)\cos\frac{\pi}{s} \tag{15.13}$$

The problem then becomes one of determining the number of such streets that are required to cover the zone of interest as a function of orbital inclination. Luders[4] (also Luders and Ginsberg[5]) used this method and a computer search over orbit inclination to solve the continuous single zonal coverage problem. Rider[6] further pursued this method to develop an analytic, closed-form solution to the inclined orbit zonal coverage problem for multiple coverage.

Results from the streets of coverage technique indicate that, if the zone (a region between two latitude values on the Earth's surface) is in the low to mid-latitudes, the optimal constellation will consist of inclined orbit planes with nodes spaced evenly through 360 deg. A typical example of an optimal zonal coverage constellation is depicted in Fig. 15.15. In this example, two planes of eight satellites each, at an altitude of 900 km and inclined at 68.5 deg provide continuous single coverage to the zone between 50° and 60° latitude. Note that, because the constellation has north/south symmetry, both the northern and southern zones of latitude are covered.

For zonal coverage at high latitudes or any zone including the pole (including global coverage), researchers Beste[7] and Rider[8] found that the streets of coverage method using polar orbits with nodes spread over 180 deg were preferable. These polar orbits required fewer satellites at the same altitude than the inclined orbits did. An example of an optimal polar constellation for continuous global coverage is shown in Fig. 15.16. In this case, four planes of eight satellites each, at an altitude of 1100 km, provide continuous single global coverage. The four polar

Fig. 15.15 Optimal street of coverage constellation for continuous zonal coverage of latitudes 50–60° (two planes, eight satellites per plane, alt = 900 km, incl = 68.5 deg).

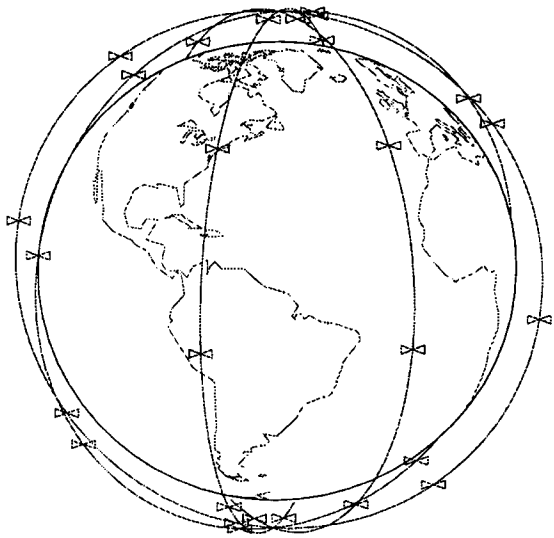


Fig. 15.16 Optimal street of coverage constellation for continuous global coverage with arbitrary interplane phasing (four polar planes separated by $\Delta\Omega = 45$ deg, eight satellites per plane, alt = 1100 km).

planes are separated by 45 deg in node, resulting in a symmetric constellation. Note that, by doubling the number of orbital planes, using polar orbits, and increasing the altitude somewhat, continuous global coverage instead of coverage of a small latitude band has been achieved. Because the street of coverage is the same regardless of the direction of rotation or relative phasing between the different orbit planes, these constellations are referred to as having arbitrary interplane phasing.

Researchers further noted that these optimal polar constellations had $2p$ interfaces between adjacent streets (see Fig. 15.16). Of these, $2p - 2$ were corotating; that is, the satellites were moving in the same direction. Only two interfaces are counterrotating with satellites moving in opposite directions. If the satellites in adjacent corotating planes are correctly phased, the coverage circles from one plane can be used to cover the cusps in the adjacent plane. Such an optimal phasing allows corotating planes to be spaced further apart than the simple half-street width would allow. The spacing of counterrotating planes remains bounded by the half-street width. The overall effect of optimally phasing the satellites between planes is to take advantage of these corotating interfaces. The result is that the polar planes are no longer evenly distributed about 180 deg, but rather a slightly larger value. The constellation is no longer symmetrical, but the resulting value of θ is reduced. As a result of optimally phasing the satellites in adjacent planes in the example of Fig. 15.16, the same single global coverage can be achieved by a nonsymmetric planar arrangement at an altitude of 900 km instead of 1100 km. This nonsymmetric, optimally phased constellation is shown in Fig. 15.17. Clearly, if the phasing between satellites in different planes can be stationkept (i.e., can maintain optimal phasing), then a significant improvement can be realized.

Adams and Rider[9] have tabulated optimal streets of coverage constellations for continuous global and various polar cap coverages for both arbitrarily and optimally phased polar arrangements. Multiple folds (from one- to fourfold) of

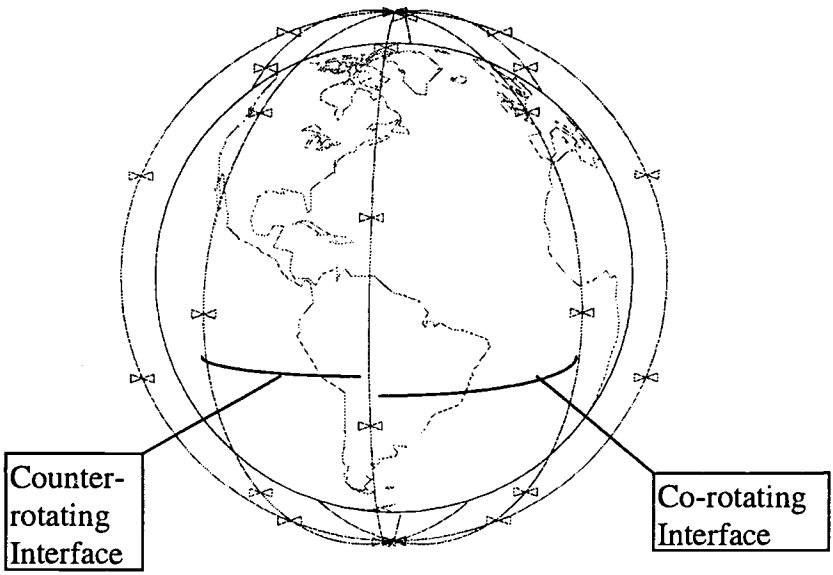


Fig. 15.17 Optimal street of coverage constellation for continuous global coverage with optimal interplane phasing (four polar planes separated by $\Delta\Omega = 47.4$ deg, eight satellites per plane, alt = 900 km).

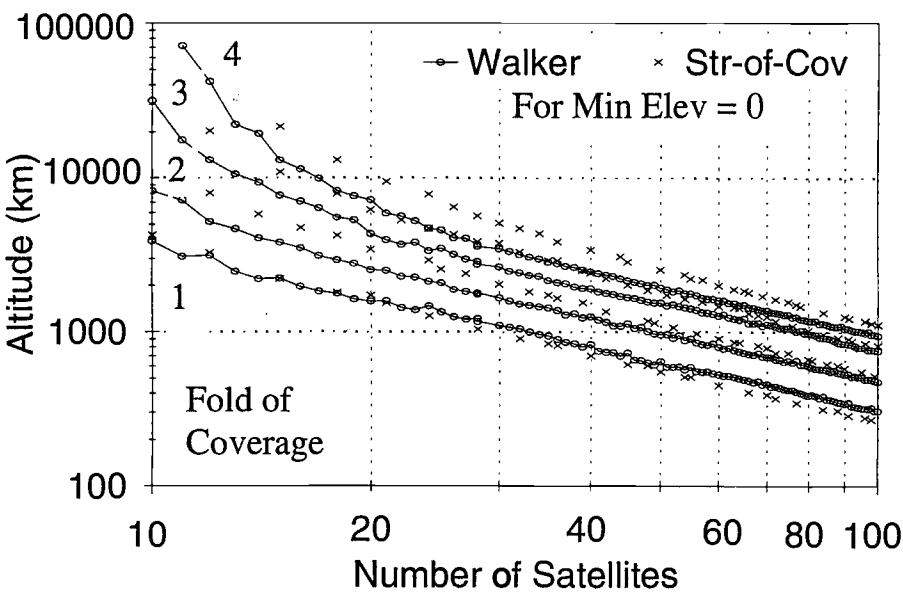


Fig. 15.18 Minimum circular orbit altitude for continuous global coverage.

coverage are examined for as many as 100 to 200 satellites. Later tables by these same authors investigate constellations of up to several thousand satellites. For as many as 100 satellites, the number of satellites required for continuous global coverage as a function of satellite altitude (h) for these optimally phased, nonsymmetric, polar constellations of satellites using the street of coverage method is shown in Fig. 15.18. As expected, the number of satellites increases steadily (although not always monotonically) as the satellite altitude decreases. Note that twofold coverage of the globe does not require twice as many satellites as onefold. Sometimes, an additional fold of coverage is available for a minor percentage increase in the number of satellites.

