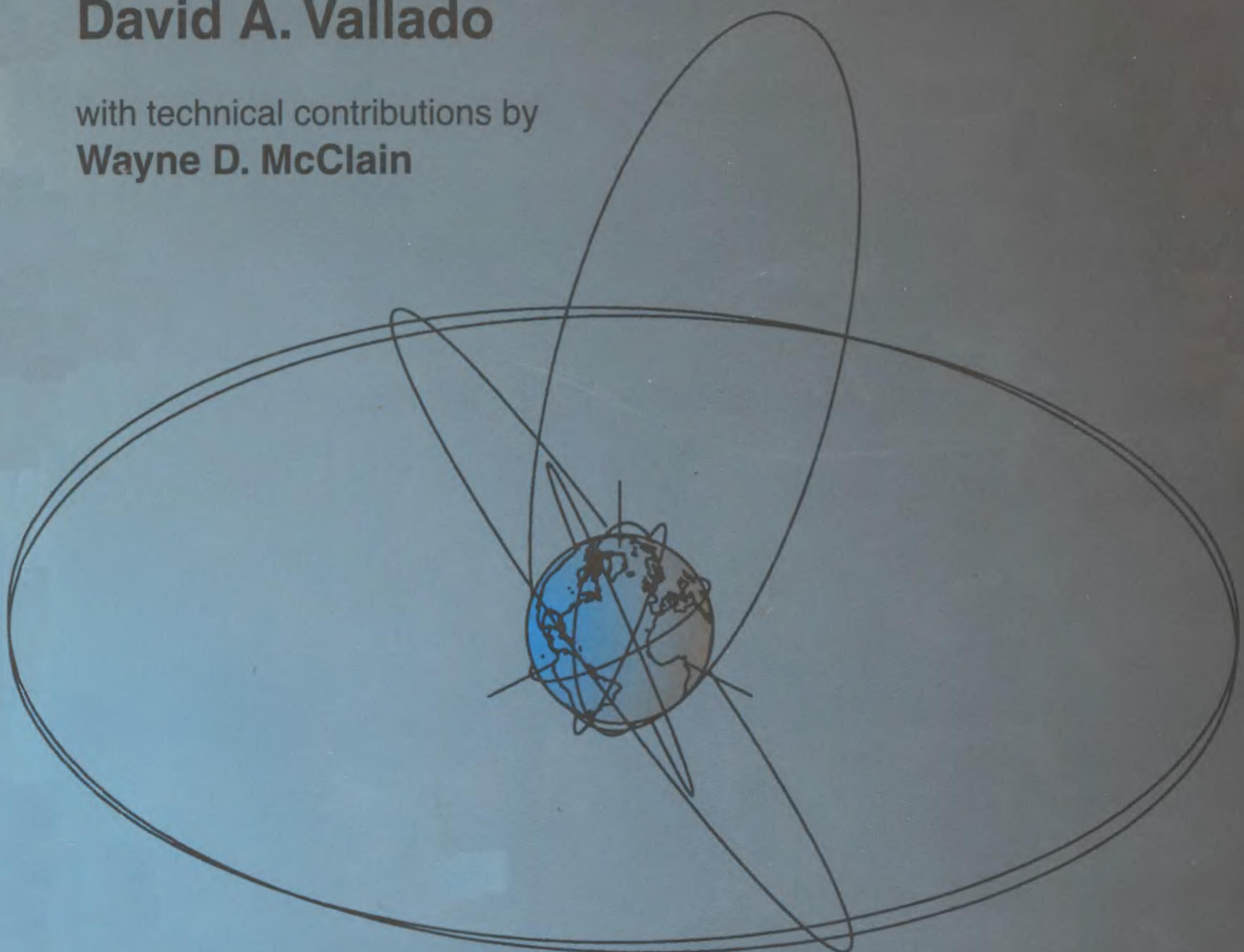


FUNDAMENTALS of ASTRODYNAMICS and APPLICATIONS

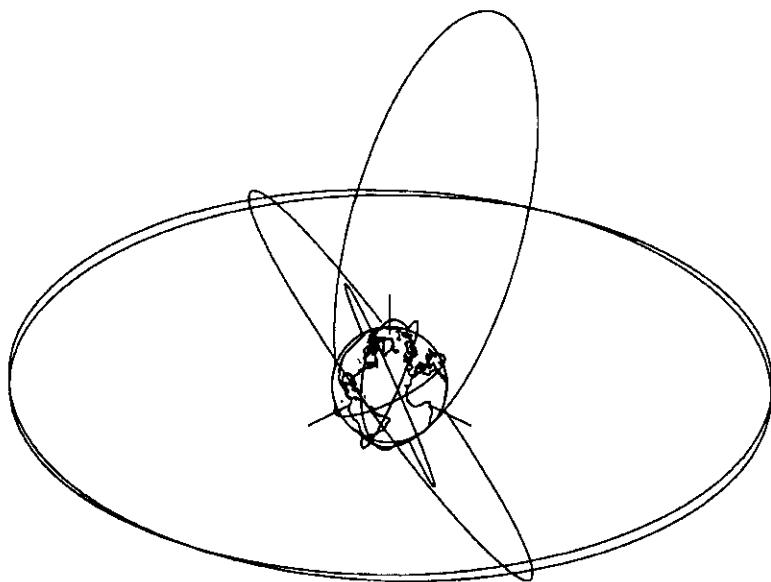
David A. Vallado

with technical contributions by
Wayne D. McClain



Space Technology Series

FUNDAMENTALS of ASTRODYNAMICS and APPLICATIONS



Space Technology Series

This book is published as part of the Space Technology Series, a cooperative activity of the United States Department of Defense and National Aeronautics and Space Administration

Wiley J. Larson
Series Editor

From McGraw-Hill: (mholland@mcgraw-hill.com)

Understanding Space: An Introduction to Astronautics by Sellers

Space Propulsion Analysis and Design by Humble, Henry, and Larson

Cost-Effective Space Mission Operations by Boden and Larson

Fundamentals of Astrodynamics and Applications by Vallado

(valladod@plk.af.mil or wmcclain@draper.com)

From Kluwer and Microcosm Publishers: (deGeus@WKAP.nl)

Space Mission Analysis and Design, Second Edition by Larson and Wertz

Spacecraft Structures and Mechanisms: From Concept to Launch by Sarafin

Reducing Costs of Space Missions by Wertz and Larson

Future Books in the Series

Modeling and Simulation: An Integrated Approach to Development and Operation
by Cloud and Rainey

Human Space Mission Analysis and Design by Connally, Giffen, Griffen, and Larson

FUNDAMENTALS of ASTRODYNAMICS and APPLICATIONS

David A. Vallado

United States Air Force Phillips Laboratory

with technical contributions by

Wayne D. McClain

Charles Stark Draper Laboratory

Series Editor

Wiley J. Larson

United States Air Force Academy

*This book is published as part of the Space Technology Series,
a cooperative activity of the United States Department of Defense
and the National Aeronautics and Space Administration.*

The McGraw-Hill Companies, Inc.
College Custom Series

New York St. Louis San Francisco Auckland Bogotá
Caracas Lisbon London Madrid Mexico Milan Montreal
New Delhi Paris San Juan Singapore Sydney Tokyo Toronto



McGraw-Hill

A Division of The McGraw-Hill Companies



FUNDAMENTALS of ASTRODYNAMICS and APPLICATIONS

All graphic illustrations copyright © 1997 by David A. Vallado. All rights reserved. Printed in the United States of America. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a database retrieval system, without prior written permission of the publisher.

1 2 3 4 5 6 7 8 9 0 DOC DOC 9 0 9 8 7

ISBN 0-07-066829-9, Hardcover
ISBN 0-07-066834-5, Softcover

Library of Congress Catalog Card Number: 97-70117

Editor: M.A. Hollander

Cover Design: M.C. Lytle

Text Design: Linda Pranke

Printer/Binder: Donnelley & Sons Company

To God from whom all blessings and wisdom flow,
To the dedication of one's life to the pursuit of wisdom,
To my parents for their devotion and example,
To my wife's support through this long task,
and especially to my lovely daughters,

Simone and Kathleen

Table of Contents

Chapter 1	Fundamentals	1
1.1	History	1
1.1.1	Ancient Era	1
1.1.2	The Advent of the Modern Era	6
1.2	Geometry of Conic Sections	9
1.2.1	Basic Parameters	9
1.3	The Earth	17
1.3.1	Location Parameters	19
1.3.2	Shape of the Earth	20
1.3.3	Geodesy	28
1.3.4	Gravitational Model	29
1.4	Coordinate Systems	31
1.4.1	Interplanetary Systems	36
1.4.2	Earth-based Systems	37
1.4.3	Satellite-based Systems	40
1.5	Coordinate Transformations	45
1.5.1	Coordinate Rotation	45
1.5.2	Rotating Transformations	51
1.5.3	Common Transformations	53
1.6	Time	55
1.6.1	Sidereal Time	58
1.6.2	Solar Time and Universal Time	63
1.6.3	Atomic Time	69
1.6.4	Dynamical Time	70
1.7	Motion of the Coordinate System	74
1.7.1	Precession (J2000)	77
1.7.2	Nutation (J2000)	79
1.7.3	Sidereal Time (J2000)	81
1.7.4	Polar Motion (J2000)	82
1.7.5	Summary (J2000)	84
1.7.6	Other Corrections	89
1.7.7	B1950 Reduction	92
1.8	Earth Models and Constants	93
1.8.1	Canonical Units	94
Chapter 2	Equations of Motion	97
2.1	Historical Background	97
2.1.1	Kepler's Laws	99
2.1.2	Newton's Laws	101
2.1.3	Other Early Astrodynamical Contributions	102
2.2	Two-body Equation	103
2.2.1	Assumptions for the Two-Body Equation	105
2.2.2	Specific Angular Momentum	106
2.2.3	Specific Mechanical Energy	108
2.2.4	Kepler's First Law (Trajectory Equation)	110
2.2.5	Kepler's Second and Third Laws	112
2.2.6	Velocity Formulas	114

2.3	Three-body and n-body Equations	116
2.3.1	Inertial, Relative, and Barycentric Formulas	116
2.3.2	Ten Known Integrals	119
2.3.3	General Three-body Problem	122
2.4	Satellite State Representations	129
2.4.1	Classical Orbital Elements (Keplerian)	130
2.4.2	2-line Element Sets	140
2.4.3	Equinoctial Elements	142
2.4.4	Canonical Variables	143
2.5	Application: Orbital Elements from r and v	144
2.6	Application: r and v from Orbital Elements	149
Chapter 3	Fundamental Applications	155
3.1	Historical Background	155
3.2	Introduction	157
3.3	Time Conversions	158
3.3.1	DMS to Rad / Rad to DMS	158
3.3.2	HMS to Rad / Rad to HMS	159
3.3.3	HMS to UT / UT to HMS	160
3.3.4	YMD to Day of Year/ Day of Year to YMD	161
3.3.5	YMDHMS to Days / Days to YMDHMS	162
3.3.6	Inverse Julian Date	163
3.4	Observation Transformations	164
3.4.1	Geocentric Right Ascension and Declination	164
3.4.2	Topocentric Right Ascension and Declination	167
3.4.3	Azimuth-Elevation	170
3.4.4	Transformations for Ecliptic Latitude and Longitude	176
3.5	Application: Sun Position Vector	181
3.6	Application: Moon Position Vector	184
3.7	Application: Planetary Ephemerides	188
3.7.1	Eclipses	192
3.8	Application: Sunrise, Sunset, and Twilight Times	194
3.9	Application: Sight and Light	198
3.10	Converting IJK To Latitude and Longitude (LatLon)	202
Chapter 4	Kepler's Equation and Kepler's Problem	207
4.1	Historical Background	207
4.2	Kepler's Equation	209
4.2.1	Alternate Formulation for Eccentric Anomaly	215
4.2.2	Formulation for the Parabolic Anomaly	217
4.2.3	Formulation for the Hyperbolic Anomaly	219
4.2.4	Universal Formulation	224
4.2.5	Solutions of Kepler's Equation	230
4.2.6	Summary and Related Formulas	244
4.3	Kepler's Problem	247
4.3.1	Solution Techniques	247
4.4	Application: Groundtracks	264
4.5	Application: Find Time of Flight (FindTOF)	268

Chapter 5	Orbital Maneuvering	273
5.1	Historical Background	273
5.2	Introduction	275
5.3	Coplanar Maneuvers	275
5.3.1	Hohmann and Bi-elliptic Transfers	278
5.3.2	Comparing Hohmann and Bi-elliptic Transfers	284
5.3.3	Transfers Using the One-Tangent Burn	287
5.3.4	General Transfers	293
5.4	Noncoplanar Transfers	293
5.4.1	Introduction	294
5.4.2	Inclination-Only Changes	298
5.4.3	Changes in the Longitude of Ascending Node	301
5.4.4	Changes to Inclination and Longitude of Ascending Node	304
5.5	Combined Maneuvers	306
5.5.1	Minimum-Inclination Maneuvers	307
5.5.2	Fixed- Δv Maneuvering	310
5.6	Circular Rendezvous	314
5.6.1	Circular Coplanar Phasing	314
5.6.2	Circular Noncoplanar Phasing	321
5.7	Continuous-Thrust Transfers	327
5.7.1	Introduction	328
5.7.2	Orbit Raising	331
5.7.3	Low-Thrust, Noncoplanar Transfers	337
5.8	Relative Motion	343
5.8.1	Position Solutions for Nearly Circular Orbits	348
5.8.2	Trend Analysis	353
5.8.3	Extending the Results	366
Chapter 6	Initial Orbit Determination	371
6.1	Historical Background	371
6.2	Obtaining Data	374
6.2.1	Quantity of Data	377
6.2.2	Types of Data	378
6.2.3	Example Applications	378
6.3	Observations of Range, Azimuth, and Elevation	379
6.3.1	Application: SITE-TRACK	380
6.4	Angles-only Observations	386
6.4.1	Laplace's Method	388
6.4.2	Gaussian Determination	392
6.4.3	Double r -iteration	397
6.5	Mixed Observations	403
6.5.1	Range and Range-Rate Processing	403
6.5.2	Range-only Processing	405
6.6	Three Position Vectors and Time	408
6.6.1	Gibbs Method	408
6.6.2	Herrick-Gibbs	415
6.7	Two Position Vectors and Time—Lambert's Problem	421
6.7.1	Lambert—Minimum Energy	423
6.7.2	Lambert—Gaussian Solution	430

6.7.3	Lambert—Universal Variables	435
6.7.4	Lambert Solution—Battin Method	441
6.8	Application: Targeting Problem	446
Chapter 7	Special Perturbation Techniques	465
7.1	Historical Background	465
7.2	Introduction to Perturbations	467
7.3	Encke's Formulation	472
7.4	Cowell's Formulation	473
7.5	Numerical Integration Methods	475
7.5.1	Implementing an Integrator and Determining Step Size	482
7.6	Disturbing Forces	485
7.6.1	Gravity Field of a Central Body	485
7.6.2	Atmospheric Drag	497
7.6.3	Third-Body Perturbations	514
7.6.4	Solar-Radiation Pressure	517
7.6.5	Other Perturbations	524
7.7	Forming Numerical Solutions	526
7.7.1	Application: Simplified Acceleration Model	527
7.7.2	Application: Complex Acceleration Model	530
7.8	Practical Considerations	533
7.8.1	Validating the Propagator	534
7.8.2	Physical Data and Sources	535
Chapter 8	General Perturbation Techniques	539
8.1	Historical Background	539
8.2	Introduction	543
8.2.1	The Method of Perturbations	547
8.3	Variation of Parameters	549
8.3.1	Lagrangian VOP (Conservative Effects)	552
8.3.2	Gaussian VOP (Nonconservative and Conservative Effects)	559
8.4	Hamilton's Formulation	569
8.5	Disturbing-Potential Formulations	572
8.5.1	Gravity Potential in Terms of the Satellite's Orbital Elements	572
8.5.2	Third-Body Potential in Terms of the Satellite's Orbital Elements	574
8.5.3	Tidal-Motion Potential in terms of the Satellite's Orbital Elements	578
8.6	Linearized Perturbations and Effects	578
8.6.1	Central-Body Analysis	579
8.6.2	Drag Analysis	603
8.6.3	Third-Body Analysis	611
8.6.4	Solar-Radiation Analysis	617
8.7	Forming Analytical Solutions	621
8.7.1	Application: Perturbed Two-body Propagation	622
8.7.2	Kozai's Method	624
8.7.3	Brouwer's Method	626
8.8	Semianalytical Solutions	628
8.8.1	The Draper Semianalytical Satellite Theory (DSST)	629
8.9	Practical Considerations	634

8.9.1	Expected Accuracy of Propagation Techniques	635
8.9.2	Initial Conditions, Sources, and Conversions	640
8.10	Summary of Perturbation Effects	644
Chapter 9	Statistical Orbit Determination	649
9.1	Historical Background	649
9.2	Introduction to Sensor Systems	651
9.3	Linear Least Squares	658
9.3.1	Error Analysis	664
9.3.2	Linear Weighted Least Squares	671
9.4	Nonlinear Least Squares	675
9.5	Application: Orbit Determination With Differential Correction	682
9.5.1	Calculating the Partial-Derivative Matrix (\mathbf{A})	685
9.5.2	Calculating the Observation Partial Derivatives	688
9.5.3	Calculating the Variational Equations	689
9.5.4	Implementing Least-Squares Techniques	696
9.6	Sequential-Batch Least Squares	702
9.7	Kalman Filtering	707
9.7.1	Kalman Filter (Linear System)	717
9.7.2	Linearized Kalman Filter (<i>LKF</i>)	722
9.7.3	Extended Kalman Filter (<i>EKF</i>)	724
9.7.4	Successful Filtering Applications	727
9.7.5	Summary of Kalman Filters	728
9.8	Practical Considerations	730
9.8.1	Data	730
9.8.2	Implementing a Differential-Correction Technique	732
9.8.3	Analyzing the Results of Differential Corrections	735
9.8.4	Applications: Detecting Maneuvers	738
Chapter 10	Real-World Mission Analysis	741
10.1	Introduction	741
10.1.1	Satellite Populations	743
10.2	Mission Orbits	743
10.2.1	Stationkeeping	746
10.3	Geometries for Surveillance and Reconnaissance	749
10.4	Designing and Maintaining Mission Orbits	757
10.4.1	Sun-Synchronous Orbits	757
10.4.2	Repeat-Groundtrack Orbits	763
10.4.3	Minimum Altitude Variation Orbits	773
10.4.4	Frozen-Orbit Eccentricity and Argument of Perigee	779
10.4.5	Application: Designing a Specialized Orbit	780
10.5	Navigation—the Global Positioning System	788
10.5.1	Historical Background	788
10.5.2	System Introduction	791
10.5.3	Signals from GPS Satellites	793
10.6	Predicting Satellite Look Angles	800
10.7	Determining Close Approaches	805

10.7.1	Finding the Close-Approach Functions	806
10.7.2	Statistical Analysis	814
10.8	Rise / Set	815
Appendix A	Dictionary of Symbols and Acronyms	825
A.1	Dictionary of Symbols	825
A.2	Dictionary of Acronyms	838
Appendix B	Modeling the Atmosphere	843
B.1	Jacchia-Roberts Atmosphere	843
B.1.1	Evaluating Temperature	843
B.1.2	Robert's Corrections to Temperature	845
B.1.3	Jacchia's Corrections for Density	846
B.1.4	Robert's Corrections to Density	846
B.2	Soviet Cosmos Atmosphere	852
Appendix C	Mathematical Fundamentals	855
C.1	Vector Fundamentals	855
C.2	Matrix Fundamentals	856
C.3	Trigonometric Fundamentals	857
C.3.1	Planar Trigonometry	860
C.3.2	Spherical Trigonometry	861
C.4	Numerical Techniques	863
C.4.1	Polynomial Solutions	863
C.4.2	Interpolation	867
C.4.3	Blending and Splining Techniques	867
Appendix D	Constants and Expansions	875
D.1	Gravitational Coefficients	875
D.2	Planetary Constants	877
D.3	1980 IAU Reduction Coefficients	880
D.4	Planetary Ephemerides	882
D.5	Data Sources	886
D.6	Computer Programming	887

Algorithm Summary

Algorithm 1	<i>LSTIME</i>	61
Algorithm 2	<i>Julian Date</i>	67
Algorithm 3	<i>ConvTime</i>	72
Algorithm 4	<i>J2000 Reduction</i>	86
Algorithm 5	<i>ELORB</i>	146
Algorithm 6	<i>RANDV</i>	151
Algorithm 7	<i>DMStoRad</i>	158
Algorithm 8	<i>RadtoDMS</i>	158
Algorithm 9	<i>HMStoRad</i>	159
Algorithm 10	<i>RadtoHMS</i>	159
Algorithm 11	<i>HMStoUT</i>	160
Algorithm 12	<i>UTtoHMS</i>	160
Algorithm 13	<i>Geocentric Radec</i>	166
Algorithm 14	<i>Topocentric</i>	168
Algorithm 15	<i>RAZEL</i>	173
Algorithm 16	<i>RadectoAzel</i>	176
Algorithm 17	<i>AzeltoRadec</i>	176
Algorithm 18	<i>Sun</i>	183
Algorithm 19	<i>Moon</i>	186
Algorithm 20	<i>PLANETRV</i>	190
Algorithm 21	<i>SunriseSet</i>	196
Algorithm 22	<i>SIGHT</i>	201
Algorithm 23	<i>IJK to LatLon</i>	204
Algorithm 24	<i>Find $c_2 c_3$</i>	230
Algorithm 25	<i>KepEqtnE</i>	232
Algorithm 26	<i>KepEqtnP</i>	236
Algorithm 27	<i>KepEqtnH</i>	237
Algorithm 28	<i>ν to Anomaly</i>	245
Algorithm 29	<i>Anomaly to ν</i>	245
Algorithm 30	<i>KeplerCOE</i>	249
Algorithm 31	<i>Kepler</i>	261
Algorithm 32	<i>FINDTOF</i>	270
Algorithm 33	<i>Hohmann Transfer</i>	281
Algorithm 34	<i>Bi-elliptic Transfer</i>	282
Algorithm 35	<i>One-Tangent Burn</i>	289
Algorithm 36	<i>Inclination Only</i>	300
Algorithm 37	<i>Change in Longitude of Ascending Node—Circular</i>	304
Algorithm 38	<i>Combined Changes to i and Ω—Circular</i>	306
Algorithm 39	<i>Minimum Combined Plane Change</i>	309
Algorithm 40	<i>Fixed-Δv Maneuvers</i>	312

Algorithm 41	<i>Circular Coplanar Phasing (Same Orbit)</i>	318
Algorithm 42	<i>Circular Coplanar Phasing (Different Orbits)</i>	319
Algorithm 43	<i>Noncoplanar Phasing</i>	324
Algorithm 44	<i>Hill's Equations</i>	351
Algorithm 45	<i>SITE-TRACK</i>	384
Algorithm 46	<i>ANGLES-ONLY GAUSS</i>	396
Algorithm 47	<i>ANGLES-DOUBLE R</i>	398
Algorithm 48	<i>GIBBS</i>	414
Algorithm 49	<i>HGIBBS</i>	420
Algorithm 50	<i>Lambert's Problem—Minimum Energy</i>	428
Algorithm 51	<i>Lambert—Gaussian Solution</i>	434
Algorithm 52	<i>Lambert—Universal Variables</i>	439
Algorithm 53	<i>Lambert—Battin Method</i>	441
Algorithm 54	<i>Hit Earth</i>	450
Algorithm 55	<i>Target</i>	451
Algorithm 56	<i>Encke</i>	473
Algorithm 57	<i>PKEPLER</i>	623
Algorithm 58	<i>NOMINAL STATE</i>	697
Algorithm 59	<i>Differential Correction</i>	699
Algorithm 60	<i>Kalman Filter—Linear System</i>	717
Algorithm 61	<i>Linearized Kalman Filter</i>	723
Algorithm 62	<i>Extended Kalman Filter</i>	725
Algorithm 63	<i>Repeat Groundtrack</i>	772
Algorithm 64	<i>Minimum Altitude Variation</i>	778
Algorithm 65	<i>PREDICT</i>	802
Algorithm 66	<i>ANCAS</i>	811
Algorithm 67	<i>RISE/SET</i>	819

Preface

My goal in writing this book is to augment some of the information other astrodynamics books contain and, perhaps more importantly, to present and describe the computational fundamentals of astrodynamics. During the early 1960s, tremendous progress took place in astrodynamics, in part due to the U.S.'s drive to reach the Moon and to stay technologically superior during the Cold War. While the theories advanced by large steps, the computer code took only a "small step" because of available computing languages and hardware. But in the 1970s and 1980s, the computer industry took gigantic steps, so now is the time to capitalize on advancements in both areas.

You're probably wondering, does this book present anything new? Well, yes and no. The dynamics have been in effect since the beginning of time, and most of our mathematics hasn't changed much since its discovery over 300 years ago. Today, advances in computer technology allow students and practitioners to determine orbits with high precision on a PC, but they need a compendium of *well-documented* algorithms to use that power. Hence, this book.

The days of studying just one discipline are nearly over. Today's astrodynamists should study computer science in *addition* to rigorous engineering principles—after all, that's how we do our business. As older managers retire and newer computer systems arrive, the need is more pressing than ever for a new look at *traditional* astrodynamics. The days of "well, it works" are long gone.

The keys for the future are standardizing and documenting, modularity, efficiency, and maintainability. This book attempts to use the new programming capabilities to exploit each of these areas. Standardizing and documenting can't be overemphasized. I've witnessed tremendous efforts in astrodynamics and, on every project, I always came away wishing that someone had written a few more steps, or had provided a test case to check how the program worked. Too often, books are written mainly for academics; writers forget that engineers need simple algorithms and examples to use in everyday operations. Operational "documentation" is often severely limited or nonexistent. Although I use standard notation as much as possible and work hard to match existing practices, it's impossible to be consistent with all sources.

I've included many derivations because I've seen too many engineers spend excessive time trying to recreate something that is "easy to verify." Recall the infamous scribbling of Fermat in the margin of a page where he omitted the proof because surely everyone could understand it and there wasn't enough room to say more (*Scientific American*, Oct., 1993, 1994 and Oct. 1978, 104–122). It remained a mystery for 350 years and though recently claimed to have been proven, still generates controversy today! Redoing fundamental problems is useful to learn the basics, but we can save steps (and therefore time and money) by presenting the material straightforwardly, with consistent notation.

Modularity is closely aligned with standardizing and documenting, but it's distinct. I've become accustomed to the old axiom "this program already does everything you need," only to find that your application requires just 5% of the code, and you can't pull out the routines you need because all routines are interconnected. In this book I present many

independent algorithms. Admittedly, some routines will share algorithms, but with a little thought, this becomes only a minor issue.

Efficiency is important, yet elusive. During the development, we can consider some simple guidelines that will make routines run much faster. I've given you lots of examples and hints throughout the book.

Finally, at times it will seem as though I'm leading you by the hand through a problem, and at others, you'll need to scratch some intermediate equations in the margin (make sure you *do* record your notes!). My goal in writing this book has been (within some reasonable bounds for the overall length) to derive and detail each algorithm to about the same level to provide a solid foundation. Because individual abilities vary widely, you may find routines you've worked with to be too simple and unknown algorithms to be confusing. I've tried to minimize the confusion, and I'll appreciate your comments for improvement.

This book tries to bridge between astrodynamics material in undergraduate courses and follow-on, graduate-level topics. It's intended for upper-level undergraduate classes in astrodynamics, as well as for some graduate-level classes. From my experience in laboratory research, operational environments, and instruction, I've tried to present the material with a practical flavor. The book covers theory, applications, and interpretations through numerous examples. The idea is to blend real-world operations with a theoretical understanding of the mechanics at work.

I've coded almost all the algorithms to verify each technique. Actual computer source code will be available via CD-ROM shortly. It will detail test cases and code in PASCAL, C++, or FORTRAN-90. I hope these languages will be diverse enough to permit relatively easy conversion to other languages.

Finally, although I've made every attempt to correct mistakes in the book, I'm sure there will be some errors. I encourage you to contact myself, or Wayne McClain by e-mail for any questions you discover. *All* comments are important!

valladod@plk.af.mil

wmcclain@draper.com

Questions regarding the publication of the book should contact

mholland@mcgraw-hill.com, or via the Internet: mcgraw-hill.com

Acknowledgments

Before leaping into Chap. 1, I want to credit the numerous people who made this project a success. First and foremost, I thank my parents who have encouraged me throughout my life. My early childhood memories are of the 1960s, when the Cold War raged, the Cuban missile crisis arose, and Vietnam erupted. To their credit, my parents somehow diverted my attention toward one of the most important events to occur in human history: the effort to land on the Moon. I'll never forget that feeling I had, glued to the TV (VCRs weren't available yet) as Neil Armstrong walked the surface of the Moon. That event, coupled with my father's example as an engineer on the Manhattan Project during World War II, compelled me to become an engineer. Space was the obvious choice as the final frontier to explore.

Any large work needs solid technical reviews, but I'm especially thankful for those of Wayne McClain. As the primary technical reviewer, he helped me refine this book into a polished presentation. His insight and extensive experience were irreplaceable. Also, his patience with my ever-changing ideas and endless questions was truly remarkable! He provided extensive technical contributions to the book resulting from his experience in the field. The Charles Stark Draper Laboratory also deserves credit for granting Wayne McClain the time necessary to provide the thorough review of this work.

I'd also like to credit the people who allowed me to adapt parts of their work. I've listed each person alphabetically to avoid any favoritism—every comment was useful!

- Salvatore Alfano provided many papers for this book on continuous-thrust modeling (with Jim Thorne), close approach (with David Negron), rise/set calculations, and the method of ratios (also with David Negron). These works add a lot of practical insight and applications to my presentations.
- Leon Blitzer graciously permitted use of his *Handbook For Orbital Perturbations* (1970) for the perturbations chapters. Although I changed the look of his handbook material, much of the content remains, and I tried to keep the original intent he achieved almost thirty years ago.
- Carole Boelitz permitted me to include sections from her thesis in the perturbations sections. She also provided valuable reviews and derivations.
- Chapter 9 was initially based on a handout Jack Ferguson developed at the U.S. Air Force Academy in 1978. Jack Ferguson, others, and I reworked this handout in 1992 and I've expanded it again for this application.
- Daniel Fonte also allowed me to use sections of his thesis and his recent technical papers to give great credibility to the perturbations section. This area has been neglected for much too long.
- Tom Logsdon and Warner Miller provided initial thoughts and insight for the GPS section for Chap. 10.
- Jim Thorne allowed me to use sections from his continuous thrust and series reversion for solution of the Lambert problem. Some examples are from his work.

I received tremendous support from many people throughout the astrodynamics community.

- Gerry Baumgartner and Michael Gabor provided a very detailed examination of the complete text. They gave extensive derivations and provided a level of confidence in many of the difficult sections.

Other people worked entirely through the text, some more than once:

- Don Danielson (and many of his students), Kim Luu, Richard Rast, and Gary Yale.

Other individuals concentrated on certain chapters, relying on their expertise, which worked especially well considering the broad range of topics covered in the book:

- Carole Boelitz, Paul Cefola, Jack Ferguson, Jim Hall, Richard Huisak, John Junkins, David Negron, Jr., Beny Neta, David Richardson, Ken Seidelmann, Wayne Smith, Sid Sridharan, and Jim Thorne.

Several people reviewed certain sections in detail:

- Cheryl Arney, Eric Beck, Eric Claunch, Steve Crumpton, Gim Der, Karl Jensen, Ron Lisowski, Anne Long, Ron Madler, Warner Miller, Steve Nerem, Dave Pohlen, Chris Sabol, David Spencer, David Suzuki, Nate Titus, Mark Tolefson, Scott Wallace, Kitt Carlton-Wippern, Gary Wirsig, and Lora Worth.

Any book written mostly by an engineer needs fine tuning to become a complete work. Daryl Boden assisted as a technical editor and provided technical insight for the text. Perry Luckett, a former professor of mine, really came to the rescue. He clarified ideas through more direct language, standardized or sharpened figures and tables, and kept my English in line!

I enjoyed continued support from the leadership at Phillips Laboratory, including Richard Davis, Eugene Dionne, Michael Heil, Henry Pugh, and Edward Duff. Their support for, and perseverance with, this long project strongly contributed to a quality product for Phillips Laboratory.

The book was formatted in FrameMaker Ver. 4.02 by several people. I began the typing and oversaw the operation throughout. Elsie Lugo deserves great credit for helping me create the first draft. Linda Pranke continued as production editor through the many intermediate drafts until the book was close to publication. She not only endured my cryptic notations but attended to every detail and caught inconsistencies that might have confused readers. She really made this book happen—despite my endless changes!

Last but certainly not least, my thanks to Margaret Hollander and Wiley Larson for getting this book published and on the streets!

CHAPTER 1 FUNDAMENTALS

- 1.1 History
- 1.2 Geometry of Conic Sections
- 1.3 The Earth
- 1.4 Coordinate Systems
- 1.5 Coordinate Transformations
- 1.6 Time
- 1.7 Motion of the Coordinate System
- 1.8 Earth Models and Constants

1.1 History

Let's begin by agreeing on a definition of *astrodynamics*. Most dictionaries don't help—we can find no specific references for astrodynamics. We can define related topics: *celestial mechanics* examines the dynamic motions of celestial objects, *orbital dynamics* studies the motion of all orbiting bodies, and *attitude dynamics* deals with the orientation of an object. But these subjects don't capture the essence of astrodynamics, so I've chosen a definition from *Space Vehicle Design* (Griffin and French, 1991, 85):

Astrodynamics is the study of the motion of man-made objects in space, subject to both natural and artificially induced forces.

Notice that this definition actually combines features of celestial mechanics with orbital and attitude dynamics. Although discussion of celestial events is necessary, this book will focus on the behavior of artificial objects in space. Examining artificial and natural forces is important because it allows us to see the effects of both on a satellite. From a historical perspective, true astrodynamics has existed for only about 40 years, so we must examine its predecessors—astronomy and (ancient) astrology, coupled with celestial and orbital dynamics. With this caveat in mind, let's look at the progression of science leading to modern astrodynamics.

1.1.1 Ancient Era

The ancient astronomers (we'll consider those before about 200 A.D.) accomplished many technical feats, considering their primitive tools. Their accomplishments probably resulted from their dedication and the tremendous amount of time available to study a particular

problem. It would take an entire book to detail all of their contributions to modern astrodynamics, so I've included only a few snapshots to capture the flavor of the times.

We can find the ancestral roots of astrodynamics before the birth of Christ. At that time developments in astrodynamics occurred in astrology. Although modern scientists cringe at the thought of being called astrologers, Berry (1961, 20) remarks:

To the Chaldeans may be assigned the doubtful honour of having been the first to develop Astrology, the false science which has professed to ascertain the influence of the stars on human affairs, . . . It should be noted, however, that if the history of astrology is a painful one, . . . the belief in it has undoubtedly been a powerful stimulus to genuine astronomical study.

Perhaps the two greatest concerns of the ancient civilizations were agriculture and religion. People were concerned about their crops, the summer rains, storms, or, in essence, their very existence. As Barcley (1956, 97) says,

The ancient world was dominated by the thought of the influence of the stars; and even the greatest and wisest men would not act without consulting them. [They] believed that all things were in the grip of an iron fatalism settled by the stars; and the science of astrology professed to provide men with the secret knowledge which would rid them of their slavery to the elemental spirits.

Because most events in nature are random (such as earthquakes and storms), astrologers sought explanations in events that were consistent and predictable over long periods of time, such as the movements of the Sun, Moon, and stars. Given the religious significance of determining heavenly motions, they attacked the problem with great vigor and made every effort to improve their accuracy.

The Sun's diurnal rising and setting motion was the easiest to observe; yet, its brightness made *precise* observations difficult. Early people studied solar eclipses in great detail because these events terrified them. They knew their crops depended on the Sun, so when it "went out," they feared disaster—an unfortunate consequence of natural disasters occurring coincidentally after each eclipse. It's interesting that a *good* event wasn't linked to the eclipse, but we seem to remember bad events more readily. In any case, studying the Sun and Moon quickly dominated astrology.

The Chaldeans determined the *Saros cycle* by noticing the regularity of certain types of eclipses over time (Green, 1988, 449). The period between eclipses is 6585.32 days, and consecutive occurrences are displaced approximately 120° to the west. An interesting side legend tells of Chinese astronomers who regularly predicted eclipses, probably using the Saros cycle (Asimov, 1971, 18–19). The night before an eclipse they had predicted, the astronomers went out drinking. Unfortunately, in their inebriated state, they forgot to tell the emperor. When the eclipse frightened the emperor, he ordered the astronomers executed. If true, this legend has important lessons for us even today, for mistakes in our field of astrodynamics can have grave consequences! Engineers must hold themselves to a higher standard of accuracy.

Observing the cycles of the Moon was slightly easier than for the Sun, and it permitted more accurate observations. The early Babylonians, who had reasonably good weather

conditions for observing the Moon's phases, developed the lunar month (about 29½ days) (Pannekoek, 1989, 21).

The stars and planets presented a greater challenge because their motions appear transient and inconsistent. The early Babylonians accomplished tremendous feats and were even able to predict the phases and period of Venus. Still, as Pannekoek (1989, 33) points out, many of their predictions were based on incorrect phenomena or intervals from erroneously copied values and rough averages.

Thales (c. 640–546 B.C.) of Miletus, was perhaps the first astronomer to receive notoriety. He is credited with determining the length of the year and founding the Ionian school of astronomy and philosophy. He predicted eclipses, probably using the Saros cycle. Moulton (1914, 30) suggests that he may also have defined the obliquity of the ecliptic. As a teacher, he was able to influence many people of the time. Thales is believed to have taught that the Earth was spherical, but people generally remained skeptical until the 15th century!

Pythagorous (569–470 B.C.) is best known for his contributions to geometry; however, he also taught astronomy and philosophy. In his courses, he taught that the Earth and comets revolved around the Sun, and the Earth rotated about its own axis. He also believed virtually every problem could be solved by properly applying numbers (Dreyer, 1953, 36). This principle resulted in the *harmony of the spheres*—the thought that each planet emitted a certain musical note. Of course, the music was audible only to certain people.

Much of Plato's (c. 427–347 B.C.) work may have come from his students, and although he is best known for his philosophical insights, he also extended the mathematical interpretation of the universe (harmony of the spheres) alluded to by Pythagorous. The musical notes were based on the

major consonances, that is the octave, perfect fifth, and perfect fourth, [which] are produced by vibrating strings whose lengths stand to one another in the ratios of 1:2, 3:2, and 4:3 respectively. (Colliers, 1992, 547)

His logic followed that if the musical world were so mathematically oriented, the cosmos must be ordered likewise.

The ancient Greeks were perhaps best known for their contribution to mathematics. Euclid (c. 330–275 B.C.) is credited with being the first to write about conic sections, but his writings were lost. Consequently, Apollonius (c. 287–212 B.C.) is credited with the first known treatise on conic sections (c. 225 B.C.) and was the first to name the sections. He probably also introduced the *excentric* and *epicycle* theories of orbital motion.

Although records from this period are scarce because of fires and destruction over time, it's commonly thought that Aristarchus (310–250 B.C.) suggested the Earth revolved around the Sun. Unfortunately, his theory met great resistance because it couldn't predict the position of Mars and didn't observe the expected angular separation from different viewing locations. It's interesting to note that Copernicus used some of Aristarchus's results almost 1800 years later to develop his own heliocentric model.

Eratosthenes (275–194 B.C.) was perhaps the first person to obtain a reasonable estimate of the Earth's radius. He did this using knowledge of the Sun's light rays during the

summer solstice in Syene (Aswan), Egypt (24° N). On the summer solstice, the light rays were known to fall vertically in Syene, but Eratosthenes lived about 805 km north of Syene in Alexandria (31° N). He measured the Sun's rays at an angle of 7.2° , or about $1/50^{\text{th}}$ of a circle (Defense Mapping Agency, 1983, 2). From this meager data and knowledge of simple geometry, he was able to determine the Earth's radius. It's difficult to determine the accuracy of his estimate because he used units of stadia, but most sources agree his result was within about 1% of today's accepted value!

Although most of his writings were lost, we know that Hipparchus (c. 161–126 B.C.) developed spherical trigonometry and taught that the Earth was the center of the universe, even though Pythagoras and Aristarchus placed the Sun at the center much earlier. This inconsistency points out the difficulties in sharing data at the time. In modern society, we often take for granted the flow and availability of information, but that flow was very limited in ancient times. To copy a few pages of calculations or a book required scarce materials, knowledgeable scribes, and days, weeks, or even years of effort. It's no surprise that scientists in the same country often didn't know of their contemporaries' accomplishments and findings in towns a few hundred miles away.

Hipparchus also noticed an increase in the longitude of the stars (Green, 1988, 51). To observe such a small phenomenon during the second century B.C. is extraordinary and illustrates the dedication, time, and diligence that the ancient Greeks applied to particular tasks. Hipparchus likewise developed the first system of cataloging star magnitudes. The list included 800 to 1000 stars, categorized by brightness. He did this by noting that the stars could be separated into six categories of brightness, each of which was separated by about 2.5 times the brightness. The brightest was a first magnitude, whereas the dimmest, about 100 times dimmer, was a sixth magnitude. It's difficult to imagine the naked eye being able to detect gradations accurately enough to place nearly 1000 stars in the correct categories.

Hipparchus also developed theories to describe orbital motion. He made very precise observations, which presented some problems when trying to describe the orbital motion. As mentioned earlier, the original science of astrology had strong ties to religion, so people thought the heavens were holy and perfect. Because the circle was the only geometric shape considered to be perfect, they relied on circular motion. Hipparchus, and probably Apollonius before him, was caught in a difficult dilemma—fitting observational data to a theory which (as we know now) didn't represent the data. Hipparchus continued developing the basic *excentric* and *epicycle* systems. The excentric system consisted of the Sun revolving about the Earth in a circular path, called the *excentric*, which is a convenient way to represent the Sun's apparently irregular motion while preserving perfect circular motion from the center of the circle. For the epicycle system, the Sun was placed on a small circle called the *epicycle*, the center of which rotated about the Earth. The circular orbit of the epicycle center was referred to as the *deferent*. Both the excentric and epicycle systems resulted in the same apparent motion of the Sun and planets about the Earth. Ptolemy used the term excentric instead of the deferent, which could cause confusion. This book uses the excentric term as a historical remnant to avoid confusion with eccentricity, which I'll use later. Figure 1-1 shows the two theories for clarity.

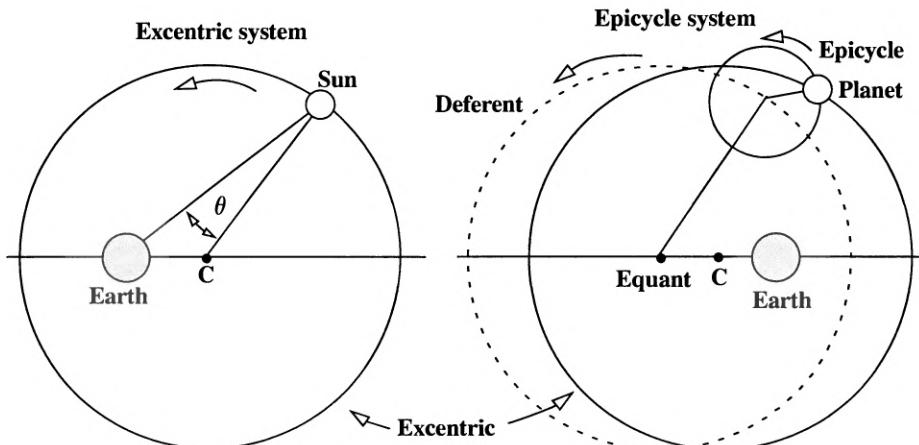


Figure 1-1. Excentric and Epicycle Systems. Both the excentric and epicycle theories adhered to an Earth-centered philosophy and used combinations of circular motion to approximate observations. The excentric system consisted of the Sun traveling around a larger circle, the excentric, with the Earth displaced to one side. In the epicycle system the Sun (or planet) moved around a smaller circle, the epicycle, the center of which moved around the equant, forming the deferent. Ptolemy placed the Earth at the equant for the Sun, but left it as shown for all other planets. (Beer and Strand, 1975b, 559–560).

Julius Caesar developed the Julian calendar in 46 B.C. as a sophisticated way to track time. That particular year had 445 days due to the tremendous error which had built up (Berry, 1961, 23). The **Julian year** of 365.25 days is quite familiar because it contains a leap day every four years to account for the Earth's nonuniform motion about the Sun (the actual number is 365.242 189 7). A **Julian century** has exactly 36,525 days. Unfortunately, the Julian year doesn't account for the remaining variation (about 11 minutes and 14 seconds per year). Thus, the system was fairly accurate over a few centuries but accumulated appreciable errors over longer periods. In 1582, Pope Gregory XIII issued a decree that ten days would be omitted from the calendar of October 1582 to correct the accumulated error. Concurrently, the small variation in the length of the year was partially corrected by establishing a **Gregorian calendar** in which century years would have a leap day *only* if they were evenly divisible by 400 (1900 has no leap day, 2000 does, 2100 doesn't). Although we use the Gregorian calendar today, be careful in historical research because some countries didn't adopt this calendar right away (Great Britain, 1752; Turkey, 1927). The error in this approach is only about 26 seconds per year.

Lastly, Claudius Ptolemaeus (usually called Ptolemy) (100–170 A.D.) published a 13-volume work, called the *The Mathematical Collection (Mathematike syntaxis)* or the *Almagest*, which contained his theory for an Earth-centered solar system. He used the results of Hipparchus but was unaware of earlier astronomers' works that declared the

Earth to be spherical and rotating around the Sun. Using some of his own observations, Ptolemy introduced the final refinements to the excentric and epicycle theories: the *moveable excentric* and the *equant*. The *moveable excentric* was a complicated arrangement in which the excentric moved about the Earth. The *equant* was similar to the earlier epicycle theory, except the Earth was placed on the opposite side from the center of the excentric. The motion of the Sun or planet was uniform about the equant but still occurred about the center of the deferent. This was actually a departure from the circular motion.

It may, however, fairly be doubted whether Hipparchus or Ptolemy ever had an abstract belief in the exclusive virtue of such [circular] motions, except as a convenient and easily intelligible way of representing certain more complicated motions, and it is difficult to conceive that Hipparchus would have scrupled any more than his great follower, in using an equant to represent an irregular motion, if he had found that the motion was therefore represented with accuracy. . . . The earlier Greeks, . . . and again many astronomers of the Middle Ages, felt that it was on a priori grounds necessary to represent the perfection of the heavenly motions by the most perfect or regular of the geometric schemes. (Berry, 1961, 71–72)

Although Ptolemy had devised unique methods to explain the motion of the Sun and planets, he still faced the problem of fitting the observations to one of his new theories. The problem was separated for the *inferior planets* (Mercury, Venus, and the Sun), and the *superior planets* (Mars, Jupiter, and Saturn). The difficulty *all* the ancient astronomers faced was explaining the observed irregular motions of the Sun and planets. In fact, some planets exhibit *retrograde* motion at times, or a backwards motion when observed from the Earth. The effect is the same as the perceived velocity change as you pass another car.

Ptolemy's explanation of orbital motion greatly increased its complexity. He even accounted for the inclination of the planets by slightly tilting the epicycles. Finally, in his *Almagest*, Ptolemy described the construction of an *astrolabe*, a fundamental tool of the time, and the basis of tools still used today. The instrument permitted calculations of the motions of the planets and stars; it was based on the theories Ptolemy presented in the rest of his work.

Over the next 1500 years, very little happened in the fields which led to astrodynamics. In the 7th century, the rise of Islam demanded more accessible tables of the stars so followers could determine the direction of Mecca; thus, the *Almagest* was translated into Arabic. In fact, the 9th-century title was created from the Arabic superlative *mageste* (meaning largest or biggest) with the added *al* prefix. Because instruments of the time couldn't measure the inaccuracies of the Ptolemaic theories, there was no real need for change, and poor distribution of data delayed revisions of the theories.

1.1.2 The Advent of the Modern Era

The early astrologers had done much work that laid the foundation for the next round of thinkers: Nicholas Copernicus (1473–1543), Galileo Galilei (1564–1642), Johann Kepler (1571–1630), and Isaac Newton (1642–1727). These men mainly took the values and constants discovered earlier and integrated them with “newly” formed orbital theories to describe the motion of the planets.

In Chapter 2, we'll completely derive two-body motion and look at Kepler's and Newton's laws of motion. Kepler's three laws are:

1. The orbit of each planet is an ellipse with the Sun at one focus.
2. The line joining the planet to the Sun sweeps out equal areas in equal times.
3. The square of the orbital period of a planet is proportional to the cube of its mean distance to the Sun.

Perhaps more than any other law, the Law of Gravitation enables the astrodynamist to model planetary and satellite motion. In the *Principia*, Newton ([1687] 1946, 415) stated that any two point masses attract one another with a force proportional to the product of their masses and inversely proportional to the square of the distance between them. We can express this law mathematically (for a satellite orbiting the Earth) in vector notation as

$$\vec{F}_g = -\frac{Gm_{\oplus}m_{sat}}{r^2} \frac{\vec{r}}{|\vec{r}|} \quad (1-1)$$

where \vec{F}_g is the force on the satellite (of mass m_{sat}) due to the Earth (of mass m_{\oplus}), and \vec{r} is the vector from m_{\oplus} to m_{sat} . As a consequence of Newton's Universal Law of Gravitation, Kepler's ellipse is extended to *any* conic section. This gives us tools to define the motion of an object in space. Eq. (1-1) assumes the entire force of gravity acts as though it were a point at the center of mass. Later, we'll need to generalize Eq. (1-1) to account for nonuniform mass distribution in finite bodies. This is called a nonspherical gravity field. The difficulty in the 1600s with this equation was how to find the *gravitational constant*, G , which is really a proportionality constant.

In 1798, Henry Cavendish (1731–1810) measured a value for this constant. He reasoned that two massive spheres would attract each other when they came close together, but he faced the problem of negating the effects of the Earth's gravity. He devised a scheme to enclose two lead spheres in an airtight box, to prevent any air currents from affecting the experiment. Inside the box, he placed two more large lead spheres on a rod suspended by a wire. They were on the same level, so gravity would affect each sphere equally. Then, using a torsional balance (developed by Coulomb in 1777), he was able to measure the twist of the wire due to the gravitational attraction of the lead spheres. The precise nature of this experiment is almost as remarkable today as it was 200 years ago.

Although the procedure described above will find a representative value for G , its accuracy is limited because the gravitational attraction between small masses is very small and thus difficult to measure accurately. Today, we commonly measure the quantity Gm_{\oplus} from satellite observations, which is much more accurate than Cavendish's method. The quantity is called the *gravitational parameter*, μ . The JGM-2 (See "Gravitational Coefficients" on page 30.) values for μ and G are

$$\mu \equiv G(m_{\oplus} + m_{sat}) \approx Gm_{\oplus} = 3.986\,004\,415 \times 10^5 \frac{\text{km}^3}{\text{s}^2}$$

$$G = 6.673 \times 10^{-20} \pm 0.001 \times 10^{-20} \frac{\text{km}^3}{\text{kg} \cdot \text{s}^2}$$

$$m_{\oplus} \approx 5.973\,332\,0 \times 10^{24} \text{ kg}$$

Although the precise definition includes the masses of the Earth and the satellite, we can neglect the satellite's mass because it's so small relative to the Earth's. Still, we can easily determine m_{sat} before launch. Whenever the objects of interest are more nearly the same (the Sun and Jupiter, for instance), both masses are necessary. Notice that, although μ and G are experimentally measured, m_{\oplus} is a derived quantity.

At the time of Cavendish, measuring the mass of the Sun and Earth was as difficult as finding G . The solution was to determine a gravitational parameter for the solar system, or more correctly, its square root, called the *Gaussian constant of gravitation*, κ . The International Astronomical Union adopted a value of κ for the Sun at their 1939 meeting and proclaimed that it couldn't be revised (Taff, 1988, 94–95).

$$\kappa_{\odot} = \sqrt{G m_{\odot}} = 0.017\,202\,098\,95$$

Actually, Karl Friedrich Gauss (1777–1855) first determined this value in 1809. Gauss recognized that

it is of no importance which body we use for determining this [gravitational parameter for] the motion of the Earth, the mean distance of which from the Sun we shall adopt for the unit of distance; the mean solar day will always be our unit of time. (Gauss [1809] 1963, 2).

With the constants listed below and Kepler's third law, Gauss determined a value for the gravitational parameter which we still use. He recognized that distance, mass, and G weren't readily available by observation, so he adopted 1 AU, 1 m_{\odot} , and the year to normalize all other values. The value of κ_{\odot} depends on the Earth's orbital period about the Sun, P_{\oplus} ; the masses of the Earth and the Sun— m_{\oplus} and m_{\odot} , respectively—measured in solar masses; and the semimajor axis of the Earth's orbit measured from the Sun, a_{\oplus} . Because he didn't know the value of this semimajor axis, he adopted it as unity and gave it the name *astronomical unit*, AU. Let's begin by examining the period relation [Kepler's third law and Eq. (2-16)].

$$P_{\oplus} = \frac{2\pi\sqrt{a_{\oplus}^3}}{\sqrt{G \sqrt{m_{\odot} + m_{\oplus}}}}$$

Substituting the Gaussian constant of gravitation and solving gives us

$$\kappa_{\odot} = \frac{2\pi\sqrt{a_{\oplus}^3}}{P_{\oplus}\sqrt{1 + \frac{m_{\oplus}}{m_{\odot}}}}$$

$$\begin{aligned} P_{\oplus} &= 365.256\,383\,5 \text{ days} & a_{\oplus} &= 1.0 \text{ AU} \\ m_{\oplus} &= \frac{1}{354,710} \text{ solar masses} & m_{\odot} &= 1.0 \end{aligned}$$

About determining the Gaussian constant of gravitation, Brouwer and Clemence (1961, 58–59) state:

As a matter of historical interest it may be noted that Gauss intended to choose K_{\odot} such that the unit of distance would turn out to be half the major axis of the Earth's orbit. Actually, because Gauss used a value for the Earth's mass we know to be too small by about 7%, it turns out that half of the major axis of the Earth's orbit, as derived from Kepler's third law with Gauss's value of K_{\odot} , is 1.00000003 a.u. Much confusion has been caused by careless usage of the words semimajor axis and mean distance in the astronomical literature.

1.2 Geometry of Conic Sections

We'll explore Kepler's laws of motion in later chapters, but first, let's look at some fundamentals. Kepler's first law states that the planets travel in conic sections, namely, ellipses. We must consider the types of conic sections because they represent *all* the possibilities allowed by Newton's law of gravity for satellite orbits. The name derives from the fact that a *conic section* is the intersection of a plane and a right circular cone.

Figure 1-2 describes generally how to find each conic section. The first four conic sections (circle, ellipse, parabola, and hyperbola) are considered to be orbits. The point at the intersection of the two cones is a special case that doesn't represent realistic satellite motion. *Rectilinear orbits* are defined for elliptical, parabolic, and hyperbolic motion. They result from the plane lying on an outer surface of the cone, a plane parallel to and infinitely far from the surface, and a plane perpendicular to and infinitely far from the center of the base, respectively. Although these aren't *really* closed orbits, they represent ideal limiting cases that we may approximate by sections of parabolic and hyperbolic orbits. We use them to represent sections of maneuvers, comets, or flights of missiles. Because we may encounter these orbits, computer software should account for them to avoid numerical problems. Later chapters treat both standard and special orbits.

1.2.1 Basic Parameters

A few parameters describe a conic section's shape and size and allow us to locate the satellite's position within the orbit. Certain geometrical concepts are important because they underlie many derivations throughout the book. Figure 1-3 shows the construction of a conic section.

The foci are the first features we should examine. Every conic section has two *foci*. In astrodynamics, the foci are important because the planetary center coincides with one focus for all motion, called the *primary focus*. From a geometrical standpoint, the foci provide the locations from which to construct a conic section. Figure 1-4 shows the two distinct foci for an ellipse, and Fig. 1-5 shows the special case of a circle in which both

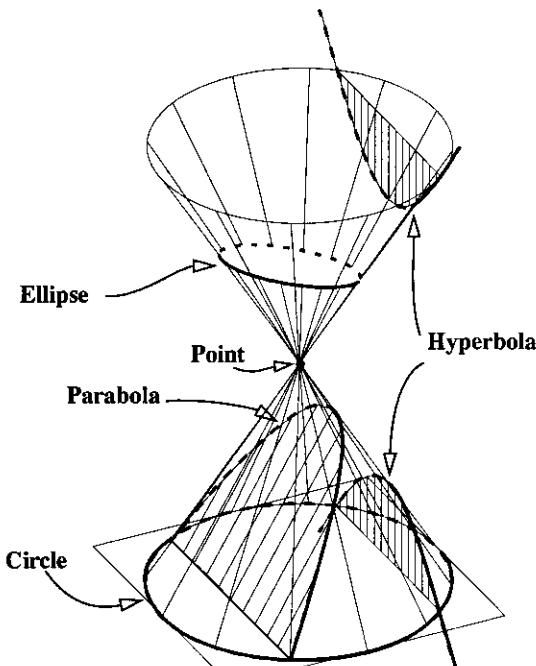


Figure 1-2. Conic Sections. Conic sections are formed by slicing a cone with a plane. When the plane is parallel to the base, a circle results. Planes which are perpendicular to the base yield hyperbolas, and planes parallel to the outer surface yield parabolas. All other sections are ellipses, except for special cases in which the plane is *on* the surface (a line or rectilinear orbit) or only through the vertex (a point).

foci coincide. The parabola has one focus at infinity (Fig. 1-6). The hyperbola has a branch associated with each focus, as shown in Fig. 1-7. Notice the divergent nature of the parabola and the hyperbola. Both of these orbits are considered ***open*** because the satellite does not repeat its position. ***Closed orbits*** are those in which a satellite ideally retraces its position over time.

Next, we determine the scale of the conic section by the value of the major axis. Other distance parameters to describe conics are the minor axis and the distance between the two foci. Typically, we use half of each of these parameters as the *semimajor axis*, a , *semiminor axis*, b , and *half the distance between the foci*, c , respectively. The semimajor axis is positive for both the circle and the ellipse, but it's infinite for a parabola and negative for a hyperbola. The sign of the semimajor axis determines the type of rectilinear orbit.

Some basic relations will help us develop conic sections. First, the sum or difference of the distance from each focus to a point on an elliptical or hyperbolic orbit is a constant. A simple means of creating an ellipse is to take a fixed-length string, two tacks, and a pencil to stretch out the string and move on a blank page. As the pencil stretches the string and

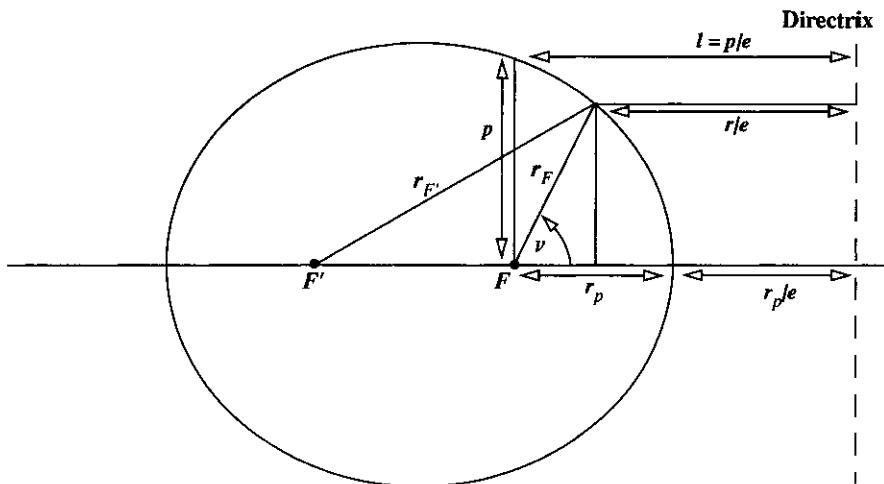


Figure 1-3. Geometry for Conic Sections. We create conic sections by recognizing that the sum of the distance from both foci to the orbit is constant. The ratio of the distance from a focus to the orbit and the distance to the directrix is also a constant called the eccentricity, e . The closest point in the orbit to the primary focus, F , is the radius of periapsis, r_p . The distance l is a standard quantity used to describe conic sections.

moves, the length from each focus changes, and the pencil traces out the ellipse. Mathematically (Prussing and Conway, 1993, 62),

$$\begin{aligned} r_{F'} + r_F &= \text{constant} = 2a \\ r_F - r_{F'} &= \text{constant} = 2ae \end{aligned} \tag{1-2}$$

Conic sections also use the distance from each focus to a stationary line called the **directrix**. The ratio of the distance from either focus to the orbit and the distance to the directrix is a constant called the eccentricity. The **eccentricity**, e , is a fixed constant for each type of conic section; it indicates the orbit's shape—its “roundness” or “flatness.” The eccentricity is *always* zero or positive, and its value determines the type of conic section for common orbits. It's 1.0 for parabolic and rectilinear orbits, less than 1.0 for ellipses, zero for circles, and greater than 1.0 for hyperbolas. The eccentricity is

$$e = \frac{c}{a} \tag{1-3}$$

Some authors use *flattening*, f , rather than eccentricity, to describe the shape. This value is

$$f = \frac{a-b}{a} \tag{1-4}$$

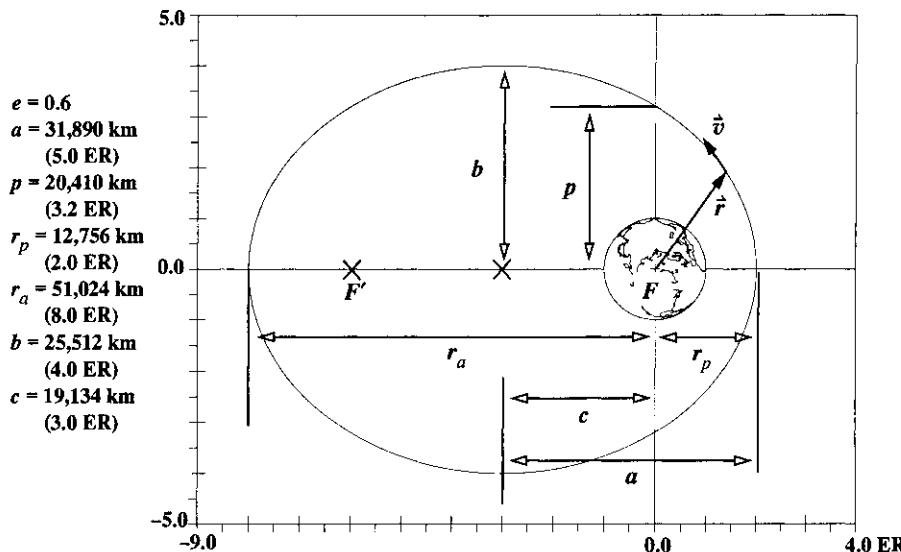


Figure 1-4. Elliptical Orbits. An elliptical orbit has two distinct foci with the primary focus, F , at the center of the Earth. The radius of apoapsis, r_a , and periapsis, r_p , denote the extreme points of the ellipse. The semimajor axis, a , and the semiminor axis, b , describe the shape of the orbit. Half the distance between the foci is c , and the semi-parameter, p , locates the orbit distance normal from the major axis.

Let's find an expression to convert between the flattening and eccentricity notations. Using Eq. (1-2) and Fig. 1-4, we find the distance from either focus to the point on the ellipse corresponding to the definition of b is given by a . Solve Eq. (1-3) for half the focal distance ($c = ae$), and substitute into the Pythagorean theorem ($b^2 + c^2 = a^2$). Solve for the semiminor axis to get

$$b = a\sqrt{1 - e^2} \quad (1-5)$$

This expression is useful because it relates any vertical part of an ellipse, y_e , to the corresponding part of a circle, y_c . The following conversion finds the vertical distance on the circle, y_c , and the vertical distance for the ellipse, y_e . We'll use this scaling factor in many derivations throughout the book.

$$y_e = y_c\sqrt{1 - e^2} \quad y_c = \frac{y_e}{\sqrt{1 - e^2}} \quad (1-6)$$

If we solve Eq. (1-5) for the square of the eccentricity,

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

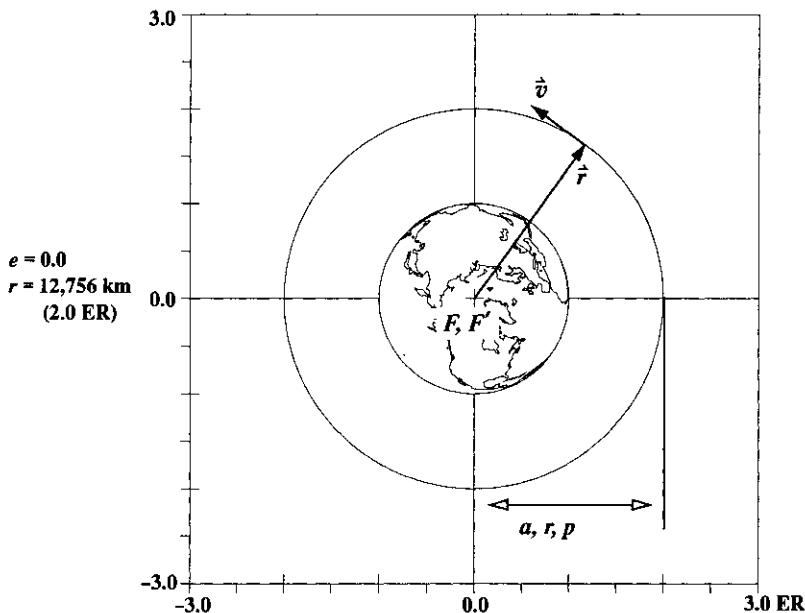


Figure 1-5. Circular Orbits. For circular orbits, the semimajor axis, semiparameter, and radius are always equal. One ER is equal to the Earth's radius. This is a special case of the ellipse, in which both foci coincide.

we get an alternate (and useful) equation for the eccentricity:

$$e = \frac{\sqrt{a^2 - b^2}}{a} \quad (1-7)$$

Notice we take only the positive value for eccentricity because eccentricity is *defined* as a positive value. Continuing the effort to develop an expression for flattening, use Eq. (1-4) and form $(2f - f^2)$:

$$2f - f^2 = \frac{2a - 2b}{a} - \frac{a^2 - 2ab + b^2}{a^2} = \frac{a^2 - b^2}{a^2}$$

Therefore, by Eq. (1-7),

$$e^2 = 2f - f^2 \quad (1-8)$$

Some authors use another expression, derived from Eq. (1-8), to relate the two notations:

$$1 - e^2 = (1 - f)^2 \quad (1-9)$$

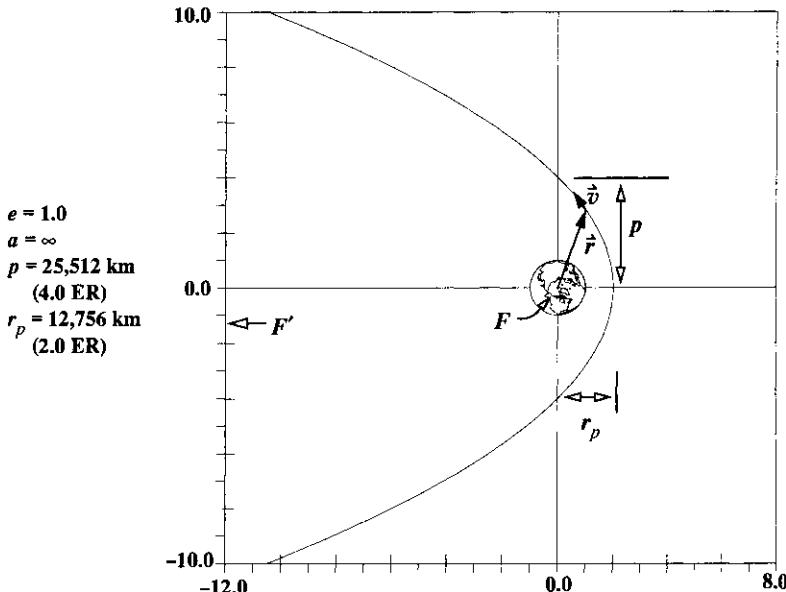


Figure 1-6. Parabolic Orbits. Parabolic orbits are considered open because they don't repeat. The semimajor axis is infinite. The second focus is at infinity.

Another useful variable is the *semiparameter*, sometimes called the *semilatus rectum* or simply the *parameter*. The *semiparameter*, p , is the distance from the primary focus to the orbit. It's measured perpendicular to the major axis except on the circle, for which it's equal to the semimajor axis. Typically, we use the semiparameter to describe the scale for a parabolic orbit because the value of the semimajor axis is infinite. The semiparameter is zero for all rectilinear orbits. Texts on analytical geometry usually define the semiparameter in terms of the semimajor and semiminor axes:

$$p = \frac{b^2}{a} \quad (1-10)$$

Using the scaling relation in Eq. (1-5), and simplifying gives us

$$p = a(1 - e^2) \quad (1-11)$$

The extreme points of an elliptical orbit are *apoapsis* and *periapsis*, representing the farthest and nearest points in the orbit, respectively. We change the endings of these words to indicate a particular planet or central body attracting the satellite—for the Sun, *aphelion* and *perihelion*; for the Earth, *apogee* and *perigee*; for the Moon, *aposelenium*, and *periselenium*, and so forth. Commonly, we refer to the distance of the closest point in the

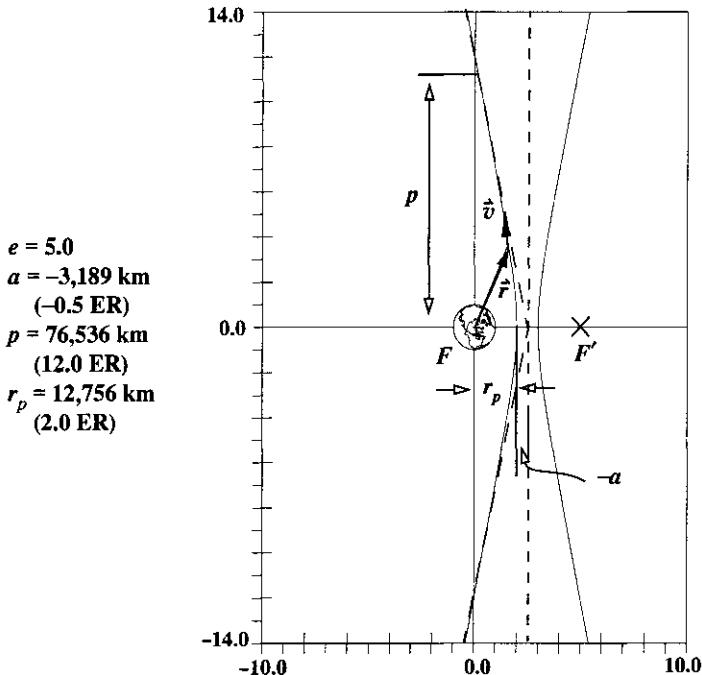


Figure 1-7. Hyperbolic Orbits. For a hyperbolic orbit, the semimajor axis is negative, and the eccentricity is larger than 1.0. Notice the rapid departure from the Earth with a *modest* eccentricity. The orbit approaches the dashed-line asymptotes. Hyperbolas occur in pairs called branches. The second branch is a mirror image of the first branch about the conjugate axis, shown as the dashed vertical line.

orbit to the primary focus as the *radius of periapsis*, r_p , and similarly, to the distance of the farthest point in the orbit as the *radius of apoapsis*, r_a , as shown in Fig. 1-8.

The two radii simply represent the satellite's distance from the center of the attracting body, or the primary focus. Don't confuse these values with the *altitudes*, which are from the *surface* of the attracting body to a location on the orbit.

The *true anomaly*, ν , is used to locate the satellite in the orbital plane and is the angular displacement measured from periapsis to the position vector along the direction of motion. As you can see in Fig. 1-8, the true anomaly will vary from 0° to 360° as the satellite travels through one revolution. True anomaly is *not* defined for circular orbits because they have no periapsis. We can overcome this limitation by selecting a point in the orbit to replace periapsis as the location for the initial measurement. Computer-software routines must account for this special case.

We'll define the *flight-path angle*, ϕ_{fpa} , as the angle from the *local horizontal* (the line perpendicular to the radius vector) *to* the velocity vector. Figure 1-9 shows the geometry for ϕ_{fpa} . You may see other definitions.

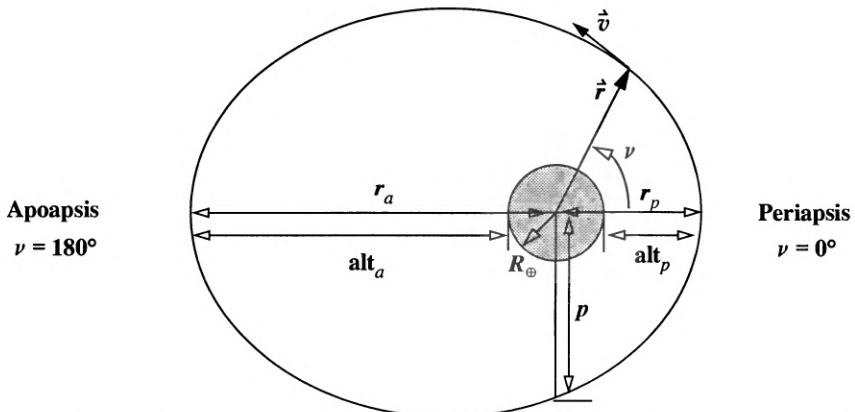


Figure 1-8. Radius of Apoapsis and Periapsis, as well as True Anomaly. Apoapsis occurs farthest from the planet; periapsis is closest. Notice the difference between altitudes and radius values. True anomaly, ν , locates the satellite's current position. We measure it in the direction of satellite motion from periapsis.

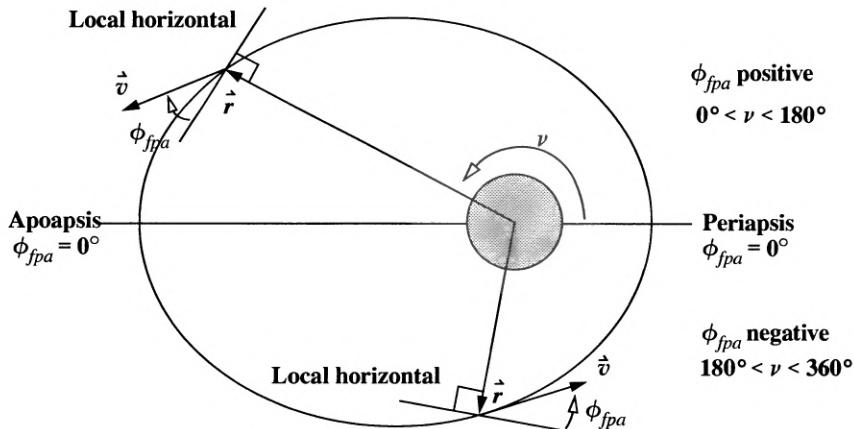


Figure 1-9. Geometry for the Flight-path Angle. The flight-path angle is always measured from the local horizontal to the velocity vector. It's always positive while the satellite travels from periapsis to apoapsis and negative for travel from apoapsis to periapsis. I've exaggerated the diagram for clarity.

As previously stated, for a circle the two foci coincide, so the satellite remains at a constant distance from the center. The velocity vector is always perpendicular to the radius vector, resulting in $\phi_{pacircle} = 0^\circ$ at all times. In the ellipse, ϕ_{fpa} may be positive or negative. From Fig. 1-9, notice ϕ_{fpa} is positive from periapsis to apoapsis and negative on the

return because we measure the angle in the *same* direction as the radius vector while the satellite travels towards apoapsis. In addition, ϕ_{fpa} is zero only at apoapsis and periapsis. The values of ϕ_{fpa} for parabolic and hyperbolic orbits may also be positive or negative depending on whether the satellite is heading away from or towards periapsis, respectively. It's zero at periapsis.

Table 1-1 summarizes values for parameters of basic conic sections. I'll develop these relations in greater detail in Sec. 2.2.

TABLE 1-1. Values for Conic Sections. This table shows characteristic values for the basic parameters. Sec. 4.2 discusses the limited values.

	a	e	p	ν	ϕ_{fpa}
Circle	$a = r$	$e = 0$	$p = r$	Undefined	$\phi_{fpa} = 0$
Ellipse	$r_p \leq a \leq r_a$	$0 < e < 1$	$r_p < p < 2r_p$	$0^\circ \leq \nu \leq 360^\circ$	$-90^\circ \leq \phi_{fpa} \leq 90^\circ$
Parabola	$a = \infty$	$e = 1$	$p = 2r_p$	Limited	Limited
Hyperbola	$a < 0$	$e > 1$	$p > 2r_p$	Limited	Limited
Rectilinear Ellipse	$r_p \leq a \leq r_a$	$e = 1$	$p = 0$		
Rectilinear Parabola	$a = \infty$	$e = 1$	$p = 0$		
Rectilinear Hyperbola	$a < 0$	$e = 1$	$p = 0$		

1.3 The Earth

Because satellites orbiting the Earth are a primary concern of this book, we must establish certain references to the Earth. We can generalize a set of fundamental parameters to specify the shape of any planet. These values allow us to specify locations, shape, the Earth's precise size, and the gravity field.

Five basic physical characteristics of the Earth are:

1. The equatorial radius, R_\oplus
2. The Earth's eccentricity or flattening, e_\oplus or f
3. The normalized, second-degree, zonal gravitational coefficient, $\bar{C}_{2,0}$
4. The rotational velocity, ω_\oplus
5. The gravitational parameter, μ

We can define more constants to represent the Earth's exact shape, but they're usually reserved for advanced applications. I've summarized these values with many conversions in the table on the inside front cover of this book.

The Earth's radius has been investigated for literally thousands of years. It has taken on many values, but "modern" efforts began to converge on an accepted value in the mid-1800s. JGM-2 (Sec. 1.8) defines the *mean equatorial radius of the Earth*, R_{\oplus} , to be

$$R_{\oplus} = 6378.1363 \text{ km}$$

Although it isn't considered a defining value, we can also calculate the semiminor axis of the Earth's ellipsoid to aid in formulas relating positions on the Earth's surface. The *semiminor axis*, b_{\oplus} , also called the polar axis, is

$$b_{\oplus} \cong 6356.751 \underline{600} 5 \text{ km}$$

Remember, this is a derived quantity, and the underline represents digits beyond the original accuracy in R_{\oplus} . Refer to Defense Mapping Agency (1987b, 3-14 to 3-15) for exact formulas and descriptions.

We must carefully convert between systems so we don't exceed the actual number of significant digits. Taff (1985, 71–73) shows several examples of this error. Unfortunately, the number of digits retained can introduce errors when trying to reverse a calculation. To preserve the original accuracy, it's usually good practice to keep as many decimal places as possible. Although these "rules" may seem contradictory, the extra digits permit more accurate reverse calculations. I'll use an underline to distinguish digits that fall outside the given accuracy.

Using Eq. (1-4), we can also derive the *flattening* and the *eccentricity of the Earth*, e_{\oplus} :

$$f = 0.003\ 352\ 813\ 178 = \frac{1}{298.257}$$

$$e_{\oplus} = 0.081\ 819\ 221\ 456$$

We can define the gravitational attraction resulting from the irregular distribution of the Earth's mass using a complex potential function, which I'll discuss fully in Sec. 7.6.1. For now, let's define the constant which gives us an idea of a planet's overall nonsphericity. Specifically, the ***normalized, second-degree, zonal gravitational coefficient***, $C_{2,0}$, represents the Earth's equatorial bulge. [Eq. (7-19) discusses normalization]. It also begins the series of parameters defining the Earth's precise gravitational shape. This constant is the second defining parameter.

$$\bar{C}_{2,0} = -484.165\ 466\ 3 \times 10^{-6}$$

This value *doesn't* include the permanent tidal deformation (Sec. 7.6.5 and Sec. 8.5.3). The conventional (unnormalized) form of this equation uses the following relation:

$$C_{l0} = \bar{C}_{l0} \sqrt{2l+1} \quad (1-12)$$

Notice this is simply a special case of general formulas which I'll present shortly. Thus the first zonal coefficient is

$$C_{2,0} = -0.001\ 082\ 626\ 9$$

Some authors use a “*J*” notation for the zonal harmonics, which actually results from a discussion in Sec. 7.6.1, where we’ll explore all these symbols. The correct use of the “*J*” notation is that it’s the negative of the *C* notation for zonal harmonics.

$$-C_{l0} = J_l \quad (1-13)$$

Though this is merely a sign change, it can be very significant in forming the perturbation equations discussed in Sec. 8.6. The second-order zonal harmonic is $J_2 = 0.001\ 082\ 626\ 9$.

The Earth’s *rotational velocity*, ω_{\oplus} , is quite often assumed to be constant, and indeed it seemed to be so for many years, given the limits of measuring devices. Most applications use the adopted constant value for the Earth’s rotation:

$$\omega_{\oplus} = 7.292\ 115 \times 10^{-5} \pm 1.5 \times 10^{-12} \text{ rad/s}$$

With the aid of precise clocks invented around the end of the 19th century, observers began to detect small variations in the value of the Earth’s rotational velocity. As we’ll see in Sec. 1.6, the perturbative effects of the Sun, Moon, and other planets, as well as the Earth’s nonspherical nature, all induce small variations in the magnitude and direction of the Earth’s angular velocity. We now recognize several types of variability in the Earth’s rotation with secular, irregular, and seasonal attributes. According to the Defense Mapping Agency (1987b, 3–13), from 1967–1985 the value ranged from $7.292\ 114\ 832 \times 10^{-5}$ to $7.292\ 115\ 099 \times 10^{-5}$ rad/s.

The final parameter is the gravitational parameter, which we’ve already discussed. It’s important to recognize that the mass of the atmosphere contributes to μ . Defense Mapping Agency (1987a, 3–2 to 3–4) lists separate values for μ that exclude the atmosphere.

1.3.1 Location Parameters

It’s now possible to introduce concepts of latitude and longitude. We define these values as *terrestrial* parameters because they refer to the Earth’s equator and prime meridian and describe locations on the Earth. Consider Fig. 1–10, which shows these location parameters.

Terrestrial latitude, ϕ_{gc} , and *longitude*, λ , are familiar to us through common use. The “*gc*” subscript on the latitude is a descriptor used to delineate geocentric latitude (see Sec. 1.3.2). *Longitude* is an east-west angular displacement measured positive to the east from a primary meridian in a plane. Measuring positive to the east is an arbitrary convention agreed upon at the International Meridian Conference in Washington D.C. in 1884 (*Explanatory Supplement*, 1961, 7). But the United States didn’t officially begin using this convention until 1981 in the *Astronomical Almanac*, so we must carefully check older programs. A *meridian* is a constant line of longitude over the Earth. The Royal Observatory at Greenwich is the *prime meridian* for the Earth, and the equator is the reference plane for measuring latitude. These reference lines are fixed to the Earth, so they’re ideal for

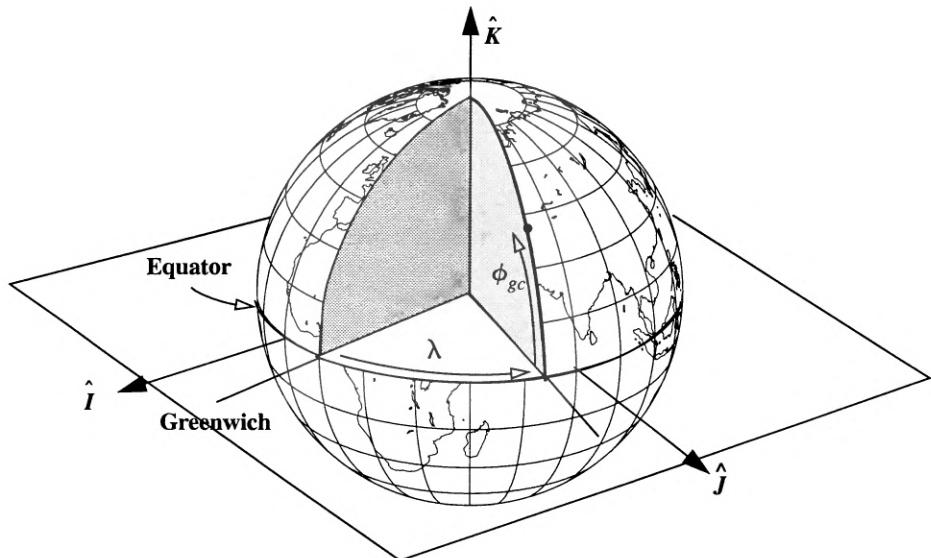


Figure 1-10. Using Latitude and Longitude to Measure Location. Terrestrial latitude and longitude, ϕ_{gc} and λ , use the Earth's equator. The Greenwich meridian is the zero point for longitude, and the Earth's equator is the reference point for latitude.

locating sites on the Earth's surface. Longitude may take on values from 0° to 360° when measured east or from 0° to $\pm 180^\circ$ if measured east and west. **Latitude** is the north-south measurement from the reference plane (the Earth's equator, for instance), with values from 0° to $\pm 90^\circ$. It's positive in the Northern Hemisphere.

1.3.2 Shape of the Earth

The Earth's shape is very important to many astrodynamical studies—locating ground stations, remote sensing, geodesy, oceanography, plate tectonics, viewing constraints, and many others. We often use a spherical Earth, which is the simplest representation with the same center of gravity and mass as the real Earth. Although it was a surprise to some ancient people, the Earth isn't a perfect sphere. The spherical model is sufficient for many studies, but several other models better represent the Earth's shape. In particular, ellipsoids of revolution (spheroids) can be very accurate. In general, ellipsoids have three axes (a, b, c) that are different. Ellipsoids of revolution (spheroids) have two equal axes. **Prolate** spheroids result from the revolution of an ellipse around its major axis (a, b, b), and **oblate** spheroids use the minor axis for rotation (a, a, b). Simple ellipsoidal models aren't always the best—a triaxial ellipsoid (a, b, c) better represents the Moon.

Much of the recent data has come from satellites, but many early studies showed the Earth to be roughly elliptical. The flattening notation introduced earlier gives us a way to explore the ellipsoidal concept shown in Fig. 1-11. Keep in mind the true scale of the

ellipsoidal Earth model as we work through this section; we've drawn many diagrams with greatly exaggerated scales for clarity. The ellipsoidal model is widely used for semi-precise calculations. It assumes the Earth's shape is approximately an *oblate spheroid* (a, b) with its semimajor axis equal to the equatorial radius and semiminor axis equal to the polar radius.

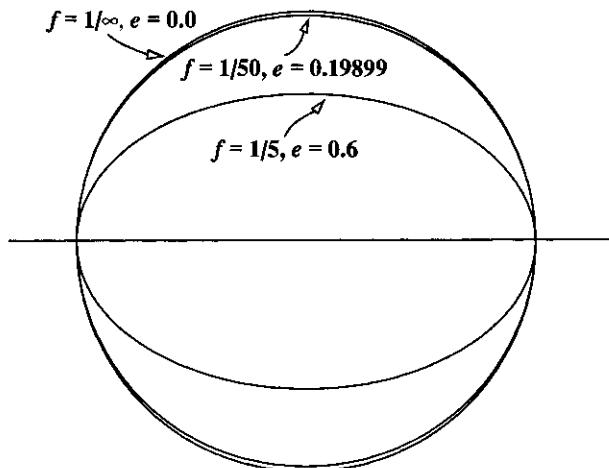


Figure 1-11. Ellipsoidal Representation of the Earth. We can approximate the Earth's shape with a flattening of about 1/298.3 to account for the 21 km difference in equatorial and polar radii. This flattening would be indistinguishable from the outer circle at this scale. Notice how closely the ellipse approaches a circle even with $f = 1/50$. In JGM-2, the eccentricity of the Earth is 0.081 819 221 456.

This reference ellipsoid is a good approximation of the hypothetical surface referred to as mean sea level. The actual mean-sea-level surface, excluding small atmospheric effects, is called the *geoid*, and it deviates from the reference ellipsoid because of the uneven distribution of mass in the Earth's interior. The *geoid* is a geopotential surface—a plumb-bob will hang perpendicular to it at every point.

Although an oblate Earth introduces no unique problems with longitude, it creates problems with latitude. Consider the Earth's cross-section in Fig. 1-12 showing the types of latitude commonly used in astrodynamics. *Geocentric latitude*, ϕ_{gc} , is the angle measured at the Earth's center from the plane of the equator to the point of interest. The *geodetic latitude*, ϕ_{gd} , is the angle between the equatorial plane and the normal to the surface of the ellipsoid. The latitude on most maps is geodetic latitude, which is normal to the reference ellipsoid. Although not shown in the figure, *astronomic latitude*, ϕ_{as} , is defined as the angle between the equatorial plane and the geoid; thus it accounts for all local irregularities such that a plumb-bob will hang perpendicular to the geoid. This angle is very close to geodetic latitude. The difference between astronomic and geodetic latitude is

called the **deflection of the vertical**. It represents the deviation of a plumb line from the ellipsoid normal. The differences between geocentric, geodetic, and astronomic latitude are usually very small, but some applications require the increased accuracy and must account for the variation.

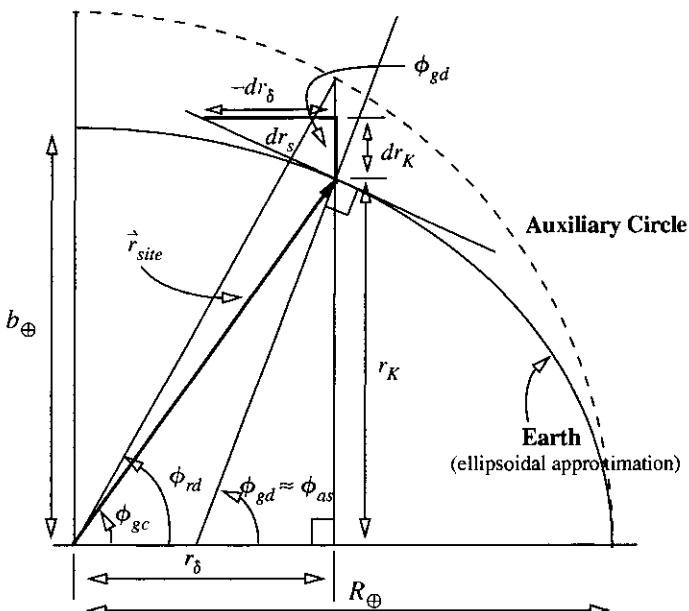


Figure 1-12. Latitude Geometry (exaggerated scale). The geodetic latitude, ϕ_{gd} , makes an angle perpendicular to the surface and the equatorial plane, whereas the geocentric latitude, ϕ_{gc} , is referenced to the center of the Earth. Astronomic latitude, ϕ_{as} , is very close to ϕ_{gd} . The center of the ellipsoid differs slightly for ϕ_{gd} and ϕ_{as} . Reduced latitude, ϕ_{rd} , is used only to derive relations.

The main problem involves transforming between geocentric and geodetic latitudes and computing the geocentric radius using the geodetic latitude. We determine astronomical latitude by observation. Station locations, such as those found on standard maps, are usually given in geodetic coordinates. We require geocentric values for calculations involving coordinate transformations (Sec. 1.5.1). Thus, we must be able to represent position vectors in terms of the latitudes, as well as to transform between each type.

Begin by examining the representation of a position vector (or station) in terms of the various latitudes. This process uses an angle called the *reduced latitude*. Measurement of this angle uses an auxiliary circle around the ellipse (see Fig. 1-12). The **reduced latitude**, ϕ_{rd} , is the geocentric angle measured to a point on the auxiliary circle resulting from the vertical projection of the site. The position vector of the site is given with geocentric components as

$$\vec{r}_{site} = \begin{bmatrix} r \cos(\phi_{gc}) \cos(\lambda) \\ r \cos(\phi_{gc}) \sin(\lambda) \\ r \sin(\phi_{gc}) \end{bmatrix}$$

Because we're interested only in the relationship between latitudes, we can consider the position as a two-dimensional problem independent of λ . In Eq. (1-14), we use r_δ as the horizontal component along the semimajor axis and r_K as the vertical component along the semiminor axis.

$$\begin{aligned} r_\delta &= \sqrt{r_I^2 + r_J^2} = r_{site} \cos(\phi_{gc}) \\ r_K &= r_{site} \sin(\phi_{gc}) \end{aligned} \quad (1-14)$$

We can also find the station coordinates in terms of the reduced latitude. We find the horizontal component similar to Eq. (1-14). To discover the vertical component, we need to use the scaling transformation [Eq. (1-6)], which allows us to determine the vertical component on an ellipse or a circle. The results are

$$\begin{aligned} r_\delta &= R_\oplus \cos(\phi_{rd}) \\ r_K &= R_\oplus \sqrt{1 - e_\oplus^2} \sin(\phi_{rd}) \end{aligned} \quad (1-15)$$

We can express the site position's magnitude as

$$\begin{aligned} r_{site} &= \sqrt{r_\delta^2 + r_K^2} \\ &= R_\oplus \sqrt{1 - e_\oplus^2 \sin^2(\phi_{rd})} \end{aligned}$$

Finding the station coordinates using the geodetic latitude is more difficult. Although the geodetic latitude is related to the site position, we can't see right away how to directly determine an expression. Notice, however, if we take the differential of Eq. (1-15), we obtain the slope or rate of change of the site latitude in terms of the reduced latitude. This is the first step in determining an expression for the site in terms of geodetic latitude.

$$\begin{aligned} dr_\delta &= -R_\oplus \sin(\phi_{rd}) d\phi_{rd} \\ dr_K &= R_\oplus \sqrt{1 - e_\oplus^2} \cos(\phi_{rd}) d\phi_{rd} \end{aligned}$$

Likewise, we can find the hypotenuse of the differential triangle:

$$dr_s = \sqrt{dr_\delta^2 + dr_K^2} = R_\oplus \sqrt{1 - e_\oplus^2 \cos^2(\phi_{rd})} d\phi_{rd}$$

Now we can use these slope values to find expressions for the geodetic latitude. Notice the connection between the components of the differentials and the geodetic latitude drawn at the site in Fig. 1-12.

$$\sin(\phi_{gd}) = -\frac{dr_\delta}{dr_s} = \frac{\sin(\phi_{rd})}{\sqrt{1-e_\oplus^2 \cos^2(\phi_{rd})}} \quad \cos(\phi_{gd}) = \frac{dr_k}{dr_s} = \frac{\sqrt{1-e_\oplus^2} \cos(\phi_{rd})}{\sqrt{1-e_\oplus^2 \cos^2(\phi_{rd})}}$$

If we multiply the first equation by $\sqrt{1-e_\oplus^2}$, then square it, and square the second equation, we get

$$\begin{aligned} (1-e_\oplus^2) \sin^2(\phi_{gd}) &= \frac{(1-e_\oplus^2) \sin^2(\phi_{rd})}{1-e_\oplus^2 \cos^2(\phi_{rd})} \\ \cos^2(\phi_{gd}) &= \frac{(1-e_\oplus^2) \cos^2(\phi_{rd})}{1-e_\oplus^2 \cos^2(\phi_{rd})} \end{aligned}$$

Now let's add these two expressions:

$$-e_\oplus^2 \sin^2(\phi_{gd}) + \sin^2(\phi_{gd}) + \cos^2(\phi_{gd}) = \frac{(1-e_\oplus^2) \sin^2(\phi_{rd}) + (1-e_\oplus^2) \cos^2(\phi_{rd})}{1-e_\oplus^2 \cos^2(\phi_{rd})}$$

and solve them for the denominator of the right-hand side:

$$1-e_\oplus^2 \cos^2(\phi_{rd}) = \frac{1-e_\oplus^2}{1-e_\oplus^2 \sin^2(\phi_{gd})}$$

From the equations of the geodetic latitude above, we'll solve for the reduced latitude term in the denominator and substitute the quantity above:

$$\begin{aligned} \sin(\phi_{rd}) &= \sin(\phi_{gd}) \frac{\sqrt{1-e_\oplus^2}}{\sqrt{1-e_\oplus^2 \sin^2(\phi_{gd})}} \\ \cos(\phi_{rd}) &= \frac{\cos(\phi_{gd})}{\sqrt{1-e_\oplus^2 \sin^2(\phi_{gd})}} \end{aligned}$$

Finally, substituting these expressions for the reduced latitude into Eq. (1-15), we get

$$r_{\delta} = \left(\frac{R_{\oplus}}{\sqrt{1 - e_{\oplus}^2 \sin^2(\phi_{gd})}} \right) \cos(\phi_{gd})$$

$$r_K = \left(\frac{R_{\oplus}(1 - e_{\oplus}^2)}{\sqrt{1 - e_{\oplus}^2 \sin^2(\phi_{gd})}} \right) \sin(\phi_{gd})$$

These formulas are exact for our ellipsoidal surface, but we know of surface features which may cause the site to vary significantly from the reference ellipsoid. For example, Death Valley (36.34°N, 117.0°W) is 86 m below sea level; yet, just over 100 km away, Mt. Whitney (36.34°N, 118.18°W) is 4417 m above sea level. It turns out these variations don't cause too much difficulty because we can add the components of the altitude into the component equations. We use the term h_{ellp} for the height; the subscript ($ellp$) specifies that the elevation is related to the ellipsoid. I'll discuss other areas we need to consider for highly accurate studies in the next section on geodesy.

Now we can define two auxiliary quantities (obtained from geometrical properties of an ellipse) and write the components for the position of a site:

$$C_{\oplus} = \frac{R_{\oplus}}{\sqrt{1 - e_{\oplus}^2 \sin^2 \phi_{gd}}} \quad S_{\oplus} = \frac{R_{\oplus}(1 - e_{\oplus}^2)}{\sqrt{1 - e_{\oplus}^2 \sin^2(\phi_{gd})}}$$

$$r_{\delta} = (C_{\oplus} + h_{ellp}) \cos(\phi_{gd}) \quad r_K = (S_{\oplus} + h_{ellp}) \sin(\phi_{gd}) \quad (1-16)$$

$$r_{site} = \sqrt{r_{\delta}^2 + r_K^2}$$

Notice the two coefficients differ only by the numerator, thus enabling some computational efficiency when programming on the computer. Also recognize that C_{\oplus} is more commonly known as the *radius of curvature in the meridian* and often has the symbol N_{ϕ} .

We can also use Fig. 1-12 to find expressions for a position's magnitude in terms of the geocentric latitude. Although not used as often, this value has certain advantages in some situations. Begin by determining the sine and cosine expressions for the geocentric latitude [as in Eq. (1-15)].

$$\sin(\phi_{gc}) = \frac{r_K}{r_{site}} = \frac{\sqrt{1 - e_{\oplus}^2} \sin(\phi_{rd})}{\sqrt{1 - e_{\oplus}^2 \sin^2(\phi_{rd})}} \quad \cos(\phi_{gc}) = \frac{r_{\delta}}{r_{site}} = \frac{\cos(\phi_{rd})}{\sqrt{1 - e_{\oplus}^2 \sin^2(\phi_{rd})}}$$

Now solve for the sine expression in terms of the reduced latitude:

$$\sin(\phi_{rd}) = \sin(\phi_{gc}) \sqrt{\frac{1 - e_{\oplus}^2 \sin^2(\phi_{rd})}{1 - e_{\oplus}^2}}$$

Next, consider the following term:

$$\sqrt{\frac{1}{1 - e_{\oplus}^2 \cos^2(\phi_{gc})}}$$

Substitute the square of the cosine expression, $\cos(\phi_{gc})$, using reduced latitude and obtaining a common denominator:

$$\sqrt{\frac{1}{1 - e_{\oplus}^2 \cos^2(\phi_{gc})}} = \sqrt{\frac{1 - e_{\oplus}^2 \sin^2(\phi_{rd})}{1 - e_{\oplus}^2 \sin^2(\phi_{rd}) - e_{\oplus}^2 \cos^2(\phi_{rd})}}$$

Replace the cosine-squared term with $(1 - \sin^2(\phi_{rd}))$ and simplify:

$$\sqrt{\frac{1}{1 - e_{\oplus}^2 \cos^2(\phi_{gc})}} = \sqrt{\frac{1 - e_{\oplus}^2 \sin^2(\phi_{rd})}{1 - e_{\oplus}^2}}$$

We can now rewrite the sine of the reduced latitude as

$$\sin(\phi_{rd}) = \frac{\sin(\phi_{gc})}{\sqrt{1 - e_{\oplus}^2 \cos^2(\phi_{gc})}}$$

so we can use this result directly in Eq. (1-15) to find the site's vertical component. Squaring and using Eq. (C-8) gives us an expression for the horizontal component:

$$\cos(\phi_{rd}) = \frac{\sqrt{1 - e_{\oplus}^2} \cos(\phi_{gc})}{\sqrt{1 - e_{\oplus}^2 \cos^2(\phi_{gc})}}$$

which we then use with Eq. (1-15) to complete the result:

$$\begin{aligned} r_{\delta} &= \left(\frac{R_{\oplus} \sqrt{1 - e_{\oplus}^2}}{\sqrt{1 - e_{\oplus}^2 \cos^2(\phi_{gc})}} + h_{ellp} \right) \cos(\phi_{gc}) \\ r_K &= \left(\frac{R_{\oplus} \sqrt{1 - e_{\oplus}^2}}{\sqrt{1 - e_{\oplus}^2 \cos^2(\phi_{gc})}} + h_{ellp} \right) \sin(\phi_{gc}) \end{aligned} \quad r_{site} = \sqrt{r_{\delta}^2 + r_K^2} \quad (1-17)$$

We also can define transformations between the geocentric and geodetic latitudes by finding an expression for the tangent of the reduced latitude, using both the geodetic and geocentric latitudes. The intermediate values for the reduced latitude in terms of the geodetic latitude permit the following expression for the tangent:

$$\tan(\phi_{rd}) = \frac{\sin(\phi_{rd})}{\cos(\phi_{rd})} = \frac{\sin(\phi_{gd}) \sqrt{1-e_\oplus^2}}{\cos(\phi_{gd})} = \tan(\phi_{gd}) \sqrt{1-e_\oplus^2}$$

Likewise, we can find expressions for the reduced latitude from Fig. 1-12 in terms of the geocentric latitude and then form an equivalent expression for the tangent:

$$\tan(\phi_{rd}) = \frac{\sin(\phi_{rd})}{\cos(\phi_{rd})} = \frac{\sin(\phi_{gc})}{\sqrt{1-e_\oplus^2} \cos(\phi_{gc})} = \frac{\tan(\phi_{gc})}{\sqrt{1-e_\oplus^2}}$$

Setting these quantities equal, we get

$$\tan(\phi_{gd}) \sqrt{1-e_\oplus^2} = \frac{\tan(\phi_{gc})}{\sqrt{1-e_\oplus^2}}$$

This equation permits transformation between latitude types in either direction, depending on the given information. Thus,

$$\begin{aligned}\tan(\phi_{gd}) &= \frac{\tan(\phi_{gc})}{1-e_\oplus^2} \\ \tan(\phi_{gc}) &= (1-e_\oplus^2) \tan(\phi_{gd})\end{aligned}\tag{1-18}$$

Remember that Eq. (1-18) is valid only for locations on the Earth's surface. Section 3.10 shows a way to convert between latitudes above this surface. The following simple example will tie together this section's concepts.

▼ Example 1-1: Determining Site Coordinates.

GIVEN: Mt. Evans, Colorado. $\phi_{gd} = 39.586\ 667^\circ$, $\lambda = -105.640^\circ$, $H_{MSL} = 4347.667\text{ m}$ (14,264 ft)

FIND: Site coordinates for the observatory near the top of the mountain.

Begin by assuming that the height above sea level is equal to the height above the ellipsoid. Now,

$$C_\oplus = \frac{R_\oplus}{\sqrt{1-e_\oplus^2 \sin^2(\phi_{gd})}} = \frac{6378.1363}{\sqrt{1-0.006\ 694\ 385 \sin^2(39.586\ 667^\circ)}} = 6386.823\ 373\text{ km}$$

$$S_\oplus = \frac{R_\oplus (1-e_\oplus^2)}{\sqrt{1-e_\oplus^2 \sin^2(\phi_{gd})}} = \frac{6378.1363 (1-0.006\ 694\ 385)}{\sqrt{1-0.006\ 694\ 385 \sin^2(39.586\ 667^\circ)}} = 6344.067\ 518\text{ km}$$

Now determine the site's coordinates.

$$r_\delta = (C_\oplus + h_{ellp}) \cos(\phi_{gd}) = (6386.823\ 373 + 4.347\ 667\ 2) \cos(39.586\ 667^\circ)$$

$$r_K = (S_\oplus + h_{ellp}) \sin(\phi_{gd}) = (6344.067\ 518 + 4.347\ 667\ 2) \sin(39.586\ 667^\circ)$$

$$r_\delta = 4925.429\ 802\ 6\text{ km} = 0.772\ 236\ 5\text{ ER}$$

$$r_K = 4045.493\ 742\ 6\text{ km} = 0.634\ 275\ 2\text{ ER}$$

1.3.3 Geodesy

Discussions of geodesy are usually tied to forming gravitational coefficients, surface topography, and tides. We'll discuss the precise mathematical representation of the Earth's gravitational field in Chap. 7. Here, I'm applying it to a site's exact coordinates.

The concept of height relies on *geops*—surfaces at which the gravity is equal at all points. This is also called an *equipotential* surface. Many people are familiar with the standard geop called the *geoid*, also known as the potential of *Mean Sea Level*, *MSL*, where gravity acts perpendicular to the geoid at all points. Unfortunately, determining *MSL* isn't as easy as we might think. Tide-gauge measurements are commonly used at various locations around a region and averaged over time. Because the Moon is the main cause of the tides, measurements are collected for about 18 years to account for periodic variations in the Moon's motion. Still, the actual sea-level surface varies widely. This causes difficulties which I'll discuss shortly. Figure 1-13 shows the relations between the different types of geopotential surfaces.

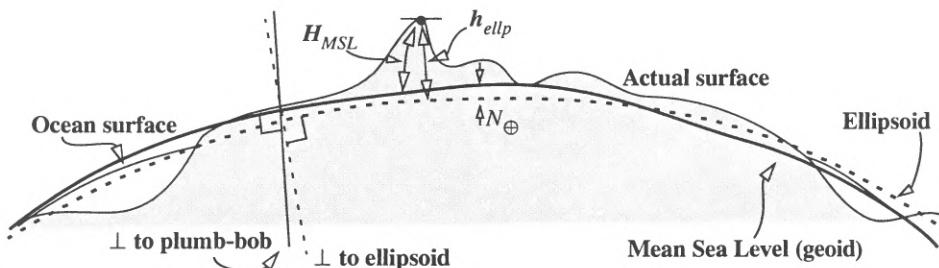


Figure 1-13. Geopotential Reference Surfaces. The height above sea level, H_{MSL} , and the geoid's undulation, N_{\oplus} , determine the ellipsoidal height, h_{ellp} . N_{\oplus} is typically very small. Notice that the ellipsoid is regular whereas the geoid has many small variations. Also, H_{MSL} is measured along a curved surface as shown.

Considering Mean Sea Level (*MSL*) introduces some more definitions. The *geoid's undulation*, N_{\oplus} , is the geoid's height above the ellipsoid, whereas the actual height above the geoid (*MSL*) is called the *orthometric height*, H_{MSL} . We're all familiar with orthometric heights because they're used for signs and maps. They are usually developed through surveying and geodetic techniques. All these heights are referenced to a starting set of values called a *vertical datum*. Determining these base values depends on the method used to obtain the location of the orthometric height, H_{MSL} —usually with tide-gauges. Thus, vertical datums are defined throughout the world. The main ones in use are the North American Datum [NAD-27 (83)]—so labeled because it was completed in 1927 and updated in 1983, South American Datum (SAD-69), European Datum (ED-50), Tokyo Datum (TD), and Australian Geodetic Datum [AGD-66 (84)]. Unfortunately, having several datums introduces complexity because we can use each datum to define a distinct Earth ellipsoid.

If we simply combine the datums, their centers don't coincide, and we need a transformation—often a form of the *Molodensky formulas* (See Defense Mapping Agency, 1987b, 7-5 to 7-28). The *ellipsoidal height*, h_{ellp} , is similar to the orthometric height, but it's measured perpendicular to the ellipsoid (as shown in Fig. 1-13) and differs somewhat. We can now define the geoid's undulation at any location (Defense Mapping Agency, 1987b, 6-4):

$$N_{\oplus} \equiv h_{\text{ellp}} - H_{\text{MSL}}$$

$$N_{\oplus} = \frac{\mu}{r g_{th}} \sum_{l=2}^{\infty} \sum_{m=0}^l \left(\frac{R_{\oplus}}{r} \right)^l \bar{P}_{lm} [\sin(\phi_{gc})] \left\{ \bar{C}_{lm}^* \cos(m\lambda) + \bar{S}_{lm} \sin(m\lambda) \right\} \quad (1-19)$$

Note here that r is the distance to the ellipsoid from the Earth's center and \bar{C}_{lm}^* indicates the even zonal harmonics, modified by subtracting their geometric (nonrotating) value (Defense Mapping Agency, 1987b, 6-7 to 6-8). We usually limit the number of harmonics used, and \bar{P}_{lm} is a normalized Legendre polynomial (Sec. 7.6.1). Eq. (1-19) is similar to the gravitational potential [Eq. (7-17)], but we add a centrifugal term and use *Brun's theorem* [Lambeck (1988, 20)] to convert from nominal values. Eq. (1-19) uses the gravitational coefficients of the Earth and the theoretical gravity, g_{th} , in Eq. (1-20). Because the geoid's undulation takes on values between +76 to -102 m, you use this formula only for very precise analysis. See Defense Mapping Agency (1987b, 6-4) for more information. Calculating a few sample values should convince you that the geoid's undulation is very small; thus, using the height above the ellipsoid for the orthometric height is reasonably accurate. Because this process is so complex, N_{\oplus} is tabulated in contour charts for grids of $10^\circ \times 10^\circ$, $1^\circ \times 1^\circ$, and so on, to provide increased speed. Specific locations not contained on a grid are found using interpolation techniques.

1.3.4 Gravitational Model

If we assume a spherical Earth of uniform density and a point mass at a given position above the Earth's surface, Newton's inverse square law [Eq. (1-1)] predicts the attraction should decrease based on the reciprocal of the square of increasing distance of the point mass from the center of the Earth. Figure 1-14 shows the magnitude of the acceleration due to gravity versus position. Although we commonly say satellites in orbit experience "zero-gravity," this is not true. Of course, we think gravity is zero because no normal force, such as the Earth's surface, is supporting the object. The force of gravity is still present, however, and actually causes the satellite to fall continually around the Earth.

A simple example demonstrates the phenomenon. Suppose a cannon rests on the Earth and fires a shell. If it's quite powerful, the shell may travel 20 km. With a higher initial horizontal velocity, the shell could travel perhaps 25 km. If we increase the initial velocity to about 7900 m/s, in the time it takes the shell to hit the Earth, it's no longer over the Earth, but over the "edge" of the Earth, and still falling. Firing the shell at ever-higher velocities will eventually place it in orbit!

The preceding discussion assumed a spherical Earth. Because the Earth isn't spherical, we may also approximate the force of gravity by recognizing differences based on an observer's location, ϕ_{gd} . The adopted value of *theoretical gravity*, g_{th} , at the Earth's surface is a function of the reference ellipsoid and the latitude of the observer. From the Defense Mapping Agency (1987b, 4-16), the closed-gravity formula of Somigliana is

$$g_{th} = g_{equator} \left\{ \frac{1 + k_g \sin^2(\phi_{gd})}{\sqrt{1 - e_\oplus^2 \sin^2(\phi_{gd})}} \right\} \quad g_{equator} = 9.780\,326\,771\,4 \text{ m/s}^2$$

$$k_g = \frac{b_\oplus g_{pole}}{R_\oplus g_{equator}} - 1 = 0.001\,931\,853\,863\,9 \quad g_{pole} = 9.832\,186\,368\,5 \text{ m/s}^2 \quad (1-20)$$

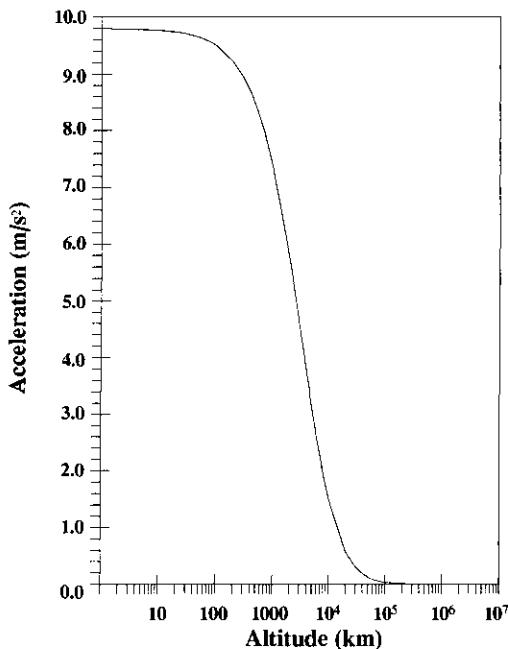


Figure 1-14. Magnitude of Gravitational Acceleration. Notice how high a satellite must be before it is in “zero gravity.” It may seem confusing that objects float on board a Space Shuttle (altitude ≈ 250 km), which has almost 90% of normal gravity, but the gravity is balanced with centrifugal forces, thus giving the perception of zero gravity.

Gravitational Coefficients

Finally, we can introduce another model of the gravitational field that's valid at *any* point. Unlike Eq. (1-20), this model assumes a complex mass distribution model of the Earth

(see Sec. 7.6.1). The general expression assumes an infinite number of gravitational coefficients, C_{lm} and S_{lm} :

$$U = \frac{\mu}{r} \sum_{l=0}^{\infty} \sum_{m=0}^l \left(\frac{R_{\oplus}}{r} \right)^l P_{lm} [\sin(\phi_{sat})] \left\{ C_{lm} \cos(m\lambda_{sat}) + S_{lm} \sin(m\lambda_{sat}) \right\}$$

We use this formula only for precise astrodynamical analyses, which we'll discuss later in Chap. 7 (geodesy, oceanography, precise location, and so on). The double summation of spherical-harmonic coefficients uses "l" and "m" indices as *degree* and *order*, respectively. The second-degree coefficient describes the contribution to the potential due to a generally elliptical shape, whereas the remaining parameters help represent the geopotential due to the central body's specific shape. Notice the similarity to Eq. (1-19).

We get data from which the coefficients are derived from two main sources: measurements taken at the Earth's surface (terrestrial) and the observed motion of satellites. Historically, only terrestrial measurements were available. The *Ohio State University* model (OSU-91A) is complete through degree and order 360. Models based on satellite observations are sometimes smaller. Models include the *Joint Gravity Model* (JGM-2, 70×70), the Defense Mapping Agency's *World Geodetic Survey* (WGS-84, 41×41), the *Goddard Earth Model* (GEM, 50×50), and others. A larger and combined (WGS/GEM/JGM/OSU) model is due in late 1996. Typically, accuracy requirements drive the maximum degree and order. For example, 4×4 fields are often adequate for deep-space orbits, whereas some low-Earth satellites need 50×50 fields. Increasing the field size isn't always enough. We usually need to consider other factors, which we'll discuss in Chap. 7 and Chap. 8. Table D-1 and Table D-2 in Appendix D show representative coefficient values from the JGM-2 model (Vetter et al., 1994, McCarthy, 1992, and Nerem et al., 1994).

1.4 Coordinate Systems

One of the first requirements for describing an orbit is to define a suitable reference frame, which usually means finding an appropriate inertial coordinate system. We define a rectangular coordinate frame by specifying its *origin*, *fundamental plane*, and the *preferred direction*. In addition, we must specify the *sense*, or the positive direction. Most systems have a right-handed sense—that is, the positive directions of each axis form an orthogonal triplet that's oriented with the thumb, index, and middle finger of the right hand. Some systems are designed with a left-handed sense of direction. This may be confusing, so I'll say when a system is left-handed. We use three unit vectors to represent three orthogonal axes and express any other vector in the coordinate system as a linear combination of the base unit vectors. The three unit vectors also form a *basis*. This book uses certain letters to label the systems; although the letters are somewhat similar, I've tried to make coordinate systems distinct and consistent.

The origin of the coordinate system is important in naming each type. In general, there are three common designations: an object's center, the system's center of mass, or **barycenter**, and a rotating system using the barycenter, called the **synodic** system.

Some concepts transcend all coordinate systems. In particular, because of the great distances to all natural celestial objects compared to the Earth's size, these objects appear to be fixed to, or move on the inner surface of, a **celestial sphere** with the observer at the center (Fig. 1-15). **Great circles** are defined as the intersection of the celestial sphere with any plane passing through the center of the sphere. The **celestial poles** (north and south) result from the intersection of the Earth's rotational axis and the celestial sphere. The **celestial equator** extends the Earth's equator onto the celestial sphere. **Hour circles** are great circles that are perpendicular to the celestial equator.

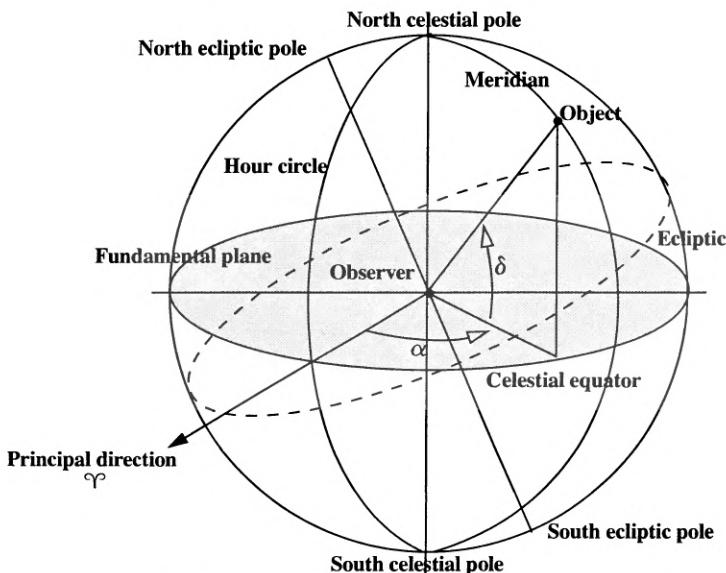


Figure 1-15. Geometry of the Celestial Sphere. The celestial sphere is based on an observer's perceived view of objects in space. A meridian, or hour circle, is any circle that passes through the observer at the center. I've shown the principal direction to help us find right ascension, α , and declination, δ .

This framework allows us to define the direction of objects in space. It also allows us to describe the position of objects closer to the Earth. We uniquely determine the direction of a point on the celestial sphere by two angular coordinates. These coordinates use the fundamental plane and a reference point. One coordinate describes the angular distance of the point of interest above or below the fundamental plane, and the other describes the angle from the reference point along the celestial equator to the hour circle passing through the object. These coordinates are analogous to latitude and longitude.

In addition to being able to define the orbit, we must be able to describe an object's position in space using a celestial coordinate frame. A celestial coordinate frame is one that is used to locate celestial objects. Let's first discuss a few fundamentals.

The Earth and its orbit around the Sun form the basis for celestial coordinate systems, which we'll discuss later. The *ecliptic* is defined as the mean plane of the Earth's orbit about the Sun. We assume it is free of periodic variations. The term comes from the fact that eclipses of the Moon occur only when the Moon is close enough to this plane and is between the Earth and the Sun (Berry, 1961, 11). When we view the Sun from the Earth, it appears to move along the ecliptic as shown in Fig. 1-16. It doesn't move exactly on the ecliptic because this path is defined as the mean plane of the Earth's orbit. The Earth's *equatorial plane* extends the equator from the Earth. The angle between the Earth's mean equator and the ecliptic is called the *obliquity of the ecliptic*, ϵ . This angle is about 23.5° , although it does vary slightly over time due to perturbations we'll discuss later. The intersection of the two planes helps us fix a principal direction. The line of intersection is called the *line of nodes*. The Sun occupies a position along this intersection twice a year, and they're called *equinoxes*: one when the Sun is at the ascending node (in the spring about March 21, *vernal equinox*) and one when the Sun is at the descending node (in the fall about September 23, *autumnal equinox*). Remember, the seasons cited are for the Northern Hemisphere! When the Sun is at an equinox, the Earth experiences equal times of day and night because the Sun's declination is zero—equinox comes from the Latin root (*aequinoctium*) meaning equal day and equal night.

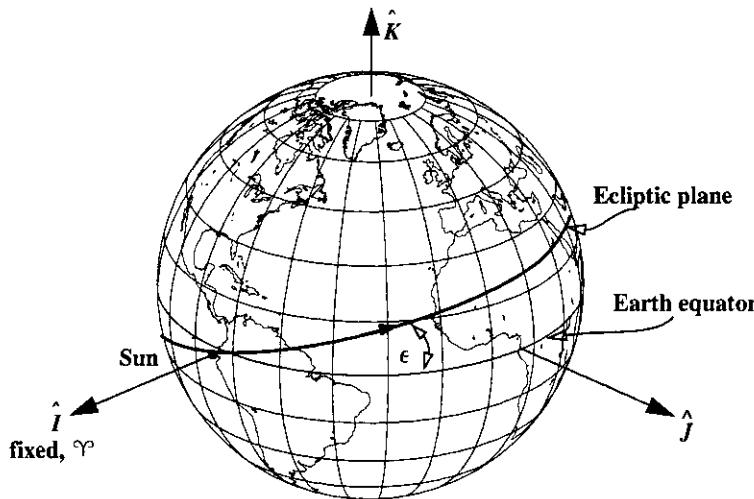


Figure 1-16. Geometry of the Vernal Equinox. The Earth's mean orbit around the Sun forms the ecliptic plane. The Earth's equatorial plane is inclined about 23.5° to the ecliptic. When the Sun is at the intersection of the two planes (has zero declination) and is at the ascending node, as viewed from Earth, it's the first day of spring.

A precise definition for the ***vernal equinox*** is that it occurs when the Sun's declination (Fig. 1-15) is 0° as it changes from negative to positive values. This point differs slightly from the intersection of the ecliptic and the equator because the ecliptic is the mean path of the Sun, not the true (or actual) path. In other words, the vernal equinox occurs at the ascending node of the Sun as viewed from the Earth. The direction of the vernal equinox is designated γ and often referred to as the *first point of Aries*. The symbol designates the ram. It comes from the fact that the direction of the vernal equinox pointed to the constellation Aries during Christ's lifetime. Today, however, due to precession (page 77), the vernal equinox points in the direction of the constellation Pisces.

We measure the ***ecliptic latitude*** and ***ecliptic longitude*** (sometimes called ***celestial latitude*** and ***longitude***) similarly to latitude and longitude for the Earth; however, we use the ecliptic as the fundamental plane and measure ecliptic longitude from the vernal equinox positive to the east, as shown in Fig. 1-10 and Fig. 1-17.

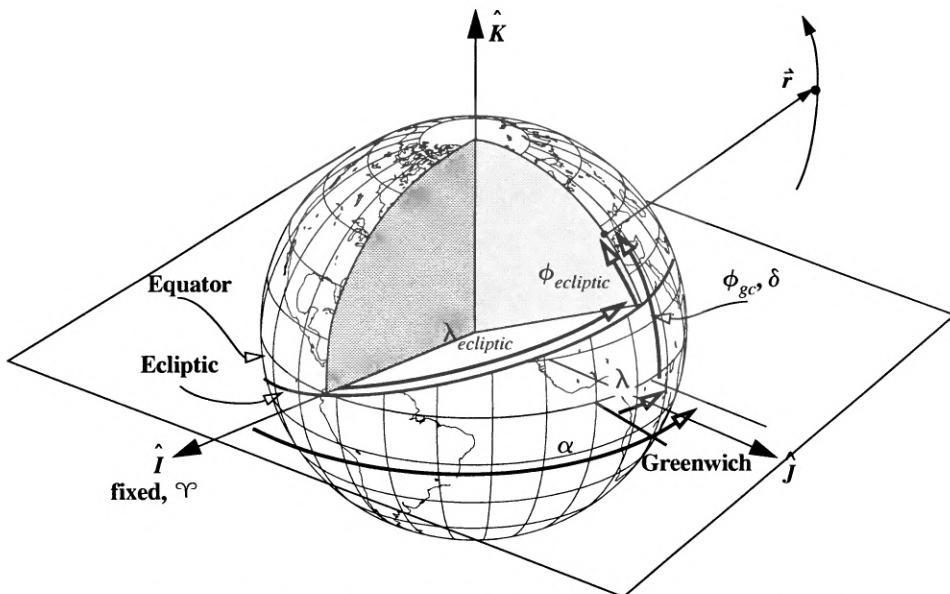


Figure 1-17. Using Right Ascension and Declination or Longitude and Latitude to Measure Location. Notice that both terrestrial latitude and longitude, ϕ and λ , and right ascension and declination, α and δ , use the Earth's equator, whereas ecliptic latitude and longitude are celestial coordinates that use the ecliptic. The declination and the terrestrial latitude are the same, but right ascension and terrestrial longitude use different starting locations.

Measurements using right ascension and declination are similar to those using longitude and latitude, but the reference point is fixed to the vernal equinox. Typically, we use the Earth's equator, but because the vernal equinox moves slightly, we often must specify

its exact location at a certain time (Fig. 1-17). The **right ascension**, α , is measured positive to the east in the equator's plane from the vernal equinox direction. Similar to longitude, the range of values may be from 0 h to 24 h (0° to 360°). Because these angles mainly depict locations of stars, fixing the initial point to the vernal equinox is convenient. The **declination**, δ , is measured northward from the equator (0° to 90°) to the object's location. Declinations which are south of the equator are negative (0° to -90°). Combining right ascension and declination defines only the object's direction. Astronomers use this system to catalog star positions. The extreme distance of stars from the Earth means that we don't have range information and thus specify only the location. We'll examine these angles in more detail in Chap. 3.

Hour Angle

We use hour angles mainly for describing observations of celestial objects relative to a local observer. Hour angles are easiest to visualize by looking at a diagram. They measure the angular distance along the celestial equator of an object, and are analogous to longitude. The hour circle that passes through the observer is 0 hours—it's called the **primary hour circle**. The **hour angle** of any object is the angle from the primary hour circle to the hour circle of the object. The units are usually hours from 0 to 24. This is a left-handed system, so it's important to follow a sign convention in which angles are measured positively *westward* to the object. We sometimes refer to the **Greenwich Hour Angle**, *GHA*, and the **Local Hour Angle**, *LHA*. The current hour angle of any object is equal to the elapsed time from when the object was overhead. This definition applies to all objects (Sun, stars, satellites) and to local observers and observers at Greenwich. Because we can use this definition for *any* object, this book will always use subscripts with *GHA* and *LHA* to identify the object of interest. From Fig. 1-18, notice that $GHA = LHA + \lambda$. Calculations are simply additions and subtractions of *GHA*, *LHA*, and terrestrial longitude; with Eq. (1-41), degrees and hours are interchangeable, although the common units are hours.

Barycentric Coordinate Systems

We often form equations of motion using a system's center of mass, or **barycenter**, as the origin. Except for the origin, the remaining characteristics are derived from the "parent" system—for instance, the heliocentric and geocentric systems I'll introduce. I'll identify barycentric systems with a *B* subscript and the letters of the "parent" system (heliocentric— $X_B Y_B Z_B$, geocentric $I_B J_B K_B$, and so on).

Synodic Coordinate Systems

We encounter the **synodic coordinate system** when studying interplanetary satellites and sometimes Earth satellites. It differs from other systems because we always use the system's barycenter. Because there are often many objects within a system, it's common to identify each by its mass. Thus, the most massive body is primary, the next is secondary, and so on. We also assume the synodic system rotates with a constant angular velocity, ω_s , relative to an inertial frame. This rate is the mean motion of the secondary object about the

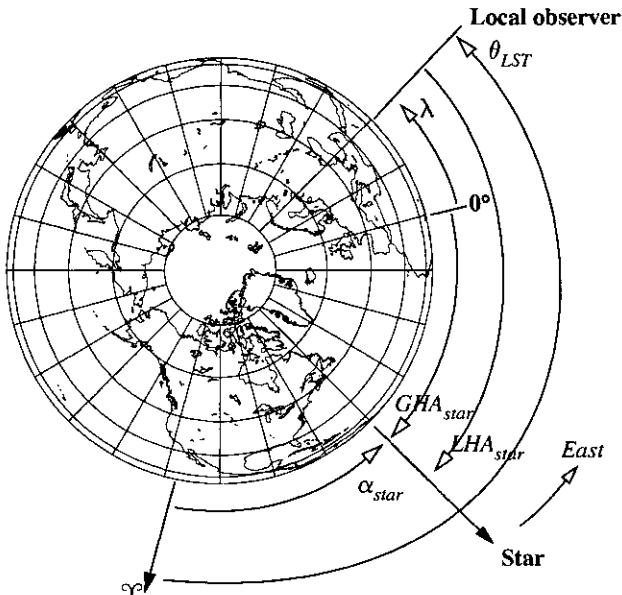


Figure 1-18. Geometry for LHA and GHA . GHA , and LHA are always measured positive to the west from the object of interest. Longitude, λ , is positive to the east. Notice here that both GHA_{star} , and LHA_{star} , would be positive. It's common to use the right ascension, α , measured from the equinox, with the hour angle. We use subscripts to avoid confusion when referring to specific objects. θ_{LST} will be discussed shortly.

primary. The fundamental plane is defined by the conservation of total angular momentum (the so-called *invariable plane* in Sec. 2.3). In the absence of perturbations, this is just the orbital plane defined by two-body motion. Just as I have for the barycentric systems, I'll use the letters of the "parent" system with a subscripted S to remind you of the rotation. The X_S axis, which is the principal axis, points toward the primary body from the center of mass. The Y_S axis is normal to the X_S axis in the invariable plane. Finally, the Z_S axis is normal to the invariable plane. This system is especially useful in studying three- and four-body problems, introduced in Chap. 2.

1.4.1 Interplanetary Systems

We'll examine only one main interplanetary system, the *Heliocentric Coordinate System*, XYZ , as shown in Fig. 1-19, because we're focusing mainly on Earth-orbiting satellites. As the name implies, this system has its origin at the center of the Sun. The letters XYZ describe the three axes and the fundamental plane is the ecliptic. The system uses ecliptic latitude and longitude coordinates to locate objects. As an aside, because the principal direction is the vernal equinox, the Earth's aphelion and perihelion don't match any particular axis of the coordinate system. The system is important when developing equations for

interplanetary flights (e.g., patched-conic approximations) and for describing planetary motion. As we mentioned earlier, variations exist for the barycenter of the solar system ($X_B Y_B Z_B$) and a rotating synodic system ($X_S Y_S Z_S$).

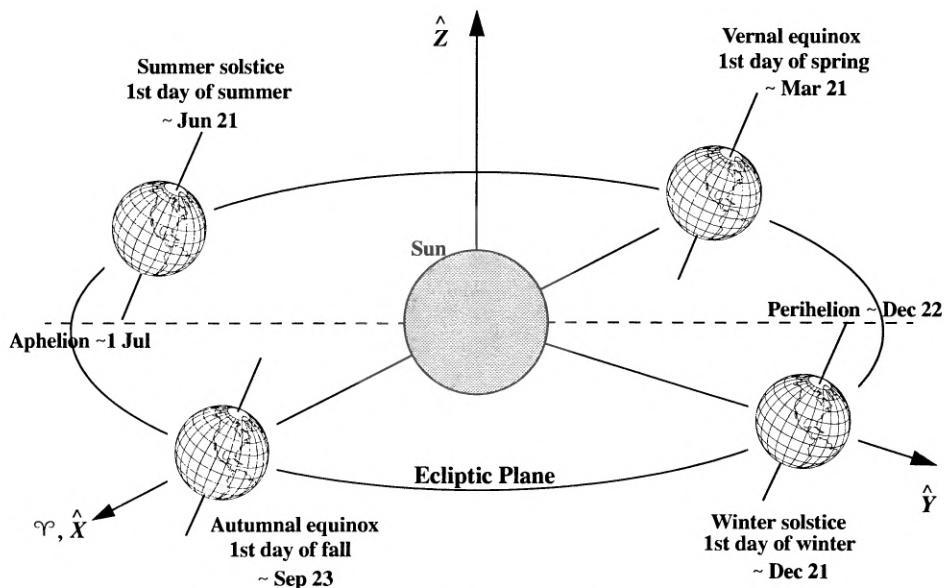


Figure 1-19. Heliocentric Coordinate System, XYZ. The primary direction for the heliocentric system is the vernal equinox, Υ , on the first day of spring. Seasonal data for the Northern Hemisphere are approximate due to small perturbational forces on the Earth.

1.4.2 Earth-based Systems

The origin of these systems may be at the Earth's center (geocentric), or at a site (viewer's location) on or off the Earth's surface (topocentric). These coordinate systems are the basis for most operations in this book. As we do for the heliocentric system, we can define a barycentric system ($I_B J_B K_B$) and a rotating synodic system ($I_S J_S K_S$) that are very useful in the restricted three-body problem.

Geocentric Equatorial Coordinate System, IJK

This system originates at the center of the Earth, as the name implies, and is designated with the letters IJK . The fundamental plane is the Earth's equator, as shown in Fig. 1-20. The I axis points towards the vernal equinox; the J axis is 90° to the east in the equatorial plane; the K axis extends through the North Pole. This coordinate system isn't rotating. It's assumed to be fixed in space and is one of the most common systems in astrodynamics. The geocentric frame, IJK , is often used interchangeably with the *Earth-Centered Inertial*

(*ECl*) system. In Sec. 1.7 we'll see that this system actually moves over time, so I'll use the *IJK* notation for general purposes, and *ECl* when we discuss precession and nutation.

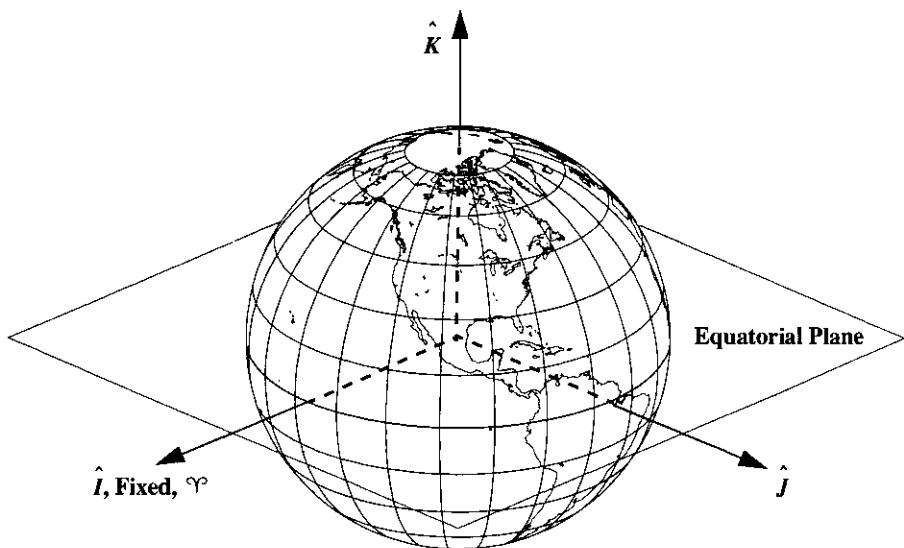


Figure 1-20. Geocentric Equatorial System (*IJK*). This system uses the Earth's equator and the axis of rotation to define an orthogonal set of vectors. It's the most common coordinate system in this book. The vernal equinox direction is fixed for most applications. It's also commonly called the Earth-Centered Inertial system (*ECl*).

As with the heliocentric coordinate system, the equinox and plane of the equator move very slightly over time, so a *truly* inertial reference frame for the Earth is impossible to realize. However, we can almost achieve an inertial coordinate system if we refer to a particular epoch and specify how to transform vectors to and from this time. The **J2000 system** is based on the *Fundamental Katalog, FK5* star catalog. This system represents our best realization of an ideal, inertial frame at a fixed epoch. It's sometimes called the *ECl* or *Conventional Inertial System (CIS)*. This system is considered inertial for precise orbit determination and calculations. The equator's and equinox's motion is precisely accounted for, so inertial frames at other times defined by the equator and equinox of date can be transformed to the J2000 frame. These other frames are called **true-of-date** inertial frames, meaning they reference the true equator and true equinox of date.

The **J2000** system is basically the Geocentric Equatorial coordinate system at a particular time, the J2000 epoch, so it fixes a common frame for all observations. Calculations that transform vectors to and from this epoch are usually called **reduction formulas**. Be careful when using these formulas; previous epochs (e.g. B1950, *FK4*) and their associated reduction formulas will produce small, but noticeable errors if you apply them to the

J2000 system. I treat the B1950 (*FK4* system) briefly in Sec. 1.7. See Seidelmann (1992, 167–187) for more information on these older systems.

Earth-Centered, Earth-Fixed Coordinate System, ECEF

We may also allow the geocentric coordinate system to rotate *with* the Earth, resulting in the *Earth-Centered, Earth-Fixed, ECEF*, coordinate system. The main difference with this system is that the primary axis is always aligned with a particular meridian. This system is especially useful in processing satellite observations from a site. The Greenwich meridian is usually chosen as the principal direction. In this case, it's called the *Earth-Fixed Greenwich System, EFG*. Be aware that *ECEF* and *EFG* are often used interchangeably—I'll adhere to the definitions above. Because the coordinate system rotates, *we must specify an epoch*. In addition, because the system references the Earth's equator, it's closely related to the true-of-date frame. We'll explore this more in Sec. 1.7. Typically, we use this system to convert actual observations to the J2000 system for use in other calculations.

Topocentric Horizon Coordinate System, SEZ

This system is useful in observing satellites and is used extensively with sensor systems. The *SEZ* system rotates with the site and is shown in Fig. 1-21. Notice the local horizon forms the fundamental plane. The *S* axis points due south from the site (even in the Southern Hemisphere). The *E* axis points east from the site and is undefined for the North or South Pole. Finally, the *Z* axis (zenith) points radially outward from the site, along the site position vector from the Earth's center.

The *SEZ* system allows us to define “look angles” to view the satellite from a ground station. The *azimuth*, β , is the angle measured from north, clockwise (as viewed from above) to the location beneath the object of interest (although other conventions exist). It may assume values from 0 to 360° . Negative values simply mean we've measured counter-clockwise. *Elevation*, el , is measured from the local horizon, positive up to the object of interest. It takes on values from -90° to 90° , and is sometimes called *altitude*. We rarely encounter negative values for sites on the Earth, but they often appear when used with an orbiting satellite. A *zenith distance* is also defined as the complementary angle, $90^\circ - el$. I'll discuss these parameters further in Chaps. 3 and 6.

Notice that a subtlety exists in the definition for azimuth. You might expect β to be measured positive, counter-clockwise from south. We get the same result using the previous definition (positive, clockwise from north) because it implicitly uses a left-handed convention. This north, east, zenith convention is the classical definition of the topocentric horizon system and it includes the same parameters as the above definition. Azimuth is still measured positive clockwise from north, but the principal axis now points north. This is an important distinction when trying to determine initial orbits (Chap. 6).

Topocentric Equatorial Coordinate System, I,J,K_t

This system is essentially the geocentric equatorial coordinate system with the origin translated from the Earth's center to the origin of the topocentric horizon system—the

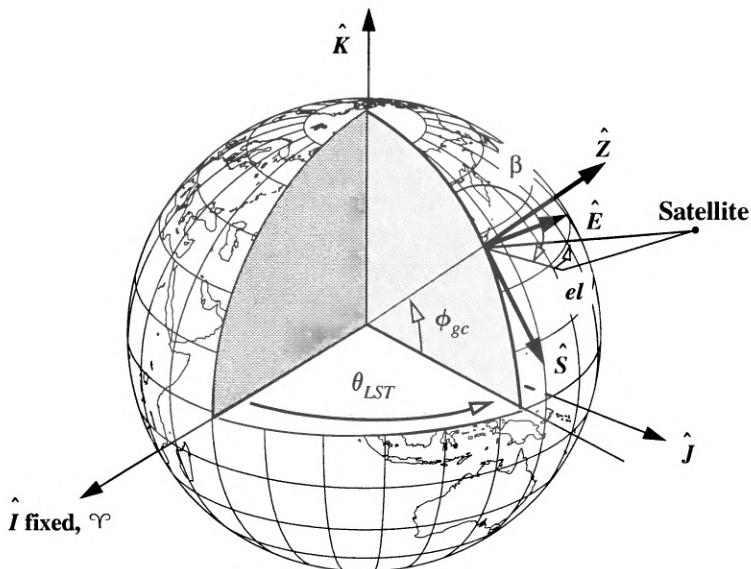


Figure 1-21. Topocentric Horizon Coordinate System, SEZ. This system moves with the Earth, so it's not fixed like the IJK system. LST is required to orient the SEZ frame to a fixed location. It's used extensively in radar observations and often involves azimuth, β , and elevation, el .

observer (see Fig. 1-22). The orientation of the axes is the same as the IJK system; the I_t axis is fixed to the vernal equinox direction, even though the center is attached to the site, which does change position as the Earth rotates. This system is used mainly for highly accurate optical systems and their observations (for instance the GEODSS and optical telescopes).

This system's usefulness may not yet be apparent. As discussed earlier, right ascension and declination use the geocentric equatorial system. The I_t, J_t, K_t system thus seems to be unnecessary, because for stars, the right ascension and declination values are virtually the same even when viewed from different points on the Earth. However, computations for satellites orbiting near Earth are conveniently performed in the topocentric equatorial system. In Fig. 1-22, the parameters are denoted as *topocentric right ascension*, α_t , and *topocentric declination*, δ_t . I'll use the "topocentric" prefix to specify the origin and the subscript to signify the reference frame.

1.4.3 Satellite-based Systems

Although other satellite-based systems exist, much of the nomenclature isn't standard, and many systems are developed for specific satellite missions. Each of these systems is based on the plane of the satellite's orbit. They all use true anomaly and the classical orbital ele-

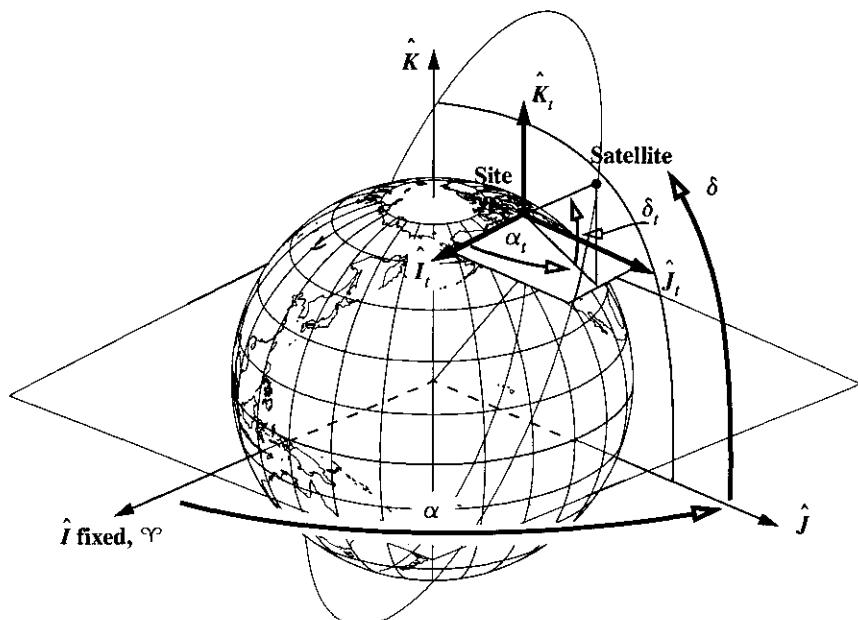


Figure 1-22. Topocentric Equatorial Coordinate System, $I_t J_t K_t$. This system uses a plane parallel to the Earth's equator but located at a particular site. The system rotates with the Earth but maintains an inertial orientation. The difference in numerical values is apparent above in the declinations. The topocentric declination, δ_p , in this figure is about 20° , whereas the declination, δ , is about 50° .

ments (Sec. 2.4.1) to describe object locations. We need many systems to cover various satellite operations.

Perifocal Coordinate System, PQW

A convenient system for processing satellite observations is the *perifocal coordinate system, PQW*. In this system, the fundamental plane is the satellite's orbit, and the origin is at the center of the Earth. The P axis points towards perigee, and the Q axis is 90° from the P axis in the direction of satellite motion. The W axis is then normal to the orbit. Figure 1-23 shows the PQW system. Don't be confused and think this system rotates with the satellite. The PQW system *always* maintains an orientation towards perigee and is therefore best suited for orbits with a well-defined eccentricity. For a perfectly circular orbit, because perigee doesn't exist, the usual convention is to use another system or to align the P axis with the orbit's ascending node (where it crosses to the Northern Hemisphere) in the equatorial plane.

We sometimes use a variant of this system in which the principal axis is in the direction of the ascending node. It's usually called the *nodal coordinate system*. Although it permits

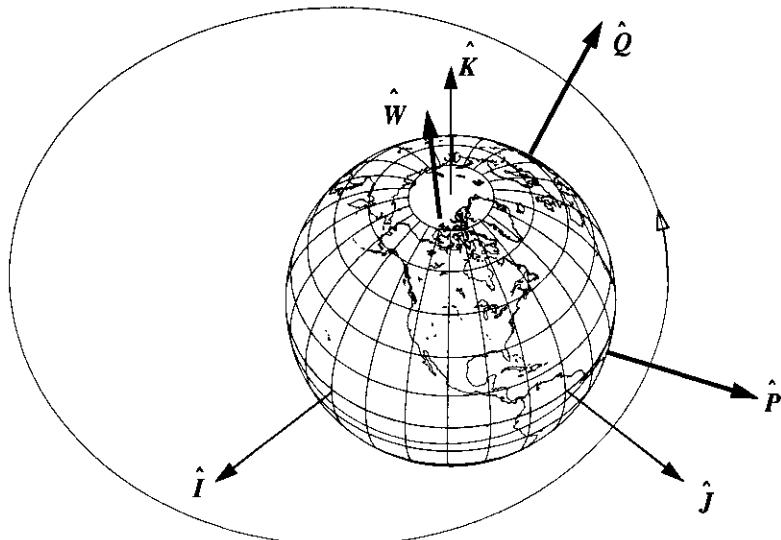


Figure 1-23. Perifocal Coordinate System, PQW. This system points towards perigee and is within the orbital plane. The W axis is aligned with the normal direction to the orbit. Although this particular orbit is near equatorial (10°), it doesn't have to be. We often view this system looking down from the W axis.

evaluation of low-eccentricity orbits, it's ill-defined for equatorial orbits. One solution orients the P axis to the vernal equinox direction.

Satellite Coordinate System, RSW

The *satellite coordinate system* applies to studies of relative motion (see the discussion of relative motion in Sec. 5.8). The system moves *with* the satellite and is sometimes called the *Gaussian coordinate system*. The R axis is defined as *always* pointing from the Earth's center along the radius vector toward the satellite as it moves through the orbit, and the W axis is fixed along the direction normal to the orbital plane. The S axis points in the direction of the velocity vector and is *perpendicular* to the radius vector—an important additional requirement. Figure 1-24 shows this system. The S axis is usually not aligned with the velocity vector except for circular orbits or for elliptical orbits at apogee and perigee. In addition, because the coordinate system is based on the satellite's present location, it applies to all orbit types.

Relative positions or displacements use the *RSW* coordinate system. **Radial** positions and displacements are parallel to the position vector (along the R axis). **Along-track** displacements are normal to the position vector and, for small e , are nearly parallel to the satellite's instantaneous velocity vector (along the S axis). Some confusion may exist with *in-track* displacements, which I'll define with the *NTW* coordinate system in the next section.

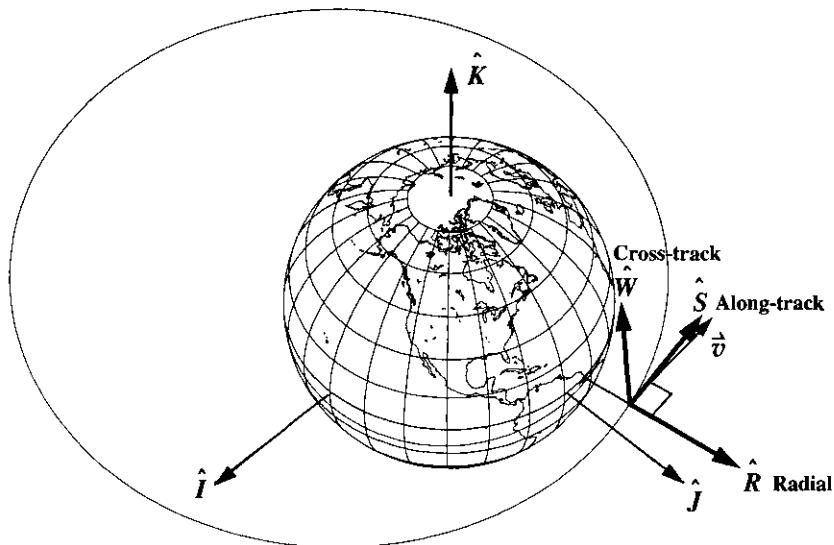


Figure 1-24. Satellite Coordinate System, RSW. This system moves with the satellite. The R axis points to the satellite, the W axis is normal to the orbital plane (and usually *not* aligned with the K axis), and the S axis is normal to the position vector. The S axis is *continuously* aligned with the velocity vector *only* for circular orbits.

Finally, ***cross-track*** positions are normal to the plane defined by the current position and velocity vectors (along the W axis). These orientations mainly provide reference points for describing the satellite's position from a sensor site.

We sometimes use a *roll-pitch-yaw coordinate system (RPY)* in attitude dynamics and instrument pointing to provide an “airplane-like” reference system for satellites. The *RPY* system is very similar to the *RSW* system and is actually only one (standard) rotation from that system. (We'll look at rotations in the next section.) The *roll* axis is the same as the S axis of the *RSW* system. The *pitch* axis is opposite to the position vector, points toward the center of the Earth, and also lies in the orbital plane. Finally, the *yaw* axis is opposite to the angular-momentum vector; it completes the orthogonal set of vectors. Equivalent axis directions are R —pitch, S —roll, and W —yaw. Many books reference the *roll* axis as being along the velocity vector, but this is true *only* for circular orbits.

Satellite Coordinate System, NTW

Figure 1-25 shows this system, which is valid for all orbit types. In this “normal” system, the primary axis lies in the orbital plane, normal to the velocity vector. The T axis is tangential to the orbit, and the W axis is normal to the orbital plane as in the *RSW* system. We define ***in-track*** displacements as deviations along the T axis. In-track errors are *not* the

same as along-track variations in the *RSW* system. One way to remember the distinction is that the in-track errors are *in* the direction of the velocity, whereas along-track variations are simply *along* the velocity vector. We use this coordinate system mainly to analyze drag effects on the orbit (Sec. 8.6.2) because drag always acts along the velocity vector. It's sometimes referred to as the *Frenet system*.

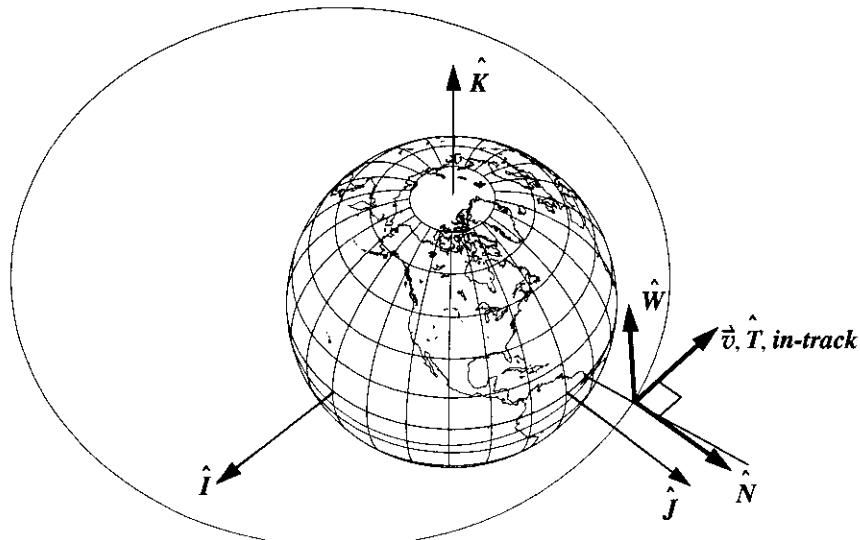


Figure 1-25. Satellite Coordinate System, *NTW*. This system has the *W* axis normal to the orbital plane (*not* aligned with the *K* axis). The *T* axis is always aligned with the velocity vector. The *N* axis is normal to the velocity vector and is *not* aligned with the radius vector, except for circular orbits and for apogee and perigee in elliptical orbits.

Equinoctial Coordinate System, EQW

This system is very useful for analyzing perturbations. The satellite's orbital plane is the fundamental plane, as shown in Fig. 1-26. We find the principal axis, *E*, by rotating in the orbital plane positively from the vernal equinox along the equator to the orbit's ascending node. Next we rotate from the equator to the orbital plane through the inclination. Then we rotate counter-clockwise from the line of nodes through the ascending node angle. The third axis is along the angular momentum vector, and the middle axis lies in the orbital plane to complete a right-handed coordinate system. As we've seen, other systems can't account for some forms of circular and elliptical orbits. The equinoctial system eliminates this difficulty and provides a way to measure the equinoctial orbital elements (Sec. 2.4.3). Remember, when we consider perturbations, the system actually moves over time.

Table 1-2 summarizes the types of coordinate systems.

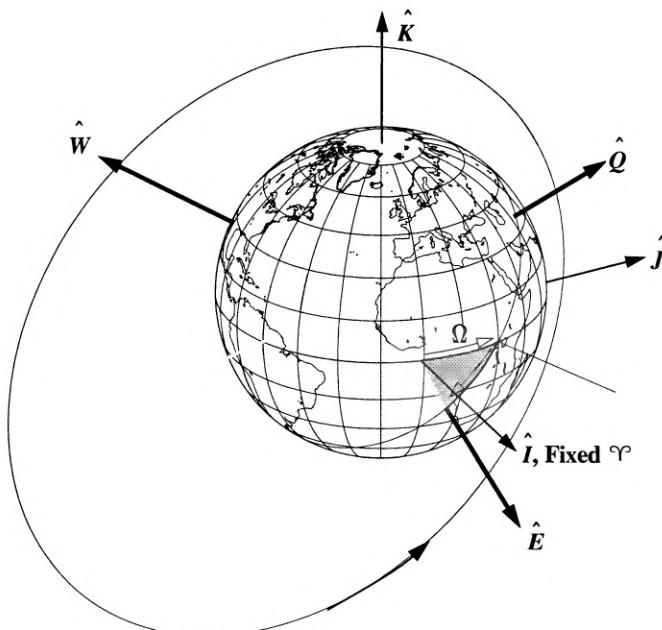


Figure 1-26. Equinoctial System, EQW. This system is formed starting at the *IJK* system (*I* axis). Rotating through the longitude of ascending node and then back through the same angle but in the orbital plane places us in the correct orientation and defines the *E* axis. The *Q* axis is perpendicular to the *E* axis in the orbital plane. The *W* axis is normal to the orbit.

1.5 Coordinate Transformations

I won't discuss the mathematical basis for coordinate transformations in detail because you're probably familiar with the basic concepts. For more information, see a standard mathematical text, such as Kreyzig (1983) or Danielson (1996). In general, two concepts apply to the broader category of transformations: corrections for rotations within a system and translations between systems having different origins. We'll examine translations later in the book because they vary for each problem.

1.5.1 Coordinate Rotation

We can express a vector's components in any coordinate system, but commonly use rectangular coordinates, although spherical coordinates occasionally simplify a particular application. Remember, a coordinate rotation changes only the vector's basis. The vector has the same length and direction, but the three components will differ. Figure 1-27 shows a single rotation about a primary axis (*principal axis*), which is the underlying principle for all coordinate transformations. The angle of rotation is called an **Euler Angle**.

TABLE 1-2. Summary of Coordinate Systems. With a few exceptions, most coordinate systems simply derive from the ones in this table.

System	Symbol	Origin	Fundamental Plane	Principal Direction	Example Use
Interplanetary systems					
Heliocentric	XYZ	Sun	Ecliptic	Vernal equinox	Patched conic
Solar system	$X_B Y_B Z_B$	Barycenter	Invariable plane	Vernal equinox	Planetary motion
Earth-based systems					
Geocentric*	IJK	Earth	Earth equator	Vernal equinox	General
Earth-Moon (synodic)	$I_S J_S K_S$	Barycenter	Invariable plane	Earth	Restricted three-body
Earth-centered Earth-fixed	$ECEF$ or $(IJK)_{ECEF}$	Earth	Earth equator	Local meridian†	Observations
Topocentric horizon	SEZ	Site	Local horizon	South	Radar observations
Topocentric equatorial	$I_P J_P K_P$	Site	Parallel to Earth equator	Vernal equinox	Optical observations
Satellite-based systems					
Perifocal**	PQW	Earth	Satellite orbit	Periapsis	Processing
Satellite radial	$RSW^{\dagger\dagger}$	Satellite	Satellite orbit	Radial vector	Relative motion, Perturbations
Satellite normal	NTW	Satellite	Satellite orbit	Normal to velocity vector	Perturbations
Equinoctial	EQW	Satellite	Satellite orbit	Calculated vector	Perturbations

*We specify a particular epoch for the J2000, or ECI system. This actually applies to all frames using Ψ .

†If we use the Greenwich meridian, we use an EFG designation

**To form the nodal coordinate system, we use the ascending node instead of the perigee direction.

††A roll-pitch-yaw system results from rotating RSW 180° about the S axis.

The easiest way to look at this process is to work through an example. Suppose we have a unit position vector \hat{r} in the SEZ frame, and we want to represent it in the IJK frame. In reality these frames have different origins, so we'll assume a translation has already occurred to align the frames. For convenience, let the position vector be aligned along the Z component of the SEZ frame. The unit vector notation simply allows the solution to be general given any combination of SEZ components. Figure 1-28 illustrates the situation.

The problem of representing each component in the new frame breaks down into two steps. First, we must represent the SEZ vectors in an intermediate frame ($ECEF$) between the SEZ frame and the IJK frame. (Remember that $ECEF$ is aligned with a particular meridian. If we used EFG , we would have to correct for both θ_{GST} and λ .) We do so by

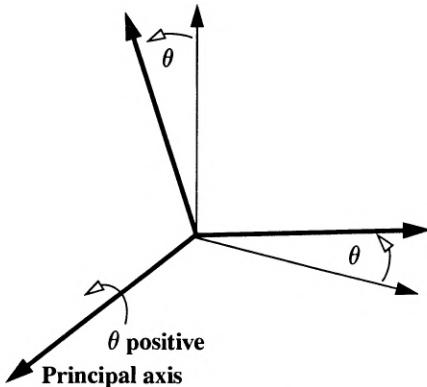


Figure 1-27. Coordinate Rotation. A coordinate rotation is a series of single-axis rotations. The sign of a rotation is considered positive if the right-hand-rule applies, i.e., if the angular motion is about the positive direction of the principal axis, as shown.

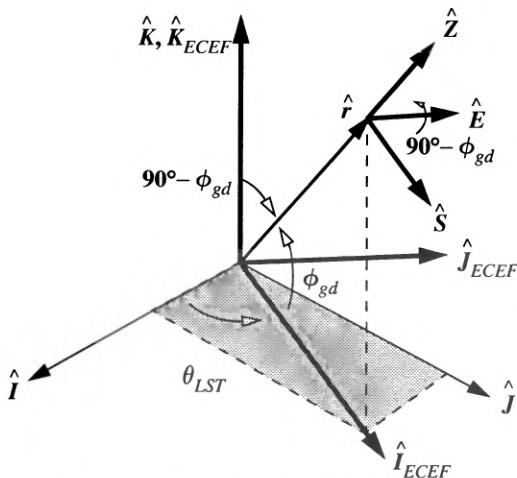


Figure 1-28. Topocentric to Geocentric Rotation #1. The first step in transforming coordinates is to rotate the vector from the *SEZ* system to the *ECEF* system. The rotation occurs about the *E* axis, as shown, through the angle $(90^\circ - \phi_{gd})$.

rotating about the *E* axis to align \hat{Z} with the \hat{K}_{ECEF} axis. The negative sign appears because the direction of the motion from \hat{Z} to \hat{K}_{ECEF} opposes the positive direction of the right-hand rule about the *E* axis. The rationale for performing a translation of the two coordinate systems so they have the same origin should now be apparent.

Mathematically, the idea is to represent each component of the position vector in the intermediate frame, *ECEF*. Symbolically, the vector \hat{r}_{ECEF} is

$$\hat{r}_{ECEF} = \cos(\phi_{gd}) \hat{I}_{ECEF} + \sin(\phi_{gd}) \hat{K}_{ECEF}$$

where again the unit vector permits a general formula. The obvious difficulty here is the component that comprises the \hat{I}_{ECEF} axis direction. Breaking this component into the \hat{I} and \hat{J} axes forces the second step, which is to rotate about the \hat{K}_{ECEF} axis to align \hat{I}_{ECEF} with \hat{I} . The sign is again due to the direction of motion from \hat{I}_{ECEF} to \hat{I} . Figure 1-29 shows the situation as viewed from along the Z axis.

If we let \hat{r} represent the resultant vector (IJK) from the rotation of just the \hat{I}_{ECEF} component,

$$\hat{r} = \cos(\phi_{gd}) \cos(\theta_{LST}) \hat{I} + \cos(\phi_{gd}) \sin(\theta_{LST}) \hat{J}$$

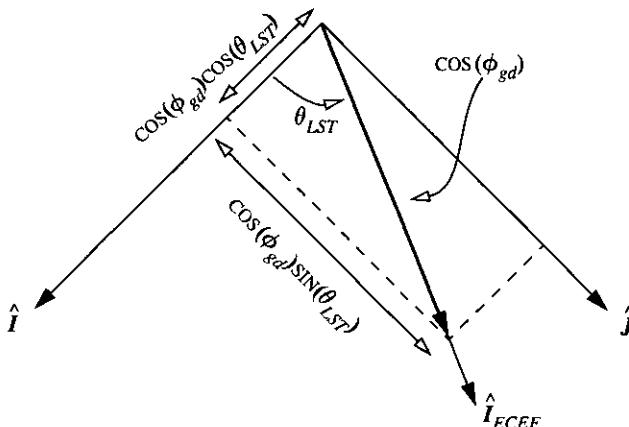


Figure 1-29. Topocentric to Geocentric Rotation #2. The final conversion involves changing from the $ECEF$ system to the IJK system, which requires the θ_{LST} angle.

To complete the process, we combine the two interim results. For this example,

$$\hat{r}_{IJK} = \cos(\phi_{gd}) \cos(\theta_{LST}) \hat{I} + \cos(\phi_{gd}) \sin(\theta_{LST}) \hat{J} + \sin(\phi_{gd}) \hat{K}$$

We can combine these equations using matrix notation. The general formulas are defined in Eq. (1-21). The number with each “ROT” signifies the principal axis of the rotation (e.g., $I-1$, $J-2$, $K-3$). Each of these rotations is for a right-hand coordinate system and, without scaling, the vector has the same magnitude throughout.

$$\begin{aligned}
 \text{ROT1}(\alpha) &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha) & \sin(\alpha) \\ 0 & -\sin(\alpha) & \cos(\alpha) \end{bmatrix} \\
 \text{ROT2}(\alpha) &= \begin{bmatrix} \cos(\alpha) & 0 & -\sin(\alpha) \\ 0 & 1 & 0 \\ \sin(\alpha) & 0 & \cos(\alpha) \end{bmatrix} \\
 \text{ROT3}(\alpha) &= \begin{bmatrix} \cos(\alpha) & \sin(\alpha) & 0 \\ -\sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{bmatrix}
 \end{aligned} \tag{1-21}$$

The matrix form is especially useful for computer software; it allows easy processing of multiple rotations. The result for the example (including the vectors, or 3×1 matrices) is

$$\vec{r}_{IJK} = \text{ROT3}(-\theta_{LST}) \text{ROT2}(-\{90^\circ - \phi_{gd}\}) \vec{r}_{SEZ}$$

where the subscripts on the vector indicate the coordinate system in which the vector is expressed. Notice the signs on the angles for each of the rotations. As discussed within the example, they follow from the conventions used in defining the rotations above. The individual symbols shown in the development include these signs, however, so they'll fit the definitions of these rotations—we need the negative signs. To determine the sign for any problem, we use the right-hand rule. Thus, if the rotation goes in the positive direction for this rule, the angle is positive. In the problem just described, the rotation goes *opposite* to the positive direction, so we have negative signs. Finally, keep in mind that the matrix multiplication is *not* commutative. This means that the *order of multiplication* is extremely important! We evaluate the rotations from right to left.

To be complete, the process of rotating vectors must work in both directions. We could use the same steps to reverse it, but we can save a lot of effort by looking at the matrices used to this point. Using matrix operations, the reverse process would use the inverses of each matrix, in the opposite order. Usually, matrix inversion is mathematically intensive. However, the rotation matrices in Eq. (1-21) are special. A three-by-three matrix is called orthogonal if the rows and columns are scalar components of mutually independent unit vectors. Here, all rows are mutually independent, so the rotation matrix is orthogonal. The inverse of an orthogonal matrix is simply the transpose, and the transpose is calculated by switching the rows and columns. In terms of the individual rotations, the order and signs of each angle are reversed, and for the complete matrix, the rows and columns are exchanged. For example,

$$\text{ROT3}(-\alpha) = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) & 0 \\ \sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{bmatrix} = [\text{ROT3}(\alpha)]^{-1} = [\text{ROT3}(\alpha)]^T$$

The inverse process is simply the reverse order of rotations.

$$\begin{aligned}\hat{\vec{r}}_{SEZ} &= \text{ROT2}(90^\circ - \phi_{gd}) \text{ROT3}(\theta_{LST}) \hat{\vec{r}}_{IJK} \\ \hat{\vec{r}}_{IJK} &= \text{ROT3}(-\theta_{LST}) \text{ROT2}(-(90^\circ - \phi_{gd})) \hat{\vec{r}}_{SEZ}\end{aligned}$$

Rotations Using Vectors

In cases that seem to require more than three rotations, it may be easier and faster to use a vector approach. This approach is not quite as easy to visualize but yields the same results. We can find unit vectors that describe a given coordinate system in terms of the desired system. When the desired system is the geocentric equatorial system, the analysis is sometimes simpler because defined vectors for the orbit are available (eccentricity, ascending node, etc., to be discussed later). The advantage with either approach for coordinate transformation is that once found, they remain constant, and we may use them continually as the satellite travels around its orbit. The following equations list transformations for the coordinate systems described earlier.

The *SEZ* coordinate system is defined for a given site location with latitude, ϕ_{gd} , and at a particular local sidereal time, θ_{LST} . For this transformation, we must first form the site position vector because it defines the direction of \hat{Z} :

$$\begin{aligned}\hat{\vec{r}}_{site} &= \cos(\phi_{gd}) \cos(\theta_{LST}) \hat{I} + \cos(\phi_{gd}) \sin(\theta_{LST}) \hat{J} + \sin(\phi_{gd}) \hat{K} \\ \hat{Z} &= \frac{\hat{\vec{r}}_{site}}{|\hat{\vec{r}}_{site}|} \quad \hat{E} = \hat{K} \times \hat{Z} \quad \hat{S} = \hat{E} \times \hat{Z}\end{aligned}\tag{1-22}$$

Notice the use of unit vectors to eliminate any problems with unintentional scaling. The actual implementation uses each of the unit vector's three components as a column, resulting in the three-by-three matrix. If we want speed, we may want to code each operation, taking into account savings through zero and unity multiplications and redundant calculations. The *values* in the matrix will be identical to those resulting from Eq. (1-33).

$$\hat{\vec{r}}_{IJK} = [\hat{S} \ : \ \hat{E} \ : \ \hat{Z}] \hat{\vec{r}}_{SEZ}\tag{1-23}$$

The reverse process is also simpler because the unit vectors form an orthogonal set; thus, the inverse is simply the transpose:

$$\hat{\vec{r}}_{SEZ} = \begin{bmatrix} \hat{S}^T \\ \cdots \\ \hat{E}^T \\ \cdots \\ \hat{Z}^T \end{bmatrix} \hat{\vec{r}}_{IJK}\tag{1-24}$$

We can also examine transformations between other coordinate frames. For the *PQW* system, we need no other trigonometric functions, but we do need the eccentricity vector to locate the direction of periapsis. We'll derive this quantity in Sec. 2.4. For now, we'll accept that it always points towards perigee. The following discussion drops the *IJK* subscripts from the position, velocity, and eccentricity vectors.

$$\begin{aligned}\hat{\mathbf{e}} &= \frac{\left(v^2 - \frac{\mu}{r}\right)\hat{\mathbf{r}} - (\hat{\mathbf{r}} \cdot \hat{\mathbf{v}})\hat{\mathbf{v}}}{\mu} \\ \hat{\mathbf{P}} &= \frac{\hat{\mathbf{e}}}{|\hat{\mathbf{e}}|} \quad \hat{\mathbf{W}} = \frac{\hat{\mathbf{r}} \times \hat{\mathbf{v}}}{|\hat{\mathbf{r}} \times \hat{\mathbf{v}}|} \quad \hat{\mathbf{Q}} = \hat{\mathbf{W}} \times \hat{\mathbf{P}}\end{aligned}$$

The actual formula is

$$\hat{\mathbf{r}}_{IJK} = [\hat{\mathbf{P}} \ : \ \hat{\mathbf{Q}} \ : \ \hat{\mathbf{W}}] \hat{\mathbf{r}}_{PQW} \quad (1-25)$$

For the *RSW* system, we don't need more calculations because the *R* axis always points to the satellite's present position. The equations are

$$\hat{\mathbf{R}} = \frac{\hat{\mathbf{r}}}{|\hat{\mathbf{r}}|} \quad \hat{\mathbf{W}} = \frac{\hat{\mathbf{r}} \times \hat{\mathbf{v}}}{|\hat{\mathbf{r}} \times \hat{\mathbf{v}}|} \quad \hat{\mathbf{S}} = \hat{\mathbf{W}} \times \hat{\mathbf{R}}$$

which allow the transformation

$$\hat{\mathbf{r}}_{IJK} = [\hat{\mathbf{R}} \ : \ \hat{\mathbf{S}} \ : \ \hat{\mathbf{W}}] \hat{\mathbf{r}}_{RSW} \quad (1-26)$$

The *NTW* coordinate system has the following equations:

$$\hat{\mathbf{T}} = \frac{\hat{\mathbf{v}}}{|\hat{\mathbf{v}}|} \quad \hat{\mathbf{W}} = \frac{\hat{\mathbf{r}} \times \hat{\mathbf{v}}}{|\hat{\mathbf{r}} \times \hat{\mathbf{v}}|} \quad \hat{\mathbf{N}} = \hat{\mathbf{T}} \times \hat{\mathbf{W}} \quad (1-27)$$

The advantage of using unit vectors is that we can find the rotation matrix components through cross and dot products instead of complex trigonometric functions. Admittedly, many of the trigonometric functions may be stored in temporary variables, but using unit vectors may be faster. This speed benefits certain applications, but the angle approach may be more useful when speed isn't as crucial as understanding.

1.5.2 Rotating Transformations

In the preceding discussion, we've focused entirely on transforming position vectors. When we transform velocity and acceleration vectors, we must also understand the frame in which the vector is determined. The terms *inertial* and *rotating* specify the motion of a particular coordinate system. These descriptors imply two things: (1) vectors with these

subscripts are written in terms of the basis vectors for the subscripted frame, and (2) all time derivatives are determined with respect to that frame.

To examine velocity and acceleration vectors in rotating coordinate systems, we must transform the coordinates and determine the magnitude of the effect caused by the rotation. To solve both problems, we can write the position, velocity, and acceleration vectors entirely in the rotating frame and then transform them using a transformation matrix. Or we can express all velocity and acceleration vectors in terms of the rotating coordinate system once we determine the position vector in the inertial frame.

Consider the case where two vectors are given to express a satellite's position in inertial and rotating coordinate systems. Let the vectors have arbitrary components, η_i and ξ_i , so

$$\vec{r}_{Inertial} = \eta_1 \hat{I} + \eta_2 \hat{J} + \eta_3 \hat{K} \quad \vec{r}_{Rot} = \xi_1 \hat{I}' + \xi_2 \hat{J}' + \xi_3 \hat{K}'$$

Of course $\vec{r}_{Inertial} = \vec{r}_{Rot}$ in *magnitude* and *direction* because they both locate the same satellite, although their components aren't identical. This is an important point—a vector is invariant under a coordinate transformation. Indeed, the rotating system's origin may be displaced from the inertial. In this case, we translate the origin, \vec{r}_{org} , so both frames coincide for complete generality. Differentiation permits analysis in each coordinate system as follows:

$$\begin{aligned} \dot{\vec{v}}_{Inertial} &= \dot{\eta}_1 \hat{I} + \dot{\eta}_2 \hat{J} + \dot{\eta}_3 \hat{K} & \dot{\vec{v}}_{Rot} &= \dot{\xi}_1 \hat{I}' + \dot{\xi}_2 \hat{J}' + \dot{\xi}_3 \hat{K}' \\ \ddot{\vec{a}}_{Inertial} &= \ddot{\eta}_1 \hat{I} + \ddot{\eta}_2 \hat{J} + \ddot{\eta}_3 \hat{K} & \ddot{\vec{a}}_{Rot} &= \ddot{\xi}_1 \hat{I}' + \ddot{\xi}_2 \hat{J}' + \ddot{\xi}_3 \hat{K}' \end{aligned} \quad (1-28)$$

Unfortunately, we must be able to transform vectors in either the inertial or rotating systems, so we need both analyses. Suppose we wish to find the *inertial* velocity vector, given the *rotating* position vector. By simple differentiation,

$$\dot{\vec{v}}_{Inertial} = \dot{\xi}_1 \hat{I}' + \dot{\xi}_2 \hat{J}' + \dot{\xi}_3 \hat{K}' + \dot{\xi}_1 \hat{I}' + \dot{\xi}_2 \hat{J}' + \dot{\xi}_3 \hat{K}'$$

I've included the derivatives of the basis vectors because the coordinate system is rotating and perhaps translating. We can simplify the result by recognizing the first three terms on the right-hand side as simply the velocity in the rotating frame. Using a more standard notation, then,

$$\dot{\vec{v}}_{Inertial} = \dot{\vec{v}}_{Rot} + \vec{\omega}_{R/I} \times \vec{r}_{Rot} + \vec{v}_{org} \quad (1-29)$$

where $\vec{\omega}_{R/I}$ is the moving system's rate of rotation with respect to the inertial system. I've included the velocity of the origin to make the formula more general. The result is sometimes referred to as the *basic kinematic equation*. Because we've used the rotating velocity vector, we must add a correction to get the complete expression in the inertial frame. Thus, I've added the cross-product term using the rotating position vector. Extending this approach gives us an acceleration

$$\ddot{\vec{a}}_{Inertial} = \ddot{\vec{a}}_{Rot} + \dot{\vec{\omega}}_{RI} \times \vec{r}_{Rot} + \vec{\omega}_{RI} \times (\dot{\vec{\omega}}_{RI} \times \vec{r}_{Rot}) + 2\dot{\vec{\omega}}_{RI} \times \vec{v}_{Rot} + \vec{a}_{org} \quad (1-30)$$

Notice the rotating acceleration is augmented by several other parameters to account for the coordinate system rotation. The supplemental terms are usually identified with specific names. The first term ($\dot{\vec{\omega}}_{RI} \times \vec{r}_{Rot}$) is called the *tangential acceleration* because ω changes. For the orbit problem, it's zero only for a true circular orbit. Next are the *Coriolis acceleration*, ($\vec{\omega}_{RI} \times (\dot{\vec{\omega}}_{RI} \times \vec{r}_{Rot})$), and the *centripetal acceleration* due to the angle between ω and \vec{v}_{Rot} . A final term covers an accelerating origin.

Another change relies on the *derivatives* of the rotation matrices and the rotating position vector. Again, let's use the example *SEZ* vector, but now it's the position relative to the *SEZ* frame (denoted by the lower-case r). Thus,

$$\begin{aligned}\dot{\vec{r}}_{Inertial} &= [\text{ROT}] \dot{\vec{r}}_{SEZr} \\ \dot{\vec{v}}_{Inertial} &= [\text{ROT}] \dot{\vec{v}}_{SEZr} + [\dot{\text{ROT}}] \dot{\vec{r}}_{SEZr} \\ \ddot{\vec{a}}_{Inertial} &= [\text{ROT}] \ddot{\vec{a}}_{SEZr} + 2[\dot{\text{ROT}}] \dot{\vec{v}}_{SEZr} + [\ddot{\text{ROT}}] \dot{\vec{r}}_{SEZr}\end{aligned}\quad (1-31)$$

Here, the derivatives of the rotation matrices transform the vector to the inertial system and account for centripetal and Coriolis effects. We typically use the first method because we don't have to determine the rates of change of the transformation matrix. It's important to know both notations so you can work with any system.

1.5.3 Common Transformations

Because the geocentric equatorial system, *IJK*, is most common, the transformations listed here will usually conform to this system. The following equations highlight some of the transformations—in different forms. Special care is sometimes required for the velocity vectors because rotating frames affect the magnitude and orientation of velocity vectors. Also, systems used with a site require extra calculations to add and subtract the site vector, which further complicates the velocity transformation. Sec. 2.4 covers orbital elements included in the transformations. I've included complete rotation matrices that result from multiplying the individual rotations together. This form is useful whenever we need computational efficiency. The combined rotation matrix is given as final/initial.

Transformations between J2000/*FK5* and *ECEF*:

See “Motion of the Coordinate System” on page 74.

Transformations between *ECEF* and *IJK* ignoring precession, nutation, and polar motion involve only θ_{AST} [Eq. (1-63)].

$$\begin{aligned}\dot{\vec{r}}_{ECEF} &= \text{ROT3}(\theta_{AST}) \dot{\vec{r}}_{IJK} \\ \dot{\vec{r}}_{IJK} &= \text{ROT3}(-\theta_{AST}) \dot{\vec{r}}_{ECEF}\end{aligned}\quad (1-32)$$

Transformations between *SEZ* and *IJK*:

$$\begin{aligned}\dot{\vec{r}}_{IJK} &= \text{ROT3}(-\theta_{LST}) \text{ROT2}(-(90^\circ - \phi_{gd})) \dot{\vec{r}}_{SEZ} \\ \dot{\vec{r}}_{SEZ} &= \text{ROT2}(90^\circ - \phi_{gd}) \text{ROT3}(\theta_{LST}) \dot{\vec{r}}_{IJK} \\ \left[\begin{array}{c} \dot{\vec{r}}_{IJK} \\ \dot{\vec{r}}_{SEZ} \end{array} \right] &= \left[\begin{array}{ccc} \sin(\phi_{gd}) \cos(\theta_{LST}) & -\sin(\theta_{LST}) & \cos(\phi_{gd}) \cos(\theta_{LST}) \\ \sin(\phi_{gd}) \sin(\theta_{LST}) & \cos(\theta_{LST}) & \cos(\phi_{gd}) \sin(\theta_{LST}) \\ -\cos(\phi_{gd}) & 0 & \sin(\phi_{gd}) \end{array} \right] \quad (1-33)\end{aligned}$$

$$\dot{\vec{v}}_{IJK} = \dot{\vec{\rho}}_{IJK} + \dot{\vec{\omega}}_\oplus \times \dot{\vec{r}}_{IJK}$$

Transformations between *I_tJK_t* and *IJK*:

$$\begin{aligned}\dot{\vec{r}}_{IJK} &= \dot{\vec{r}}_{I_t J_t K_t} + \dot{\vec{r}}_{siteIJK} & \dot{\vec{v}}_{IJK} &= \dot{\vec{v}}_{I_t J_t K_t} + \dot{\vec{v}}_{siteIJK} \\ \dot{\vec{r}}_{I_t J_t K_t} &= \dot{\vec{r}}_{IJK} - \dot{\vec{r}}_{siteIJK} & \dot{\vec{v}}_{I_t J_t K_t} &= \dot{\vec{v}}_{IJK} - \dot{\vec{v}}_{siteIJK} \quad (1-34)\end{aligned}$$

Note: no rotations are required because the two systems are parallel.