Optimal Satellite Constellations Using Walker's Method

Researchers such as Walker,[10,11,14–16,18] Mozhaev,[12,13] Ballard,[17] and Lang[19,20] have sought constellations of satellites in common-altitude, generally inclined, circular orbits that provide continuous global (single or multiple) coverage with a minimum of satellites. Only symmetric arrangements of satellites are considered.

Using Walker's notation, symmetric constellations of satellites can be described by the parameters $T/P/F$ and i, where

T = total number of satellites in the constellation
P = number of commonly inclined orbital planes
F = relative phasing parameter
i = common inclination for all satellites

In order to have a symmetric arrangement, the T/P satellites in a given orbital plane are equally spaced in central angle (phasing), and the P orbital planes are evenly spaced through 360 deg of right ascension of ascending node. The phasing parameter F relates the satellite positions in one orbital plane to those in an adjacent plane (i.e., interorbit phasing). The units of F are $360/T$ deg.

Table 15.1 Initial satellite locations for example 12/3/2 constellation

Satellite No.	Plane No.	RAAN, deg	Initial mean anomaly, deg
1	1	0	0
2	1	0	90
3	1	0	180
4	1	0	270
5	2	120	60
6	2	120	150
7	2	120	240
8	2	120	330
9	3	240	120
10	3	240	210
11	3	240	300
12	3	240	30

Fig. 15.19 Walker 12/3/2 constellation (three planes separated by $\Delta\Omega$ = 120 deg, four satellites per plane, alt = 1000 km, incl = 60 deg).

As an example, consider $T/P/F = 12/3/2$. The 12 satellites are located such that four satellites are evenly spaced in each of three orbit planes. The three orbit planes are 120 deg apart in right ascension of ascending node (RAAN). If there is a satellite at its ascending node (mean anomaly = 0) in orbit plane 1, then, in orbit plane 2 (adjacent plane to the east), a satellite will be at a mean anomaly of $0 + F \times 360/T = 60$ deg. For this example, the arrangement of all 12 satellites in terms of right ascension of ascending node and initial mean anomaly is given in Table 15.1. An illustration of this Walker 12/3/2 constellation with an inclination of $i = 60$ deg is shown in Fig. 15.19.

The parameters $T/P/F$ and the orbital inclination i are sufficient definition of a constellation to allow the determination of the Earth central angle radius of coverage θ required for any specified level of continuous global coverage. To accomplish this, Walker devised the circumcircle approach. Using this method, the satellite positions are computed at small time steps. At each time step, all combinations of three satellites are examined, and the radius of the spherical circle that contains them (the *circumcircle*) is determined. Figure 15.20 shows the circumcircle approach applied at a single time step. The largest circumcircle over all time intervals that does not include another satellite is equal to the size of the coverage footprint θ necessary to achieve continuous global coverage. For double or twofold coverage, the largest circumcircle that includes one satellite is sought, and so on for multiple folds of coverage.

To optimize the constellation, Walker numerically varied the orbital inclination so as to minimize θ. With this done, the optimal value of θ was obtained for a specific set of $T/P/F$. For a specified value of T total satellites, this procedure must be repeated for all values of the number of orbit planes P (i.e., all the factors of T) and for all values of the phasing parameter F (for each value of P, F can take a value of 1 through P) in order to determine the overall best set of $T/P/F$ and i.

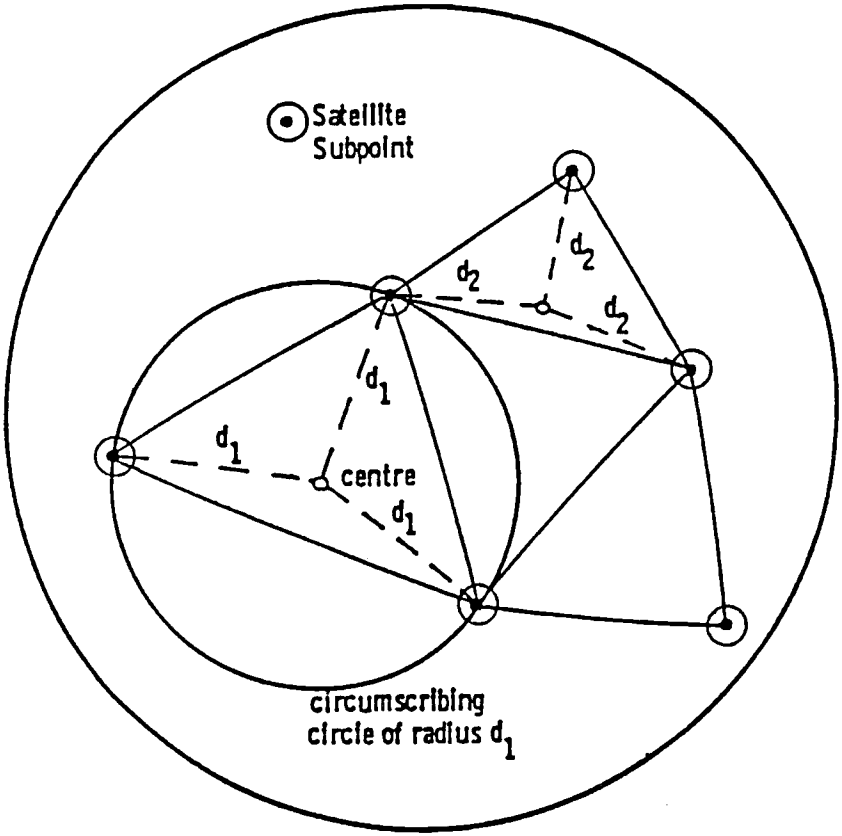


Fig. 15.20 Walker's circumcircle approach. At each instant in time the largest circumcircle defined by three satellite subpoints is determined.

Walker's circumcircle approach is numerically intensive because its accuracy depends on the size of the time step used. Additionally, as the number of satellites T increases, the number of combinations of three satellites used to define each circumcircle increases as $[T \times (T-1) \times (T-2)]$. For these reasons, this method is computationally demanding for large values of T. Walker was able to use this approach to optimize constellations of as many as 24 satellites for onefold through fourfold continuous global coverage. Lang[19] abandoned the circumcircle approach and used some symmetry simplifications to reduce the computational requirements for large constellations. In this manner, optimal Walker-type constellations have been determined for up to 100 satellites.