Transformations between *PQW* and *IJK*:

$$\begin{aligned}\dot{\vec{r}}_{IJK} &= \text{ROT3}(-\Omega) \text{ROT1}(-i) \text{ROT3}(-\omega) \dot{\vec{r}}_{PQW} \\ \dot{\vec{r}}_{PQW} &= \text{ROT3}(\omega) \text{ROT1}(i) \text{ROT3}(\Omega) \dot{\vec{r}}_{IJK} \quad (1-35)\end{aligned}$$

$$\left[\begin{array}{c} \dot{\vec{r}}_{IJK} \\ \dot{\vec{r}}_{PQW} \end{array} \right] = \left[\begin{array}{ccc} \cos(\Omega) \cos(\omega) - \sin(\Omega) \sin(\omega) \cos(i) & -\cos(\Omega) \sin(\omega) - \sin(\Omega) \cos(\omega) \cos(i) & \sin(\Omega) \sin(i) \\ \sin(\Omega) \cos(\omega) + \cos(\Omega) \sin(\omega) \cos(i) & -\sin(\Omega) \sin(\omega) + \cos(\Omega) \cos(\omega) \cos(i) & -\cos(\Omega) \sin(i) \\ \sin(\omega) \sin(i) & \cos(\omega) \sin(i) & \cos(i) \end{array} \right]$$

Transformations between *PQW* and *RSW*:

$$\begin{aligned}\dot{\vec{r}}_{RSW} &= \text{ROT3}(\nu) \dot{\vec{r}}_{PQW} \\ \dot{\vec{r}}_{PQW} &= \text{ROT3}(-\nu) \dot{\vec{r}}_{RSW} \\ \left[\begin{array}{c} \dot{\vec{r}}_{RSW} \\ \dot{\vec{r}}_{PQW} \end{array} \right] &= \left[\begin{array}{ccc} \cos(\nu) & \sin(\nu) & 0 \\ -\sin(\nu) & \cos(\nu) & 0 \\ 0 & 0 & 1 \end{array} \right] \quad (1-36)\end{aligned}$$

Transformations between *EQW* and *IJK* use a retrograde factor, f_r (+1 when $0^\circ \leq i \leq 90^\circ$ and -1 when $90^\circ < i \leq 180^\circ$):

$$\begin{aligned}\hat{\vec{r}}_{EQW} &= \text{ROT3}(-f_r\Omega) \text{ROT1}(-i) \text{ROT3}(\Omega) \hat{\vec{r}}_{IJK} \\ \hat{\vec{r}}_{IJK} &= \text{ROT3}(\Omega) \text{ROT1}(i) \text{ROT3}(-f_r\Omega) \hat{\vec{r}}_{EQW}\end{aligned}\quad (1-37)$$

$$\left[\begin{matrix} \frac{\text{EQW}}{\text{IJK}} \end{matrix} \right] = \begin{bmatrix} \cos^2(\Omega) + f_r \cos(i) \sin^2(\Omega) & \cos(\Omega) \sin(\Omega) \{1 - f_r \cos(i)\} & -f_r \sin(i) \sin(\Omega) \\ f_r \cos(\Omega) \sin(\Omega) \{1 - f_r \cos(i)\} & f_r \sin(\Omega)^2 + f_r^2 \cos(i) \cos^2(\Omega) & \sin(i) \cos(\Omega) \\ \sin(i) \sin(\Omega) & -\sin(i) \cos(\Omega) & \cos(i) \end{bmatrix}$$

The next three transformations aren't very common, so I won't show combined matrices.

Transformations between *NTW* and *IJK*:

$$\hat{\vec{r}}_{IJK} = \text{ROT3}(-\Omega) \text{ROT1}(-i) \text{ROT3}(-u) \text{ROT3}(-\phi_{fp}) \hat{\vec{r}}_{NTW} \quad (1-38)$$

Transformations between *RSW* and *IJK*:

$$\hat{\vec{r}}_{IJK} = \text{ROT3}(-\Omega) \text{ROT1}(-i) \text{ROT3}(-u) \hat{\vec{r}}_{RSW} \quad (1-39)$$

Transformations between *PQW* and *SEZ*:

$$\hat{\vec{r}}_{SEZ} = \text{ROT2}(90^\circ - \phi_{gd}) \text{ROT3}(\theta_{LST}) \text{ROT3}(-\Omega) \text{ROT1}(-i) \text{ROT3}(-\omega) \hat{\vec{r}}_{PQW} \quad (1-40)$$

The last few examples point out some problems with coordinate transformations. Recall from mathematics that any coordinate transformation requires, at most, three independent Euler angles. These few transformations seem to disobey this rule, but the key is in the independence of the angles. All angles in the last three transformations aren't independent. Although they help us see the problem and are usually available through computation, a set of three independent angles does exist for these transformations. But these angles aren't often named. The solution resides with quaternions, which are sometimes used in these analyses but are so specialized that we won't consider them in this book.

1.6 Time

According to Newcomb ([1906] 1960, 114), "the main purpose of time is to define with precision the moment of a phenomenon." This moment is referred to as the epoch of the event; thus, the *epoch* designates a particular instant described as a *date*. To determine the epoch of an event, we also need the concept of a precise time interval. Once a fundamental epoch is agreed upon, we readily determine other epochs by counting the number of intervals from the original.

Time is a fundamental dimension in almost every branch of science. In astrodynamics, time is especially critical because objects move so far so quickly. To have a practical time system, we need a precise, repeatable time interval based on some physical phenomenon that we can readily measure. We must agree on a fundamental epoch from which to count

intervals. The commonly accepted fundamental epoch is the beginning of the Christian era, although others exist. Finding a precise repeatable time interval is more problematic. Four time systems now provide timekeeping for scientific, engineering, and general purposes: sidereal time, solar and universal time, dynamical time, and atomic time. Sidereal time and universal time are based on the Earth's rotation and are related through mathematical relationships. Dynamical and atomic time are truly independent from the other forms. They're far very precise timekeeping.

From ancient times, the day has been the precise repeatable time interval. We determine the day from observations of the apparent diurnal motions of the Sun and stars due to the Earth's rotation. We're most accustomed to the usual divisions of hour-minute-second, which we commonly use to describe time. But it's equally correct to describe time as an angle because the Earth rotates through one revolution (360°) every day. Consequently, the Earth is divided into 24 time zones, each containing 15° of longitude. The 24-hour day has been in use for hundreds of years, but the 1884 International Meridian Conference formally acknowledged the concept of 24 hours per day and 360° of longitude (*Explanatory Supplement*, 1961, 7). Having two forms of representation requires the conversions shown in Eq. (1-41). Time divides into hours (h), minutes (min), and seconds (s); angular measure has three similar distinctions: degrees ($^\circ$), arcminutes ($'$), and arcseconds ($''$). Although the names are similar, the symbols aren't:

$$1 \text{ h} = 60 \text{ minutes (60 min)} = 3600 \text{ seconds (3600 s)}$$

$$1^\circ = 60 \text{ arcminutes (60')} = 3600 \text{ arcseconds (3600'')}$$

$$\begin{aligned} 1 \text{ h} &= 15^\circ & 1^\circ &= \frac{1 \text{ h}}{15} = 4 \text{ min} \\ 1 \text{ min} &= 15' & 1' &= \frac{1 \text{ min}}{15} = 4 \text{ s} \\ 1 \text{ s} &= 15'' & 1'' &= \frac{1 \text{ s}}{15} \end{aligned} \tag{1-41}$$

In modern times, the second has replaced the day as the basic interval. Ephemeris time was based on the ephemeris second—a quantity determined as a fraction of the year. By 1976, the *Système International (SI)* second was recognized as the newest standard. Consult Seidelmann (1992, 73–87) for more information on the history of various time systems.

The Earth's diurnal rotation with respect to the stars or with respect to the Sun gives rise to the concepts of sidereal time and solar time, respectively. We define them by successive transits of the local meridian by a given star and by the Sun. The lengths of these days differ mainly because of the Earth's annual orbital motion about the Sun, which causes an apparent displacement of the Sun in the sky of roughly one degree per day. Thus, the Earth must rotate almost one degree extra per solar day compared to the sidereal day. This causes the sidereal day to be about four minutes shorter than the solar day.

More than anything else, the Sun governs our daily activity, and because its motion is (at first glance) apparently regular, we base solar time on that motion. **Solar time** is loosely defined by successive transits of the Sun over a local meridian of longitude, although we need a fixed reference point on the Earth to define the beginning of each day. The Royal Observatory was established in 1675 at Greenwich, England. The Meridian Conference in 1884 recognized its importance in obtaining accurate observations and adopted the Greenwich meridian as the 0° longitude point. Figure 1-30 shows an exaggerated view of how solar time is measured. Notice the Earth rotates a little more than 360° during one solar day because of the Earth's motion around the Sun.

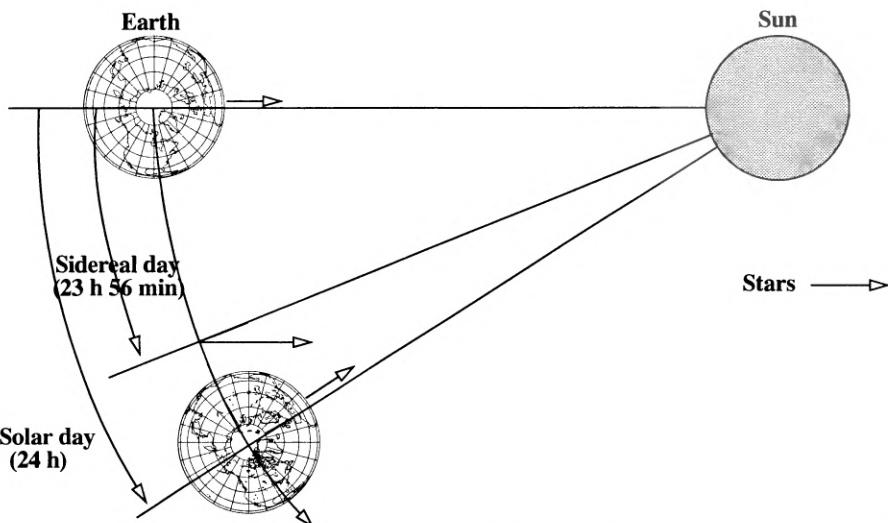


Figure 1-30. Measuring Solar and Sidereal Time. Distance and angular motion are greatly exaggerated for clarity. One solar day is the time required for an observer on the Earth to revolve once *and* observe the Sun at the same location. A sidereal day uses observations of the stars instead of the Sun. Notice that the sidereal day is slightly shorter than the solar day because the stars remain essentially in the same location over time.

We define **sidereal time** as the time between successive transits of the *stars* over a particular meridian. Because the stars are several orders of magnitude more distant than the Sun, their relative locations as seen from Earth don't change much even during a year (Fig. 1-30).

However, other irregularities in the Sun's actual motion make it difficult to use for reckoning time. As a result, the concept of **universal time**, *UT*, was adopted years ago. It's based on a *fictitious mean Sun* exhibiting uniform motion in right ascension along the equator. This fictitious mean Sun is defined mathematically as a function of the sidereal time. So, ultimately, we derive *UT* from sidereal time. There are actually three variations of *UT* (*UT0*, *UT1*, and *UT2*), which I'll discuss shortly. I'll drop the identifiers in general

discussions but keep them when they're important. There are also seasonal and irregular variations in the Earth's rotation that were discovered in the late 19th century. These variations demonstrated the need for a more stable reference for time intervals. Thus, other time systems have been developed to provide stable, highly precise timing measurements.

We can also measure time by the dynamical motion of bodies, such as the Earth's motion about the Sun or, more usually, the Moon's motion about the Earth. This is **dynamical time**, which depends on the fact that time is the independent variable in the equations of motion describing the motion. By observing the motion of the body from one point to another, we deduce the elapsed time from the mathematical description of the motion. The precise determination of dynamical time requires us to include relativistic corrections in modeling the Earth's motion.

Atomic time is the newest and most precise time standard. It's based on the specific quantum transition of electrons in a cesium-133 atom. The transition causes the emission of photons of a known frequency that we can count. We define the **atomic second** by a fixed number of cycles.

1.6.1 Sidereal Time

Sidereal time is a direct measure of the Earth's rotation and it's measured positively in the counter-clockwise direction when viewed from the North pole. Ideally, the observations of any star would suffice for determining sidereal time. The motion of the *true celestial pole* with respect to the rigid Earth, or **polar motion**, causes the local meridian to continually change its orientation. This produces a small difference in the time of meridian transits, depending on the star's declination. Because this effect vanishes at the equator, it's better to use stars with small declinations. Few stars meet this requirement, but remember we define the vernal equinox to be always on the equator. Thus, we define **sidereal time** as the hour angle of the vernal equinox relative to the local meridian. Because the vernal equinox is the reference point, the sidereal time associated with the Greenwich meridian is termed **Greenwich Sidereal Time**, θ_{GST} or *GST*. The sidereal time at a particular longitude is called **Local Sidereal Time**, θ_{LST} or *LST*. In this context, time is an angle measured from the observer's longitude to the equinox. Occasionally, when we use stars, I'll introduce appropriate subscripts. Figure 1-31 shows the relation between the *GST* and *LST*.

Be careful when comparing *GHA* and *GST* because hour angles are measured positive in a clockwise direction in a right-handed system, while sidereal time is measured positive in the counter-clockwise direction in that same system. Similar to *LHA* and *GHA*, *LST* and *GST* may be expressed in degrees or hours and may have positive or negative values. It helps to make a simple sketch like Fig. 1-18 or Fig. 1-31 when working problems to ensure the answer is correct. Also, you may need to check quadrants to correct the angle.

We can convert between *GST* and *LST* at a particular longitude, λ , using

$$\theta_{LST} = \theta_{GST} + \lambda \quad (1-42)$$

This formula requires a convention for east and west longitudes. As discussed earlier, the convention for this book is positive for east longitudes, negative for west longitudes. The

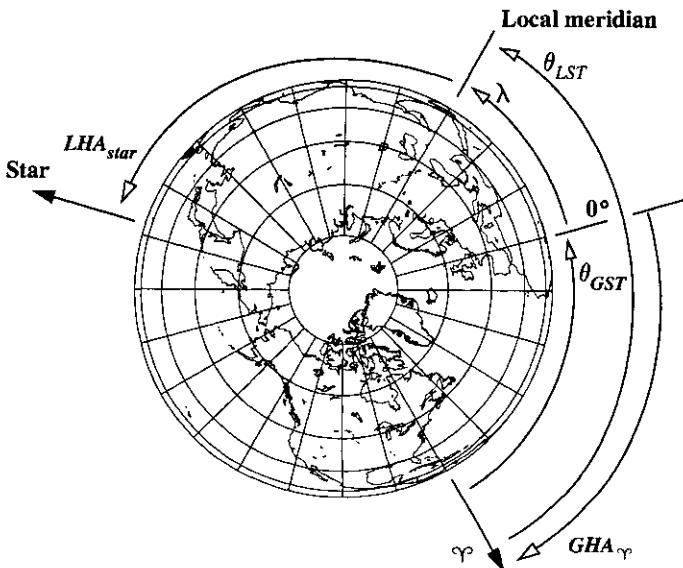


Figure 1-31. Geometry for *LST* and *GST*. Sidereal time is measured positive to the east from the vernal equinox to the location of interest. If a star is used, we can also use hour angles with an appropriate subscript (LHA_{star}). In this example, both sidereal times are positive in a right-handed system, whereas GHA_γ is positive and LHA_{star} is negative in a left-handed system.

opposite convention will work as long as you are *consistent* and *document* the notation—but remember, you’re implicitly corrupting a right-handed coordinate convention.

The concept of sidereal time would be perfect except that the equinox moves very slowly due to precession, the apparent locations of the stars *do* change a little, and polar motion causes small deviations in the exact longitudes of sites. Recall that the equinox results from the intersection of the Earth’s equator and the ecliptic, and both planes are moving. A distinction is made for mean and apparent time. **Mean sidereal time**, which is most commonly used, refers to a mean equinox that moves only with secular motion (precession). We measure **apparent sidereal time**, AST, from the true vernal equinox which includes secular and periodic contributions to the motion of the vernal equinox. The difference between the mean equinox and the apparent equinox is the **equation of the equinoxes**. It gives us the difference between the mean and apparent sidereal times, and is expressed in Eq. (1-63). For this book, I’ll consider all *LST* and *GST* values to be mean sidereal times and will identify occurrences of apparent sidereal time.

We can determine the longitude, λ , of a site by measuring the difference in local and Greenwich sidereal times. The term *reduced longitude* is sometimes given to indicate that polar motion corrections have been made to establish the exact local meridian. Uncorrected polar motion causes variations in the site’s longitude that can affect our accuracy.

In practice, the formal definition of sidereal time presents some difficulty because we must infer the hour angle of the vernal equinox (apparent or mean). We can determine sidereal time by observing the upper transit of stars. From Fig. 1-18, we can develop relations between systems. For any object of interest, the following formula permits various permutations. Remember that θ_{LST} always references the vernal equinox.

$$\theta_{LST} = \alpha_{star} + GHA_{star} + \lambda = \alpha_{star} + LHA_{star} \quad (1-43)$$

$$LHA_{star} = GHA_{star} + \lambda$$

Finally, I'll introduce two techniques to determine sidereal time when we know universal time, $UT1$. One uses an adopted epoch such as the J2000 epoch; the other uses an epoch at the beginning of each year. Both methods yield the mean sidereal time. Remember, this mean time is referenced to the mean equinox, and apparent sidereal time is referenced to the true equinox. (We'll introduce these concepts in Sec. 1.7.)

We prefer the first method using the Julian date because it contains all the information of the desired time (*YMD-HMS*) in one variable and doesn't require us to know a reference value. In addition, the formulas permit very accurate *GST* calculations for the period from 1901–2199 and are ideally suited for the computer. The limitation on applicable time is because of leap years which are discussed with the *JD*.

Recall that astronomers (Simon Newcomb) developed a mathematical expression for the fictitious mean Sun and that sidereal time is equal to the hour angle of this fictitious mean Sun. Thus the same mathematical expression can also be used to describe the sidereal time. Because it isn't practical to develop this mathematical expression for each local observer, a single expression references Greenwich, providing the Greenwich mean sidereal time at midnight, θ_{GST0} , (0 h 0 min 0 s) $UT1$. Expressions for θ_{GST0} at a desired time (in seconds, degrees, and radians), using the epoch of J2000 as a basis (*Astronomical Almanac*, 1984, S15, for seconds), are

$$\begin{aligned}\theta_{GST0} &= 24,110.548\,41\,s + 8,640,184.812\,866T_{UT1} + 0.093\,104\,T_{UT1}^2 - 6.2 \times 10^{-6}T_{UT1}^3 \\ \theta_{GST0} &= 100.460\,618\,4^\circ + 36,000.770\,053\,61T_{UT1} + 0.000\,387\,93\,T_{UT1}^2 - 2.6 \times 10^{-8}T_{UT1}^3 \quad (1-44)\end{aligned}$$

in radians;

$$\theta_{GST0} = 1.753\,368\,560 + 628.331\,970\,688\,9T_{UT1} + 6.7707 \times 10^{-6}T_{UT1}^2 - 4.5 \times 10^{-10}T_{UT1}^3$$

where T_{UT1} is the number of Julian centuries elapsed from the epoch J2000 and is computed using the Julian day numbers for the date of interest (JD_0) and for the epoch J2000. Julian day numbers are simply the integer part of the Julian date (JD), i.e., the JD at 0 h 0 min 0 s of the day.

The last form given in Eq. (1-44) above is most useful for computer software. Some applications drop the cubed term because it's required only for very precise calculations (on the order of 1×10^{-8} seconds). The $UT1$ subscript designation is required to be consistent with the precise notation necessary for certain applications. To complete the process, we must add the elapsed $UT1$ time on the day of interest:

$$\theta_{GST} = \theta_{GST0} + \omega_{\oplus} UT1 \quad (1-45)$$

where ω_{\oplus} is the Earth's mean angular rotation in degrees (or radians in the table on the inside front cover) per solar second and $UT1$ is the universal time in solar seconds. Note that units may need to be changed to be consistent with θ_{GST0} . The following algorithm for GST and LST and an example illustrate this method.

ALGORITHM 1: LSTIME (JD_0 , $UT1$, $\lambda \Rightarrow \theta_{LST}$, θ_{GST})

$$T_{UT1} = \frac{JD_0 - 2,451,545.0}{36,525}$$

$$\theta_{GST0} = 100.460\,618\,4^\circ + 36,000.770\,053\,61 T_{UT1} + 0.000\,387\,93 T_{UT1}^2 - 2.6 \times 10^{-8} T_{UT1}^3$$

$$\theta_{GST} = \theta_{GST0} + \omega_{\oplus} UT1$$

$$\theta_{LST} = \theta_{GST} + \lambda$$

▼ Example 1-2. Finding GST and LST (Method 1).

GIVEN: August 20, 1992, 12:14 h $UT1$ at 104° west longitude

FIND: $GST(\theta_{GST})$ and $LST(\theta_{LST})$

Start by finding the Julian date at the beginning of the day of interest (0 h $UT1$).

$$JD_0 = 367(1992) - \text{INT} \left\{ \frac{7\{1992 + \text{INT}\left(\frac{8+9}{12}\right)\}}{4} \right\} + \text{INT}\left(\frac{275(8)}{9}\right) + 20 + 1,721,013.5$$

$$JD_0 = 731,064.0 - 3487 + 244 + 20 + 1,721,013.5 = 2,448,854.5$$

$$T_{UT1} = \frac{JD_0 - 2,451,545.0}{36,525} = -0.073\,661\,875$$

$$\theta_{GST0} = 100.460\,618\,4^\circ + 36,000.770\,053\,61 T_{UT1} + 0.000\,387\,93 T_{UT1}^2 - 2.6 \times 10^{-8} T_{UT1}^3$$

$$= -2551.423\,618\,46^\circ$$

Reduce this quantity to a result within the range of 360° ($-31.423\,618\,46^\circ$). Next, update GST for the current $UT1$ (ω_{\oplus} is from the table on the inside front cover).

$$\theta_{GST} = \theta_{GST0} + \omega_{\oplus}(UT1) = -31.423\,618\,46^\circ + 0.250\,684\,477\,337 \{(12)(60) + 14\}$$

$$= 152.578\,787\,886^\circ$$

Find LST given $\theta_{GST} + \lambda = 152.578\,787\,886^\circ + (-104.0^\circ) = 48.578\,787\,886^\circ$

▲ $\theta_{LST} = 48.578\,787\,886^\circ$

The second method works by determining how far the Earth has turned from the epoch at the beginning of each year. Notice this method requires us to know the Greenwich sidereal time on January 1, 0 h $UT1$ (θ_{GST0}) of the year of interest but doesn't provide the

means to find it. We can find values for θ_{GST0} in the *Astronomical Almanac* for a given year or calculate them for *any* year by the first method [Algorithm 1]. Then, by knowing the day of the year and the fraction of a day, we find the *GST*, θ_{GST} using the following formula in radians:

$$\theta_{GST} = \theta_{GST0} + 1.002\,737\,909\,350\,795 (2\pi) (ED) \quad (1-46)$$

where *ED* is the total elapsed time in solar days from the epoch January 1, 0 h *UT1*. See Table 1-5 for a list of values. Be careful *not* to use the day of the year, which is one *more* than the *elapsed days* from January 1, (i.e., January 4 is only 3 *ED* from the epoch). We can also form the equation in degrees by replacing the 2π with 360° . But we must know θ_{GST0} before we can calculate anything, so we're limited in planning future missions using this method. We must reduce the result to values between 0° and 360° because the Earth turns a full 360° each day from the epoch. Note that this approximation neglects the contributions of the quadratic and cubic terms in Eq. (1-44). These contributions are usually small but measurable for precision applications. An example illustrates this method.

▼ **Example 1-3. Finding *GST* and *LST* (Method 2).**

GIVEN: August 20, 1992, 12:14 P.M. *UT* at 104° west longitude

FIND: $GST(\theta_{GST})$ and $LST(\theta_{LST})$

From the *Astronomical Almanac* (1992) for January 1, 1992, 0 h *UT*,

$$\theta_{GST0} = 6\text{ h }39\text{ min }37.4863\text{ s} = 99.906\,192\,9^\circ$$

Next find *ED* = 233 from Table 1-5 (1992 is a leap year). Then, update *ED* with $12/24 + 14/1440$ for *UT*.

$$ED = 233.509\,722\,222 \text{ from January 0, or } 232.509\,722\,222 \text{ elapsed days from January 1}$$

Find the *GST* and be sure to eliminate whole revolutions.

$$\theta_{GST} = \theta_{GST0} + 1.002\,737\,909\,350\,795 (360^\circ) (ED) = 84,032.578\,788\,27 \text{ MOD } 360^\circ$$

$$\theta_{GST} = 152.578\,788\,27^\circ = 10\text{ h }10\text{ min }18.909\text{ s}$$

$$\text{Find } LST \text{ as } \theta_{GST} + \lambda = 152.578\,788\,27 + (-104.0) = 48.578\,788\,27^\circ$$

$$\blacktriangleleft \quad \theta_{LST} = 48.578\,788\,27^\circ = 3\text{ h }14\text{ min }18.909\text{ s}$$

Although the computations are less complex than Example 1-2, notice the similarity in results, which suggests we can compare them for accuracy. To do so, use *mean sidereal times* tabulated in the *Astronomical Almanac*. Consider Table 1-3.

We've used values from the *Astronomical Almanac* as the "truth data" for every day of the year (1992). The maximum and minimum errors represent the amount each method differed from the *Astronomical Almanac*. Although the *Astronomical Almanac* lists only four decimal places (therefore $\pm 0.000\,05\text{ s} = 3.6 \times 10^{-9}$), we can still compare values. Notice that the minimum errors are very close, but the maximum errors are about three times different. As would be expected, the method using an epoch at the start of each year is very accurate towards the beginning of the year, but the errors continue to increase throughout the year. The Julian date method is very consistent throughout the year, so we prefer it.

TABLE 1-3. Comparing Methods for Calculating *GST* (0 h *UT1*). We've compared methods based on data for each day of 1992 and data from the *Astronomical Almanac* (1992). Though these differences seem small, they correspond to 2.6×10^{-5} s and 8.3×10^{-5} s for the average values. For a satellite travelling at 7.5 km/s, this error amounts to almost a meter.

Method	Maximum Difference (°)	Minimum Difference (°)	Average Difference (°)	Comments
<i>Astronomical Almanac</i>				Assumed truth
1. Julian date	$2.496\ 815\ 6 \times 10^{-7}$	4.0687×10^{-10}	$1.084\ 033\ 2 \times 10^{-7}$	Very consistent
2. θ_{GST} on Jan 1	$7.915\ 385\ 9 \times 10^{-7}$	7.9065×10^{-10}	$3.468\ 787\ 0 \times 10^{-7}$	Best performance at beginning of year

1.6.2 Solar Time and Universal Time

Solar time is based on the interval between successive transits of the Sun over a local meridian, which establishes the basic solar day. This concept has served adequately for ages. The Sun's apparent motion combines the Earth's rotation on its axis and its annual orbital motion about the Sun.

Kepler's first law implies that the distance of the Earth from the Sun varies according to its position on the ellipse. The Earth moves with a variable speed in the orbit, as shown in Kepler's second law. These factors affect the Sun's apparent annual motion in the sky, causing the Sun to exhibit nonuniform motion along the ecliptic. In addition, the ecliptic is inclined about 23.5° to the celestial equator; thus, the solar motion on the ecliptic appears as a sinusoidal motion about the celestial equator. The Sun's declination varies from about -23.5° on December 22 at the winter solstice to $+23.5^\circ$ on June 22, when the Sun is at the summer solstice (see Fig. 1-19).

Apparent Solar Time

The Earth's orbit has a small eccentricity and inclination, causing the length of each day to differ by a small amount. **Apparent solar time** is the interval for successive transits which we observe from a particular site:

$$\text{local apparent solar time} = \text{LHA}_{\odot} + 12 \text{ h}$$

The 12-hour increment ensures that 0 h corresponds to midnight and 12 h to noon, respectively. Greenwich apparent solar time is formed similarly. We can find apparent solar time using *any* star and the right ascension of that star (providing we use a small correction), but here I show the Sun. Now we must incorporate the right ascension of the Sun:

$$\text{Greenwich apparent solar time} = \text{GHA}_{\odot} - \alpha_{\odot} + 12 \text{ h} \quad (1-47)$$

Even if the Sun had moved uniformly along the ecliptic, the projection of the solar motion along the ecliptic onto the celestial equator contributes some nonuniform motion along the celestial equator. The variation in the Sun's apparent motion in right ascension (along the celestial equator) makes it a poor choice for establishing a precise time system because its observed length varies throughout the year. Consequently, in the late 19th century, astronomers replaced apparent solar time with mean solar time as the primary reference for time keeping.

Mean Solar Time and Universal Time

Historically, to avoid the variation in the Sun's apparent motion along the celestial equator, we've introduced a *Fictitious Mean Sun* proposed by Simon Newcomb (1835–1909) in 1895. He defined it to have nearly uniform motion along the celestial equator. The right ascension of this fictitious mean Sun is

$$\alpha_{FMS} = \theta_{GST} = 18^\circ 38' 45.836'' + 8,640,184.542'' T + 0.0929'' T^2$$

where T is measured in Julian centuries from Greenwich at noon, January 0, 1900. The diurnal motion of the fictitious mean Sun as measured by its hour angle is essentially identical to the average diurnal motion of the apparent Sun, neglecting variations in the local meridian due to polar motion and in the Earth's rotation rate. We can develop mathematical relations from Eq. (1-47) using the fictitious mean Sun. This practice was discontinued in 1984 when new standards were adopted.

Today, the U.S. Naval Observatory determines the value of mean solar time through GST and corrections for the Earth's rotation. ***Mean solar time*** still refers to an equinox that has only secular motions, but it's based on GST , which ultimately depends on observations of the radio galaxies. We define the mean solar time at Greenwich as ***universal time***, UT , and the difference between apparent and mean solar time as the ***equation of time***, EQ_{time} . This is essentially the difference in the Sun's true right ascension and the right ascension of the fictitious mean Sun. Seidelmann (1992, 484) gives the complete expression using the mean anomaly of the Sun, M_\odot , from Sec. 4.2:

$$EQ_{time} = -1.914\,666\,471^\circ \sin(M_\odot) - 0.019\,994\,643 \sin(2M_\odot) + 2.466 \sin(\lambda_{ecliptic}) - 0.0053 \sin(4\lambda_{ecliptic}) \quad (1-48)$$

During the year, the difference in apparent and mean solar time varies from minus 14 minutes to plus 16 minutes (Seidelmann, 1992, 6).

There are three important distinctions for UT . For precise applications, we must distinguish $UT0$, $UT1$, and $UT2$, all of which I'll discuss shortly. The differences are small. ***Greenwich mean time***, the local time in England, is often incorrectly assumed to be UT .

It's useful to introduce procedures that determine the various forms of UT and to distinguish their differences. $UT0$ is the observation of UT at a particular ground station. Determining $UT0$ depends on the apparent diurnal motion of the stars. At first this seems to contradict the concept of solar time already introduced; however, instruments can't mea-

sure the Sun's motion at a particular location with enough precision. The practice is to measure the locations of the radio galaxies and apply that information to determine solar time. For an observer at a known longitude, λ , $UT0$ is sometimes calculated as 12 hours plus the *Greenwich Hour Angle*, GHA_{\odot} :

$$UT0 = 12 \text{ h} + GHA_{\odot} = 12 \text{ h} + LHA_{\odot} - \lambda$$

In this manner, if the object is on the meridian at Greenwich ($GHA = 0^\circ$), it's 1200 local time (Green, 1988, 37). Supposing $GHA = 2$ hours, $UT0$ is then 1400, which makes sense because the Sun has already passed the vertical at noon. To stay *consistent* with established practice, the convention for this book will be to measure GHA positive west and negative east. Beware—this is a left-handed system.

We correct $UT0$ for polar motion so time is independent of station location (Green, 1988, 245) to obtain $UT1$. The second factor is basically a correction applied to the station's nominal longitude, λ .

$$UT1 = UT0 - (x_p \sin(\lambda) + y_p \cos(\lambda)) \tan(\phi_{gc})$$

The longitude, λ , and the x_p and y_p are coefficients of the instantaneous positions of the pole and are discussed on page 82. The geocentric latitude, ϕ_{gc} , is the latitude of the observing site. The USNO calculates $UT1$ by correlating several worldwide observations of radio galaxies. Differences from $UT0$ are typically about 30 milliseconds.

Finally, $UT1$ is corrected for seasonal variations to yield $UT2$. Modern programs seldom use $UT2$, but the USNO uses it to evaluate clock performance.

Converting between Solar Time and Sidereal Time

Recall Fig. 1-30 showing the general geometry of the solar and sidereal days. A *mean sidereal day* consisting of 24 *sidereal* hours is defined as the time for the Earth to rotate once on its axis relative to the stars. You can convert from one system to the other as shown in the *Astronomical Almanac* (1984, S15).

$$1 \text{ solar day} = 1.002\,737\,909\,350\,795 \text{ sidereal day}$$

$$1 \text{ sidereal day} = 0.997\,269\,566\,329\,084 \text{ solar day}$$

$$1 \text{ solar day} = 24 \text{ h } 3 \text{ min } 56.555\,367\,8 \text{ s sidereal time}$$

$$1 \text{ sidereal day} = 23 \text{ h } 56 \text{ min } 4.090\,524 \text{ s solar time}$$

Although these conversions between times appear simple, differences occur because of the moving equinox. An expression which relates the motion of the equinox, and therefore the variations in these relations (*Astronomical Almanac*, 1984, S16), is

$$\begin{aligned} 1 \text{ solar day} &= 1.002\,737\,909\,350\,795 + 5.9006 \times 10^{-11} T_{UT1} \\ &\quad - 5.9 \times 10^{-15} T_{UT1}^2 \text{ sidereal day} \end{aligned} \tag{1-49}$$

where T_{UT1} is the number of Julian centuries ($UT1$) since the J2000 epoch, such that T_{UT1} will be negative for all times occurring before the year 2000.

Coordinated Universal Time

The most commonly used time system is *coordinated universal time*, *UTC*, which is derived from atomic time. It's designed to follow *UT1* within ± 0.9 s ($\Delta UT1 = UT1 - UTC$). *UTC* is the basis of civil time systems and is on ordinary clocks. This definition of *UTC* was introduced in January, 1972, to provide a bridge between *TAI* and *UT1*. It's sometimes called *Zulu* time, but you should use *UTC* to be precise. Because *UT1* varies irregularly due to variations in the Earth's rotation, we must periodically add *leap seconds* to keep the two time scales in close agreement. The U.S. Naval Observatory adds these seconds at the end of June or December and publishes changes through the news media. *UTC* always differs by an integer number of seconds from *TAI*. *UTC* and a $\Delta UT1$ estimate have been broadcast on WWV since January 1, 1972.

UTC originated in 1962 as a cooperative international effort to provide a consistent time standard for broadcasting. Raw observations were collected from around the world and statistically processed (including seasonal variations) to provide a time standard which was accurate to within 0.1 seconds of *UT2*. This necessitated periodic step adjustments of 0.1 seconds. This process was changed in 1972 when the current *UTC* was adopted.

For civilian timekeeping, we define a time zone for a particular region. Using the exact longitude for calculations will be more accurate because it defines the time *for the specific longitude*—not just a general region. Table 1-4 lists common time zones in the United States and their conversion to *UTC*. Be aware that *UTC* does *not* change for daylight savings time. (As an aside, Benjamin Franklin suggested daylight savings time, and the U.S. started using it during World War I to save energy used to produce electricity.)

TABLE 1-4. *UTC* Offsets for Selected Time Zones. The central meridian defines the difference in *UTC* from the Greenwich meridian. Hawaii doesn't use daylight savings time.

Time Zone	Standard	Daylight Savings (April to October)	Central Meridian
Atlantic	<i>UTC</i> – 4 h	<i>UTC</i> – 3 h	-60°
Eastern	<i>UTC</i> – 5 h	<i>UTC</i> – 4 h	-75°
Central	<i>UTC</i> – 6 h	<i>UTC</i> – 5 h	-90°
Mountain	<i>UTC</i> – 7 h	<i>UTC</i> – 6 h	-105°
Pacific	<i>UTC</i> – 8 h	<i>UTC</i> – 7 h	-120°
Alaska	<i>UTC</i> – 9 h	<i>UTC</i> – 8 h	-135°
Hawaii	<i>UTC</i> – 10 h	<i>UTC</i> – 10 h	-150°

Julian Date

A very important concept in astrodynamics is the Julian date. The **Julian date**, JD , is the continuous amount of time measured in days from the epoch January 1, 4713 B.C., 1200. The actual determination was made by Joseph Scaliger in 1582. He combined the solar cycle (28 years), the Metonic cycle (19 years), and the Roman Indication (15 years) to create a **Julian period** consisting of 7980 **Julian** years (365.25 days). Because the three cycles were already established, the only common point for these cycles was 4713 B.C.—hence, the epoch value. Interestingly enough, the Julian period was named for Scaliger's father and not Julius Caesar as the name might suggest (Wertz, 1978, 20). The convention to start the JD at noon each day benefits astronomers (who often work at night) because they can make all their observations on a single *day*. It's also the time that we could best use to reckon solar days. To find the Julian date from a known date and time within the period March 1, 1900 to February 28, 2100, we use Algorithm 2, as shown below:

ALGORITHM 2: Julian Date ($yr, mo, d, h, min, s \Rightarrow JD$)

$$JD = 367(yr) - \text{INT} \left\{ \frac{7 \{ yr + \text{INT} \left(\frac{mo + 9}{12} \right) \}}{4} \right\} + \text{INT} \left(\frac{275mo}{9} \right) + d + \\ \frac{\left(\frac{s}{60} + min \right)}{60} + h \\ 1,721,013.5 + \frac{24}{24}$$

where the year, month, day, hour, minute, and second are known. The INT relation denotes *real* truncation (see Sec. A.2). The year must be four digits and not the commonly abbreviated two-digit value. This empirical formula is designed to be accurate for the period described above. Its use is restricted because the JD is continuous whereas the calendar contains periodic steps through the addition of leap years and seconds. Algorithm 2 is valid for any of the time systems we've discussed and should be identified to avoid confusion (JD_{UT1} , JD_{TDT} , JD_{TDB} , ...). Unless specified, the standard Julian date is JD_{UT1} .

The values are typically quite large, so we may be tempted to use only a few decimal places. However, because the units are days, I recommend retaining at least 8 decimal digits to provide reasonable accuracy (about 4×10^{-4} s). Some authors choose to use a *Modified Julian Date*, MJD , calculated as follows:

$$MJD = JD - 2,400,000.5$$

Although this adjustment reduces the size of the date (about two significant digits), it adds a layer of potential confusion to calculations. Meeus (1991, 61) gives a general formula without restriction. We treat January and February as months 13 and 14, respectively. If

$mo = 1$ or 2 , $yr = yr - 1$ and $mo = mo + 12$. The daily time is added as in the final term of Algorithm 2. For Julian calendar dates, set $B = 0$.

$$B = 2 - \text{INT}\left(\frac{yr}{100}\right) + \text{INT}\left(\frac{\text{INT}\left(\frac{yr}{100}\right)}{4}\right)$$

$$JD = \text{INT}\{365.25(yr + 4716)\} + \text{INT}\{30.6001(mo + 1)\} + d + B - 1524.5$$

Because the JD is often used as an epoch value for observations, a shorthand notation such as J1900 or J2000 is common. To be precise, I've shown the J2000 epoch using TDB (or TT), although we often use $UT1$ for practical problems. Some commonly used epochs are

$$J2000 = 2,451,545.0 = \text{January 1, 2000 12:00 } TDB$$

$$J1900 = 2,415,021.0 = \text{January 1, 1900 12:00 } UT1$$

$$B1900 = 2,415,020.313\ 52 = \text{December 31, 1899 19:31:28.128 } UT1$$

$$B1950 = 2,433,282.423\ 459\ 05 = \text{December 31, 1949 22:09:46.862 } UT1$$

The notations of B1900 and B1950 are for *Besselian years*. Besselian years are sometimes denoted by the year with a decimal point, e.g., 1950.0 which is the same as B1950. I've included them as a reference to previous systems; highly precise studies typically don't use them today. Example 1-4 shows how to find the Julian date.

▼ Example 1-4. Determining the Julian Date.

GIVEN: October 26, 1996, at 2:20 P.M. UT

FIND: JD

From Algorithm 2, $JD = 367(1996)$

$$JD = 732,532.0 - 3494.0 + 305.0 + 26 + 1,721,013.5 + 0.597\ 222 + 2,450,383.097\ 222\ 22 \\ - \text{INT}\left\{\frac{7\{1996 + \text{INT}\left(\frac{10+9}{12}\right)\}}{4}\right\} + \text{INT}\left(\frac{275(10)}{9}\right) + 26 + 1,721,013.5 + \frac{\frac{0}{60} + 20}{\frac{60}{24}} + 14$$

$$JD = 732,532.0 - 3494.0 + 305.0 + 26 + 1,721,013.5 + 0.597\ 222 + 2,450,383.097\ 222\ 22$$

Notice the use of 14 hours, not 2, for 2 P.M. This is probably the most common error when doing calculations with the Julian date. Keep as many digits as possible, so finding the input values will be easier.

▲ Remember that one second = 1.1574×10^{-5} days!

The JD provides a continuous, simple, concise method of preserving year-month-day-hour-minute-second information in one variable, which is especially nice for computer applications. In addition, many relations for astronomical equations of motion use the number of Julian centuries from a particular epoch. It's important to distinguish particular types of time with subscripts (T_{UT1} , T_{TDB} , ...). The general formula is

$$T_{xxx} = \frac{JD_{xxx} - 2,451,545.0}{36,525} \quad (1-50)$$

Day of the Year

The **day of the year** has integer values from 1 to 365 (366 in a leap year). Table 1-5 shows the values throughout the year, as well as elapsed time from January 1.

TABLE 1-5. Day of the Year. This table lists numbers for certain days of the year. The elapsed days from the beginning of the year are useful to calculate time. Leap-year values are in parentheses.

Date (0 h)	Day of the Year	Elapsed Days from Jan 1, 0 h	Date (0 h)	Day of the Year	Elapsed Days from Jan 1, 0 h
January 1	1	0	July 1	182 (183)	181 (182)
January 31	31	30	July 31	212 (213)	211 (212)
February 28	59	58	August 31	243 (244)	242 (243)
March 31	90 (91)	89 (90)	September 30	273 (274)	272 (273)
April 30	120 (121)	119 (120)	October 31	304 (305)	303 (304)
May 31	151 (152)	150 (151)	November 30	334 (335)	333 (334)
June 30	181 (182)	180 (181)	December 31	365 (366)	364 (365)

Notice that the days are recorded for the end of each month, so we can determine a particular day by adding to the last day of the previous month. For example, May 14 in a leap year is day $121 + 14 = 135$. A fractional part of a day based on *UTC* is often added. We'll examine routines to process these times in Chap. 3.

1.6.3 Atomic Time

We still need to define a highly accurate time system which is independent of imperfectly realized coordinate frames. With the small variations of all coordinate systems, we'd like to avoid attaching the time system to any coordinate system, but this isn't possible. Instead, we can base it on a highly *regular* occurrence. **International Atomic Time**, *temps atomique international*, or *TAI*, is based on counting the cycles of a high-frequency electrical circuit maintained in resonance with a cesium-133 atomic transition. It's considered the most accurate time because it is independent of variations in the Earth's rotation. One *second*, *Système International* or *SI*, equals the duration of 9,192,631,770 periods of the wavelength associated with the radiation emitted by electron transition between two hyperfine levels of the ground state of cesium-133. The official introduction of *TAI* dates to January 1972 but it's been available since 1955. *TAI* is the basis for interpolating and extrapolating all other time scales today.

TAI is a product of the Bureau International des Poids et Mesures in Sèvres, France and is based on the analysis of individual standards for commercial frequencies plus the primary frequency standards maintained by several countries. Data from roughly 200 frequency standards are weighted and statistically combined with known corrections to maintain a unit of time as close to the ideal *SI* second as possible. The adjusted time scale is published as *TAI* along with corrections to each of the contributing time standards.

TAI achieves a precision that permits the observation of relativistic effects for clocks in motion or under a variation in the local gravitational field. It's consistent for an observer on the reference ellipsoid associated with a geocentric coordinate frame, and with a unit of time specified by the *SI* second as measured on the rotating geoid.

1.6.4 Dynamical Time

Both solar and sidereal time fluctuate under the combined effects of the Sun, Moon, and planets. In the early 1960s, *ephemeris time*, *ET*, was created to provide a more stable time reference than those based on the Earth's variable rotation. As additional effects were discovered, it became apparent that relativistic effects were significant. Many of the astronomical equations of motion now reference the barycenter of the solar system. *Terrestrial dynamical time*, *TDT*, and *barycentric dynamical time*, *TDB*, were adopted in 1976 to replace *ET* and provide a link to the barycentric-referenced equations of motion. In 1991, a further refinement was adopted which defined *terrestrial time*, *TT*, *geocentric coordinate time*, *TCG*, and *barycentric coordinate time*, *TCB*. Consult Seidelmann (1992, 41–48) and McCarthy (1992) concerning the development and use of these highly accurate representations of time.

Terrestrial dynamical time, *TDT*, is the “*theoretical timescale of apparent geocentric ephemerides of bodies in the solar system*” (Seidelmann, 1992, 42). *TDT* is independent of equation of motion theories and uses the *SI* second as the fundamental interval. It derives directly from *TAI*, as seen in Eq. (1-51) and is related to other times as follows:

$$\begin{aligned} UTC &= UT1 - \Delta UT1 \\ TAI &= UTC + \Delta AT \\ TDT &= TAI + 32.184 \text{ s} \end{aligned} \quad (1-51)$$

where ΔAT and $\Delta UT1$ are the accumulated time differences from the specified times. We can get both values from periodic publications of Circular D of the Bureau International de l'Heure (BIH), the *Earth Orientation Bulletin Series 7*—IERS published by the USNO, or the *Astronomical Almanac* (Annual, Section K). Other sources exist—check Section D. The 32.184-second difference with the atomic time is the predefined constant offset between the two systems. It arises because it was the current value in 1977 when the change of systems was adopted. Representative values are in Table 1-6.

Barycentric dynamical time, *TDB*, is defined as the “*independent variable of [the] equations of motion with respect to the barycenter of the solar system*” (Seidelmann, 1992, 42). The difference from *TDT* is that *TDB* refers to the solar system's barycenter, as opposed to the Earth's center, and it depends on dynamical theory. *TDB* is more difficult to represent because it includes relativistic effects and requires *TDT*, *UT1*, gravitational constants, and many other parameters. Green (1988, 253–254) shows the development. Begin with a conversion between barycentric and terrestrial time:

$$TDB = TDT + \frac{2r_s}{a_e n_e} e_e \sin(E) + \text{other} \quad (1-52)$$

TABLE 1-6. Timing Coefficients. This table shows values for the difference in atomic time ($\Delta AT = TAI - UTC$) as well as coordinated and universal time ($\Delta UT1 = UT1 - UTC$). ΔAT remains constant until changed, whereas $\Delta UT1$ changes continuously. Offsets to the GPS satellites are important when determining navigational information. $GPS - UTC = \Delta AT - 19$ s. (Source: *Astronomical Almanac*, 1995, K9, and USNO)

Date	ΔAT (s) (After Jan 1)		$\Delta UT1$	Date	ΔAT (s) (After Jan 1)		$\Delta UT1$
			$UT1 - UTC$ (s)				$UT1 - UTC$ (s)
1984 Jan 1	22		0.395 000	1990 Jul 1	25		-0.038 670
1984 Jul 1	22		0.097 500	1991 Jan 1	26		0.618 670
1985 Jan 1	22		-0.158 500	1991 Jul 1	26		0.226 070
1985 Jul 1	23		0.549 200	1992 Jan 1	26		-0.125 300
1986 Jan 1	23		0.312 600	1992 Jul 1	27		0.442 910 0
1986 Jul 1	23		0.707 000	1993 Jan 1	27		0.062 020 0
1987 Jan 1	23		-0.137 900	1993 Jul 1	28		0.598 830 0
1987 Jul 1	23		-0.397 600	1994 Jan 1	28		0.199 280 0
1988 Jan 1	24		0.364 300	1994 Jul 1	29		0.782 670 0
1988 Jul 1	24		0.089 700	1995 Jan 1	29		0.398 500 0
1989 Jan 1	24		-0.116 600	1995 Jul 1	29		-0.061 470 0
1989 Jul 1	24		-0.386 170	1996 Jan 1	30		0.555 290 0
1990 Jan 1	25		0.328 640	1996 Jul 1	30		0.187 120 0

using the *Schwarzschild radius* ($r_s = 1.478$ km), the Earth's semimajor axis about the Sun ($a_e = 149,598,023$ km), the Earth's mean motion about the Sun ($n_e = 1.991 \times 10^{-7}$ rad/s), and the eccentricity of the Earth's orbit ($e_e = 0.016\ 708\ 617$). "Other" includes small effects contributed by the Moon and planets, as well as the Earth's diurnal motion. The Earth's eccentric anomaly, E , is obtained from the mean anomaly, more often encountered in computations. See Eq. (1-59). Later, I'll show a way to convert between the two values [Eq. (4-51)], but we can use Eq. (4-51) now to find a "second-order" result:

$$E \equiv M_{\oplus} + e_e \sin(M_{\oplus}) + \frac{e_e^2}{2} \sin(2M_{\oplus})$$

Be careful to use radians in this expression! We also know from Eq. (4-3) that

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

so we can now use these two results to find an approximate expression:

$$e_e \sin(E) \approx e_e \sin(M_{\oplus}) + \frac{e_e^2}{2} \sin(2M_{\oplus})$$

Substituting into Eq. (1-52), we get

$$\begin{aligned}
 TDB &= TDT + 0.001\,658^s \sin(M_{\oplus}) + 0.000\,013\,85 \sin(2M_{\oplus}) \\
 &\quad + \text{lunar/planetary terms} + \text{daily terms} \\
 M_{\oplus} &\equiv 357.527\,723\,3^{\circ} + 35,999.050\,34 T_{TDB}
 \end{aligned} \tag{1-53}$$

Recognize the mean anomaly from Eq. (1-59) is used interchangeably for the Earth and the Sun (*Astronomical Almanac*, 1984, S26), and T_{TDB} is the Julian centuries of TDB . I'll introduce the mean anomaly in Sec. 4.2. Be aware that some formulations use a "first-order" approximation, which results in

$$TDB \cong TDT + 0.001\,658\,s \sin(M_{\oplus} + 0.016\,708\,617 \sin(M_{\oplus}))$$

and isn't quite as accurate as Eq. (1-53). Rigorous formulas including relativistic effects aren't necessary for this book, but you can find them in Seidelmann (1992, 42–44). It's common practice to drop the lunar or planetary (order $1\times 10^{-5}\,s$) and daily terms (order $1\times 10^{-6}\,s$) because they're small. Notice, however, the interdependence of the TDB calculation with T_{TDB} . This naturally suggests some form of iteration to determine a solution. Frank (1984, 5–6) shows that the approximate equations

$$\begin{aligned}
 TDB &\cong TDT + 0.001\,658\,s \sin(M_{\oplus}) + 0.000\,013\,85 \sin(2M_{\oplus}) \\
 M_{\oplus} &\cong 357.527\,723\,3^{\circ} + 35,999.050\,34 T_{TDT} \\
 &\text{in radians;} \\
 M_{\oplus} &\cong 6.240\,035\,939 + 628.301\,956\,0 T_{TDT}
 \end{aligned} \tag{1-54}$$

will yield accuracy comparable to that using the neglected lunar and planetary terms while eliminating the need to iteratively determine TDB . The radian expression is included for computer software. The usual process determines the Julian date referenced to $UT1$ because $UT1$ is most commonly available. However, we must convert to TDB , TDT , and TAI systems for many applications.

While the preceding equations give us a way to transfer between the time scales, we often must determine the number of centuries for use with precession, nutation, and other calculations involving series expansions. The following formulas provide a means to determine the number of centuries elapsed from the epoch J2000. Notice this equation is similar to Eq. (1-50); however, the basic time unit is different. A summary algorithm and an example are useful at this point to tie the concepts of time together.

ALGORITHM 3: *ConvTime* (*yr, mo, day, UTC, ΔUT1, ΔAT* ⇒ *UT1, TAI, TDT, TDB*)

$$UT1 = UTC + \Delta UT1$$

$$TAI = UTC + \Delta AT$$

$$TDT = TAI + 32.184\,s$$

Julian Date (*yr, mo, day, h, min, s_{TDT}* \Rightarrow *JD_{TDT}*)

$$T_{TDT} = \frac{JD_{TDT} - 2,451,545.0}{36,525}$$

$$M_{\oplus} \equiv 6.240\,035\,939 + 628.301\,956\,0 T_{TDT}$$

$$TDB \equiv TDT + 0.001\,658\text{ s } \sin(M_{\oplus}) + 0.000\,013\,85 \sin(2M_{\oplus})$$

Julian Date (*yr, mo, day, h, min, s_{TDB}* \Rightarrow *JD_{TDB}*)

$$T_{TDB} = \frac{JD_{TDB} - 2,451,545.0}{36,525}$$

▼ Example 1-5. Calculating Dynamical Time.

GIVEN: May 14, 1990, 10:43 Mountain Standard Time (UTC based)

FIND: $UT1, TAI, TDT, TDB, T_{UT1}, T_{TDT}, T_{TDB}$

Obtain ΔAT (25.0 s) directly from the *Astronomical Almanac*. Unfortunately, $\Delta UT1$ changes continuously and isn't listed directly. Appendix D lists sources to find values. You must find correct values as $\Delta UT1$ varies about ± 0.9 s. For precise work, a timing error this large can easily result in almost a 10 km error! From the USNO's IERS final data, $\Delta UT1 = 0.046\,94$ s for May 14, 1990. Remember that in May, Daylight Savings Time exists in the Mountain Standard Zone. Therefore, from Table 1-4, we must add six hours to UTC. We find $UT1$ as

$$UT1 = UTC + \Delta UT1 = 16:43 + 0.046\,94 = 16:43:0.046\,94$$

As a check, UTC and $UT1$ should be within 0.9 seconds. Next, find the atomic time:

$$TAI = UTC + \Delta AT = 16:43:0.00 + 25\text{ s} = 16:43:25.0$$

Notice that TAI is always within an integer number of seconds of UTC. Dynamical time is then

$$TDT = TAI + 32.184\text{ s} = 16:43:25.0 + 32.184\text{ s} = 16:43:57.184$$

TDB depends on the Julian centuries of TDT , which you can calculate knowing the Julian date of TDT :

$$DMY\,HMS \Rightarrow JD_{TDT} = 2,448,026.197\,189\,629\,6$$

Now find the Julian centuries of TDT :

$$T_{TDT} = \frac{JD_{TDT} - 2,451,545.0}{36,525} = \frac{2,448,026.197 - 2,451,545.0}{36,525} = -0.096\,339\,570\,441\,3$$

Find the mean anomaly for the Earth. Be sure to reduce the result to the interval of $\pm 360^\circ$.

$$M_{\oplus} \equiv 357.527\,723\,3^\circ + 35,999.050\,34 T_{TDT} = -230.605\,322\,7^\circ$$

Now, determine TDB :

$$TDB \equiv TDT + 0.001\,658\text{ s } \sin(M_{\oplus}) + 0.000\,013\,85 \sin(2M_{\oplus}) = 16:43:57.185\,267\,7$$

Finally, determine additional Julian centuries for specific applications. Remember to first find the Julian date for each time before determining the centuries. ($JD_{TDB} = 2,448,026.197\,189\,644\,3$; $JD_{TAI} = 2,448,026.196\,817\,129\,6$)

$$T_{TDB} = \frac{JD_{TDB} - 2,451,545.0}{36,525} = -0.096\,339\,570\,440\,9$$

In summary, the results are as follows. Note that we use Eq. (1-50) to find the Julian centuries for $UT1$.

$$UT1 = 16:43:00.046\ 94T_{UT1} = -0.096\ 339\ 588\ 546\ 9$$

$$UTC = 16:43:00.000\ 00$$

$$TAI = 16:43:25.000\ 00$$

$$TDT = 16:43:57.184\ 00T_{TDT} = -0.096\ 339\ 570\ 441\ 3$$

$$TDB = 16:43:57.185\ 27T_{TDB} = -0.096\ 339\ 570\ 440\ 9$$

▲ You can reverse this entire process, as long as you attend carefully to the signs.

1.7 Motion of the Coordinate System

An observer makes observations at some time from a given site. The observations are usually referred to the local horizon system (*SEZ*), e.g., range, range rate, azimuth, and elevation angles. To use the observations for calculations, we must tie them to an inertial frame. Thus we must transform to the *ECEF* frame and then to the *ECI* frame. The *ECEF* frame depends on the Earth's equator, as does the *ECI* frame. However, the Earth's equator is not fixed in inertial space. The *ECI* frame also depends on the vernal equinox, which is tied to the ecliptic plane. This plane also moves due to perturbations by the planets on the Earth's orbital plane, which defines the ecliptic. Hence, the equator and equinox are time-dependent parameters, and the *ECI* frame is time dependent with a wobbling equator and a sliding vernal equinox. The Moon contributes additional motion, called astronomical nutation, to the *ECI* frame. Thus the *ECI* frame is not very "inertial" unless we fix it at a specified time.

In practical terms the motion is quite small and can be ignored over a few days. This is not true for longer intervals. If observations from sufficiently different epochs are to be combined for processing, we need a way to transform between the respective *ECI* frames, so the observations can be brought to a common frame. These different *ECI* frames are designated by the epoch or date of the equator and equinox for that frame. To do so, we must develop expressions for the effects of precession and nutation and designate a standard *ECI* frame from which all others may be referred. We can perform these reductions on the current inertial system (*J2000, FK5*) or its predecessor, *B1950*, or any other systems of choice. I'll talk briefly about the *B1950 (FK4)* epoch in the upcoming section on *FK4* reduction. The *B1950* system has noticeable errors, so we should avoid it whenever possible. See Seidelmann (1992, 167–187) and Moritz and Mueller (1987, 525–567) for other transformations.

A truly inertial coordinate frame is difficult, if not impossible, to realize. For precise applications, the *J2000 (FK5)* system is considered an inertial frame. Most analyses use a coordinate system which is *sufficiently inertial* for the particular application. These systems are ones in which Newton's laws of motion are valid. In the case of orbits around the Sun—planets, asteroids, comets and some deep space probes—the heliocentric coordinate system is usually sufficiently inertial. For satellites orbiting the Earth, the geocentric equatorial system is sufficient.

The coordinate system must be *inertial* and not accelerating. In developing his law of gravitation, Newton described this inertial reference frame as fixed in "*absolute space*,

which in its own nature, without relation to anything external, remains always similar and immovable" (Newton [1687] 1962, 6). He also realized the frame includes systems moving with *constant* velocity. But he didn't show how to find such a frame. This turned out to be a rather difficult problem. In fact, only the extra-galactic coordinate system (origin at the center of the galaxy) approaches Newton's definition. The ideal solution relies on finding **pseudoinertial** reference frames, which we'll define as an ideal system for which Newtonian mechanics hold. Sec. 1.4 discusses the existence of many different coordinate systems, some of which may be considered pseudoinertial systems. Depending on the particular problem, we can consider most of the coordinate systems in Sec. 1.4 as pseudoinertial because the small noninertial effects (1) can't be measured with current technology, or (2) are assumed to be negligible during the period of interest. Properly choosing one of these pseudoinertial coordinate systems is important for astrodynamics. In practice, the J2000 system is the standard inertial coordinate system used for all precise Earth-centered applications.

In a stricter sense, determining truly inertial systems is different when we account for relativity. Solutions in this area use **quasi-inertial** systems, which have special relativistic forms. Don't confuse these systems with the pseudoinertial systems sufficient for Newtonian mechanics. Seidelmann (1992, 96) gives insight concerning quasi-inertial systems:

There are finite regions with respect to a chosen space of reference where material particles move freely without acceleration and in which the laws of special relativity hold with remarkable accuracy, which can be regarded as quasi-inertial frames. Thus the center of mass of the Earth-Moon system falling in an elliptic orbit around the Sun in a relatively weak gravitational field is an example of such a finite region. Therefore, we may assume quasi-inertial frames of reference, and any violation of the principles when using classical mechanics can be taken into account with small corrections applied to the observations and by an appropriate coordinate-time reference. The effects of special relativity for a system moving with the Earth around the Sun are of the order 10^{-8} ; likewise those with general relativity are of order 10^{-8} . It is necessary to distinguish between various types of quasi-inertial frames.

To understand the motion of a coordinate system, we begin with the geometry of the Earth's orbit. The equatorial planes of the Earth and the ecliptic are not actually fixed relative to the stars. Several forces cause these discrepancies. First, the gravitational forces of the planets affect the Earth's orbit through **planetary precession**, which results in a very slow secular change in the ecliptic's orientation. This causes the equinox to precess to the west about $0.0033^\circ(12'')$ per century and decreases the obliquity of the ecliptic about $0.013\ 05^\circ(47'')$ per century. Next, because the Earth isn't spherical (significant equatorial bulge), the Sun's and Moon's gravitational fields produce a very small torque on the Earth causing **luni-solar precession**. This torque results in a smooth wobbling or precessional motion over a very long period—similar to a spinning top. Because the obliquity of the ecliptic is about 23.5° , you can see from Fig. 1-32 that the Earth's axis of rotation traces out a roughly circular shape over each period of precession. The half-cone angle equals about 23.5° . The period of the luni-solar precession is about 26,000 years, and the angle precesses about $0.013\ 846^\circ(49.846'')$ per year. The combined effects of planetary and luni-

solar precession are collectively called ***general precession***, which amounts to about 0.013° 889° (50") per year in longitude.

The Moon is mainly responsible for producing an additional torque on the Earth's equatorial bulge causing ***nutation***—small oscillations in the Earth's rotation axis. This is because the Moon's orbit about the Earth is inclined about 5°, which causes monthly variations in the periodic torque the Moon exerts on the Earth's equatorial bulge. More importantly, the Moon's orbital plane itself precesses with a period of about 18.6 years due to solar perturbations, so nutation from the Moon has a primary period of about 18.6 years. This is more commonly called the ***regression of the node*** because the motion is in a negative direction. Another important contribution is the advance of perigee of the lunar orbit due to solar perturbations. The eccentricity of the Moon's orbit causes the direction of maximum torque to vary. Because the overall nutation effect is actually a complicated result of more than one disturbance, it has many smaller periods, is periodic, and appears as a superimposed motion with a maximum amplitude of about 0.0025°(9") in the obliquity of the ecliptic, and about 0.004° 72°(17") in longitude. Figure 1-32 shows the result of the perturbing forces.

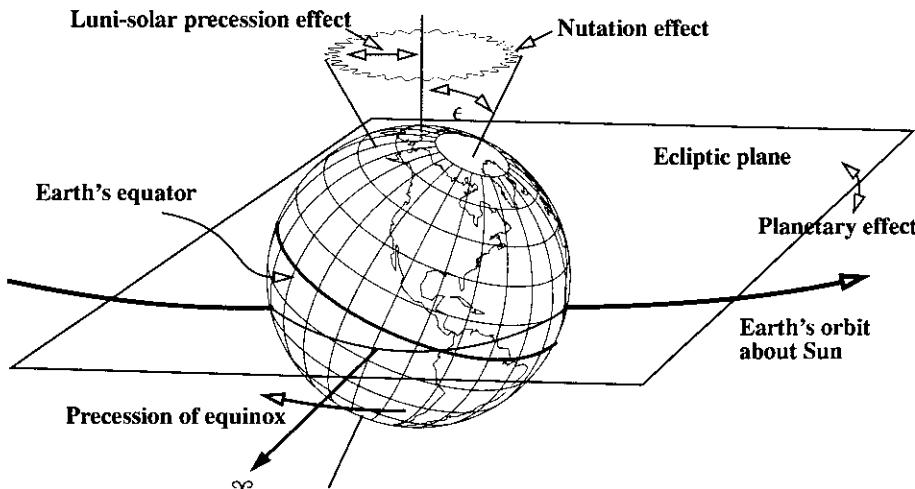


Figure 1-32. Precession of the Earth's Equatorial Plane. This figure shows the combined effects of perturbing forces on the Earth. The general motion of the equinox is to the west. Planetary and nutation effects are small compared to the luni-solar effects.

Because of the many forces inducing perturbations to the Earth's orientation in space, we must introduce a new set of terms to completely specify the frame of reference. The motion of the ecliptic plane due to planetary precession causes the equinox to move along the equator, and we associate its orientation with an ideal inertial frame on a specific date, the ***ecliptic of date***. The Earth's rotation axis also precesses secularly about the North ecliptic pole causing the equator to wobble about the ecliptic. We model this effect as a

mean equator of date measuring the large-amplitude motion of the Earth's rotation axis. The intersection of these two planes on a given date defines the **mean equinox of date**. By including the nutation (periodic small oscillation) effects of the luni-solar perturbations on the equator, we get the **true equator of date**. For precise applications, we must specify the correct reference frame for all measurements (Green, 1988, 54).

Let's examine the transformation of the inertial J2000 system into Earth-Centered, Earth-Fixed, *ECEF*, coordinates of an arbitrary date. Recall that J2000 is the location of an inertial frame at the J2000 epoch, and *ECEF* is a geocentric system at a particular epoch. It's common to compute planetary positions, stellar positions, and precise vectors for satellite orbits in the J2000 frame. We'll look at an example in which the initial location is known in the J2000 system. Then, we'll convert to the *ECEF* system which is at an epoch of our choice. I've adapted this process from Frank (1984), Seidelmann (1992, 99–145), the *Astronomical Almanac* (1992, B18–B19), and the Defense Mapping Agency (1987b, A15–A31).

1.7.1 Precession (J2000)

Resulting from perturbations of the Sun, Moon, and planets, general precession alters the ecliptic's orientation as shown in Fig. 1-33.

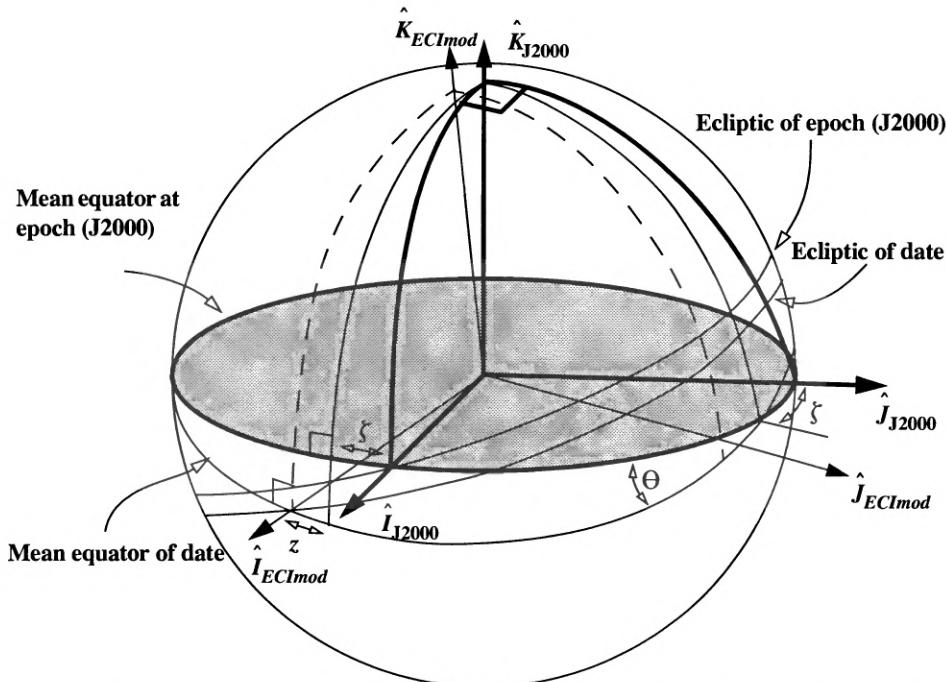


Figure 1-33. Transformation Geometry Due to Precession. Accounting for precession moves from the mean epoch values to mean of date values. Typically, the epoch is J2000. The diagram is greatly exaggerated. In order, the three angles required are ζ , z , Θ .

The transformation is obtained by recognizing that a common point exists between the mean equator of J2000 and the ecliptic of date. This permits a motion along the mean equator of J2000, a translation to the mean equator of date, and a final motion to the location of the equinox of date. The process converts an *ECI* vector in the J2000 system, having a mean equinox of J2000, to an *ECI* vector having the mean equinox of a desired date. The angles ζ , z , and Θ represent the combined effects of general precession. Fortunately, these angles represent secular values and are easily calculated (Seidelmann, 1992, 104) and *Astronomical Almanac* (1984, S19). Eq. (1-55) presents the formulas necessary to determine the precession angles. Notice we typically use a base epoch (J2000 in this case), an arbitrary fixed epoch, JD_{start} , and the epoch of date, JD_{end} . Then, we use these Julian dates (all in *TDB*) to determine the number of Julian centuries required for the formulas.

$$\begin{aligned} T_o &= \frac{JD_{start} - J2000}{36,525} & T_\Delta &= \frac{JD_{end} - JD_{start}}{36,525} \\ \zeta &= (2306.2181'' + 1.39656T_o - 0.000139T_o^2) T_\Delta + (0.30188 - 0.000344T_o) T_\Delta^2 + 0.017998T_\Delta^3 \\ z &= (2306.2181'' + 1.39656T_o - 0.000139T_o^2) T_\Delta + (1.09468 + 0.000066T_o) T_\Delta^2 + 0.018203T_\Delta^3 \\ \Theta &= (2306.2181'' - 0.85330T_o - 0.000217T_o^2) T_\Delta + (-0.42665 - 0.000217T_o) T_\Delta^2 - 0.041833T_\Delta^3 \end{aligned} \quad (1-55)$$

Eq. (1-55) allows direct transformation from a starting to ending date, without having to do an intermediate transformation to the J2000 epoch. In most cases, however, the fixed epoch and the base epoch are the same ($T_o = 0$), and we can simplify the formulas:

$$\begin{aligned} \zeta &= 0.6406161^\circ T_{TDB} + 0.0000839T_{TDB}^2 + 5.0 \times 10^{-6}T_{TDB}^3 \\ z &= 0.6406161^\circ T_{TDB} + 0.0003041T_{TDB}^2 + 5.1 \times 10^{-6}T_{TDB}^3 \\ \Theta &= 0.5567530^\circ T_{TDB} - 0.0001185T_{TDB}^2 - 1.16 \times 10^{-5}T_{TDB}^3 \\ &\quad \text{in radians;} \\ \zeta &= 0.011180860T_{TDB} + 1.464 \times 10^{-6}T_{TDB}^2 + 8.7 \times 10^{-8}T_{TDB}^3 \\ z &= 0.011180860T_{TDB} + 5.308 \times 10^{-6}T_{TDB}^2 + 8.9 \times 10^{-8}T_{TDB}^3 \\ \Theta &= 0.009717173T_{TDB} - 2.068 \times 10^{-6}T_{TDB}^2 - 2.02 \times 10^{-7}T_{TDB}^3 \end{aligned} \quad (1-56)$$

T_{TDB} [Eq. (1-50)] represents the number of Julian centuries of *TDB* from the base epoch (J2000). The order of the rotations is important, and to complete the transformation due to precession, we have

$$\vec{r}_{ECImod} = \text{ROT3}(-z) \text{ROT2}(\Theta) \text{ROT3}(-\zeta) \vec{r}_{J2000mod} \quad (1-57)$$

1.7.2 Nutation (J2000)

The next transformation accounts for the periodic effects contributed primarily by the Moon. The nutation is more lengthy to determine because of the various contributing factors. It consists of over 100 trigonometric terms and is periodic, rather than secular like the precession effect. Adding the effects of nutation transforms the *mean* equator of date to the *true* equator of date.

The first step is to determine the mean obliquity of the ecliptic. In general, the obliquity of the ecliptic (at J2000) is

$$\bar{\epsilon} = 23^\circ 26' 21.448'' = 23.439\ 291^\circ$$

We can express the effect on $\bar{\epsilon}$ as a function of time due to the precession effects on the Earth. Thus, a more precise determination is

$$\bar{\epsilon} = 23.439\ 291^\circ - 0.013\ 004\ 2T_{TDB} - 1.64 \times 10^{-7} T_{TDB}^2 + 5.04 \times 10^{-7} T_{TDB}^3 \quad (1-58)$$

in radians;

$$\bar{\epsilon} = 0.409\ 092\ 80 - 0.000\ 226\ 966T_{TDB} - 2.86 \times 10^{-9} T_{TDB}^2 + 8.80 \times 10^{-9} T_{TDB}^3$$

where T_{TDB} is the Julian centuries from the epoch (J2000) (*Astronomical Almanac*, 1984, S21). Again the radian expression applies to computer software. Notice that the coefficient of the first term of the expansion is the approximate decrease in the obliquity of the ecliptic resulting from planetary precession ($0.013\ 056^\circ$ [$47''$] per century).

Next, we must determine the mean anomalies (Sec. 4.2) for the Moon and Sun, $M_{\mathbb{C}}$, M_{\odot} , the mean argument of latitude of the Moon, $u_{M\mathbb{C}}$, measured on the ecliptic from the mean equinox of date, the mean elongation from the Sun, D_{\odot} , and the longitude of the ascending node of the mean lunar orbit, $\Omega_{\mathbb{C}}$.

$$r = 360^\circ$$

$$\begin{aligned} M_{\mathbb{C}} &= 134.962\ 981\ 4^\circ + (1325r + 198.867\ 398\ 1) T_{TDB} + 0.008\ 697\ 2T_{TDB}^2 + 1.778 \times 10^{-5} T_{TDB}^3 \\ M_{\odot} &= 357.527\ 723\ 3^\circ + (99r + 359.050\ 34) T_{TDB} - 0.000\ 160\ 28T_{TDB}^2 - 3.33 \times 10^{-6} T_{TDB}^3 \\ u_{M\mathbb{C}} &= 93.271\ 910\ 3^\circ + (1342r + 82.017\ 538\ 1) T_{TDB} - 0.003\ 682\ 5T_{TDB}^2 + 3.06 \times 10^{-6} T_{TDB}^3 \\ D_{\odot} &= 297.850\ 363\ 1^\circ + (1236r + 307.111\ 480) T_{TDB} - 0.001\ 914\ 17T_{TDB}^2 + 5.28 \times 10^{-6} T_{TDB}^3 \quad (1-59) \\ \Omega_{\mathbb{C}} &= 125.044\ 522\ 2^\circ - (5r + 134.136\ 260\ 8) T_{TDB} + 0.002\ 070\ 8T_{TDB}^2 + 2.22 \times 10^{-6} T_{TDB}^3 \\ \text{in radians, } r &= 2\pi; \\ M_{\mathbb{C}} &= 2.355\ 548\ 394 + (1325r + 3.470\ 890\ 872) T_{TDB} + 0.000\ 151\ 795T_{TDB}^2 + 3.103 \times 10^{-7} T_{TDB}^3 \\ M_{\odot} &= 6.240\ 035\ 9 + (99r + 6.266\ 610\ 6) T_{TDB} - 0.000\ 002\ 797\ 4T_{TDB}^2 - 5.81 \times 10^{-8} T_{TDB}^3 \end{aligned}$$

$$u_{M\zeta} = 1.627\ 901\ 934 + (1342r + 1.431\ 476\ 084) T_{TDB} - 0.000\ 064\ 272 T_{TDB}^2 + 5.34 \times 10^{-8} T_{TDB}^3$$

$$D_{\odot} = 5.198\ 469\ 514 + (1236r + 5.360\ 106\ 50) T_{TDB} - 0.000\ 033\ 408\ 6 T_{TDB}^2 + 9.22 \times 10^{-8} T_{TDB}^3$$

$$\Omega_{\zeta} = 2.182\ 438\ 624 - (5r + 2.341\ 119\ 397) T_{TDB} + 0.000\ 036\ 142 T_{TDB}^2 + 3.87 \times 10^{-8} T_{TDB}^3$$

Once we find these values, we evaluate the trigonometric series to find the nutation in longitude, $\Delta\Psi$, and the nutation in obliquity, $\Delta\epsilon$. Table D-6 lists the coefficients.

$$\begin{aligned}\Delta\Psi &= \sum_{i=1}^{106} (A_i + B_i T_{TDB}) \sin \left\{ a_{1i} M_{\zeta} + a_{2i} M_{\odot} + a_{3i} u_{M\zeta} + a_{4i} D_{\odot} + a_{5i} \Omega_{\zeta} \right\} \\ \Delta\epsilon &= \sum_{i=1}^{106} (C_i + D_i T_{TDB}) \cos \left\{ a_{1i} M_{\zeta} + a_{2i} M_{\odot} + a_{3i} u_{M\zeta} + a_{4i} D_{\odot} + a_{5i} \Omega_{\zeta} \right\}\end{aligned}\quad (1-60)$$

We can improve accuracy by using four terms from the IAU 1980 (Seidelmann, 1992, 117), which add to the expressions in Eq. (1-60). We can also include corrections for planetary effects on the nutation and the obliquity of the ecliptic. This 85-term trigonometric series is discussed in Seidelmann (1992, 116–119), and it also adds to Eq. (1-60).

The correction values from the IAU theory of nutation allow us to calculate the true obliquity of the ecliptic, ϵ .

$$\epsilon = \bar{\epsilon} + \Delta\epsilon \quad (1-61)$$

Fig. 1-34 shows the transformation.

The rotations needed to rigorously transform coordinates for nutation are

$$\vec{r}_{ECI_{tod}} = \text{ROT1}(-\epsilon) \text{ROT3}(-\Delta\Psi) \text{ROT1}(\bar{\epsilon}) \vec{r}_{ECI_{mod}} \quad (1-62)$$

Figure 1-34 shows both the mean and true obliquity of the ecliptic values. Applying nutation transforms the *ECI* vectors in the mean-of-date frame to the true-of-date frame.

We can get a good approximation to the rigorous nutation transformation by expanding in a power series. Because these quantities are small, we can neglect all terms of degree two and higher, which yields

$$[\text{ROT}] = \begin{bmatrix} 1 & -\Delta\Psi \cos(\epsilon) & -\Delta\Psi \sin(\epsilon) \\ \Delta\Psi \cos(\epsilon) & 1 & -\Delta\epsilon \\ \Delta\Psi \sin(\epsilon) & \Delta\epsilon & 1 \end{bmatrix}$$

The neglected terms contribute an error of only about 10^{-12} ; thus, it's suitable for most computations and reduces the number of trigonometric evaluations.

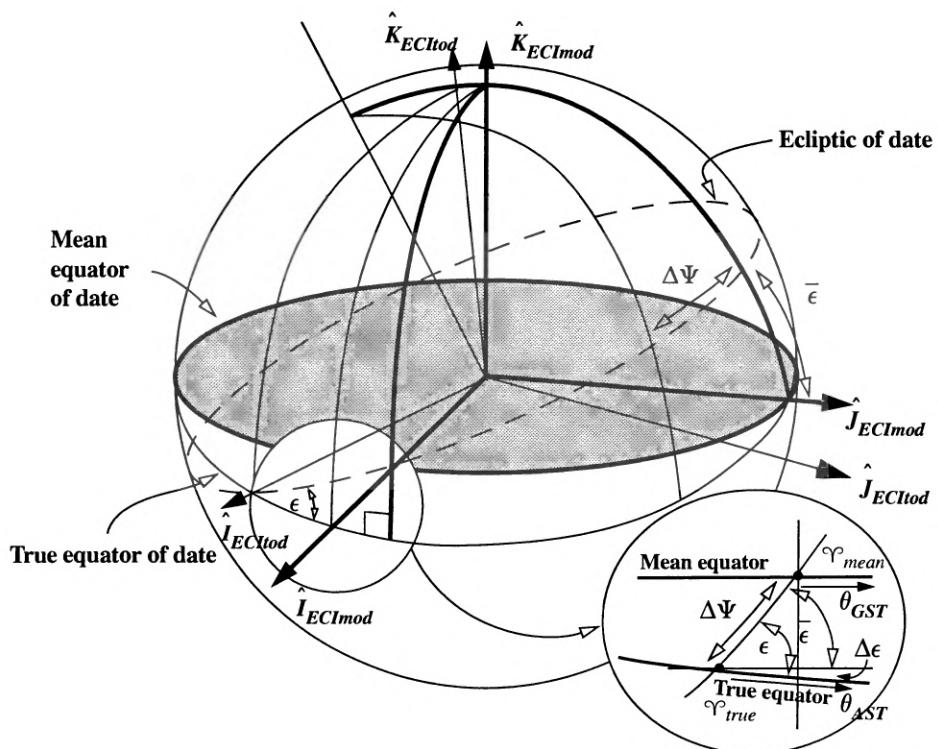


Figure 1-34. Transformation Geometry Due to Nutation. The transformation for nutation moves the state from the mean equinox to the true equinox of date. The starting point is the mean equator and ecliptic of date found from the precession calculations.

1.7.3 Sidereal Time (J2000)

The third transformation usually provides the largest numerical change in the vector components. It transforms the nonrotating true of date frame to the *Earth-fixed* coordinate system. This section addresses the variability in sidereal time and requires the Greenwich *apparent* sidereal time because the *ECEF* frame references the Earth's actual equator of date, which is the true equator. Greenwich apparent sidereal time is measured on the true equator relative to the true equinox, and Greenwich mean sidereal time is measured along the mean equator relative to the mean equinox. The difference between these two times is called the equation of the equinoxes. It's obtained by projecting the difference between the mean and true equinoxes onto the true equator. The formula for the *equation of the equinoxes* is $\Delta\Psi\cos(\epsilon)$, which we saw in the approximate form of the nutation matrix. We use the equation of the equinoxes to convert between apparent and mean sidereal times. We can easily derive this relation from the right triangle having a central angle ϵ . Be aware that after (about) February 26, 1997, there will be two extra terms, $+0.002\ 64'' \sin(\Omega_{\zeta}) +$

$0.000\,063 \sin(2\Omega_\zeta)$, shown in McCarthy (1992, 30). From Fig. 1-34, we find the Greenwich apparent sidereal time, θ_{AST} , using the Greenwich mean sidereal time [Eq. (1-51) and Eq. (1-45)], plus the equation of the equinoxes:

$$\begin{aligned} EQ_{equinox} &= \Delta\Psi \cos(\epsilon) \\ \theta_{AST} &= \theta_{GST} + EQ_{equinox} \end{aligned} \quad (1-63)$$

The rotation is then

$$\dot{\vec{r}}_{ECEFw/OPM} = \text{ROT3}(\theta_{AST}) \dot{\vec{r}}_{ECIItod} \quad (1-64)$$

1.7.4 Polar Motion (J2000)

The final transformation accounts for the North Pole's changing location. The coefficients are measured by the International Polar Motion Service (IPMS) and published by Bureau International de l'Heure (BIH), as well as in the *Astronomical Almanac* (1994, K10, B60). Appendix D shows other sources, and Table 1-7 shows some recent values. The actual displacements are given in arcseconds in the "x" and "y" direction, measured positive south along the 0° and 90°W (270°) longitude meridians, respectively. The *Celestial Ephemeris Pole* (CEP) is the axis of Earth rotation, which is normal to the true equator. The *Conventional International Origin* (CIO) is the *mean* location of the pole agreed upon by international committees. It results from observed values between 1900 and 1905. Deflections locate the CIO from the CEP. The motion is roughly a circular spiral about the North Pole. The maximum amplitude of the variations is about 9 m in any direction. Figure 1-35 shows the geometry for the effect of polar motion.

To do this transformation rigorously involves two rotations:

$$\dot{\vec{r}}_{ECEFwPM} = \text{ROT2}(-x_p) \text{ROT1}(-y_p) \dot{\vec{r}}_{ECEFw/OPM} \quad (1-65)$$

Because these values are very small—typically only a few meters in each direction—we usually avoid the trigonometry for a rigorous coordinate transformation. Instead, we use the trigonometric small-angle approximation to simplify the results. The resultant matrix (with x_p and y_p values in radians) takes the form

$$\dot{\vec{r}}_{ECEFwPM} = \begin{bmatrix} 1 & 0 & x_p \\ 0 & 1 & -y_p \\ -x_p & y_p & 1 \end{bmatrix} \dot{\vec{r}}_{ECEFw/OPM} \quad (1-66)$$

Be aware of two important limitations: (1) The change in values over about one week is significant enough to require updates, assuming you need the correction at all, and (2) we can't predict values with high accuracy over long periods because the motion isn't fully understood. An example may illustrate the importance of including this effect.

TABLE 1-7. Pole Positions. This table shows the coordinates of the North Pole over ten years. Get current values for highly precise studies. The source is the *Astronomical Almanac* (1994, K10). All units are arcseconds ("). Remember that $0.1'' = 0.000\ 028^\circ$, which is about 3.09 m for the surface of the Earth. x_p is measured along the 0° longitude meridian while y_p refers to the 90°W longitude meridian.

Year	Jan 1		Apr 1		Jul 1		Oct 1	
	x_p''	y_p''	x_p''	y_p''	x_p''	y_p''	x_p''	y_p''
1984	-0.125	0.089	-0.211	0.410	0.119	0.543	0.313	0.246
1985	0.051	0.025	-0.196	0.220	-0.044	0.482	0.214	0.404
1986	0.187	0.072	-0.041	0.139	-0.075	0.324	0.062	0.395
1987	0.146	0.315	0.096	0.212	-0.003	0.208	-0.053	0.295
1988	-0.023	0.414	0.134	0.407	0.171	0.253	0.011	0.132
1989	-0.159	0.316	0.028	0.482	0.238	0.369	0.167	0.106
1990	-0.132	0.165	-0.154	0.469	0.161	0.542	0.297	0.243
1991	0.023	0.069	-0.217	0.281	-0.033	0.560	-0.250	0.436
1992	0.182	0.168	-0.083	0.162	-0.142	0.378	0.055	0.503
1993	0.208	0.359	0.115	0.170	-0.062	0.209	-0.095	0.370
1994	0.010	0.476	0.174	0.391	0.137	0.212	-0.066	0.199
1995	0.154	0.418	0.032	0.558	0.280	0.384	0.138	0.106
1996	-0.176	0.191	-0.152	0.506	0.179	0.546	0.267	0.227

The GPS control station receives polar motion and UT corrections in polynomial form once per week. These are *predicted* values. When a succeeding week's polynomial is received, it can't be introduced without incurring a step function either in *ECEF* coordinates or in J2000 coordinates. If we introduce the polar-motion parameters by substituting values and keep the internal files for ephemerides in J2000 coordinates, we introduce a step function into the *ECEF* positions of all of the GPS satellites. But the GPS system is designed to deliver high-quality *ECEF* locations for the user; it can't introduce a step function in *ECEF* coordinates. GPS *ECEF* positions (and velocities) must be continuous. This requirement is met by using complete transformations, including polar motion, and forming the complete derivatives of the J2000-to-*ECEF* transformation. J2000 positions and velocities are transformed to *ECEF* coordinates with the "old" polar motion and ΔUT coefficients at midnight and then transformed back to J2000 coordinates with the "new" coefficients. This introduces a step function in the internal files for J2000 coordinates but preserves continuity of *ECEF* ephemerides. This GPS approach has another benefit. If we don't use it, orbit determination must somehow fit across a small step function (tracking data is collected from stations fixed to the Earth's surface), which degrades accuracy.

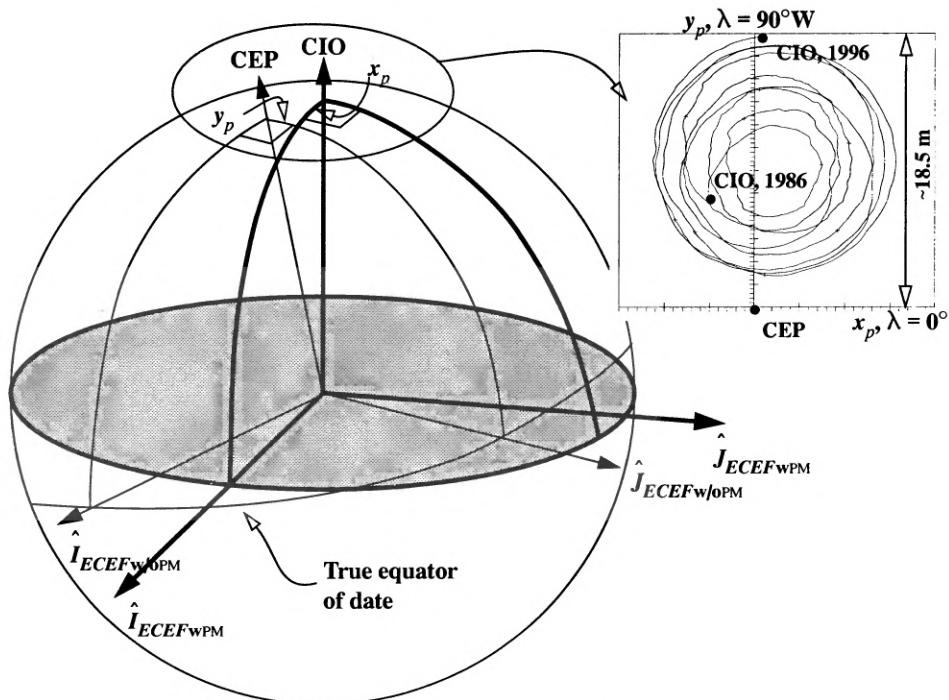


Figure 1-35. Transformation Geometry Due to Polar Motion. Accounting for polar motion takes into account the actual location of the Celestial Ephemeris Pole (CEP) over time. It moves from an ECEF system without polar motion through the CEP, to an ECEF system with polar motion using the Conventional International Origin (CIO). This correction changes the values very little, but highly accurate studies should include it. The inset plot shows the motion for the CIO from May 1986 to May 1996.

The variation in the Pole's exact location also causes a slight shift in terrestrial coordinates of latitude and longitude (considered as mean values in this case). To estimate the offset, Seidelmann (1992, 140) provides these equations:

$$\Delta\phi_{gd} = x_p \cos(\bar{\lambda}) - y_p \sin(\bar{\lambda})$$

$$\Delta\lambda = \{x_p \sin(\bar{\lambda}) - y_p \cos(\bar{\lambda})\} \tan(\bar{\phi}_{gd})$$

1.7.5 Summary (J2000)

We can apply the transformation equations in either direction, of course, recognizing that the order of rotations is important for specific steps of the transformation. To transform completely from one frame to the other, we'll need to do four separate groups of rotations

and conversions: precession, nutation, sidereal time, and polar motion. For the most precise calculations, we must do each of these transformations. Less precise calculations sometimes use just the third transformation (sidereal time). This option is often exercised and seldom documented!

It's important to be able to transform velocity vectors so we can move the complete satellite state vector. However, an important concern arises: there are two ways to account for the transformation of the velocity vector. The first method assumes we have both position and velocity vectors and requires us to add angular corrections. The other method is more common for sites taking observations. In this case, the observations are processed before the position and velocity vectors are formed, so we don't have to correct the transformation of the velocity vector. For this section, we'll discuss the first technique.

The procedure is similar to that already presented, except we change the sidereal time rotation, [ST], to reflect the *ECEF* system's rotation, ω_{\oplus} , with respect to the J2000 system. Abbreviating each rotation, we have

$$\dot{\vec{v}}_{ECEFwPM} = [\text{PM}] \left\{ [\text{ST}] [\text{NUT}] [\text{PREC}] \vec{v}_{J2000} - \vec{\omega}_{\oplus} \times \vec{r}_{ECEFwPM} \right\} \quad (1-67)$$

where ST = sidereal time

PM = polar motion

NUT = nutation

PREC = precision

Notice we adjust the step to correct for sidereal time for the rotating coordinate systems and complete the entire operation *before* doing the final transformation for polar motion. The negative sign in the cross product occurs because the transformation is proceeding *from* the inertial system *to* the noninertial, Earth-fixed system. Also, we need the *ECEF* position vector to complete the process. The literature contains alternate notations. Using the J2000 position vector and the time rate of change of rotation matrix for sidereal time (because it's the only parameter that varies much) gives us the same results [see also Eq. (1-31)]:

$$\begin{aligned} \dot{\vec{v}}_{ECEFwPM} &= [\text{PM}] [\dot{\text{ST}}] [\text{NUT}] [\text{PREC}] \vec{r}_{J2000} + [\text{PM}] [\text{ST}] [\text{NUT}] [\text{PREC}] \vec{v}_{J2000} \\ [\dot{\text{ST}}] &= \begin{bmatrix} -\omega_{\oplus} \sin(\theta_{AST}) & \omega_{\oplus} \cos(\theta_{AST}) & 0 \\ -\omega_{\oplus} \cos(\theta_{AST}) & -\omega_{\oplus} \sin(\theta_{AST}) & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{aligned} \quad (1-68)$$

Although this approach yields identical answers, it doesn't show the dependence on the rotating coordinate systems. Finally, we must reverse the transformation. Be careful of the order of operations and remember to use transposes of rotation matrices.

It's advisable to integrate the equations of motion in an inertial frame. We can integrate in a rotating coordinate system, but the process becomes a bit messy because rotating the

coordinate system introduces additional terms (called Coriolis effects). One inertial frame is the mean equinox, mean equator of J2000, *FK5*, which was defined earlier as the J2000 frame. However, many space centers use other inertial frames to integrate the equations of motion, one being the true equator and mean equinox of date. These other frames are a legacy of earlier computers, when saving a few matrix operations (for precession and nutation) greatly accelerated system throughput. In fact, before computers, extraneous hand calculations were avoided at all costs! Unfortunately, these other frames are more complex because we must keep track of the epoch and frame of the integration and must transform to a new epoch and new frame when updating the reference epoch of an element set or state.

The process for these reductions is rather lengthy, but it's not prohibitive for individual problems. However, for large-scale reductions and long simulations, involving hundreds and thousands of reductions, we use either approximate relations or a numerical technique. The numerical approach separately determines the complete expansions and stores the combined rotation matrices before running the main program. As the larger program runs, the matrices are interpolated numerically to provide highly accurate results without the large expense of doing all the individual calculations during runtime. As always, we must check the software documentation to ensure compatibility.

Algorithm 4, below, supports *FK5* reductions to and from the J2000 epoch.

ALGORITHM 4: J2000 Reduction

$$(\dot{\vec{r}}_{J2000}, \dot{\vec{v}}_{J2000}, z, \Theta, \zeta, \epsilon, \Delta\Psi, \bar{\epsilon}, \theta_{AST}, x_p, y_p, \dot{\vec{r}}_{ECEFwPM}, \dot{\vec{v}}_{ECEFwPM})$$

Find precession and nutation angles $z, \Theta, \zeta, \epsilon, \Delta\Psi, \bar{\epsilon}, \theta_{AST}$

$$[\text{PREC}] = \text{ROT3}(-z) \text{ROT2}(\Theta) \text{ROT3}(-\zeta)$$

$$[\text{NUT}] = \text{ROT1}(-\epsilon) \text{ROT3}(-\Delta\Psi) \text{ROT1}(\bar{\epsilon})$$

$$[\text{ST}] = \text{ROT3}(\theta_{AST})$$

$$[\text{PM}] = \begin{bmatrix} 1 & 0 & x_p \\ 0 & 1 & -y_p \\ -x_p & y_p & 1 \end{bmatrix}$$

$$\dot{\vec{r}}_{J2000} = [\text{PREC}]^T [\text{NUT}]^T [\text{ST}]^T [\text{PM}]^T \dot{\vec{r}}_{ECEFwPM}$$

$$\dot{\vec{v}}_{J2000} = [\text{PREC}]^T [\text{NUT}]^T [\text{ST}]^T \left\{ [\text{PM}]^T \dot{\vec{v}}_{ECEFwPM} + \vec{\omega}_{\oplus} \times \dot{\vec{r}}_{ECEFwPM} \right\}$$

$$\begin{cases} \vec{r}_{ECEFwPM} = [\text{PM}] [\text{ST}] [\text{NUT}] [\text{PREC}] \vec{r}_{J2000} \\ \vec{v}_{ECEFwPM} = [\text{PM}] \left\{ [\text{ST}] [\text{NUT}] [\text{PREC}] \vec{v}_{J2000} - \vec{\omega}_{\oplus} \times \vec{r}_{ECEFwPM} \right\} \end{cases}$$

An example is useful to tie all the concepts together.

▼ Example 1-6. Performing a J2000 Reduction.

GIVEN: $\vec{r}_{J2000} = 5102.5096 \hat{i} + 6123.01152 \hat{j} + 6378.1363 \hat{k}$ km

$$\vec{v}_{J2000} = -4.743\,219\,6 \hat{i} + 0.790\,536\,6 \hat{j} + 5.533\,756\,19 \hat{k}$$
 km/s

FIND: $\vec{r}_{ECEFwPM}, \vec{v}_{ECEFwPM}$ on April 6, 1991, 07:51:28.788 53 UT1

Let's use metric units for clarity in this example. Calculate the time quantities before starting the reductions. Get ΔAT (26.0 s) directly from the *Astronomical Almanac* (1994, K9). From the USNO's IERS final data, $\Delta UT1 = 0.402\,33$ s.

$$UTC = UT1 - \Delta UT1 = 07:51:28.788\,53 - 0.402\,33 = 07:51:28.3862$$

$$TAI = UTC + \Delta AT = 07:51:28.3862 + 26\text{ s} = 07:51:54.3862$$

$$TDT = TAI + 32.184\text{ s} = 07:51:54.3862 + 32.184\text{ s} = 07:52:26.5702$$

$$DMY\ HMS \Rightarrow JD_{TDT} = 2,448,352.828\,085\,303\,2 \text{ (Be sure to keep all digits for accuracy.)}$$

$$T_{TDT} = \frac{JD_{TDT} - 2,451,545.0}{36,525} = \frac{2,448,352.828\,085\,303\,2 - 2,451,545.0}{36,525} = -0.873\,969\,04$$

$$M_{\oplus} \equiv 357.527\,723\,3^\circ + 35,999.050\,34^\circ T_{TDT} = -268.677\,819\,5^\circ$$

$$TDB \cong TDT + 0.001\,658 \sin(M_{\oplus}) + 0.000\,013\,85 \sin(2M_{\oplus}) = 07:52:26.571\,856\,9$$

$$JD_{TDB} = 2,448,352.828\,085\,322\,4$$

$$T_{TDB} = \frac{JD_{TDB} - 2,451,545.0}{36,525} = -0.087\,396\,90$$

Now determine the rotation angles for precession from Eq. (1-56).

$$\xi = -0.055\,987\,22^\circ$$

$$z = -0.055\,985\,54^\circ$$

$$\Theta = -0.048\,659\,38^\circ$$

The rotations for precession produce an intermediate vector, still in a mean-equinox, mean-equator system of date:

$$\vec{r}_{ECImod} = 5119.880\,902\,16 \hat{i} + 6113.022\,741\,15 \hat{j} + 6373.795\,530\,72 \hat{k}$$
 km

$$\vec{v}_{ECImod} = -4.736\,964\,26 \hat{i} + 0.799\,800\,13 \hat{j} + 5.537\,781\,80 \hat{k}$$
 km/s

Determine the angles for the Earth's nutation ($r = 360^\circ$):

$$\begin{aligned}
 M_{\zeta} &= 134.962\,981\,4^\circ + (1325r + 198.867\,398\,1) T_{TDB} + 0.008\,697\,2 T_{TDB}^2 + 1.778 \times 10^{-5} T_{TDB}^3 \\
 M_{\odot} &= 357.527\,723\,3^\circ + (99r + 359.050\,34) T_{TDB} - 0.000\,160\,28 T_{TDB}^2 - 3.33 \times 10^{-6} T_{TDB}^3 \\
 u_{M\zeta} &= 93.271\,910\,3^\circ + (1342r + 82.017\,538\,1) T_{TDB} - 0.003\,682\,5 T_{TDB}^2 + 3.06 \times 10^{-6} T_{TDB}^3 \\
 D_{\odot} &= 297.850\,363\,1^\circ + (1236r + 307.111\,480) T_{TDB} - 0.001\,914\,17 T_{TDB}^2 + 5.28 \times 10^{-6} T_{TDB}^3 \\
 \Omega_{\zeta} &= 125.044\,522\,2^\circ - (5r + 134.136\,260\,8) T_{TDB} + 0.002\,070\,8 T_{TDB}^2 + 2.22 \times 10^{-6} T_{TDB}^3 \\
 M_{\zeta} &= -170.740\,503\,83^\circ \\
 M_{\odot} &= -268.677\,820\,7^\circ \\
 u_{M\zeta} &= -17.088\,405\,322^\circ \\
 D_{\odot} &= -97.116\,600\,08^\circ \\
 \Omega_{\zeta} &= 294.082\,058\,9^\circ
 \end{aligned}$$

Now find the obliquity of the ecliptic and the corrections for a true system, using the 106 coefficients from the 1980 IAU theory [Eq. (1-58) to Eq. (1-61)]:

$$\bar{\epsilon} = 23.440\,427\,525^\circ$$

$$\Delta\Psi = 0.004\,185\,849^\circ$$

$$\Delta\epsilon = 0.001\,170\,66^\circ$$

$$\epsilon = \bar{\epsilon} + \Delta\epsilon = 23.441\,598\,18^\circ$$

Check that the true and mean obliquities are very close. If not, recheck your calculations! Then, do the required rotations for nutation and find the vectors for true equinox and true equator of date:

$$\vec{r}_{ECItod} = 5119.285\,913\,05 \hat{i} + 6113.235\,662\,37 \hat{j} + 6374.069\,220\,71 \hat{k} \text{ km}$$

$$\vec{v}_{ECItod} = -4.737\,178\,79 \hat{i} + 0.799\,369\,47 \hat{j} + 5.537\,766\,046 \hat{k} \text{ km/s}$$

Determine the *GST* using the *JD_{UT1}*, Eq. (1-44), and Eq. (1-45):

$$GST = 312.297\,677\,780\,19^\circ.$$

Find the *AST* using

$$\theta_{AST} = \theta_{GST} + \Delta\Psi \cos(\epsilon) = 312.301\,518\,154\,51^\circ$$

After doing the rotation for sidereal time, you'll get these initial *ECEF* vectors:

$$\vec{r}_{ECEFW/OPM} = -1076.108\,623\,08 \hat{i} + 7900.679\,005\,91 \hat{j} + 6374.069\,220\,71 \hat{k} \text{ km}$$

$$\vec{v}_{ECEFW/OPM} = -3.203\,369\,83 \hat{i} - 2.887\,216\,76 \hat{j} + 0.616\,715\,09 \hat{k} \text{ km/s}$$

Finally, determine polar motion. Find the x_p and y_p coefficients from Table 1-7 as $x_p = -0.217''$ and $y_p = 0.281''$. Actual daily values from USNO's IERS final data are $x_p = -0.2206''$ and $y_p = 0.3014''$. The approximate values in Table 1-7 are often adequate. Convert to radians and rotate for polar motion to find the final Earth-centered, Earth-fixed coordinates as

$$\vec{r}_{ECEFWPM} = -1076.115\,440\,15 \hat{i} + 7900.669\,691\,94 \hat{j} + 6374.079\,614\,51 \hat{k} \text{ km}$$

$$\hat{v}_{ECEF_{WPM}} = -3.203\ 325\ 81 \hat{i} -2.887\ 267\ 24 \hat{j} + 0.616\ 707\ 44 \hat{k} \text{ km/s}$$

The length of this example and the complexity reveal the importance of knowing three things: (1) how to do the operations, (2) when you don't need them, and (3) how to program a computer! Now examine the results of the sample problem. Table 1-8 shows the position and velocity vectors through each of the reduction rotations.

TABLE 1-8. Summary of J2000 Reduction Values for Example 1-6. This table shows the position and velocity vectors after each transformation. The main difference occurs with the correction for sidereal time—the transition between Earth-fixed and true coordinates. The terminology for mean and true values is not always followed. Mean equinox values are sometimes substituted for the true values listed.

Position (km)	\hat{i}	\hat{j}	\hat{k}	Terminology
	5102.509 600 00	6123.011 520 00	6378.136 300 00	\hat{r}_{J2000} , mean equator and mean equinox of J2000
Precession 20,503*	5119.880 902 16	6113.022 741 15	6373.795 530 72	\hat{r}_{ECImod} , mean equator and mean equinox of date
Nutation 689*	5119.285 913 05	6113.235 662 37	6374.069 220 71	\hat{r}_{ECItod} , true equator and true equinox of date
Sidereal time 6,448,090*	-1076.108 623 08	7900.679 005 91	6374.069 220 71	$\hat{r}_{ECEF_{w/oPM}}$, ECEF
Polar motion 16*	-1076.115 440 15	7900.669 691 94	6374.079 614 51	$\hat{r}_{ECEF_{WPM}}$, ECEF
Velocity (km/s)				
	-4.743 219 60	0.790 536 60	5.533 756 19	\hat{v}_{J2000}
Precession	-4.736 964 26	0.799 800 13	5.537 781 80	\hat{v}_{ECImod}
Nutation	-4.737 178 79	0.799 369 47	5.537 660 46	\hat{v}_{ECItod}
Sidereal time	-3.203 325 15	-2.887 266 33	0.616 715 09	$\hat{v}_{ECEF_{w/oPM}}$
Polar motion	-3.203 325 81	-2.887 267 24	0.616 707 44	$\hat{v}_{ECEF_{WPM}}$

*RMS magnitudes from the previous vector are listed in meters. This shows the relative effect of each change.

1.7.6 Other Corrections

We can apply many corrections to the motion and use of coordinate systems. Many of the corrections are mainly for observations of stars, but some may be of interest for Earth-orbiting satellites. Consult Seidelmann (1992, 21–24, 121–139, and 721, 730) for more information.

Aberration

Aberration is the angular displacement of an apparent position due to the motion of the object and the observer. Because the speed of light is finite, we must account for the travel time when we measure the direction of an object. We call this correction for *light time*, because the path is no longer an ideal straight line between the points. For stars, we ignore light-time correction. Accounting for the observer's motion, called *stellar aberration*, consists of three types: *diurnal*—about the Earth's center, *annual*—about the barycenter of the solar system, and *secular*—due to the motion of the solar system. *Planetary aberration* includes correction for both light time and stellar aberration.

Parallax

Parallax—a topic usually referring to the stars—is also an important concept in the study of Earth satellites. As mentioned earlier, we define *parallax*, ϑ , as the angular difference in the apparent direction of an object when viewed from different locations. There are three main types of parallax. *Geocentric parallax*, ϑ_{geo} , also called *diurnal parallax*, is the angular difference between observations from the center of the Earth and those from a site (topocentric observation). When the object lies on the site's horizontal plane, we use the term *horizontal parallax*, ϑ_{horiz} . *Heliocentric parallax*, ϑ_{helio} , also called *annual parallax*, is the angular difference between observations taken from the centers of the Sun and the Earth. For instance, we find ϑ_{helio} for a star by calculating the difference between the angle measured from the Earth to the star versus the angle measured from the Sun to the star. We can find this angle with the following general formula (see Fig. 1-36):

$$\sin(\vartheta) \approx \frac{a}{b}$$

Because the magnitude of the distance to the object, r , is usually much larger than the semi-diameter, a ,

$$\sin(\vartheta) \approx \frac{a}{r} \quad (1-69)$$

where the semi-diameter, a , is at the viewing location, and r is the distance to the star being observed. The equation is actually an approximation because the sine requires the slant range to the star, but because the distance is much larger than the semi-diameter, the approximation is very accurate. Note that we can't take observations from the Earth's center or the Sun. However, the difference between observations of a star taken 12 hours apart gives us twice the geocentric parallax. Similarly, observations taken six months apart give us twice the annual parallax.

Parallax isn't a new concept. Hipparchus is credited with using the geocentric parallax for the Moon in about 129 B.C. to determine the Moon's approximate distance from the Earth (Pannekoek, 1989, 129). During an eclipse year, Hipparchus noticed that the solar eclipse at Hellespont was total, while at Alexandria (only 800 km away), the eclipse was

only about 80%. This information enabled Hipparchus to estimate the distance to the Moon to be 60 to 74 Earth radii. Today we know the distance to be about 60 Earth radii. For the geocentric parallax of the Sun, the semimajor axis is that of the Earth (6378 km), and the distance is the distance from the Earth to the Sun (149,597,870 km), with a resulting parallax of about 0.002 44°. An example will show how parallax works.

▼ Example 1-7. Determining Parallax.

GIVEN: Neptune on May 14, 1994, $r = 29.664\ 361$ AU (1 AU = 149,597,870 km)

FIND: ϖ_{helio} , and ϖ_{geo}

For Neptune's orbit, the radius is 29.664 361 AU, and the semimajor axis of the Earth's orbit is 1 AU. Because the orbit planes of Neptune and the Earth nearly coincide, we'll assume they are identical. For the heliocentric parallax,

$$\sin(\varpi_{helio}) \equiv \frac{a}{r} = \frac{1}{29.664\ 361}$$

$$\varpi_{helio} = 1.931\ 83^\circ = 6954.604''.$$

This angle doesn't seem very small, and it isn't, because Neptune is relatively "close" to the Earth compared to the distance of the Earth from the Sun.

Consider the geocentric parallax measured at the Earth's center and at its surface. Here,

$$\sin(\varpi_{geo}) \equiv \frac{a}{r} = \frac{6378}{149,597,870\ (29.664\ 361)}$$

$$\varpi_{geo} = 0.000\ 082\ 346^\circ = 0.296\ 446''.$$

▲ This value is much smaller due to the larger distance to the planet compared to the radius of the Earth.

Example uses of parallax in astrodynamics studies appear in two main areas. First, in calculating the position of the Moon and planets, parallax is often an argument in the solution. Second, parallax is useful in determining angular accuracies for sensor systems.

Refraction

The atmosphere bends light rays (and other electromagnetic radiation) as they pass through it. We're quite familiar with this concept; it's the explanation for distortion of an underwater object viewed from above. The deflection causes distortion in observations, which can be appreciable, and we can correct it using various models. Two main regions in the atmosphere cause refraction: the troposphere and the ionosphere. Tropospheric refraction occurs due to local temperature, humidity, and pressure conditions over a site. Ionospheric refraction is caused by changes in electromagnetic waves as they pass through the electron plasma of the ionosphere. We can correct both effects by comparing two separate wavelengths. Unfortunately, we often have only one measurement frequency. Consult Seidelmann (1992, 24, 140–145) for more information.

Deflection of Light

We know from relativity theory that light bends as it travels near massive objects. This is a factor in position measurements involving distant galaxies and quasars because their light

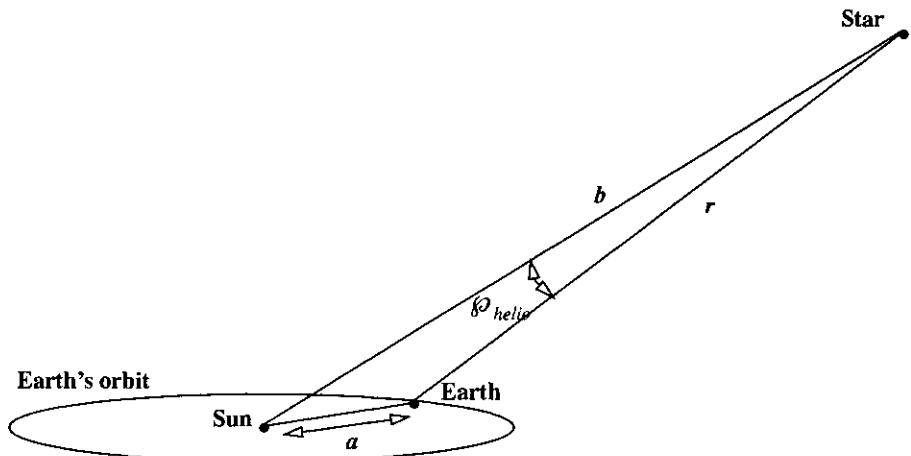


Figure 1-36. Parallax Geometry. Notice the greatly exaggerated scale to show the parallax. Depending on the accuracy, even from different points in the Earth's orbit, stars can appear to be at virtually the same location. The heliocentric parallax angle shown here defines the apparent difference for observations from the Sun and Earth.

often passes near galaxies and objects which deflect it. Remember that we use these types of observations to realize the FK5 inertial frame. For satellite applications, deflection usually isn't important.

Proper Motion and Radial Velocity

Proper motion is defined as the motion of an object with respect to an observer. Although stars are distant from the Earth, they usually have some relative motion with respect to it. The velocity in the radial direction is called **radial velocity**; it's determined from Doppler measurements. The proper motion values are typically determined from observations spanning long periods of time—usually decades. Values are given in terms of rates (seconds of arc) per year or per century. We can combine them with parallax to accurately transform standard reference frames. Proper motion slowly degrades our position estimates for individual stars, so we must periodically update their positions. More distant stellar objects, such as radio galaxies, have smaller values of proper motion, making them better references for realizing inertial frames. Unfortunately, their extreme distance and extended structure make observing them a challenge.

1.7.7 B1950 Reduction

I mention reduction using the B1950 system because many programs still use this former system even though it's no longer considered precise. It relies on the **Besselian year**, defined as the time for the ephemeris mean Sun's right ascension to increase by 24 hours

(Green, 1988, 249). Unlike the complex determinations of precession and nutation angles needed for the J2000 transformation, the conversion with B1950 uses a matrix operation. Seidelmann (1992, 313) shows the process for any position or velocity vectors, \vec{x} .

$$\vec{x}_{J2000} \equiv \begin{bmatrix} 0.999\,925\,679\,495\,687\,7 & -0.011\,181\,483\,220\,466\,2 & -0.004\,859\,003\,815\,359\,2 \\ 0.011\,181\,483\,239\,171\,7 & 0.999\,937\,484\,893\,313\,5 & -0.000\,027\,162\,594\,714\,2 \\ 0.004\,859\,003\,772\,314\,3 & -0.000\,027\,170\,293\,744\,0 & 0.999\,988\,194\,602\,374\,2 \end{bmatrix} \vec{x}_{B1950}$$

The reverse transformation is

$$\vec{x}_{B1950} \equiv \begin{bmatrix} 0.999\,925\,679\,495\,687\,7 & 0.011\,181\,483\,239\,171\,7 & 0.004\,859\,003\,772\,314\,3 \\ -0.011\,181\,483\,220\,466\,2 & 0.999\,937\,484\,893\,313\,5 & -0.000\,027\,170\,293\,744\,0 \\ -0.004\,859\,003\,815\,359\,2 & -0.000\,027\,162\,594\,714\,2 & 0.999\,988\,194\,602\,374\,2 \end{bmatrix} \vec{x}_{J2000}$$

Be aware that converting between the B1950 and J2000 systems introduces two important errors—UT1 has an updated secular rate in J2000, and the origins of the systems don't coincide.

1.8 Earth Models and Constants

Engineering demands that we correctly use constants and conversions within a certain model. In addition, as satellite technology and experiments become more sophisticated, we learn more about the physical properties of the Earth and planets. With ever-changing constants, we must understand what each value is, what the reference frame is, and which of these values a particular program or algorithm requires.

Perhaps the best information on various reference frames comes from the publications of the International Earth Rotation Service (McCarthy, 1992). McCarthy distinguishes between *celestial* and *terrestrial* reference frames. *Celestial reference frames* are barycentric frames determined from VLBI measurements, whereas *terrestrial frames* are bound to the Earth. We'll mainly use terrestrial frames for the satellite motions in this book. The suggested Earth reference frame is the *International Terrestrial Reference Frame* (ITRF-92) model described by McCarthy (1992). This frame is a particular definition of the J2000 system. As mentioned earlier, many datums further define terrestrial reference frames. McCarthy (1992, Chap. 5) shows ways to convert between terrestrial reference frames (e.g. ITRF-90) and datums (e.g. WGS-84).

We've already discussed coordinate systems and datums, so we'll concentrate on the parameters that describe the physical shape of the Earth. Constants are sometimes designed with a new reference frame but are often designed to be compatible with an existing frame. One source of physical constants is the *World Geodetic Survey* (WGS), which tries to define many of the Earth's physical parameters, mostly for the military. It has several different dates of release through the Defense Mapping Agency (1960, 1966, 1972, 1984). Each revision is referenced as WGS-##, with the appropriate year inserted. Although the newer revisions yield more accurate results, they sometimes include changes

in the reference frame, so be careful when using these values. The Global Positioning System in Chap. 10 uses WGS-84.

Although the WGS values are usually acceptable, we have more accurate models, such as the *Joint Gravity Model* (JGM-2 and JGM-3) developed through collaboration between NASA/Goddard and the University of Texas at Austin (Nerem et al., 1994). The JGM models actually update NASA's Goddard Earth Model (GEM-T3) and have improvements using International Earth Rotation Service (IERS) constants, analysis of long-wavelength data, and TOPEX altimeter data. Because of the JGM-2 system's high accuracy, this book will use its data. See Nerem et al. (1995), McCarthy (1992), and Vetter et al., (1994) for discussions of the joint models. A joint model combining the Defense Mapping Agency's efforts and joint gravity models is due in 1997.

Unfortunately, although we have various models for the Earth, some programming practices complicate attempts to modernize many older systems with newer or different models. For the average amateur astronomer, the differences between these data sets are insignificant; however, for precise satellite calculations, we must have the most accurate constants. In addition, high-accuracy programs *must be consistent* among themselves. Perhaps the most common error is using conversion factors derived from different systems or sources, which can cause unexpected and erroneous results.

1.8.1 Canonical Units

Discussing the Gaussian constant of gravitation is particularly important because it introduces the concept of *canonical units*, which scale the units of a problem with respect to a particular coordinate system. Even Gauss recognized he couldn't measure all the quantities needed and decided to fix certain values to unity. Indeed, astrodynamical calculations may become unwieldy. Imagine using a position vector for the Moon in feet or the mass of the Earth in kilograms! As computers have become more powerful, canonical units may appear to be less important, but we can keep the code much simpler by using a standardized set of values in each algorithm. We can work faster by eliminating useless conversions between routines and by not changing validated routines to incorporate new constants. The overall effect is programs that are faster, more accurate, and easier to maintain. This book uses systems of canonical units because they

1. Reduce the size of the numbers involved in calculations, so we can verify them more easily in our minds. Of course, this ease depends on what we learn and are familiar with.
2. Can make certain operations more mathematically stable.
3. Often speed an algorithm by eliminating unnecessary mathematical operations because unity values replace many large constants.
4. Allow different organizations to use standard values.
5. Reduce required maintenance programming to incorporate future updates to constants and therefore save a lot of time (and money). They often per-

mit changes to constants used in conversion routines while using the same code for technical algorithms.

The added consistency of standardization permits great savings between organizations. The continual debate on using English or metric units illustrates this application. Canonical units let us use English or metric data. They are a way to solve orbit problems in *any* organization with *whatever specific units are desired* because the conversions are independent from the mathematical algorithms.

Most computer-software code contains unit conversions within each routine of the program, and the constants are often embedded and combined with other values. To compound the problem, the constants may be from very old sources, are usually undocumented, and are often inconsistent with newer routines from the same program. Updating these constants embedded in almost all engineering code is a continual drain of time and money. As new data becomes available, it's often impossible to update the values because they're undocumented or combined with other constants. The high cost of reworking many of these programs is prohibitive, so we often simply *tolerate* the inefficiency and inaccuracy.

The Earth's physical parameters (Sec. 1.3) provide a way to develop canonical units for the planet. These constants will allow us to normalize units and simplify future changes. Unlike the fixed value of κ_{\odot} , attempts to standardize a similar constant for the Earth, κ_{\oplus} , have largely been ignored because computers can quickly recalculate ephemerides as new constants are determined. Let's define one ***Earth radius***, ER, unit to be the Earth's equatorial radius, R_{\oplus} :

$$\text{ER} \equiv R_{\oplus} \quad (1-70)$$

The *gravitational parameter*, μ_{\oplus} , is also defined as unity, so

$$\frac{\text{ER}^3}{\text{TU}^2} \equiv \mu_{\oplus} \quad (1-71)$$

These definitions provide two of the three elements necessary for a complete set of canonical constants to convert whatever data we have for a particular problem. The remaining task is to relate the *time unit*, TU, to known quantities. [The Navy calls this quantity a *Herg*, He, in honor of Paul Herget (1908–1981)]. We can do so by recognizing that one ***time unit*** is the time it takes a hypothetical satellite to travel one radian on its way around an orbit skimming the surface of a spherical Earth of radius $r = R_{\oplus}$ or by using the definitions above. Solving Eq. (1-71) for TU and using constant values, we find that

$$\text{TU} \equiv \sqrt{\frac{R_{\oplus}^3}{\mu_{\oplus}}} \quad (1-72)$$

If we take the Earth-value units on the right-hand side as km^3 and km^3/s^2 for the radius cubed and gravitational parameter, respectively, we can define one TU in seconds. We can immediately see the advantage of this approach. To use a different or newer version of the

Earth's physical parameters, we simply recalculate the above conversions and adjust the conversion constants of the input and output routines. Routines that do the actual mathematical operations *remain unchanged!* The table on the inside front cover lists the defined and derived constants of JGM-2. Again, be aware of the need for consistency between values in this table. Don't mix conversions from JGM-2, WGS-72, WGS-84, etc. when doing highly accurate studies. We can derive each value in the table from the other values.

Problems

1. What is the Julian date on May 5, 2004 at 14:26 local time at the Naval Post-graduate School ($\lambda = -121.53^\circ$)? What is *UTC* and *TAI*? What other information would you need?
2. The Astrodynamics division at Phillips Laboratory, located at $\phi_{gc} = 35.05^\circ$, $\lambda = -106.40^\circ$, $h_{ellp} = 1866.03$ m, is located on the vernal equinox on May 2, 1996. What are the *LHA*, *GHA*, *LST*, and *GST*?
3. What are numerical values of canonical units for Jupiter? Use Table D-5.
4. You're briefing a general about a new polar-orbiting satellite. The general asks, "How many g's does the satellite pull as it travels under the South Pole?" Wishing to remain employed, how do you answer?
5. What are the J2000 position and velocity vectors for Example 1-6 in the *PQW*, *SEZ*, *RSW*, *NTW*, and *EQW* coordinate systems for a site at $\phi_{gd} = 40^\circ\text{N}$ and $\lambda = 100^\circ\text{W}$?
6. Calculate the Mean of 1950 position and velocity vectors (B1950) for the inertial Mean of J2000 vectors given in Table 1-8. Why are the two vectors so different?
7. You're working at the site ($\lambda = 100^\circ\text{W}$) of a new, highly precise sensor and have just received some vectors (true equator of date, true equinox) that tell you where to look for the Hubble Space Telescope (circular orbit, altitude ≈ 590 km, $i = 28.5^\circ$). Your system calculates the topocentric α_t and δ_t from vectors in the J2000 system but assumes the input is true equator, mean equinox of date. How much error is there for your system if the observation takes place directly overhead at noon local time on April 26, 1994?
8. What are the values for the geoid's undulation and h_{ellp} for a site at $\phi_{gc} = 28.00^\circ$, $\lambda = 86.57^\circ$. As you add terms to 70th order, do you see any variations? Explain your answers.
9. For the position and velocity vectors we determined in Example 1-6 (*ECEF-wPM*), what are the vectors on August 20, 1994? (Hint: only one transformation is required.)
10. Eq. (1-6) relates the vertical distance of a circle and an ellipse. What are the corresponding relations for a hyperbola?

CHAPTER 2 EQUATIONS OF MOTION

- 2.1 Historical Background
- 2.2 Two-body Equation
- 2.3 Three-body and n -body Equations
- 2.4 Satellite State Representations
- 2.5 Application: Orbital Elements from r and v
- 2.6 Application: r and v from Orbital Elements

2.1 Historical Background

The long period of inactivity in the roots of astrodynamics from the end of ancient times began to change with Nicholas Copernicus (1473–1543), though the reasons for this change are open to debate. Perhaps it was the resurgence of relatively large and stable governments, which again permitted the time for scientific thought to flourish. Was it Christopher Columbus’s voyage to the New World, or Galileo’s invention of the telescope, or Johann Gutenberg’s development of the printing press in 1440 that sparked the renewal? Small ([1804] 1963, 81) has suggested that it was because the Ptolemaic tables (and also the Alphonsine tables derived from them in 1252 by Castilian King Alphonso) became unreliable and that the instruments of the day were beginning to detect the errors produced by the older system. Whatever the case, the foundation of astrodynamics was about to undergo perhaps the most significant change in history.

It may be difficult for us to imagine today that the basic notions of the universe could be in error; yet, that’s exactly what Copernicus attempted to prove. Copernicus was the first scientist to bridge the gap between antiquity and modern times. Besides being deeply interested in mathematics, Copernicus earned a degree in medicine during a tour in Italy in 1501. He worked more than 31 years to resolve the fundamental motions of the solar system (Beer and Strand, 1975a, 138). The completion of this work is often cited as the beginning of a scientific, as well as a cultural, revolution. In his *De Revolutionibus Orbium Coelestium* [*On the Revolutions of the Celestial Sphere* (1543)] he proposed a Sun-centered system. Figure 2-1 shows the usual depiction.

Notice the continued reliance on circular motion. Indeed, some texts give the impression that all the orbits were circular. This wasn’t quite true, as we’ll see shortly. Copernican theory differed from Ptolemaic theory in three main ways: heliocentric theory, new numbers and data, and details of planetary motion. Pannekoek gives an excellent account of Copernicus’s arguments, mostly in his own words.

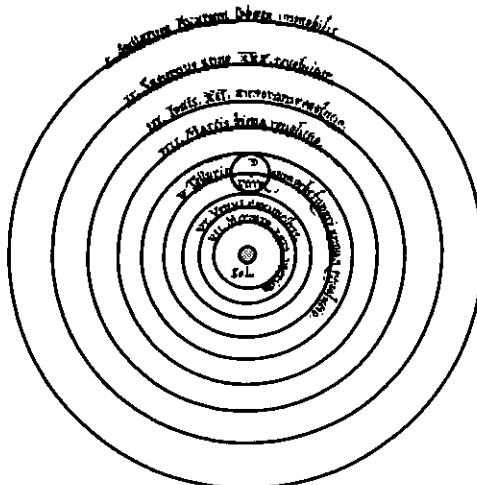


Figure 2-1. Copernicus's Solar System. (Pannekoek, 1989, 191) In addition to defining a Sun-centered system, Copernicus also introduced smaller epicycles than the previous Ptolemaic system.

If a motion of the Earth is assumed, this must appear, although in the opposite direction, in all that is outside, as if things past her; and this holds especially for the daily motion. Since heaven contains all, it is not conceivable that motion should not be attributed to what is contained therein rather than to what contains all. If then someone should deny that the Earth occupies the centre of the world but admits to her a distance not large enough to be measured against the spheres of the fixed stars, but comparable to the orbits of the Sun and the planets, he could perhaps indicate the cause for the irregularities in the different motions as due to another centre than the Earth's. (Pannekoek, 1989, 190)

Copernicus's arguments continue at length to illustrate each of his three main divergences from Ptolemaic theory. The mode of argument in the 16th century was much more philosophical than modern arguments but just as valid. Copernicus did have difficulty in removing the equant from Ptolemaic theory: his theory includes small epicycles and slight modifications to pure circular motion about a central point. He needed them because the *correct* orbital theory didn't exist. The Sun-centered system alleviated some of the difficulties but didn't solve the problem.

Copernicus knew his theories were controversial, so he resisted publication of this work until he was on his deathbed. Were it not for a well-intentioned student of Copernicus named Georg Rheticus (1514–1576), the world may have missed the great discoveries for years. Rheticus started the publication process, and after he left to teach in Leipzig in 1542, Andreas Osiander (1498–1552, a Lutheran theologian) continued it through the printing in Nürnberg in 1543. A copy from the first printing was delivered to Copernicus on May 24, 1543, the day he died (Dreyer, 1953, 319). Beer and Strand (1975a, 205) suggest that a section discussing elliptical motion was left out of the original publication. His-

tory blames Rheticus; however, Osiander may have had a greater motive to omit it because he had strong views concerning the religious implications of Copernicus's theories. Osiander *did* change the preface to lessen the work's impact, which shows the importance of authors maintaining control over their manuscripts! In any case, if Copernicus did intend to include a discussion of elliptical motion, it could have further revolutionized the study of planetary motion. But would the world have been ready to accept this much change?

In many respects, Galileo Galilei (1564–1642) picked up where Copernicus left off. In fact, Berry (1961, 149) mentions that Galileo (he preferred his first name) adopted the Copernican ideas a few years before 1597, even though his famous works were published more than ten years later. The main advantage of his research was his use of the telescope for regular and dedicated scientific research.

Galileo is perhaps best known for his support of theories which opposed the religious doctrine of the time. The initial controversy erupted about 1610 over a strictly scientific discussion to determine whether Sun spots were on the surface of the Sun or were actually planets passing between the Sun and the Earth. Galileo had observed the spots, as had Christopher Scheiner (1575–1650). But Scheiner thought the spots were planets between the Earth and the Sun. They exchanged letters to resolve the problem—a common practice in the 1600s. Galileo apparently didn't like the results, and, being a devout Catholic, decided around 1611 to go to Rome for a "second opinion." Although he was warmly received at first, these events ultimately led to the famous call from Rome in 1616 for Galileo's formal reprimand. The Vatican also placed Galileo's writings, and even some by Copernicus, on the prohibited list. Berry (1961, 160) suggests that the reprimand may not have been as severe as is sometimes thought, but it was a serious impediment to Galileo's career.

The publication of Galileo's *Dialogue on the Two Chief Systems of the World, the Ptolemaic and Copernican* represented another marvelous feat by Galileo and another opportunity for controversy. His own findings from the telescope inclined his results to the Copernican system, which he defended in the work. He again had to go through Rome to obtain the necessary permission for printing, which finally occurred in 1632. It's interesting that Galileo wrote in Italian instead of the scholarly Latin. Perhaps his use of a common language was a factor in Galileo's troubles. If he had written in Latin, the clergy may have passed off his ideas because they would have remained uninterpreted for most of the common people. However, texts written in Italian posed a dual challenge: Galileo and the common people would begin to question the Church's fundamental precepts. Galileo served a valuable role in continuing the new thought which was about to take solid shape under Tycho Brahe, Johann Kepler, and Isaac Newton.

2.1.1 Kepler's Laws

Scientific change accelerated when Johann Kepler (1571–1630) and Tycho Brahe (1546–1601) combined forces. Tycho Brahe was born into an aristocratic, reasonably wealthy family. His gifts were extreme diligence and persistence. In contrast, Johann Kepler was

born two months prematurely, spent much of his life in poor health, and suffered from poor eyesight as a result of smallpox when he was four years old. But Kepler was also diligent. A strongly devout man, Kepler actually aspired to become a theologian; however, in 1594 he chose to accept a teaching position at Graz. Part of Kepler's duties in this position was to compile annual almanacs.

Several years and many positions later, Kepler became the imperial court mathematician for Emperor Rudolf II in Prague in 1601. Tycho Brahe had died shortly before, leaving Kepler with all his very precise observational data. This was perfect for Kepler. Unfortunately, producing horoscopes for the Emperor wasn't thrilling work, but it provided income and allowed him to pursue technical interests on the side.

The relatively "large" eccentricity of the Martian orbit attracted Kepler's interest. After many years of work, he published *Astronomia Nova* (*New Astronomy*) in 1609. This was a huge work containing his first two laws. It's worth pointing out that Kepler completed the manuscript in 1605 but couldn't print it for four more years.

Kepler's theological aspirations were ultimately responsible for his third law because he believed resolving planetary motion was a divine calling. Stephenson comments on *Astronomia Nova* that "Kepler closed the chapter, one of the most beautiful in the entire book, with the hope that it would please God to allow the human race sufficient time on this world to resolve these matters" (1987, 137). Indeed, Kepler completed several theological writings in 1614 and 1615. These works paved the way for *De Cometis* (On Comets) in 1618.

Finally, Kepler realized his life-long goal to describe physical motion by publishing in 1619 his *Harmonices Mundi Libri V* (*Harmony of the World*). But Kepler buried his third law with 13 other theorems, which he included only as background for the greater goal—describing the harmonic motion of the planets. In the preface, Kepler remarks, "So I throw the dice and write a book for my contemporaries or for posterity. I do not care. It may take a hundred years for my book to find readers, it even took God 6000 years to find an observer." (Beer and Strand, 1975b, 524)

Kepler's third law now receives particular attention in the literature, but all three are important:

1. The orbit of each planet is an ellipse with the Sun at one focus.
2. The line joining the planet to the Sun sweeps out equal areas in equal times.
3. The square of the period of a planet is proportional to the cube of its mean distance to the Sun.

Beer and Strand (1975b, 526) summarize Kepler's own view of what he had accomplished by quoting the concluding remarks from *Harmonices Mundi*: "See I have now completed the work to which I was called. In it I have used all the Talents Thou hast lent to my spirit. I have revealed the majesty of Thy works to those men who will read my words, insofar as my narrow understanding can comprehend their infinite richness."

Kepler's modest view of his efforts differs greatly from our perception today. We'll examine more details of these laws when we discuss Kepler's equation and Kepler's prob-

lem in Chap. 4. These laws mark the beginning of a remarkable period in the history of mathematics and lead directly to the accomplishments of Isaac Newton.

2.1.2 Newton's Laws

As remarkable as Kepler's laws were, they didn't completely solve planetary motion. They captured the kinematics of motion, but the dynamics of motion remained unsolved until Isaac Newton (1642–1727) unlocked them.

In 1665, Newton was a student at the University of Cambridge when a devastating outbreak of bubonic plague forced the university to shut down until the spring of 1667. The unintended 18-month vacation at his home in Woolsthorpe was very productive for Newton. With little to do, he busied himself with interesting problems of the day. During this period, he observed the famous falling apple and pondered what force could cause this motion. Extending his thoughts to the Moon, he wondered how strong and how far such a force would extend. He knew an approximate value (for gravity) from Galileo's famous Leaning Tower of Pisa experiments, but his rough calculations for the Moon were about 15% off. Frustrated, he abandoned the problem and went on to develop the binomial theorem, early calculus, and the foundations for the study and composition of white light. When Cambridge reopened after the plague had subsided, Newton finished his Ph.D.; yet, it contained *none* of these results! His instructor at the University was so impressed with Newton's performance that he resigned his post and allowed Newton to take his teaching position at Cambridge in 1667.

This could have been the end of the story were it not for Edmond Halley (1656–1742), the discoverer of Halley's Comet, and other members of the Royal Society of London. One day in January, 1684, Halley, Sir Christopher Wren (1632–1723), and Robert Hooke (1635–1703) were discussing current theories for planetary motion. These theories were combinations of proposals by Huygens and Descartes; they envisioned combinations of eddies and pools in which the planets orbited. When Hooke boldly exclaimed he had demonstrated the laws of celestial motion, Wren was so interested he challenged either Halley or Hooke to produce convincing results within two months. In addition, Wren offered to pay 40 shillings to the winner. Halley wrote Newton about the offer and asked for advice but never received a response.

In August, 1684, Halley was visiting Newton in Cambridge and he casually posed the question of what law of attraction would account for the elliptical motion of the planets. Newton surprised Halley by immediately responding with the answer of an inverse square law. He added that he had already calculated the results but couldn't find the papers. Halley persuaded Newton to reaccomplish the results. On December 10, 1684, Newton delivered a short answer, *De Motu*, to the Royal Society.

Halley was ultimately responsible for convincing Newton to finish his work on planetary motion and publish the results. The process took almost two years. Newton's health suffered because he worked so diligently on the problem that he often neglected sleep and food. The work was complete in 1686, and Halley himself paid for the printing of the manuscript in 1687 as *Philosophiae Naturalis Principia Mathematica* (*Mathematical*

Principles of Natural Philosophy, or simply *Principia*). In Book I of his *Principia*, Newton introduced his three laws of motion:

1. Every body continues in its state of rest, or of uniform motion in a right [straight] line, unless it is compelled to change that state by forces impressed upon it.
2. The change of motion is proportional to the motive force impressed and is made in the direction of the right line in which that force is impressed.
3. To every action there is always opposed an equal reaction: or, the mutual actions of two bodies upon each other are always equal and directed to contrary parts. (Newton [1687] 1962, 13)

We should reflect on the time in which these observations were made. Newton continued his explanation for the third law as follows:

If a horse draws a stone tied to a rope, the horse (if I may say so) will be equally drawn back towards the stone; for the distended rope, by the same endeavor to relax or unbend itself, will draw the horse as much towards the stone as it does the stone towards the horse, and will obstruct the progress of the one as much as it advances that of the other. (Newton [1687] 1962, 14)

We often forget that, although Newton achieved great technical success, he lived in a world which was much simpler than our modern society. Yet this simplistic approach is valuable even today.

2.1.3 Other Early Astrodynamical Contributions

Although this chapter deals mainly with the two-body problem, the multi-body problem has great importance in celestial mechanics, as well as a rich history. Joseph Louis Lagrange (1736–1813) showed the first solutions for the three-body problem in *Essai sur le problème des trois corps* in 1772. In the same year, Leonhard Euler (1707–1783) arrived at a similar conclusion for equilibrium points. Carl Gustav Jacob Jacobi (1804–1851) found an exact integral in 1836 (*Comptes rendus de l'Académie des Sciences de Paris*), and Henri Poincaré (1854–1912) proved this was the *only* exact integral for the three-body problem in 1899. This work even won a prize from the King of Sweden. The technique was only of mathematical interest until 1906 when the Trojan planets were discovered near Jupiter. These small planetoids, named after heroes in the Trojan war, remain near equilibrium positions described by the restricted three-body problem. History has favored Lagrange's contributions, so the equilibrium positions bear his name.

The achievements of the people discussed in this section may seem small, but they're actually momentous in the history of modern astrodynamics. Together with many other scientists of the time, these pioneers broke a 1500-year logjam and created the basic tools needed to advance astrodynamical thought and study. The mathematical techniques, instruments, and means of distribution had caught up to the scientific thought of the time. Indeed, even with the advent of the computer, much of our work in astrodynamics uses the results these men formed long before the modern era!

2.2 Two-body Equation

This chapter explores the two-body equation of motion and initial analyses. Complete knowledge of the two-body problem and its assumptions is crucial to astrodynamics. These techniques are extremely useful for back-of-the-envelope calculations and often serve as the starting point for more complex study.

Newton's second law and his universal law of gravitation are the starting points for virtually any study of orbital motion, especially when combined with Kepler's Laws. His first law states that bodies tend to stay at *rest* or in *uniform motion* unless they are acted upon by an external force. Although this idea is easily dismissed in modern times as trivial, it was revolutionary in 1687. Previously, concepts of motion were based on Aristotelian philosophy. Most scientists believed that an object's natural state was at rest, because they knew that friction causes objects to slow down and eventually stop. The concept of bodies staying in *motion* was new.

Most people are familiar with the second law, which states that the time rate of change of linear momentum is proportional to the force applied. For a fixed-mass system,

$$\sum \vec{F} = \frac{d(m\vec{v})}{dt} = m\vec{a} \quad (2-1)$$

This simply means we assume the mass is constant and the sum of all the forces, F , acting on a body is equal to the mass, m , times the acceleration, a , of that body. Although the equation is written as a vector relation, it also represents three scalar magnitudes ($F_i = ma_i$, $i = 1$ to 3) for an individual force. Newton's law of gravitation [Eq. (1-1)] provides the means to find the components of the force (and therefore the acceleration through his second law) if only gravity affects the body. Figure 2-2 shows the geometry for a fixed-mass system, in which we consider only the gravity from the central body (mass m). The coordinate system is at the center of mass of the entire system.

For the derivation of the two-body equation of motion, we'll assume an ideal inertial system that's fixed in inertial space or has a fixed orientation with an origin moving at constant velocity. Consider a system of only two bodies consisting of the Earth, m_{\oplus} , and a satellite, m_{sat} . Figure 2-3 shows this arrangement.

The coordinate system ABC is defined as an ideal inertial system displaced from the geocentric IJK coordinate system. In the IJK system, Newton's law of gravitation for the force of gravity of the Earth acting on the satellite takes on the form

$$\vec{F}_g = -\frac{Gm_{\oplus}m_{sat}}{r^2} \left(\frac{\vec{r}}{r} \right) \quad (2-2)$$

To determine the equation of motion, we now use Fig. 2-3 and Eq. (2-2). The position vectors of the Earth and satellite with respect to the origin of the XYZ coordinate system are \vec{r}_{\oplus} and \vec{r}_{sat} , respectively. Thus, a vector from the Earth to the satellite is

$$\vec{r}_{\oplus sat} = \vec{r}_{sat} - \vec{r}_{\oplus}$$

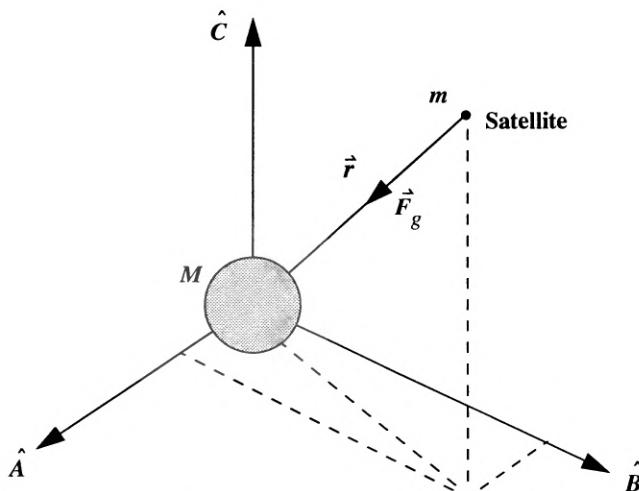


Figure 2-2. **Gravitational Forces Acting on a Satellite.** In a perfectly inertial system, the ideal coordinate system ABC doesn't rotate or accelerate and is often at the center of mass. Gravity acts on each object as though it were a spherically symmetrical point mass.

The importance of the inertial coordinate system is apparent now because it allows us to differentiate this vector equation without considering the derivatives of each axis of the coordinate system. Thus, we obtain the vector derivative by simply differentiating the coordinates of each vector component. The second derivative yields the satellite's acceleration relative to the center of the Earth:

$$\ddot{\vec{r}}_{\oplus \text{sat}} = \ddot{\vec{r}}_{\text{sat}} - \ddot{\vec{r}}_{\oplus}$$

Newton's second law and his gravitational law permit us to write the *inertial* forces:

$$\hat{\vec{F}}_{g \text{sat}} = m_{\text{sat}} \ddot{\vec{r}}_{\text{sat}} = -\frac{Gm_{\oplus}m_{\text{sat}}}{r^2} \left(\frac{\vec{r}}{r} \right) \quad \hat{\vec{F}}_{g \oplus} = m_{\oplus} \ddot{\vec{r}}_{\oplus} = \frac{Gm_{\oplus}m_{\text{sat}}}{r^2} \left(\frac{\vec{r}}{r} \right)$$

Notice the positive sign on the gravitational force in the right-hand equation. The force of the Earth is opposite to the direction of the satellite's force. Solving for the individual accelerations and using the differentiated form of the acceleration above, we can now group these equations and solve them for the relative acceleration $\ddot{\vec{r}}$ (without subscripts):

$$\ddot{\vec{r}} = -\frac{Gm_{\oplus}}{r^2} \frac{\vec{r}}{r} - \frac{Gm_{\text{sat}}}{r^2} \frac{\vec{r}}{r}$$

or

$$\ddot{\vec{r}} = -\frac{G(m_{\oplus} + m_{\text{sat}})}{r^2} \frac{\vec{r}}{r} \quad (2-3)$$

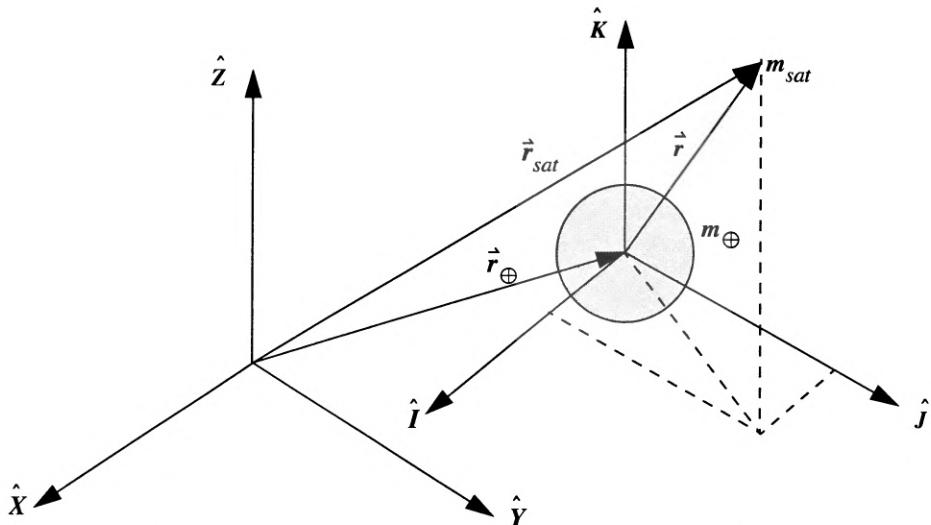


Figure 2-3. Geometry for Two Bodies in an Inertial Reference Frame. XYZ is assumed to be an inertial coordinate system. IJK is displaced from XYZ , but does not rotate or accelerate with respect to XYZ .

It's convenient now to use the gravitational constant μ derived in Sec. 1.8. If we assume the satellite's mass is significantly smaller (by several orders of magnitude) than the mass of the attracting body (the Earth in this case), we may ignore the satellite's mass and replace Gm_{\oplus} with μ .

$$\ddot{\vec{r}} = -\frac{\mu}{r^2} \hat{r} \quad (2-4)$$

This is the basic two-body equation. It's often called the *relative* form, because the motion is referenced to the primary attracting body. Remember, the results for this equation are valid only with certain assumptions. It is a second-order, nonlinear, vector, differential equation which serves as the basis of our development.

2.2.1 Assumptions for the Two-Body Equation

To develop the two-body equation, we assume:

1. The mass of the satellite is negligible compared to that of the attracting body. This is reasonable for artificial satellites in the foreseeable future.
2. The coordinate system chosen for a particular problem is inertial. The importance of this assumption becomes apparent in the derivations which follow; in essence, it removes derivatives of the coordinate system itself

when differentiating vectors. For most cases examined in this book, the heliocentric system satisfies the requirement for interplanetary probes, and the geocentric equatorial system serves for satellites orbiting the Earth. Remember, these are really pseudoinertial frames.

3. The bodies of the satellite and attracting body are spherically symmetrical, with uniform density. This allows us to treat each as a point mass.
4. No other forces act on the system except for gravitational forces that act along a line joining the centers of the two bodies.

For two-body motion, these assumptions are adequate and even necessary. But as later chapters will describe, perturbations significantly change the orbit, so we can't ignore them for real-world simulations. A lot of analysis is possible using two-body equations, and we'll explore many applications in subsequent chapters. We can sometimes gain insight into more complex problems using two-body motion, but it's still approximate.

2.2.2 Specific Angular Momentum

We can develop an expression for angular momentum that is *independent* of mass and constant for the orbit. To derive the *specific angular momentum*, \hat{h} , begin by cross-multiplying the two-body equation with the position vector, \hat{r} .

$$\hat{r} \times \ddot{\hat{r}} + \dot{\hat{r}} \times \frac{\mu}{r^3} \hat{r} = 0$$

Because $\hat{r} \times \dot{\hat{r}} = 0$, the second term vanishes. Now, consider the following differential:

$$\frac{d}{dt}(\hat{r} \times \dot{\hat{r}}) = \dot{\hat{r}} \times \dot{\hat{r}} + \hat{r} \times \ddot{\hat{r}} = \hat{r} \times \ddot{\hat{r}}$$

We can substitute this differential into the original equation, but because this results in the derivative being equal to zero, the internal quantity must be a constant, \hat{h} . If we rearrange and substitute the position derivative with the velocity, the original expression becomes

$$\hat{h} = \hat{r} \times \hat{v} = \text{constant} \quad (2-5)$$

Notice the lack of mass from the usual connotation of momentum, hence the *specific* qualifier. Because \hat{h} is the vector cross product between \hat{r} and \hat{v} , it must lie perpendicular to the plane of the orbit. For two-body motion, the satellite's motion is always confined to this plane, called the orbital plane. An important observation results. Any position and velocity vector pair taken at the same point in time will uniquely determine the specific angular momentum. This is an important factor when solving orbital problems.

Let's also develop equations that determine the magnitude of the specific angular momentum if the position and velocity vectors are unknown. Figure 2-4 illustrates this situation. The standard cross product allows us to investigate the angular momentum using

the auxiliary angle, θ , between the position and velocity vectors. But we measure the flight-path angle, ϕ_{fpa} , as the complementary angle, so we use a standard trigonometric reduction:

$$h = r v \cos(\phi_{fpa}) \quad (2-6)$$

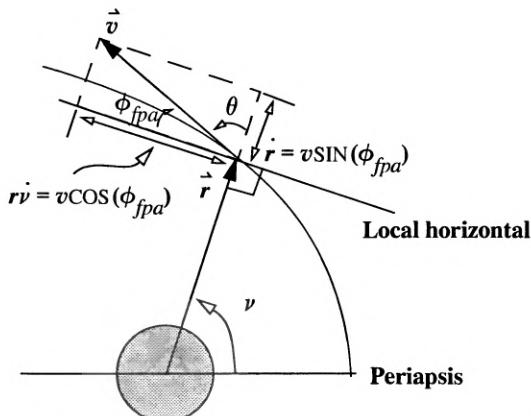


Figure 2-4. Geometry of the Flight-Path Angle. The flight-path angle, ϕ_{fpa} , is measured from the local horizontal to the velocity vector, which we can break into radial and transverse components. The auxiliary angle θ is sometimes used in derivations.

At apoapsis and periapsis, the flight-path angle is zero, and the respective angular momenta are given by

$$\begin{aligned} h &= r_a v_a \\ h &= r_p v_p \end{aligned} \quad (2-7)$$

We find another useful expression for calculating angular rate as follows. Using the cosine expression of the flight-path angle in Fig. 2-4,

$$\cos(\phi_{fpa}) = \frac{r \dot{v}}{v}$$

Substituting the result of Eq. (2-6) solved for $\cos(\phi_{fpa})$ into this equation gives us

$$h = r^2 \dot{v} \quad (2-8)$$

We can also get an expression relating the semiparameter and the angular momentum by combining the trajectory equation [Eq. (2-14)] and appropriate formulas from analytical geometry. The standard quantity in the numerator of a polar form of a conic (or trajec-

tory equation) is given as le , where the eccentricity is known and l represents the distance from the directrix to the primary focus (see Fig. 1-3). The numerator term becomes

$$le = \left(\frac{a}{e} - \sqrt{a^2 - b^2} \right) e$$

Notice that the square root is part of the formula for eccentricity [Eq. (1-7)]. When we replace it by the quantity ae ,

$$le = a - ae^2$$

By Eq. (1-11), $le = p$. If the numerator of the polar-trajectory equation is the semiparameter,

$$p = \frac{h^2}{\mu} \quad h = \sqrt{\mu p} \quad (2-9)$$

2.2.3 Specific Mechanical Energy

To derive the energy constant of motion that is also independent of mass—the *specific mechanical energy* (SME) or ξ —begin by dot multiplying the two-body equation [Eq. (2-4)] with the velocity vector. Recognize that $\dot{\vec{v}} = \dot{\vec{r}}$.

$$\dot{\vec{v}} \cdot \dot{\vec{v}} + \dot{\vec{v}} \cdot \frac{\mu}{r^3} \vec{r} = 0$$

Next, eliminate the vectors. The general form of a dot product is $\dot{r} \cdot \dot{v} = r v \cos(\theta)$, which is a scalar quantity. However, from Fig. 2-4, notice $v \cos(\theta) = \dot{r}$, where \dot{r} is the *radial component* of change of the position vector (don't confuse it with the magnitude of velocity, v). Substitution yields $\dot{r} \cdot \dot{v} = r \dot{r}$. An analogous operation for the velocity and acceleration vectors enables us to write the previous equation as

$$v \dot{v} + \dot{r} \frac{\mu}{r^3} r = 0$$

Again, we must recognize certain derivatives as

$$\frac{d}{dt} \left(\frac{v^2}{2} \right) = v \dot{v} \quad \frac{d}{dt} \left(-\frac{\mu}{r} \right) = \frac{\mu}{r^2} \dot{r}$$

Substitution allows

$$\frac{d}{dt} \left(\frac{v^2}{2} \right) + \frac{d}{dt} \left(-\frac{\mu}{r} \right) = 0$$

Notice if the sum of the time rate of change is zero, the quantity must be a constant. If we integrate both sides with respect to time, the differentials disappear, and the right-hand side also becomes a constant. The constant is called the *specific mechanical energy*, ξ . This equation is often referred to as the *energy integral*, or the *vis-viva equation*. Rearranging then,

$$\xi = \frac{v^2}{2} - \frac{\mu}{r} + c \quad (2-10)$$

The choice of a value for c is arbitrary, allowing us to designate the condition for $\xi = 0$. The physics community should recognize this equation where $c = \mu/R_{\oplus}$. ξ varies with height and speed and is zero at the Earth's surface. In astrodynamics c is defined to be zero. Using this definition, the same concept still applies: the higher and faster you go, the more your energy increases. However, you start with a negative value for energy when at rest at the surface of the Earth. We'll use this definition throughout the book. You should also recognize the energy as the sum of kinetic and potential energy (less the mass). The form of **potential energy** ($-\mu/r$) appears throughout the book.

A consequence of this definition is the concept of a *sphere of influence* (SOI). The **sphere of influence** for a central body is an imaginary sphere within which the gravity of the object is primarily responsible for all orbital motion. Outside this sphere, other bodies influence most of the satellite's motion. Although most scenarios we'll examine take place within the Earth's SOI, missions can cross (and have crossed) the boundary continuing to other planets. The SOI is more a concept than a physical reality. At an infinite distance ($r = \infty$ at the SOI), only the velocity determines the energy. It may be zero or some positive value, $v^2/2$. If the energy is positive, further motion is possible, whereas zero energy means the satellite remains stationary. The quantity $v^2/2$ is often called *v-infinity*, v_{∞} , because it's the excess velocity at "infinity." It is sometimes given the symbol c_3 . Using Eq. (2-11) we can also solve for the **hyperbolic excess velocity** which is $v_{\infty}^2 = -\mu/a$. Different conventions exist.

It's also useful to define ξ in terms of the semimajor axis, a , whenever we don't know the position and velocity vectors. Use the periapsis location to find this result. Solving Eq. (2-7) for the velocity at periapsis and substituting into Eq. (2-10) yield

$$\xi = \frac{v^2}{2} - \frac{\mu}{r} = \frac{h^2}{2r_p^2} - \frac{\mu}{r_p}$$

We already know that $r_p = a(1-e)$ and $h = \sqrt{\mu p} = \sqrt{\mu a (1-e^2)}$, so

$$\xi = \frac{\mu a (1-e^2)}{2a^2 (1-e)^2} - \frac{\mu}{a(1-e)}$$

The result (for orbits other than parabolic) is then

$$\xi = -\frac{\mu}{2a} \quad (2-11)$$

The previous equations are sometimes combined to give an alternate formulation of the energy interval. (It's really Eq. (2-21) squared.)

$$v^2 = \mu \left(\frac{2}{r} - \frac{1}{a} \right) \quad (2-12)$$

2.2.4 Kepler's First Law (Trajectory Equation)

Kepler's first law states that each planet travels in an ellipse, or conic section. The two-body equation describes the path of a small mass orbiting a large central body. Although simple, the trajectory equation gives us great insight into orbital motion. Begin by crossing the two-body equation [Eq. (2-4)] into the specific angular momentum vector, \hat{h} :

$$\ddot{\vec{r}} \times \hat{h} + \frac{\mu \vec{r}}{r^3} \times \hat{h} = 0$$

Notice the first term is simply the derivative,

$$\frac{d}{dt} (\dot{\vec{r}} \times \hat{h}) = \ddot{\vec{r}} \times \hat{h} + \dot{\vec{r}} \times \dot{\hat{h}} = \ddot{\vec{r}} \times \hat{h}$$

because $\dot{\hat{h}} = 0$. A useful vector identity [Eq. (C-6)] allows us to expand the equation above. This identity, along with the definition of the angular momentum vector [Eq. (2-5)], permits

$$\frac{\mu}{r^3} (\vec{r} \times \hat{h}) = \frac{\mu}{r^3} \vec{r} \times (\vec{r} \times \vec{v}) = \frac{\mu}{r^3} ((\vec{r} \cdot \vec{v}) \vec{r} - (\vec{r} \cdot \vec{r}) \vec{v}) = \frac{\mu}{r^3} ((\vec{r} \cdot \vec{v}) \vec{r} - r^2 \vec{v})$$

From our discussion of the vector dot product while deriving ξ , we know we can simplify the dot product to get

$$\frac{\mu}{r^3} (\vec{r} \times \hat{h}) = \frac{\mu}{r^2} \vec{r} \cdot \vec{r} - \frac{\mu}{r} \vec{v}$$

The key is to recognize this quantity as the following time derivative:

$$-\mu \frac{d}{dt} \left(\frac{\vec{r}}{r} \right) = \frac{\mu}{r^2} \vec{r} \cdot \vec{r} - \frac{\mu}{r} \vec{v}$$

We can substitute into the starting equation and reverse the equation:

$$\frac{d}{dt} (\dot{\vec{r}} \times \hat{h}) = \mu \frac{d}{dt} \left(\frac{\vec{r}}{r} \right)$$

Integrating both sides,

$$\dot{\vec{r}} \times \vec{h} = \mu \frac{\vec{r}}{r} + \vec{B} \quad (2-13)$$

where \vec{B} is the vector constant of integration. We know \vec{B} lies in the orbital plane because the right-hand side is normal to the angular momentum. Now, dot multiply this equation by \vec{r} to get a scalar relation:

$$\vec{r} \cdot (\dot{\vec{r}} \times \vec{h}) = \vec{r} \cdot \left(\mu \frac{\vec{r}}{r} + \vec{B} \right)$$

Using another vector identity, rearrange the first term as $(\dot{\vec{r}} \times \vec{r}) \cdot \vec{h} = h^2$ and make the right-hand side into a scalar relation with the dot-product formula. We must introduce an angle from the dot-product relation. For now, let this angle be the true anomaly, ν , and express the relation as

$$h^2 = \mu r + r B \cos(\nu)$$

Solving for the position yields the solution:

$$r = \frac{\frac{h^2}{\mu}}{1 + \frac{B}{\mu} \cos(\nu)}$$

Students of analytical geometry will recognize this expression as the polar form of a conic section. The standard form given in books on analytical geometry defines the quantity B/μ as the eccentricity, e , which I've already introduced. Standard texts also reference the angle from periaxis to the current location on the conic section; thus, our previous choice of ν was correct. We've shown that the quantity h^2/μ is equal to the semiparameter, p , in Eq. (2-9). The *trajectory equation* is then

$$r = \frac{p}{1 + e \cos(\nu)} \quad (2-14)$$

Notice the use of eccentricity, e , which defines the shape of the orbit. This equation verifies and *extends* Kepler's first law because it doesn't restrict the motion to an ellipse.

It's useful to develop an equation for the radial-velocity component of the velocity vector, \dot{r} , when we examine perturbations. Taking the derivative of Eq. (2-14),

$$\dot{r} = \frac{p \dot{v} e \sin(\nu)}{(1 + e \cos(\nu))^2}$$

which simplifies to

$$\dot{r} = \frac{r\dot{\nu}e \sin(\nu)}{1 + e \cos(\nu)} \quad (2-15)$$

Remember, this value *isn't* equal to the magnitude of the velocity vector!

2.2.5 Kepler's Second and Third Laws

We now derive the *period*, \mathcal{P} , of a satellite. Recall that the period relies on Kepler's second and third laws, stating that equal areas are swept out in equal intervals of time and that the square of the period of the planet is proportional to the cube of the semimajor axis. Figure 2-5 shows the geometries used to develop the period equation.

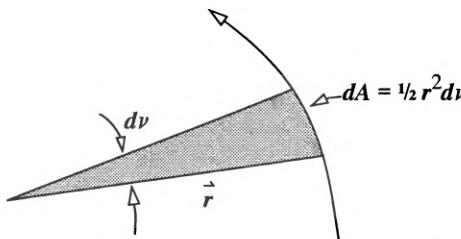


Figure 2-5. Area Swept Out by a Satellite. According to Kepler's second law, satellites sweep out equal areas in equal time intervals.

Recall from Fig. 2-4 the transverse component of velocity is simply $v \cos(\phi_{fpa})$, which we can also express in terms of simple rotational velocity, $\dot{\nu}$. Rewriting Eq. (2-8) gives us

$$h = \frac{r^2 d\nu}{dt}$$

from which we rearrange to find

$$dt = \frac{r^2 d\nu}{h}$$

Recall from calculus that the differential of area (dA) swept out by the radius vector as it moves through a differential angle ($d\nu$) is given by the expression

$$dA = \frac{1}{2} r^2 d\nu$$

After solving each of the two previous two relations for $d\nu$ and equating, we have

$$dt = \frac{2}{h} dA$$

This equation proves Kepler's second law that equal areas are swept out in equal times because h is constant for the orbit. During one orbit period, P , the radius vector sweeps out the entire area of the ellipse (or circle). Integrating the equation above over 2π radians of ν gives us an expression for one period:

$$P = \frac{2\pi ab}{h}$$

where πab is the area of the ellipse (and for a circle, πa^2). From Chap. 1, recall the scaling formulas of Eq. (1-5) and Eq. (1-6). These formulas allow us to obtain

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

and using Eq. (2-9), we find an expression of Kepler's third law:

$$P = 2\pi \sqrt{\frac{a^3}{\mu}} \quad (2-16)$$

The period often categorizes the orbit. For instance, the U.S. Air Force's Space Command defines *deep-space* satellites as those whose periods are greater than or equal to 225 minutes. Unfortunately, this definition was developed decades ago, when most satellites were near the Earth and few satellites had large eccentricities. A satellite in a circular orbit with a 225-minute period is at only about 5870 km in altitude ($a = 12,248$ km). You can argue that this point marks a transition from gravitational effects to those induced by third bodies, but we can form better categories (see also Sec. 8.10). An important distinction is made for *geosynchronous* satellites which have periods equal to the Earth's rotation (24 sidereal hours). *Geostationary* satellites remain over one spot on the Earth at all times. They also have 24-hour periods but *must* be equatorial. Don't confuse these orbits with the more general class of geosynchronous satellites, which can be nonequatorial. *Semi-synchronous* satellites have 12-hour periods (like the Global Positioning Satellites), and *super-synchronous* satellites have periods of more than 24 hours. This book defines several general categories of orbits by the forces, including gravity, which influence their motions: low-Earth orbits (LEO—to about 800 km altitude, $a < 7178$ km); mid-Earth orbits (MEO—from about 800 km to 30,000 km altitude, $a = 7178$ km to 36,378 km), and geosynchronous orbits (GEO—at about 35,780 km altitude, $a \approx 42,158$ km). The first category is mainly affected by atmospheric drag, and the two latter categories are influenced chiefly by solar-radiation pressure and third-body effects. *Deep space* refers to satellites that orbit near or above geosynchronous altitude.

An alternate way to express Kepler's third law proves very useful for analysis. Let

$$n = \frac{2\pi}{P} = \sqrt{\frac{\mu}{a^3}} \quad (2-17)$$

where n represents the mean motion with units of radians per unit of time. Thus it's an angular rate or the mean angular rate of change of the object in orbit. Substituting into Eq. (2-16) squared (i.e., Kepler's third law) and rearranging yields an alternate form:

$$n^2 a^3 = \mu \quad \mu = G(m_1 + m_2) \quad (2-18)$$

Thus the product of the square of the mean motion with the cube of the semimajor axis is constant. That constant is the gravitational parameter, μ .

Eq. (2-16) permits rapid determination of the size of a particular orbit (semimajor axis). If we solve Eq. (2-16) for the semimajor axis,

$$a = \left(\mu \left(\frac{\rho}{2\pi} \right)^2 \right)^{1/3} = \sqrt[3]{\frac{\mu}{n^2}} \quad (2-19)$$

Or we can solve for mean motion, n , which I'll discuss later. Consider the following example.

▼ Example 2-1: Determining the Semimajor Axis from the Period.

GIVEN: A geosynchronous orbit

FIND: Semimajor axis, a

ρ is equal to 24 sidereal hours (or 106.795 869 7 TU in canonical units) because the Earth completes one revolution through 2π radians in this amount of time. Be careful not to use 24 solar hours, which is a little *more* than one complete revolution.

$$a = \left(\mu \left(\frac{\rho}{2\pi} \right)^2 \right)^{1/3} = \left(1 \left(\frac{106.795 \ 869 \ 7}{2\pi} \right)^2 \right)^{1/3} = 6.610 \ 734 \ 645 \text{ ER}$$

Determine the semimajor axis in metric units by using the conversion from the table on the inside front cover for the Earth's radius (6378.1363 km).

$$a = 6.610 \ 734 \ 645 (6378.1363) = 42,164.17124 \text{ km}$$

From these results, we can derive several important points:

- (1) Using canonical units permits a savings in the calculation because μ is 1.0 ER³/TU² versus 398,600.4 km³/s².
- (2) The period in the numerator is the number of TUs in one *sidereal* day, and it comes from the table on the inside front cover ($1.0/0.009 \ 363 \ 658 \ 228 \ 3 = 106.795 \ 869 \ 7$). Don't confuse this quantity with the TUs in one *solar* day (107.088 268 9).
- (3) It may not be quite as obvious, but the complete quantity in the inner brackets is the reciprocal of the Earth's rotation rate ($1/\omega_{\oplus}$).

2.2.6 Velocity Formulas

Although we analyze orbits mostly by using the classical orbital elements, we can generate expressions for the velocity's magnitude to augment the available information on the satellite. I'll show several forms. Velocity equations come from solving the equation for specific mechanical energy in terms of the velocity's magnitude. From Eq. (2-10),

$$\xi = \frac{v^2}{2} - \frac{\mu}{r}$$

Solving for the velocity gives us

$$v = \sqrt{2\left(\frac{\mu}{r} + \xi\right)} \quad (2-20)$$

Because ξ is defined for *all* orbits, this equation is good for all conic sections. To find an alternate form, replace the equation for specific mechanical energy with Eq. (2-11):

$$v = \sqrt{\frac{2\mu}{r} - \frac{\mu}{a}} \quad (2-21)$$

When writing computer software to use this formula, be careful. Remember, the semimajor axis is infinite for parabolic orbits, so you may get indeterminate results. Sometimes, it's useful to express this relation in terms of the true anomaly and not the semimajor axis. Using the trajectory equation [Eq. (2-13)] and the definition of the semiparameter gives us

$$\begin{aligned} r &= \frac{p}{1 + e \cos(\nu)} = \frac{a(1 - e^2)}{1 + e \cos(\nu)} \\ \frac{1}{a} &= \frac{(1 - e^2)}{r(1 + e \cos(\nu))} \end{aligned}$$

Substituting this expression into the velocity equation [Eq. (2-21)] yields

$$v = \sqrt{\frac{\mu}{r} \left(2 - \frac{1 - e^2}{1 + e \cos(\nu)} \right)} \quad (2-22)$$

The circle is just a special case of the ellipse, so we can derive the relevant formulas from the general case [Eq. (2-21)]. Because the radius and the semimajor axis are identical, let's replace the semimajor axis with the position and simplify:

$$v_{circle} = \sqrt{\frac{\mu}{r}} \quad (2-23)$$

The parabola represents another special form of a conic section, but the orbit isn't closed—an important distinction because the satellite won't return. The specific mechanical energy is defined in Eq. (2-10); however, because the semimajor axis is infinite, Eq. (2-11) isn't computationally valid. In addition, with an infinite semimajor axis, the specific mechanical energy is zero. So

$$\xi_{parabola} = \frac{v^2}{2} - \frac{\mu}{r} = 0$$

The parabola represents the case for which the satellite will just escape the Earth and not return. This case gives us the opportunity to define the *escape velocity*, v_{esc} , the velocity at which a satellite can just escape from a planet. Although this book won't focus on interplanetary missions, I introduce the concept to completely examine the problem.

Because the parabolic orbit has zero specific mechanical energy, as shown above, the satellite reaches infinity ($r = \infty$) with zero velocity. Solving for velocity, we find

$$v_{\text{parabola}} = v_{\text{esc}} = \sqrt{\frac{2\mu}{r}} \quad (2-24)$$

Eq. (2-24) gives the velocity on a parabolic orbit at any given radius, r .

2.3 Three-body and n -body Equations

The standard two-body relations underlie much of the important work in astrodynamics, but sometimes we need to model the real world by including other bodies. The next logical step, then, is to derive formulas for a three-body system, using the same initial concepts as in Sec. 2.2. A further generalization of the three-body problem is the n -body problem. Of course, the notion of adding bodies has limits. For instance, the effects of Pluto's non-spherical shape on the orbit of a satellite in low-Earth orbit are real, but accounting for them is unnecessary. In the following development, I'll use the subscript "sat" whenever I reference the satellite to avoid possible ambiguity in the summations. In addition, I'll use a double subscript to indicate vector differences, e.g., $\vec{r}_{\oplus\odot} = \vec{r}_\odot - \vec{r}_\oplus$.

2.3.1 Inertial, Relative, and Barycentric Formulas

Eq. (2-1) still applies when determining the equations of motion for multiple bodies, except that additional forces arise from the interactions between the bodies. Figure 2-6 shows the general geometry.

For a three-body system, (Sun, Earth, and satellite, for instance), we examine the individual forces. The relationship between the vectors in Fig. 2-6 allows us to solve for the satellite's acceleration *relative to* the Earth:

$$\vec{r}_{\oplus\text{sat}} = \vec{r}_{\text{sat}} - \vec{r}_\oplus$$

Remember, this is an inertial system XYZ. Successive differentiation yields

$$\ddot{\vec{r}}_{\oplus\text{sat}} = \ddot{\vec{r}}_{\text{sat}} - \ddot{\vec{r}}_\oplus$$

Based on Newton's second law and the law of gravitation, the sum of the forces acting on the Earth is

$$\sum \vec{F}_{g\oplus} = m_\oplus \ddot{\vec{r}}_\oplus = \frac{Gm_\oplus m_{\text{sat}} \vec{r}_{\oplus\text{sat}}}{r_{\oplus\text{sat}}^3} + \frac{Gm_\oplus m_\odot \vec{r}_{\oplus\odot}}{r_{\oplus\odot}^3}$$

The result is a combination of the satellite pulling on the Earth and the Sun pulling on the Earth (left and right terms respectively). The acceleration, $\ddot{\vec{r}}_\oplus$, represents the acceleration

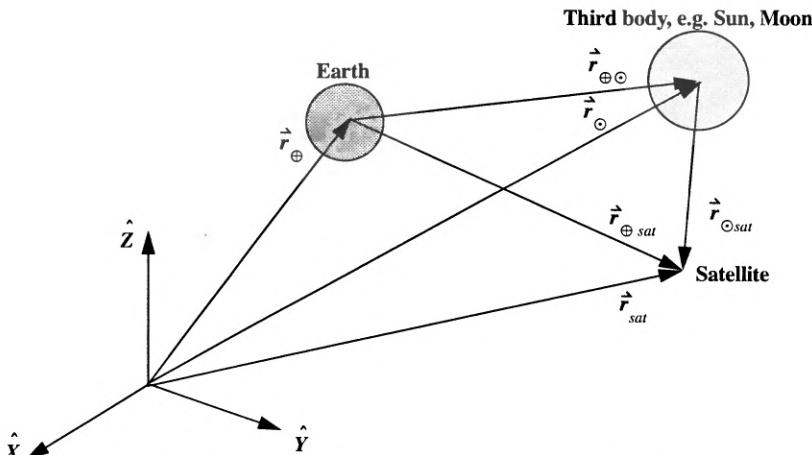


Figure 2-6. Three-body Geometry. The inertial coordinate system XYZ is used to measure all the accelerations. The double subscripts are necessary to completely define the direction of each vector. The satellite is intended to be orbiting the Earth, but for clarity, I've greatly exaggerated the figure.

an observer would see from the origin of the inertial XYZ system. (One such inertial system is the heliocentric system or a variant with the origin at the barycenter of the solar system.) For the satellite, the sum of the gravitational forces is

$$\sum \vec{F}_{gsat} = m_{sat} \ddot{\vec{r}}_{sat} = -\frac{Gm_{\oplus}m_{sat}\vec{r}_{\oplus sat}}{r_{\oplus sat}^3} - \frac{Gm_{\odot}m_{sat}\vec{r}_{\odot sat}}{r_{\odot sat}^3}$$

Notice both forces are negative because they are in a direction opposite that of the vectors to the satellite. Appropriate substitutions produce

$$\ddot{\vec{r}}_{\oplus sat} = -\frac{Gm_{\oplus}\vec{r}_{\oplus sat}}{r_{\oplus sat}^3} - \frac{Gm_{\odot}\vec{r}_{\odot sat}}{r_{\odot sat}^3} - \frac{Gm_{sat}\vec{r}_{\oplus sat}}{r_{\oplus sat}^3} - \frac{Gm_{\odot}\vec{r}_{\oplus \odot}}{r_{\oplus \odot}^3}$$

and the final result (Escobal, [1965] 1985, 37) using $\vec{r}_{sat\odot} = -\vec{r}_{\odot sat}$ is

$$\ddot{\vec{r}}_{\oplus sat} = -\frac{G(m_{\oplus} + m_{sat})\vec{r}_{\oplus sat}}{r_{\oplus sat}^3} + Gm_{\odot} \left(\frac{\vec{r}_{sat\odot}}{r_{sat\odot}^3} - \frac{\vec{r}_{\oplus \odot}}{r_{\oplus \odot}^3} \right) \quad (2-25)$$

This form is sometimes referred to as the “relative” form, but this term may be misleading. We have only transformed the acceleration to a different origin. Therefore, the acceleration is measured relative to the Earth and the subscripts are necessary. Several important terms are in this relation. The first term is the two-body acceleration of the Earth

acting on the satellite. The second term has two parts and it represents the *perturbation*, or the additional forces beyond simple two-body motion. The left-hand term is called the *direct effect* because it's the acceleration of the Sun (or third body) directly on the satellite. The last term is called the *indirect effect* because it's the force per unit mass (acceleration) of the Sun (or third body) on the Earth. As the position of the Earth changes, the subsequent acceleration on the satellite changes. For some applications, the direct and indirect effects may have nearly the same magnitude—a numerical challenge for some problems. More often, the indirect term may not be very large. In Chap. 7, we'll see that expanding the direct term in Legendre polynomials contains a term that exactly cancels the indirect term.

A finite sum for the total acceleration of the i th-body due to gravitational attraction from n bodies is given in Eq. (2-26). The three-body problem is a special case where $n = 3$.

$$\ddot{\vec{r}}_i = -G \sum_{\substack{j=1 \\ j \neq i}}^n \frac{m_j}{r_{ji}^3} \vec{r}_{ji} \quad \vec{r}_{ji} = \vec{r}_i - \vec{r}_j \quad i = 1, \dots, n \quad (2-26)$$

The inertial form is more useful if viewed from the center of mass of the primary body [as in Eq. (2-25)]. The general series representation of this relative acceleration (1 is the primary body, 2 is replaced with satellite) is

$$\ddot{\vec{r}}_{1sat} = -\frac{G(m_1 + m_{sat})}{r_{1sat}^3} \vec{r}_{1sat} + G \sum_{j=3}^n m_j \left(\frac{\vec{r}_{satj}}{r_{satj}^3} - \frac{\vec{r}_{1j}}{r_{1j}^3} \right) \quad (2-27)$$

Although the form of Eq. (2-26) is desirable to explore the ten known integrals of the motion (see Sec. 2.3.2), Eq. (2-27) is useful in real world applications because it's more convenient to reference the Earth's center.

Another useful form of equations is the *barycentric* form which originates at the barycenter of the system. Although greatly exaggerated, Fig. 2-7 shows the general geometry. This relationship is very similar to developing the relative equations of motion with the origin being simply a fixed point at the system's barycenter. A *synodic* frame (which rotates about the barycenter) is useful in studies of the restricted three-body problem (Sec. 2.3.3). This frame requires additional terms from Eq. (1-30).

Our ultimate goal is to obtain an expression for the satellite's acceleration as seen from the barycenter [center of mass in Eq. (2-29)]. The position of the i th body in the barycentric frame is $\vec{r}_{Bi} = \vec{r}_i - \vec{r}_B$, so by Eq. (2-26), \vec{r}_{ji} in the barycentric frame is

$$\vec{r}_{Bji} = \vec{r}_i - \vec{r}_B - (\vec{r}_j - \vec{r}_B) = \vec{r}_i - \vec{r}_j = \vec{r}_{ji}$$

Therefore, the acceleration of the i th body in the barycentric frame is (see also Fig. 2-7)

$$\ddot{\vec{r}}_{Bi} = \ddot{\vec{r}}_i - \ddot{\vec{r}}_B = \ddot{\vec{r}}_i$$

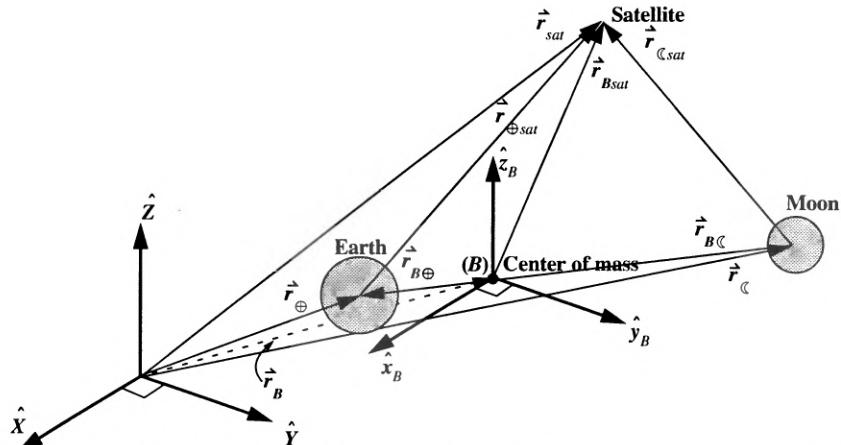


Figure 2-7. Geometry for the Barycentric Equations of Motion. Remember that the barycentric coordinate system is aligned with the inertial XYZ system. The Z and z_B axes are aligned.

because $\ddot{\vec{r}}_B = 0$ due to conservation of linear momentum. Thus we see that the equations of motion are independent of a particular origin and therefore of any particular inertial frame. They depend only on the relative position vectors, \vec{r}_{ji} , and the second derivatives, which are also independent of the inertial frame's origin. Consequently, Eq. (2-26) also applies to the barycentric formulation. The generalized barycentric equation of motion is

$$\ddot{\vec{r}}_{Bsat} = \ddot{\vec{r}}_{sat} = -G \sum_{\substack{j=1 \\ j \neq 3}}^n \frac{m_j \vec{r}_{jsat}}{r_{jsat}^3} \quad (2-28)$$

2.3.2 Ten Known Integrals

In general, solving differential equations requires a fixed number of integration constants. Consider a simple gravity problem in which we have constant acceleration over time, $a(t) = a_o$. If we integrate this equation, we obtain the velocity, $v(t) = a_o t + v_o$. Integrating once more provides, $r(t) = \frac{1}{2}a_o t^2 + v_o t + r_o$. To complete the solution, we must know the initial conditions. This example is a straight-forward analytical solution using the initial values, or a function of the time and constants of integration, called *integrals of the motion*. Unfortunately, this isn't always the case. When initial conditions alone don't provide a solution, integrals of the motion can reduce the order of the differential equation, also called the *degrees of freedom* of the dynamical system. Ideally, if the number of integrals equals the order of the differential equations, we can reduce it to order zero. These integrals are constant functions of the initial conditions, as well as the position and velocity at any time, hence the term *constants of the motion*.