The resulting optimal Walker-type constellations are shown in Fig. 15.18 in terms of the number of satellites required to achieve the desired level of continuous global coverage as a function of satellite altitude. Optimal constellations for both the streets of coverage and Walker methods are included on this plot. Recall that the constellations from the streets of coverage technique involve optimally phased satellites in polar orbits spread over roughly 180 deg in node. Typically, the number of satellites per plane far exceeds the number of planes. The Walker technique

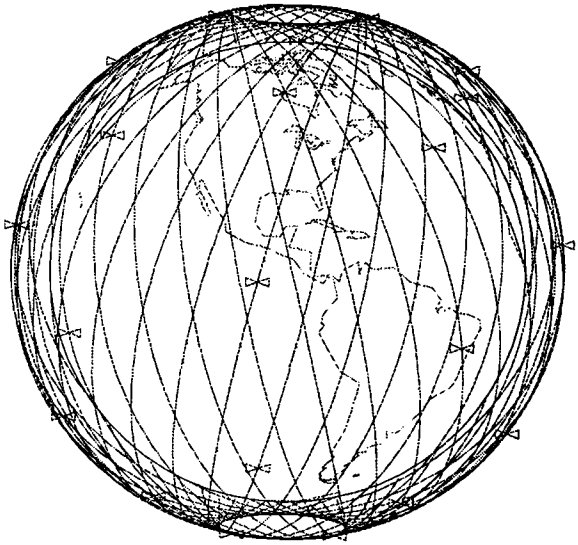


Fig. 15.21 Optimal Walker-type constellation for continuous global coverage ($T/P/F$ = 32/32/28) (32 planes separated by $\Delta\Omega$ = 11.25 deg, 1 satellite per plane, alt = 1048 km, incl = 76.9 deg).

produces symmetric constellations of inclined orbit planes with nodes distributed around 360 deg. Often, there is only one satellite per plane.

Figure 15.21 depicts the optimal 32-satellite Walker-type constellation for continuous global coverage. It employs 32 separate orbital planes inclined at 76.9 deg with one satellite per plane ($T/P/F = 32/32/28$). The orbital altitude is 1048 km to achieve continuous global coverage with a minimum elevation angle of 0 deg. This altitude is nearly 150 km higher than the corresponding optimal polar constellation (see Fig. 15.17), thereby making the polar constellation the more efficient. For an orbit planner, however, efficiency may not be the only consideration in selecting between such constellations. Despite its higher altitude, the Walker-type constellation has a lower inclination (76.9 instead of 90 deg), which might make its orbit easier to achieve, depending on the launch vehicle and launch site. On the other hand, planners might prefer the polar constellation because they can minimize their launch costs by lofting two, four, or even eight satellites together to the same orbital plane using a single launch vehicle. In contrast, this particular Walker-type constellation would require a separate launch to each of the 32 orbital planes. Less optimal (higher-altitude) Walker-type constellations are likely to be found that use fewer planes to achieve the same coverage. Clearly, arriving at the most cost-effective constellation often involves a detailed tradeoff analysis.

In comparing the optimal streets of coverage and Walker-type constellations as shown in Fig. 15.18, several conclusions can be drawn. For single continuous global coverage with 20 or fewer satellites, the symmetric, inclined Walker-type constellations are more efficient. For the same number of satellites, they offer continuous global coverage at a lower altitude (correspondingly lower θ). Conversely, at the same altitude, they can perform the same job with fewer satellites.

For single continuous global coverage with more than 20 satellites, the optimally phased, nonsymmetric polar constellations (via the streets of coverage method) are more efficient. For double or higher folds of continuous global coverage, the Walker-type constellations are always more efficient. In fact, in the region of 30 satellites, the inclined Walker-type constellations achieve fourfold coverage at altitudes for which the polar constellations cannot even achieve full threefold coverage.

Optimal Satellite Constellations Using Draim's Method

Earlier, we saw how two or three highly eccentric orbit (HEO) satellites could be arranged in a single groundtrack with apogee at the northernmost point so as to provide concentrated coverage over the northern hemisphere. Similarly, placing apogee at the southernmost point in the orbit would concentrate the coverage in the southern hemisphere. Using elliptical orbit satellite constellations with apogees located half in the north and half in the south presents an alternate way of obtaining global coverage. Draim[21-24] has used this technique to arrange satellites in high-altitude, moderately elliptical, moderately inclined orbits in a polyhedral fashion so as to obtain single and multiple continuous global coverage. In this manner, he has achieved one-, two-, three-, and fourfold continuous global coverage with 4, 6, 8, and 10 elliptical orbit satellites, respectively. This is one satellite fewer than would be required by the best circular orbit constellations even at the highest altitudes. The orbital elements of the 4, 6, 8, and 10 satellite constellations are shown in Table 15.2.

The values given for semimajor axis are the lowest values that yield the specified fold of continuous global coverage using a ground elevation angle ε of 0 deg. Note that these Draim constellations are all supersynchronous; that is, the orbital periods (26.49 h for onefold, 102 h for twofold, 272 h for threefold, and 568 h for fourfold) are all greater than 24 h. Since the eight- and ten-satellite constellations operate at such high altitudes, the severe lunar perturbations limit the practical value of these constellations.

15.3 Considerations in Selecting Satellite Constellations

The foregoing discussions have investigated the coverage offered by satellites in different types of orbits and methods for arranging multiple satellites in constellations so as to optimize their collective coverage. Although the coverage provided from a satellite constellation is a strong driver in the constellation selection process, it may not be the primary driver. Other factors that influence constellation selection will be addressed in this section.

Most often, the ultimate goal of satellite constellation selection is to achieve the desired coverage task with a minimum overall system cost. Although this objective is simple to state, the process involved can be quite complicated. Figure 15.22 illustrates a typical constellation selection process. The selection process pictured here is simplified and highlights the role played by the orbital constellation in the overall mission. The mission objectives usually dictate the coverage required of the system (e.g., regional, zonal, global, continuous, part-time) and certain sensor constraints (e.g., minimum elevation angle, lighting conditions, maximum slant range). These requirements and constraints become the inputs to the constellation

Table 15.2 Satellite orbital elements for Draim global coverage constellations

Sat. no.	Semimajor axis, km	Eccentricity	Inclination, deg	RAAN, deg	Arg of periaosis, deg	Mean anomaly, deg
	Four-satellite, onefold continuous global coverage constellation:					
1	45,033	0.263	31.3	0	270	0
2	45,033	0.263	31.3	90	90	270
3	45,033	0.263	31.3	180	270	180
4	45,033	0.263	31.3	270	90	90
	Six-satellite, twofold continuous global coverage constellation:					
1	110,630	0.233	27.5	0	270	0
2	110,630	0.233	27.5	60	90	300
3	110,630	0.233	27.5	120	270	240
4	110,630	0.233	27.5	180	90	180
5	110,630	0.233	27.5	240	270	120
6	110,630	0.233	27.5	300	90	60
	Eight-satellite, threefold continuous global coverage constellation:					
1	212,742	0.218	25.0	0	270	0
2	212,742	0.218	25.0	45	90	315
3	212,742	0.218	25.0	90	270	270
4	212,742	0.218	25.0	135	90	225
5	212,742	0.218	25.0	180	270	180
6	212,742	0.218	25.0	225	90	135
7	212,742	0.218	25.0	270	270	90
8	212,742	0.218	25.0	315	90	45
	Ten-satellite, fourfold continuous global coverage constellation:					
1	347,567	0.205	24.0	0	270	0
2	347,567	0.205	24.0	36	90	324
3	347,567	0.205	24.0	72	270	288
4	347,567	0.205	24.0	108	90	252
5	347,567	0.205	24.0	144	270	216
6	347,567	0.205	24.0	180	90	180
7	347,567	0.205	24.0	216	270	144
8	347,567	0.205	24.0	252	90	108
9	347,567	0.205	24.0	288	270	72
10	347,567	0.205	24.0	324	90	36

selection process. Often, these inputs themselves may be changed as the process continues, perhaps as a result of improved technology, but most often because certain requirements or constraints prove too costly.

These inputs to the selection process are first used to establish a candidate satellite constellation. The constellation may be selected from the various methods described earlier (e.g., streets of coverage, Walker-type, Draim). By whatever method, the number of satellites in the constellation and their altitude, inclination, and number of orbital planes are determined.