Consider the two-body problem—a twelfth-order system represented by three second-order, differential equations for each body [$(3 \times 2) \times 2$ bodies]. Conservation of linear momentum provides six constants from Eq. (2-29) with the initial position and velocity vectors for the center of mass. We use this information to transform the origin to the barycenter of the primary body because we assume the mass of the satellite is negligible. This results in a system of three second-order equations describing the motion of the satellite. We then use the conservation of energy to find a constant ($\xi = -\mu/2a$). The conservation of angular momentum gives us three components in the angular momentum vector, and this maps to i and Ω . Using Kepler's first law, e and ω are also found to be constants. Finally, a combination of Kepler's second law and the polar component of the angular momentum vector define time of perigee passage, T . Thus, we have twelve constants for the solution.

For the n -body problem, a system of $3n$ second-order differential equations, we need $6n$ integrals of the motion for a complete solution. Conservation of linear momentum provides six, conservation of energy one, and conservation of total angular momentum three, for a total of ten. There are no laws analogous to Kepler's first two laws to obtain additional constants, thus we are left with a system of order $6n - 10$ for $n \geq 3$.

These equations for n -bodies, $n \geq 3$, defy all attempts at closed-form solutions. H. Brun showed that there were no other algebraic integrals in 1887 (Szebehely (1967, 43)). Although Poincaré later generalized Brun's work, we still have only the *ten known integrals*. They give us insight into the motions within the three-body and n -body problems. Conservation of total linear momentum assumes no external forces are on the system. Thus, if we sum all the forces (mass times acceleration) in Eq. (2-26), the total is zero (a consequence of Newton's third law):

$$\sum_{i=1}^n m_i \ddot{\mathbf{r}}_i = -G \sum_{i=1}^n \sum_{j=1, j \neq i}^n \frac{m_i m_j}{r_{ji}^3} \hat{\mathbf{r}}_{ji} = 0$$

The zero result is also seen mathematically because ($\ddot{\mathbf{r}}_{ij} = -\ddot{\mathbf{r}}_{ji}$). We can integrate the left-hand term to yield position- and velocity-vector constants for the system:

$$\sum_{i=1}^n m_i \dot{\mathbf{r}}_i = \dot{\mathbf{v}}_{sys} \quad \sum_{i=1}^n m_i \ddot{\mathbf{r}}_i = \ddot{\mathbf{v}}_{sys} t + \dot{\mathbf{r}}_{sys}$$

From the definition of the system's center of mass, let's solve for the position, $\dot{\mathbf{r}}_{cm}$:

$$\dot{\mathbf{r}}_{cm} = \sum_{i=1}^n m_i \dot{\mathbf{r}}_i / \sum_{i=1}^n m_i = \frac{\dot{\mathbf{v}}_{sys} t + \dot{\mathbf{r}}_{sys}}{m_{total}} \quad (2-29)$$

The first six constants (two vectors) of the motion are now defined by ($\dot{\mathbf{r}}_{sys}/m_{total}$) and ($\dot{\mathbf{v}}_{sys}/m_{total}$), which represent the initial position and velocity of the system's center of mass. The form of the constants clearly indicates that the center of mass moves with *constant* (or zero) velocity and is not influenced by external forces.

We find the next three constants through conservation of total angular momentum. The process is virtually identical to the two-body case (Sec. 2.2.2). Crossing Eq. (2-26) into the product of the mass, m_i , with the position vector, \vec{r}_i , and summing over the n -bodies ($i = 1 \dots n$) gives us

$$\sum_{i=1}^n m_i \ddot{\vec{r}}_i \times \vec{r}_i = -G \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \frac{m_i m_j}{r_{ji}^3} \vec{r}_{ji} \times \vec{r}_i = 0$$

The right-hand side vanishes from Eq. (2-26), $\vec{r}_{ji} \times \vec{r}_i = \vec{r}_i \times \vec{r}_i - \vec{r}_j \times \vec{r}_i = -\vec{r}_j \times \vec{r}_i$, and because there is a double summation, we'll have $\vec{r}_i \times \vec{r}_j$ and $-\vec{r}_i \times \vec{r}_j$ terms, which will all cancel. For the left-hand side, we can use the relation

$$\ddot{\vec{r}}_i \times \vec{r}_i = \frac{d}{dt} (\dot{\vec{r}}_i \times \vec{r}_i)$$

(assuming the masses, m_i , are constant) to obtain

$$\sum_{i=1}^n m_i \ddot{\vec{r}}_i \times \vec{r}_i = \sum_{i=1}^n \frac{d}{dt} (\dot{\vec{r}}_i \times \vec{r}_i) = \frac{d}{dt} \left[\sum_{i=1}^n m_i (\dot{\vec{r}}_i \times \vec{r}_i) \right] = 0$$

Integrating this result yields the total angular momentum of the system:

$$\vec{h}_{sys} = \sum_{i=1}^n m_i \dot{\vec{r}}_i \times \vec{r}_i = \text{constant} \quad (2-30)$$

As with the two-body equations, the fact that angular momentum is conserved means that a plane perpendicular to \vec{h}_{sys} stays fixed in inertial space. For the n -body problem and three-body problem, this plane is sometimes called the *invariable plane* of Laplace. People have suggested using the invariable plane instead of the ecliptic for realizing our coordinate frames, but accurate knowledge of the plane requires long-term ephemerides for all significant bodies in the solar system. This is a source of practical difficulty, so the idea hasn't caught on. Danby (1992, 274) lists approximate parameters for the invariable plane of the solar system as $\Omega = 107^\circ$ and $i = 1.583^\circ$. Section 2.4.1 introduces Ω and i .

We obtain the final constant by examining conservation of total energy. The process follows the development for specific mechanical energy in Sec. 2.2.3, but with a slight variation. Multiply Eq. (2-26) by m_i to find the force acting on each mass.

$$m_i \ddot{\vec{r}}_i = -G \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \frac{m_i m_j}{r_{ji}^3} \vec{r}_{ji}$$

Because we have many bodies, the potential energy (μ/r) is now

$$U = G \sum_{i=1}^n \sum_{j=1}^n \frac{m_i m_j}{r_{ji}}$$

The gradient of the potential function gives the force acting on the i th body (Sec. 7.2),

$$\frac{\partial U}{\partial \vec{r}_i} = -G \sum_{i=1}^n \sum_{j=1}^n \frac{m_i m_j}{r_{ji}^3} \vec{r}_{ji}$$

and the equations of motion are

$$m_i \ddot{\vec{r}}_i = \frac{\partial U}{\partial \vec{r}_i}$$

Multiplying through by $\ddot{\vec{r}}_i$ and summing over the individual bodies ($i = 1, 2, \dots, n$) gives us

$$\sum_{i=1}^n m_i \ddot{\vec{r}}_i \cdot \ddot{\vec{r}}_i = \sum_{i=1}^n \frac{\partial U}{\partial \vec{r}_i} \cdot \vec{r}_i = \frac{dU}{dt}$$

Integrating with respect to time results in $\frac{1}{2}mv^2 = U + \text{constant}$. But the first term is the kinetic energy, and we find the same result as in Eq. (2-10). We can now write the system energy as

$$\xi_{sys} = \frac{1}{2} \sum_{i=1}^n m_i v_i^2 - G \sum_{i=1}^n \frac{m_i m_j}{r_{ji}} = \text{constant} \quad (2-31)$$

2.3.3 General Three-body Problem

Despite the availability of ten constants of the motion, no one has solved the general three-body problem in closed form—it may not be possible. Therefore, work has focused on simplifying the general problem. Sundmann found a power series solution in 1912. When this solution is taken with initial conditions, it yields numerical evaluations of the trajectories over a limited time interval. Consult Szebehely (1967) for an extensive treatment of the problem. One special analytical solution—the restricted three-body problem—has been known since the time of Euler and Lagrange. More recently, numerical techniques have been used to generate solutions.

Restricted Three-body Problem

This problem has received significant attention. It makes two simplifying assumptions:

1. The primary and secondary bodies move in circular orbits about the center of mass, which lies between the two objects.

2. The mass of the third body (satellite) is negligible compared to that of the major bodies.

The resulting problem is usually called the *circular restricted three-body problem*. Sometimes, an additional assumption requires motion only in the orbital plane of the primary and secondary objects. For this *planar* circular restricted three-body problem, the z -component of the equations vanishes. Finding equations to express this solution is difficult. Entire books cover this topic (Szebehely, 1967). Our goal here is to describe qualitative motions and highlight several well-known, classical solutions.

First, let's use a *synodic coordinate frame* for the three-body problem, as shown in Fig. 2-8. I'll introduce equations for a barycentric (fixed) frame shortly. This frame uses lower-case symbols x , y , and z for individual components. The subscripted S with the axes indicates the origin is at the center of mass (the system's barycenter) and the frame is rotating with angular velocity ω_S .

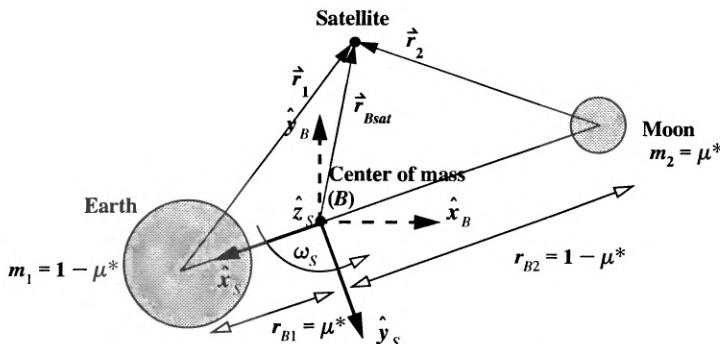


Figure 2-8. Geometry for the Restricted Three-body Problem. Notice the x_S -axis is in the direction of the primary object (Earth), and the y_S -axis lies in the plane of the satellite-Earth-Moon. Remember, this synodic coordinate system is rotating with respect to the fixed barycentric frame (x_B , y_B , z_B). The z_B and z_S axes are aligned. The distances from the barycenter to each object (r_{B1} , r_{B2}) are important in developing the equations of motion. The mass ratio (μ^*) allows us to normalize the problem.

The choice of coordinate system begins by using the barycentric form of the equations of motion. From Eq. (2-28), with $n = 3$, the inertial acceleration is

$$\ddot{\vec{r}}_{Bsat} = -Gm_1 \frac{\vec{r}_{1sat}}{r_{1sat}^3} - Gm_2 \frac{\vec{r}_{2sat}}{r_{2sat}^3}$$

The inherent rotation of the coordinate system requires us to examine the additional terms which arise in the acceleration. Eq. (1-30) delineates inertial and rotating accelerations (I've used symbols from this section and the rotating synodic frame, \vec{r}_S , \vec{v}_S , \vec{a}_S):

$$\ddot{\vec{r}}_{Bsat} = \dot{\hat{a}}_S + \dot{\hat{\omega}}_S \times \vec{r}_S + \vec{\omega}_S \times (\vec{\omega}_S \times \vec{r}_S) + 2\vec{\omega}_S \times \vec{v}_S + \vec{a}_{org}$$

Because the rotation acts about the z_S axis, we can evaluate each of the terms. We've assumed circular motion, so the rate of change of the rotation is zero. Also, the coordinate system is not accelerating with respect to the inertial origin; thus, $\dot{\vec{a}}_{org} = 0$. Switching the relation above to follow common practice, of using cartesian coordinates in the synodic frame for the satellite.

$$\ddot{\vec{r}}_S + \omega_S^2 (x\hat{x}_S + y\hat{y}_S) + 2\omega_S(\dot{y}\hat{x}_S - \dot{x}\hat{y}_S) = \ddot{\vec{r}}_{Bsat}$$

Now we must find simplified expressions for the inertial acceleration. Eq. (2-28) gives the general expression. At this point it's convenient to introduce gradients. We'll use gradients often in Chap. 7. Here, they simplify the notation for the acceleration.

$$\nabla R = \frac{\partial R}{\partial x}\hat{x}_S + \frac{\partial R}{\partial y}\hat{y}_S + \frac{\partial R}{\partial z}\hat{z}_S \quad \ddot{\vec{r}}_{Bsat} = \nabla\left(\frac{\mu_1}{r_1} + \frac{\mu_2}{r_2}\right)$$

The result of a gradient on a potential function, R , is an acceleration in the direction of each vector. Remembering the form of Eq. (1-30), we assume circular orbits to simplify the various acceleration terms. Breaking the vector relations into components gives us the equations of motion using dimensional (not unitized) quantities. We'll use \ddot{x} , \ddot{y} , and \ddot{z} to represent the components of the rotating acceleration in the synodic frame, $\ddot{\vec{r}}_S$.

$$\begin{aligned} \ddot{x} - 2\omega_S \dot{y} - \omega_S^2 x &= \frac{\partial}{\partial x}\left(\frac{\mu_1}{r_1} + \frac{\mu_2}{r_2}\right) = -\frac{\mu_1(x - r_{B1})}{r_1^3} - \frac{\mu_2(x + r_{B2})}{r_2^3} \\ \ddot{y} + 2\omega_S \dot{x} - \omega_S^2 y &= \frac{\partial}{\partial y}\left(\frac{\mu_1}{r_1} + \frac{\mu_2}{r_2}\right) = -\frac{\mu_1 y}{r_1^3} - \frac{\mu_2 y}{r_2^3} \\ \ddot{z} &= \frac{\partial}{\partial z}\left(\frac{\mu_1}{r_1} + \frac{\mu_2}{r_2}\right) = -\frac{\mu_1 z}{r_1^3} - \frac{\mu_2 z}{r_2^3} \end{aligned} \quad (2-32)$$

$$r_1 = \sqrt{(x - r_{B1})^2 + y^2 + z^2} \quad r_2 = \sqrt{(x + r_{B2})^2 + y^2 + z^2}$$

If we multiply each of these equations by $2\dot{x}$, $2\dot{y}$, and $2\dot{z}$, respectively, and sum them,

$$\begin{aligned} 2\dot{x}\ddot{x} + 2\dot{y}\ddot{y} + 2\dot{z}\ddot{z} - 2\omega_S^2(x\dot{x} + y\dot{y}) &= 2\dot{x}\frac{\partial}{\partial x}\left(\frac{\mu_1}{r_1} + \frac{\mu_2}{r_2}\right) \\ &\quad + 2\dot{y}\frac{\partial}{\partial y}\left(\frac{\mu_1}{r_1} + \frac{\mu_2}{r_2}\right) + 2\dot{z}\frac{\partial}{\partial z}\left(\frac{\mu_1}{r_1} + \frac{\mu_2}{r_2}\right) \end{aligned}$$

Integrating this equation results in the well-known *Jacobian integral* equation:

$$\dot{x}^2 + \dot{y}^2 + \dot{z}^2 - \omega_S^2 (x^2 + y^2) = \frac{2\mu_1}{r_1} + \frac{2\mu_2}{r_2} - C \quad (2-33)$$

The constant of integration, C , is known as *Jacobi's constant*, sometimes referred to as a pseudo-integral that exists only in the synodic frame for the restricted problem. Extensive literature describes its use for analysis of possible motion including points of stability and instability.

It's useful to introduce a nondimensional set of units so we can apply the results to many systems. Choose the unit of mass so the sum of the primary and secondary masses is unity:

$$m_1 = 1 - \mu^* \text{ and } m_2 = \mu^*$$

μ^* is often called the *mass ratio* of the restricted three-body problem—it's always less than $\frac{1}{2}$. The subscripts distinguish it from the gravitational parameter, μ . If we use unity for the distance between the two masses, the larger and smaller masses are at distances from the origin μ^* and $1 - \mu^*$ respectively. We can now write the relative vectors from the primary and secondary masses as

$$r_1 = \sqrt{(x - \mu^*)^2 + y^2 + z^2}$$

$$r_2 = \sqrt{(x - \mu^* + 1)^2 + y^2 + z^2}$$

With the angular velocity ($\omega_S = 1$), Jacobi's integral is now

$$v^2 = x^2 + y^2 + \frac{2(1 - \mu^*)}{r_1} + \frac{2\mu^*}{r_2} - C + \mu^*(1 - \mu^*) \quad (2-34)$$

Eq. (2-33) or Eq. (2-34) is called the *Jacobian integral* for the restricted three-body problem. Most analyses use Eq. (2-34) as their starting point. Next we solve the equations of motion. We begin with Eq. (2-32) and simply do the partial differentiation. Using the mass ratio instead of μ_1 and μ_2 gives us the rotating, nondimensional equations.

$$\ddot{x} - 2\dot{y} - x = -\frac{(1 - \mu^*)(x - \mu^*)}{r_1^3} - \frac{\mu^*(x + 1 - \mu^*)}{r_2^3}$$

$$\ddot{y} + 2\dot{x} - y = -\frac{(1 - \mu^*)y}{r_1^3} - \frac{\mu^*y}{r_2^3}$$

$$\ddot{z} = -\frac{(1 - \mu^*)z}{r_1^3} - \frac{\mu^*z}{r_2^3} \quad (2-35)$$

Some applications require the barycentric equations of motion—those in a nonrotating frame. Szebehely (1967, 21) shows the formulas.

$$\begin{aligned}
 \ddot{x} &= -\frac{\mu_1(x - r_{B1}\cos(\omega_S t))}{r_1^3} - \frac{\mu_2(x + r_{B2}\cos(\omega_S t))}{r_2^3} \\
 \ddot{y} &= -\frac{\mu_1(y - r_{B1}\sin(\omega_S t))}{r_1^3} - \frac{\mu_2(y + r_{B2}\sin(\omega_S t))}{r_2^3} \\
 \ddot{z} &= -\frac{\mu_1 z}{r_1^3} - \frac{\mu_2 z}{r_2^3} \\
 r_1 &= \sqrt{(x - r_{B1}\cos(\omega_S t))^2 + (y - r_{B1}\sin(\omega_S t))^2 + z^2} \\
 r_2 &= \sqrt{(x - r_{B2}\cos(\omega_S t))^2 + (y - r_{B2}\sin(\omega_S t))^2 + z^2}
 \end{aligned} \tag{2-36}$$

Three general areas are of interest: special cases called *equilibrium solutions*, regions of motion which are related to the Jacobian constants, and stability of motions from analyzing the *phase space*. To examine the equilibrium positions, we first set the velocity and acceleration equal to zero in Eq. (2-35). The equilibrium positions, or the **Lagrange points**, are shown in Fig. 2-9. It's important to remember that these points exist only in the rotating coordinate system. Immediately, z must be zero, or an out-of-plane acceleration induces oscillatory motion that doesn't represent an equilibrium state.

$$x - \frac{(1 - \mu^*) (x - \mu^*)}{r_1^3} - \frac{\mu^* (x + 1 - \mu^*)}{r_2^3} = 0 \quad y \left(1 - \frac{1 - \mu^*}{r_1^3} - \frac{\mu^*}{r_2^3} \right) = 0 \tag{2-37}$$

If we set $y = 0$, we can solve x using a quintic equation to yield the three *colinear* Lagrange points, L_1 , L_2 , and L_3 . These points lie on the x_S axis and were first obtained by Euler. In reality, three quintics are required (Szebehely, 1967, 135–138):

$$\begin{aligned}
 x^5 + (3 - \mu^*) x^4 + (3 - 2\mu^*) x^3 - \mu^* x^2 - 2\mu^* x - \mu^* &= 0 \\
 x^5 - (3 - \mu^*) x^4 + (3 - 2\mu^*) x^3 - \mu_{2S} x^2 + 2\mu^* x - \mu^* &= 0 \\
 x^5 + (2 + \mu^*) x^4 + (1 + 2\mu^*) x^3 - (1 - \mu^*) x^2 - 2(1 - \mu^*) x - (1 - \mu^*) &= 0
 \end{aligned}$$

Eq. (2-37) is also used to find the two *triangular* equilibrium points, L_4 and L_5 , which lie off the x_S axis. They're much simpler to find because $r_1 = r_2 = 1$. Lagrange discovered these *equilateral* points (with x , y coordinates):

$$L_4 = \mu^* - \frac{1}{2}, \frac{\sqrt{3}}{2} \quad L_5 = \mu^* - \frac{1}{2}, -\frac{\sqrt{3}}{2}$$

Fig. 2-9 shows the Lagrange points for the Earth-Moon system, $\mu^* = 0.01214$. Another common depiction is for the Sun-Jupiter system, in which $\mu^* = 0.00095$. This value is

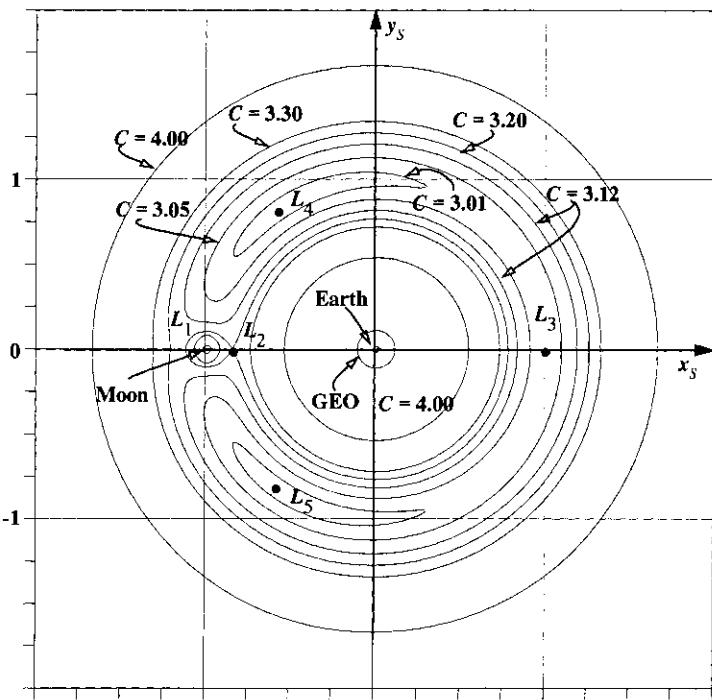


Figure 2-9. Regions of x - y Motion for the Earth-Moon System, $\mu^* = 0.012\ 14$. This representation of motion in the Earth-Moon system shows the Lagrange points, as well as the positions of the Earth and Moon. A geosynchronous orbit is shown about the Earth to give a sense of scale. Notice the significant differences from Fig. 2-10 due to the smaller mass ratio of the Earth-Moon system. Motion across curves of higher C values is possible only with additional thrust.

especially useful for describing the motion of the Trojan asteroids because they occupy the L_4 and L_5 positions of the Sun-Jupiter system.

We obtain zero-velocity curves using the Jacobian integral [Eq. (2-34)]. Setting $v^2 = 0$, we find $2U = C$ where U is the terms in Eq. (2-34). Note that U is a function of x , y , r_1 , and r_2 . Each zero-velocity curve separates distinct regions of motion. For a given value of C , we determine the locus of points that satisfy $2U = C$. Essentially, we can find limiting values of C by substituting the Lagrange points into Eq. (2-34). The goal of determining the values of the Jacobian constant permits us to examine the regions of motion about either the primary or secondary mass, or both. Figure 2-10 shows plots for a system in which $\mu^* = 0.3$. Notice how different the shape of the curve is compared to the case of $\mu^* = 0.012\ 14$ in Fig. 2-9. Values this high are often encountered in astronomy with the study of binary stars, although they're not very useful for planets in our solar system.

Figure 2-11 shows a vertical “slice” (the x - z plane) for zero-velocity curves associated with the Earth-Moon system. This depiction isn't quite as common, but it does illustrate

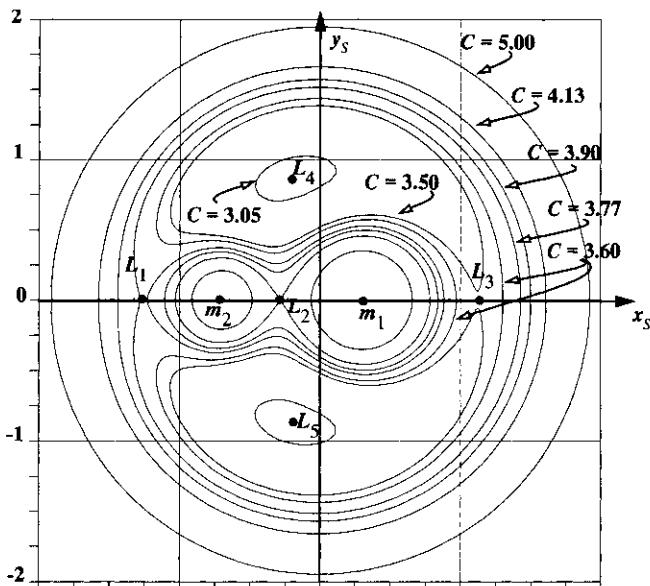


Figure 2-10. Regions of x - y Motion for a System with $\mu^* = 0.3$. This view of the motion in the x - y plane shows the five Lagrange points. This large mass ratio applies best to binary star systems.

the restricted nature of regions of motion outside the horizontal plane. Three-dimensional figures are possible by combining Fig. 2-9 and Fig. 2-11. Consult Lundberg et al. (1985) for additional information.

Stability of the motion has always sparked considerable interest. When a satellite is near a stable point, it will remain there. The limiting values of C give us boundaries which mark transitions between stable and unstable regions. Much of the analysis in this area deals with nonlinear dynamics and *phase space* or *manifold* techniques. Consult Szebehely (1967) for details. It's important to recognize that the colinear Lagrange points L_1 , L_2 , and L_3 are unstable, whereas the equilateral points L_4 and L_5 are stable. A satellite can't cross a zero-velocity curve unless we apply an external force such as thrusting or perturbations from a fourth body! Close examination of Fig. 2-9 and Fig. 2-11 reveals that, as we select C values that intersect the unstable Lagrange points (L_2), it's possible to "cross over" to the other section of zero velocity. This fact was actually used during the Apollo program—the trips to the Moon were so-called *free-return trajectories*, in which the spacecraft could pass from one region to another with a small thrusting force.

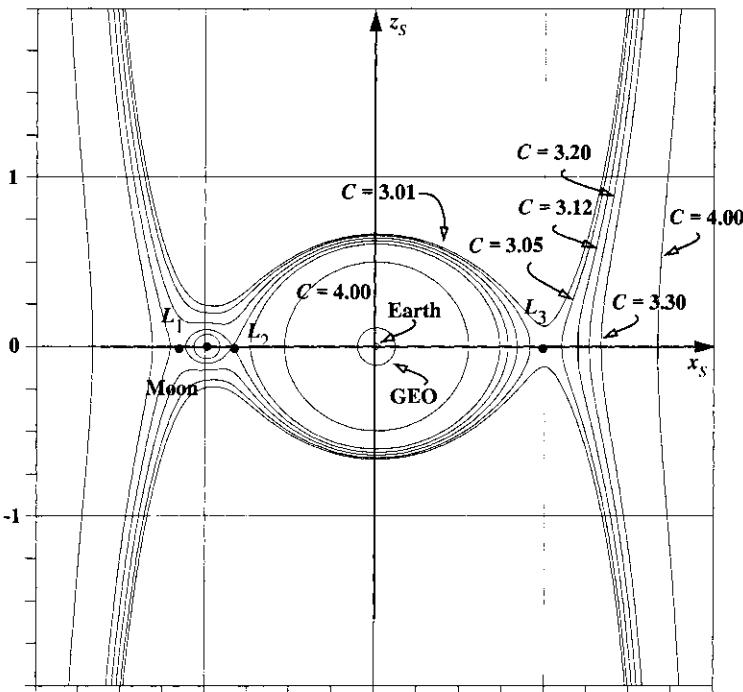


Figure 2-11. Regions of x - z Motion for the Earth-Moon System, $\mu^* = 0.012\ 14$. This equatorial view shows the zero-velocity curves for the Earth-Moon system. The Earth and geosynchronous orbit are shown for scaling purposes. Notice how restricted out-of-plane motion becomes near the Earth.

2.4 Satellite State Representations

We need six quantities to define what we'll call the *state* of a satellite in space. These quantities may take on many equivalent forms. Whatever the form, we call the collection either a *state vector*, X , usually associated with position and velocity vectors, or an *element set*, *elset*, typically used with scalar magnitude and angular representations of the orbit called *orbital elements*. Either set of quantities completely specify the two-body orbit and provide a complete set of initial conditions for solving an initial value problem class of differential equations. Time is always associated with a state vector and is often considered a seventh component. We'll explore concepts about state spaces in Chap. 9.

State vectors are referenced to a particular coordinate frame. This is also true for element sets. Element sets have many forms due to the variety of orbital elements with which to form the vector. The most common element sets you'll encounter are the *classical orbital elements*, often called *Keplerian elements*, *two-body elements*, or *osculating elements*. We'll explore the terminology behind osculating elements when we introduce perturbations in Sec. 8.2. Several other element sets have been developed for convenience or

to avoid the difficulties the classical orbital elements suffer for certain orbital geometries: 2-line, equinoctial, Delaunay, and Poincaré.

2.4.1 Classical Orbital Elements (Keplerian)

Figure 2-12 shows some of the classical orbital elements. Although the semimajor axis is listed as the first element, other parameters may be necessary to use in certain cases: the semiparameter, radii of apoapsis and periapsis, and the mean motion. All these variations help describe the *size* of the orbit. The *semimajor axis*, a , has already been introduced in Sec. 1.2. For most orbits, it's easiest to find the semimajor axis from the energy integral [Eq. (2-12)], but we can also use the extreme points of the orbit—the radii of apoapsis and periapsis.

$$a = \left(\frac{2}{r} - \frac{v^2}{\mu} \right)^{-1} = \frac{r_a + r_p}{2} \quad (2-38)$$

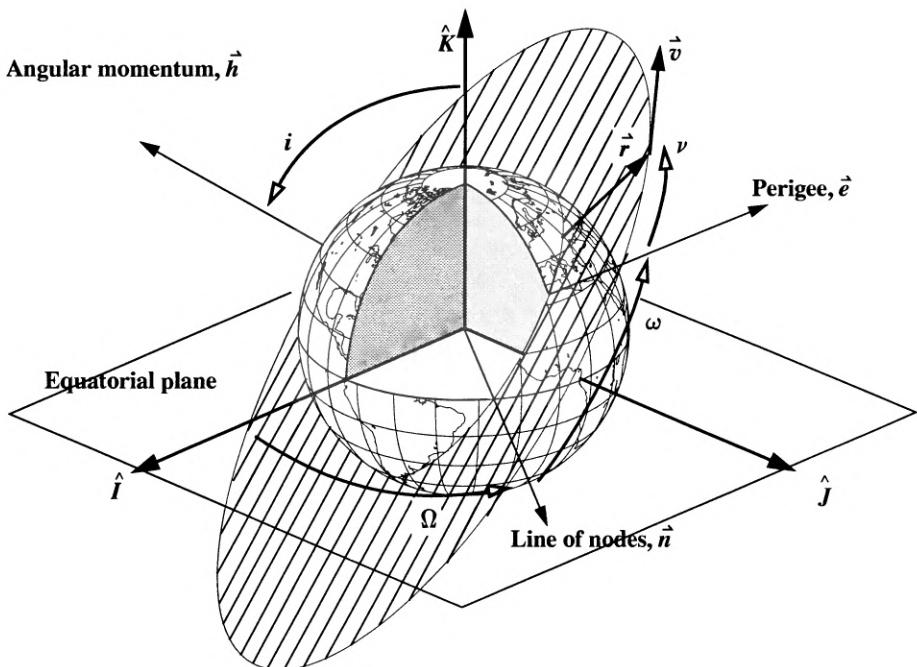


Figure 2-12. Classical Orbital Elements. The six classical orbital elements are the *semimajor axis*, a ; *eccentricity*, e ; *inclination*, i ; *longitude of ascending node*, Ω —often referred to as simply the *node*; *argument of perigee*, ω ; and *true anomaly*, ν . I haven't shown scale and shape elements (a and e) because I've introduced them in Chap. 1.

As discussed before, the semimajor axis is infinite for parabolic orbits, and Eq. (2-38) is indeterminate. For this reason, the *semiparameter*, p , describes the size of the conic section by defining the width at the primary focus. Chapter 1 introduced formulas for the semiparameter: Eq. (1-10) and Eq. (1-11).

$$p = \frac{b^2}{a} = a(1 - e^2)$$

The most general form of the semiparameter is Eq. (2-9) because angular momentum is always defined. The semiparameter is useful to derive the extreme points in an orbit. We substitute values of 0° and 180° into the trajectory equation [Eq. (2-14)] to yield

$$\begin{aligned} \nu &= 0^\circ & r_p &= \frac{p}{1 + e \cos(\nu)} = \frac{p}{1 + e} \\ \nu &= 180^\circ & r_a &= \frac{p}{1 + e \cos(\nu)} = \frac{p}{1 - e} \end{aligned}$$

For *most* orbits, we can simplify to get useful results:

$$\begin{aligned} r_p &= \frac{a(1 - e^2)}{1 + e} = a(1 - e) \\ r_a &= \frac{a(1 - e^2)}{1 - e} = a(1 + e) \end{aligned} \tag{2-39}$$

Many of the preceding formulas are indeterminate for a parabola because the semimajor axis is infinite and $e = 1$. Obviously, a parabola has a value for the semiparameter and radius of perigee. From Eq. (2-9), we can find the semiparameter. If we substitute the result into the trajectory equation [Eq. (2-14)], the perigee radius may also determine the semiparameter.

$$P_{\text{parabola}} = 2r_p \tag{2-40}$$

We often want to use an element that describes the satellite's average angular rate of motion over one orbit, or the *mean motion*, n . The mean motion is defined as

$$n = \sqrt{\frac{\mu}{a^3}} \tag{2-41}$$

The second classical orbital element is the *eccentricity*, e , referring to the *shape* of the orbit. The eccentricity is the magnitude of a vector quantity. The derivation of the trajectory equation contains the starting point for deriving the eccentricity vector. Begin with Eq. (2-13) to introduce the vector constant of integration, \hat{B}/μ , or the eccentricity vector

$$\hat{e} = \frac{\hat{v} \times \hat{h}}{\mu} - \frac{\hat{r}}{r}$$

Replace the angular momentum with the cross product,

$$\hat{\mathbf{e}} = \frac{\hat{\mathbf{v}} \times (\hat{\mathbf{r}} \times \hat{\mathbf{v}})}{\mu} - \frac{\hat{\mathbf{r}}}{r}$$

Then, expand the vector triple cross product to get

$$\hat{\mathbf{e}} = \frac{(\hat{\mathbf{v}} \cdot \hat{\mathbf{v}}) \hat{\mathbf{r}} - (\hat{\mathbf{r}} \cdot \hat{\mathbf{v}}) \hat{\mathbf{v}}}{\mu} - \frac{\hat{\mathbf{r}}}{r}$$

Simplify the final equation:

$$\hat{\mathbf{e}} = \frac{v^2 \hat{\mathbf{r}} - (\hat{\mathbf{r}} \cdot \hat{\mathbf{v}}) \hat{\mathbf{v}}}{\mu} - \frac{\hat{\mathbf{r}}}{r}$$

Combine to get

$$\hat{\mathbf{e}} = \frac{\left(v^2 - \frac{\mu}{r}\right) \hat{\mathbf{r}} - (\hat{\mathbf{r}} \cdot \hat{\mathbf{v}}) \hat{\mathbf{v}}}{\mu} \quad (2-42)$$

The eccentricity vector always points to periapsis. We can show $\hat{\mathbf{e}}$ points to perigee mathematically by substituting position and velocity vectors in the *PQW* system into Eq. (2-42) and simplifying (Sellers, 1994, 605–607). In orbits for which periapsis doesn't exist (circles), the eccentricity vector is zero.

We can also calculate the magnitude of eccentricity from Eq. (2-42) or, with many fewer calculations, by using $e = c/a$ (as shown in Sec. 1.2). Notice, for parabolic orbits, calculations are impossible because the semimajor axis is ∞ and $c = \infty$. We can find a more general formula using the semiparameter equation and solving it for the eccentricity:

$$e = \sqrt{1 - \frac{p}{a}} \quad (2-43)$$

Knowing that $h = \sqrt{\mu p}$ and $\xi = -\mu/2a$, let's substitute appropriately into these formulas to get an equation defined for all conic sections:

$$e = \sqrt{1 + \frac{2\xi h^2}{\mu^2}} \quad (2-44)$$

We can categorize the ellipse in terms of the orbital elements, semimajor axis and eccentricity. Ellipses may be large or small; thus, no limit is placed on the semimajor axis (other than its being larger than zero and less than ∞) because these values define the bounds of different conic sections. The value of eccentricity for ellipses is bounded between 0.0 and 1.0. For hand calculations, use 0 and 1. For computer applications, be very cautious as to what represents zero. Most operational programs mention circular orbits, but the code is vague as to what constitutes "0.0" (0.000 001, 0.0002, etc.). In addition to specifying your definition for a circular orbit, remember that orbits close to the cir-

cular or parabolic boundary may have certain angles that change rapidly. For applications in which this is a problem, it's wise to choose an alternate set of orbital elements, such as equinoctial elements. We'll discuss these applications later.

The *inclination*, i , refers to the tilt of the orbit plane. This angle is measured from the unit vector \hat{K} to the angular momentum vector \hat{h} . It ranges from 0° to 180° . Inclinations of 0° and 180° are *equatorial orbits*, whereas all others are *inclined orbits*. In addition, the direction of measuring the angle means that orbits from 0° to 90° will travel "with the Earth"; those from 90° to 180° will oppose the Earth's rotation. The first are therefore *direct* or *prograde orbits*, and the latter are *retrograde orbits*. One final distinction involves orbits with inclinations of exactly 90° . These orbits travel over the poles, so they're called *polar orbits*. Mathematically,

$$\cos(i) = \frac{\hat{K} \cdot \hat{h}}{|\hat{K}| |\hat{h}|} \quad (2-45)$$

Because the arccosine function will always return an angle between 0° and 180° , no further checks are necessary for computer applications.

The *longitude of the ascending node*, Ω , the *right ascension of the node*, or simply the *node*, is the angle in the equatorial plane measured positively from the \hat{I} unit vector to the location of the ascending node. The *ascending node* is the point on the equatorial plane at which the satellite crosses the equator from south to north. All inclined orbits also have a *descending node*, at which the satellite crosses from north to south across the equatorial plane. The line segment connecting both nodes defines a *line of nodes*. There is also a vector associated with the ascending node, \hat{n} . Don't confuse the magnitude of \hat{n} with the mean motion, n . They aren't the same! We find the node vector as

$$\hat{n} = \hat{K} \times \hat{h} \quad (2-46)$$

Equatorial orbits have no nodes, and thus the longitude of the ascending node is undefined. The direction of measurement results from a right-handed coordinate convention—it's a counter-clockwise motion when viewed from above the North Pole. Node values may range from 0° to 360° because we must account for all locations in the I - J plane. Again, we get the formula from the result of a dot product:

$$\cos(\Omega) = \frac{\hat{I} \cdot \hat{n}}{|\hat{I}| |\hat{n}|} \quad (2-47)$$

$$\text{IF } (n_J < 0) \text{ THEN } \Omega = 360^\circ - \Omega$$

But this time we must check quadrants to ensure we have the correct answer. From Fig. 2-12, notice the node vector along the intersection of the equatorial plane and the orbital plane. If the " J " component of \hat{n} is less than zero, the Ω must be adjusted because it lies between 180° and 360° .

An obvious question arises for computer applications and for using standard functions such as the FORTRAN ATAN2 routine, which resolves the quadrant "automatically." We can use any computer language to capture the basic function of this type of routine because we simply have to input both the sine and cosine arguments so we can use the signs to determine the correct quadrant during processing. These functions are appropriate in many cases and are the preferred method described in several places throughout the book. Still, when calculating the orbital elements, it's usually easier (and faster) to use the traditional angle method. This method avoids the increased calculations required by the operations embedded in the ATAN2 function. If you wish, you can derive sine expressions for each angle to complete the alternate process.

The *argument of perigee*, ω , measured from the ascending node, locates the closest point of the orbit (periapsis). Again, this quantity varies from 0° to 360° and is obtained by extending a dot product:

$$\cos(\omega) = \frac{\hat{n} \cdot \hat{e}}{|\hat{n}| |\hat{e}|} \quad (2-48)$$

$$\text{IF } (e_K < 0) \text{ THEN } \omega = 360^\circ - \omega$$

This angle is undefined for perfectly circular ($\hat{e} = 0$) or equatorial ($\hat{n} = 0$) orbits because periapsis doesn't exist for these special cases. We must check to obtain the correct value. This check depends on the " K " component of the eccentricity vector, which points to periapsis. If negative, the location of perigee is below the equator, and therefore the argument of perigee should be between 180° and 360° .

The final classical orbital element is the *true anomaly*, which I've already briefly introduced. The *true anomaly*, ν , determines the satellite's current position relative to the location of periapsis. Again, using the dot product is convenient because the angle is between the eccentricity and position vectors:

$$\cos(\nu) = \frac{\hat{e} \cdot \hat{r}}{|\hat{e}| |\hat{r}|} \quad (2-49)$$

$$\text{IF } (\hat{r} \cdot \hat{v} < 0), \text{ THEN } \nu = 360^\circ - \nu$$

True anomaly varies from 0° to 360° , requires a quadrant check, and is undefined for perfectly circular orbits. This check is perhaps not as obvious as the previous checks because you're actually trying to determine if the satellite is between periapsis and apoapsis or vice versa. From previous discussions about the flight-path angle, remember that the flight-path angle is always positive when going from periapsis to apoapsis and is negative on the return trip. A dot product of the position and velocity vectors is the solution because it finds the projected component of the two vectors; it will also be positive going away from periapsis and negative going towards periapsis.

Special Cases

The preceding variables can define the orbit and represent its location at any time except under certain geometric conditions. Alternate definitions are required for perfectly circular and equatorial orbits. We never see perfectly circular and equatorial orbits, but orbits close to these limits may cause problems with computer solutions. The existence of these problems forces us to define three additional variables to account for each special orbit: elliptical equatorial ($\tilde{\omega}_{true}$), circular inclined (ω), and circular equatorial (λ_{true}). The true subscript reminds us that these elements are valid for any orbit. Astronomers use two similar angles ($\tilde{\omega}$ and λ_{true}) for real-world cases which approach the theoretical circular and equatorial orbits.

An elliptical equatorial orbit has no ascending node. An example of this type of orbit is the Earth's orbit about the Sun because it has a small eccentricity and is therefore elliptical. However, it's defined to have a zero inclination because the ecliptic is used as the fundamental plane to measure planetary inclinations. Remember the ecliptic is the mean plane of the Earth's orbit about the Sun, and the instantaneous orbital plane is always close to the mean plane. Thus, the inclination value is very small. How do we find the ascending node and the argument of periapsis? Our solution is the *true longitude of periapsis*, $\tilde{\omega}_{true}$, which combines Ω and ω to remove the ambiguity and is the angle measured eastward from the vernal equinox (\hat{I} unit vector in the geocentric coordinate system) to the eccentricity vector. We can find the angle as

$$\cos(\tilde{\omega}_{true}) = \frac{\hat{I} \cdot \hat{e}}{|\hat{I}| |\hat{e}|} \quad \text{Elliptical Equatorial} \quad (2-50)$$

$$\text{IF } (e_J < 0) \text{ THEN } \tilde{\omega}_{true} = 360^\circ - \tilde{\omega}_{true}$$

Because $\tilde{\omega}_{true}$ varies from 0° to 360° , we must check quadrants. The check determines whether or not the " J " component of the eccentricity vector is negative, meaning the angle should range from 180° to 360° . An approximate relation for $\tilde{\omega}_{true}$ exists when the inclination of the orbit is small ($\tilde{\omega}_{true} \approx \Omega + \omega$).

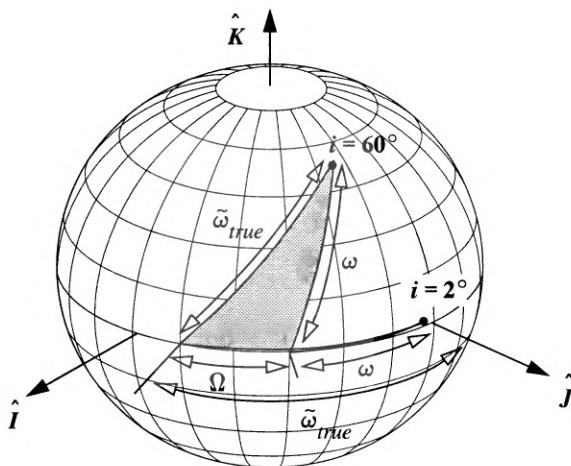
Students of astronomy should recognize a similar quantity by its more common symbol, $\tilde{\omega}$. This element has the similar name, *longitude of periapsis*, and refers to the location of periapsis from the primary axis:

$$\tilde{\omega} = \Omega + \omega \quad (2-51)$$

Notice that $\tilde{\omega}$ is *equal* to this result and ends up being a summation of two angles in different planes if the inclination is significant. In this case, Eq. (2-50) *won't* equal $\tilde{\omega}$ because it's designed to determine the angle between the two positions and not the sum of two angles in different planes.

Consider Fig. 2-13. For orbits which are very nearly equatorial (direct or retrograde), the relation holds true and will yield reasonably good results. For the inclined case in Fig. 2-13, we must use rules from spherical trigonometry to determine the validity. Notice the

large discrepancy in the values. You can see how quickly the results degrade as the inclination departs from 0° or 180° . As a general rule, for cases where the inclination is within about 5° of being equatorial, $\tilde{\omega}_{true}$ will be fairly close to $\tilde{\omega}$. The rationale for the true subscript should now be apparent. We'll use it in this book to minimize confusion about the angles.



$$i = 2^\circ, \quad e = 0.000\ 01$$

$$\Omega = 30^\circ, \omega = 40^\circ$$

$$\tilde{\omega}_{true} = 69.988\ 27^\circ$$

$$i = 60^\circ, \quad e = 0.000\ 01$$

$$\Omega = 30^\circ, \omega = 40^\circ$$

$$\tilde{\omega}_{true} = 59.820\ 08^\circ$$

Figure 2-13. Representation of Longitude of Periapsis. Notice the effect inclination has on the values for the longitude of periapsis. The values of $(\Omega + \omega)$ and $\tilde{\omega}_{true}$ are close when the inclination is small but very different as the inclination increases.

The second special orbit is the circular inclined orbit because there is no periapsis from which to measure ω and ν . The **argument of latitude**, u , is the angle measured between the ascending node and the satellite's position vector. An example of this type of orbit is provided by a moon of Neptune. Triton is inclined 157.345° to the equator of Neptune and has an eccentricity of about $0.000\ 016$ (*Astronomical Almanac*, 1992, F2). Thus,

$$\cos(u) = \frac{\hat{n} \cdot \hat{r}}{|\hat{n}| |\hat{r}|} \quad \text{Circular Inclined} \quad (2-52)$$

$$\text{IF } (r_K < 0) \text{ THEN } u = 360^\circ - u$$

A relation that's always valid exists between classical orbital elements and the argument of latitude:

$$u = \omega + \nu \quad (2-53)$$

Because all these angles are in the same plane, the relation is always true. However, be aware that as the orbit's inclination approaches 0° or 180° , the node vector approaches zero and Eq. (2-52) becomes indeterminate.

A variant is sometimes used called the ***mean argument of latitude***, u_M . It determines location based on the ***mean anomaly***, M , which we'll discuss in Sec. 4.2:

$$u_M = \omega + M \quad (2-54)$$

The *mean* subscript is intended to remind the user of the presence of mean anomaly. Later, we'll show that $M = \nu$ only for circular orbits. Thus, for circular orbits, the argument of latitude and the *mean* argument of latitude are the same. In all other cases, the two quantities are different.

Finally, for circular equatorial orbits, we define the ***true longitude***, λ_{true} , as an angle measured eastward from the I -axis to the position of the satellite. This angle is the same as the right ascension of the satellite that's sometimes used. Don't confuse this concept with terrestrial longitude, which is measured from Greenwich. Again, we're using the *true* subscript to avoid confusion:

$$\cos(\lambda_{true}) = \frac{\hat{I} \cdot \hat{r}}{|\hat{I}| |\hat{r}|} \quad \text{Circular Equatorial} \quad (2-55)$$

$$\text{IF } (r_J < 0) \text{ THEN } \lambda_{true} = 360^\circ - \lambda_{true}$$

The moon of Uranus, Oberon, provides an example of this type of orbit. It has an inclination of about 0.1° and an eccentricity of 0.0008 (*Astronomical Almanac*, 1992, F2). An approximate relation for λ_{true} is $\lambda_{true} \approx \Omega + \omega + \nu$ because Ω and $(\omega + \nu)$ are in different planes when an inclined orbit is involved.

As with the longitude of periapsis, astronomers tend to use a variation for this angle. The ***mean longitude***, λ_M , is the location of the satellite from periapsis; it's analogous to the true anomaly but not equal to it. The expression uses the mean anomaly, M , which we'll discuss later, and is measured in two separate planes if the orbit has some inclination. Thus,

$$\lambda_M = \Omega + \omega + M = \tilde{\omega} + M \quad (2-56)$$

Flight-path Angle

Although we usually don't consider the flight-path angle as a classical orbital element, it's very useful for specifying a satellite's orientation or attitude. This orientation is crucial to determining the effective cross-sectional area required for both drag and solar-radiation perturbations, which may be very important depending on the satellite's mission. Because we measure the flight-path angle from the local horizontal to the velocity vector, it's always zero for circular motion,

$$\phi_{fpa} = 0^\circ \quad \text{circle}$$

For general applications, it's possible to derive an expression for the flight-path angle. Referring to Fig. 2-4, notice that

$$\cos(\phi_{fpa}) = \frac{r\dot{v}}{v}$$

But we've already developed several alternate formulas. In addition, if we use the velocity formula for an ellipse [Eq. (2-21) and Eq. (2-8)],

$$\cos(\phi_{fpa}) = \frac{\frac{h}{r}}{\sqrt{\frac{2\mu}{r} - \frac{\mu}{a}}}$$

Because $h = \sqrt{\mu r}$, using Eq. (1-11) permits

$$\cos(\phi_{fpa}) = \frac{\sqrt{a^2(1-e^2)}}{\sqrt{r(2a-r)}} \quad (2-57)$$

We can similarly develop an expression for the sine of the flight-path angle because, from Fig. 2-4,

$$\sin(\phi_{fpa}) = \frac{\dot{r}}{v} = \frac{\dot{r}}{\sqrt{\frac{2\mu}{r} - \frac{\mu}{a}}}$$

Now substitute \dot{r} from Eq. (2-15):

$$\sin(\phi_{fpa}) = \frac{\frac{r\dot{v}e \sin(v)}{1+e \cos(v)}}{\sqrt{\frac{2\mu}{r} - \frac{\mu}{a}}}$$

Notice the dependence on the true anomaly, which makes developing a relation for the sine easier. First, use Eq. (2-8) to replace the derivative of the true anomaly in the numerator and find

$$\sin(\phi_{fpa}) = \frac{h \left(\frac{e \sin(v)}{1+e \cos(v)} \right)}{\sqrt{\frac{2a-r}{\mu \frac{ra}{r}}}}$$

Substitute definitions for the trajectory equation, Eq. (2-14) and Eq. (2-9):

$$\sin(\phi_{fpa}) = e \sin(\nu) \sqrt{\left(\frac{1}{p}\right)\left(\frac{ra}{2a - r}\right)}$$

Next, substitute the definition of the semiparameter, Eq. (1-11):

$$\sin(\phi_{fpa}) = e \sin(\nu) \sqrt{\frac{r}{(1 - e^2)(2a - r)}}$$

Substitute the trajectory equation again and simplify to get the final result:

$$\sin(\phi_{fpa}) = \frac{e \sin(\nu)}{\sqrt{1 + 2e \cos(\nu) + e^2}} \quad (2-58)$$

To determine the cosine expression in terms of true anomaly, use Eq. (2-58) and Eq. (C-8). Recognize that the numerator of the cosine squared term must equal $(1 + 2e \cos(\nu) + e^2 - e^2 \sin^2(\nu))$. Using the Pythagorean relation again on the sine term, reduce the equation to

$$\cos(\phi_{fpa}) = \frac{1 + e \cos(\nu)}{\sqrt{1 + 2e \cos(\nu) + e^2}} \quad (2-59)$$

We use both sine and cosine expressions to aid computer programs and employ ATAN2 to help find the correct quadrant. The expressions in Eq. (2-58) and Eq. (2-59) permit a graphical analysis of the flight-path angle. Figure 2-14 shows various values for the flight-path angle. Notice the relatively small values for orbits with an eccentricity of 0.5 or less.

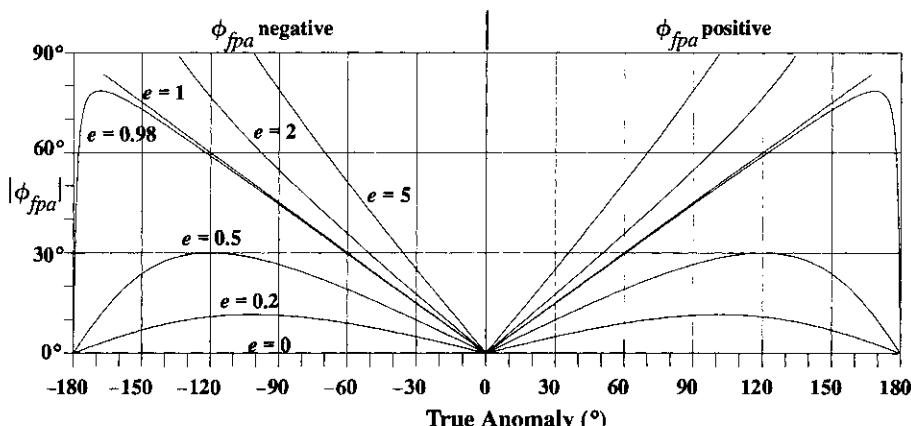


Figure 2-14. Values for the Flight-path Angle. Notice how the behavior of the flight-path angle changes dramatically as the eccentricity increases. I've shown the absolute value for clarity. For parabolic and hyperbolic orbits, the range of true anomaly is limited. See Sec. 4.2.3 for details.

2.4.2 2-line Element Sets

The classical orbital elements are widely used throughout the scientific community, but they don't apply universally. For instance, the military uses a *2-line element set* (*2-line elset*), which is similar but not identical to the classical elements. Because these elsets are available to the general public through NASA/Goddard, we'll describe their formation. Other sources for the data include the U.S. Air Force's Space Command, the Celestial Bulletin Board Service at the Air Force Institute of Technology in Ohio, and the Canadian Space Society. It's important to understand how these element sets are formed regardless of their source because we can get accurate results only if we use standard procedures and elements based on the same assumptions.

The elements of the 2-line elset are shown in Eq. (2-60). Although ten values are listed, the first six represent the independent quantities required for calculations. The remaining variables (mean motion rate, mean motion acceleration, and B^* , a drag-like parameter) are required to describe the effect of perturbations on satellite motion. We'll discuss them in Sec. 8.7.1. The mean anomaly, M , is used instead of the true anomaly; we'll discuss this angle in Sec. 4.2. The time is based on *UTC*.

$$\begin{array}{ccccccc} \bar{n} & = & \sqrt{\frac{\mu}{a^3}} & e & i & \Omega & \omega \\ \dot{\bar{n}} & = & \frac{\dot{n}}{2} & \ddot{n} & \frac{6}{6} & B^* & = \frac{1}{2} \frac{c_D A}{m} \rho_o \\ & & & & & & UTC \end{array} \quad (2-60)$$

Here, ρ_o is the atmospheric density at perigee of the orbit (assumed to be 2.461×10^{-5} kg/m²/s^{1/2}), and n has units of rev/day. The bars on the mean motion and semimajor axis denote Kozai *mean* values, which we'll discuss in Chap. 8. In fact, all the values are means. We must recover the true *ballistic coefficient*, BC , from the B^* term in the 2-line element set. From Hoots and Roehrich (1980, 56),

$$\frac{R_\oplus \rho_o}{2B^*} = BC$$

With an Earth radius of 6378.135 km defined in Hoots and Roehrich (1980), a constant conversion is

$$BC = \frac{1}{12.741\,621 B^*} \frac{\text{kg}}{\text{m}^2} \quad (2-61)$$

Be aware that the value of B^* is *always* modified. It's really an arbitrary free parameter in differential correction. Chapter 9 will introduce how to estimate a drag parameter. The estimated value of B^* may be completely unrelated to drag effects in the presence of satellite maneuvers, significant solar pressure and atmospheric perturbations, large third-body effects from the Sun or Moon, or large deflections caused by mismodeling of the Earth's gravitational field. B^* can even appear as a negative number! For highly accurate studies, we need the *correct BC*.

Let's examine the format in which these elements arrive—it's rather cryptic. The following is an example of a 2-line elset. The two numbered lines are found commonly on bulletin boards and in messages.

Dataset for the 2-line example

```
1 16609U 86017A 93352.53502934 .00007889 00000 0 10529-3      34
2 16609   51.6190 13.3340 0005770 102.5680 257.5950 15.59114070 44786
```

Modern programmers gasp at the poor readability and inherent high maintenance created by so little internal file documentation. But we should note that the format for elsets originated decades ago, when good documentation and transfers of large files over communication and computer systems didn't exist. Fortunately, these limitations are long gone, but the format remains as a historical legacy. Figure 2-15 shows a translation key.

By applying the translation to the dataset, we can get a usable result:

Epoch December 18, 1993, 12 h 50 min 26.5350 s UTC

$$\bar{n} = 15.591\ 140\ 70 \text{ rev / day} \Rightarrow \bar{a} = 1.061\ 180\ 87 \text{ ER} = 6768.357 \text{ km}$$

$$\frac{\dot{n}}{2} = 7.889 \times 10^{-5} \frac{\text{rev}}{\text{day}^2} \quad \frac{\ddot{n}}{6} = 0.0 \frac{\text{rev}}{\text{day}^3}$$

$$B^* = 1.0529 \times 10^{-4}$$

$$e = 0.000\ 577\ 0$$

$$i = 51.6190^\circ$$

$$\Omega = 13.3340^\circ$$

$$\omega = 102.5680^\circ$$

$$M = 257.5950^\circ$$

We find the estimated ballistic coefficient from Eq. (2-61) and $BC = 745.397\ 95 \text{ kg/m}^2$.

Many people often incorrectly assume these element sets have no errors. In fact, all element sets contain errors based on their mathematical formation. Section 8.9.1 and Fig. 8-15 suggest that errors in the formation of element sets can be on the order of a kilometer or more. For now, let's assume that the data in the element set is perfectly accurate and examine what's required to maintain this precision.

All highly accurate propagation techniques transmit elements using as many digits as possible to minimize unnecessary errors—we want to transmit with the same precision we use to determine the element set. Unfortunately, formats like the 2-line element set introduce uncertainty by arbitrarily limiting the number of decimal digits for each parameter. Because the element set has finite digits for each variable, we can determine the maximum accuracy available at the epoch through the transmission format. Recognize that, with

Satellite Number	International Designator	Epoch		$\frac{i}{2}$	$\frac{\pi}{6}$	BStar	Element Number	
116609U	85017A	YY	PP	000000000000	S.	000007889	S	10529-3
		93352	53502934	000007889	0000000	0	34	
	Inclination	Right Ascension of node	Eccentricity	Argument of perigee	Mean Anomaly	Mean Motion	Epoch Rev	
216609	51	6190	13.3340	0005770	102.5680	257.5950	15.5911407044786	

Figure 2-15. Transmission Format for the 2-line Element Set. This example 2-line element set uses data from the previous example in the text. Note the use of implied decimal points. S is the sign of the values; E is the exponent.

only eight decimal places given in the epoch time value, the epoch above is accurate only to about 0.0004 seconds. We determine this accuracy by assuming the original value is unknown to $\pm 5 \times 10^{-9}$ days, which is about $\pm 4.3 \times 10^{-4}$ seconds.

Now consider a roughly circular orbit with an altitude of 390 km and a velocity of about 7.6 km/s. The satellite travels about 4 m in 0.000 43 s. The eccentricity is given to seven decimal places. This introduces an error at epoch which for a GEO satellite is about 2 m ($r \approx a\Delta e$). The angles provide only four decimal places, which introduce an uncertainty at epoch of about 6 m for LEO satellites and about 35 m for GEO satellites. Remember that these estimates are based solely on the given data and have nothing to do with their mathematical formation and subsequent use. For precise orbit determination, even these errors probably aren't acceptable.

2.4.3 Equinoctial Elements

Let's discuss equinoctial elements as a basis for the perturbation studies we'll analyze later. These elements are also useful to overcome difficulties with the special geometry cases described in Sec. 2.4.1. Note that we use the longitude of periapsis, $\tilde{\omega}$. Remember, this angle is measured in two *different* planes whenever the satellite's orbit is nonequatorial! We use the longitude of periapsis to relieve the mathematical difficulties arising from equatorial orbits. *True* retrograde equatorial orbits ($i = 180^\circ$) are the only cases that cause problems because p_e and q_e are undefined. We solve this by introducing a *retrograde factor*, f_r . It's +1 for direct orbits, and -1 for retrograde orbits. This set is used in virtually all large, high-accuracy programs.

$$\begin{aligned}
 k_e &= e \cos(\omega + f_r \Omega) & h_e &= e \sin(\omega + f_r \Omega) \\
 a && \lambda_M &= M + \omega + f_r \Omega \\
 p_e &= \tan^{f_r} \left(\frac{i}{2} \right) \sin(\Omega) & q_e &= \tan^{f_r} \left(\frac{i}{2} \right) \cos(\Omega)
 \end{aligned} \tag{2-62}$$

Notice that λ_M is the only fast variable. In other words, it's the only variable that changes value rapidly with respect to the orbital period.

2.4.4 Canonical Variables

Canonical variables have special characteristics and relationships which make them useful for astrodynamics. Referred to as *conjugate coordinates* and *generalized momenta* in physics, they appear in the Lagrangian and Hamiltonian forms of dynamical systems.

To give you an idea of their use, consider the equations of motion for two-body motion. We could write the equations as the product of a 6×6 matrix times the 6×1 state vector containing the position and velocity vectors. Using standard elements, the 6×6 matrix contains mainly nonzero terms on and off the diagonal. Using canonical variables, the 6×6 matrix is purely diagonal, which simplifies the construction of perturbation solutions (Chap. 8).

Delaunay Variables

The Delaunay variables are the canonical counterpart to the classical, or Keplerian, orbital elements. As such, they contain singularities for small values of eccentricity and inclination. They are

$$\begin{aligned} M \\ \omega \\ \Omega \\ L_d &= \sqrt{\mu a} \\ h \\ H_d &= \sqrt{\mu a (1 - e^2)} \cos(i) \end{aligned} \tag{2-63}$$

Recognize H_d is really the component of angular momentum, h , in the equatorial plane. Don't confuse the Delaunay variables with fundamental quantities already in use throughout this text. The subscripts are used to distinguish similar variables. These formulas permit some rather elegant perturbation analysis, discussed in Chap. 8. In this book, we'll address these variables mainly with Brouwer's theory concerning treatment of orbital perturbations (Sec. 8.7.3).

Poincaré Variables

As the canonical counterpart of equinoctial elements, Poincaré variables are nonsingular elements and are good for small eccentricities and inclinations. Because they're canonical, this set is especially useful for treating orbit problems with Hamiltonian dynamics. The Poincaré variables are

$$\begin{aligned}
 \lambda_M &= M + \omega + \Omega \\
 g_p &= \sqrt{-2(\sqrt{\mu a}(1 - \sqrt{1 - e^2})) \sin(\omega + \Omega) \cos(\omega + \Omega)} \\
 h_p &= \sqrt{-2(\sqrt{\mu a}(1 - e^2)) (\cos(i) - 1) \sin(\Omega) \cos(\Omega)} \\
 L_p &= \sqrt{\mu a} \\
 G_p &= \sqrt{-2(\sqrt{\mu a}(1 - \sqrt{1 - e^2})) \sin(\omega + \Omega)} \\
 H_p &= \sqrt{-2(\sqrt{\mu a}(1 - e^2)) (\cos(i) - 1) \sin(\Omega)}
 \end{aligned} \tag{2-64}$$

2.5 Application: Orbital Elements from r and v

Converting between position and velocity vectors and orbital elements is one of the most common problems in astrodynamics. Remember, we chose to develop in the *IJK* coordinate system, which we assume as a starting point. We'll refer to this problem as the *ELORB* routine.

Observations of a satellite usually reduce to position and velocity vectors. Although convenient for numerical processing and usually easiest for orbit determination, these vectors aren't terribly descriptive to an analyst trying to determine information about the mission or status of the satellite. If we know the classical orbital elements, our analysis is much easier. Thus, we need to determine the orbital elements from the position and velocity vectors. We've presented the equations to do so, but we must group them to form a complete algorithm and solve the problem.

We need several intermediate vectors to begin. Let's start by finding the angular-momentum vector. This vector is defined for *all* orbits and may always be calculated from the general cross-product equation:

$$\hat{h} = \hat{r} \times \hat{v}$$

Next, find the vector pointing to the node. The formula is shown below; however, the magnitude may be zero, which would indicate the orbit is equatorial.

$$\hat{n} = \hat{K} \times \hat{h}$$

Find the eccentricity vector using Eq. (2-42). Because the initial position and velocity vectors are known, use the vector form for the eccentricity. Recognize that this vector is zero for circular orbits. Computer applications must provide a means to avoid mathematically indeterminate equations, which may result from improper division operations using the eccentricity. The magnitude of this vector provides the orbit's eccentricity.

Find the semimajor axis starting with the specific mechanical energy. The preferred equation uses the magnitudes of the position and velocity vectors because it's valid for all orbits:

$$\xi = \frac{v^2}{2} - \frac{\mu}{r}$$

Many applications attempt to find the semimajor axis. Although there is no immediate problem as long as the orbits are *all* circular or elliptical, the possibility for unintended use or expanded tasks suggests caution. If we really need the semimajor axis for an application, we can use eccentricity to do a pre-filter check to eliminate problems in calculations for parabolic orbits in which the semimajor axis is infinite. Any computer program must include this check, and, of course, the tolerance with which the eccentricity of "1.0" is important. Usually, 1.0×10^{-6} is enough for a tolerance. For other than parabolic cases, we can use Eq. (2-37), or

$$a = -\frac{\mu}{2\xi} = \left(\frac{2}{r} - \frac{v^2}{\mu} \right)^{-1}$$

Likewise, we can find the semiparameter for all orbits using Eq. (2-9).

The final step uses the dot product to determine the remaining classical orbital elements. These equations have all been presented before (Sec. 2.4). We can save a lot of program run time if we drop the unity values and do some of the dot products before running the program. The standard approach would use a routine to determine the angle between two vectors, but this adds many unnecessary calculations. Another benefit from doing each calculation independently is that the quadrant checks are done for each element. It would be difficult, if not impossible, to use standard techniques to resolve quadrants for each angle. I've reduced the following equations and omitted the quadrant-resolution procedures because they remain the same as in the original discussion. Computer programs should save the magnitudes of eccentricity, node, and position because they are used often throughout the process and can save many unnecessary operations. The remaining elements are found from Eqs. (2-45) and (2-47) to (2-49). We can determine whether we must evaluate special cases based on the particular needs of the program.

Implementing ELORB

Several substitutions permit computational efficiency. Using canonical units eliminates four mathematical operations because the gravitational parameter is unity. Also notice we can store the quantity $1/r$ as a temporary variable to eliminate another three to five mathematical operations. It may be advisable to store the quantities $1/n$ and $1/e$ to simplify the angle calculations. We can also be more efficient by ordering conditional tests with the most likely occurrence at the beginning. Usually, elliptical orbits fall into this category because perfect circular orbits are actually a mathematical consequence and are rarely, if ever, seen in real operations.

ELORB contains various possible situations. To fully test the algorithm, we must check *each* orbit type we may encounter. For rigorous applications, we must also test the boundary conditions between orbits, such as the eccentricity in the parabolic case. Although certain applications may not require the robustness of the complete algorithm, it's prudent to

design for all known test cases. The small increase in code size is rarely important to the modern programmer.

ALGORITHM 5: $ELORB (\dot{r}_{IJK}, \dot{v}_{IJK} \Rightarrow p, a, e, i, \Omega, \omega, \nu (u, \lambda_{true}, \tilde{\omega}_{true}))$

$$\hat{h} = \dot{r} \times \dot{v} \quad h = |\hat{h}|$$

$$\hat{n} = \hat{K} \times \hat{h}$$

$$\hat{e} = \frac{\left(v^2 - \frac{\mu}{r}\right)\dot{r} - (\dot{r} \cdot \dot{v})\dot{v}}{\mu} \quad e = |\hat{e}|$$

$$\xi = \frac{v^2}{2} - \frac{\mu}{r}$$

$$\begin{cases} \text{IF } e \neq 1.0 \text{ THEN} \\ \quad a = -\frac{\mu}{2\xi} \\ \quad p = a(1 - e^2) \\ \text{ELSE} \\ \quad p = \frac{h^2}{\mu} \text{ and } a = \infty \end{cases}$$

$$\cos(i) = \frac{h_K}{|\hat{h}|}$$

$$\cos(\Omega) = \frac{n_J}{|\hat{n}|} \quad \text{IF } (n_J < 0) \text{ THEN } \Omega = 360^\circ - \Omega$$

$$\cos(\omega) = \frac{\hat{n} \cdot \hat{e}}{|\hat{n}| |\hat{e}|} \quad \text{IF } (e_K < 0) \text{ THEN } \omega = 360^\circ - \omega$$

$$\cos(\nu) = \frac{\hat{e} \cdot \dot{r}}{|\hat{e}| |\dot{r}|} \quad \text{IF } (\dot{r} \cdot \dot{v} < 0) \text{ THEN } \nu = 360^\circ - \nu$$

Special Cases

Elliptical equatorial:

$$\cos(\tilde{\omega}_{true}) = \frac{e_I}{|\hat{e}|} \quad \text{IF } (e_J < 0) \text{ THEN } \tilde{\omega}_{true} = 360^\circ - \tilde{\omega}_{true}$$

Circular Inclined:

$$\cos(u) = \frac{\hat{n} \cdot \hat{r}}{|\hat{n}| |\hat{r}|} \quad \text{IF } (r_K < 0) \text{ THEN } u = 360^\circ - u$$

Circular equatorial:

$$\cos(\lambda_{true}) = \frac{\hat{r}_I}{|\hat{r}|} \quad \text{IF } (r_J < 0) \text{ THEN } \lambda_{true} = 2\pi - \lambda_{true}$$

An example shows how to convert position and velocity vectors into orbital elements.

▼ Example 2-2. Finding the Orbital Elements with ELORB.

GIVEN: $\hat{r}_{IJK} = 6524.834 \hat{i} + 6862.875 \hat{j} + 6448.296 \hat{k}$ km
 $\hat{v}_{IJK} = 4.901327 \hat{i} + 5.533756 \hat{j} - 1.976341 \hat{k}$ km/s

FIND: Classical orbital elements

Convert units to find

$$\hat{r}_{IJK} = 1.023 \hat{i} + 1.076 \hat{j} + 1.011 \hat{k}$$
 ER

$$\hat{v}_{IJK} = 0.62 \hat{i} + 0.7 \hat{j} - 0.25 \hat{k}$$
 ER/TU

The magnitudes are $r = 1.796226$ ER and $v = 0.967936$ ER/TU

Begin by finding the specific angular momentum:

$$\hat{h} = \hat{r} \times \hat{v} = \begin{bmatrix} 1.023 & 1.076 & 1.011 \\ 0.62 & 0.7 & -0.25 \end{bmatrix} = -0.9767 \hat{i} + 0.88257 \hat{j} + 0.04898 \hat{k}$$
 ER²/TU

The magnitude of angular momentum is 1.317297.

Find the node vector (not the mean motion) using a cross product:

$$\hat{n} = \hat{K} \times \hat{h} = -0.88257 \hat{i} - 0.9767 \hat{j} \quad |\hat{n}| = 1.316386$$
 ER²/TU

$$\hat{e} = \frac{1}{\mu} \left[\left(v^2 - \frac{\mu}{r} \right) \hat{r} - (\hat{r} \cdot \hat{v}) \hat{v} \right] = -0.3146 \hat{i} - 0.38523 \hat{j} + 0.66804 \hat{k}$$

$$e = 0.832853$$

$$\xi = \frac{v^2}{2} - \frac{\mu}{r} = \frac{-0.967936}{2} - \frac{1}{1.79623} = -0.088273$$
 ER²/TU²

Because the orbit isn't parabolic (the eccentricity is not 1.0), find the semimajor axis by $a = -\frac{\mu}{2\xi}$:
 $a = 5.664247$ ER = 36,127.343 km

Then, find the semiparameter using the specific angular momentum:

$$p = \frac{h^2}{\mu} = \frac{1.317386^2}{1} = 1.735272$$
 ER = 11,067.790 km

The next task is to determine the angles. Find the inclination with a cosine expression. No quadrant check is necessary.

$$\cos(i) = \frac{h_K}{|\vec{h}|} = \frac{0.048\,98}{1.317\,297}$$

$$i = 87.870^\circ$$

Find the longitude of the ascending node. Notice the quadrant check here affects the final value.

$$\cos(\Omega) = \frac{n_I}{|\vec{n}|} = \frac{-0.882\,57}{1.316\,386} \quad \Omega = 132.10^\circ$$

$$\text{IF } (n_J < 0) \text{ THEN } \Omega = 360^\circ - \Omega \quad \Omega = 227.89^\circ$$

Find the argument of perigee similarly but without modification for quadrants:

$$\cos(\omega) = \frac{\vec{n} \cdot \vec{e}}{|\vec{n}| |\vec{e}|} = \frac{0.653\,96}{(1.316\,386)(0.832\,853)} \quad \omega = 53.38^\circ$$

$$\text{IF } (e_K < 0) \text{ THEN } \omega = 360^\circ - \omega$$

Finally, find the true anomaly:

$$\cos(\nu) = \frac{\vec{e} \cdot \vec{r}}{|\vec{e}| |\vec{r}|} = \frac{-0.060\,956\,6}{(0.832\,853)(1.796\,226)} \quad \nu = 92.335^\circ$$

$$\text{IF } (\vec{r} \cdot \vec{v} < 0) \text{ THEN } \nu = 360^\circ - \nu$$

Evaluate the special angles recalling the limitations imposed for planes in each calculation. Determine the variables relating to the location of perigee. Notice the difference in the values of the true longitude of perigee and the longitude of perigee!

$$\cos(\tilde{\omega}_{true}) = \frac{e_I}{|\vec{e}|} = \frac{-0.3146}{0.832\,853}, \text{ and } \tilde{\omega}_{true} = 112.194^\circ$$

$$\text{IF } (e_J < 0) \text{ THEN } \tilde{\omega}_{true} = 360^\circ - \tilde{\omega}_{true}, \text{ and } \tilde{\omega}_{true} = 247.806^\circ$$

$$\tilde{\omega} = \Omega + \omega = 227.89^\circ + 53.38^\circ = 281.27^\circ$$

Find the argument of latitude:

$$\cos(u) = \frac{\vec{n} \cdot \vec{r}}{|\vec{n}| |\vec{r}|} \quad \text{IF } (r_K < 0) \text{ THEN } u = 360^\circ - u$$

$$u = \cos^{-1} \left(\frac{(-0.882\,57)(1.023) + (-0.9767)(1.076)}{(1.316\,386)(1.796\,226)} \right) = 145.605\,49^\circ$$

Finally, determine the true longitude:

$$\cos(\lambda_{true}) = \frac{r_I}{|\vec{r}|} = \frac{1.023}{1.796\,226}$$

$$\text{IF } (r_J < 0) \text{ THEN } \lambda_{true} = 360^\circ - \lambda_{true}$$

$$\lambda_{true} = 55.282\,587^\circ$$



2.6 Application: r and v from Orbital Elements

We've seen how to find the orbital elements from the position and velocity vectors, but we often need the reverse process to complete certain astrodynamical studies. We'll call the process **RANDV** to indicate that we're determining the position and velocity vectors. The overall idea is to determine the position and velocity vectors in the perifocal coordinate system, PQW , and then rotate to the geocentric equatorial system. Although the orbit may not be elliptical, and therefore the PQW system would actually be undefined, we can elegantly work around this limitation. We can also make the method completely generic through several short, simple substitutions.

First, we must use the semiparameter instead of the semimajor axis. As previously mentioned, the semimajor axis is infinite for the parabola, whereas the semiparameter is defined for all orbits. The second requirement concerns how we treat the auxiliary classical orbital elements for the special cases of circular and equatorial orbits.

Let's begin by finding the position and velocity vectors in the perifocal coordinate system. We've developed and presented these equations previously but show them here coupled with the trajectory equation. Notice the use of the semiparameter to replace dependence on the semimajor axis.

$$\vec{r}_{PQW} = \begin{bmatrix} \frac{p \cos(\nu)}{1 + e \cos(\nu)} \\ \frac{p \sin(\nu)}{1 + e \cos(\nu)} \\ 0 \end{bmatrix} \quad (2-65)$$

An immediate difficulty arises when attempting to define the true anomaly for circular orbits. It turns out that the orbital elements may be *temporarily* replaced with the alternate elements to provide the necessary values for the calculations. Although you can design a change like this so it's transparent to users, make sure any changes or alternate codings use temporary variables and don't alter the original elements. It's possible to substitute values:

$$\begin{aligned} &\text{IF Circular Equatorial} \\ &\text{let } \omega = 0.0, \Omega = 0.0, \text{ and } \nu = \lambda_{true} \\ &\text{IF Circular Inclined} \\ &\text{let } \omega = 0.0 \text{ and } \nu = u \end{aligned} \quad (2-66)$$

The rationale for assigning ω and Ω to zero will be clear shortly; however, we haven't violated any assumptions because ω and Ω are undefined for circular orbits. Be careful not to return any changed variables in computer applications.

Find the velocity vector by differentiating the position vector:

$$\dot{\vec{v}}_{PQW} = \begin{bmatrix} \dot{r} \cos(\nu) - r\dot{\nu} \sin(\nu) \\ \dot{r} \sin(\nu) + r\dot{\nu} \cos(\nu) \\ 0 \end{bmatrix}$$

Remembering the geometry from Fig. 2-4, solve Eq. (2-8) as

$$r\dot{\nu} = \frac{h}{r}$$

Now, substitute the definitions of position and angular momentum:

$$r\dot{\nu} = \frac{\sqrt{\mu p} (1 + e \cos(\nu))}{p} = \sqrt{\frac{\mu}{p}} (1 + e \cos(\nu))$$

Using Eq. (2-15) and the equation above, write

$$\dot{r} = \sqrt{\frac{\mu}{p}} (e \sin(\nu))$$

Substituting these results into the differentiated vector gives us the final solution:

$$\dot{\vec{v}}_{PQW} = \begin{bmatrix} -\sqrt{\frac{\mu}{p}} \sin(\nu) \\ \sqrt{\frac{\mu}{p}} (e + \cos(\nu)) \\ 0 \end{bmatrix} \quad (2-67)$$

The next step is to rotate the position and velocity vectors to the geocentric equatorial frame. Although this is relatively easy for standard, elliptical, inclined orbits, we'll need to take certain precautions in order to account for special cases, as described with the true anomaly above. We've discussed two of these special cases; the third is the elliptical equatorial case:

$$\begin{aligned} &\text{IF Elliptical Equatorial} \\ &\text{set } \Omega = 0.0 \text{ and } \omega = \tilde{\omega}_{true} \end{aligned} \quad (2-68)$$

The assumptions remain intact because Ω is undefined for elliptical equatorial orbits.

We can now do the coordinate transformations using Eq. (1-35). We may want to multiply out these operations to reduce trigonometric operations. The rationale for setting certain variables to zero should now be apparent. For the special cases, a zero rotation causes the vector to remain unchanged, whereas a desired angular value causes a change.

Implementing RANDV

Computational efficiency results from assigning the trigonometric terms [$\sin(\nu)$, $\cos(\nu)$] and (μ/p) to temporary variables. This saves *many* transcendental operations and requires very little extra work. There are also some savings in treating special-case orbits if we reuse the same rotation matrices, but there may be some redundancy in special cases.

As with the **ELORB** algorithm, we may run many test cases to verify the routine. Because **RANDV** is simply designed to be a mirror calculation of the **ELORB** routine, we can use the same set of test reference data. But we must test several limiting cases. Algorithm 6 summarizes the process.

ALGORITHM 6: $RANDV(p, e, i, \Omega, \omega, \nu | u, \lambda_{true}, \tilde{\omega}_{true}) \Rightarrow \vec{r}_{IJK} \vec{v}_{IJK}$

IF Circular Equatorial

SET $(\omega, \Omega) = 0.0$ and $\nu = \lambda_{true}$

IF Circular Inclined

SET $\omega = 0.0$ and $\nu = u$

IF Elliptical Equatorial

SET $\Omega = 0.0$ and $\omega = \tilde{\omega}_{true}$

$$\vec{r}_{PQW} = \begin{bmatrix} \frac{p \cos(\nu)}{1 + e \cos(\nu)} \\ \frac{p \sin(\nu)}{1 + e \cos(\nu)} \\ 0 \end{bmatrix} \quad \vec{v}_{PQW} = \begin{bmatrix} -\sqrt{\frac{\mu}{p}} \sin(\nu) \\ \sqrt{\frac{\mu}{p}} (e + \cos(\nu)) \\ 0 \end{bmatrix}$$

$$\vec{r}_{IJK} = [\text{ROT3}(-\Omega)] [\text{ROT1}(-i)] [\text{ROT3}(-\omega)] \vec{r}_{PQW} = \left[\frac{IJK}{PQW} \right] \vec{r}_{PQW}$$

$$\vec{v}_{IJK} = [\text{ROT3}(-\Omega)] [\text{ROT1}(-i)] [\text{ROT3}(-\omega)] \vec{v}_{PQW} = \left[\frac{IJK}{PQW} \right] \vec{v}_{PQW}$$

$$\left[\frac{IJK}{PQW} \right] = \begin{bmatrix} \cos(\Omega)\cos(\omega) - \sin(\Omega)\sin(\omega)\cos(i) & -\cos(\Omega)\sin(\omega) - \sin(\Omega)\cos(\omega)\cos(i) & \sin(\Omega)\sin(i) \\ \sin(\Omega)\cos(\omega) + \cos(\Omega)\sin(\omega)\cos(i) & -\sin(\Omega)\sin(\omega) + \cos(\Omega)\cos(\omega)\cos(i) & -\cos(\Omega)\sin(i) \\ \sin(\omega)\sin(i) & \cos(\omega)\sin(i) & \cos(i) \end{bmatrix}$$

An example demonstrates the technique.

▼ Example 2-3. Finding Position and Velocity Vectors (RANDV Test Case).

GIVEN: $p = 11,067.790 \text{ km} = 1.735\ 27 \text{ ER}$, $e = 0.832\ 85$, $i = 87.87^\circ$, $\Omega = 227.89^\circ$,
 $\omega = 53.38^\circ$, $\nu = 92.335^\circ$

FIND: \vec{r}_{IJK} , \vec{v}_{IJK}

We have to change the rotation angles if we're using special orbits (equatorial or circular), but this orbit doesn't have special cases. From the given information, form the PQW position and velocity vectors:

$$\begin{aligned}\vec{r}_{PQW} &= \begin{bmatrix} p \cos(\nu) \\ 1 + e \cos(i) \\ p \sin(\nu) \\ 1 + e \cos(i) \\ 0 \end{bmatrix} = \begin{bmatrix} 1.735\ 27 \cos(92.336)^\circ \\ 1 + 0.832\ 84 \cos(92.336)^\circ \\ 1.735\ 27 \sin(92.336) \\ 1 + 0.832\ 84 \cos(92.336)^\circ \\ 0 \end{bmatrix} = \begin{bmatrix} -0.073\ 186\ 7 \\ 1.794\ 733\ 9 \\ 0 \end{bmatrix} \text{ ER} \\ \vec{v}_{PQW} &= \begin{bmatrix} -\sqrt{\frac{\mu}{p}} \sin(\nu) \\ \sqrt{\frac{\mu}{p}} (e + \cos(i)) \\ 0 \end{bmatrix} = \begin{bmatrix} -\sqrt{\frac{1}{1.735\ 27}} \sin(92.336) \\ \sqrt{\frac{1}{1.735\ 27}} (0.832\ 84 + \cos(92.336)) \\ 0 \end{bmatrix} = \begin{bmatrix} -0.758\ 499\ 8 \\ 0.601\ 313\ 6 \\ 0 \end{bmatrix} \text{ TU}\end{aligned}$$

Rotate these vectors to the geocentric equatorial system using the following rotation matrices:

$$\vec{r}_{IJK} = [\text{ROT3}(-\Omega)] [\text{ROT1}(-i)] [\text{ROT3}(-\omega)] \vec{r}_{PQW}$$

$$\vec{v}_{IJK} = [\text{ROT3}(-\Omega)] [\text{ROT1}(-i)] [\text{ROT3}(-\omega)] \vec{v}_{PQW}$$

Or, use the expanded matrix with a computer to do the many trigonometric operations, which result in the transformation matrix

$$\begin{bmatrix} IJK \\ PQW \end{bmatrix} = \begin{bmatrix} -0.377\ 736\ 47 & 0.554\ 597\ 39 & -0.741\ 442\ 44 \\ -0.462\ 538\ 21 & 0.580\ 670\ 14 & 0.669\ 985\ 52 \\ 0.802\ 105\ 71 & 0.596\ 023\ 42 & 0.037\ 182\ 20 \end{bmatrix}$$

Finally, multiply each vector to apply the transformation:

$$\begin{aligned}\vec{r}_{IJK} &= \begin{bmatrix} IJK \\ PQW \end{bmatrix} \vec{r}_{PQW} = \begin{bmatrix} -0.377\ 736\ 47 & 0.554\ 597\ 39 & -0.741\ 442\ 44 \\ -0.462\ 538\ 21 & 0.580\ 670\ 14 & 0.669\ 985\ 52 \\ 0.802\ 105\ 71 & 0.596\ 023\ 42 & 0.037\ 182\ 20 \end{bmatrix} \begin{bmatrix} -0.073\ 186\ 7 \\ 1.794\ 733\ 9 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} 1.023 \\ 1.076 \\ 1.011 \end{bmatrix} \text{ ER} = \begin{bmatrix} 6524.834 \\ 6862.875 \\ 6448.296 \end{bmatrix} \text{ km}\end{aligned}$$

$$\begin{aligned}\vec{v}_{IJK} &= \begin{bmatrix} IJK \\ PQW \end{bmatrix} \vec{v}_{PQW} = \begin{bmatrix} -0.377\ 736\ 47 & 0.554\ 597\ 39 & -0.741\ 442\ 44 \\ -0.462\ 538\ 21 & 0.580\ 670\ 14 & 0.669\ 985\ 52 \\ 0.802\ 105\ 71 & 0.596\ 023\ 42 & 0.037\ 182\ 20 \end{bmatrix} \begin{bmatrix} -0.758\ 499\ 8 \\ 0.601\ 313\ 6 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} 0.62 \\ 0.70 \\ -0.25 \end{bmatrix} \text{ ER/TU} = \begin{bmatrix} 4.901\ 320 \\ 5.533\ 756 \\ -1.976\ 341 \end{bmatrix} \text{ km/s}\end{aligned}$$

Problems

1. A satellite is in a circular parking orbit about the Earth (120 km altitude). To begin a voyage to Mars, what velocity does it need to escape from the Earth?

Would your answer be different for an elliptical orbit ($e = 0.3$, $\nu = 145^\circ$, a the same)? Why or why not?

If you assume both burns are tangent to the existing velocity, what are the required increases in velocity for each case?

2. Is there a common point at which the velocity of a satellite in elliptical orbit equals the velocity in a circular orbit? If so, where?
3. Identify the type of orbits given the following data:
 - $\hat{r} = 6372.0\hat{i} + 6372.0\hat{j}$ km, $\hat{v} = -4.7028\hat{i} + 4.7028\hat{j}$ km/s
 - $\hat{e} = 0.07\hat{i} + 0.021\hat{j} + 0.021\hat{k}$
 - $\hat{h} = 2.3\hat{j}$ ER²/TU
 - $\hat{h} = -3.5\hat{j}$ ER²/TU
4. Calculate the Delaunay, Poincaré, and equinoctial orbital elements for the satellite in Example 2-2.
5. Show how the restricted three-body solution reduces to the two-body motion. Plot the orbit for a satellite at 50,000 km above the Earth.
6. The opportunity to do a Grand Tour of the planets occurred in a window between August 20, 1977 to September 23, 1977 when all the planets “lined up” to permit a sequential flyby. You decide to recreate the process in a simulation. (Hint: Voyager 2 was launched on August 20, and Voyager 1 on September 5)
 - How do you proceed?
 - If you need reduction, what epoch might you choose?
 - Does the n -body formula apply?
7. Find colinear Lagrange points for Jupiter. (Use data from Table D-5.)
8. Last night while you were asleep, all the distances in the universe were reduced by 50%. How could you tell?
9. A common approximation for interplanetary flights is the **patched-conic trajectory** that pieces three conic sections together. The concept relies on the Sphere of Influence about a planet. We can estimate the radius of the SOI (r_{SOI}) using the following formula [Roy (1988, 385)]. We must know the radius and mass of the planet as well as the mass of the Sun.

$$r_{SOI} = \left(\frac{m_{planet}}{m_{\odot}} \right)^{2/5} r_{planet}$$

You decide to calculate the parameters for a flight from Earth to Jupiter using the patched-conic approximation. Assume the satellite leaves Earth from a 200 km altitude, circular orbit, and arrives around Jupiter in a 10,000 km altitude circular orbit. What are the semimajor axis, eccentricity, and change in velocity values for this mission? (Hint: You'll need to review the Hohmann transfer in Sec. 5.3.1)

CHAPTER 3 FUNDAMENTAL APPLICATIONS

- 3.1 Historical Background
- 3.2 Introduction
- 3.3 Time Conversions
- 3.4 Observation Transformations
- 3.5 Application: Sun Position Vector
- 3.6 Application: Moon Position Vector
- 3.7 Application: Planetary Ephemerides
- 3.8 Application: Sunrise, Sunset, and Twilight Times
- 3.9 Application: Sight and Light
- 3.10 Converting *IJK* to Latitude and Longitude (LatLon)

3.1 Historical Background

Contributions to modern astrodynamics by Copernicus (c. 1540), Kepler (c. 1610), Newton (c. 1690), and others were made possible by some interesting advancements. Pannekoek (1989, 199) remarks:

The growth of science consists not only in the development of ideas and theoretical explanations but also in the improvement of the practical working methods. The practical work of the astronomer is twofold—observation and computation. From the numbers read on the instruments, which are the direct results of observation, the desired values of the astronomical quantities must be derived by computation. Thus in the fifteenth and sixteenth centuries the construction of the mathematical apparatus was as important a part of scientific progress as was the construction of the technical apparatus, the instruments for observation.

A glance back at the historical technologies and mechanical constraints of the fifteenth and sixteenth centuries will show that many of today's conventions are products of the ingenuity and methods adopted at the time.

The Ptolemaic tables (c. 150) in the *Almagest* were the last major astronomical endeavour for many centuries after the death of Ptolemy in 170. Indeed, the fall of the Roman Empire in 476 eliminated financial support for many scientific efforts. Centuries later, the Catholic Church was very powerful (as Galileo found out!). Religious festivals became increasingly important, and in particular, the correct day for Easter was a prime concern. According to the resolutions of the Nicene Council in 325 A.D., Easter was to occur on the first Sunday following a full Moon after the first day of spring (the vernal equinox). This

meant the date could fall anywhere between March 22 and April 25. Determining this day was very difficult for most scientists of the fifteenth and sixteenth centuries and continued to perplex scientists until almost 1800!

The *Julian calendar*, instituted about the time of Christ, provided a means to determine the occurrence for Easter, but the Julian year has 365.25 days, which is a little too long. Although the error (about 0.006 36 days in a sidereal year) doesn't seem too large, it amounts to one day in about 157 years and about eight days over 1300 years. By 1300, calculations for the vernal equinox showed it occurring on March 13, when it actually fell on March 21. Something had to be done. The Council of Trent (1545–1563) ultimately proposed changing the calendar. In 1582, Gregory XIII adopted their changes, including a 10-day step function to correct accumulated errors. The Gregorian calendar we use today derives from the action of Gregory XIII, plus the insight of a librarian, Aloisius Lilius, who suggested having leap years occur only in centuries evenly divisible by 400 (Pannekoek, 1961, 220). This changed the error to about 0.001 14 days in a sidereal year.

Limits in observational tools also affected the calendar's accuracy. Observatories were very important in verifying the accuracy of the stellar tables. Although observational techniques and equipment hadn't progressed much since the time of Christ, supporting technologies had developed that set the stage for many new devices to follow. Landgrave William IV (1532–1592) was always very interested in astronomy. In 1561, he oversaw construction of the Cassel observatory. Among the new features was a revolving roof, a basic concept of modern observatories. The observatory used by Tycho Brahe (1546–1601) usually receives much more credit because its discoveries are more widely known. Wherever the observatories were built, the actual machinery was state-of-the-art. The ability of the craftsmen of the fifteenth and sixteenth centuries is truly remarkable, as the accuracy of their machines attests.

Although we tend to think increasingly in digital terms, the pioneers in astronomy began with the Greek number system. Because the Greek system used letters to represent numbers, computations were tedious. The next choice was Roman numerals, but they were also clumsy in handling small fractions of degrees. The thirteenth-century manual of Leonardo of Pisa (1170–1230), called *Fibonacci*, first used Arabic numerals in computations and calculations and was quite successful (Pannekoek, 1961, 199–200). The Babylonian *sexagesimal* system, used with the Arabic numerals, effectively recorded and computed the results of observations. The *sexagesimal* system is essentially a degree-arc-minute-arcsecond format in which the interval of division is 60.

These computing practices nicely matched the limited measuring devices of the time, which had no digital readouts. In fact, clocks didn't even appear until the seventeenth century. In any case, it was relatively easy to inscribe a wheel or other large apparatus affixed to the optical device. The wheel could then be etched with markings for degrees, arc-minutes, and arcseconds, each with an even 60 gradations. The human operator could interpret the actual measurement through an optical sighting glass or simply by looking at the markings. Many observatories still use this technique because most published values of declination, latitude, and elevation angles are in the sexagesimal degrees, arcminutes, arcseconds format. As an aside, the hour-minute-second format arises similarly. As clocks

became more popular, their gears offered an effective (and easy) way to record the “precise” time of an observation. This convention was very accurate and is the reason that right ascension, sidereal time, and hour angles maintain an hour-minute-second format.

As clocks and measurements of each observatory became more accurate, observations at different locations began to disagree. Frederick Bessel (1748–1846) stated the reason for the differences: parallax. **Parallax** is the angular difference in the apparent direction of an object when viewed from two different locations. The study of parallax didn’t proceed with vigor until instruments could detect the differences. Most objects studied were stars, all of which are at great distances from the Earth. The parallax for these stars is so small as to be imperceptible in historical times. As studies of closer objects—moons, comets, and asteroids—progressed, converting observations became increasingly important. The discovery of parallax could have proved quite helpful to Copernicus and Galileo because it proves the Earth revolves around the Sun. Remember that Copernicus was criticized in the mid-1500s for not being able to account for parallax. Observers weren’t able to detect parallax until 1838, when Bessel first published some of the distances to the stars using parallax calculations. Even today, parallax measurements stress the abilities of modern machinery. Typical example values are $0.002\ 44^\circ$ for the Sun ($8.8''$), $0.008\ 89^\circ$ for the Moon ($32''$) and $0.000\ 28^\circ$ for Alpha Centauri ($1''$), so Copernicus shouldn’t feel too bad!

3.2 Introduction

This chapter describes utilities we need for analysis. Most texts don’t discuss these techniques in detail, but they’re important when developing a complete system.

Formats and systems of time vary widely, and many standards exist. Some systems use seconds from January 1, 1970, 0 h UTC or seconds from the beginning of the year; others use modified Julian dates (several forms); still others use seconds from an arbitrary epoch. Although these units and formats may be useful in particular applications, we must be able efficiently to make them compatible with routines discussed in this text. For computer programs in particular, we’d like to have a *common* time system to pass between all routines. Accurate time is vital to space systems. Consider a satellite traveling in a low-Earth orbit, say, 200 km in altitude. Circular velocity for this satellite is 7784 m/s. An error of a tenth of a second leads to a large error in position because the satellite travels over 778 m! Depending on the accuracy requirements for your system, this error could be significant.

Observation relations are extremely important when calculating orbits. Right ascension and declination are often used to locate the stars, but if we apply them to Earth-orbiting satellites, the coordinate frame (topocentric or geocentric) in which the angles are measured is crucial. If we don’t do simple transformations, we’ll produce incorrect results.

We require position vectors for the Sun, Moon, and planets to analyze perturbation forces on the satellite. I’ll show low-precision formulas, with references to the original expansions to permit tailoring for certain applications. Sunrise or sunset conditions may be important when designing a particular mission for sensor viewing. Using expansions developed in the reduction of coordinates, we can produce a method to determine the local time of sunrise and sunset. Sensor viewing is especially important for optical sensors.

Accurately determining visibility between satellites is extremely useful for communications. We'll discuss a routine which allows us to determine line-of-sight very quickly. Finally, we require conversions to determine sub-latitude points, which help us construct ground-tracks for operational analysis. This routine has many applications.

3.3 Time Conversions

Various conversions are required to account for the different types of time, angles, and input data. The algorithms in this section fall into two main types of routines: one to convert between time systems in hours or degrees and the other relating to the time of an event expressed in days or Julian dates.

3.3.1 DMS to Rad / Rad to DMS

These two routines convert *angles* expressed in a degree-arcminute-arcsecond (*DMS*) format to a single quantity in degrees and fractions of a degree, to one in radians (*Rad*), and vice versa. The algorithm for *DMStoRad* contains only one equation which performs the operation. Be sure each component is negative for negative values.

ALGORITHM 7: *DMStoRad* ($^{\circ}, ', '' \Rightarrow \alpha_{DMS}$)

$$\alpha_{DMS} = \left(deg + \frac{'}{60} + \frac{''}{3600} \right) \frac{\pi}{180^{\circ}}$$

The reverse procedure (*RadtoDMS*) requires more steps to separate each of the components from the given α_{DMS} . The first step is to convert α_{DMS} to degrees because most computer programs will carry this value in radians. We do the actual separation using truncation formulas (taking the integer part of a real value; see Sec. A.2 for explanations).

ALGORITHM 8: *RadtoDMS* ($\alpha_{DMS} \Rightarrow ^{\circ}, ', ''$)

$$\begin{aligned} Temp &= \alpha_{DMS} \left(\frac{180^{\circ}}{\pi} \right) \\ deg &= \text{TRUNC}(Temp) \\ ' &= \text{TRUNC}((Temp - deg) 60) \\ '' &= \left(Temp - deg - \frac{'}{60} \right) 3600 \end{aligned}$$

An example illustrates both processes. To use them in computer software, be sure you understand how the TRUNC process works for negative numbers.

▼ **Example 3-1: Finding DMS.**

GIVEN: $-35^\circ - 15' - 53.63''$

FIND: α_{DMS}

First, determine the degrees in a fractional form:

$$\alpha_{DMS} = -35 - \frac{15}{60} - \frac{53.63}{3600} = -35.264\ 897^\circ = -0.615\ 488\ 6 \text{ rad}$$

The reverse process is

$$^\circ = \text{TRUNC}(-35.264\ 897) = -35$$

$$' = \text{TRUNC}(-35.264\ 897 - -35)60 = -15$$

$$'' = \text{TRUNC}(-35.264\ 897 - -35 - -15/60)3600 = -53.6299$$



3.3.2 HMS to Rad / Rad to HMS

These routines convert times in a hour-minute-second (*HMS*) format to one in radians and vice versa. As with the *DMS* routines, *HMStoRad* contains only one equation which carries out the operation; however, recall the basic relations in Eq. (1-41):

ALGORITHM 9: HMStoRad ($h, min, s \Rightarrow \tau_{HMS}$)

$$\tau_{HMS} = 15\left(h + \frac{min}{60} + \frac{s}{3600}\right)\frac{\pi}{180^\circ}$$

The reverse procedure (*RadtoHMS*) requires more steps to separate each of the components from the given τ_{HMS} . We convert τ_{HMS} to degrees because most computer programs carry this value in radians, and then to hours by dividing by 15. Don't confuse this generic conversion value with one for the Earth's rotation, which is quite close: $0.250\ 684\ 4^\circ/\text{min}$ (60) = $15.041^\circ/\text{h}$. We do the actual separation using truncation formulas.

ALGORITHM 10: RadtoHMS ($\tau_{HMS} \Rightarrow h, min, s$)

$$Temp = \tau_{HMS} \frac{180^\circ}{15\pi}$$

$$h = \text{TRUNC}(Temp)$$

$$min = \text{TRUNC}((Temp - h)60)$$

$$s = \left(Temp - h - \frac{min}{60}\right)3600$$

An example shows the operation of these two routines. Be careful of negative signs.

▼ **Example 3-2. Finding HMS.**

GIVEN: 15 h 15 min 53.63 s

FIND: τ_{HMS}

First, determine the time in hours:

$$\tau_{HMS} = 15 + \frac{15}{60} + \frac{53.63}{3600} = 15.264\ 90\text{ h}$$

Now recall the conversion to degrees involves multiplying by $15^\circ/\text{h}$.

$$\tau_{HMS} = 228.9735^\circ$$

To use this value in computer software, convert to radians:

▲ $\tau_{HMS} = 228.9735 \left(\frac{\pi}{180} \right) = 3.996\ 341\ \text{rad}$

3.3.3 HMS to UT / UT to HMS

We often use universal time (*UT*) in calculations, although it may be inconvenient to send three separate variables (h-min-s) throughout a program. Two approaches exist for determining this conversion: (1) a routine which converts the time (h-min-s) to elapsed seconds from the beginning of the day of interest and (2) a method to combine the time (h-min-s) into one value representing time as it is usually displayed in text (14:13:35.46). These two routines combine all the time information in one variable for distribution.

The first method converts the h-min-s time into the number of seconds in a day.

ALGORITHM 11: HMStoUT (*h, min, s* $\Rightarrow \tau_{UT}$)

$$\tau_{UT} = 3600h + 60min + s$$

The reverse procedure uses truncation routines. Be sure to enter negative values correctly.

ALGORITHM 12: UTtoHMS ($\tau_{UT} \Rightarrow h, min, s$)

$$Temp = \frac{\tau_{UT}}{3600}$$

$$h = \text{TRUNC}(Temp)$$

$$min = \text{TRUNC}((Temp - h) 60)$$

$$s = \left(Temp - h - \frac{min}{60} \right) 3600$$

The second conversion yields a format which looks like HHMinMin.ssssss. You may encounter this type of conversion for systems using older computer-programming techniques with limited data structures:

$$\tau_{UT} = 100h + min + (0.01)s$$

The reverse procedure again uses various truncation routines. FRAC is an operation which extracts just the decimal part of a real number (See Sec. A.2). Thus,

$$\begin{aligned} h &= \text{TRUNC}(0.01\tau_{UT}) \\ min &= \text{TRUNC}(\tau_{UT} - 100h) \\ s &= 100\text{FRAC}(\tau_{UT}) \end{aligned}$$

Consider the following example.

▼ **Example 3-3. Converting HMS to UT.**

GIVEN: 13 h 22 min 45.98 s

FIND: τ_{UT}

Method 1 begins by converting the given time to seconds:

$$\tau_{UT} = 3600(13) + 60(22) + 45.98 = 48,165.98$$

In reverse, we carefully use the truncation routines:

$$Temp = 13.379\,438\,89$$

$$h = \text{TRUNC}\left(\frac{48,165.98}{3600}\right) = 13$$

$$min = \text{TRUNC}((13.379\,438\,89 - 13)60) = 22$$

$$s = \left(13.379\,438\,89 - 13 - \frac{22}{60}\right)3600 = 45.979\,999$$

Notice the small errors in regeneration that occur with all computer routines.

By Method 2 this problem is trivial and solved by inspection, so

$$\tau_{UT} = 1322.4598$$

▲ Computer implementations can be tricky in shifting the values into the appropriate decimal locations.
Reverse the process carefully.

3.3.4 YMD to Day of Year/ Day of Year to YMD

We derive the day of the year from the day and month of a particular year (*YMD*). The solution requires knowledge of the year to determine if a leap year exists. The process is much easier on a computer because many of the operations are repetitive.

The first step is to set up an array, *LMonth*, containing the length of each month in days. Then, modify this array if the year of interest is a leap year. For leap years, the length of days for February changes to 29. Use

```
IF (Yr MOD 4) = 0 THEN
  LMonth[2] = 29
```

Leap years are irregular, meaning they occur only in centuries evenly divisible by 400 (the year 2000 is a leap year while 1900 and 2100 are not). Thus, the formula is valid only for years between 1900 and 2100.

Then, find the day of the year by adding the days for each month up to the month of interest and then the number of days in the month of interest. The inverse process must also determine the presence of leap years. Set up a temporary variable to add the days of each month until the total with the next month exceeds the given day of the year. The month and day are then available.

An example will demonstrate both methods.

▼ Example 3-4. Finding the Day of the Year.

GIVEN: May 8, 1992

FIND: Day of the year

Determine if 1992 is a leap year by (Year MOD 4). 1992 MOD 4 = 0, so 1992 is a leap year. Set the length of February to 29. Now, add the days of each month from the beginning of the year until you reach the month of interest:

$$31 + 29 + 31 + 30 = 121.$$

Add the specific day of the month to this result and get the day of the year:

$$\text{Day of Yr} = 121 + 8 = 129.$$

Using the reverse process, again find that 1992 is a leap year and set the length of February to 29 days. Then set a temporary variable at zero and add the days of each month until the result plus the next month *exceeds* the given day of the year. Thus,

$$\text{Temp} = \text{Temp} + \text{LMonth} \} 5 \text{ times.}$$

▲ $\text{Temp} = 121 + 31$ for June = 152 which is greater than 129. Thus the month is May, and the day is $129 - 121 = 8$.

3.3.5 YMDHMS to Days / Days to YMDHMS

The year-month-day-hour-minute-second (**YMDHMS**) method finds the days and fractions of a day from the beginning of the year. For example, January 1 at noon is day 1.5. This particular problem is especially useful in converting epochs given on the 2-line element sets. In general, the routine could find the Julian date of the desired time and subtract the Julian date of the beginning of the year to find the elapsed time. However, this approach would require excessive computer operations. The approach below is slightly longer but much easier in terms of calculations.

This process follows the *YMD* to day of the year method (Sec. 3.3.4) but adds time information to the final result. At this point, add the *HMS* information:

YMD to Day of Year (*YMD* \Rightarrow *Day of Yr*)
 $\text{days} = \text{Day of Yr} + h/24 + \text{min}/1440 + s/86400$

We can develop test cases from Table 1-5. The reverse process is more difficult because it requires many truncation routines. Begin by determining if the year is a leap year. Then, convert the day of the year to the individual month, day, hour, minute, and second. Quickly find the day of the year as the integer part of days. The remainder determines the day and month using the routine previously discussed:

$$\text{Day of Yr} = \text{TRUNC}(Days)$$

$$\text{YMD to Day of Yr } (Day \text{ of Yr} \Rightarrow \text{YMD})$$

Finally, find the hours, minutes, and seconds for the *UT* once you've converted to seconds:

$$\tau_{UT} = (Days - \text{Day of Yr})86,400$$

$$\text{UTtoHMS } (\tau_{UT} \Rightarrow h, \text{min}, s)$$

At this point, reverse the process (*DaystoYMDHMS*) to complete the solution. An example demonstrates each approach.

▼ Example 3-5. Converting *YMDHMS* to *Days*.

GIVEN: March 18, 2001 12:14 P.M. *UT*

FIND: Days

Begin by checking for a leap year. If 2001 MOD 4 = 0, it's a leap year. It's not, so leave the value for February as 28 days. Now add the days up to the month of interest:

$$\text{YMD to Day of Year } (2001, 3, 18 \Rightarrow \text{Day of Yr})$$

$$\text{Day of Yr} = 77$$

Then, correct for the *UT*:

$$Days = \text{Day of Yr} + 12/24 + 14/1440 = 77.509\bar{722}2$$

For the reverse process, first determine the leap-year status. Then, find the *day of the year* through truncation simply as Day 77. Using the *Day of Year* to *YMD* routine, discover the 18th of March. Finally, convert back to *UT*:

$$\tau_{UT} = (Days - \text{Day of Yr})12 = (77.509\bar{722}2 - 77)24 = 12.233\bar{332}8\text{ h}$$

$$\blacktriangle \quad \text{UTtoHMS } (\tau_{UT} \Rightarrow h, \text{min}, s) \text{ and } h = 12, \text{min} = 13, s = 59.998\bar{08}.$$

3.3.6 Inverse Julian Date

The process for finding the Julian date is cumbersome but straightforward using Algorithm 2. But reversing it isn't as simple. We must carefully consider the year, month, and day.

The first step is to find the year of the observation by determining the number of Julian years from an epoch. Notice we use the 1900 epoch, T_{1900} , because it isn't a leap year. Don't confuse this formula with the Julian centuries used in determining sidereal time and other formulas. In this case,

$$T_{1900} = \frac{JD - 2,415,019.5}{365.25}$$

$$Yr = 1900 + \text{TRUNC}(T_{1900})$$

The days are simply the remainder from the epoch (1900), correcting for leap years.

$$\text{LeapYrs} = \text{TRUNC}((Yr - 1900 - 1)(0.25))$$

$$\text{Days} = (JD - 2,415,019.5) - ((Yr - 1900)(365.0) + \text{LeapYrs})$$

Check for case of beginning of a year

IF Days < 1.0 THEN

$$Yr = Yr - 1$$

$$\text{LeapYrs} = \text{TRUNC}((Yr - 1900 - 1)(0.25))$$

$$\text{Days} = JD - 2,415,019.5 - ((Yr - 1900)(365.0) + \text{LeapYrs})$$

▼ Example 3-6. Finding the Inverse Julian Date.

GIVEN: $JD = 2,449,877.345\ 876\ 2$

FIND: YMD, HMS

Begin by finding the Julian centuries (times 100) from 1900:

$$T = (2,449,877.345\ 876\ 2 - 2,415,019.5) / 365.25 = 34,857.845\ 876\ 2 / 365.25 = 95.435\ 580\ 7$$

To find the year, use $Yr = 1900 + \text{TRUNC}(T) = 1995$. Determine leap years by

$$\text{LeapYrs} = \text{TRUNC}((1995 - 1900 - 1))0.25 = 23$$

$$\begin{aligned} \text{Days} &= (2,449,877.345\ 876\ 2 - 2,415,019.5) - ((1995 - 1900)(365) + 23) \\ &= 34,857.845\ 876\ 2 - 34,698 = 159.845\ 876\ 2 \end{aligned}$$

Notice the integer part of this number is the day of the year, 159. Next, use

$$\text{YMD to Day of Year} (1995, Day of Yr \Rightarrow YMD)$$

to find $Yr = 1995$, $Mo = \text{June}$, and $D = 8$. The final calculations determine the hours, minutes, and seconds. Using the day of the year and the days values determined previously, set up UT :

$$\tau_{UT} = (\text{Days} - \text{Day of Yr})12 = (159.845\ 876\ 2 - 159)24 = 20.301\ 028\ 8\ h$$

$$UT \text{toHMS } (\tau_{UT} \Rightarrow h, \text{min}, s) \text{ and } h = 20\ h, \text{min} = 18\ \text{min}, s = 3.703\ 68\ s.$$

▲ We designed the problem from June 8, 1995, 20:18:03.70370. Notice the error introduced by not keeping all digits. These routines are designed for computers!

3.4 Observation Transformations

Chapter 1 introduced location parameters (Sec. 1.3.1), and various coordinate systems (Sec. 1.4) containing the concepts of right ascension and declination, azimuth and elevation, as well as ecliptic latitude and longitude. To study satellite motion, we must be able to convert between these systems. Differences between topocentric and geocentric values are often negligible for the stars but can be large for satellites in low-Earth orbits. We must address two main functions: determining the specific angles from the position and velocity vectors and converting between angles of different systems.

3.4.1 Geocentric Right Ascension and Declination

Our problem is to determine these parameters when given the position in geocentric coordinates and vice versa. Figure 3-1 shows the geometry.

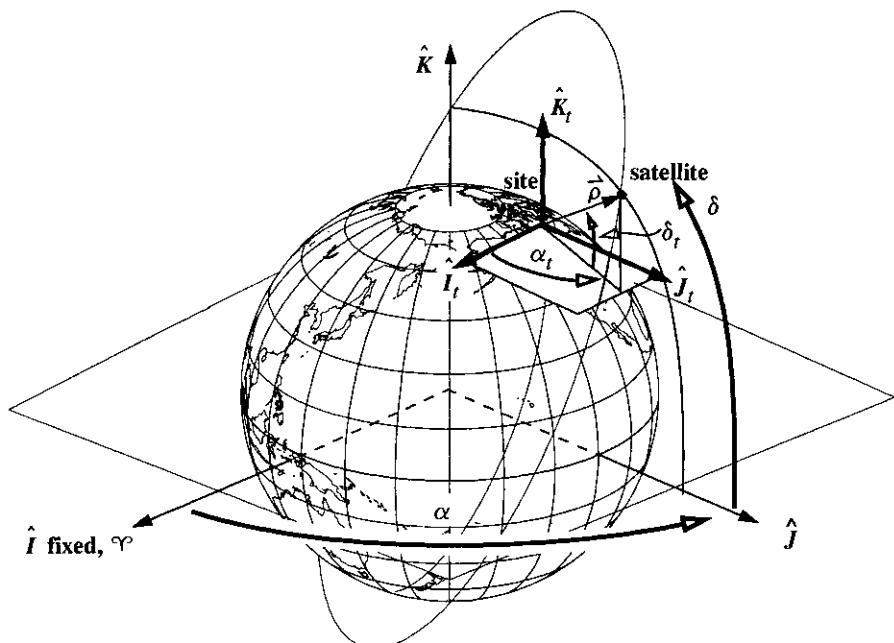


Figure 3-1. Right Ascension and Declination. Standard right ascension and declination use the Earth's equatorial plane. Topocentric right ascension and declination values use a plane parallel to the Earth's equator but located at a particular site. Transformations rely on determining the slant-range vector (ρ) in the IJK system.

This is the easiest of the transformations because we can find the position and velocity vectors directly without any transformations. If we write the position vector in terms of right ascension and declination directly from Fig. 3-1,

$$\mathbf{r}_{IJK} = \begin{bmatrix} r \cos(\delta) \cos(\alpha) \\ r \cos(\delta) \sin(\alpha) \\ r \sin(\delta) \end{bmatrix} \quad (3-1)$$

We find the velocity vector by direct differentiation. Note that, when we work with stars, the angular rates usually come from the proper motion coefficients. (See “Proper Motion and Radial Velocity” on page 92.)

$$\dot{\mathbf{v}}_{IJK} = \begin{bmatrix} \dot{r} \cos(\delta) \cos(\alpha) - r \sin(\delta) \cos(\alpha) \dot{\delta} - r \cos(\delta) \sin(\alpha) \dot{\alpha} \\ \dot{r} \cos(\delta) \sin(\alpha) - r \sin(\delta) \sin(\alpha) \dot{\delta} + r \cos(\delta) \cos(\alpha) \dot{\alpha} \\ \dot{r} \sin(\delta) + r \cos(\delta) \dot{\delta} \end{bmatrix} \quad (3-2)$$

The usual occurrence of this transformation is the reverse process of determining the angles from the position and velocity vectors. From Fig. 3-1, we can find the formulas for right ascension and declination. Singular points exist when the satellite is directly over the North or South Pole because the equatorial component of the position vector ($\sqrt{r_I^2 + r_J^2}$) is zero. The right ascension formula is undefined in this event, but we have another alternative if we know the velocity vector because it won't be zero. Because we're trying to determine only the orientation of the satellite's plane, the components of the velocity vector indicate the direction of the satellite for the special case when it's directly overhead. We use the I and J components of the velocity vector just as we use the position vector for "normal" cases. Algorithm 13 shows the result. Include the velocity vector if possible, even though you need the velocity only for error prevention.

Although the rates are less common, a few simple equations make the transformation complete. Find the range rate using results from Sec. 2.2.3. Recall that $\dot{r} \cdot \vec{v} = r\dot{r}$, where \dot{r} was the radial velocity. Solving this relation gives the result for this problem:

$$\dot{r} = \frac{\vec{r} \cdot \vec{v}}{r}$$

Find the declination rate by solving the K -component of the velocity vector. The rate is

$$\dot{\delta} = \frac{v_K - \dot{r} \sin(\delta)}{r \cos(\delta)}$$

The right ascension is more difficult and requires more formulas. Consider the negative K -component of $\vec{r} \times \vec{v}$,

$$\begin{aligned} v_I r_J - v_J r_I &= \{ \dot{r} \cos(\delta) \cos(\alpha) - r \sin(\delta) \cos(\alpha) \dot{\delta} - r \cos(\delta) \sin(\alpha) \dot{\alpha} \} r_J \\ &\quad - \{ \dot{r} \cos(\delta) \sin(\alpha) - r \sin(\delta) \sin(\alpha) \dot{\delta} + r \cos(\delta) \cos(\alpha) \dot{\alpha} \} r_I \end{aligned}$$

Notice that the terms involving \dot{r} and $\dot{\delta}$ drop out, and the coefficients in the remaining terms are actually r_J and r_I , respectively. Thus, we can solve for $\dot{\alpha}$:

$$\dot{\alpha} = \frac{v_I r_J - v_J r_I}{-r_J^2 - r_I^2}$$

The complete algorithm is

ALGORITHM 13: Geocentric Radec ($\vec{r}_{IJK}, \vec{v}_{IJK} \Rightarrow \alpha, \delta$)

$$r = |\vec{r}|$$

$$\sin(\delta) = \frac{r_K}{r} \quad \cos(\delta) = \frac{\sqrt{r_I^2 + r_J^2}}{r}$$

$$\left[\begin{array}{l} \text{IF } \sqrt{r_I^2 + r_J^2} \neq 0 \\ \sin(\alpha) = \frac{r_J}{\sqrt{r_I^2 + r_J^2}} \quad \cos(\alpha) = \frac{r_I}{\sqrt{r_I^2 + r_J^2}} \\ \text{ELSE} \\ \sin(\alpha) = \frac{v_J}{\sqrt{v_I^2 + v_J^2}} \quad \cos(\alpha) = \frac{v_I}{\sqrt{v_I^2 + v_J^2}} \end{array} \right]$$

$$\dot{r} = \frac{\vec{r} \cdot \vec{v}}{r}$$

$$\dot{\alpha} = \frac{v_I r_J - v_J r_I}{-r_J^2 - r_I^2}$$

$$\dot{\delta} = \frac{v_K - \dot{r} \sin(\delta)}{\sqrt{r_I^2 + r_J^2}}$$

Notice we need only the sine expression for the declination when it ranges within $\pm 90^\circ$. The only singularity occurs for the right ascension when the satellite is directly overhead. Solution is possible only if the velocity vector is available. You may notice the velocity vector could also have no equatorial component whenever a satellite is traveling on a straight-line trajectory away from the North or South Poles.

3.4.2 Topocentric Right Ascension and Declination

Operations with the topocentric right ascension and declination present an additional problem in that the coordinate system moves with time. Due to this translation, we need the site location to complete the transformation. Transformations involving the right ascension and declination angles and their rates use Eq. (3-1) and Eq. (3-2), but remember we're in the topocentric coordinate system, so we need a transformation to get geocentric coordinates. The position and velocity vectors are

$$\vec{\rho}_{I_t J_t K_t} = \begin{bmatrix} \rho \cos(\delta_t) \cos(\alpha_t) \\ \rho \cos(\delta_t) \sin(\alpha_t) \\ \rho \sin(\delta_t) \end{bmatrix}$$

$$\dot{\vec{r}}_{IJK} = \dot{\vec{\rho}}_{I_t J_t K_t} + \dot{\vec{r}}_{SiteIJK}$$

$$\dot{\vec{\rho}}_{I_t J_t K_t} = \begin{bmatrix} \dot{\rho} \cos(\delta_t) \cos(\alpha_t) - \rho \sin(\delta_t) \cos(\alpha_t) \dot{\delta}_t - \rho \cos(\delta_t) \sin(\alpha_t) \dot{\alpha}_t \\ \dot{\rho} \cos(\delta_t) \sin(\alpha_t) - \rho \sin(\delta_t) \sin(\alpha_t) \dot{\delta}_t + \rho \cos(\delta_t) \cos(\alpha_t) \dot{\alpha}_t \\ \dot{\rho} \sin(\delta_t) + \rho \cos(\delta_t) \dot{\delta}_t \end{bmatrix}$$

$$\begin{aligned}\dot{\vec{v}}_{SiteIJK} &= \dot{\vec{\omega}}_{\oplus} \times \dot{\vec{r}}_{SiteIJK} \\ \dot{\vec{v}}_{IJK} &= \dot{\vec{\rho}}_{I_t J_t K_t} + \dot{\vec{v}}_{SiteIJK}\end{aligned}$$

The reverse process begins by determining the slant-range vector, $\dot{\vec{\rho}}_{IJK}$, from the site to the satellite based on given information. Because the resulting vector is in the geocentric frame, no additional calculations are necessary, and we determine the topocentric values directly. Only the sine expression is required to uniquely determine the topocentric declination. We determine the rates as in Algorithm 12 but we use the slant-range vectors.

ALGORITHM 14: *Topocentric ($\dot{\vec{r}}_{IJK}, \dot{\vec{v}}_{IJK}, \dot{\vec{r}}_{SiteIJK} \Rightarrow \rho, \alpha_t, \delta_t, \dot{\rho}, \dot{\alpha}_t, \dot{\delta}_t$)*

$$\dot{\vec{\rho}}_{IJK} = \dot{\vec{r}}_{IJK} - \dot{\vec{r}}_{SiteIJK}$$

$$\rho = |\dot{\vec{\rho}}|$$

$$\sin(\delta_t) = \frac{\rho_K}{\rho}$$

$$\left[\begin{array}{l} \text{IF } \sqrt{\rho_I^2 + \rho_J^2} \neq 0 \\ \\ \sin(\alpha_t) = \frac{\rho_J}{\sqrt{\rho_I^2 + \rho_J^2}} \quad \cos(\alpha_t) = \frac{\rho_I}{\sqrt{\rho_I^2 + \rho_J^2}} \\ \\ \text{ELSE} \\ \\ \sin(\alpha_t) = \frac{\dot{\rho}_J}{\sqrt{\dot{\rho}_I^2 + \dot{\rho}_J^2}} \quad \cos(\alpha_t) = \frac{\dot{\rho}_I}{\sqrt{\dot{\rho}_I^2 + \dot{\rho}_J^2}} \end{array} \right]$$

$$\dot{\vec{v}}_{SiteIJK} = \dot{\vec{\omega}}_{\oplus} \times \dot{\vec{r}}_{SiteIJK}$$

$$\dot{\vec{\rho}}_{IJK} = \dot{\vec{v}}_{IJK} - \dot{\vec{v}}_{SiteIJK}$$

$$\dot{\rho} = \frac{\dot{\rho} \cdot \dot{\rho}}{\rho}$$

$$\dot{\alpha}_t = \frac{\dot{\rho}_I \rho_J - \dot{\rho}_J \rho_I}{-\rho_J^2 - \rho_I^2}$$

$$\dot{\delta}_t = \frac{\dot{\rho}_K - \dot{\rho} \sin(\delta_t)}{\sqrt{\rho_I^2 + \rho_J^2}}$$

The most common transformation to other systems takes place with the geocentric right ascension and declination system. Carefully examining the problem reveals that we must know the slant range to accurately transform the coordinates. We don't need to go through the details here. See Roy (1988, 64–67) for approximate and rigorous techniques to determine the difference between the geocentric and topocentric values. Notice we must know the topocentric values because observations usually come from a particular site and are thus in the topocentric frame. Try not to use approximate formulas; inappropriate use invites problems. The rigorous formulas to determine geocentric values (r , α , δ) from given topocentric values and the site position magnitude (ρ , α_t , δ_t , r_{site}) (from Roy, 1988, 66–67) are as follows:

$$\tan(\alpha_t - \alpha) = \frac{\frac{r_{site}}{\rho} \cos(\phi_{gc}) \sin(\alpha_t - \theta_{LST})}{\cos(\delta_t) + \frac{r_{site}}{\rho} \cos(\phi_{gc}) \cos(\alpha_t - \theta_{LST})}$$

Roy identifies a temporary variable, γ , used in the following formulas:

$$\tan(\gamma) = \frac{\tan(\phi_{gc}) \cos\left(\frac{\alpha_t - \alpha}{2}\right)}{\cos\left(\frac{\alpha_t + \alpha}{2} - \theta_{LST}\right)}$$

$$\tan(\delta_t - \delta) = \frac{\frac{r_{site}}{\rho} \sin(\phi_{gc}) \sin(\delta_t - \gamma)}{\sin(\gamma) + \frac{r_{site}}{\rho} \sin(\phi_{gc}) \cos(\delta_t - \gamma)}$$

$$r = \frac{\rho}{\cos(\delta_t - \delta)} \left[1 + \frac{\frac{r_{site}}{\rho} \sin(\phi_{gc}) \cos(\delta_t - \gamma)}{\sin(\gamma)} \right]$$

A crucial data point is the slant range, ρ . If ρ is available, the preceding equations give the differences directly. If ρ is unknown, as in many optical systems, we must iterate because

we can't uniquely determine the values based solely on the angular data. Solution ultimately rests on choosing initial guesses and successively evaluating the equations. In this case, we have four equations (including γ) and four unknowns (ρ , r , α , and δ). More complex algorithms in subsequent chapters will often require iteration.

3.4.3 Azimuth-Elevation

Conversions with azimuth and elevation are similar to those for geocentric right ascension and declination but have some subtle differences. Consider first the problem of determining the position and velocity vectors from the range, ρ , azimuth, β , elevation, el , and their rates. As in the development of Eq. (3-1), begin with the position vector:

$$\hat{\rho}_{SEZ} = \begin{bmatrix} -\rho \cos(el) \cos(\beta) \\ \rho \cos(el) \sin(\beta) \\ \rho \sin(el) \end{bmatrix} \quad (3-3)$$

Notice the negative sign in the I axis component because, by definition, we measure azimuth positive clockwise from north. Some texts define azimuth differently and therefore have slightly different relations; be consistent! Also, we determine the result in the SEZ coordinate system, so we may need to rotate it to another system, such as the geocentric equatorial system, IJK . As before, find the velocity as the derivative of Eq. (3-3):

$$\dot{\hat{\rho}}_{SEZ} = \begin{bmatrix} -\dot{\rho} \cos(el) \cos(\beta) + \rho \sin(el) \cos(\beta) \dot{el} + \rho \cos(el) \sin(\beta) \dot{\beta} \\ \dot{\rho} \cos(el) \sin(\beta) - \rho \sin(el) \sin(\beta) \dot{el} + \rho \cos(el) \cos(\beta) \dot{\beta} \\ \dot{\rho} \sin(el) + \rho \cos(el) \dot{el} \end{bmatrix} \quad (3-4)$$

Application: Razel

In this section, I'm focusing on transformations of position and velocity vectors into range, azimuth, and elevation angles and their rates because they're quite common and occur in several places throughout this book. This process applies mainly to statistical orbit determination (Chap. 9), in which it's important to be able to determine the predicted value of the measurements as the state vector (position and velocity vectors) changes. Thus, the process finds the range, azimuth and elevation, and the associated rates, given the position and velocity vectors, the site, and the time of the observation.

Later we'll see the overall process simply reverses the **SITE-TRACK** (Algorithm 45) equations. Begin by finding the slant-range vector from the site to the satellite:

$$\hat{\rho}_{IJK} = \hat{r}_{IJK} - \hat{r}_{SiteIJK} \quad (3-5)$$

Notice that, if we need *only* the magnitude of the range, the process is complete at this point because the magnitude of the slant-range vector remains constant through the fol-

lowing rotations. We find the correction for the velocity vector due to the rotating coordinate frames as before. Subtract this quantity from the known velocity vector to find the slant-range velocity:

$$\dot{\vec{r}}_{IJK} = \vec{v}_{IJK} - \vec{\omega}_{\oplus} \times \vec{r}_{IJK} \quad (3-6)$$

The slant range is in the geocentric equatorial system at this point, so we have to rotate it to the topocentric-horizon system to determine azimuth and elevation. Figure 3-2 shows this transformation along with the azimuth and elevation.

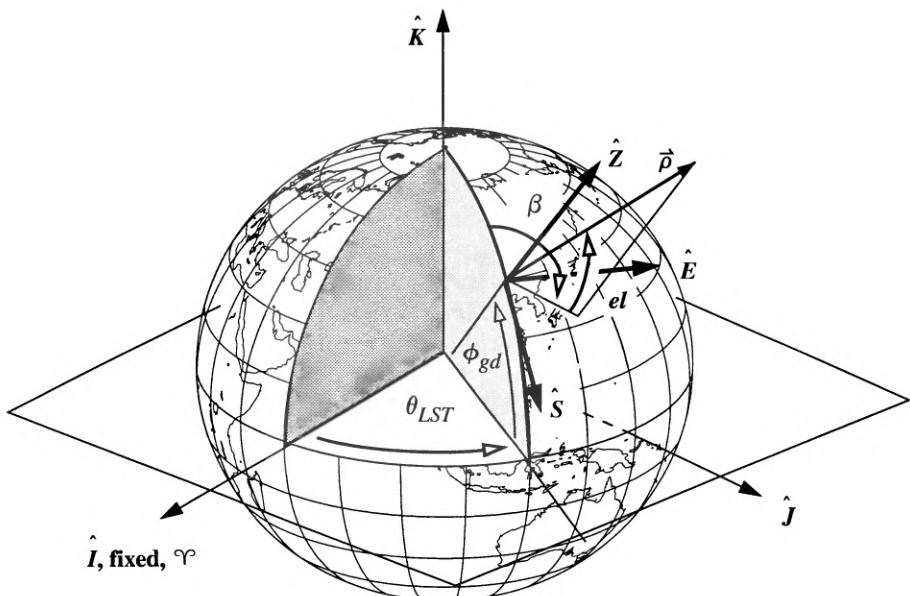


Figure 3-2. Topocentric-Horizon Coordinate System (SEZ). We measure azimuth, β , and elevation, el , in the topocentric-horizon coordinate system.

This process uses the relations in Eq. (1-33):

$$\begin{aligned}\vec{\rho}_{SEZ} &= [\text{ROT2}(90^\circ - \phi_{gd})] [\text{ROT3}(\theta_{LST})] \vec{\rho}_{IJK} \\ \dot{\vec{\rho}}_{SEZ} &= [\text{ROT2}(90^\circ - \phi_{gd})] [\text{ROT3}(\theta_{LST})] \dot{\vec{\rho}}_{IJK}\end{aligned}$$

The next task is to find the individual observation quantities (range, azimuth and elevation). Once we know the slant-range vector, the magnitude is available. Determine the elevation from looking at Fig. 3-2. The sine permits unique determination because elevation varies from -90° to 90° . Both expressions are given for completeness.

$$\sin(\text{el}) = \frac{\rho_Z}{\rho} \quad \cos(\text{el}) = \frac{\sqrt{\rho_S^2 + \rho_E^2}}{\rho}$$

We also find the azimuth from Fig. 3-2. Due to the direction of positive measurement (positive clockwise from north), the relation is somewhat different from the formulas for right ascension (Algorithm 12).

IF Elevation $\neq 90^\circ$

$$\sin(\beta) = \frac{\rho_E}{\sqrt{\rho_S^2 + \rho_E^2}} \quad \cos(\beta) = \frac{-\rho_S}{\sqrt{\rho_S^2 + \rho_E^2}}$$

IF Elevation $= 90^\circ$

$$\sin(\beta) = \frac{\dot{\rho}_E}{\sqrt{\dot{\rho}_S^2 + \dot{\rho}_E^2}} \quad \cos(\beta) = \frac{-\dot{\rho}_S}{\sqrt{\dot{\rho}_S^2 + \dot{\rho}_E^2}}$$

Notice several things. First, we need both the sine and cosine expressions to correctly determine the quadrant. The case for 90° elevation is provided to permit continuous coverage through all possible combinations of data. For the case of a satellite directly over the site, we can't determine azimuth unless the velocity vector is available.

Find the rate terms by rearranging relations in Eq. (3-4). The range rate is easiest to find, but it requires knowledge of the discussion of rates in Sec. 2.2.3.

$$\dot{\rho} = \frac{\dot{\rho} \cdot \dot{\rho}}{\rho} \quad (3-7)$$

We find the elevation rate by examining the Z component in Eq. (3-4). Specifically,

$$\dot{el} = \frac{\dot{\rho}_Z - \dot{\rho} \sin(el)}{\rho \cos(el)}$$

Recognize that $\rho \cos(el)$ is also $\sqrt{\rho_S^2 + \rho_E^2}$; thus, the rate of change of elevation is

$$\dot{el} = \frac{\dot{\rho}_Z - \dot{\rho} \sin(el)}{\sqrt{\rho_S^2 + \rho_E^2}} \quad (3-8)$$

The rate of change in azimuth is a more interesting task. Careful examination of Eq. (3-4) shows we may be able to eliminate some terms in the S and E components of velocity. Consider multiplying $\dot{\rho}_S \rho_E - \dot{\rho}_E \rho_S$. If we substitute terms shown in Eq. (3-4),

$$\begin{aligned}\dot{\rho}_S \rho_E - \dot{\rho}_E \rho_S &= \{-\dot{\rho} \cos(\text{el}) \cos(\beta) + \rho \sin(\text{el}) \cos(\beta) \dot{\text{el}} + \rho \cos(\text{el}) \sin(\beta) \dot{\beta}\} \rho_E \\ &\quad - \{\dot{\rho} \cos(\text{el}) \sin(\beta) - \rho \sin(\text{el}) \sin(\beta) \dot{\text{el}} + \rho \cos(\text{el}) \cos(\beta) \dot{\beta}\} \rho_S\end{aligned}$$

If we use the expressions from Eq. (3-3), the terms involving $\dot{\rho}$ and $\dot{\text{el}}$ in the expression above cancel. Also, recognize that the coefficients for the rate-of-azimuth term are the east and south components of the position vector. The remaining expression is

$$\dot{\rho}_S \rho_E - \dot{\rho}_E \rho_S = (\rho_S^2 + \rho_E^2) \dot{\beta}$$

and solving for the rate of change of azimuth gives us

$$\dot{\beta} = \frac{\dot{\rho}_S \rho_E - \dot{\rho}_E \rho_S}{\rho_S^2 + \rho_E^2} \quad (3-9)$$

Again, there are singular points where the *S* and *E* components of the slant-range vector are zero. Unless the acceleration vector is available, we have no other vectors to examine, so we must adopt a convention to arbitrarily assign a value for this special case. I mention this only for completeness; we're usually interested only in the ρ , β , and *el* values discussed previously.

Implementing the RAZEL Algorithm

Algorithm 15 shows the complete result. We can compute efficiently by saving trigonometric variables in temporary locations. The square root of the range-vector component also helps us eliminate operations.

ALGORITHM 15: RAZEL

$$(\hat{r}_{IJK}, \hat{v}_{IJK}, \hat{r}_{SiteIJK}, \phi_{gd}, \theta_{LST} \Rightarrow \rho, \beta, el, \dot{\rho}, \dot{\beta}, \dot{el})$$

$$\hat{\rho}_{IJK} = \hat{r}_{IJK} - \hat{r}_{SiteIJK}$$

$$\dot{\hat{\rho}}_{IJK} = \hat{v}_{IJK} - \hat{\omega}_{\oplus} \times \hat{r}_{IJK}$$

$$\dot{\hat{\rho}}_{SEZ} = [\text{ROT2}(90^\circ - \phi_{gd})] [\text{ROT3}(\theta_{LST})] \dot{\hat{\rho}}_{IJK}$$

$$\dot{\hat{\rho}}_{SEZ} = \underbrace{[\text{ROT2}(90^\circ - \phi_{gd})] [\text{ROT3}(\theta_{LST})]}_{\begin{bmatrix} \text{SIN}(\phi_{gd}) \text{COS}(\theta_{LST}) & \text{SIN}(\phi_{gd}) \text{SIN}(\theta_{LST}) & -\text{COS}(\phi_{gd}) \\ -\text{SIN}(\theta_{LST}) & \text{COS}(\theta_{LST}) & 0 \\ \text{COS}(\phi_{gd}) \text{COS}(\theta_{LST}) & \text{COS}(\phi_{gd}) \text{SIN}(\theta_{LST}) & \text{SIN}(\phi_{gd}) \end{bmatrix}} \dot{\hat{\rho}}_{IJK}$$

$$\rho = |\vec{\rho}_{SEZ}|$$

$$\sin(ei) = \frac{\rho_Z}{\rho}$$

IF Elevation $\neq 90^\circ$

$$\sin(\beta) = \frac{\rho_E}{\sqrt{\rho_S^2 + \rho_E^2}} \quad \cos(\beta) = \frac{-\rho_S}{\sqrt{\rho_S^2 + \rho_E^2}}$$

IF Elevation $= 90^\circ$

$$\sin(\beta) = \frac{\dot{\rho}_E}{\sqrt{\dot{\rho}_S^2 + \dot{\rho}_E^2}} \quad \cos(\beta) = \frac{-\dot{\rho}_S}{\sqrt{\dot{\rho}_S^2 + \dot{\rho}_E^2}}$$

$$\dot{\rho} = \frac{\dot{\rho} \cdot \dot{\rho}}{\rho}$$

$$\dot{\beta} = \frac{\dot{\rho}_S \rho_E - \dot{\rho}_E \rho_S}{\rho_S^2 + \rho_E^2}$$

$$\dot{el} = \frac{\dot{\rho}_Z - \dot{\rho} \sin(ei)}{\sqrt{\dot{\rho}_S^2 + \dot{\rho}_E^2}}$$

RAZEL simply reverses **SITE-TRACK** (Algorithm 45), so the test cases are just the reverse of those for **SITE-TRACK**. See Example 3-7 following this section.

Additional Az-El Changes

Transformations between topocentric right ascension and declination values with azimuth and elevation are sometimes useful, especially when range information is unavailable. To develop the relations, consider Fig. 3-3, in which the spherical triangle shows all the necessary angles. (See Sec. C.3.2 for a discussion of spherical trigonometry.) Notice we must perform additional processing if we need to use geocentric values, but it's impossible to measure "true" geocentric angles from the Earth's center.

Suppose we want to find azimuth and elevation values when we know topocentric right ascension and declination. The first task is to define the *LHA* of the object, which from Chap. 1 is

$$LHA = \theta_{LST} - \alpha_t$$

The cosine law for sides will yield the desired result for elevation:

$$\begin{aligned} \cos(90^\circ - el) &= \cos(90^\circ - \phi_{gc}) \cos(90^\circ - \delta_t) \\ &\quad + \sin(90^\circ - \phi_{gc}) \sin(90^\circ - \delta_t) \cos(LHA) \end{aligned}$$

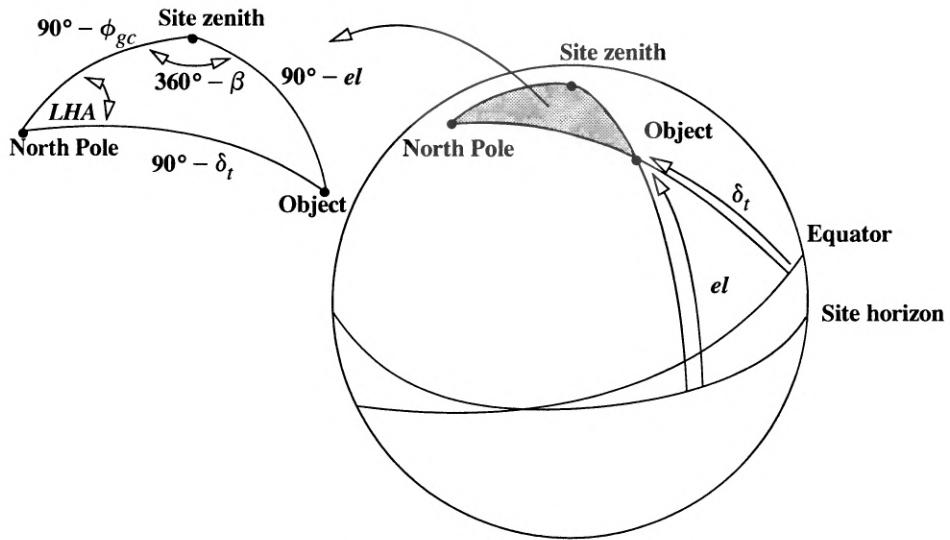


Figure 3-3. Az-el and Ra-dec Geometry. The spherical triangles shown are necessary to develop relations between the two sets of angles. Don't confuse the site *horizon* with the equator shown in many diagrams.

Simplifying gives us

$$\sin(e) = \sin(\phi_{gc}) \sin(\delta_t) + \cos(\phi_{gc}) \cos(\delta_t) \cos(LHA)$$

The sine expression is all we need because it will produce answers within the $\pm 90^\circ$ range of possible elevation values. The azimuth will require both sine and cosine expressions because it may assume values from 0° to 360° . Use the sine law to determine

$$\sin(\beta) = -\frac{\sin(LHA) \cos(\delta_t) \cos(\phi_{gc})}{\cos(e) \cos(\phi_{gc})}$$

in which we've multiplied an extra cosine of latitude on the top and bottom to maintain consistency with the cosine expression. Use the cosine law to find the remaining azimuth expression:

$$\cos(\beta) = \frac{\sin(\delta_t) - \sin(e) \sin(\phi_{gc})}{\cos(e) \cos(\phi_{gc})}$$

Now, summarize the algorithm:

ALGORITHM 16: RadectoAzel ($\alpha_r, \delta_r, \phi_{gc}, \theta_{LST} \Rightarrow \beta, el$)

$$\begin{aligned} LHA &= \theta_{LST} - \alpha_r \\ \sin(el) &= \sin(\phi_{gc}) \sin(\delta_r) + \cos(\phi_{gc}) \cos(\delta_r) \cos(LHA) \\ \sin(\beta) &= -\frac{\sin(LHA) \cos(\delta_r) \cos(\phi_{gc})}{\cos(el) \cos(\phi_{gc})} \\ \cos(\beta) &= \frac{\sin(\delta_r) - \sin(el) \sin(\phi_{gc})}{\cos(el) \cos(\phi_{gc})} \end{aligned}$$

Form the reverse problem similarly. Rearrange the cosine of the azimuth to find the sine of the declination. Find the right ascension using the *LHA*. The sine expression comes from the sine of the azimuth relation (Algorithm 16), keeping an extra latitude term. Find the cosine expression from the sine of the elevation in Algorithm 16. The algorithm is

ALGORITHM 17: AzeltoRadec ($\beta, el, \phi_{gc}, \theta_{LST} \Rightarrow \alpha_r, \delta_r$)

$$\begin{aligned} \sin(\delta_r) &= \sin(el) \sin(\phi_{gc}) + \cos(el) \cos(\phi_{gc}) \cos(\beta) \\ \sin(LHA) &= -\frac{\sin(\beta) \cos(el) \cos(\phi_{gc})}{\cos(\delta_r) \cos(\phi_{gc})} \\ \cos(LHA) &= \frac{\sin(el) - \sin(\phi_{gc}) \sin(\delta_r)}{\cos(\delta_r) \cos(\phi_{gc})} \\ \alpha_r &= \theta_{LST} - LHA \end{aligned}$$

3.4.4 Transformations for Ecliptic Latitude and Longitude

Representing values in the ecliptic latitude and longitude system is the starting point for determining position vectors of the Sun and Moon, which are very important for perturbation analyses. Find the position vector in the heliocentric coordinate system by looking at Fig. 3-4. Because the geocentric and heliocentric (ecliptic) systems differ only by the obliquity of the ecliptic, it's simple to transform between the systems. [Note the similarity to Eq. (3-1)].

$$\vec{r}_{XYZ} = \begin{bmatrix} r \cos(\phi_{ecliptic}) \cos(\lambda_{ecliptic}) \\ r \cos(\phi_{ecliptic}) \sin(\lambda_{ecliptic}) \\ r \sin(\phi_{ecliptic}) \end{bmatrix}$$

$$\dot{\vec{r}}_{IJK} = \text{ROT1}[-\epsilon] \dot{\vec{r}}_{XYZ}$$

To determine the position vector of any object with position magnitude, r , use

$$\vec{r}_{IJK} = r \begin{bmatrix} \cos(\phi_{ecliptic}) \cos(\lambda_{ecliptic}) \\ \cos(\epsilon) \cos(\phi_{ecliptic}) \sin(\lambda_{ecliptic}) - \sin(\epsilon) \sin(\phi_{ecliptic}) \\ \sin(\epsilon) \cos(\phi_{ecliptic}) \sin(\lambda_{ecliptic}) + \cos(\epsilon) \sin(\phi_{ecliptic}) \end{bmatrix} \quad (3-10)$$

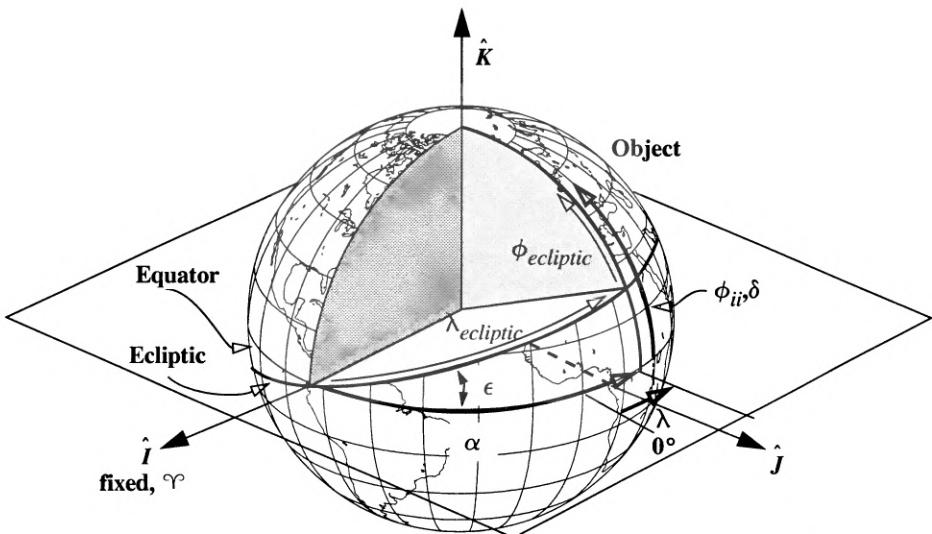


Figure 3-4. **Transforming the Ecliptic and Equatorial Systems.** Ecliptic coordinates and geocentric coordinates differ only by the obliquity of the ecliptic, ϵ . This permits a simple rotation between systems.

Solve the problem of converting to right ascension and declination values by looking at unit vectors in each system. In the ecliptic system, we find a unit vector in the same way as the position vector without the magnitude:

$$\hat{L}_{XYZ} = \begin{bmatrix} \cos(\phi_{ecliptic}) \cos(\lambda_{ecliptic}) \\ \cos(\phi_{ecliptic}) \sin(\lambda_{ecliptic}) \\ \sin(\phi_{ecliptic}) \end{bmatrix}$$

Likewise, in the geocentric system, use the right ascension and declination:

$$\hat{L}_{XYZ} = \begin{bmatrix} \cos(\delta) \cos(\alpha) \\ \cos(\delta) \sin(\alpha) \\ \sin(\delta) \end{bmatrix} \quad (3-11)$$

Now rotate the geocentric system to the ecliptic system (or vice versa) depending on the desired formulation for the problem:

$$\hat{L}_{XYZ} = \text{ROT1}(\epsilon) \hat{L}_{IJK} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\epsilon) & \sin(\epsilon) \\ 0 & -\sin(\epsilon) & \cos(\epsilon) \end{bmatrix} \begin{bmatrix} \cos(\delta) \cos(\alpha) \\ \cos(\delta) \sin(\alpha) \\ \sin(\delta) \end{bmatrix}$$

This process yields an immediate way to transform the values. Equating gives us

$$\cos(\phi_{ecliptic}) \cos(\lambda_{ecliptic}) = \cos(\delta) \cos(\alpha)$$

$$\cos(\phi_{ecliptic}) \sin(\lambda_{ecliptic}) = \cos(\delta) \sin(\alpha) \cos(\epsilon) + \sin(\delta) \sin(\epsilon)$$

$$\sin(\phi_{ecliptic}) = -\cos(\delta) \sin(\alpha) \sin(\epsilon) + \sin(\delta) \cos(\epsilon)$$

and because the ecliptic latitude is in the range of $\pm 90^\circ$, the arcsine is sufficient. The sine and cosine expressions are given for the longitude to resolve the quadrant:

$$\begin{aligned} \sin(\phi_{ecliptic}) &= -\cos(\delta) \sin(\alpha) \sin(\epsilon) + \sin(\delta) \cos(\epsilon) \\ \sin(\lambda_{ecliptic}) &= \frac{\cos(\delta) \sin(\alpha) \cos(\epsilon) + \sin(\delta) \sin(\epsilon)}{\cos(\phi_{ecliptic})} \\ \cos(\lambda_{ecliptic}) &= \frac{\cos(\delta) \cos(\alpha)}{\cos(\phi_{ecliptic})} \end{aligned} \quad (3-12)$$

Solve the inverse problem by transforming the ecliptic unit vector, \hat{L}_{XYZ} , to the equatorial frame ($\text{ROT1}(-\epsilon)$) and equating. The result is

$$\cos(\delta) \cos(\alpha) = \cos(\phi_{ecliptic}) \cos(\lambda_{ecliptic})$$

$$\cos(\delta) \sin(\alpha) = -\sin(\phi_{ecliptic}) \sin(\epsilon) + \cos(\phi_{ecliptic}) \cos(\epsilon) \sin(\lambda_{ecliptic})$$

$$\sin(\delta) = \sin(\phi_{ecliptic}) \cos(\epsilon) + \cos(\phi_{ecliptic}) \sin(\epsilon) \sin(\lambda_{ecliptic})$$

Remember to be careful of the quadrants for α ; that's why we present both sine and cosine expressions. Find the solution:

$$\begin{aligned} \sin(\delta) &= \sin(\phi_{ecliptic}) \cos(\epsilon) + \cos(\phi_{ecliptic}) \sin(\epsilon) \sin(\lambda_{ecliptic}) \\ \sin(\alpha) &= \frac{-\sin(\phi_{ecliptic}) \sin(\epsilon) + \cos(\phi_{ecliptic}) \cos(\epsilon) \sin(\lambda_{ecliptic})}{\cos(\delta)} \\ \cos(\alpha) &= \frac{\cos(\phi_{ecliptic}) \cos(\lambda_{ecliptic})}{\cos(\delta)} \end{aligned} \quad (3-13)$$

Implementing Right Ascension and Declination

The right ascension and declination actually change over time due to the small motions of the equinox. To be very precise, coordinates should use reduction techniques from Sec. 1.7. An example will illustrate how we combine values.

▼ Example 3-7. Finding Right Ascension and Declination, Azimuth, Elevation, Ecliptic Lat, Lon.

GIVEN: Neptune on May 14, 1994

FIND: \vec{r} , \vec{v} , α , δ , α_p , δ_p , β , el , $\phi_{ecliptic}$, $\lambda_{ecliptic}$, and rates

The *Astronomical Almanac* provides data for our analysis. Rates of right ascension and declination aren't usually given, but we can derive them from data on the following day. On May 14, 1994,

$$r = 29.664\,361 \text{ AU} = 4,437,725,220.51 \text{ km}$$

$$\alpha = 19^\circ 39' 57.395'' = 294.989\,145^\circ$$

$$\delta = -20^\circ 49' 24.58'' = -20.823\,494\,4^\circ$$

Using data from May 14 to May 15, 1994 gives us

$$\dot{r} = \frac{(149,598,023) (29.649\,616 - 29.664\,361)}{86,400} = -25.530\,330\,94 \text{ km/s}$$

$$\dot{\alpha} = \frac{15 (54.856 - 57.395) \text{ s}}{(3600) (86,400)} = -0.000\,000\,122\,44^\circ/\text{s}$$

$$\dot{\delta} = \frac{(30.16 - 24.58)}{(3600) (86,400)} = -0.000\,000\,017\,94^\circ/\text{s}$$

Eq. (3-1) and Eq. (3-2) allow us to determine position and velocity vectors:

$$\vec{r}_{IJK} = \begin{bmatrix} r \cos(\delta) \cos(\alpha) \\ r \cos(\delta) \sin(\alpha) \\ r \sin(\delta) \end{bmatrix} = \begin{bmatrix} 1,752,246,215 \\ -3,759,563,433 \\ -1,577,568,105 \end{bmatrix} \text{ km}$$

$$\vec{v}_{IJK} = \begin{bmatrix} \dot{r} \cos(\delta) \cos(\alpha) - r \sin(\delta) \cos(\alpha) \dot{\delta} - r \cos(\delta) \sin(\alpha) \dot{\alpha} \\ \dot{r} \cos(\delta) \sin(\alpha) - r \sin(\delta) \sin(\alpha) \dot{\delta} + r \cos(\delta) \cos(\alpha) \dot{\alpha} \\ \dot{r} \sin(\delta) + r \cos(\delta) \dot{\delta} \end{bmatrix} = \begin{bmatrix} -18.324 \\ 18.332 \\ 7.777 \end{bmatrix} \text{ km/s}$$

If we know a particular ground site, we can do the topocentric calculations. Consider a site at the U.S. Air Force Academy having these coordinates:

$$\phi_{gc} = 39.007^\circ, \lambda = -104.883^\circ, h_{ellp} = 2194.56 \text{ m}$$

The epoch time of the date permits evaluation of LST , which we need for site calculations. For this example, consider

$$LST = -35.0^\circ$$

Using Algorithm 45 (I'll present it later), find the site vector:

$$\vec{r}_{siteIJK} = \begin{bmatrix} 4066.716 \\ -2847.545 \\ 3994.302 \end{bmatrix} \text{ km}$$

Because this section requires many formulas, I'll just present results, using the terms *lateral* for right ascension, azimuth, and longitude, and *vertical* for declination, elevation, and latitude.

	range (km)	lateral (°)	vertical (°)
Geocentric	$r = 4,437,725,220.511$	$\alpha = 294.989\ 145\ 8$	$\delta = -20.823\ 494\ 4$
Topocentric	$\rho = 4,437,722,622.307$	$\alpha_t = 294.989\ 111\ 5$	$\delta_t = -20.823\ 562\ 3$
Horizon	$\rho = 4,437,722,622.307$	$\beta = 210.753\ 732\ 5$	$el = 23.902\ 574\ 5$
Ecliptic	$r = 4,437,725,220.511$	$\lambda_{ecliptic} = 293.258\ 210\ 8$	$\phi_{ecliptic} = 0.620\ 750\ 5$

Note the geocentric and topocentric right ascensions are very close but different. Likewise, the declination values are similar due to the extreme range. Azimuth and elevation are typically different from their analogous counterparts, whereas ecliptic latitude and longitude are different from all the rest. The rate terms also have similarities:

	range rate (km/s)	lateral rate (°/s)	vertical rate (°/s)
Geocentric	$\dot{r} = -25.530\ 330\ 9$	$\dot{\alpha} = -0.000\ 000\ 122\ 444$	$\dot{\delta} = -0.000\ 000\ 017\ 940$
Topocentric	$\dot{\rho} = -25.530\ 324\ 3$	$\dot{\alpha}_t = -0.000\ 000\ 122\ 444$	$\dot{\delta}_t = -0.000\ 000\ 017\ 940$
Horizon	$\dot{\rho} = -25.361\ 082\ 1$	$\dot{\beta} = 0.003\ 866\ 399\ 952$	$\dot{el} = -0.001\ 660\ 238\ 846$
Ecliptic	$\dot{r} = -25.530\ 330\ 9$	$\dot{\lambda}_{ecliptic} = -0.000\ 000\ 115\ 840$	$\dot{\phi}_{ecliptic} = 0.000\ 000\ 001\ 548$

Because there are so many similarities, we may question the need for so many representations. We can answer this good question by recalculating the entire problem assuming a satellite at a range of 12,756 km (2.0 ER) with a velocity of 6.798 614 km/s (0.86 ER/TU). Because the satellite orbits at a much closer range, we expect the values to differ more dramatically. The results are

	range (km)	lateral (°)	vertical (°)
Geocentric	$r = 12,756.273$	$\alpha = 294.989\ 145\ 83$	$\delta = -20.823\ 494\ 4$
Topocentric	$\rho = 11,706.270$	$\alpha_t = 276.949\ 244\ 50$	$\delta_t = -46.768\ 043\ 9$
Horizon	$\rho = 11,706.270$	$\beta = 210.806\ 346\ 93$	$el = -5.899\ 579\ 9$
Ecliptic	$r = 12,756.273$	$\lambda_{ecliptic} = 293.258\ 210\ 85$	$\phi_{ecliptic} = 0.620\ 750\ 4$

The rates are

	range rate (km/s)	lateral rate (°/s)	vertical rate (°/s)
Geocentric	$\dot{r} = 6.798\ 614\ 888$	$\dot{\alpha} = -0.000\ 000\ 122\ 444$	$\dot{\delta} = -0.000\ 000\ 017\ 940$
Topocentric	$\dot{\rho} = 5.899\ 476\ 243$	$\dot{\alpha}_t = 0.014\ 061\ 544\ 168$	$\dot{\delta}_t = 0.013\ 444\ 113\ 526$
Horizon	$\dot{\rho} = 6.083\ 898\ 607$	$\dot{\beta} = 0.003\ 843\ 602\ 081$	$\dot{el} = 0.014\ 968\ 654\ 960$
Ecliptic	$\dot{r} = 6.798\ 614\ 888$	$\dot{\lambda}_{ecliptic} = -0.000\ 000\ 115\ 840$	$\dot{\phi}_{ecliptic} = 0.000\ 000\ 001\ 548$

▲ Notice how the differences between geocentric and topocentric values are more pronounced for the satellite case than for the example using stars.

3.5 Application: Sun Position Vector

We often want to find the position vector from the Earth to the Sun for uses such as solar-panel illumination and viewing. Very precise ephemerides are available through the Jet Propulsion Laboratory (see Appendix D); however, it's often more convenient to place a mathematical algorithm inside the particular program using a less precise formula from the *Astronomical Almanac* (1992, C24). The accuracy is given as 0.01° and is valid from 1950 to 2050 because of the truncation of the expansions; however, because many of the formulas have been expanded slightly, the accuracy should always exceed these values. Meeus (1991, 154–164) shows a more rigorous method which yields much higher accuracy. Using higher-precision algorithms requires us to examine the epoch and equinox. Meeus discusses how to select and calculate vectors referencing a particular epoch. Figure 3-5 shows the geometry necessary to visualize the problem.

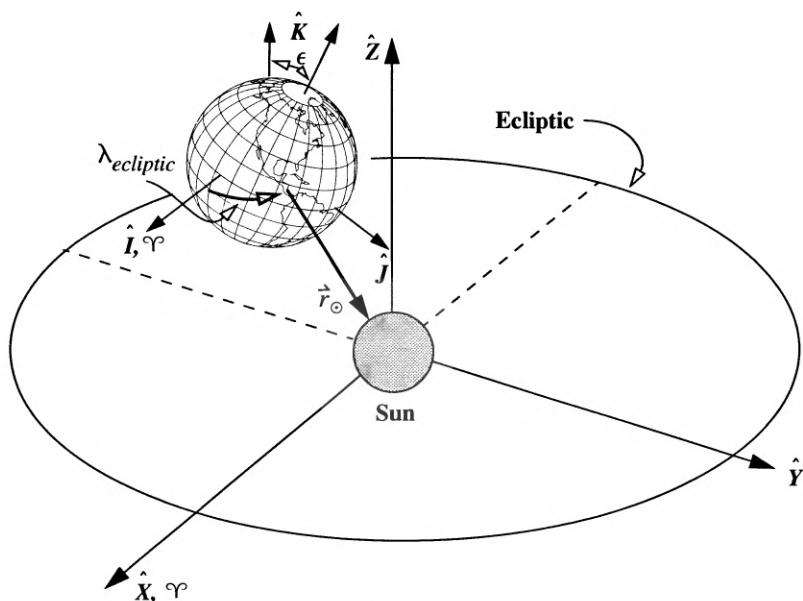


Figure 3-5. Geometry for the Sun Position Vector. Solution of the position vector to the Sun rests on determining the ecliptic longitude (ecliptic latitude is 0°) and the range. The X and I axes are parallel.

The solution rests in part on expansions derived from the J2000 epoch. First, find the number of Julian centuries, T_{UT1} , from the epoch (J2000), using Eq. (1-50). Find the mean longitude of the Sun, λ_{M_\odot} , similarly to Eq. (1-44), but with a 180° offset because the Sun—not the Earth—is the point of interest.

$$\lambda_{M_\odot} = 280.460\,618\,4^\circ + 36,000.770\,053\,61T_{UT1}$$

Find the mean anomaly (see Sec. 4.2) for the Sun, M_{\odot} , from Eq. (1-59). We show T_{TDB} to be precise, but it's quite common to use T_{UT1} because this is only a low-precision formula.

$$M_{\odot} = 357.527\,723\,3^\circ + 35,999.050\,34T_{TDB}$$

Reduce both $\lambda_{M_{\odot}}$ and M_{\odot} to the range of 0° to 360° . Approximate the ecliptic longitude and latitude by applying the equation of center [see Eq. (4-51)]:

$$\nu = M_{\odot} + 2e \sin(M_{\odot}) + \frac{5e^2}{4} \sin(2M_{\odot}) + \dots$$

where we use the eccentricity of the Earth's orbit ($e_e = 0.016\,708\,617$) around the Sun. Don't confuse this term with the eccentricity used to determine site locations, e_{\oplus} . Meeus (1991, 151) shows a more accurate expression for the eccentricity of the Earth's orbit:

$$e_e = 0.016\,708\,617 - 0.000\,042\,037T_{TDB} - 0.000\,000\,123\,6T_{TDB}^2$$

This algorithm does *not* use this variable formula for eccentricity because it's so small, but I'll include it for completeness. Also, because the Earth's orbit is approximately circular, we can assume the true anomaly is close enough to the longitude for the following expression to hold (remember to convert to units of degrees!). The ecliptic latitude of the Sun never exceeds $0.000\,333^\circ$ and is almost always set to 0.0° .

$$\begin{aligned}\lambda_{ecliptic} &= \lambda_{M_{\odot}} + 1.914\,666\,471^\circ \sin(M_{\odot}) + 0.019\,994\,643 \sin(2M_{\odot}) \\ \phi_{ecliptic} &= 0^\circ\end{aligned}$$

This equation contains important information. The trigonometric terms in the ecliptic longitude expression are part of the equation of time, introduced in Chap. 1 [Eq. (1-48)].

We can approximate the obliquity of the ecliptic using only the first two terms in Eq. (1-58) and letting $T_{TDB} \approx T_{UT1}$. Thus,

$$\epsilon = 23.439\,291^\circ - 0.013\,004\,2T_{TDB}$$

Find the distance in AU from the Earth to the Sun using an expansion of elliptic motion (Taff, 1985, 60):

$$\begin{aligned}r &= a \left[1 + \frac{e^2}{2} + \left\{ -e + \frac{3e^3}{8} - \frac{5e^5}{192} + \frac{7e^7}{9216} + \dots \right\} \cos(M) \right. \\ &\quad + \left\{ -\frac{e^2}{2} + \frac{e^4}{3} - \frac{e^6}{16} + \dots \right\} \cos(2M) + \left\{ -\frac{3e^3}{8} + \frac{45e^5}{128} - \frac{567e^7}{5120} + \dots \right\} \cos(3M) \\ &\quad + \left\{ -\frac{e^4}{3} + \frac{2e^6}{5} + \dots \right\} \cos(4M) + \left\{ -\frac{125e^5}{384} + \frac{4375e^7}{9216} + \dots \right\} \cos(5M) \\ &\quad \left. + \left\{ -\frac{27e^6}{80} + \dots \right\} \cos(6M) + \left\{ -\frac{16087e^7}{46,080} + \dots \right\} \cos(7M) \right]\end{aligned}$$

Find the position magnitude using values of the Earth from Table D-3 ($a = 1.000\,001\,001\,78$ AU) and $e = 0.016\,708\,617$:

$$r_{\odot} = 1.000\,140\,612 - 0.016\,708\,617 \cos(M_{\odot}) - 0.000\,139\,589 \cos(2M_{\odot})$$

If we use Eq. (3-10) and remember that the ecliptic latitude of the Sun is zero, the position vector in geocentric equatorial coordinates is

$$\vec{r}_{\odot} = r_{\odot} \cos(\lambda_{ecliptic}) \hat{i} + r_{\odot} \cos(\epsilon) \sin(\lambda_{ecliptic}) \hat{j} + r_{\odot} \sin(\epsilon) \sin(\lambda_{ecliptic}) \hat{k}$$

We find the right ascension and declination using Eq. (3-13) and other known values as necessary. These calculations add several mathematical operations, but they're unnecessary if we want only to get a position vector for the Sun.

ALGORITHM 18: Sun ($JD_{UT1} \Rightarrow \vec{r}_{\odot}$)

$$T_{UT1} = \frac{JD_{UT1} - 2,451,545.0}{36,525}$$

$$\lambda_{M_{\odot}} = 280.460\,618\,4^{\circ} + 36,000.770\,053\,61 T_{UT1}$$

$$\text{LET } T_{TDB} \equiv T_{UT1}$$

$$M_{\odot} = 357.527\,723\,3^{\circ} + 35,999.050\,34 T_{TDB}$$

$$\lambda_{ecliptic} = \lambda_{M_{\odot}} + 1.914\,666\,471^{\circ} \sin(M_{\odot}) + 0.019\,994\,643 \sin(2M_{\odot})$$

$$r_{\odot} = 1.000\,140\,612 - 0.016\,708\,617 \cos(M_{\odot}) - 0.000\,139\,589 \cos(2M_{\odot})$$

$$\epsilon = 23.439\,291^{\circ} - 0.013\,004\,2 T_{TDB}$$

$$\vec{r}_{\odot} = \begin{bmatrix} r_{\odot} \cos(\lambda_{ecliptic}) \\ r_{\odot} \cos(\epsilon) \sin(\lambda_{ecliptic}) \\ r_{\odot} \sin(\epsilon) \sin(\lambda_{ecliptic}) \end{bmatrix} \text{AU}$$

▼ **Example 3-8. Finding the Sun Position Vector.**

GIVEN: April 2, 1994, 12:00 UT1

FIND: Geocentric position vector of the Sun

Begin by finding the Julian date using Algorithm 1, $JD_{UT1} = 2,449,444.5$. The Julian centuries are

$$T_{UT1} = \frac{JD_{UT1} - 2,451,545.0}{36,525} = -0.057\,508\,555$$

Let $T_{TDB} \equiv T_{UT1}$. The mean longitude of the Sun and other parameters are

$$\lambda_{M_\odot} = 280.460\,618\,4^\circ + 36.000\,770\,053\,61 T_{UT1} = 10.108\,325^\circ$$

$$M_\odot = 357.527\,723\,3^\circ + 35.999.050\,34 T_{TDB} = 87.274\,328^\circ$$

$$\lambda_{ecliptic} = \lambda_{M_\odot} + 1.914\,666\,471^\circ \sin(M_\odot) + 0.019\,994\,643 \sin(2M_\odot) = 12.022\,725^\circ$$

$$\phi_{ecliptic} = 0^\circ$$

As a check, the *Astronomical Almanac* lists the values of ecliptic latitude as $0.12''$ or $0.000\,033\,3^\circ$ and longitude as $12^\circ 01' 20.39''$ or $12.022\,330\,56^\circ$. The obliquity of the ecliptic is

$$\epsilon = 23.439\,291^\circ - 0.013\,004\,2 T_{TDB} = 23.440\,038\,8^\circ$$

The magnitude of the distance to the Sun is

$$r_\odot = 1.000\,140\,612 - 0.016\,708\,617 \cos(M_\odot) - 0.000\,139\,589 \cos(2M_\odot) = 0.999\,485\,0 \text{ AU}$$

Now find the vector (referred to the epoch of date).

$$\vec{r}_\odot = \begin{bmatrix} r_\odot \cos(\lambda_{ecliptic}) \\ r_\odot \cos(\epsilon) \sin(\lambda_{ecliptic}) \\ r_\odot \sin(\epsilon) \sin(\lambda_{ecliptic}) \end{bmatrix} = \begin{bmatrix} 0.977\,561\,4 \\ 0.191\,011\,7 \\ 0.082\,816\,7 \end{bmatrix} \text{ AU} = \begin{bmatrix} 146,241,097 \\ 28,574,940 \\ 12,389,196 \end{bmatrix} \text{ km}$$

The *Astronomical Almanac* lists the vector as

$$\vec{r}_\odot = \begin{bmatrix} 0.977\,276\,6 \\ 0.192\,263\,5 \\ 0.083\,361\,3 \end{bmatrix} \text{ AU} = \begin{bmatrix} 146,241,097 \\ 28,762,210 \\ 12,470,672 \end{bmatrix} \text{ km}$$

Find the right ascension and declination from Eq. (3-13). The formula is simplified because $\phi_{ecliptic} = 0^\circ$.

$$\tan(\alpha) = \cos(\epsilon) \tan(\lambda_{ecliptic}) = 11.056\,071^\circ$$

$$\sin(\delta) = \sin(\epsilon) \sin(\lambda_{ecliptic}) = 4.752\,939\,3^\circ$$

The *Astronomical Almanac* again lists the apparent values:

$$\alpha = 0\,h\,44\,\text{min}\,13.00\,\text{s} = 11.054\,166\,63^\circ$$

$$\delta = 4^\circ 45' 6.80'' = 4.751\,888\,89^\circ$$



3.6 Application: Moon Position Vector

We often need to find the position vector from the Earth to the Moon. The ephemerides of the Jet Propulsion Laboratory (see Appendix D) are usually regarded as the most accurate, but we can use a less precise formula from the *Astronomical Almanac* (1992, D46) and Green (1988, 173–174). The *Almanac* accuracy is given as 0.3° in ecliptic longitude, $\lambda_{ecliptic}$, 0.2° in ecliptic latitude, $\phi_{ecliptic}$, and 1275 km (0.2 ER) in distance. Consult Meeus (1991, 307–314) for more accurate algorithms. If you use *any* higher-precision the-

ory, the complexity of the Moon's motion becomes evident. Figure 3-6 is useful to visualize the transformations.

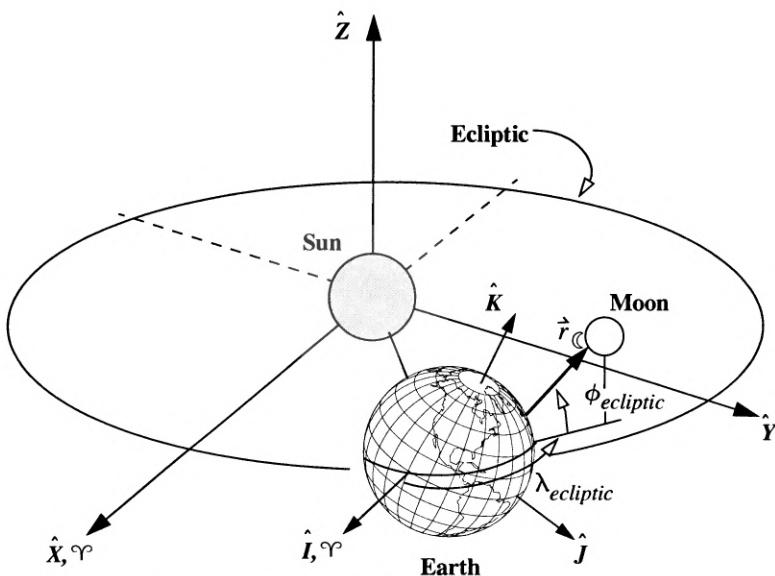


Figure 3-6. Geometry for the Moon Position Vector. Notice the ecliptic latitude is not zero for the Moon. The ecliptic coordinates determine the position vector. This figure is extremely exaggerated. X and I are parallel.

The Moon's motion is very complex, and exact formulas require a lot of computations. Ernest Brown (1866–1938) is credited with developing one of the best numerical methods to determine the position of the Moon. He did this in 1900! The actual theory consists of a series expansion involving hundreds of terms. The less precise formula in the *Astronomical Almanac* is based on Brown's theory but includes only the first few terms. Because this algorithm is so complex, I'll present the shortened equations and briefly describe their meanings. As with the Sun, we find the ecliptic latitude, longitude, and parallax using the following equations (Green, 1988, 174):

$$\begin{aligned}\lambda_{ecliptic} = & \lambda_{\zeta} + 6.29 \sin(M_{\zeta}) - 1.27 \sin(M_{\zeta} - 2D_{\odot}) + 0.66 \sin(2D_{\odot}) \\ & + 0.21 \sin(2M_{\zeta}) - 0.19 \sin(M_{\odot}) - 0.11 \sin(2u_{M_{\zeta}})\end{aligned}$$

$$\begin{aligned}\phi_{ecliptic} = & 5.13^\circ \sin(u_{M_{\zeta}}) + 0.28 \sin(M_{\zeta} + u_{M_{\zeta}}) \\ & - 0.28 \sin(u_{M_{\zeta}} - M_{\zeta}) - 0.17 \sin(u_{M_{\zeta}} - 2D_{\odot})\end{aligned}$$

$$\varphi = 0.9508^\circ + 0.0518 \cos(M_\zeta) + 0.0095 \cos(M_\zeta - 2D_\odot) \\ + 0.0078 \cos(2D_\odot) + 0.0028 \cos(2M_\zeta)$$

Find expressions for the mean anomalies, M_ζ , the mean argument of latitude, u_{M_ζ} , and elongation, D_\odot , from Eq. (1-59). Meeus (1991, 132) lists the Moon's longitude as

$$\lambda_\zeta = 218.32^\circ + 481,267.8813 T_{TDB}$$

Let's identify several terms by the source of their changes. $(2u_{M_\zeta})$ is due to the inclination of the Moon's orbit to the ecliptic ($\sim 5^\circ$), and $2D_\odot$ is the variation accounting for the Sun's gravitational attraction on the Moon. $M_\zeta - 2D_\odot$ is the perturbation term for the evection with a period of 31.8 days (there are periodic variations in eccentricity and longitude of perigee). Lastly, we recognize the constant term in the parallax expression (0.9508°) as the mean value of the horizontal parallax.

Begin by finding the number of Julian centuries from the epoch (J2000) using Eq. (1-50). Notice we use TDB exclusively for this algorithm. Eq. (1-59) gives the correct expressions for each of the parameters in the expansion for the Moon. Find the ecliptic latitude and longitude and parallax through the series expansions of Brown's Lunar Theory, as shown in the algorithm.

Find the distance in Earth radii from the Earth to the Moon using Eq. (1-69). Then, use Eq. (3-10) to find the position vector in Earth radii in the geocentric equatorial frame. This is similar to determining the Sun's position vector, but we need extra terms because a non-zero component of ecliptic latitude is present. The *Astronomical Almanac* replaces the sine and cosine values of the obliquity of the ecliptic with numerical terms rather than trigonometric terms. You may prefer these terms in some situations, but you'll get higher accuracy from the trigonometric expressions. Determine the obliquity of the ecliptic from Eq. (1-58) and then find

$$\vec{r}_\zeta = r_\zeta \begin{bmatrix} \cos(\phi_{ecliptic}) \cos(\lambda_{ecliptic}) \\ \cos(\epsilon) \cos(\phi_{ecliptic}) \sin(\lambda_{ecliptic}) - \sin(\epsilon) \sin(\phi_{ecliptic}) \\ \sin(\epsilon) \cos(\phi_{ecliptic}) \sin(\lambda_{ecliptic}) + \cos(\epsilon) \sin(\phi_{ecliptic}) \end{bmatrix}$$

As with the algorithm for the Sun position vector, we can determine the right ascension and declination if necessary using Eq. (3-13).

ALGORITHM 19: Moon ($JD_{TDB} \Rightarrow \vec{r}_\zeta$)

$$T_{TDB} = \frac{JD_{TDB} - 2,451,545.0}{36,525}$$

$$\begin{aligned}\lambda_{ecliptic} = & 218.32^\circ + 481,267.8813T_{TDB} + 6.29 \sin(134.9 + 477,198.85T_{TDB}) \\ & - 1.27 \sin(259.2 - 413,335.38T_{TDB}) + 0.66 \sin(235.7 + 890,534.23T_{TDB}) \\ & + 0.21 \sin(269.9 + 954,397.70T_{TDB}) - 0.19 \sin(357.5 + 35,999.05T_{TDB}) \\ & - 0.11 \sin(186.6 + 966,404.05T_{TDB})\end{aligned}$$

$$\begin{aligned}\phi_{ecliptic} = & 5.13^\circ \sin(93.3 + 483,202.03T_{TDB}) + 0.28 \sin(228.2 + 960,400.87T_{TDB}) \\ & - 0.28 \sin(318.3 + 6003.18T_{TDB}) - 0.17 \sin(217.6 - 407,332.20T_{TDB}) \\ \varphi = & 0.9508^\circ + 0.0518 \cos(134.9 + 477,198.85T_{TDB}) \\ & + 0.0095 \cos(259.2 - 413,335.38T_{TDB}) + 0.0078 \cos(235.7 + 890,534.23T_{TDB}) \\ & + 0.0028 \cos(269.9 + 954,397.70T_{TDB})\end{aligned}$$

$$r_{\mathbb{C}} = \frac{1}{\sin(\varphi)}$$

$$\dot{r}_{\mathbb{C}} = r_{\mathbb{C}} \left[\begin{array}{l} \cos(\phi_{ecliptic}) \cos(\lambda_{ecliptic}) \\ \cos(\epsilon) \cos(\phi_{ecliptic}) \sin(\lambda_{ecliptic}) - \sin(\epsilon) \sin(\phi_{ecliptic}) \\ \sin(\epsilon) \cos(\phi_{ecliptic}) \sin(\lambda_{ecliptic}) + \cos(\epsilon) \sin(\phi_{ecliptic}) \end{array} \right]$$

An example shows the process.