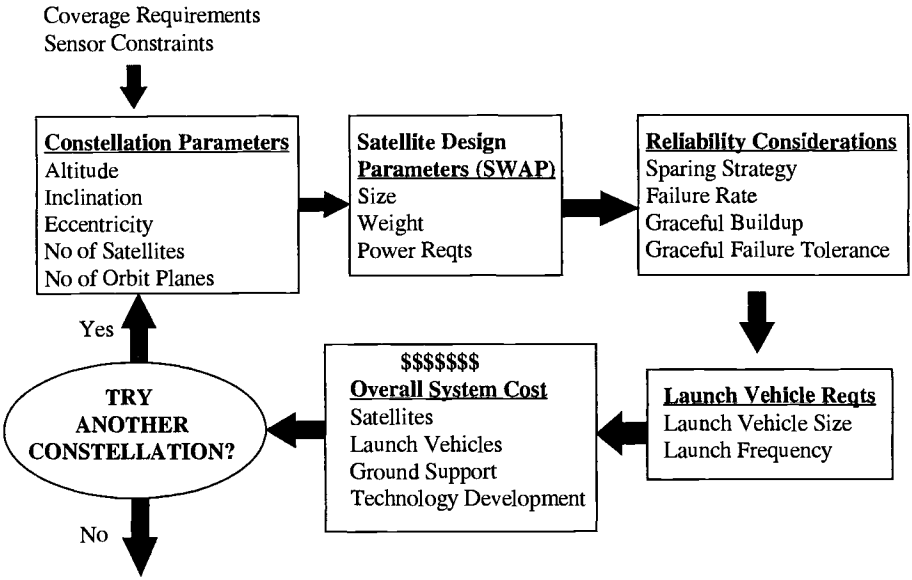


Fig. 15.22 Constellation selection process.

With the constellation defined, the next step is to derive the basic satellite design parameters. This step, often referred to as SWAP, determines the satellite's size, weight, and power requirements. Obviously, the sensor maximum range required from the previous step is a primary driver here.

Reliability considerations come next. The constellation is often rated on how well it performs during its buildup phase in orbit. That is, how useful is the constellation at satisfying mission goals between the time the first satellite is placed in orbit and the time the final satellite is launched? The buildup phase can be important, as evidenced by the GPS/NAVSTAR program, which placed its 24 satellites in orbit with an aggressive launch schedule spanning five years. During this period, satellites were located so as to provide maximum customer utility. Similarly, we seek a constellation whose performance degrades gracefully as satellite failures are experienced. Built-in failure protection can be provided by including on-orbit spares in the constellation. Usually, an on-orbit spare is provided for each orbital plane in the constellation. If a satellite fails, the spare is maneuvered by phasing so as to take the place of the failed satellite. This method favors constellations using fewer orbital planes. Another sparing option is to increase the constellation size so that twofold instead of onefold coverage is achieved. In this option, several satellite failures can usually be tolerated before the one fold coverage is affected. Although this method sounds extreme, there are many cases in which the tradeoff is favorable. As part of this step, the expected failure rate of the satellites is determined.

At this point in the process, the launch vehicle requirements are determined. The number of satellites and spares in the constellation, along with the anticipated failure replacement rate, will determine a launch rate required to support the buildup and replenishment phases of the program. This information, combined

with the weight of the satellite and the orbit involved (both orbit altitude and inclination play a part in this), helps to determine the size of the launch vehicle and the launch frequency required to support the mission. Sometimes, it is economical to fly several satellites on a single expensive launcher (remember that the satellites on a single launcher are generally destined for the same orbital plane) and, at other times, the best solution is for satellites to be launched individually by inexpensive launch vehicles.

With information at hand on the satellite design, launch vehicle type, and frequency of launch, the next step is to evaluate the total mission cost, including satellites, launch vehicles, ground support (e.g., uplink/downlink stations), and any technology development required (i.e., non-off-the-shelf components). The overall objective is to reduce this cost.

The final step in the process is to use the information gained in the previous steps to re-evaluate the satellite constellation parameters. The sensor performance might be a function of altitude. Would a lower orbit altitude help even though a larger constellation size would result? Perhaps raising the orbit altitude would reduce costs. Will a lower orbit inclination reduce the launch vehicle requirements? Would eccentric orbits help? Because of multiple satellites per launch vehicle or a satellite sparing philosophy, is it preferable to have fewer orbit planes? These are the types of questions often asked in the selection of the next candidate satellite constellation to be evaluated. After numerous iterations in this process, it is hoped that a constellation is arrived at that minimizes the overall system cost.

It is interesting to note that different groups of researchers often apply the constellation selection process to the same problem with differing results. Consider the problem of selecting a constellation of satellites for the purpose of continuous global communication coverage. Some groups claim that this problem is best solved by the use of a few large GEO satellites lofted by expensive launch vehicles. Others claim that the best solution is many (even hundreds) of smallsats launched into LEO on cheap (relatively speaking) launch vehicles. Still others claim that a moderate number of satellites in a MEO constellation is the right answer. The correct answer to this problem may well vary with time as the various orbit regimes (LEO, MEO, and GEO) deal with the problems of satellite crowding and space debris.

Despite various disagreements over which satellite constellations are best for which tasks, there is general agreement in many areas. Table 15.3 attempts to categorize the various satellite constellations and to link them to specific satellite functions or applications.

15.4 Nontypical Coverage Patterns

The previous discussions have assumed a simple conical sensor that views points on the Earth's surface. Sometimes, the sensor field of view is not simply a cone, or the objects of interest are not on the Earth's surface. This section will investigate several such situations and show how the earlier coverage analyses can be applied in these cases.

Satellite-to-Satellite Viewing

In this situation, the objective is to analyze the ability of one or more viewing satellites to cover a set of one or more target satellites. The target satellites are

Table 15.3 Typical satellite constellation applications

Constellation Type	Characteristics	Applications
1 to several LEO	Short sensor range	Earth resources
	Heavy satellites	Manned platforms
	Low ΔV reqts	Space Shuttle
	Re-entry/reuse	Space Station
	Revisit time	Surveillance (periodic)
		Weather (periodic)
10s to 100s LEO	Same as above	Contin global comm?
	Contin zonal covg	Space Defense?
	Contin global covg	
10's of MEO	Medium sensor range	Navigation
	Contin global covg	Contin global comm?
2–3 HEO	Contin N hem covg	Surveillance of N hem
	Highly elliptical	Contin comm in N hem
1 GEO	Contin regional covg	Contin regional comm
	Large sensor range	Weather (regional)
		Surveillance (regional)
3+ GEO	Contin global covg	Contin global comm

assumed to be in circular orbits at the same altitude. The problem is analogous to the Earth coverage problem, except that we are now trying to cover a sphere about the Earth at the target satellite altitude. The viewing geometry is shown in Fig. 15.23. Quite often, the viewing satellite is not allowed to look at the target through the Earth limb, so that a minimum graze height constraint for the line of sight is established. Define the following parameters:

h = viewing satellite altitude
h_t = target satellite altitude
h_g = minimum graze height for line of sight
r_e = Earth equatorial radius

Then the following equations can be used to find the Earth central angles of coverage for this viewing geometry:

$$\begin{aligned} \theta_A &= \cos^{-1}\left(\frac{r_e + h_g}{r_e + h}\right) \\ \theta_B &= \cos^{-1}\left(\frac{r_e + h_g}{r_e + h_t}\right) \\ \theta_1 &= \theta_A + \theta_B \\ \theta_2 &= \theta_A - \theta_B \end{aligned} \tag{15.14}$$

The Earth central angles θ_1 and θ_2 define the coverage circles for this satellite-to-satellite viewing geometry. The viewing satellite can cover the target satellite