▼ Example 3-9. Finding the Moon Position Vector.

GIVEN: April 28, 1994, 12:00 UT1

FIND: Geocentric position vector of the Moon

For precision, we would rigorously find TDB using Algorithm 3, but for this example, let $TDB \cong UT1$. Next, determine the Julian date ($JD = 2,449,470.5$). The Julian centuries are

$$T_{TDB} = \frac{2,449,470.5 - 2,451,545.0}{36,525.0} = -0.057\ 508\ 555$$

$$\begin{aligned}\lambda_{ecliptic} = & 218.32^\circ + 481,267.8813T_{TDB} + 6.29 \sin(134.9 + 477,198.85T_{TDB}) \\ & - 1.27 \sin(259.2 - 413,335.38T_{TDB}) + 0.66 \sin(235.7 + 890,534.23T_{TDB}) \\ & + 0.21 \sin(269.9 + 954,397.70T_{TDB}) - 0.19 \sin(357.5 + 35,999.05T_{TDB}) \\ & - 0.11 \sin(186.6 + 966,404.05T_{TDB}) \\ = & -111.762\ 976\ 7^\circ = 248.237\ 023\ 3^\circ\end{aligned}$$

The ecliptic latitude isn't zero, so we find it as

$$\begin{aligned}\phi_{ecliptic} &= 5.13^\circ \sin(93.3 + 483,202.03T_{TDB}) + 0.28 \sin(228.2 + 960,400.87T_{TDB}) \\ &\quad - 0.28 \sin(318.3 + 6003.18T_{TDB}) - 0.17 \sin(217.6 - 407,332.20T_{TDB}) \\ &= 1.218\ 504\ 8^\circ\end{aligned}$$

The *Astronomical Almanac* lists $\lambda_{ecliptic} = 248.18^\circ$, and $\phi_{ecliptic} = 1.31^\circ$. The horizontal parallax is

$$\begin{aligned}\varphi &= 0.9508^\circ + 0.0518 \cos(134.9 + 477,198.85T_{TDB}) \\ &\quad + 0.0095 \cos(259.2 - 413,335.38T_{TDB}) + 0.0078 \cos(235.7 + 890,534.23T_{TDB}) \\ &\quad + 0.0028 \cos(269.9 + 954,397.70T_{TDB}) \\ &= 1.009\ 152\ 7^\circ\end{aligned}$$

The *Astronomical Almanac* lists $\varphi = 1.009\ 936\ 11^\circ$. Next, find the magnitude of the position vector:

$$r_{\mathbb{Q}} = \frac{1}{\sin(\varphi)} = 56.779\ 057\ 5\text{ ER}$$

$$\begin{aligned}\hat{r}_{\mathbb{Q}} &= r_{\mathbb{Q}} (\cos(\phi_{ecliptic}) \cos(\lambda_{ecliptic})) \hat{I} \\ &\quad r_{\mathbb{Q}} (\cos(\epsilon) \cos(\phi_{ecliptic}) \sin(\lambda_{ecliptic}) - \sin(\epsilon) \sin(\phi_{ecliptic})) \hat{J} \\ &\quad r_{\mathbb{Q}} (\sin(\epsilon) \cos(\phi_{ecliptic}) \sin(\lambda_{ecliptic}) + \cos(\epsilon) \sin(\phi_{ecliptic})) \hat{K}\end{aligned}$$

Now determine the position vector:

$$\hat{r}_{\mathbb{Q}} = \begin{bmatrix} -21,047\ 085\ 1 \\ -48,849\ 901\ 0 \\ -19,863\ 754\ 1 \end{bmatrix} \text{ER} = \begin{bmatrix} -134,241,192 \\ -311,571,362 \\ -126,693,745 \end{bmatrix} \text{km}$$

If right ascension and declination are required, use Eq. (3-13):

$$\begin{aligned}\sin(\delta) &= \sin(\phi_{ecliptic}) \cos(\epsilon) + \cos(\phi_{ecliptic}) \sin(\epsilon) \sin(\lambda_{ecliptic}) \\ &= -20.477\ 711^\circ \\ \sin(\alpha) &= \frac{-\sin(\phi_{ecliptic}) \sin(\epsilon) + \cos(\phi_{ecliptic}) \cos(\epsilon) \sin(\lambda_{ecliptic})}{\cos(\delta)} \\ &= 113.308\ 898^\circ \text{ or } 246.691\ 102^\circ\end{aligned}$$

A quadrant check is needed to determine the correct right ascension. The *Astronomical Almanac* again lists the apparent values:

$$\alpha = 16\ h\ 26\ min\ 34.60\ s = 246.644\ 165\ 91^\circ$$



$$\delta = -20^\circ 22' 27.00'' = -20.374\ 166\ 67^\circ$$

3.7 Application: Planetary Ephemerides

We often want to know the approximate position of the planets. Although the *Astronomical Almanac* provides abundant data on past locations, we must calculate future locations. Determining a planet's location combines several topics discussed so far. First, we have to form a position and velocity vector at the given epoch (usually the mean equator and mean equinox of J2000). Usually, we do so using polynomials in terms of the classical orbital

elements and the elapsed Julian centuries [Eq. (1-50)]. Some applications may also specify mean or osculating elements. I'll discuss these two types of elements further in Sec. 8.2. Finally, we must select an output coordinate system. Because the polynomial expansions are virtually always given for the heliocentric coordinate system, we must carefully convert to other systems. Table D-4 and Table D-5 (in Appendix D) show data for all the planets for mean elements for the epoch J2000 and the epoch of date. A sample appears below for the mean elements of the mean equinox of date (Meeus, 1991, 200–204):

Heliocentric elements of Jupiter, mean equinox of date

$$\begin{aligned}a &= 5.202\,603\,191 + 1.913 \times 10^{-7} T_{TDB} \\e &= 0.048\,494\,85 + 0.000\,163\,244 T_{TDB} - 4.719 \times 10^{-7} T_{TDB}^2 - 1.97 \times 10^{-9} T_{TDB}^3 \\i &= 1.303\,270^\circ - 0.005\,496\,67 T_{TDB} + 4.65 \times 10^{-6} T_{TDB}^2 - 4 \times 10^{-9} T_{TDB}^3 \\\Omega &= 100.464\,441^\circ + 1.020\,955\,07 T_{TDB} + 0.000\,401\,17 T_{TDB}^2 + 5.69 \times 10^{-7} T_{TDB}^3 \\\tilde{\omega} &= 14.331\,309^\circ + 1.612\,666\,87 T_{TDB} + 0.001\,031\,27 T_{TDB}^2 - 4.569 \times 10^{-6} T_{TDB}^3 \\\lambda_M &= 34.351\,484^\circ + 3036.302\,788\,97 T_{TDB} + 0.000\,223\,74 T_{TDB}^2 + 2.5 \times 10^{-8} T_{TDB}^3\end{aligned}$$

Notice the orbital elements use alternate elements that refer to the *longitude of perihelion* and the *mean longitude*; thus, some calculations are necessary before forming the position and velocity vectors. We obtain the argument of perihelion directly because the quantity is measured in two planes. Because the inclination of each of the planets (except Pluto) is relatively small, there is little difference between $\tilde{\omega}$ and $\tilde{\omega}_{true}$, and the overall accuracy isn't affected.

$$\omega = \tilde{\omega} - \Omega$$

$$M = \lambda_M - \tilde{\omega}$$

Once we find the “standard” classical orbital elements, Algorithm 6 (**RANDV**) finds the position and velocity vectors at the epoch of the input data. Remember that because we use T_{TDB} as an argument, the velocity vector will be in units of AU/day. Convert to AU/TU by dividing the velocity vector by one solar TU (Table D-3). To form the geocentric values, we must rotate the vectors through the obliquity of the ecliptic, ϵ , because the initial values are given with respect to the ecliptic. Thus,

$$\vec{r}_{IJK} = \text{ROT1}[-\epsilon] \vec{r}_{XYZ}$$

We may also need to do some vector translations if we want the vectors relative to the Earth.

Remember that the vectors formed are inertial (J2000, *ECI*). Some applications may require the reduction formulas of Algorithm 4 to convert the position and velocity vectors

to an Earth-centered frame at the time of interest (*ECEF*). The important point to remember is that the vectors must be in the correct coordinate system for maximum accuracy. Selection of one method over another depends on the desired accuracy and specific problem under consideration.

Meeus (1991, 205–208) introduces the method of Bretagnon and the “*Variations séculaires des orbites planétaires*” (VSOP82 and VSOP87) for determining the precise (oscillating or true time-varying) locations of the planets. This solution rests on solving long polynomial equations to determine the orbital elements. It uses equinoctial elements. Consult either Bretagnon (1982), Bretagnon and Francou (1988), or Meeus (1991) for more information.

ALGORITHM 20: *PLANETRV* (*Planet, JD* $\Rightarrow \vec{r}_{IJK}, \vec{v}_{IJK}, \vec{r}_{XYZ}, \vec{v}_{XYZ}$)

TIMECONV (*Yr, Mo, D, UT1, ΔUT1, ΔAT* \Rightarrow *UTC, TAI, TDT, TDB*)

$$T_{TDB} = \frac{JD_{TDB} - 2,451,545.0}{36,525}$$

Find elements $a, e, i, \Omega, \tilde{\omega}, \lambda_M$ from Table D-4 or Table D-5:

$$M = \lambda_M - \tilde{\omega}$$

$$\omega = \tilde{\omega} - \Omega$$

KEPEQTNE ($M, e \Rightarrow v$)

$$p = a(1 - e^2)$$

RANDV ($\rho, e, i, \Omega, \omega, v (u, \lambda_{true}, \tilde{\omega}_{true}) \Rightarrow \vec{r}_{XYZ}, \vec{v}_{XYZ}$)

$$\vec{r}_{IJK}, \vec{v}_{IJK} = \text{ROT1}(-\epsilon) \vec{r}_{XYZ}, \vec{v}_{XYZ}$$

This algorithm is particularly complicated, especially if it includes reduction (from Sec. 1.7), so we'll run through an example to show how it works.

▼ Example 3-10. Calculating a Planetary Ephemeris.

GIVEN: Table D-4 and Table D-5

FIND: Position and velocity vectors on Jupiter on May 20, 1994 at 20:00 h *UT1*, using the mean equator and equinox of J2000.

First, determine the Julian date. If you use Algorithm 2,

$$JD = 2,449,493.333$$

Use Eq. (1-50) to find the Julian centuries of TDB:

$$T_{TDB} = -0.056\,171\,58$$

From Table D-5, find values for the orbital elements of Jupiter. These values allow you to find the heliocentric elements for the mean equator, mean equinox of J2000.

$$a = 5.202\,603\,191 + 1.913 \times 10^{-7} T_{TDB} = 5.202\,603 \text{ AU}$$

$$\begin{aligned} e &= 0.048\,494\,85 + 0.000\,163\,244 T_{TDB} - 4.719 \times 10^{-7} T_{TDB}^2 - 1.97 \times 10^{-9} T_{TDB}^3 \\ &= 0.048\,486 \end{aligned}$$

$$\begin{aligned} i &= 1.303\,270^\circ - 0.005\,496\,6 T_{TDB} + 4.65 \times 10^{-6} T_{TDB}^2 - 4 \times 10^{-9} T_{TDB}^3 \\ &= 1.303\,579^\circ \end{aligned}$$

$$\begin{aligned} \Omega &= 100.464\,441^\circ + 1.020\,955\,0 T_{TDB} + 0.000\,401\,17 T_{TDB}^2 + 5.69 \times 10^{-7} T_{TDB}^3 \\ &= 100.407\,082^\circ \end{aligned}$$

$$\begin{aligned} \tilde{\omega} &= 14.331\,309^\circ + 1.612\,666\,8 T_{TDB} + 0.001\,031\,27 T_{TDB}^2 - 4.569 \times 10^{-6} T_{TDB}^3 \\ &= 14.240\,708^\circ \end{aligned}$$

$$\begin{aligned} \lambda_M &= 34.351\,484^\circ + 3036.302\,788\,9 T_{TDB} + 0.000\,223\,74 T_{TDB}^2 + 2.5 \times 10^{-8} T_{TDB}^3 \\ &= -136.236\,16^\circ \end{aligned}$$

Convert to classical elements:

$$M = \lambda_M - \tilde{\omega} = -150.476\,867^\circ$$

$$\omega = \tilde{\omega} - \Omega = -86.166\,374^\circ$$

Kepler's equation transforms elements to yield true anomaly:

$$KEPEQTNE(M, e \Rightarrow \nu), \nu = -147.650\,09^\circ$$

$$p = a(1 - e^2) = 5.190\,372 \text{ AU}$$

Use Algorithm 6 to find the position and velocity vectors in the heliocentric frame (don't forget to divide the velocity vector by one solar TU = 58.132 440 9):

$$RANDV(\rho, e, i, \Omega, \omega, \nu(u, \lambda_{true}, \tilde{\omega}_{true}) \Rightarrow \hat{r}_{XYZ} \hat{v}_{XYZ})$$

$$\hat{r}_{XYZ} = -4.075\,932 \hat{X} - 3.578\,306 \hat{Y} + 0.105\,970 \hat{Z} \text{ AU}$$

$$\hat{v}_{XYZ} = 0.004\,889 \hat{X} - 0.005\,320 \hat{Y} - 0.000\,087 \hat{Z} \text{ AU/TU}$$

Because the numbers are very large when we convert to km and km/s, the answer is often left in AU and AU/TU. A simple rotation produces the geocentric vector, still referencing the mean equator and mean equinox of J2000:

$$\epsilon = 23.440\,021^\circ$$

$$\hat{r}_{IJK} \hat{v}_{IJK} = \text{ROT1}(-\epsilon) \hat{r}_{XYZ} \hat{v}_{XYZ}$$

$$\begin{aligned} \hat{r}_{IJK} &= -4.075\,932 \hat{I} - 3.325\,167 \hat{J} - 1.326\,185 \hat{K} \text{ AU} \\ &= -609,750,812 \hat{I} - 497,437,970 \hat{J} - 198,394,489 \hat{K} \text{ km} \end{aligned}$$

$$\begin{aligned}\hat{v}_{IJK} &= 0.004\,889 \hat{i} - 0.004\,846 \hat{j} - 0.002\,197 \hat{k} \text{ AU/TU} \\ &= 0.145\,626 \hat{i} - 0.144\,349 \hat{j} - 0.065\,424 \hat{k} \text{ km/s}\end{aligned}$$

Remember this is the vector from the Sun to Jupiter in the *IJK* frame. If you need the vector from the Earth to Jupiter, you must subtract the vector from the Sun to the Earth. You can compare your answers with those from DE245 for the mean equator and mean equinox of J2000 (see Appendix D for information on obtaining these vectors).

$$\hat{r}_{IJK} = -4.068\,697 \hat{i} - 3.335\,668 \hat{j} - 1.330\,683 \hat{k} \text{ AU}$$

▲ $\hat{v}_{IJK} = 0.004\,897 \hat{i} - 0.004\,840 \hat{j} - 0.002\,194 \hat{k} \text{ AU/TU}$

3.7.1 Eclipses

Although lunar eclipses aren't of utmost interest in studying Earth-orbiting satellites, the concept of eclipses is important to complete mission profiles. Satellites routinely enter eclipse periods during their orbits, and this can cause a great fluctuation in the amount of available solar radiation for the solar arrays. This section presents the fundamentals of eclipse geometries and types, permitting analysis of solar and lunar eclipses. Figure 3-7 shows the basic eclipse geometry.

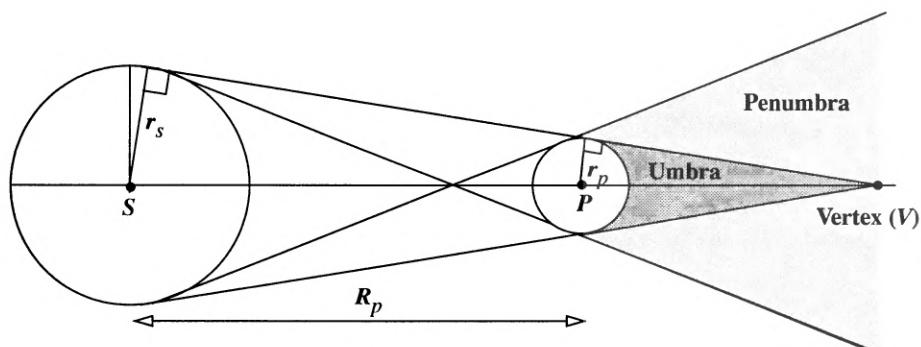


Figure 3-7. Eclipse Geometry. The general geometry for eclipses involves a primary object (Earth) and a secondary object (Sun). Eclipses are formed within the umbra and the penumbra regions. Notice the scales are greatly exaggerated; the penumbral region is actually rather small. For the Sun-Earth system, $PV \approx 1.384 \times 10^6 \text{ km}$, or about four times the distance to the Moon.

We're interested in two main regions. The **umbra** is totally eclipsed by the primary (Earth); the **penumbra** is only partially obscured by the primary. The similar triangles allow us to solve the distance from the planet to the vertex. Thus,

$$\frac{PV}{r_p} = \frac{SV}{r_s} \quad \Rightarrow \quad PV = \frac{SVr_p}{r_s}$$

We also know the distance between the secondary (Sun) and primary (Earth):

$$SV = PV + R_p$$

Substituting for SV and solving gives us

$$PV = \frac{PVr_p}{r_s} + \frac{R_p r_p}{r_s}$$

Now let's combine terms to get

$$PV \left(1 - \frac{r_p}{r_s} \right) = \frac{R_p r_p}{r_s}$$

Finally,

$$PV = \frac{R_p r_p}{r_s - r_p} \quad (3-14)$$

Using variables for the Earth and the Sun, we find PV to be 1.384×10^6 km. Notice this is almost four times the distance from the Earth to the Moon (3.8268×10^5 km). Hence, it comes as little surprise that the Earth sometimes completely obscures the Moon, causing a *lunar eclipse*. These eclipses vary based on the amount of coverage. The same phenomenon affects satellites.

Lunar Eclipses

1. Total eclipse—the Moon is completely within the umbral region of the Earth's shadow.
2. Partial—the Moon is partially covered by the penumbra and umbra regions.
3. Penumbral—the Moon is only within the penumbra of the Earth.

We can do similar calculations for the Moon and the Sun. The result from Eq. (3-14) for the Moon yields a total distance of the umbral cone of about 3.735×10^5 km. Notice the value is surprisingly similar to the Moon's distance from the Earth. When the Moon obscures part of the Earth with its shadow, it causes a *solar eclipse*. The closeness in values is the reason solar eclipses are less common. The geometry of the Moon and the Earth must have certain configurations to permit the shadow of the Moon to fall correctly. The Moon can also shadow a satellite given the correct geometry, but this is much less common than lunar eclipses. Solar eclipses take three forms:

Solar Eclipses

1. Total eclipse—the Moon's umbral shadow falls directly on the Earth.

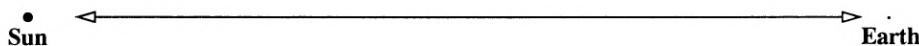
2. Annular eclipse—the Moon's umbral cone doesn't fall directly on the Earth, but the axis of the cone intersects the Earth.
3. Partial eclipse—only the Moon's penumbra region is involved in the shadowing.

Several other factors are involved. First, the Earth's gravity affects the light through its gravity field. Atmospheric refraction also tends to decrease (slightly) the size of the umbra and penumbra regions.

3.8 Application: Sunrise, Sunset, and Twilight Times

The times at which sunrise, sunset, and twilight occur will vary during the year based on the Earth's rotation and its revolutions around the Sun. We distinguish *sunrise* and *sunset* as the times when the upper limb of the Sun is on the horizon of an observer on the Earth while *twilight* is the time at which the Sun has a particular angular separation from the observer. The *Almanac for Computers* (1990, B5–B7) gives a concise formula to find these times for latitudes on Earth below about 65° . Above 65° latitude, many days have no rising or setting phenomenon. Be aware that I've made some changes to maintain notation and to improve accuracy. An important parameter in the discussion is the angle between the site and the Sun, ζ , shown in Fig. 3-8.

To scale:



Exaggerated scale:

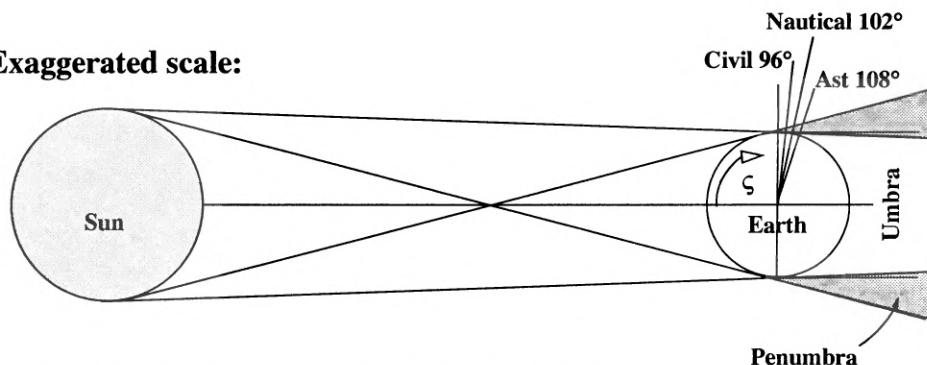


Figure 3-8. Geometry for Solar Illumination. I've shown an exact and an exaggerated scale to highlight the immense distances between the Sun and Earth. Sunrise, sunset, and twilight times depend on the angle between the site and the Sun, ζ , which defines the position and timing of an event. The various angles define the different types of illumination.

To start, we have various definitions for what constitutes solar illumination. The simplest method is to define the time when the site is exactly 90° away from the vector that runs from the Earth's center to the center of the Sun. Designate this description simply as a sunrise or sunset time. Illustrated in Fig. 3-8 is the fact that the Sun's light actually "curves" around the Earth a little, due to the Sun's immense size. Although the light doesn't actually bend (except for refraction and small relativistic effects), the rays from the upper limb of the Sun extend about $15'45''$ beyond 90° ; thus, we can modify our simple definition. The usual convention uses $\zeta = 90^\circ 50'$. The extra minutes above the geometric deflection ($15'45''$) account for atmospheric refraction. Civilian authorities have adopted 96° as *civil twilight* to define "darkness" for legal purposes. Mariners define *nautical twilight* as 102° because no mountain ranges or land masses block the rays. Finally, astronomers need the sky to be very "dark" to make observations, so they use 108° and call it *astronomical twilight*.

Figure 3-8 shows two shadowed regions. These regions are very similar to the definitions in eclipses but arise from different geometry. The *umbra* region receives no direct light from the Sun and is in total darkness. The *penumbra* receives partial light and is quite small. Analysis of entrance and exit times from these regions is important when evaluating the availability of satellite power.

The procedure to find the sunrise, sunset, and twilight times involves several concepts of time from Chap. 1. In summary, we must determine all parameters for the Sun at both a rising or setting phenomenon—*independent* of the site of interest. Once calculated, the solar parameters yield values of sidereal time, which we may subtract from and add to the site longitude in order to determine the *UT* for the phenomenon. I don't specify a particular *UT* in this section because the method is only approximate. I'll indicate formulas that use a specific type of *UT* (*UTC*, *UT1*, *TDB*) with the appropriate subscript, but for this section, they are interchangeable.

The formal procedure begins identically to the first few steps in Algorithm 18 (*Sun*) except for the *JD*. For this case, we must calculate the *JD* on the day of interest for either a rise or set phenomenon. In a perfect system, the phenomenon would occur at local times of 06:00 and 18:00. We can therefore calculate the offset at a particular site as follows:

$$JD_{\text{sunrise}} = JD_{0\text{h}} + 6 \text{ h}/24 \text{ h/day} - \lambda/15^\circ/\text{day}$$

$$JD_{\text{sunset}} = JD_{0\text{h}} + 18 \text{ h}/24 \text{ h/day} - \lambda/15^\circ/\text{day}$$

Once we find the mean longitude, mean anomaly, and ecliptic longitude from Algorithm 18, we can determine the right ascension and declination values using Eq. (3-13). Remember that the ecliptic latitude is 0° —which allows us to simplify—and be sure to check that ecliptic longitude and right ascension are in the same half plane.

$$\text{TAN}(\alpha_\odot) = \text{COS}(\epsilon) \text{TAN}(\lambda_{\text{ecliptic}})$$

$$\text{SIN}(\delta_\odot) = \text{SIN}(\epsilon) \text{SIN}(\lambda_{\text{ecliptic}})$$

Now find the local hour angle of the site. Algorithm 17 (*AzeltoRadec*) provides the relevant equation. Notice you must recalculate the resulting equation because the argument ζ is necessary to determine the type of phenomenon. Using $\sin(90 - \zeta) = \cos(\zeta)$, find this result:

$$\cos(LHA_{\odot}) = \frac{\cos(\zeta) - \sin(\delta_{\odot}) \sin(\phi_{gc})}{\cos(\delta_{\odot}) \cos(\phi_{gc})}$$

As always, the cosine provides two solutions to the inverse trigonometric operation. Don't use the two values from the cosine operation for both rise and set values of LHA . You must recalculate the parameters for the Sun using the appropriate JD value. The primary answer is for sunset calculations, and the secondary answer ($360^\circ - LHA$) is for sunrise. If you find GST (Eq. (1-44)) at 6 hours and 18 hours on the day of interest, Eq. (1-43) allows you to determine the longitudinal difference. This represents the UT of the phenomena. The universal time is

$$UT = LHA_{\odot} + \alpha_{\odot} - GST$$

Be careful of the units when determining UT with the equation above. I show the algorithm using the Julian date, implying you must know the year. In practice, the times change very little from year to year and can be determined for any year if one isn't specified.

ALGORITHM 21: SunriseSet ($JD_{UT1}, \phi_{gc}, \lambda, \zeta \Rightarrow UT_{sunrise}, UT_{sunset}$)

Do separate calculations for sunrise and sunset using

$$JD_{sunrise} = JD_{0h} + 6/24 - \lambda/15$$

$$JD_{sunset} = JD_{0h} + 18/24 - \lambda/15$$

$$T_{UT1} = \frac{JD_{UT1} - 2,451,545.0}{36,525} \quad \text{Let } T_{TDB} = T_{UT1}$$

$$\lambda_{M_{\odot}} = 280.460\,618\,4^\circ + 36,000.770\,053\,61 T_{UT1}$$

$$M_{\odot} = 357.527\,723\,3^\circ + 35,999.050\,34 T_{TDB}$$

$$\lambda_{ecliptic} = \lambda_{M_{\odot}} + 1.914\,666\,471^\circ \sin(M_{\odot}) + 0.019\,994\,643 \sin(2M_{\odot})$$

$$\epsilon = 23.439\,291^\circ - 0.013\,004\,2 T_{TDB}$$

$$\tan(\alpha_{\odot}) = \cos(\epsilon) \tan(\lambda_{ecliptic})$$

$$\begin{aligned}\sin(\delta_{\odot}) &= \sin(\epsilon) \sin(\lambda_{ecliptic}) \\ \cos(LHA_{sunset}) &= \frac{\cos(\zeta) - \sin(\delta_{\odot}) \sin(\phi_{gc})}{\cos(\delta_{\odot}) \cos(\phi_{gc})}\end{aligned}$$

IF a sunrise calculation

$$LHA_{sunrise} = 360^\circ - LHA_{sunset}$$

$$UT = LHA_{\odot} + \alpha_{\odot} - GST$$

Consider the following example.

▼ **Example 3-11. Calculating Sunrise-Sunset Times.**

GIVEN: Site at 40°N, 0°E on March 23, 1996

FIND: Sunrise and sunset times

Begin by finding the Julian dates of the sunrise and sunset occurrences.

$$JD_{sunrise} = JD_{0h} + 6/24 - \lambda/15$$

$$JD_{sunset} = JD_{0h} + 18/24 - \lambda/15$$

Next, find several intermediate values. First, consider the sunrise case, for which

$$T_{UT1} = \frac{JD_{UT1} - 2,451,545.0}{36,525} = \frac{2,450,165.75 - 2,451,545.0}{36,525} = -0.037762$$

$$\lambda_{M_{\odot}} = 280.460\,618\,4^\circ + 36,000.770\,053\,61 T_{UT1} = 1.006\,470^\circ$$

Let $T_{TDB} \equiv T_{UT1}$

$$M_{\odot} = 357.527\,723\,3^\circ + 35,999.050\,34 T_{TDB} = 78.139^\circ$$

$$\lambda_{ecliptic} = \lambda_{M_{\odot}} + 1.914\,666\,471^\circ \sin(M_{\odot}) + 0.019\,994\,643 \sin(2M_{\odot}) = 2.888^\circ$$

$$\epsilon = 23.439\,291^\circ - 0.013\,004\,2 T_{TDB} = 23.439\,978^\circ$$

$$\tan(\alpha_{\odot}) = \cos(\epsilon) \tan(\lambda_{ecliptic}), \alpha_{\odot} = 2.650^\circ$$

$$\sin(\delta_{\odot}) = \sin(\epsilon) \sin(\lambda_{ecliptic}), \delta_{\odot} = 1.149^\circ$$

$$\cos(LHA_{sunset}) = \frac{\cos(\zeta) - \sin(\delta_{\odot}) \sin(\phi_{gc})}{\cos(\delta_{\odot}) \cos(\phi_{gc})}, LHA_{sunset} = 92.052\,31^\circ$$

Because you're considering sunrise, adjust the value of the LHA :

$$LHA_{sunrise} = 360^\circ - LHA_{sunset} = 267.947\,69^\circ$$

Use Eq. (1-44) to find the GST at 6 h on March 23, 1996:

$$GST_{\odot} = -178.994^\circ$$

Find $UT_{sunrise}$ as

$$UT_{sunrise} = LHA_{\odot} + \alpha_{\odot} - GST = 89.59152^\circ \Rightarrow 5 \text{ h } 58 \text{ min } 21.9659 \text{ s}$$

Similar operations for sunset [remember to use JD_{sunset} ($T_{UT1} = -0.037748$)] lead to

$$\lambda_{M_{\odot}} = 1.499312^\circ, M_{\odot} = 78.631^\circ, \lambda_{ecliptic} = 3.384^\circ, \epsilon = 23.440411^\circ$$

$$\alpha_{\odot} = 3.105^\circ, \delta_{\odot} = 1.346^\circ, LHA_{sunset} = 92.05231^\circ, GST_{\odot} = -178.501^\circ \text{ (at 18 h UT)}$$

▲ $UT_{sunset} = LHA_{\odot} + \alpha_{\odot} - GST_{\odot} = 273.824006^\circ \Rightarrow 18 \text{ h } 15 \text{ min } 17.7614 \text{ s}$

3.9 Application: Sight and Light

The primary problem to solve for **SIGHT** and **LIGHT** is whether or not *line of sight* (LOS) exists between two given position vectors. In the case of **SIGHT**, the two vectors usually refer to satellites, whereas the **LIGHT** routine determines whether or not a satellite has LOS with the Sun and is therefore in the sunlight. We have two ways to determine the **SIGHT** problem: a simple angular derivation and a more complex algebraic manipulation.

Assume a problem in which we know two arbitrary vectors (\vec{a} and \vec{b}). Using the dot product, find the angle between the two vectors:

$$\vec{a} \cdot \vec{b} = |\vec{a}| |\vec{b}| \cos(\theta)$$

If you now fix the maximum perpendicular distance between the central point and the line-of-sight vector to 1.0 ER, you can determine the two half-angles shown in Fig. 3-9. The two angles (using canonical units) for the constrained problem are

$$\cos(\theta_1) = \frac{1}{|\vec{a}|} \quad \cos(\theta_2) = \frac{1}{|\vec{b}|}$$

No quadrant check is required because all the angles are less than 180° . If the sum of $\theta_1 + \theta_2 \leq \theta$, there is no LOS. Although this approach is relatively straightforward, it does contain a number of costly calculations (divisions, square roots, and trigonometric functions).

An alternate method by Alfano (1991) removes the trigonometric operations but requires a little more development. Figure 3-9 shows the situation. Start with a parametric representation of a line between the two position vectors:

$$\vec{c}(\tau) = \vec{a} + (\vec{b} - \vec{a})\tau \quad (3-15)$$

where τ varies from 0 (\vec{a}) to 1.0 (\vec{b}). We want to determine the value of τ which represents the minimum distance to the central body. Because we're really interested only in the magnitude, the first thought is to square $\vec{c}(\tau)$ and take the square root to obtain the magnitude. But we can get the same value by examining the squared result and avoiding the difficulties caused by the square root (Let $\vec{a} = a_1\hat{I} + a_2\hat{J} + a_3\hat{K}$ and $\vec{b} = b_1\hat{I} + b_2\hat{J} + b_3\hat{K}$).

$$|\vec{c}(\tau)|^2 = (a_1 + (b_1 - a_1)\tau)^2 + (a_2 + (b_2 - a_2)\tau)^2 + (a_3 + (b_3 - a_3)\tau)^2$$

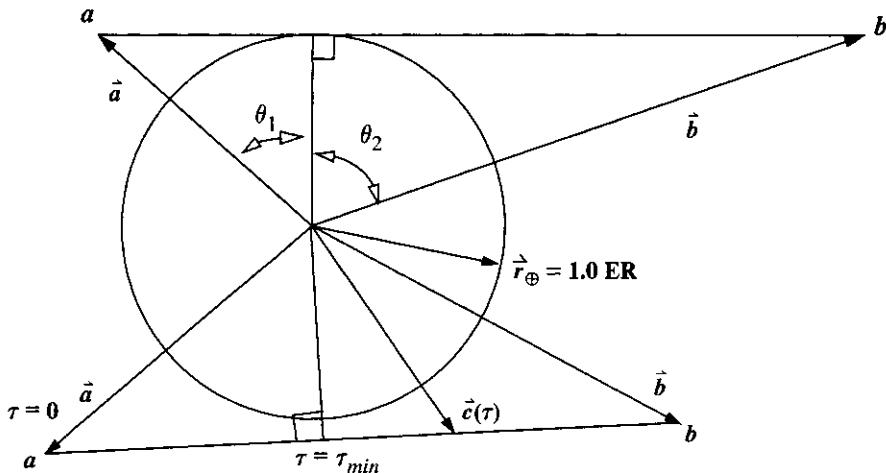


Figure 3-9. Line-of-Sight (LOS) Geometry. LOS exists between the two positions if the sum of the two angles ($\theta_1 + \theta_2$) is larger than the calculated value for θ . We can also use a parametric approach in which we find the location for the minimum distance to the Earth. Each vector can be represented as a function of τ .

Take the derivative with respect to τ and set it equal to zero to determine the minimum point for the value of τ . Remember to write each component separately in order to do the required differentiation.

$$\frac{d|\vec{c}(\tau)|^2}{d\tau} = 2 [(a_1 + (b_1 - a_1)\tau)(b_1 - a_1) + (a_2 + (b_2 - a_2)\tau)(b_2 - a_2) + (a_3 + (b_3 - a_3)\tau)(b_3 - a_3)]$$

Equate this expression to zero and solve for the minimum value of τ , representing the closest approach with the central body:

$$\tau_{min} = \frac{a_1(a_1 - b_1) + a_2(a_2 - b_2) + a_3(a_3 - b_3)}{(b_1 - a_1)^2 + (b_2 - a_2)^2 + (b_3 - a_3)^2}$$

Multiply terms and simplify the previous equation using vectors:

$$\tau_{min} = \frac{|\vec{a}|^2 - \vec{a} \cdot \vec{b}}{|\vec{a}|^2 + |\vec{b}|^2 - 2 \vec{a} \cdot \vec{b}} = \frac{\vec{a} \cdot (\vec{a} - \vec{b})}{|\vec{a} - \vec{b}|^2} \quad (3-16)$$

Squaring Eq. (3-15) yields

$$|\vec{c}(\tau_{min})|^2 = |\vec{a}|^2 + |\vec{b} - \vec{a}|^2 \tau_{min}^2 + 2(\vec{b} - \vec{a}) \cdot \vec{a} \tau_{min}$$

Expanding gives you

$$|\hat{c}(\tau_{min})|^2 = |\hat{a}|^2 + (|\hat{b}|^2 + |\hat{a}|^2 - 2\hat{a} \cdot \hat{b})\tau_{min}^2 + 2(-|\hat{a}|^2 + \hat{a} \cdot \hat{b})\tau_{min}$$

Now, separate terms and substitute τ_{min}^2 from Eq. (3-16):

$$|\hat{a} - \hat{b}|^2 = |\hat{a}|^2 + |\hat{b}|^2 - 2\hat{a} \cdot \hat{b}$$

$$|\hat{c}(\tau_{min})|^2 = |\hat{a}|^2(1 - 2\tau_{min}) + 2(\hat{a} \cdot \hat{b})\tau_{min} + \frac{(|\hat{a}|^2 - \hat{a} \cdot \hat{b})^2}{|\hat{a}|^2 + |\hat{b}|^2 - 2\hat{a} \cdot \hat{b}}$$

Notice that τ_{min} is in the last term, so

$$|\hat{c}(\tau_{min})|^2 = |\hat{a}|^2(1 - 2\tau_{min}) + 2(\hat{a} \cdot \hat{b})\tau_{min} + (|\hat{a}|^2 - \hat{a} \cdot \hat{b})\tau_{min}$$

The final result is

$$|\hat{c}(\tau_{min})|^2 = (1 - \tau_{min})|\hat{a}|^2 + (\hat{a} \cdot \hat{b})\tau_{min} \quad (3-17)$$

To use this procedure, first find τ_{min} . If $\tau_{min} < 0$ or $\tau_{min} > 1.0$, there is line-of-sight between the two vectors. From Fig. 3-9, this is reasonable because both vectors are in the same quadrant with respect to the attracting body. If τ_{min} falls between 0 and 1, substitute its value into Eq. (3-17) to determine the square of the magnitude at the minimum location. If this squared magnitude is larger than (or equal) to 1.0, LOS exists. In all remaining cases, LOS doesn't exist.

An immediate question arises when using this routine for an ellipsoidal Earth model. The spherical approach will yield conservative values (fewer lines of sight than actual) because its radii are “larger.” Solution is relatively easy because we can scale the K component of the position vectors using Eq. (1-6), and use the **SIGHT** algorithm exactly as before. We may want to do the actual scaling transformation within a separate variable because this will help remind the programmer not to pass the scaled component back to the main program.

The **LIGHT** algorithm employs this routine to determine whether the Sun is lighting the satellite. The procedure is the same as with **SIGHT**; however, one position vector is now the *Sun* rather than another satellite. Be cautious of the units involved—some routines give the Sun position vector in AU, whereas the satellite is often in ER. The process works well for most cases. If you need extreme precision—including aberration, distortion, etc.—use a more detailed process.

Implementing the SIGHT Algorithm

The **SIGHT** algorithm appears below. Store the dot product to improve performance in a system that makes repetitive calls to this routine.

ALGORITHM 22: SIGHT (\vec{r}_1, \vec{r}_2 , in IJK \Rightarrow LOS)

$$\tau_{min} = \frac{|\vec{r}_1|^2 - \vec{r}_1 \cdot \vec{r}_2}{|\vec{r}_1|^2 + |\vec{r}_2|^2 - 2 \vec{r}_1 \cdot \vec{r}_2}$$

LOS = FALSE

IF $(\tau_{min} < 0)$ or $(\tau_{min} > 1)$

LOS = TRUE

ELSE IF $|c(\tau_{min})|^2 = (1 - \tau_{min})|\vec{r}_1|^2 + (\vec{r}_1 \cdot \vec{r}_2)\tau_{min} \geq 1.0$

LOS = TRUE

Let's work an example to test the **SIGHT** and **LIGHT** routines.**▼ Example 3-12. Solving for SIGHT and LIGHT Conditions.**GIVEN: $\vec{r}_1 = -4464.696\hat{J} - 5102.509\hat{K}$ km, $\vec{r}_2 = 5740.323\hat{J} + 3189.068\hat{K}$ km on February 15, 1995, 12:00 h UT1.

FIND: LOS and whether satellite 1 is in the sunlight.

First, convert to canonical units:

$$\vec{r}_1 = -0.7\hat{J} - 0.8\hat{K} \quad \vec{r}_2 = 0.9\hat{J} + 0.5\hat{K}$$

Then, find the minimum parametric value:

$$\tau_{min} = \frac{|\vec{r}_1|^2 - \vec{r}_1 \cdot \vec{r}_2}{|\vec{r}_1|^2 + |\vec{r}_2|^2 - 2 \vec{r}_1 \cdot \vec{r}_2} = 0.508\ 224\ 9$$

Now find the value of the function using this minimum value:

$$|c(\tau_{min})|^2 = (1 - \tau_{min})|\vec{r}_1|^2 + (\vec{r}_1 \cdot \vec{r}_2)\tau_{min} = 0.030\ 864\ 2$$

Because the value is less than 1.0, we have no LOS.

Next, determine the position vector for the Sun. Using Algorithm 18, find the Sun's position vector. For February 15, 1995, 0 h UT1, $JD = 2,449,763.5$, so

$$\vec{r}_{\odot} = 0.817\ 078\ 6\hat{I} - 0.509\ 033\ 8\hat{J} - 0.220\ 697\ 6\hat{K}$$
 AU

Be sure to convert to consistent (canonical in this case) units:

$$\vec{r}_1 = -0.7\hat{J} - 0.8\hat{K} \quad \vec{r}_{\odot} = 19,164.5\hat{I} - 11,939.1\hat{J} - 5,176.4\hat{K}$$
 ER

Apply the **SIGHT** algorithm to produce the minimum parametric value:

$$\tau_{min} = \frac{|\vec{r}_1|^2 - \vec{r}_1 \cdot \vec{r}_{\odot}}{|\vec{r}_1|^2 + |\vec{r}_{\odot}|^2 - 2 \vec{r}_1 \cdot \vec{r}_{\odot}} = 1.000\ 023\ 3$$

▲ Because the value is larger than 1.0, we have LOS.

3.10 Converting IJK To Latitude and Longitude (LatLon)

A procedure to convert the position vector for a satellite to the corresponding latitude and longitude is important, especially in determining ground tracks (Sec. 4.4). For this problem, we know the elements of the satellite's position vector in the geocentric equatorial system but must find the geocentric or geodetic latitude, the longitude, and the height of the satellite above sea level. The difference between geocentric and geodetic latitude appears small, but the difference can cause errors up to about 20 km. This procedure also is especially important for surveillance, and it provides background for the discussions in Chap. 10.

We have two main ways to do the transformation: one iterative and one analytical. Figure 3-10 below shows the situation. At first glance, this problem appears to be trivial because simple trigonometric relations will immediately yield a "value" for latitude and longitude. Unfortunately, the Earth isn't a perfect sphere, so the exact point on the Earth which lies along the position vector is *not* usually the same as the point which lies directly (perpendicular) below the satellite. As the right-hand side of Fig. 3-10 suggests, we sometimes need an iterative procedure to determine both quantities. I'll present two methods: one for speed and one for simplicity.

The first method uses spherical trigonometry. We find the right ascension directly from the cartesian position vector. Let the equatorial projection of the satellite's position vector be $r_{\delta_{sat}} = \sqrt{r_I^2 + r_J^2}$. Don't confuse this value with the equatorial projection of the site used in Chap. 1. First, we find the right ascension through sine and cosine expressions:

$$\sin(\alpha) = \frac{r_J}{r_{\delta_{sat}}} \quad \cos(\alpha) = \frac{r_I}{r_{\delta_{sat}}}$$

The *GST* allows us to determine the longitude:

$$\lambda = \alpha - \theta_{GST}$$

The difficult part of finding the geodetic latitude is that it usually requires iteration. To determine a starting value for the iteration, we can use the position vector as a rough guess because the declination and geocentric latitude are equal. Thus,

$$\sin(\delta) = \frac{r_{K_{sat}}}{r_{\delta_{sat}}}$$

At this point, we can use various techniques. Escobal ([1965] 1985, 398–399) shows an iterative method to determine geodetic latitude. Although reasonably accurate, this method is about 25% slower than one presented in the *Astronomical Almanac* (1992, K12). I'll present the latter technique because it's accurate and efficient.

To develop the method from the *Astronomical Almanac*, we must find an expression for geodetic latitude. Unfortunately, we now have the satellite coordinates and not the site

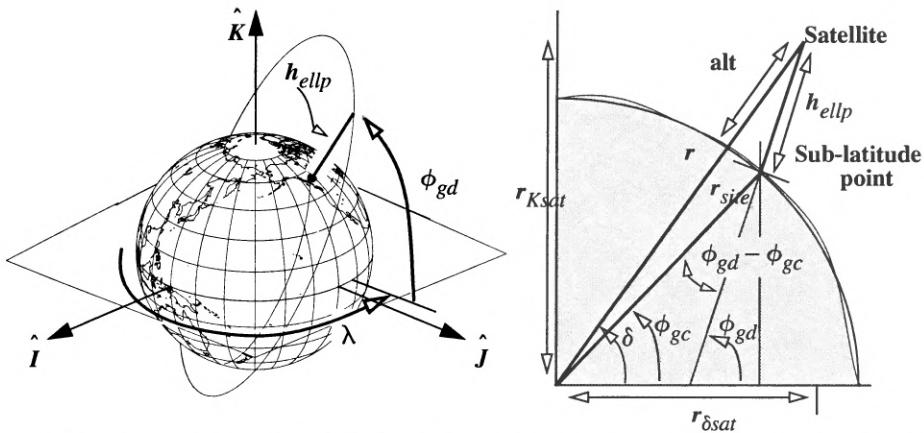


Figure 3-10. Determining a Satellite's Sub-latitude Point. Notice h_{ellp} is used for the altitude, and not H_{MSL} . The difference between the geodetic and geocentric angles is helpful in setting up the iteration for the problem.

coordinates. Assume for the first guess that $\phi_{gd} \approx \delta$. Using Eq. (1-16), find the sine and cosine expressions:

$$\sin(\phi_{gd}) = \frac{r_K}{S_{\oplus} + h_{\text{ellp}}} \quad \cos(\phi_{gd}) = \frac{r_{\delta}}{C_{\oplus} + h_{\text{ellp}}}$$

Solving the sine expression for h_{ellp} gives you

$$h_{\text{ellp}} = \frac{r_K}{\sin(\phi_{gd})} - S_{\oplus}$$

Now determine a tangent expression and substitute h_{ellp} to get

$$\tan(\phi_{gd}) = \frac{\sin(\phi_{gd})}{\cos(\phi_{gd})} = \frac{r_K \left(C_{\oplus} + \frac{r_K}{\sin(\phi_{gd})} - S_{\oplus} \right)}{\left(S_{\oplus} + \frac{r_K}{\sin(\phi_{gd})} - S_{\oplus} \right) r_{\delta}}$$

Also remember from Eq. (1-16) that $C_{\oplus} = \frac{S_{\oplus}}{1 - e_{\oplus}^2}$, so you can simplify

$$\tan(\phi_{gd}) = \frac{r_K \left(\frac{S_{\oplus}}{1 - e_{\oplus}^2} \sin(\phi_{gd}) + r_K - S_{\oplus} \sin(\phi_{gd}) \right)}{r_{\delta} r_K}$$

A common denominator yields

$$\tan(\phi_{gd}) = \frac{r_K(1-e_\oplus^2) + e_\oplus^2 S_\oplus \sin(\phi_{gd})}{r_\delta(1-e_\oplus^2)}$$

Now convert back to C_\oplus and determine the final result:

$$\begin{aligned}\tan(\phi_{gd}) &= \frac{r_K(1-e_\oplus^2) + (1-e_\oplus^2)e_\oplus^2 C_\oplus \sin(\phi_{gd})}{r_\delta(1-e_\oplus^2)} \\ \tan(\phi_{gd}) &= \frac{r_K + e_\oplus^2 C_\oplus \sin(\phi_{gd})}{r_\delta}\end{aligned}$$

The notation in this section is specific. We have a solution by rearranging equations from Chap. 1, but for this particular problem, we don't have the site coordinates (r_δ and r_K). Careful examination of Fig. 3-10 shows that, although the central angles for the site and satellite are δ and ϕ_{gc} respectively, a relation of similar triangles exists. This relation allows us to use $r_{\delta sat}$ and r_{Ksat} in the formula. We find the height above mean sea level by solving Eq. (1-16) for h_{ell} . The algorithm appears below.

ALGORITHM 23: IJK to LatLon ($\vec{r}_{IJK}, \theta_{GST} \Rightarrow \phi_{gd}, \lambda$)

$$r_{\delta sat} = \sqrt{r_I^2 + r_J^2}$$

$$\sin(\alpha) = \frac{r_I}{r_{\delta sat}} \quad \cos(\alpha) = \frac{r_I}{r_{\delta sat}}$$

$$\lambda = \alpha - \theta_{GST}$$

$$\sin(\delta) = \frac{r_{Ksat}}{r_{\delta sat}}$$

$$\text{Let } \phi_{gd} = \delta \quad r_\delta = r_{\delta sat} \quad r_K = r_{Ksat}$$

LOOP

$$C_\oplus = \frac{R_\oplus}{\sqrt{1-e_\oplus^2 \sin^2(\phi_{gd})}}$$

$$\tan(\phi_{gd}) = \frac{r_K + C_\oplus e_\oplus^2 \sin(\phi_{gd})}{r_\delta}$$

UNTIL $\phi_{gd} - \phi_{gd_{old}} < \text{Tolerance}$

$$h_{ellp} = \frac{r_\delta}{\cos(\phi_{gd})} - C_\oplus$$

This method often requires an extra iteration over that of Escobal, but the formula's simplicity actually results in a quicker routine.

▼ Example 3-13. Converting IJK to Lat Lon.

GIVEN: $\vec{r}_{IJK} = 6524.834 \hat{i} + 6862.875 \hat{j} + 6448.296 \hat{k}$ km on February 24, 1995 at 12:00 h UT.

FIND: Latitude and longitude

Convert units to find

$$\vec{r}_{IJK} = 1.023 \hat{i} + 1.076 \hat{j} + 1.011 \hat{k}$$
 ER

$$r_{\delta sat} = \sqrt{r_I^2 + r_J^2} = 1.4847$$
 ER

$$\sin(\alpha) = \frac{r_J}{r_{\delta sat}} \quad \alpha = 46.4464^\circ$$

The right ascension requires a quadrant check (using the cosine). It's correct as shown. Use the Day of Year to find the JD:

JulianDate (1995, 2, 24, 12, 0, 0.0 \Rightarrow JD)

JD = 2,449,773.0

Now, find the GST:

LSTIME (JD, UT, $\lambda \Rightarrow \theta_{GST}, \theta_{LST}$)

$\theta_{GST} = 333.893\ 486^\circ$

Determine the longitude without iteration:

$$\lambda = \alpha - \theta_{GST} = -287.447\ 086^\circ = 72.552\ 29^\circ\text{W}$$

Now the iteration begins. Three iterations are shown:

$$C_\oplus = \frac{R_\oplus}{\sqrt{1 - e_\oplus^2 \sin^2(\phi_{gd})}} = 1.001\ 062\ 1, 1.001\ 067\ 5, 1.001\ 067\ 5$$

$$\tan(\phi_{gd}) = \frac{r_K + R_\oplus C_\oplus e_\oplus^2 \sin(\phi_{gd})}{r_\delta} \Rightarrow \phi_{gd} = 34.352\ 242, 34.352\ 495, 34.352\ 496$$

Finally, the geodetic latitude is

$$\phi_{gd} = 34.352\ 496^\circ$$

You can use Eq. (1-18) to find the geocentric latitude: $\phi_{gc} = 34.173\ 429^\circ$. The height is

$$h_{ellp} = \frac{r_\delta}{\cos(\phi_{gd})} - C_\oplus = \frac{1.4847}{\cos(34.352\ 496)} - 1.001\ 067\ 5 = 0.797\ 289$$
 I ER = 5085.22 km

Problems

1. How many significant digits of decimal degrees are required to match the accuracy of 12 h 41 min 37.4567 s? Determine the answer for degrees, arc-minutes, and arcseconds using the same numerical values.
2. What day is it for the following Julian dates? 2,363,592.5, 2,391,598.5, 2,418,781.5, 2,446,470.5, 2,474,033.5. Are these dates significant?
3. You've just received observations of a new comet from hikers on Mt. Washington (44.15°N , 71.15°W , 1916.58 m) at dusk ($\zeta = 90^{\circ} 50'$) on January 31, 1996. If they obtain values of $\beta = 55.6^{\circ}$ and $el = 52.3^{\circ}$, what are the topocentric right ascension and declination? Do you need to make any assumptions?
4. On April 15, 1995, 10:00 *UT*, can the Moon "see" the Sun? Explain your answer.
5. What is the *UT* of sunrise/sunset for a site at 40°N on June 15, 1996? What is the local time of sunrise and sunset for a site at 105°W longitude? Would your answer differ if you were at 97.3°W ? Explain your answers.
6. We know Eq. (1-18) relates geocentric and geodetic latitude of the satellite. Why can't we just take a satellite-position vector, determine ϕ_{gc} , and use Eq. (1-18) to find the geodetic latitude of the satellite?
7. Another method provides the geodetic latitude and ellipsoidal height but requires no iteration. The method is from Borkowski (1989). Essentially, the procedure relies on a fourth-order polynomial to evaluate the results. Although this method provides only the geodetic latitude, we can find the geocentric latitude in one step using Eq. (1-18). Investigate the efficiency of this routine compared to Algorithm 23. Hint: the geodetic latitude and the height above the ellipsoid are found analytically. Seidelmann (1992, 206–207) shows the algorithm.

CHAPTER 4 KEPLER'S EQUATION AND KEPLER'S PROBLEM

- 4.1 Historical Background
- 4.2 Kepler's Equation
- 4.3 Kepler's Problem
- 4.4 Application: Groundtracks
- 4.5 Application: Find Time of Flight (*FindTOF*)

4.1 Historical Background

Although Isaac Newton (1642–1727) produced much of the mathematics required to solve the orbital problem, Kepler had already made many important discoveries. Kepler determined how to relate mean and true anomalies in the orbit to time. Solving this problem enabled him to predict future occurrences for planets (and ultimately, satellites). How did Kepler arrive at these conclusions? He took many paths, sometimes even using erroneous data! As Pannekoek reports, Ptolemy encountered Kepler's difficulties centuries earlier:

As Ptolemy tried to resolve the dimensions required for his equant and moveable excentrics, he pondered “Now we found, however, with continued exact comparison of the course given by observations and the results from combinations of these hypotheses, that the progress of the motion cannot be quite so simple... The epicycles cannot have their centres proceed along such eccentric circles that, seen from the centre [of these circles], they describe equal angles in equal time... But the latter bisect the distance between the point from which the motion appears to be uniform, and the centre of the ecliptic.” (Pannekoek, 1989, 137–138)

We've introduced Kepler's laws in Chap. 2. The first two laws (elliptical motion and equal areas in equal times) provide the necessary conditions for all two-body motion. As we saw in Chap. 2, an impressive array of tools is available with this “Keplerian” motion. We'll examine Kepler's first two laws in a little greater detail because they're the basis for Kepler's equation [Eq. (4-4)], which we'll derive shortly.

Kepler began his duties in 1601, having acquired more than 20 years of highly accurate observations from Tycho Brahe. Bate, Mueller, and White (1971, 177; citation from Koesstler, 1959) suggest that Kepler admitted, “I confess that when Tycho died, I quickly took advantage of the absence, or lack of circumspection, of the heirs, by taking the observations under my care, or perhaps usurping them.”

Without the telescope (remember, it wasn't applied to astronomy until Galileo did so in 1609), Kepler could only use Brahe's measurements. His data was accurate to about 0.033° ($\sim 2'$), which is remarkable for the early 1600s! Kepler's task is often called the conquest of Mars. Kepler was drawn to Mars due to the relatively "large" eccentricity of its orbit. The formal problem was one of fitting an orbit to the observational data of Brahe. Beer and Strand (1975b, 545–546) suggest that the large (by today's standards) inaccuracy of the observations was actually beneficial to Kepler. We now know that the masses of the other planets cause noticeable perturbations in the orbit of Mars. If Brahe's observations had been more accurate, Kepler may never have satisfied his desire to accurately fit the data—the study of perturbations was still over a hundred years away!

Kepler used the Copernican theory as a starting point. Use of a Sun-centered system was beneficial, but the system still used small excentrics to model the motion correctly. Kepler determined that the Earth moved faster and slower at different parts of its orbit, which led him directly to forming the second law. Kepler placed this law second, even though it was discovered first.

For almost the entire first year, Kepler pursued the system of excentrics and was able to fit the data to within 0.133° ($\sim 8'$). This wasn't good enough! The next geometric shape Kepler tried was ovals. He used a polar equation (Beer and Strand, 1975b, 554):

$$r = 1 + e \cos(\nu) + Ae^2 \sin^2(\nu) + Be^3 \sin^2(\nu) \cos(\nu) + (O)e^4 \quad (4-1)$$

where $e \approx 1/11$. A was the departure from roundness (about $-1/2$) and B was the oval symmetry.

By 1603, Kepler had improved the formula of the oval and determined that $B = -3/2$, but he couldn't find a natural cause to explain this value, so he abandoned the approach. If he had recognized the reduced solution, he would have saved many years of effort. Beer and Strand (1975b, 579) describe Kepler's frustration on July 4, 1603: "if [only] the figure were a perfect ellipse, then Archimedes and Apollonius' [work] would suffice." Some of Kepler's difficulty arose because he tried to use a "central anomaly," γ , in the radius equation ($r = a(1 + e\cos(\gamma))$). By Easter, 1605, Kepler recognized that his "central anomaly" was actually the eccentric anomaly, E , we use today. This angle actually represented the true ellipse, and the problem was solved.

"Why should I mince my words? The truth of nature, which I had rejected and chased away, returned by stealth through the backdoor, disguising itself to be accepted. That is to say, I laid [the original equation] aside and fell back on ellipses, believing that this was a quite different hypothesis, whereas the two...are one in the same. I thought and searched, until I went nearly mad, for a reason why the planet preferred an elliptical orbit (to mine)... Ah what a foolish bird I have been!" Bate, Mueller, and White (1971, 181)

In the end, Kepler wrote empirical expressions for the time-of-flight of a planet from one point in its orbit to another, although he still did not know the true reason why it should move in an orbit at all.

It's interesting to wonder if Kepler recognized the immense importance of the two solutions as he worked through the problems just discussed. Literally hundreds of technical treatises have discussed Kepler's equation, but we'll focus on deriving several forms of the solution to permit excursions when necessary. Although I'll discuss some multiple-solution techniques, I'll mainly explore general-purpose uses to help engineers in virtually all phases of simulation and analysis.

4.2 Kepler's Equation

Kepler's equation allows us to determine the relation of the time and angular displacement within an orbit. We'll consider **Kepler's problem** later because it deals with propagating a satellite through its orbit and requires us to know the solution to Kepler's equation. Many books use the terms interchangeably, but I'll discuss them separately.

While resolving the orbit of Mars, Kepler's first headache came from fitting an oval to the orbit. When he finally selected the ellipse, he encountered another problem: how to relate angular position within this conic section to time. From his second law, he knew that the area swept out in equal times at different points of the orbit was constant. This is the starting point for our discussion. Figure 4-1 shows the relationship between the area swept out, A_1 , and the associated angular displacement.

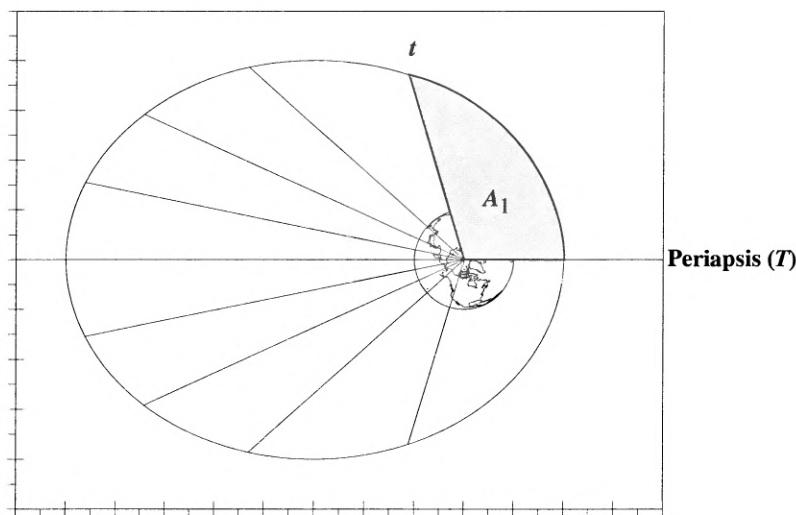


Figure 4-1. Kepler's Equation—Equal Areas. According to Kepler's second law, a satellite will sweep out equal areas in equal time. This figure shows positions for ten equal spacings in an orbit. For this particular orbit, $a = 31,890$ km (5.0 ER), $e = 0.6$, and the space between lines is about 94.4 minutes.

From basic geometry we know the total area of an ellipse is defined by the product of the semimajor and semiminor axes and pi (πab). For now, suppose the satellite starts at periapsis and continues through its orbit as shown. The second law states that equal areas are swept out in equal times, or

$$\frac{t - T}{A_1} = \frac{\rho}{\pi ab} \quad (4-2)$$

where the relation $(t - T)$ represents the change in time, and ρ is the period of one orbital revolution. The only unknown in this equation is the area, A_1 , which we can determine through geometric operations. Consider Fig. 4-2. The circle drawn around the ellipse is called the *auxiliary circle*, and the new angle to be defined is the *eccentric anomaly*, E , which is related to the true anomaly, ν , as shown. It's convenient to express the scaling factor between circle and ellipses [Eq. (1-6)] a bit differently for this derivation. Let

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

To find an expression for the area, A_1 , in terms of the eccentric anomaly, begin by determining the area of the elliptical segment, PCB , minus the smaller side triangle, A_2 , or

$$A_1 = \text{Area } PCB - \text{Triangle } A_2 \quad (4-3)$$

Use standard geometric and trigonometric relations to find the area, A_2 :

$$A_2 = \frac{1}{2} (ae - a\cos(E)) \left(\frac{b}{a} a \sin(E) \right) = \frac{ab}{2} (e \sin(E) - \cos(E) \sin(E))$$

Find the area of the segment of the ellipse, PCB , by using the area of a circle, POB' , subtracting the triangle, $OB'C$, and scaling the result with Eq. (1-6), which relates the vertical distances of an ellipse and a circle. Thus,

$$\text{Area } PCB = \frac{a^2 E}{2} \left(\frac{b}{a} \right) - \frac{1}{2} (a \cos(E)) (a \sin(E)) \frac{b}{a}$$

Reducing gives you

$$\begin{aligned} \text{Area } PCB &= \frac{abE}{2} - \frac{a^2}{2} (\cos(E)) (\sin(E)) \frac{b}{a} \\ &= \frac{ab}{2} (E - \cos(E) \sin(E)) \end{aligned}$$

After substituting into Eq. (4-3), you'll see the area

$$A_1 = \frac{ab}{2} (E - e \sin(E))$$

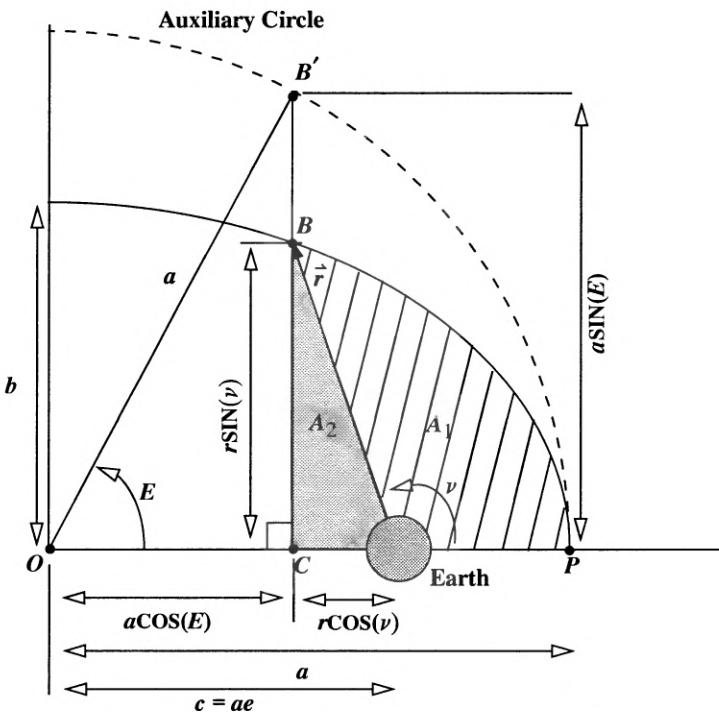


Figure 4-2. Geometry of Kepler's Equation. The eccentric anomaly uses an auxiliary circle as shown. The ultimate goal is to determine the area, A_1 , which allows us to calculate the time.

Substitute back into Eq. (4-2) and solve for the period

$$\rho = \frac{2\pi(t-T)}{E - e \sin(E)}$$

Recall the definition for the period of a satellite developed in Eq. (2-16), and note that

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

This equation—known as **Kepler's equation**—relates the time of flight from periapsis to the eccentric anomaly, semimajor axis, and eccentricity. Kepler then introduces notation for the *mean anomaly*, M , as

$$M = E - e \sin(E) = \sqrt{\frac{\mu}{a^3}}(t-T) \quad (4-4)$$

and also introduces the *mean motion*, n , or mean angular velocity. The *mean motion* is the mean angular rate of the orbital motion as defined in Eq. (2-17). Kepler's third law relates ρ to the semimajor axis, so n is a function of a .

$$n \equiv \sqrt{\frac{\mu}{a^3}} \quad (4-5)$$

For instance, a satellite completing 14 revolutions per day has $n = 5040$ °/day, or 0.8214 rad/TU. An angle defined by

$$M = \int_T^t n dt = n(t - T)$$

measured from perigee is called the *mean anomaly*, M . It corresponds to uniform angular motion on a circle of radius a .

We can't picture mean anomaly as a physical element of the motion like the eccentric anomaly; rather it's related to time. You'll see in Sec. 4.3 that relations between the eccentric and mean anomalies are possible. Mean, eccentric, and true anomaly are all equal at periapsis, $E = M = \nu = 0^\circ$. Likewise, all three elements are equal to 180° at apoapsis. In general, we can write Kepler's equation as

$$M = n(t - T) = E - e \sin(E) \quad (4-6)$$

Few applications begin at periapsis and end within one revolution, so we want to generalize this relation for orbital motion over *any* time interval. Figure 4-3 shows the general situation.

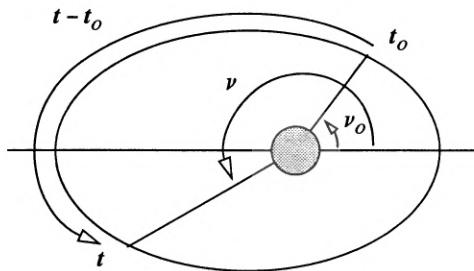


Figure 4-3. General Time of Flight. The time of flight between two specified locations has two basic forms: those of less than one period (as shown) and those with more than one period. Multiple periods include one or more complete revolutions in addition to the displacement shown here.

Again, t and t_o represent the times of the event. The satellite in the orbit shown doesn't complete any extra revolutions, so our change in time, Δt , is simply $t - t_o$. Transfers of

more than one period must add time to account for each additional revolution. k represents the number of period(s) the satellite completes in addition to moving between the two locations on the orbit. Thus,

$$\Delta t = kP + t - t_o$$

which allows us to develop a completely general form of Kepler's equation (shown here in radians):

$$\frac{M - M_o}{n} = t - t_o = \sqrt{\frac{a^3}{\mu}} \left\{ 2\pi k + E - e \sin(E) - (E_o - e \sin(E_o)) \right\} \quad (4-7)$$

To finish developing Kepler's equation, let's derive a set of conversions between ν , M , and E . It's important to remember that each of these relations is intended for elliptical orbits, and therefore we assume the eccentricity is less than 1. Many of the following formulas become indeterminate if e is outside this range. These cases require special treatment, which we'll explore later. From Fig. 4-2, we can directly obtain expressions relating eccentric anomaly and true anomaly. Remember to divide by the scaling factor when converting from elliptical quantities to corresponding circular ones [Eq. (1-6)]. Eq. (4-8) is sometimes solved for $r \sin(\nu)$ or $r \cos(\nu)$. These are considered basic relations because they provide a basis for many of the following derivations.

$$\sin(E) = \frac{r \sin(\nu)}{a \sqrt{1 - e^2}}, \quad \cos(E) = \frac{ae + r \cos(\nu)}{a} \quad (4-8)$$

We want to find relations that are independent of the position magnitude. Begin by substituting the trajectory equation for the radius, r , into Eq. (4-8). Doing so results in

$$\begin{aligned} \sin(E) &= \frac{a(1 - e^2) \sin(\nu)}{a(1 + e \cos(\nu)) \sqrt{1 - e^2}} \\ \cos(E) &= e + \frac{\frac{a(1 - e^2)}{a(1 + e \cos(\nu))} \cos(\nu)}{a} = \frac{e(1 + e \cos(\nu)) + (1 - e^2) \cos(\nu)}{(1 + e \cos(\nu))} \end{aligned}$$

Reducing gives you

$$\sin(E) = \frac{\sqrt{1 - e^2} \sin(\nu)}{1 + e \cos(\nu)} \quad \cos(E) = \frac{e + \cos(\nu)}{1 + e \cos(\nu)} \quad (4-9)$$

I've included the sine and cosine expressions so a computer can run ATAN2 (See Appendix A2) and eliminate quadrant checks. For hand calculations, either formula will work, but you must remember to resolve the angle to the proper quadrant. The simple rule

is that ν , E , and M will always be in the same half-plane— 0° to 180° or 180° to 360° —but not necessarily in the same quadrant.

Let's turn around the conversions between E and ν so we'll have a way to find the true anomaly given the eccentric anomaly. Solving the cosine relation in Eq. (4-9) for ν yields an immediate expression:

$$\cos(\nu) = \frac{\cos(E) - e}{1 - e\cos(E)} \quad (4-10)$$

The trajectory equation relates position and true anomaly. Another useful expression for the position relates semimajor axis and eccentric anomaly. Rearranging the cosine expression in Eq. (4-8) gives you

$$r = \frac{a\cos(E) - ae}{\cos(\nu)}$$

Using Eq. (4-10) above, note that

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

and

$$r = a(1 - e\cos(E)) \quad (4-11)$$

You can also determine this quantity directly from Fig. 1-3.

Find the sine expression in terms of E and ν by solving Eq. (4-8) for $\sin(\nu)$:

$$\sin(\nu) = \frac{a\sqrt{1-e^2}\sin(E)}{r}$$

Now replace the position, r , with Eq. (4-11) and simplify:

$$\sin(\nu) = \frac{\sqrt{1-e^2}\sin(E)}{1 - e\cos(E)} \quad (4-12)$$

A single equation for true anomaly is useful. Divide Eq. (4-10) by Eq. (4-12) and use the tangent half-angle formula to expand the result:

$$\tan\left(\frac{\nu}{2}\right) = \frac{\frac{\sqrt{1-e^2}\sin(E)}{1 - e\cos(E)}}{\frac{\cos(E) - e}{1 - e\cos(E)} + 1}$$

$$\tan\left(\frac{\nu}{2}\right) = \frac{\sqrt{1-e^2}\sin(E)}{\cos(E)-e+1-e\cos(E)}$$

Now recognize the denominator as the product of $(1-e)(\cos(E)+1)$; thus,

$$\tan\left(\frac{\nu}{2}\right) = \frac{\sqrt{1-e^2}\sin(E)}{(1-e)(\cos(E)+1)}$$

Reduce the terms containing eccentric anomaly with a formula for the tangent half-angle,

$$\tan\left(\frac{\nu}{2}\right) = \sqrt{\frac{1+e}{1-e}} \tan\left(\frac{E}{2}\right) \quad (4-13)$$

Then, solving for eccentric anomaly yields

$$\tan\left(\frac{E}{2}\right) = \sqrt{\frac{1-e}{1+e}} \tan\left(\frac{\nu}{2}\right) \quad (4-14)$$

4.2.1 Alternate Formulation for Eccentric Anomaly

Let's derive Kepler's equation [Eq. (4-4)] using an analytical method to prepare for an explanation of the parabolic and hyperbolic formulas. Begin with the differential equation from the derivation of specific mechanical energy [Eq. (2-8)]. This equation defines the constant value for angular momentum:

$$h = \frac{r^2 d\nu}{dt} \quad (4-15)$$

By integrating both sides, we get an expression for the time of flight. From the trajectory equation, we know position depends on true anomaly. Thus,

$$\int_T^t h dt = \int_0^\nu r^2 d\nu \quad (4-16)$$

where the limits of integration are from the time of periapsis passage (T and $\nu = 0^\circ$) to the current time (t and ν). Let's change the variable from true anomaly to eccentric anomaly to simplify the integral. From Eq. (4-10), we can differentiate the cosine expression with respect to the eccentric anomaly.

$$\begin{aligned} -\sin(\nu) d\nu &= [-\sin(E) (1-e\cos(E))^{-1} \\ &\quad + (\cos(E)-e)(-1)(1-e\cos(E))^{-2}(e\sin(E))] dE \\ &= \frac{-\sin(E) (1-e\cos(E)) - (\cos(E)-e)(e\sin(E))}{(1-e\cos(E))^2} dE \end{aligned}$$

By simplifying and separating, you'll discover that

$$d\nu = \frac{\sin(E)(1-e^2)}{\sin(\nu)(1-e\cos(E))^2} dE$$

Next, solve Eq. (4-8) for $\sin(\nu)$ and substitute Eq. (4-11) for the squared term in the denominator:

$$d\nu = \frac{\sin(E)(1-e^2)}{a\sqrt{1-e^2}\sin(E)\frac{r^2}{a^2}} dE$$

Simplify to get

$$d\nu = \frac{a\sqrt{1-e^2}}{r} dE \quad (4-17)$$

Eq. (4-17) is very important when developing analytical and semianalytical routines. It's the key relation that allows us to change variables.

Substitute Eq. (4-17) into the integral [Eq. (4-16)] and recall that angular momentum is constant, so

$$h(t-T) = \int_0^E r^2 \frac{a\sqrt{1-e^2}}{r} dE$$

Because some parameters are constant, you can remove them from the integral and substitute Eq. (4-11) to get

$$h(t-T) = a^2\sqrt{1-e^2} \int_0^E 1 - e\cos(E) dE = a^2\sqrt{1-e^2}(E - e\sin(E))$$

Remember that $h = \sqrt{\mu p} = \sqrt{\mu a(1-e^2)}$ and use Eq. (4-5) to get

$$h = na^2\sqrt{1-e^2}$$

so the solution reduces to

$$(t-T) = \sqrt{\frac{a^3}{\mu}}(E - e\sin(E)) \quad (4-18)$$

4.2.2 Formulation for the Parabolic Anomaly

For a complete analysis of Kepler's equation, we must also derive expressions for the parabolic and hyperbolic anomalies. Indeed, some of the earliest uses of the relations were for hyperbolic orbits. The possibility of an operational program requiring these formulas is small, but they'll give you an extra safeguard. Find the parabolic anomaly beginning with Eq. (4-15):

$$h = \frac{r^2 d\nu}{dt} = \sqrt{\mu p}$$

Unlike the eccentric-anomaly case, you can't change the variable to eccentric anomaly because the semimajor axis is infinite. An alternate form of the trajectory equation provides the necessary solution. The first step is to replace part of the trajectory equation with the identity $[1 + e\cos(\nu) = 1 + \cos(\nu) = 2\cos^2\left(\frac{\nu}{2}\right)$ (because $e = 1.0$)]:

$$r = \frac{p}{1 + \cos(\nu)} = \frac{p}{2\cos^2\left(\frac{\nu}{2}\right)} = \frac{p}{2} \left\{ \frac{\cos^2\left(\frac{\nu}{2}\right) + \sin^2\left(\frac{\nu}{2}\right)}{\cos^2\left(\frac{\nu}{2}\right)} \right\}$$

Thus,

$$r = \frac{p}{2} \left[1 + \tan^2\left(\frac{\nu}{2}\right) \right] \quad (4-19)$$

Substituting this result into Eq. (4-15) and rearranging gives you

$$\sqrt{\mu p} dt = \frac{p^2}{4} \left[1 + \tan^2\left(\frac{\nu}{2}\right) \right]^2 d\nu$$

Now, separate terms and set the limits of integration:

$$\int_T^t 4 \sqrt{\frac{\mu}{p^3}} dt = \int_0^\nu \left(1 + 2\tan^2\left(\frac{\nu}{2}\right) + \tan^4\left(\frac{\nu}{2}\right) \right) d\nu$$

Integrate or find the integrals from standard tables to get

$$4 \sqrt{\frac{\mu}{p^3}} (t - T) = \nu + 4 \tan\left(\frac{\nu}{2}\right) - 2\nu + \frac{2 \tan^3\left(\frac{\nu}{2}\right)}{3} - 2 \tan\left(\frac{\nu}{2}\right) + \nu$$

The result is the parabolic time equation:

$$2\sqrt{\frac{\mu}{p^3}}(t-T) = \tan\left(\frac{\nu}{2}\right) + \frac{\tan^3\left(\frac{\nu}{2}\right)}{3} \quad (4-20)$$

Now you can define *parabolic anomaly*, B , and mean motion, n_p :

$$B \equiv \tan\left(\frac{\nu}{2}\right) \quad n_p \equiv 2\sqrt{\frac{\mu}{p^3}} \quad (4-21)$$

Especially note the form of mean angular velocity, n_p , or mean motion, and parabolic anomaly, B . Although the letter P would be the obvious choice for parabolic anomaly, B was chosen for two reasons. First, the semiparameter, p , occurs with the parabolic anomaly in many equations, thus inviting confusion. Second, B is a kind of tribute to Thomas Barker (1722–1809), who developed the first tables of solution for parabolic orbits. Note that Herrick (1971, 191) suggests another possibility. He uses B for the mean anomaly only for parabolic orbits. This book uses M for all occurrences of mean anomaly and reserves B for parabolic orbits.

Several alternate formulas for parabolic anomaly and mean motion exist. For instance, some authors include the semiparameter with the parabolic anomaly. Although each is correct as long as subsequent derivations come from this given information, I chose the above definitions to (1) simplify the substitutions into the parabolic time equation, (2) keep the form of mean motion similar to that from elliptical and hyperbolic motion, and (3) preserve the correct units for the mean motion. The similarity to elliptical and hyperbolic motion is important because, by including the semiparameter in the mean-motion expression, we maintain the relationship between constant angular momentum and equal areas of Kepler's second law. Finally, note that we usually use radian units for B because it's not associated with trigonometric functions.

We can now simplify the time equation by writing

$$n_p(t-T) = B + \frac{B^3}{3} \quad (4-22)$$

As in the elliptical case, we generalize the results to include initial and final locations which aren't directly at periapsis:

$$\frac{M - M_o}{n_p} = t - t_o = \frac{1}{2}\sqrt{\frac{p^3}{\mu}} \left\{ B + \frac{B^3}{3} - \left(B_o + \frac{B_o^3}{3} \right) \right\} \quad (4-23)$$

Also with the elliptical case, we can develop expressions allowing us to convert between parabolic anomaly, B , true anomaly, ν , and mean anomaly, M . The previous equations provide some, but not all, of the conversions. Because B is defined in terms of ν , we don't need complex expressions. Eq. (4-24) is useful for determining the position magnitude and may be expressed in terms of B . From Eq. (4-19),

$$r = \frac{p}{2} [1 + B^2] \quad (4-24)$$

Find sine and cosine expressions using Eq. (4-22) and the formula for the tangent half-angle:

$$\tan\left(\frac{\nu}{2}\right) = \frac{\sin(\nu)}{\cos(\nu) + 1} \quad (4-25)$$

Using the trajectory equation (with $e = 1$) and solving for $\cos(\nu)$ gives you

$$\cos(\nu) = \frac{p - r}{r} \quad (4-26)$$

Substituting Eq. (4-26) into Eq. (4-25) and rearranging terms leads to

$$\sin(\nu) = \frac{PB}{r} \quad (4-27)$$

Mathematically, ν may change from -180° to 180° ; however, in practice, the angle is limited for parabolic orbits. Consider solving Eq. (4-21) when B is larger than about π . The true anomaly changes very slowly with increasing time (parabolic anomaly) and, for all practical purposes, never reaches 180° .

4.2.3 Formulation for the Hyperbolic Anomaly

The hyperbolic anomaly is similar to the eccentric anomaly (see Fig. 4-4).

Recall the equation for an orbit's specific mechanical energy and our discussion of energy. The hyperbola is the only orbit in which ξ is positive, and

$$\xi = \frac{v^2}{2} - \frac{\mu}{r} = -\frac{\mu}{2a}$$

If the energy is *always* positive, the semimajor axis must *always* be negative. This would seem to be contrary to our understanding of distance, but it's only a mathematical consequence. Still, it does present some problems when developing time-of-flight equations for a hyperbola.

Farther from the planet, the hyperbola approaches the asymptotes. The location of these asymptotes is described by a *turning angle*, φ , which measures their direction from the horizontal. We calculate the turning angle using a and the semiminor axis, b , which is the vertical distance from perigee to the asymptote. Battin (1987, 166) shows

$$\tan(\varphi) = \frac{b}{a} = \sqrt{e^2 - 1}$$

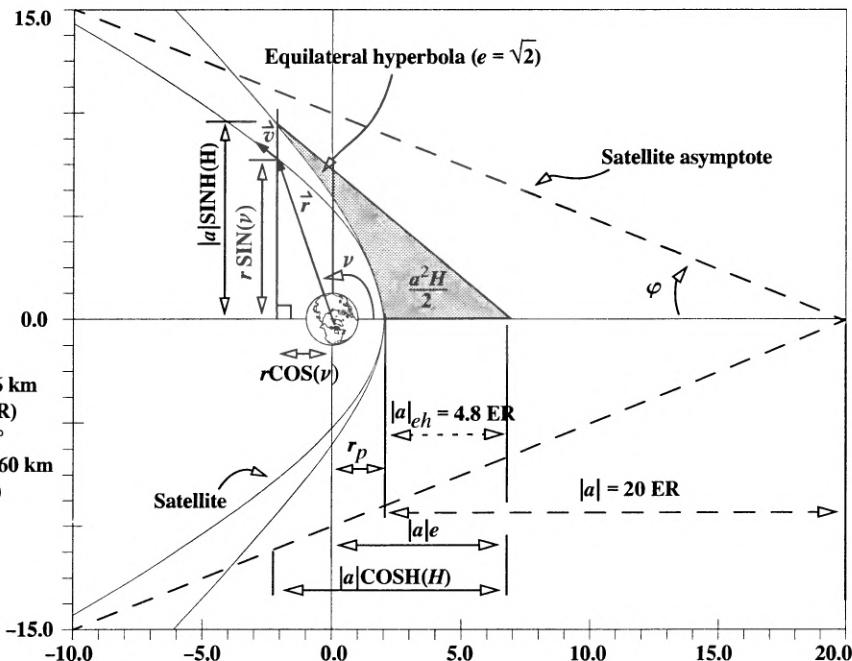


Figure 4-4. Hyperbolic Anomaly. The hyperbolic anomaly, H , is actually a measure of area, shown as the shaded region in the figure. Notice we use an equilateral hyperbola to determine H , just as we use a circle for elliptical orbits. Semimajor axis quantities are shown as absolute values to keep distances positive.

We define the **hyperbolic anomaly**, H , as we do the eccentric anomaly, E —by using an auxiliary conic section. Unfortunately, a circle won't work as it does with ellipses, so we use an equilateral hyperbola. This is simply a special case of a hyperbola with eccentricity of $\sqrt{2}$. You can scale between the two conic sections as with the circle and ellipse [Eq. (1-6)]; however, because the eccentricity is always greater than one, reverse the signs to be consistent with the hyperbolic formula. Thus, $(y_h = y_{eh}\sqrt{e^2 - 1})$ and $(y_{eh} = y_h/\sqrt{e^2 - 1})$. Incidentally, the scaling process works because each set of conic sections are really special cases of *one* type—the circle is a special case of the ellipse, and the equilateral hyperbola is a special case of the hyperbola. If you use these conversions with the example in the figure (because the eccentricity is less than $\sqrt{2}$), the vertical component of the equilateral hyperbola is greater than the hyperbola of interest. Finally, because the trajectory approaches the asymptotes at infinite distance, the true anomaly is limited within a certain range for all hyperbolic orbits. These values of true anomaly are supplementary with φ , so

$$-180^\circ + \cos^{-1}\left(\frac{1}{e}\right) < \nu < 180^\circ - \cos^{-1}\left(\frac{1}{e}\right) \quad (4-28)$$

The following table shows some representative values of the limitations on the true anomaly of a hyperbolic orbit whose eccentricity varies.

TABLE 4-1. Hyperbolic Limitations on True Anomaly. The true anomaly may assume only a limited range of values in hyperbolic orbits.

Eccentricity	Minimum ν	Maximum ν
1.001	-177.4387°	177.4387°
1.414	-135.0086°	135.0086°
1.600	-128.6822°	128.6822°
2.000	-120.0000°	120.0000°
4.000	-104.4775°	104.4775°
12.000	-94.7802°	94.7802°
25.000	-92.2924°	92.2924°

Instead of an angular measure, the hyperbolic anomaly represents the *area* from an equilateral hyperbola. From Fig. 4-4, we can directly obtain expressions for the hyperbolic anomaly. Remember to divide by the scaling factor when converting from the hyperbola to the equilateral hyperbola.

$$\text{SINH}(H) = \frac{-r \sin(\nu)}{a\sqrt{e^2 - 1}} \quad \text{COSH}(H) = \frac{ae + r \cos(\nu)}{a} \quad (4-29)$$

As in the elliptical case, we want to find relations that are independent of the position magnitude. Let's begin by substituting the trajectory equation for the radius, r :

$$\begin{aligned} \text{SINH}(H) &= \frac{-a(1-e^2) \sin(\nu)}{a\sqrt{e^2 - 1}(1+e\cos(\nu))} \\ \text{COSH}(H) &= \frac{ae + \frac{a(1-e^2)}{(1+e\cos(\nu))} \cos(\nu)}{a} = \frac{e(1+e\cos(\nu)) + (1-e^2)\cos(\nu)}{(1+e\cos(\nu))} \end{aligned}$$

Reducing and distributing the minus sign in the hyperbolic sine term gives us

$$\text{SINH}(H) = \frac{\sqrt{e^2 - 1} \sin(\nu)}{1 + e \cos(\nu)} \quad \text{COSH}(H) = \frac{e + \cos(\nu)}{1 + e \cos(\nu)} \quad (4-30)$$

Although I'm providing both expressions, we prefer the hyperbolic sine term because it's not double valued and therefore doesn't require quadrant checks.

We can also determine relations for the true anomaly given the hyperbolic anomaly. First, directly solve the cosine term in Eq. (4-30) for an expression in terms of ν :

$$\cos(\nu) = \frac{\cosh(H) - e}{1 - e\cosh(H)} \quad (4-31)$$

To find an alternate expression for the position magnitude, begin with Eq. (4-29), and solve the cosine expression for the position magnitude:

$$r = \frac{a\cosh(H) - ae}{\cos(\nu)}$$

Now use Eq. (4-31) to substitute for $\cos(\nu)$:

$$r = \frac{(a\cosh(H) - ae)(1 - e\cosh(H))}{\cosh(H) - e}$$

Simplifying results in

$$r = a(1 - e\cosh(H)) \quad (4-32)$$

Find the sine expression for true anomaly by solving the sine expression of Eq. (4-29) and then using the position magnitude from Eq. (4-32):

$$\sin(\nu) = \frac{-a\sqrt{e^2 - 1}\sinh(H)}{r} = \frac{-a\sqrt{e^2 - 1}\sinh(H)}{a(1 - e\cosh(H))}$$

The result is

$$\sin(\nu) = \frac{-\sqrt{e^2 - 1}\sinh(H)}{1 - e\cosh(H)} \quad (4-33)$$

A single equation for true anomaly is useful. Dividing Eq. (4-31) and Eq. (4-33) provides a tangent expression. Use the formula for the tangent half-angle to expand the result:

$$\tan\left(\frac{\nu}{2}\right) = \frac{-\sqrt{e^2 - 1}\sinh(H)}{1 - e\cosh(H)} \cdot \frac{1 + \frac{\cosh(H) - e}{1 - e\cosh(H)}}{1 + \frac{\cosh(H) - e}{1 - e\cosh(H)}}$$

$$\tan\left(\frac{\nu}{2}\right) = \frac{-\sqrt{e^2 - 1}\sinh(H)}{1 - e\cosh(H) + \cosh(H) - e}$$

By substitution, notice the denominator is the product of $(1 - e)(\cosh(H) + 1)$, so

$$\tan\left(\frac{\nu}{2}\right) = \frac{-\sqrt{e^2 - 1} \sinh(H)}{(1-e)(\cosh(H) + 1)}$$

Then, reduce the terms containing H with a hyperbolic formula for the tangent half-angle and distribute the negative sign to the denominator:

$$\tan\left(\frac{\nu}{2}\right) = \sqrt{\frac{e+1}{e-1}} \tanh\left(\frac{H}{2}\right) \quad (4-34)$$

Solving for hyperbolic anomaly results in

$$\tanh\left(\frac{H}{2}\right) = \sqrt{\frac{e-1}{e+1}} \tan\left(\frac{\nu}{2}\right) \quad (4-35)$$

With the conversions identified, solve Kepler's equation. Differentiate Eq. (4-31) with respect to the hyperbolic anomaly:

$$\begin{aligned} -\sin(\nu) d\nu &= \sinh(H) (1 - e \cosh(H))^{-1} \\ &\quad + (\cosh(H) - e) (-1) (1 - e \cosh(H))^{-2} - e \sinh(H) dH \\ &= \frac{\sinh(H) \{1 - e \cosh(H)\} + \{\cosh(H) - e\} e \sinh(H)}{(1 - e \cosh(H))^2} dH \end{aligned}$$

Simplifying and separating results in

$$d\nu = \frac{\sinh(H) (1 - e^2)}{-\sin(\nu) (1 - e \cosh(H))^2} dH$$

Next, solve Eq. (4-29) for $\sin(\nu)$ and substitute Eq. (4-32) for the squared term in the denominator:

$$d\nu = \frac{\sinh(H) (1 - e^2)}{\frac{a \sqrt{e^2 - 1} \sinh(H)}{r} \frac{r^2}{a^2}} dH$$

Simplify to get

$$d\nu = \frac{a(1 - e^2)}{r \sqrt{e^2 - 1}} dH = \frac{-a \sqrt{e^2 - 1}}{r} dH \quad (4-36)$$

Remember that a is intrinsically negative here, so be careful of the signs while simplifying. Now, substitute Eq. (4-36) into the integral [Eq. (4-16)] to get

$$h(t-T) = \int_0^H r^2 \frac{-a \sqrt{e^2 - 1}}{r} dH$$

Separate some terms:

$$h(t-T) = -a\sqrt{e^2-1} \int_0^H r dH$$

Now use the expression of angular momentum ($h = \sqrt{\mu p}$) and Eq. (4-32) to reduce the equation. It becomes

$$(t-T) = \frac{-a\sqrt{e^2-1}}{\sqrt{\mu p}} \int_0^H a(1-e\cosh(H)) dH$$

After integration,

$$(t-T) = \frac{-a^2\sqrt{e^2-1}}{\sqrt{\mu p}} (H - e \sinh(H))$$

Now distribute the negative sign to the terms in the parentheses and reduce the common term. Then insert the semiparameter equation to get

$$(t-T) = \frac{a^4(e^2-1)}{\sqrt{\mu a(1-e^2)}} (e \sinh(H) - H)$$

Reduce and keep track of the negative sign to write

$$(t-T) = \sqrt{\frac{-a^3}{\mu}} (e \sinh(H) - H) \quad (4-37)$$

Notice that, even though the semimajor axis is negative for hyperbolic orbits, the above formula negates it, allowing us to take the square root. The general solution is

$$\frac{M-M_o}{n} = t-t_o = \sqrt{\frac{-a^3}{\mu}} \left\{ e \sinh(H) - H - (e \sinh(H_o) - H_o) \right\} \quad (4-38)$$

4.2.4 Universal Formulation

Developing the preceding anomalies is convenient when processing data from one satellite. But some applications require extensive changes to satellite orbits and ultimately use all the orbit types, so we need a universal approach.