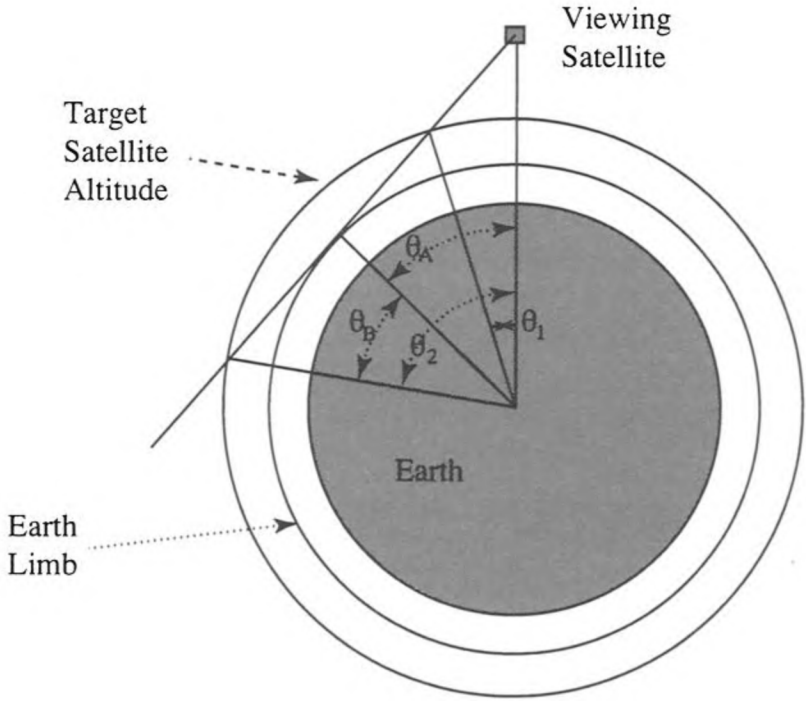


Fig. 15.23 Satellite-to-satellite viewing geometry.

sphere out to a central angle θ_1 from the viewing satellite's subpoint. These values of central angle can be used just as we did for the simple conical field of view sensor to the Earth's surface as depicted in Fig. 15.2. Equation (15.2) can be used to find the corresponding elevation angles as if the coverage were on the Earth's surface. In this manner, the coverage pattern can be projected from the target satellite sphere down onto the Earth's surface.

If simple line-of-sight visibility is the only requirement, then the coverage circle is defined by the central angle θ_1. If the viewing satellite is required to view the target against an Earth background, then the coverage circle extends only out to a central angle of θ_2 from the subpoint. This type of coverage is often referred to as *below-the-horizon*, or *BTH*, coverage. For both of these cases, the coverage pattern for each viewing satellite is simply a circular footprint about the satellite subpoint. The determination of optimal constellations of such satellites has already been dealt with earlier in this chapter.

If the viewing satellite is required to view the target against a space background, then two limiting coverage circles exist. This type of coverage is often referred to as *above-the-horizon*, or *ATH*, coverage. The outer limiting coverage circle is defined by the central angle θ_1 and the inner limiting coverage circle by θ_2. The coverage provided by a single satellite now has an inner and outer boundary. This is often referred to as a coverage pattern with a *nadir hole*. Instead of a simple circular coverage pattern projected onto the Earth, a pattern with a nadir hole appears as a doughnut-shaped, or annular, projection on the Earth as depicted in Fig. 15.24. The synthesis of optimal constellations of satellites using sensors with nadir holes will be addressed later in this section.

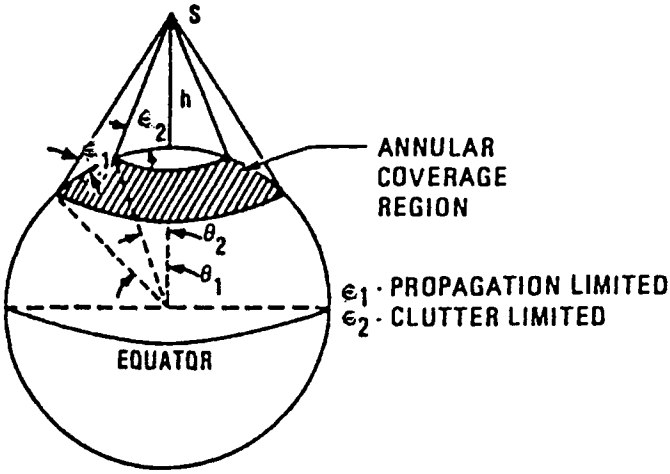


Fig. 15.24 Coverage footprint for satellite sensor with a nadir hole.

Satellite-Based Radar Sensors

The outer reach of a radar sensor illuminating the Earth is limited by atmospheric considerations. There is also a signal clutter problem that prevents the sensor from viewing targets too close to nadir. In this region, the sensor has trouble discerning the target from its background. Again, this is referred to as a *nadir hole*, and the coverage pattern is doughnut-shaped, or annular, as shown in Fig. 15.24. An excellent source of information on satellite-based radar systems is given in Cantafio.[25] Despite the nadir hole, optimal arrangments of these satellites can be determined so as to satisfy coverage requirements.

Optimal Constellations of Satellites with Nadir Hole Sensors

As seen earlier, satellite coverage patterns with nadir holes (as shown in Fig. 15.24) can arise in several situations. Satellite-based radar sensors and satellite-to-satellite sensors requiring a space background are two examples. The determination of optimal constellations of such satellite-borne sensors is complicated by the existence of the nadir hole. However, both the streets of coverage method and the Walker method of generating optimal constellations can be modified to account for the nadir hole.

For the streets of coverage approach, Rider[26] noted that the nadir hole of one satellite could be covered by the adjacent satellites in its street (or orbit plane). For this to happen without adjacent nadir holes overlapping requires the following three conditions to be met:

$$\begin{aligned} &N \geq 5 \\ &\theta_2 \leq \pi/N \\ &\theta_1 \geq \cos^{-1}(\cos\theta_2 \cos 2\pi/N) \end{aligned} \tag{15.15}$$

where N is the number of satellites per plane and θ_1 and θ_2 are the outer and inner limits, respectively, of coverage central angles. Once orbit planners have determined the coverage central angles within which the sensor must operate, the

above three conditions are used to determine constraints on the value of N, the number of satellites per plane. Having done this, the planner simply refers to the tables of optimized streets of coverage constellations contained in Adams and Rider,[9] seeking the best constellation that meets the constraints on N. Rider[27,28] has analyzed constellations of polar satellites for ATH coverage of orbital targets.

Similar methods exist for modifying Walker's method to account for the nadir hole. As already noted, the orbit planner can seek only Walker-type constellations that meet the constraints of Eq. (15.15) on $N = T/P$, the number of satellites per plane. This approach excludes all constellations with only one satellite (or a few) per plane, which are often the favored Walker-type arrangements.

Another option is to use the existing tabulated Walker-type constellations (see Refs. 14, 18, and 19) but for a fold of coverage that is one higher than required. For example, if onefold, continuous global coverage is desired, the planner would seek a constellation that offers twofold, continuous global coverage. In this manner, the nadir hole can be covered by another satellite. For this method, it is also necessary to require that the value of $D_{\min}$ (Walker's notation for the closest approach central angle between any two satellites) for a constellation is larger than the central angle θ_1, so that nadir holes cannot overlap. The value of $D_{\min}$ is often tabulated with the Walker constellation or can be computed separately.

Probably the best way to account for the nadir hole in Walker's method is to include the hole in the numerical iteration that solves for the optimal arrangements. The approach used by Lang[19] lends itself easily to this modification.

With these modifications, both the streets of coverage and the Walker approaches to optimal constellation synthesis can accommodate the coverage footprint that includes a nadir hole. Existing results for sensors without nadir holes can usually be used to generate optimal constellations for sensors with nadir holes.

References

[1]Hanson, J. M., Evans, M. J., and Turner, R. E., "Designing Good Partial Coverage Satellite Constellations," AIAA Paper 90-2901-CP, AIAA/AAS Astrodynamics Conference, Portland, OR, Aug 20–22,1990.

[2]Hanson, J. M., and Higgins, W. B., "Designing Good Geosynchronous Constellations," *The Journal of the Astronautical Sciences,* Vol. 38, No. 2, April–June 1990, pp. 143–159.

[3]Draim, J. E., and Kacena, T. J., "Populating the Abyss-Investigating More Efficient Orbits," 6th AIAA Conference on Small Satellites, Utah State University, Logan, UT, Sept. 21–24, 1992.

[4]Luders, R. D., "Satellite Networks for Continuous Zonal Coverage," *American Rocket Society Journal,* Vol. 31, Feb. 1961, pp. 179–184.

[5]Luders, R. D., and Ginsberg, L. J., "Continuous Zonal Coverage—A Generalized Analysis," AIAA Paper 74–842, AIAA Mechanics and Control of Flight Conference, Anaheim, CA, Aug 5–9, 1974.

[6]Rider, L., "Analytic Design of Satellite Constellations for Zonal Earth Coverage Using Inclined Circular Orbits," *The Journal of the Astronautical Sciences,* Vol. 34, No.1, Jan.–March 1986, pp. 31–64.

[7]Beste, D. C., "Design of Satellite Constellations for Optimal Continuous Coverage," *IEEE Transactions on Aerospace and Electronics Systems,* May 1978.

[8]Rider, L., "Optimized Polar Orbit Constellations for Redundant Earth Coverage," *The Journal of the Astronautical Sciences,* Vol. 33, April–June 1985, pp. 147–161.

[9]Adams, W. S., and Rider, L., "Circular Polar Constellations Providing Continuous Single or Multiple Coverage Above a Specified Latitude," *The Journal of the Astronautical Sciences,* Vol. 35, No.2, April–June 1987, pp. 155–192.

[10]Walker, J. G., "Circular Orbit Patterns Providing Continuous Whole Earth Coverage," Royal Aircraft Establishment Technical Report 70211, Nov. 1970.

[11]Walker, J. G., "Some Circular Orbit Patterns Providing Continuous Whole Earth Coverage," *Journal of the British Interplanetary Society,* Vol. 24, pp. 369–384, 1971.

[12]Mozhaev, G. V., "The Problem of Continuous Earth Coverage and Kinematically Regular Satellite Networks, I," *Kosmicheskie Issledovaniya,* Vol. 10, No. 6, Nov.–Dec. 1972, pp. 833–840.

[13]Mozhaev, G. V., "The Problem of Continuous Earth Coverage and Kinematically Regular Satellite Networks, II," *Kosmicheskie Issledovaniya,* Vol. 11, No.1, Jan.–Feb. 1973; pp. 59–69 trans. *Cosmic Research,* Vol. 11, No.1, Jan.–Feb. 1973, pp. 52–61.

[14]Walker, J. G., "Continuous Whole Earth Coverage by Circular Orbit Satellites," Royal Aircraft Establishment Technical Memorandum Space 194, April 1973.

[15]Walker, J. G., "Continuous Whole Earth Coverage by Circular Orbit Satellite Patterns," Royal Aircraft Esablishment Technical Report 77044, March 1977.

[16]Walker, J. G., "Satellite Patterns for Continuous Multiple Whole- Earth Coverage," Maritime and Aeronautical Satellite Communication and Navigation, IEE Conference Publication 160, March 1978, pp. 119–122.

[17]Ballard, A. H., "Rosette Constellations of Earth Satellites," *IEEE Transactions on Aerospace and Electronic Systems,* Vol. AES-16, No. 5, Sept. 1980, pp. 656–673.

[18]Walker, J. G., "Coverage Predictions and Selection Criteria for Satellite Constellations," Royal Aircraft Establishment Technical Report 82116, Dec. 1982.

[19]Lang, T. J., "Symmetric Circular Orbit Satellite Constellations for Continuous Global Coverage," AAS Paper 87-499, AAS/AIAA Astrodynamics Specialist Conference, Kalispell, MN, Aug. 10–13, 1987.

[20]Lang, T. J., "Optimal Low Earth Orbit Constellations for Continuous Global Coverage," AAS Paper 93–597, AAS/AIAA Astrodynamics Specialist Conference, Victoria, BC, Canada, Aug. 16–19, 1993.

[21]Draim, J. E., "Three- and Four-Satellite Continuous Coverage Constellations," *Journal of Guidance, Control, and Dynamics,* Vol. 6, Nov.–Dec. 1985, pp. 725–730.

[22]Draim, J. E., "A Common Period Four-Satellite Continuous Global Coverage Constellation," AIAA Preprint 86-2066-CP, AIAA/AAS Astrodynamics Conference, Williamsburg, VA, Aug. 18–20, 1986.

[23]Draim, J .E., "A Six Satellite Continuous Global Double Coverage Constellation," AAS Paper 87-497, AAS/AIAA Astrodynamics Specialist Conference, Kalispell, MN, Aug. 10–13, 1987.

[24]Draim, J. E., "Continuous Global N-Tuple Coverage with (2N + 2) Satellites," *Journal of Guidance, Control, and Dynamics,* Jan.–Feb. 1991.

[25]Cantafio, L. J., et al., "Space-Based Radar Handbook," Artec House, Inc., 1989.

[26]Rider, L., "Nadir Hole-Fill by Adjacent Satellites in a Single Orbit," Technical Note, *Journal of the Astronautical Sciences,* Vol. 28, No.3, July–Sept. 1980, pp. 299–305.

[27]Rider, L., "Optimal Orbital Constellations for Global Viewing of Targets Against a Space Background," *Optical Engineering,* March–April 1980, Vol. 19, No. 2, pp. 219–223.

[28]Rider, L., "Design of Low to Medium Altitude Surveillance Systems Providing Continuous Multiple Above-the-Horizon Viewing," *Optical Engineering,* Jan. 1989, Vol. 28, No.1, pp. 25–29.

Problems

15.1. A Comsat is capable of communicating to users on the ground down to a minimum elevation angle of $\varepsilon = 0$ deg. For two values of circular orbit altitude,

$$h = 35786 \text{ km (GEO)}$$

$$h = 1000 \text{ km (LEO)}$$

compute the following parameters:

a) The angular size (α) of the satellite antenna field of view required to cover the Earth down to $\varepsilon = 0$.

b) The Earth central angle (θ) of coverage provide by each satellite.

c) The maximum value of slant range (ρ) required to cover the Earth.

15.2. A constellation of such Comsats is envisioned to provide onefold, continuous global coverage. At GEO, this can be accomplished by a Walker constellation of five satellites in five orbital planes (Walker notation 5/5/1), inclined at 44 deg to the equator.

a) What is the separation in right ascension of ascending node (RAAN) between the orbit planes?

b) When the satellite in orbit plane 1 is at its ascending node (northbound), how far past its ascending node (measured in Earth central angle) is the satellite in orbit plane 2 (the orbit plane immediately east of it)?

15.3. For problem 15.2, a streets of coverage approach to the continuous global coverage task at GEO altitude requires a minimum constellation of six satellites. The constellation consists of two polar ($i = 90$ degrees) planes, each containing three satellites. The orbit planes are separated by 90 deg in RAAN (or by 104.5 deg if the satellites in one plane are optimally phased with respect to the other plane).

a) What is the half-width size (in degrees of Earth central angle) of the street of coverage provided by a single plane of the three satellites?

b) Why does the streets of coverage method not permit two satellites per plane?

15.4. To provide the same onefold continuous global coverage using a constellation of Comsats at an altitude of $h = 1000$ km clearly requires many more satellites. Referring to Fig. 15.18, the constellations with the lowest number of satellites below the $h = 1000$ km line are a 32 satellite street of coverage constellation (specifically, four polar planes of eight satellites each) and a 33-satellite Walker-type constellation (specifically, a Walker notation 33/33/29 arrangement, inclined at 77 deg).

a) Find the angular size (α) of the satellite antenna field of view required to cover the Earth down to $\varepsilon = 0$ at $h = 1000$ km.

b) Find the Earth central angle (θ) of coverage provided by each satellite at this altitude.

c) Find the maximum value of slant range (ρ) required to cover the Earth at this altitude.

d) For the street of coverage constellation at $h = 1000$ km, what is the half-width size of the street of coverage for a single plane of satellites?

e) What is the separation in RAAN between adjacent planes of the Walker-type constellation?

15.5. For the Comsat, the antenna size and solar array power required are a strong function of the maximum slant range (ρ). As a result, the satellite weight required to perform the mission is approximately proportional to $\rho^{0.5}$. If the LEO Comsat weighs 400 kg, how much will the GEO Comsat weigh?

15.6. Consider a simple cost tradeoff study for the four constellation options derived above. Our system cost model will include only the cost of the satellites and the launch vehicles required to orbit them. Assume that a satellite's cost is simply its weight (in kg) times $80,000. Four different launch vehicles are available (LVSM, LVMD, LVLG, and LVXL), ranging in size from small to extra large as implied by their designations. The two smaller launch vehicles can be used to boost satellites to LEO, whereas the two larger vehicles are used for GEO. The maximum payload weights for the launch vehicles to the mission orbits are shown in the table in this problem, along with the per launch cost of each vehicle. Note that the payload weight is lower to the polar orbits than to the corresponding Walker orbits. This reflects the fact that the closer a launch is to due east, the more advantage it can take of the Earth's rotational rate.

Finish filling out the table in this problem. For each mission, compute the total cost of all satellites. Based on the launch vehicle payload weight to each mission orbit, compute the number of launches required to fill the constellation. Remember, a single launch can carry one or more satellites but only to a single orbit plane. With the number of launches and the cost per launch (given), the launch vehicle total cost can be determined. What is the most cost-effective approach for this Comsat system? Which satellite constellation? Which launch vehicle?

Typical satellite constellation applications

	Satellite Constellation			
	LEO		GEO	
	Walker 33 sats 33 planes	Street of coverage, 32 sats, 8 planes	Walker 5 sats 5 planes	Street of coverage, 6 sats, 2 planes
Satellite costs				
Sat wt (kg)	400	400	?	?
Sat cost ($M/kg)	0.08	0.08	0.08	0.08
No. of sats	33	32	5	6
Total sat cost ($M)	1056	1024	?	?
Launch vehicle costs				
LVXS				
Payload capacity (kg)	450	400	—	—
LV cost ($M/launch)	10	10	—	—
No. of launches req'd	?	?	—	—
Total LV cost ($M)	?	?	—	—
LVMD				
Payload capacity (kg)	4000	3500	—	—

Typical satellite constellation applications (*Cont.*).

	Satellite Constellation			
	LEO		GEO	
	Walker 33 sats 33 planes	Street of coverage, 32 sats, 8 planes	Walker 5 sats 5 planes	Street of coverage, 6 sats, 2 planes
LV cost ($M/launch)	60	60	—	—
No. of launches req'd	?	?	—	—
Total LV cost ($M)	?	?	—	—
LVLG				
Payload capacity (kg)	—	—	1450	1350
LV cost ($M/launch)	—	—	80	80
No. of launches req'd	—	—	?	?
LVXL				
Payload capacity (kg)	—	—	4500	4300
LV cost ($M/launch)	—	—	280	280
No. of launches req'd	—	—	?	?
Total LV cost ($M)	—	—	?	?
Total system cost				
sat + launch vehicle	?	?	?	?

Selected Solutions

15.1. a) From Eq. (15.1),

$$\sin\alpha = \frac{\cos\varepsilon}{1 + h/r_e}$$

for $\varepsilon = 0$ deg, $h = 35786$ km, $r_e = 6378$ km, we get

$$\alpha = 8.7 \text{ deg}$$

b) From Eq. (15.1),

$$\cos(\theta + \varepsilon) = \sin\alpha$$

$$\text{or} \quad \theta + \varepsilon + \alpha = 90 \text{ deg}$$

for $\varepsilon = 0$ deg, $\alpha = 8.7$ deg, we get

$$\theta = 81.3 \text{ deg}$$

c) From Eq. (15.4),

$$\rho^2 = r_e^2 + (r_e + h)^2 - 2r_e(r_e + h)\cos\theta$$

solving for ρ yields

$$\rho = 41{,}679 \text{ km}$$

15.2. a) $\Delta\text{RAAN} = 360/P = 360/5 = 72$ deg
b) ΔMean anomaly $= F \times 360/T = 1 \times 360/5 = 72$ deg

15.3. a) From Eq. (15.13),

$$\cos\theta = (\cos c)\cos\pi/s$$

for $\theta = 81.3$ deg, $s =$ three sats/plane, solving for the half-street width c gives

$$c = 72.4 \text{ deg}$$

b) Since $\theta < 90$ deg for any satellite viewing the Earth's surface, it takes at least three satellites to give a street of coverage.

15.4. a) From Eq. (15.1),

$$\sin\alpha = \frac{\cos\varepsilon}{1 + h/r_e}$$

for $\varepsilon = 0$ deg, $h = 1000$ km, $r_e = 6378$ km, we get

$$\alpha = 59.8 \text{ deg}$$

b) From Eq. (15.1),

$$\cos(\theta + \varepsilon) = \sin\alpha$$

$$\text{or} \quad \theta + \varepsilon + \alpha = 90 \text{ deg}$$

for $\varepsilon = 0$ deg, $\alpha = 59.8$ deg, we get

$$\theta = 30.2 \text{ deg}$$

c) From Eq. (15.4),

$$\rho^2 = r_e^2 + (r_e + h)^2 - 2r_e(r_e + h)\cos\theta$$

solving for ρ yields

$$\rho = 3708 \text{ km}$$

d) From Eq. (15.13),

$$\cos\theta = (\cos c)\cos\pi/s$$

for $\theta = 30.2$ deg, $s = 8$ sats/plane, solving for the half–street width c gives

$$c = 20.7 \text{ deg}$$

e) $\Delta\text{RAAN} = 360/P = 360/33 = 10.9$ deg

15.5.

$$\frac{W_{\text{GEO}}}{W_{\text{LEO}}} = \frac{\rho_{\text{GEO}}^{0.5}}{\rho_{\text{LEO}}^{0.5}}$$

so that, for a LEO weight of 400 kg, the GEO weight will be

$$W_{\text{GEO}} = 1341 \text{ kg}$$

15.6. The completed table follows.

	Satellite Constellation			
	LEO		GEO	
	Walker 33 sats 33 planes	Street of coverage, 32 sats, 8 planes	Walker 5 sats 5 planes	Street of coverage, 6 sats, 2 planes
Satellite costs				
Sat wt (kg)	400	400	1341	1341
Sat cost ($M/kg)	0.08	0.08	0.08	0.08
No. of sats	33	32	5	6
Total sat cost ($M)	**1056**	**1024**	**536**	**644**
Launch vehicle costs				
LVXS				
Payload capacity (kg)	450	400	—	—
LV cost ($M/launch)	10	10	—	—
No. of launches req'd	33	32	—	—
Total LV cost ($M)	**330**	320	—	—
LVMD				
Payload capacity (kg)	4000	3500	—	—
LV cost ($M/launch)	60	60	—	—
No. of launches req'd	33	4	—	—
Total LV cost ($M)	1980	**240**	—	—
LVLG				
Payload capacity (kg)	—	—	1450	1350
LV cost ($M/launch)	—	—	80	80
No. of launches req'd	—	—	5	6
Total LV cost ($M)	—	—	**400**	**480**
LVXL				
Payload capacity (kg)	—	—	4500	4300
LV cost ($M/launch)	—	—	280	280
No. of launches req'd	—	—	5	2
Total LV cost ($M)	—	—	1400	560
Total system cost				
sat + launch vehicle	**1386**	**1264**	**936**	**1124**

The most cost-effective Comsat approach in this example is the five-satellite Walker constellation, launched singly on the large launch vehicle (LVLG). Note that the solutions are fairly close and that changing some of the launch vehicle capabilities or costs could significantly affect the answer.

INDEX

Acceleration, 5, 6
  centripetal, 6
  Coriolis, 6
Aeroassisted orbit transfer, 229
Altitude:
  geocentric, 71, 72
  geodetic, 71, 72
Angular momentum, 24
Apsidal rotation rate, 217, 218
Atmospheric density models:
  ARDC 1959, 228
  Jacchia 1964, 228
  U.S. Standard Atmosphere of 1962, 228
Auxiliary circle, 38, 39

Ballast, 118, 119
Ballistic coefficient, 227, 229, 230
Bi-elliptic transfer, 96–99

Cassini-Huygens, 296–298
Conversion:
  geodetic latitude, altitude, and longitude to x, y, z, 75
  perigee/apogee altitudes to perigee/apogee radii, 77–79
  perigee/apogee radii to perigee/apogee altitudes, 76–77
Coordinate systems, 11
  azimuth-elevation, 15
  Earth-centered (ECI), 11–13
  Earth-relative spherical (LDBARV), 70
  geographic, 14
  inertial, 5
  heliocentric, 11, 12
  rectangular, 62, 68, 69
  spherical (ADBARV), 67–70
Critical inclination, 217

Deep resonance, 219–220
Drag coefficient, 227–228

Eclipse:
  Earth, 248–250
  geosynchronous, 250–252
  sun angle, 248–260
Element sets, 30

Energy, 23, 87
  conservation of, 23
  excess, 118
  kinetic, 24, 87
  mechanical (total), 23, 87
  potential, 24, 87
Equilibrium longitude points, 187, 205

Finite burn transfer, 126–130
Fixed impulse transfer, 117–124
f and g expression, 59
  universal variables, 56–59

Galileo, 270–273, 294–295
General perturbation:
  Gaussian method, 198
  Lagrangian method, 198
  osculating orbit, 198
  Poisson's bracket, 201
  variation of parameters, 198
Geodetic, 71
  altitude, 71–75
  conversion to x, y, z, 75
  horizontal, 71
Geometric dilution of precision (GDOP), 255
Geosynchronous orbit, 221, 225
  definition, 307
  inclination perturbation, 221
  libration, 220, 221
  stable inclination, 233–236
Gravity:
  assist, 278, 279
  constant, 21
  harmonics, 203–207
  loss, 126–128
  sphere of gravitation, 274–275
  sphere of influence, 275, 276
Ground trace:
  characteristics of, 254, 255
  for Molniya type orbit, 256–258
  for navigation system, 255–257

Hohmann transfer, 94, 120–123, 135

Impulse, 7
  specific, 8

J_2 effects and frozen orbit, 220, 221

Kepler, 2
Kepler's equation, 27
  method of successive approximation, 40
  number of iterations required, 41, 42
  solving, 40
  starting value, 41
  starting value error, 43, 45
Kepler's problem, 35

Laguerre:
  root-finding method, 41
  solution of Kepler's equation, 43
Lambert's problems, 144
Launch azimuth, 241
Launch window, 124–126, 241, 252
Longitude stationkeeping, 186, 221
Low thrust transfer, 335

Mars Odyssey, 298
Mauna Kea, 295
Method of Averaging, 209–210
Modified Julian date, 71
Momentum balance, 7, 127

Newton, 2–4
  laws of, 4–5
N-impulse maneuver, 117

Oblateness (J_2), 207, 215
Offloading propellant, 118, 119
Optimal Low-Thrust Orbit Transfer:
  Continuous constant acceleration, 372, 382
  Edelbaum's analysis, 335–344
  equinoctial element formulation, 354
  Euler-Lagrange equations, 345, 370
  Hamiltonian, 345, 370, 379, 382
  LEO to GEO-averaging, 386
  LEO to GEO-Edelbaum, 350
  LEO to GEO-minimum time, 373
  LEO to GEO-optimized
    departure/arrival, 377
  optimal control theory formulation, 344–350
    Partial derivatives, 400–409
    Poisson brackets, 365, 368
    Pontryagin's maximum principle, 393
    variable specific impulse, 389
Orbit:
  elements of 28–31
  frozen, 259–263
  geometry, 25–26
  Keplerian, 21
  lifetime, 230, 334
  parameters, 25
  period, 27
  posigrade, 31
  resonance, 232
  retrograde, 31
  systems of, 241, 429, 445
  velocity, 27, 85
Orbital coverage:
  earth central angle, 414
  optimal satellite constellation, 429, 431, 434, 437, 445
  satellite-to-satellite viewing, 442–444
  single eccentric satellite, 421
  single GEO satellite, 417
  single LEO satellite, 413
  single MEO satellite, 425
  single satellite geometry, 411–413
  street of coverage, 431
  Walker notation, 435
Orbit perturbation:
  atmospheric drag, 189, 227, 228
  Earth gravity harmonics, 186, 204, 215, 221
  lunisolar gravitational attraction, 187, 222
  mutual gravitational attraction, 190, 232
  radiation pressure effect, 188, 223
  tidal friction effect, 190, 230
Orbit transfer, 84, 117, 335
  between coplanar ellipses, 107–109
  bi-elliptic, 96, 137
  bi-elliptic with split-plane change, 107, 151
  general three-impulse plane change, 103–104
  Hohmann, 94, 120–123, 135
  Hohmann with split-plane change, 104–107
  modified Hohmann, 150
  radial, 168
  restricted three-impulse plane change, 99
  semitangential, 137, 141
  single impulse, 89

Planetary data, 267–269
Plane change, 99–108, 119–122
Potential theory:
  Legendre polynomials, 203, 205
  spheroid, spheroid of revolution, 203, 204
  WGS84 model (4×4), 204, 205
Poynting-Robertson force, 188

Regional navigation system, 255, 403
Rendezvous:
  co-elliptic, 162, 164
  optimal multiple-impulse, 178
  phasing, 135
  solution to linearized equations, 159
  terminal, 155
  three-dimensional, 150
  two-impulse maneuver, 158
Rocket equation, 7, 117, 120, 128, 381

Secular nodal regression rate, 215, 216
Shuttle orbit, 119–124
Solar flux index, 190, 192, 228
Solar sailing, 226

Solid propellant motor, 117
Space debris:
  collision avoidance, 330–332
  collision hazard, 320
  density of, 310
  detection of, 306
  geosynchronous, 307
  lifetime, 327
  low Earth orbit, 302
  meteoroid, 308
  shielding, 329–330
Special perturbation:
  Adams-Moulton, 197
  Cowell's method, 196
  Encke's method, 197
  Gauss-Jackson, 197
  Runge-Kutta, 197
Sun synchronism, 250–254
Sun-synchronous orbit, 218

Tesseral (J_2) harmonics, 186, 204–206
Time systems, 17
  atomic time, 20
  ephemeris time, 20
  solar time, 19
Trace repetition parameter, 256, 415
Trajectories:
  escape, 89, 281
  interplanetary, 265, 287
  lunar, 265, 279
  modification, 118
Transformation:
  modified classical to Cartesian, 62
  rectangular to modified classical, 66
  rectangular to spherical, 68
  spherical to rectangular, 69
  Sundman, 55
Two-body problem:
  general, 21, 22
  restricted, 22, 280

Universal approach:
  general, 55
  summary, 59
  using Laguerre's method, 58
  variable, 56
Upper-stage vehicle, 117

Zonal (J_2) harmonics, 186, 207