The change of variable to a *universal variable* is known as the *Sundman transformation*, after Karl F. Sundman (1873–1949), who first proposed the idea in 1912. Many authors have applied his original presentation to the problem of unifying the various theories. A notable form is from Battin (1987, 174–187). We'll use his notation with the derivation of Bate, Mueller, and White (1971, 191–198). The overall idea is to find an

expression relating the time and the orbit's properties. Although the result is unified, the intermediate steps may seem unusual as we obscure ties to commonly known references by substituting variables. Begin with the specific mechanical energy [Eq. (2-8)],

$$\xi = \frac{v^2}{2} - \frac{\mu}{r} = -\frac{\mu}{2a}$$

Resolving the velocity into radial and transverse components results in

$$\xi = \frac{\dot{r}^2}{2} + \frac{(r\dot{\nu})^2}{2} - \frac{\mu}{r} = -\frac{\mu}{2a}$$

Now, recalling angular momentum ($h = \sqrt{\mu p} = r^2\dot{\nu}$), and therefore ($\mu p = r^4\dot{\nu}^2$), solve the above equation for \dot{r}^2 :

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

The solution isn't immediately apparent; it requires defining an independent, *universal variable*, χ , to replace time as the independent variable. Define the derivative of the universal variable as

$$\dot{\chi} = \frac{\sqrt{\mu}}{r} \quad (4-39)$$

Now you need to develop a general equation for position, r , in terms of this universal variable. Squaring Eq. (4-39) and dividing it into the previous equations results in

$$\left(\frac{dr}{d\chi}\right)^2 = -p + 2r - \frac{r^2}{a}$$

which separates to

$$d\chi = \frac{dr}{\sqrt{-p + 2r - \frac{r^2}{a}}}$$

Solve by integrating both sides of the equation using c_o as the constant of integration. Notice the right-hand-side denominator is a quadratic in r , so searching a table in Selby (1975, 430 #238) becomes easier for the integral. The semimajor axis may be positive or negative. First consider the case in which the semimajor axis is positive and the term, $-r^2/a$, has a value less than zero. The solution is

$$\chi + c_o = -\sqrt{a} \sin^{-1} \left\{ \frac{-2\frac{r}{a} + 2}{\sqrt{-4\frac{p}{a} + 4}} \right\}$$

which you can reduce and reorganize with $-\sin(-\alpha) = \sin(\alpha)$,

$$\chi + c_o = \sqrt{a} \sin^{-1} \left\{ \frac{\frac{r}{a} - 1}{\sqrt{1 - \frac{p}{a}}} \right\}$$

But by substituting Eq. (2-43), you'll find

$$\chi + c_o = \sqrt{a} \sin^{-1} \left\{ \frac{\frac{r}{a} - 1}{e} \right\}$$

Now, solve this equation for the radial distance in terms of the universal variable:

$$r = a \left(1 + e \sin \left(\frac{\chi + c_o}{\sqrt{a}} \right) \right) \quad (4-40)$$

The second case arises when the semimajor axis is negative and the term $-r^2/a$ is positive. This case results in a hyperbolic sine solution, which we won't use in this book.

To obtain a relationship for time and the universal variable, substitute Eq. (4-40) into Eq. (4-39).

$$\frac{d\chi}{dt} = \frac{\sqrt{\mu}}{a \left(1 + e \sin \left(\frac{\chi + c_o}{\sqrt{a}} \right) \right)}$$

Separate and integrate the equation to find the result, assuming that $\chi = 0$ at the initial time:

$$\sqrt{\mu} \Delta t = a \chi - ae \sqrt{a} \left\{ \cos \left(\frac{\chi + c_o}{\sqrt{a}} \right) - \cos \left(\frac{c_o}{\sqrt{a}} \right) \right\} \quad (4-41)$$

Notice that these equations [Eq. (4-41) and Eq. (4-40)] give the position and timing information for a satellite in terms of the universal variable but contain an undetermined constant, c_o . The next task, then, is to determine this constant of integration and find a way to get the universal variable. Start by expanding Eq. (4-41) above with trigonometric identities for the cosine of a sum:

$$\sqrt{\mu}\Delta t = a\chi - ae\sqrt{a} \left\{ \cos\left(\frac{\chi}{\sqrt{a}}\right) \cos\left(\frac{c_o}{\sqrt{a}}\right) - \sin\left(\frac{\chi}{\sqrt{a}}\right) \sin\left(\frac{c_o}{\sqrt{a}}\right) - \cos\left(\frac{c_o}{\sqrt{a}}\right) \right\} \quad (4-42)$$

Now, determine expressions for the terms having the integration constant c_o . To eliminate these terms, use the position equation [Eq. (4-40)] and separate as follows:

$$e \sin\left(\frac{\chi + c_o}{\sqrt{a}}\right) = \frac{r}{a} - 1 \quad (4-43)$$

However, remember at the initial time, $\chi = t = 0$, and the position is the magnitude of the original position vector. So

$$e \sin\left(\frac{c_o}{\sqrt{a}}\right) = \frac{r_o}{a} - 1$$

Next, differentiate the position equation [Eq. (4-40)] to see the effect of the universal variable on the velocity. Be sure to remember the derivative of the universal variable in Eq. (4-39). Differentiating leads to

$$\dot{r} = \frac{ae}{\sqrt{a}} \left(\frac{\sqrt{\mu}}{r} \right) \cos\left(\frac{\chi + c_o}{\sqrt{a}}\right)$$

Write this equation as

$$\frac{r\dot{r}}{\sqrt{\mu a}} = e \cos\left(\frac{\chi + c_o}{\sqrt{a}}\right)$$

and at the initial conditions, using the relation $r\dot{r} = \dot{r} \cdot \dot{v}$ developed in Sec. 2.2.3, note that

$$e \cos\left(\frac{c_o}{\sqrt{a}}\right) = \frac{\dot{r}_o \cdot \dot{v}_o}{\sqrt{\mu a}} \quad (4-44)$$

Now, substitute these two expressions containing c_o , Eq. (4-43) and Eq. (4-44), into the expanded time equation [Eq. (4-42)]:

$$\sqrt{\mu}\Delta t = a\chi - a\sqrt{a} \left\{ \cos\left(\frac{\chi}{\sqrt{a}}\right) \frac{\dot{r}_o \cdot \dot{v}_o}{\sqrt{\mu a}} - \sin\left(\frac{\chi}{\sqrt{a}}\right) \left(\frac{r_o}{a} - 1 \right) - \frac{\dot{r}_o \cdot \dot{v}_o}{\sqrt{\mu a}} \right\}$$

Although not obvious now, it's useful to rearrange the equation:

$$\sqrt{\mu}\Delta t = a \left\{ \chi - \sqrt{a} \sin\left(\frac{\chi}{\sqrt{a}}\right) \right\} + \frac{\dot{r}_o \cdot \dot{v}_o}{\sqrt{\mu}} a \left\{ 1 - \cos\left(\frac{\chi}{\sqrt{a}}\right) \right\} + r_o \sqrt{a} \sin\left(\frac{\chi}{\sqrt{a}}\right) \quad (4-45)$$

Use a trigonometric identity to expand the position equation [Eq. (4-40)]:

$$r = a + ae \left\{ \sin\left(\frac{\chi}{\sqrt{a}}\right) \cos\left(\frac{c_o}{\sqrt{a}}\right) + \cos\left(\frac{\chi}{\sqrt{a}}\right) \sin\left(\frac{c_o}{\sqrt{a}}\right) \right\}$$

Make the same substitutions for the expressions containing c_o , [Eq. (4-43) and Eq. (4-44)]:

$$r = a + a \left\{ \sin\left(\frac{\chi}{\sqrt{a}}\right) \frac{\dot{r}_o \cdot \dot{v}_o}{\sqrt{\mu a}} + \cos\left(\frac{\chi}{\sqrt{a}}\right) \left(\frac{r_o}{a} - 1 \right) \right\} \quad (4-46)$$

At this point, we have the time [Eq. (4-45)] and position [Eq. (4-46)] of the satellite expressed as some value of the universal variable without the constant of integration. But we still haven't discovered a way to find the universal variable. Notice the coupling in the trigonometric terms of χ and semimajor axis. It's convenient to introduce another variable, ψ , which will yield a solution:

$$\psi \equiv \frac{\chi^2}{a} \quad (4-47)$$

Solve Eq. (4-47) for the semimajor axis and substitute it into the time and position equations [Eq. (4-45) and Eq. (4-46)] to get

$$\begin{aligned} \sqrt{\mu} \Delta t &= \frac{\chi^2}{\psi} \left\{ \chi - \frac{\chi}{\sqrt{\psi}} \sin(\sqrt{\psi}) \right\} + \frac{\dot{r}_o \cdot \dot{v}_o \chi^2}{\sqrt{\mu}} \frac{1 - \cos(\sqrt{\psi})}{\psi} + \left(\frac{r_o \chi}{\sqrt{\psi}} \right) \sin(\sqrt{\psi}) \\ r &= \frac{\chi^2}{\psi} + \frac{\chi^2}{\psi} \left\{ \sin(\sqrt{\psi}) \frac{\dot{r}_o \cdot \dot{v}_o \sqrt{\psi}}{\chi \sqrt{\mu}} + \cos(\sqrt{\psi}) \left(\frac{r_o \psi}{\chi^2} - 1 \right) \right\} \end{aligned}$$

With a little rearranging, find

$$\begin{aligned} \sqrt{\mu} \Delta t &= \chi^3 \left\{ \frac{\sqrt{\psi} - \sin(\sqrt{\psi})}{\sqrt{\psi^3}} \right\} + \frac{\dot{r}_o \cdot \dot{v}_o}{\sqrt{\mu}} \chi^2 \left\{ \frac{1 - \cos(\sqrt{\psi})}{\psi} \right\} + \frac{r_o \chi \sin(\sqrt{\psi})}{\sqrt{\psi}} \\ r &= \chi^2 \left\{ \frac{1 - \cos(\sqrt{\psi})}{\psi} \right\} + \frac{\dot{r}_o \cdot \dot{v}_o}{\sqrt{\mu}} \chi \left\{ 1 - \psi \frac{\sqrt{\psi} - \sin(\sqrt{\psi})}{\sqrt{\psi^3}} \right\} + r_o \left\{ 1 - \psi \frac{1 - \cos(\sqrt{\psi})}{\psi} \right\} \end{aligned}$$

Leave the last term of the position expression unsimplified to highlight a similar term. Notice we now have expressions for both time and position and have removed the constant of integration. The universal variable still remains, and its presence requires some sort of iteration, which we'll discuss shortly. For now, let's separate and calculate the common terms with the following definitions (Battin, 1987, 181):

$$c_2 \equiv \frac{1 - \cos(\sqrt{\psi})}{\psi} \quad c_3 \equiv \frac{\sqrt{\psi} - \sin(\sqrt{\psi})}{\sqrt{\psi^3}} \quad (4-48)$$

Substituting these terms into the position and time equations gives us the desired result. Notice we change the last term of the time equation so we can use the c_3 function.

$$\begin{aligned}\sqrt{\mu}\Delta t &= \chi^3 c_3 + \frac{\vec{r}_o \cdot \vec{v}_o}{\sqrt{\mu}} \chi^2 c_2 + r_o \chi (1 - \psi c_3) \\ r &= \chi^2 c_2 + \frac{\vec{r}_o \cdot \vec{v}_o}{\sqrt{\mu}} \chi (1 - \psi c_3) + r_o (1 - \psi c_2)\end{aligned}\quad (4-49)$$

These equations represent the universal-variable formula for Kepler's equation and Kepler's problem, respectively. It's possible to reform the time equation to match previous formulations. Solving Eq. (2-17) for $\sqrt{\mu}$, we can divide through and obtain

$$n\Delta t = \left(\frac{\chi}{\sqrt{a}}\right)^3 + \frac{\vec{r}_o \cdot \vec{v}_o}{\sqrt{\mu a}} \left(\frac{\chi}{\sqrt{a}}\right)^2 + \frac{r_o}{a} \left(\frac{\chi}{\sqrt{a}}\right) (1 - \psi c_3)$$

This form of Kepler's equation shows a "universal angle." We'll see later that $\chi/(\sqrt{a}) = \Delta E$.

Let's discuss formulations of the $c(\psi)$ functions. We determine both these functions with either a series representation or the trigonometric relations. I recommend the trigonometric relations because the penalty from transcendental operations is not too great. For the series, we have to truncate at eight to ten terms (a lot of multiplication and division in itself) to avoid the possibility of erroneous answers when truncating the series with too few terms. Recognize that ψ may be positive or negative; thus, we define hyperbolic functions for cases in which ψ is less than zero. The equations are

$$\begin{aligned}c_2 &= \frac{1 - \cos(\sqrt{\psi})}{\psi} \text{ for } \psi \geq 0 & c_2 &= \frac{1 - \cosh(\sqrt{-\psi})}{\psi} \text{ for } \psi < 0 \\ &= \sum_{k=0}^{\infty} \frac{(-\psi)^k}{(2k+2)!} = \frac{1}{2!} - \frac{\psi}{4!} + \frac{\psi^2}{6!} - \frac{\psi^3}{8!} + \frac{\psi^4}{10!} - \frac{\psi^5}{12!} + \dots\end{aligned}$$

$$\begin{aligned}c_3 &= \frac{\sqrt{\psi} - \sin(\sqrt{\psi})}{\sqrt{\psi^3}} \text{ for } \psi \geq 0 & c_3 &= \frac{\sinh(\sqrt{-\psi}) - \sqrt{-\psi}}{\sqrt{-\psi^3}} \text{ for } \psi < 0 \\ &= \sum_{k=0}^{\infty} \frac{(-\psi)^k}{(2k+3)!} = \frac{1}{3!} - \frac{\psi}{5!} + \frac{\psi^2}{7!} - \frac{\psi^3}{9!} + \frac{\psi^4}{11!} - \frac{\psi^5}{13!} + \dots\end{aligned}$$

Whenever you use the series implementation, you must include *at least* the six terms shown above, but preferably eight to ten. This is especially true for orbits in which the transfer is hyperbolic because ψ is negative for these cases. When ψ is negative, you'll no longer receive improved convergence from an alternating series, so you must rely on absolute convergence. Figure 4-5 shows these functions.

Remember that ψ is negative whenever the orbit is hyperbolic. Although this case is not the most common, we can arrange Algorithm 24 (below) to permit an efficient formulation of the $c(\psi)$ functions by placing the elliptical cases first. It also allows a smooth transition over all ranges of ψ_n and tries to increase efficiency by placing the most common situations *first* (ψ positive) and the least common *last* (ψ negative).

ALGORITHM 24: Find $c_2 c_3 (\psi \Rightarrow c_2, c_3)$

IF $\psi > 1 \times 10^{-6}$

$$c_2 = \frac{1 - \cos(\sqrt{\psi})}{\psi} \quad c_3 = \frac{\sqrt{\psi} - \sin(\sqrt{\psi})}{\sqrt{\psi^3}}$$

ELSE

IF $\psi < -1 \times 10^{-6}$

$$c_2 = \frac{1 - \cosh(\sqrt{-\psi})}{\psi} \quad c_3 = \frac{\sinh(\sqrt{-\psi}) - \sqrt{-\psi}}{\sqrt{(-\psi)^3}}$$

ELSE

$$c_2 = \frac{1}{2} \quad c_3 = \frac{1}{6}$$

4.2.5 Solutions of Kepler's Equation

Working with Kepler's equation results in two main operations. First, determining eccentric (and true) anomaly given the mean anomaly is a transcendental operation and is the form most commonly identified as Kepler's equation. The inverse problem—if the eccentric anomaly (or true anomaly) and eccentricity are known—allows us to directly calculate the mean anomaly (and therefore time) using Eq. (4-6) and Eq. (4-7). We won't examine the second case because it doesn't require iteration.

Solving Kepler's equation has intrigued scientists for centuries. With literally hundreds of published techniques, it may seem impossible to choose one method. In fact, you can further break down the solutions into series and recursion methods. The series method gives good results over the span of eccentricity and angular values. I recommend Newton's iteration method because it converges to the order of 10^{-8} in three or four steps for

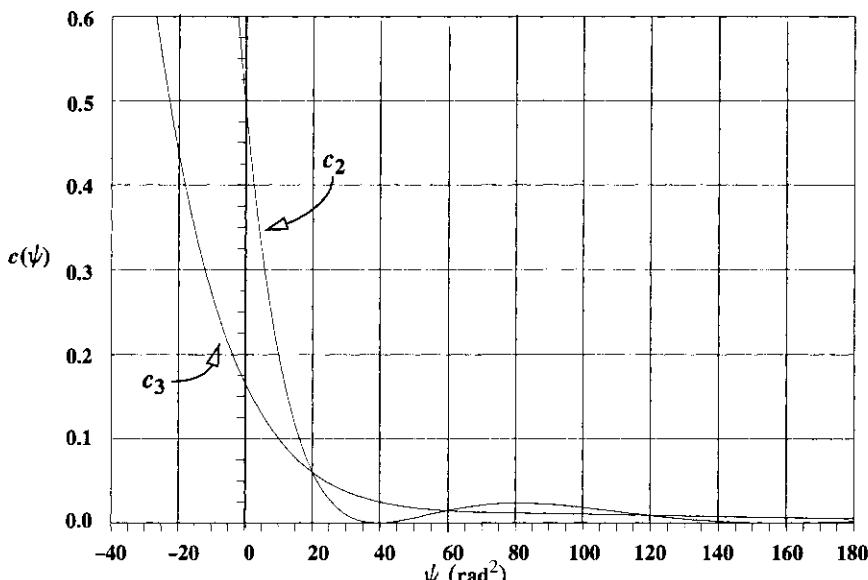


Figure 4-5. $c(\psi)$ Functions. Notice how quickly the variables tend towards zero. The zero points for c_2 occur where $\cos(\sqrt{\psi})$ equals one when ψ equals squared even multiples of 2π — $(2\pi)^2, (4\pi)^2, \dots$

most applications, which is extremely desirable. The bottom line is to use a method which gives correct results for your problem.

Elliptical Solutions

The simplest way to solve Kepler's equation is by successive approximation, but it's very time-consuming and inefficient on the computer. We'd like a way to get faster convergence—for example, using the slope to refine the next guess. This is *Newton's method*; it's also called the *Newton-Raphson* iteration because it uses successive derivatives of the function in the solution. We want to solve a function, $f(y) = 0$, for y . Let's assume $y = x + \delta$, where x is an approximate estimate for y , and δ is a small correction. If so, we can expand in a Taylor series:

$$0 = f(y) = f(x + \delta) \approx f(x) + f'(x)\delta + \frac{f''(x)\delta^2}{2!} + \dots$$

If we assume we can neglect second-order terms and higher (leaving only a slope term and hence a linear relationship), we can solve for δ .

$$\delta \approx -\frac{f(x)}{f'(x)}$$

By dropping the higher-order terms, we'll have to iterate to arrive at the correct solution. Now, $y_n = x_n + \delta_n$, which improves our estimate of y . In this case, we label subsequent values with a subscript n and repeat the process until $\delta_n < \epsilon$, where ϵ is the desired tolerance. The ending criterion is usually about $\epsilon = 1 \times 10^{-8}$. The "traditional" Newton-Raphson result is

$$x_{n+1} = x_n + \delta_n = x_n - \frac{f(x_n)}{f'(x_n)} \quad (4-50)$$

Note from Eq. (4-6) that $f(y) = f(E) = M - E + e \sin(E)$ for solving Kepler's equation. Therefore, $f'(E) = -1 + e \cos(E)$. I'll discuss how to select the initial guess for the eccentric anomaly in a later section on hybrid solutions. Initial estimates must be close enough to the true solution so we don't violate the linear assumption of the Newton-Raphson method. We also have to make sure input values of mean anomaly are within the range $\pm 2\pi$. Algorithm 25 presents the results.

ALGORITHM 25: *KepEqtnE* ($M, e \Rightarrow E$)

```

[ IF -  $\pi < M < 0$  or  $M > \pi$ 
    let  $E = M - e$ 
 ELSE
    let  $E = M + e$ 
]
```

LOOP

$$E_{n+1} = E_n + \frac{M - E_n + e \sin(E_n)}{1 - e \cos(E_n)}$$

UNTIL $|E_{n+1} - E_n| < \text{tolerance}$

By carefully selecting a starting value, we can make these computations more efficient. This process will continue until successive values of eccentric anomaly are close enough (usually within 10^{-8}). The only challenge is to pick an initial value for E_n . As you'll notice in the Fig. 4-6 plot of eccentric versus mean anomaly, the difference between the two angles isn't great, but it's more pronounced at higher eccentricities and around $\pm 90^\circ$. See *Hybrid Solutions*, page 241, for techniques.

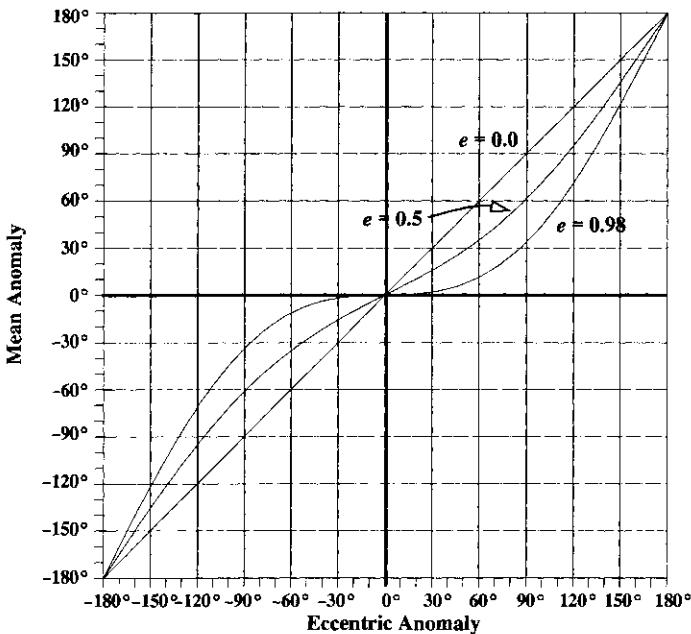


Figure 4-6. Mean Anomaly vs. Eccentric Anomaly. Although the difference between the eccentric and mean anomaly isn't great, as the eccentricity increases, the differences become larger.

Although not common in the literature, it's also useful to examine a plot of the mean anomaly versus the true anomaly. This allows us to compare all conic sections because mean anomaly and true anomaly are defined for all conic sections. Figure 4-7 plots the same eccentricities as those in Fig. 4-6. For now, consider an example:

▼ Example 4-1: Using Kepler's Equation.

GIVEN: $M = 235.4^\circ, e = 0.4$

FIND: E

Begin by letting $E_n = M - e = 235.4^\circ - 0.4 = 235.0^\circ$. Notice we subtracted the eccentricity. If we had added it, our first guess would have been wrong. Notice also the use of radians in the answer to this equation. A common error when doing these calculations by hand is to interchange degrees and radians. The first estimate of the eccentric anomaly is

$$\begin{aligned}
 E_{n+1} &= E_n + \frac{M - E_n + e \sin(E_n)}{1 - e \cos(E_n)} \\
 &= 235.0^\circ + \frac{235.4^\circ - 235.0^\circ + (0.4) \sin(235.0^\circ)}{1 - 0.4 \cos(235.0^\circ)} \\
 &\approx 3.848697 \text{ rad}
 \end{aligned}$$

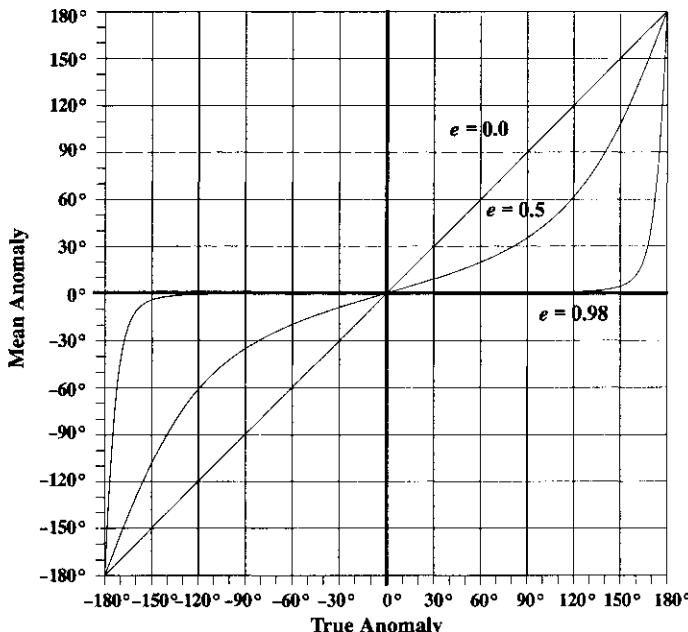


Figure 4-7. Mean Anomaly vs. True Anomaly. To obtain this plot, we first solve Kepler's equation and obtain the eccentric anomaly, E . We can then convert E to the true anomaly, ν . Notice the rapid slope changes as the eccentricity approaches 1.0. Because E , M , and ν are related through Eq. (4-7) and Eq. (4-9), these rapid changes translate to increased difficulty in solving Kepler's equation whenever e is near 1.0.

After the first iteration, the difference between E_n and E_{n+1} is still large. Therefore, reassign E_n and repeat the process:

$$E_{n+1} = E_n + \frac{M - E_n + e \sin(E_n)}{1 - e \cos(E_n)}$$

$$= 3.848\,697\,11 + \frac{4.108\,505\,06 - 3.848\,697\,11 + 0.4 \sin(3.848\,697\,11)}{1 - 0.4 \cos(3.848\,697\,11)}$$

$$= 3.848\,661\,46 \text{ rad}$$

▲ Notice that the values of E_n and E_{n+1} are much closer. Subsequent iterations will yield the results in Table 4-2. When the process is complete, $E = 220.512\,074\,767\,522^\circ$.

TABLE 4-2. Kepler's Equation—Example Solution. Results are for a test with $e = 0.4$, $M = 235.4^\circ = 4.108\ 505\ 059\ 194\ 65$ rad. We find the initial guess by subtracting the eccentricity from the mean anomaly in radians. ($E_n = 3.708\ 505\ 059\ 194\ 65$ rad). Notice how quickly the process converges.

Iteration (n)	E_{n+1}	$E_{n+1} - E_n$
1	3.846 971 106 002 65	$1.384\ 660\ 5 \times 10^{-1}$
2	3.848 661 460 814 60	$1.690\ 354\ 8 \times 10^{-3}$
3	3.848 661 745 097 16	$2.842\ 825\ 7 \times 10^{-7}$
4	3.848 661 745 097 17	$8.051\ 068\ 5 \times 10^{-15}$

Parabolic Solutions

To be complete, we should examine the separate cases for parabolic and hyperbolic motion. We can't readily graph the parabolic and hyperbolic anomalies (as we can the eccentric anomaly) because the eccentric anomaly represents an angle whereas the other two represent areas. We can graph these area measurements using the true anomaly and the resultant mean anomaly. For the parabolic case, the main difficulty in solving Kepler's problem is to find the parabolic anomaly, B , when the time is known. When we rewrite the parabolic time equation, the solution becomes obvious because the equation is simply a third-order polynomial:

$$\frac{B^3}{3} + B - n_p(t - T) = 0$$

An analytical solution of a third-order polynomial will result in at least one real root, which is the answer. See "Cubic Roots" on page 863.

Another solution uses trigonometric substitutions to solve for the parabolic anomaly. This formula is more common, but it's about 10% slower due to the numerous trigonometric substitutions. The clever part of this process is to recognize a relation that allows us to solve this problem. Thomas Barker first proposed an answer in 1757 when he published tables of solutions; it's commonly known as Barker's solution. In actuality, Barker wasn't the first to discover this process—Halley (1705) and Euler (1744) had also tackled it (Yeomans, 1991, 142–143)—but history favors Barker's accomplishment. The following identity will allow us to rewrite the parabolic time equation:

$$\frac{1}{3} \left(\lambda^3 - \frac{1}{\lambda^3} \right) = \left(\lambda - \frac{1}{\lambda} \right) + \frac{1}{3} \left(\lambda - \frac{1}{\lambda} \right)^3$$

Recognize that, if $B = \lambda - \frac{1}{\lambda}$, the right side looks very similar to the right side of Eq. (4-22). Substituting the identity into Eq. (4-22) gives us

$$n_p(t-T) = \frac{1}{3} \left(\lambda^3 - \frac{1}{\lambda^3} \right)$$

Now, letting $\lambda^3 = -\tan(s)$ and using trigonometric identities, change the right-hand side to

$$\begin{aligned} \frac{1}{3} \left(\lambda^3 - \frac{1}{\lambda^3} \right) &= \frac{1}{3} \left(-\tan(s) + \frac{1}{\tan(s)} \right) = \frac{1}{3} \left(\frac{-1}{\cot(s)} + \cot(s) \right) \\ &= \frac{1}{3} \frac{\cot^2(s) - 1}{\cot(s)} = \frac{2}{3} \cot(2s) \end{aligned}$$

Substituting this result into the parabolic time equation results in

$$n_p(t-T) = \frac{2}{3} \cot(2s)$$

Because the change in time is known, you can uniquely determine s . Remember that s is really cubic in λ , so to find the *single* solution, let $(-\tan(w)) = \lambda$. Using the relation,

$$\tan^3(w) = \tan(s)$$

find the angle w . The final step is to find the value of parabolic anomaly; thus, you must use the original substitution and trigonometric identities:

$$B = \tan\left(\frac{v}{2}\right) = \left(\lambda - \frac{1}{\lambda}\right) = -\tan(w) + \frac{1}{\tan(w)} = 2 \cot(2w)$$

The algorithm for this solution is

ALGORITHM 26: *KepEqtnP* ($\Delta t, p \Rightarrow B$)

$$n_p = 2 \sqrt[3]{\frac{\mu}{p^3}}$$

Solve cubic or Solve Barker's equation

$$\frac{B^3}{3} + B - n_p(\Delta t) = 0 \quad \cot(2s) = \frac{3}{2} n_p \Delta t$$

$$\tan^3(w) = \tan(s)$$

$$B = 2 \cot(2w)$$

Let's consider an example for parabolic orbits:

▼ **Example 4-2. Using Kepler's Equation for Parabolic Orbits.**

GIVEN: $\Delta t = 53.7874 \text{ min (4.0 TU)}$, $p = 25,512 \text{ km (4.0 ER)}$, $e = 1$

FIND: B

Begin by finding the mean motion:

$$n_p = 2 \sqrt{\frac{\mu}{p^3}} = 2 \sqrt{\frac{1}{4^3}} = 0.25 \text{ rad/TU}$$

Now, use the time equation:

$$\cot(2s) = \frac{3}{2} n_p \Delta t = \frac{3}{2} (0.25) (4.0) = 1.5 \text{ rad}$$

which means $s = 16.845\ 03^\circ$

$$\tan^3(w) = \tan(s) = \tan(16.845\ 03^\circ)$$

$$w = 33.881\ 012^\circ$$

Finally, find $B = 2\cot(2w) = 2 \cot(2(33.881\ 012^\circ))$

▲ $B = 0.817\ 731\ 6 \text{ rad}$

Figure 4-8 shows a graph of the mean anomaly and the true anomaly for a parabolic orbit. Notice the rapid growth of mean anomaly as true anomaly increases. Because $M = n_p \Delta t$, mean anomaly continues to grow as Δt increases. For most practical purposes, Fig. 4-8 seems to indicate that the true anomaly for a parabolic orbit is limited. Although the true anomaly theoretically exists to $\pm 180^\circ$, it becomes a slowly varying function of time past values of about 130° to 150° .

Hyperbolic Solutions

Finally, let's examine solutions to Kepler's equation for the hyperbolic case. As in the elliptical case, the hyperbolic solution is transcendental and requires iteration. We use the Newton-Raphson iteration method for general convergence. Here, $f(H) = M - e \sinh(H) + H$, and $f'(H) = -e \cosh(H) + 1$. The result, as shown in Algorithm 27 (below), is very similar to the elliptical case except for the initial guess. See *Hybrid Solutions*, page 241, for details on the initial guess.

ALGORITHM 27: KepEqtnH ($M, e \Rightarrow H$)

```

    IF  $e < 1.6$ 
        IF  $-\pi < M < 0$  or  $M > \pi$ 
             $H = M - e$ 
        ELSE
             $H = M + e$ 
    ELSE

```

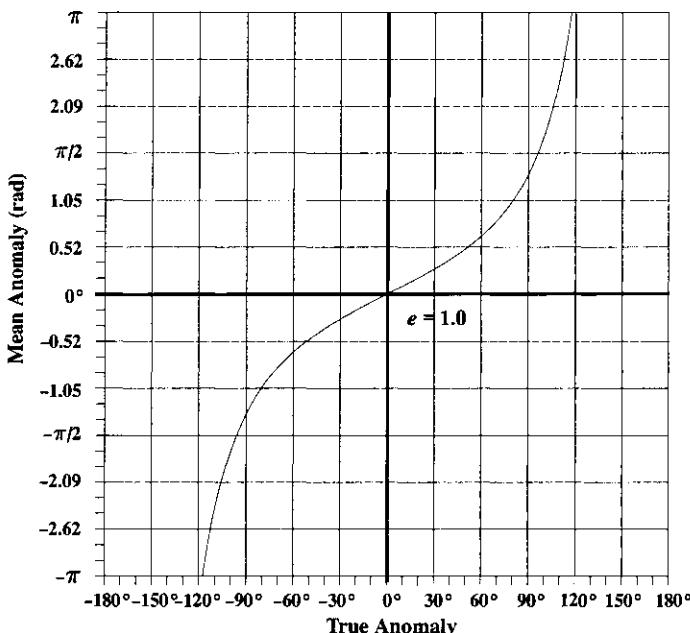


Figure 4-8. Mean Anomaly vs. True Anomaly for Parabolic Orbits. For a parabolic orbit, the mean anomaly is best described as a radian measure. Although the true anomaly appears to be constrained to certain values, at infinite distance true anomaly will equal $\pm 180^\circ$.

```

    IF  $e < 3.6$  and  $|M| > \pi$ 
         $H = M - \text{SIGN}(M) e$ 
    ELSE
         $H = \frac{M}{e - 1}$ 
    LOOP

```

$$H_{n+1} = H_n + \frac{M - e \sinh(H_n) + H_n}{e \cosh(H_n) - 1}$$

UNTIL $|H_{n+1} - H_n| <$ tolerance

▼ **Example 4-3. Using Kepler's Equation for Hyperbolic Orbits.**

GIVEN: $M = 235.4^\circ$, $e = 2.4$

FIND: H

Begin with an initial guess. Because the eccentricity is greater than 1.6 and less than 3.6, and the mean anomaly is greater than π ,

$$H_n = M - \text{SIGN}(M)e = (4.108\ 505\ 059) - (+) 2.4 = 1.708\ 505\ 059 \text{ rad}$$

Then, making sure you use radians, you'll find the hyperbolic anomaly is

$$H_{n+1} = H_n + \frac{M - e \sinh(H_n) + H_n}{e \cosh(H_n) - 1} = 1.607\ 435\ 562\ 6 \text{ rad}$$

The difference in iteration values is still too large, and after another iteration,

$$H_{n+1} = 1.601\ 396\ 28 \text{ rad}$$

▲ The process is converging to the answer: $H = 1.601\ 376\ 144 \text{ rad}$. Table 4-3 summarizes the results.

TABLE 4-3. Values for Kepler's Equation—Hyperbolic Example 4-3. Each iterative value is in radians—from the starting values of $e = 2.4$, $M = 235.4^\circ = 4.108\ 505\ 059\ 194\ 65 \text{ rad}$. The initial guess is $H_n = 1.708\ 505\ 059$.

Iteration <i>n</i>	H_{n+1}	$H_{n+1} - H_n$
1	1.607 435 562 6	$-1.010\ 69 \times 10^{-1}$
2	1.601 396 281 5	$-6.039\ 28 \times 10^{-3}$
3	1.601 376 145 15	$-2.013\ 638 \times 10^{-5}$
4	1.601 376 144 9	$-2.228\ 866\ 8 \times 10^{-10}$

In the elliptical case, we saw that Fig. 4-7 gave us insight into solving Kepler's equation. Because M , H , and ν are related via Eq. (4-30) and Eq. (4-38), we can use the elliptical results to estimate the qualitative behavior of the hyperbolic case. I haven't shown a plot of M vs. H for hyperbolic orbits, but it's very similar to the elliptical case. Thus, plots of mean anomaly vs. true anomaly (Fig. 4-9) indirectly help us gain insight into the solution of Kepler's equation for hyperbolic orbits. Notice, as with the parabolic case, the anomalies change slowly as mean anomaly increases. You can also see the limits on ν as the orbit approaches the vertical lines (asymptotes).

Series Solutions

Although computers can efficiently solve Kepler's equation through iteration, many sources apply series solutions to the problem. Modern computers make these solutions

less useful, but astrodynamists still need to know them. Although I'm presenting several approximations, remember that their very nature implies an infinite series which we must truncate at some point. For instance, Plummer ([1918] 1960, 37) shows a series representation

$$E = M + 2 \sum_{k=1}^{\infty} \frac{\sin(kM)}{k} (J_k(ke))$$

where J_k are modified Bessel functions of the first kind. Incidentally, Bessel functions were developed to solve Kepler's equation (Taff, 1985, 34). Remember that Bessel functions are actually the coefficients of the Fourier sine series (Battin, 1987, 206–207) and we find them as

$$J_k(ke) = \sum_{j=0}^{\infty} \frac{(-1)^j}{j!(k+j)!} \left(\frac{ke}{2}\right)^{k+2j}$$

If $k > 0, j = 0, 1, 2, \dots$

If $k < 0, j = -k, -k+1, -k+2, \dots$

Depending on how many terms you take and the desired accuracy, this method *may* be very fast. Danby (1992, 198–200) shows how to develop the series approach for the eccentric anomaly, and Taff (1985, 61) lists a number of series for the eccentric, mean, and true anomalies:

$$\begin{aligned} E &= M + \left(e - \frac{e^3}{8} + \frac{e^5}{192}\right) \sin(M) + \left(\frac{e^2}{2} - \frac{e^4}{6}\right) \sin(2M) + \left(\frac{3e^3}{8} - \frac{27e^5}{128}\right) \sin(3M) + \left(\frac{e^4}{3}\right) \sin(4M) + \dots \\ E &= \nu - \left(e + \frac{e^4}{12}\right) \sin(\nu) + \left(\frac{e^2}{4} + \frac{e^4}{8}\right) \sin(2\nu) - \left(\frac{e^3}{12}\right) \sin(3\nu) + \left(\frac{e^4}{32}\right) \sin(4\nu) + \dots \\ \nu &= E + \left(e + \frac{3e^3}{12}\right) \sin(E) + \left(\frac{e^2}{4} + \frac{e^4}{24}\right) \sin(2E) + \left(\frac{e^3}{12}\right) \sin(3E) + \left(\frac{e^4}{96}\right) \sin(4E) + \dots \\ \nu &= M + \left(2e - \frac{e^3}{4} + \frac{5e^5}{96}\right) \sin(M) + \left(\frac{5e^2}{4} - \frac{11e^4}{24}\right) \sin(2M) + \left(\frac{13e^3}{12} - \frac{43e^5}{64}\right) \sin(3M) \\ &\quad + \left(\frac{103e^4}{96}\right) \sin(4M) + \dots \\ M &= \nu - 2e \sin(\nu) + \left(\frac{3e^2}{4} + \frac{e^4}{8}\right) \sin(2\nu) - \left(\frac{e^3}{3} + \frac{e^5}{8}\right) \sin(3\nu) + \left(\frac{5e^4}{32}\right) \sin(4\nu) + \dots \end{aligned} \tag{4-51}$$

Some of these equations have their roots in the early history of astrodynamics (Sec 1.1.1). While developing the deferent and epicycle theory, the early Greeks termed the angle between the circle and the offset circle as the *equation of center*. Today, we define the *equation of the center* as the difference $\nu - M$.

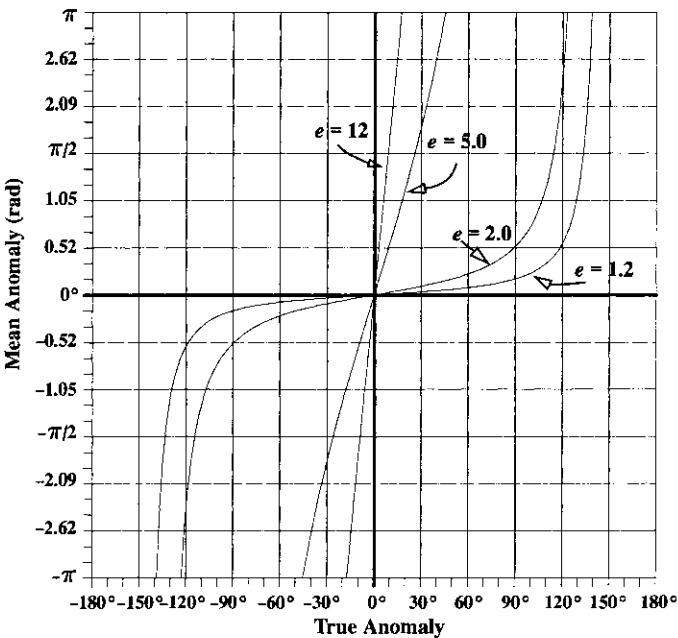


Figure 4-9. Mean Anomaly vs. True Anomaly for Hyperbolic Orbits. Notice the effective limits on true anomaly. Unlike the parabola, these limits are real because all hyperbolic orbits are limited by the asymptotes. High eccentricities produce nearly straight lines.

Hybrid Solutions

When we use the Newton-Raphson iteration to solve Kepler's equation, a key concern is selecting an initial guess that will require fewer iterations *and* shorter computing times—for all orbit types. Many initial-guess schemes require too much time for the initial calculation compared to operations for the actual iteration. We could program Eq. (4-51) to include 20 or 30 terms, but the expense of evaluating the long transcendental equations wouldn't yield a more accurate solution for *most* cases.

As a baseline, consider the initial guess of letting the eccentric (and hyperbolic) anomaly equal the mean anomaly. Of course, an exact solution is always available for the parabolic case by solving a cubic equation. Figure 4-10 shows the number of iterations for these cases. The time it requires is the baseline for comparison to the other methods.

Because Eq. (4-51) shows the series expansion for the eccentric anomaly in terms of the mean anomaly, it is reasonable to explore the added benefits of including higher-order terms in the expansion for the initial guess. Of course, as we include more terms, the complexity rises, so increased computational time may offset the fewer iterations. Now consider a case for which the initial guess is equal to the mean anomaly plus the

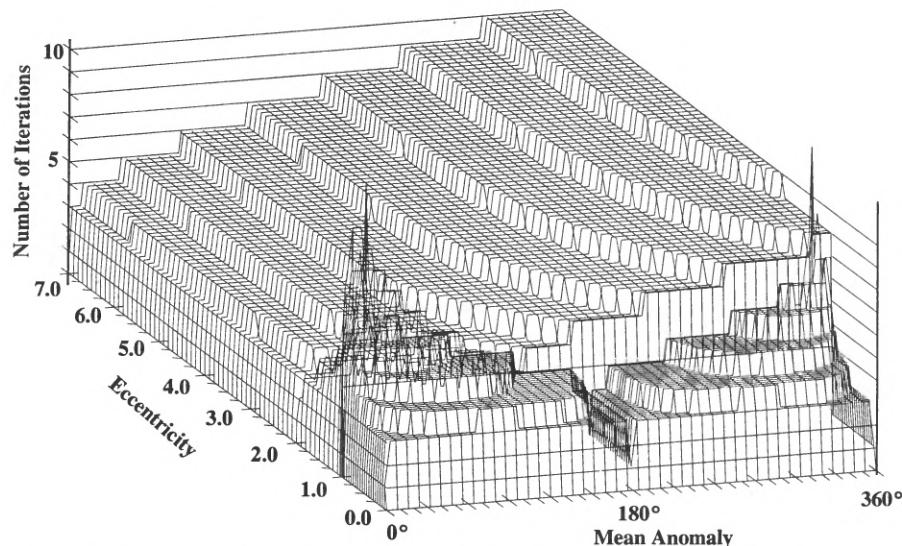


Figure 4-10. Solution of Kepler's Equation, $E, H = M$. This figure shows a complete spectrum of values through $\pm 360^\circ$, which enables correct sign determination and is crucial to effective computations. Notice the difficulty this method has at low values of mean anomaly and eccentricity near 1.0. Although we may not need a robust solution for everyday operations, some situations may require it.

eccentricity. Examination of Fig. 4-11 shows results in which the eccentricity was added or subtracted to produce symmetrical results over the complete range of eccentric anomaly. The solution is to subtract the eccentricity in cases for which the mean anomaly is less than zero (but larger than -180°) or larger than 180° and to add the eccentricity elsewhere. This precaution presumes we've already reduced the value of mean anomaly to the range of $\pm 360^\circ$. Equivalently,

$$\text{IF } -\pi < M < 0 \text{ or } M > \pi$$

let $E, H = M - e$

ELSE

let $E, H = M + e$

Figure 4-11 shows the results. Although not evident in the figure, this particular method is about 15% faster than the original approach. It converges in fewer iterations in almost all cases. Note the improved performance in the region where the eccentricity is less than 1.0.

Lastly, consider the case in which we include the first-order terms of eccentricity:

$$E = M + e \sin(M) + \frac{e^2}{2} \sin(2M)$$

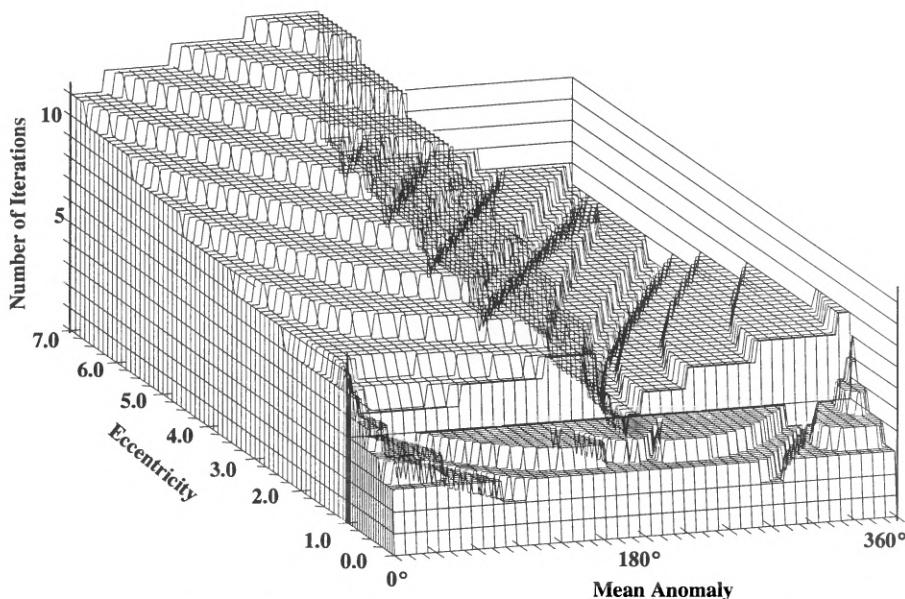


Figure 4-11. Solution of Kepler's Equation, $E, H = M + e$. Note the improvement in the eccentric case by simply adding the eccentricity when $0^\circ < M < 360^\circ$ and subtracting elsewhere. The hyperbolic case shows some improvement over Fig. 4-10, but only from 180° to 360° .

Figure 4-12 shows the required number of iterations with this initial guess for elliptical orbits. The hyperbolic solution changes dramatically because we've used a straight-line approximation. From Fig. 4-9, you can see that, as eccentricity gets large, the M vs. H curve becomes almost linear, permitting very effective first guesses. I experimentally determined the numerical divisions for eccentricity in Algorithm 27, so consider them only an initial baseline.

The results in Fig. 4-12 may be difficult to interpret. Although this method solves the problem of high eccentricity and low mean anomaly, the processing time is only about 2% faster than the original baseline. Clearly, the extra mathematical calculations, while speeding the number of iterations, have taken their toll in computing time. Be aware that if we choose a solution based solely on average iterations, the results could be misleading because just a few cases which take 30–50 iterations can easily skew the results. Thus, we should choose an approach based on overall computing times for the entire range of values. The iteration for the hyperbolic case in Algorithm 27 is efficient, and the limits of eccentricity give “reasonable” performance.

For the best efficiency over a broad range of values for eccentricity and mean anomaly, use the eccentric-anomaly conditions in Fig. 4-11 (Algorithm 25), the cubic solution for

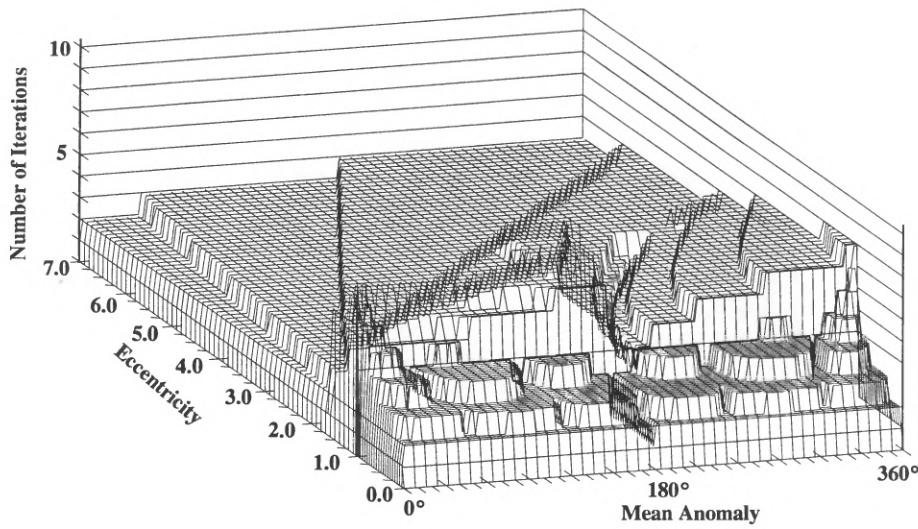


Figure 4-12. Solution of Kepler's Equation, $E = M + e \sin(M) + \frac{e^2}{2} \sin(2M)$. Including more terms in the initial guess improves the performance, but not significantly. The hyperbolic solution is actually a hybrid approach combining several desirable features of solution from the linearity of the problem. Notice that the hyperbolic solution actually improves as eccentricity increases.

parabolic cases (Algorithm 26), and the technique for solving hyperbolic cases in Fig. 4-12 (Algorithm 27).

4.2.6 Summary and Related Formulas

Using the basics of Kepler's equation, we may summarize the formulas and develop alternate expressions for the flight-path angle in terms of the various anomalies. As mentioned earlier, the flight-path angle is crucial to accurately determining the satellite's attitude. It's especially important for accurate perturbation analysis. We can represent the flight-path angle in terms of the eccentric anomaly, E , for elliptical orbits; the parabolic anomaly, B , for parabolic orbits; and the hyperbolic anomaly, H , for hyperbolic orbits.

First, we'll summarize the results of Kepler's equations. These algorithms will be useful in future operations.

ALGORITHM 28: *v to Anomaly* ($e, v \Rightarrow E, B, H$)

if $e < 1.0$	if $e = 1.0$	if $e > 1.0$
$\text{SIN}(E) = \frac{\sqrt{1-e^2} \text{SIN}(v)}{1+e\text{COS}(v)}$	$B = \text{TAN}\left(\frac{v}{2}\right)$	$\text{SINH}(H) = \frac{\sqrt{e^2-1} \text{SIN}(v)}{1+e\text{COS}(v)}$
$\text{COS}(E) = \frac{e+\text{COS}(v)}{1+e\text{COS}(v)}$		$\text{COSH}(H) = \frac{e+\text{COS}(v)}{1+e\text{COS}(v)}$

ALGORITHM 29: *Anomaly to v* ($e, E, B(p,r), H \Rightarrow v$)

if $e < 1.0$	if $e = 1.0$	if $e > 1.0$
$\text{SIN}(v) = \frac{\sqrt{1-e^2} \text{SIN}(E)}{1-e\text{COS}(E)}$	$\text{SIN}(v) = \frac{pB}{r}$	$\text{SIN}(v) = \frac{-\sqrt{e^2-1} \text{SINH}(H)}{1-e\text{COSH}(H)}$
$\text{COS}(v) = \frac{\text{COS}(E)-e}{1-e\text{COS}(E)}$	$\text{COS}(v) = \frac{p-r}{r}$	$\text{COS}(v) = \frac{\text{COSH}(H)-e}{1-e\text{COSH}(H)}$

The derivation for the cosine of the flight-path angle begins with Eq. (2-59). Substitute Eq. (4-11).

$$\text{COS}(\phi_{fpa}) = \sqrt{\frac{a^2(1-e^2)}{a(1-e\text{COS}(E)) \{2a-a(1-e\text{COS}(E))\}}}$$

Rearrange the denominator and simplify the equation to get

$$\text{COS}(\phi_{fpa}) = \sqrt{\frac{1-e^2}{(1-e\text{COS}(E))(1+e\text{COS}(E))}}$$

After expanding the denominator, you'll have the final result:

$$\text{COS}(\phi_{fpa}) = \sqrt{\frac{1-e^2}{1-e^2\text{COS}^2(E)}} \quad (4-52)$$

Find the sine expression of the flight-path angle using the Pythagorean relation of Eq. (C-8), which requires you to solve

$$\frac{1-e^2}{1-e^2\text{COS}^2(E)} + \text{SIN}^2(\phi_{fpa}) = 1$$

Once you expand the $\cos^2(E)$ term with Eq. (C-8), you can quickly get

$$\sin(\phi_{fpa}) = \frac{e \sin(E)}{\sqrt{1 - e^2 \cos^2(E)}} \quad (4-53)$$

The solution for parabolic orbits is a little different. First, find the tangent expression for ϕ_{fpa} by using Eq. (2-58) and Eq. (2-59):

$$\tan(\phi_{fpa}) = \frac{e \sin(\nu)}{1 + e \cos(\nu)}$$

Because eccentricity is identically 1.0 for parabolic orbits, the formula is apparent for the tangent half-angle. Solve by taking the arc-tangent of both sides:

$$\phi_{parabolic} = \frac{\nu}{2} \quad (4-54)$$

Finally, hyperbolic orbits are similar to the elliptical case. Begin with Eq. (2-58) and substitute Eq. (4-33) for ($\sin(\nu)$):

$$\sin(\phi_{fpa}) = \frac{-e \sqrt{e^2 - 1} \sinh(H)}{\sqrt{1 + 2e \cos(\nu) + e^2(1 - e \cosh(H))}}$$

Now, use Eq. (4-31) to replace the ($\cos(\nu)$) term. After you find a common denominator,

$$\sin(\phi_{fpa}) = -e \sinh(H) \sqrt{\frac{(e^2 - 1)(1 - e \cosh(H))}{(1 - e \cosh(H))^2(1 + e \cosh(H) - e^2 - e^3 \cosh(H))}}$$

Simplify the denominator by rearranging the last term in parentheses to equal

$$\sin(\phi_{fpa}) = -e \sinh(H) \sqrt{\frac{e^2 - 1}{(1 - e \cosh(H))(1 - e^2)(1 + e \cosh(H))}}$$

Then, multiply the denominator and distribute the negative sign to both denominator terms. The final result is

$$\sin(\phi_{fpa}) = \frac{-e \sinh(H)}{\sqrt{e^2 \cosh^2(H) - 1}} \quad (4-55)$$

Develop the cosine expression using Eq. (C-8). For the solution, you'll need

$$\frac{e^2 \sinh^2(H)}{e^2 \cosh^2(H) - 1} + \cos^2(\phi_{fpa}) = 1$$

Now use Eq. (C-9) to expand the hyperbolic sine term and get the final result, remembering the denominator must be the same as for the sine relation:

$$\cos(\phi_{fpa}) = \sqrt{\frac{e^2 - 1}{e^2 \cosh^2(H) - 1}} \quad (4-56)$$

4.3 Kepler's Problem

Actually, two classes of problems arise from Kepler's equation: the time to travel between two known points on any type of orbit and the location of a satellite in orbit after a certain amount of time. Kepler's equation captures the first problem. But the second class leads to what we know as *Kepler's problem* or more generally, *propagation*. Simply stated, this problem is to find a satellite's future location given the last known position and velocity vectors at a particular time. Small ([1804] 1963, 296–297) says:

This problem has, ever since the time of Kepler, continued to exercise the ingenuity of the ablest geometers; but no solution of it which is rigorously accurate has been obtained. Nor is there much reason to hope that the difficulty will ever be overcome; because, as Kepler remarks, a circular arch, and any of the lines drawn about it, as the sines, are quantities of different kinds; and it transcends the powers of geometry to express their mutual relation with perfect accuracy.

This statement may give you the wrong impression. Although there are no exact *algebraic* solutions, there are rigorous complete solutions using circular transcendental functions and the solutions of Kepler's equation. Siewert and Burniston (1972) claim these methods find a closed-form solution in the complex plane. This approach contrasts with perturbation approximations (Chap. 8) which aren't rigorous but are very useful. Fortunately, the modern computer allows calculations of these quantities with almost arbitrary accuracy. Figure 4-13 shows the general situation.

Although at first glance this seems to be an easy problem, it is extremely important to astrodynamics because its solution is the basis of virtually all planning and operational simulations. The method uses the results of Kepler's equation from Sec. 4.2.

4.3.1 Solution Techniques

We'll discuss three broad ways to solve Kepler's problem for two-body motion: an orbital-element technique, the broad series method using f and g functions, and the universal-variable techniques. Chapter 8 discusses the effects of perturbations on this process.

Orbital Elements

The simplest way to solve Kepler's problem involves classical orbital elements. In the time between successive positions in the orbit, the only variable to change (no perturbations included yet!) is the true anomaly, ν , or associated parameters for the special orbits (u , λ_t , $\tilde{\omega}_{true}$). Knowing the conversion between the anomalies and the mean motion

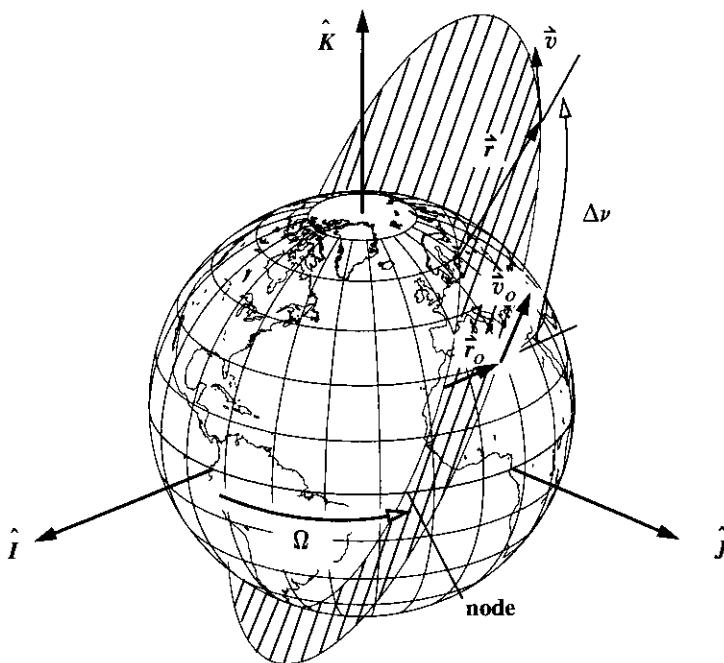


Figure 4-13. Geometry for Kepler's Problem. The concept behind Kepler's problem is the propagation of the satellite's state vector from an initial epoch to a future time.

solved with Kepler's equation, we can simply update the individual anomalies and then update the position and velocity vectors.

The procedure first finds the classical orbital elements, updates the mean anomaly through the applicable time-of-flight equation, and then re-forms the position and velocity vectors. The usual problem scenario is to know the original vectors \vec{r}_o , \vec{v}_o , and the amount of time desired to move the satellite's state vector into the future, Δt . Solving this scenario requires several steps.

First, through techniques described in Algorithm 5 (**ELORB**), find the original value of the mean anomaly, M_o , and update it. The reason we chose our definition for the parabolic motion is now apparent because it allows *one* expression for the value of mean anomaly for all orbit types ($M = M_o + n \Delta t$).

Next, knowing the mean anomaly and eccentricity, solve Kepler's *equation* for the intermediate determination of the eccentric, parabolic, or hyperbolic anomaly. Then use the results of Sec. 4.2 to find the true anomaly and, in turn, use the true anomaly at the *new* position to find the updated position and velocity vectors (Algorithm 6, **RANDV**).

You may wonder about solving the special cases of circular-inclined and equatorial orbits. Elliptical equatorial orbits present no difficulties because true anomaly is defined.

For the circular orbits, set the eccentric anomaly equal to the argument of latitude or true longitude; in the reverse cases, set each angle equal to the eccentric anomaly. Notice also that the iteration for Kepler's equation is unnecessary with circular orbits because the two quantities are equal! Solve as shown in Algorithm 30, below.

ALGORITHM 30: *KeplerCOE* ($\vec{r}_o, \vec{v}_o, \Delta t \Rightarrow \vec{r}, \vec{v}$)

ELORB ($\vec{r}_o, \vec{v}_o \Rightarrow a, e, i, \Omega, \omega, \nu, M(u, \lambda_{true}, \tilde{\omega}_{true})$)

```

[ IF  $e \neq 0$ 
    vtoAnomaly ( $e, \nu \Rightarrow E, B, H$ )
 ELSE
     $E_o = u$  or  $E_o = \lambda_{true}$ 
    if  $e < 1.0$            if  $e = 1.0$            if  $e > 1.0$ 
 $M_o = E_o - e\sin(E_o)$        $\vec{h} = \vec{r}_o \times \vec{v}_o$        $M_o = e\sinh(H_o) - H_o$ 
 $M = M_o + n\Delta t$            $p = \frac{h^2}{\mu}$            $M = M_o + n\Delta t$ 
KepEqtnE ( $M, e \Rightarrow E$ )       $M_o = B_o + \frac{B_o^3}{3}$       KepEqtnH ( $M, e \Rightarrow H$ )
                                         KepEqtnP ( $\Delta t, p \Rightarrow B$ )

```

```

[ IF  $e \neq 0$ 
    Anomalytov ( $e, E, B, H \Rightarrow \nu$ )
 ELSE
     $u = E$  or  $\lambda_{true} = E$ 

```

RANDV ($a, e, i, \Omega, \omega, \nu(u, \lambda_{true}, \tilde{\omega}_{true}) \Rightarrow \vec{r}, \vec{v}$)

Classical Formulas Using f and g Functions

This method allows multiple propagations from an original epoch using the same functions. For cases in which we don't consider thrusting and maneuvering, this is a fast alternative to other methods. Specific formulas depend on the type of orbit being propagated.

To use the *f* and *g* functions, assume a solution that linearly combines the initial position and velocity vectors. In the perifocal coordinate system (Fig. 4-14), write an expression for the position and velocity vectors simply from the components *x* and *y* in each axis:

$$\vec{r} = x\hat{P} + y\hat{Q} \quad \vec{v} = \dot{x}\hat{P} + \dot{y}\hat{Q}$$

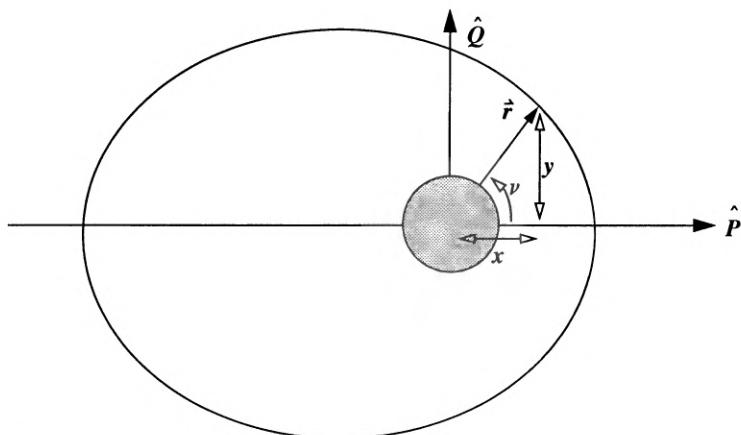


Figure 4-14. Geometry for the Perifocal Coordinate System. We can break a satellite's position into two components (x, y) in the orbital plane.

If you assume a linear combination of the initial position and velocity vectors and use arbitrary functions, f and g , you'll get

$$\begin{aligned}\dot{\vec{r}} &= f\vec{r}_o + g\vec{v}_o \\ \dot{\vec{v}} &= \dot{f}\vec{r}_o + \dot{g}\vec{v}_o\end{aligned}\tag{4-57}$$

To find the specific f and g functions, start by crossing $\dot{\vec{r}}$ into \vec{v}_o to isolate the f expression:

$$\dot{\vec{r}} \times \vec{v}_o = (f\vec{r}_o + g\vec{v}_o) \times \vec{v}_o = f(\vec{r}_o \times \vec{v}_o) + g(\vec{v}_o \times \vec{v}_o)$$

Because the last cross product is zero and the first has a value only in the \hat{W} direction, you can consider the term a magnitude. Also, recognize the angular-momentum as the term multiplying f . Therefore,

$$xy_o - \dot{x}_o y = fh$$

Solving results in

$$f = \frac{xy_o - \dot{x}_o y}{h}\tag{4-58}$$

You can directly find the derivative of f because \dot{x}_o and \dot{y}_o are components of the initial velocity and are therefore constants with zero derivatives. Remember that the derivative of the angular momentum is also zero. Thus,

$$\dot{f} = \frac{\dot{x}\dot{y}_o - \dot{x}_o\dot{y}}{h} \quad (4-59)$$

Next, to find the g expression, expand $\dot{\vec{r}}_o \times \dot{\vec{r}}$:

$$\dot{\vec{r}}_o \times \dot{\vec{r}} = \dot{\vec{r}}_o \times (f\dot{\vec{r}}_o + g\dot{\vec{v}}_o) = f(\dot{\vec{r}}_o \times \dot{\vec{r}}_o) + g(\dot{\vec{r}}_o \times \dot{\vec{v}}_o)$$

Examine only the component in the \hat{W} direction (the first term is zero):

$$x_o y - x \dot{y}_o = gh$$

Solving yields

$$g = \frac{x_o y - x \dot{y}_o}{h} \quad (4-60)$$

Differentiating gives the dot term (x_o and y_o are initial constants):

$$\dot{g} = \frac{x_o \dot{y} - \dot{x} \dot{y}_o}{h} \quad (4-61)$$

Eq. (4-57) is a linear combination of the f and g expressions above. This formula's advantage is that it works for all f and g functions we can derive. Because the f and g functions combine the original vectors and angular momentum—a constant for two-body motion—we can check their accuracy. Now look at the cross product:

$$\begin{aligned} \dot{\vec{h}} &= \dot{\vec{r}} \times \dot{\vec{v}} = (f\dot{\vec{r}}_o + g\dot{\vec{v}}_o) \times (f\dot{\vec{r}}_o + g\dot{\vec{v}}_o) \\ &= f\dot{f}\dot{\vec{r}}_o \times \dot{\vec{r}}_o + f\dot{g}\dot{\vec{r}}_o \times \dot{\vec{v}}_o + f\dot{g}\dot{\vec{v}}_o \times \dot{\vec{r}}_o + g\dot{g}\dot{\vec{v}}_o \times \dot{\vec{v}}_o \\ &= f\dot{g}\dot{\vec{h}} - f\dot{g}\dot{\vec{h}} \end{aligned}$$

This can only be true (for $\dot{\vec{h}} \neq 0$) if

$$1 = f\dot{g} - f\dot{g} \quad (4-62)$$

We can also adopt a matrix notation and write the update in Eq. (4-57) with Φ as a matrix containing scalar 3×3 matrices of the f and g functions.

$$\bar{X}_{k+1} = \begin{bmatrix} f & g \\ \dot{f} & \dot{g} \end{bmatrix} \bar{X}_k = \Phi \bar{X}_k$$

The Φ matrix is called the ***state-transition matrix*** because it moves the state through time:

$$\bar{X}(t_{k+1}|t_k) = \int_{t_k}^{t_{k+1}} \dot{\bar{X}}_k dt + \bar{X}_k$$

The advantage of this approach is the ease with which we can transform position and velocity vectors through time, once we've calculated the f and g expressions. It's particularly useful for Kalman filtering applications, which we'll discuss in Chap. 9.

Now we can use the f and g equations with various solutions of the PQW components, depending on what information is available for the problem. To determine the f and g functions for various sets of initial data, we must first find the x and y components. Herrick (1971, 67–75 and 207–219) develops many of these relations, and Bate, Mueller, and White (1971, 212–220) show several derivations. I've used these references but changed the parabolic formulas to match notations in this book.

If the true anomaly, ν , is known, define the x and y terms based on the final value of ν :

$$x = r \cos(\nu) \quad y = r \sin(\nu)$$

Find the derivatives using the trajectory equation, Eq. (2-8) and Eq. (2-9), and simple differentiation (of course, no perturbations).

$$\dot{x} = -\sqrt{\frac{\mu}{p}} \sin(\nu) \quad \dot{y} = \sqrt{\frac{\mu}{p}} (e + \cos(\nu))$$

To find the f and g functions, substitute each of the x and y expressions into Eq. (4-58) to Eq. (4-61). Carefully examine the equations to discover that the change in true anomaly must be known for the f and g functions because both *initial* and *final* values of x and y are used to form the f and g functions. For the special case in which the initial point is at perigee, you can further simplify the equations, but this is usually *not* the case. After substituting, you get

$$f = 1 - \left(\frac{r}{p} \right) (1 - \cos(\Delta\nu)) \quad g = \frac{rr_o \sin(\Delta\nu)}{\sqrt{\mu p}} \\ \dot{f} = \sqrt{\frac{\mu}{p}} \tan\left(\frac{\Delta\nu}{2}\right) \left(\frac{1 - \cos(\Delta\nu)}{p} - \frac{1}{r} - \frac{1}{r_o} \right) \quad \dot{g} = 1 - \left(\frac{r_o}{p} \right) (1 - \cos(\Delta\nu)) \quad (4-63)$$

If the eccentric anomaly, E , is known, use relations developed in the solution to Kepler's equation [Eq. (4-8)] in Sec. 4.2 [see also Battin (1987, 162)]:

$$x = a(\cos(E) - e) \quad y = a\sqrt{1-e^2}\sin(E) \quad \dot{E} = \frac{1}{r}\sqrt{\frac{\mu}{a}} \\ \dot{x} = -a\dot{E}\sin(E) \quad \dot{y} = a\dot{E}\sqrt{1-e^2}\cos(E)$$

To find the overall expressions for the eccentric anomaly, use the same procedure we used for the true anomaly.

$$\begin{aligned} f &= 1 - \frac{a}{r_o} (1 - \cos(\Delta E)) & g &= (t - t_o) - \sqrt{\frac{a^3}{\mu}} (\Delta E - \sin(\Delta E)) \\ \dot{f} &= \frac{-\sqrt{\mu a} \sin(\Delta E)}{r_o r} & \dot{g} &= 1 - \frac{a}{r} (1 - \cos(\Delta E)) \end{aligned} \quad (4-64)$$

Assume the parabolic anomaly, B , is known. Then, by Eqs. (4-24), (4-26), and (4-27), you have

$$\begin{aligned} x &= \left(\frac{p}{2}\right)(1 - B^2) & y &= pB \\ \dot{x} &= \frac{-\sqrt{\mu p}}{r} B & \dot{y} &= \frac{-\sqrt{\mu p}}{r} \end{aligned}$$

The resulting f and g functions are

$$\begin{aligned} f &= \frac{1 - B^2 + 2BB_o}{1 + B_o^2} & g &= \frac{p^2 \Delta B (1 + BB_o)}{2h} \\ \dot{f} &= \frac{4h \Delta B}{p^2 (1 + B^2) (1 + B_o^2)} & \dot{g} &= \frac{1 - B_o^2 + 2BB_o}{1 + B^2} \end{aligned} \quad (4-65)$$

If the hyperbolic anomaly, H , is known, you can use Eq. (4-29), Eq. (4-32), and Eq. (4-37) [see also Battin (1987, 170)] to write

$$\begin{aligned} x &= a(\cosh(H) - e) & y &= -a\sqrt{e^2 - 1} \sinh(H) & \dot{H} &= -\frac{1}{r} \sqrt{\frac{\mu}{-a}} \\ \dot{x} &= \frac{-\sqrt{-\mu a}}{r} \sinh(H) & \dot{y} &= \frac{\sqrt{\mu p}}{r} \cosh(H) \end{aligned}$$

The resulting functions are

$$\begin{aligned} f &= 1 - \frac{a}{r_o} (1 - \cosh(\Delta H)) & g &= (t - t_o) - \sqrt{\frac{(-a)^3}{\mu}} (\sinh(\Delta H) - \Delta H) \\ \dot{f} &= \frac{-\sqrt{-\mu a} \sinh(\Delta H)}{r_o r} & \dot{g} &= 1 - \frac{a}{r} (1 - \cosh(\Delta H)) \end{aligned} \quad (4-66)$$

Finally, if universal variables χ and ψ are known, you have (Bate, Mueller, and White, 1971, 200–201)

$$\begin{aligned}x &= -ae - a \sin\left(\frac{x + c_o}{\sqrt{a}}\right) & y &= a\sqrt{1-e^2} \cos\left(\frac{x + c_o}{\sqrt{a}}\right) \\ \dot{x} &= -\frac{\sqrt{\mu a}}{r} \cos\left(\frac{x + c_o}{\sqrt{a}}\right) & \dot{y} &= -\frac{h}{r} \sin\left(\frac{x + c_o}{\sqrt{a}}\right)\end{aligned}$$

Substituting to determine the f and g functions yields (see Battin, 1987, 179–180)

$$\begin{aligned}f &= 1 - \frac{x_o^2}{r_o} c_2 & g &= (t - t_o) - \frac{x_o^3}{\sqrt{\mu}} c_3 \\ \dot{f} &= \frac{\sqrt{\mu}}{r_o r} x_o (\psi c_3 - 1) & \dot{g} &= 1 - \frac{x_o^2}{r} c_2\end{aligned}\tag{4-67}$$

For a general method of solution, we prefer the universal-variable method because it allows us to solve “standard” orbits, as well as difficult orbits, with one set of equations. We’ll develop and discuss this form of solution shortly.

Series Forms of f and g

The series approach permits us to find expressions for f and g whenever we don’t know all the orbital elements. This is especially useful for some techniques of initial orbit determination in Chap. 6, for which we know only the position magnitudes. Refer to the derivations in either Bate, Mueller and White (1971, 251–258) or Escobal ([1965] 1985, 107–111) for further information. In each case, we attempt to represent the position vector at time $t = \tau$ with a Taylor series about the position vector at $t = 0$.

$$\hat{r}(\tau) = \hat{r}(t) + \dot{\hat{r}}(t) \tau + \frac{\ddot{\hat{r}}(t) \tau^2}{2!} + \frac{\dddot{\hat{r}}(t) \tau^3}{3!} + \dots$$

Notice this requires additional derivatives, evaluated at $t = 0$.

$$\begin{aligned}\dot{\hat{r}} &= \hat{v} \\ \ddot{\hat{r}} &= -\frac{\mu}{r^3} \hat{r} = -u \hat{r} \\ \dddot{\hat{r}} &= -\dot{u} \hat{r} - u \dot{\hat{r}}\end{aligned}$$

We have introduced a new variable u into the relations. Don’t confuse this series coefficient, $u = \frac{\mu}{r^3}$, with the gravitational parameter, μ . Next, group terms of the initial position and velocity vectors after substituting the derivatives into the Taylor series. This yields the

f and g functions we seek. Use the general form given by Escobal ([1965] 1985, 108), dropping the subscripts for clarity:

$$\begin{aligned}
 f = & 1 - \frac{u}{2}\tau^2 - \frac{\dot{u}}{6}\tau^3 - \frac{\ddot{u} - u^2}{24}\tau^4 - \frac{\ddot{u} - 4u\dot{u}}{120}\tau^5 - \frac{\overset{\text{iv}}{u} - 7u\ddot{u} + u^3 - 4\dot{u}^2}{720}\tau^6 \\
 & - \frac{\overset{\text{v}}{u} - 15\dot{u}\ddot{u} - 11u\ddot{u} + 9u^2\dot{u}}{5040}\tau^7 \\
 & - \frac{\overset{\text{vi}}{u} - 15\ddot{u}^2 - 26u\ddot{u} - 16\overset{\text{iv}}{u} + 28u\dot{u}^2 + 22u^2\ddot{u} - u^4}{40,320}\tau^8 \quad (4-68) \\
 g = & \tau - \frac{u}{6}\tau^3 - \frac{\dot{u}}{12}\tau^4 - \frac{3\ddot{u} - u^2}{120}\tau^5 - \frac{2\ddot{u} - 3u\dot{u}}{360}\tau^6 \\
 & - \frac{5\overset{\text{iv}}{u} - 13u\ddot{u} - 10\dot{u}^2 + u^3}{5040}\tau^7 - \frac{6\overset{\text{v}}{u} - 48u\ddot{u} - 24u\dot{u} + 12u^2\dot{u}}{40,320}\tau^8
 \end{aligned}$$

Some programs truncate the above expressions to the first order—only the first two terms—depending on the application. Of course, truncation limits the time interval over which this approach is useful. The final solution uses Eq. (4-57) to find the position and velocity vectors.

Universal Variables

Until now, we've explored separate solutions of Kepler's problem to solve for each type of conic section. The overall importance of Kepler's problem, plus the fact that simulations which consider *all* orbit types often use it, means we'd like *one* algorithm to solve all cases. Although there are separate formulas for solving Kepler's equation, the boundary points make computation difficult. Although this section is rather long, I've included the steps so you'll firmly understand the process.

An example will illustrate the difficulties in using different methods for each type of conic section. Suppose we want to find the time of flight between periapsis and a point at which $\nu = 60^\circ$. If the eccentricity is 0.999 and the semimajor axis is 1500.0 ER, the solution (using the elliptical formula) is

$$E = \cos^{-1} \left(\frac{e + \cos(\nu)}{1 - e \cos(\nu)} \right) = \cos^{-1} \left(\frac{0.999 + \cos 60^\circ}{1 + 0.999 \cos 60^\circ} \right) = 1.479\,658\,4^\circ$$

$$t - t_o = \frac{1}{n} \{ E - e \sin(E) \}$$

$$= \sqrt{\frac{1500^3}{1}} (0.0258\,249\,11 - 0.999 \sin(1.479\,658\,4)) = 1.666\,89 \text{ TU} = 22.414 \text{ min}$$

This answer is actually reasonable because an elliptical orbit with a radius of perigee of 1.5 ER (1500 (1 - 0.999)) is relatively close to the Earth and the satellite will be traveling very fast. Another solution might result from using a parabolic formula. If we use the same radius of perigee (1.5 ER), the semiparameter is 3.0 ER for the parabolic orbit, so

$$B = \tan\left(\frac{\nu}{2}\right) = \tan\left(\frac{60^\circ}{2}\right) = 0.577\ 350\ 27$$

$$\begin{aligned} (t - T) &= \frac{1}{2} \sqrt{\frac{p^3}{\mu}} \left(B + \frac{B^3}{3} \right) = \frac{1}{2} \sqrt{\frac{3^3}{1}} \left(0.577\ 35 + \frac{0.577\ 35^3}{3} \right) \\ &= 1.666\ 67 \text{ TU} = 22.411 \text{ min} \end{aligned}$$

This answer is slightly more accurate and proves to be correct when verified through numerical techniques we'll discuss later.

Notice the sensitivity of this problem to determining the correct solution. The question immediately arises, "Where should the 'cut-off' between orbits be and what errors will result?" Obviously, not using enough significant digits will greatly affect the results. Truncation used to introduce error into the solutions, but modern computers have actually reduced some of these errors through better floating-point operations.

By developing a single set of equations for all the conic sections (a universal-variable formula), we can eliminate these difficulties. The classical formulas for time-of-flight—involved the various anomalies, E , B , and H —don't work very well when the orbits are nearly parabolic. In addition, it's very difficult to solve for the anomalies when a , e , ν , and $t - t_o$ are known. Using the universal variable overcomes these problems as well.

The first equation in Eq. (4-49) relates the change in time to the universal variable χ . Thus, it constitutes the universal variable form of Kepler's equation. The ultimate goal is to find χ when Δt is known. Unfortunately, we can't isolate χ in Eq. (4-49). From Eq. (4-39), we'll use a Newton-Raphson technique in which $f(\chi) = \sqrt{\mu} \Delta t - \int r d\chi$. Notice that $\int r d\chi$ is the right-hand side of Eq. (4-49). Differentiating with respect to χ gives $f'(\chi) = -r$. Then, using Eq. (4-50), we find subsequent values of χ are

$$\chi_{n+1} = \chi_n - \frac{\sqrt{\mu} \Delta t - \chi_n^3 c_3 - \frac{\overset{\curvearrowleft}{r}_o \cdot \overset{\curvearrowleft}{v}_o}{\sqrt{\mu}} \chi_n^2 c_2 - r_o \chi_n (1 - \psi c_3)}{-r}$$

I've replaced $\int r d\chi$ with the right-hand side of Eq. (4-49). As an aside, you can also verify this relation by differentiating Eq. (4-49) with respect to χ and using the derivatives

$$\begin{aligned}\frac{dc_2}{d\chi} &= \frac{1}{2\chi} \{ 1 - \chi c_3 - 2c_2 \} \\ \frac{dc_3}{d\chi} &= \frac{1}{2\chi} \{ c_2 - 3c_3 \}\end{aligned}\quad (4-69)$$

The result is r in Eq. (4-49), as shown in Eq. (4-39). The final update equation is

$$\chi_{n+1} = \chi_n + \frac{\sqrt{\mu} \Delta t - \chi_n^3 c_3 - \frac{\hat{r}_o \cdot \hat{v}_o}{\sqrt{\mu}} \chi_n^2 c_2 - r_o \chi_n (1 - \psi c_3)}{r}$$

Solving this equation enables us to determine Kepler's problem using universal variables. Three issues remain: the behavior of the function, $f(\chi)$, the physical meaning of the various symbols, and the initial guess for χ .

Let's first look at plots of the time, t , versus χ in Fig. 4-15. Although discontinuities appear between hyperbolic, parabolic, and elliptical orbits, we'll always know the orbit type before beginning the iteration. The "smooth" nature of the curves in Fig. 4-15 suggests the Newton-Raphson iteration will perform well on all orbit types.

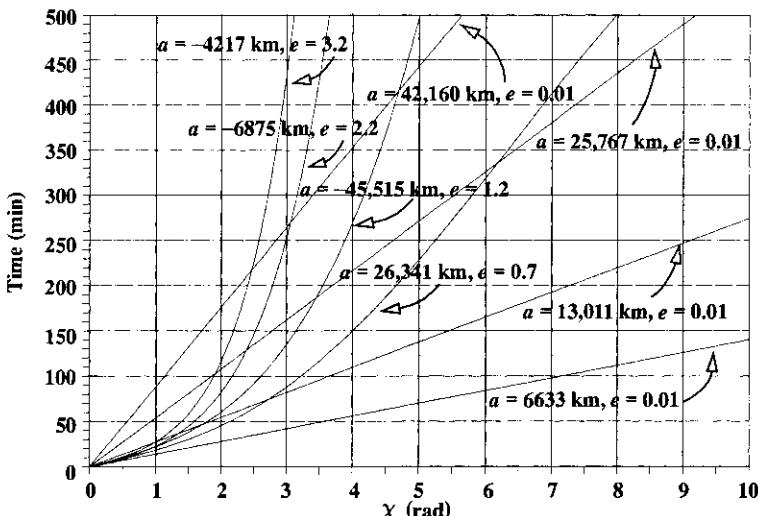


Figure 4-15. Plot of Time versus χ . Several different orbits are shown with their equivalent values of χ . Because the orbit type is always known in advance, the iteration will not have discontinuities. Notice that elliptical orbits ($e > 0.01$) have a slight curve compared to circular orbits, and it becomes more pronounced as the eccentricity increases. This result is analogous to the elliptical results in Figs. 4-6 and 4-7.

Up to this point, these variables have merely been definitions to simplify solving the problem. Following Bate, Mueller, and White (1971, 203–205), compare the expressions for the position with the eccentric anomaly and the universal variable, χ :

$$r = a(1 - e \cos(E)) = a\left(1 + e \sin\left(\frac{\chi + c_o}{\sqrt{a}}\right)\right)$$

$$-e \cos(E) = e \sin\left(\frac{\chi + c_o}{\sqrt{a}}\right)$$

Expand the sine term by recognizing that it's offset only 90° from the cosine:

$$-e \cos(E) = e \sin\left(\frac{\chi + c_o}{\sqrt{a}}\right) = -e \cos\left\{\frac{\pi}{2} + \frac{\chi + c_o}{\sqrt{a}}\right\}$$

Dividing by eccentricity and taking the arc cosine of each side gives you

$$E = \frac{\pi}{2} + \frac{\chi + c_o}{\sqrt{a}}$$

Now find the difference between the eccentric anomaly at the current time and the epoch time, where $c_o = 0$:

$$E - E_o = \frac{\pi}{2} + \frac{\chi + c_o}{\sqrt{a}} - \left\{\frac{\pi}{2} + \frac{c_o}{\sqrt{a}}\right\} = \frac{\chi}{\sqrt{a}}$$

Bate, Mueller, and White (1971, 204) show the relation for hyperbolic orbits: $\chi/\sqrt{-a} = \Delta H$.

To determine the value of χ for a parabolic orbit, begin with Eq. (4-49):

$$r = \chi^2 c_2 + \frac{\vec{r}_o \cdot \vec{v}_o}{\sqrt{\mu}} \chi (1 - \psi c_3) + r_o (1 - \psi c_2)$$

Recall the parabola has special parameters: $a = \infty$, so $\psi = 0$, $c_2 = \frac{1}{2}$, and $c_3 = 1/6$. Thus,

$$r = \frac{\chi^2}{2} + \frac{\vec{r}_o \cdot \vec{v}_o}{\sqrt{\mu}} \chi + r_o$$

Use interim results from the **RANDV** derivation and the trajectory equation to find an expression for the dot product:

$$\vec{r}_o \cdot \vec{v}_o = r \dot{r} = \frac{p}{1 + \cos(\nu)} \sqrt{\frac{\mu}{p}} \sin(\nu) = \sqrt{\mu p} \tan\left(\frac{\nu}{2}\right) = \sqrt{\mu p} B$$

Now, recall Eq. (4-24) and substitute into the universal-variable relation for the position above, being careful in identifying initial and final time values:

$$\frac{p + pB^2}{2} = \frac{\chi^2}{2} + \frac{\sqrt{\mu p} B_o}{\sqrt{\mu}} \chi + \frac{p + pB_o^2}{2}$$

$$\frac{\chi^2}{2} + \sqrt{p} B_o \chi + \frac{p + pB_o^2}{2} - \frac{p + pB^2}{2} = 0$$

Because this equation is quadratic in χ , solve and find

$$\chi = \frac{-\sqrt{p} B_o \pm \sqrt{p B_o^2 - 2 \left\{ \frac{p + pB_o^2}{2} - \frac{p + pB^2}{2} \right\}}}{2 \left(\frac{1}{2} \right)}$$

$$\chi = -\sqrt{p} B_o \pm \sqrt{p B_o^2 - p - pB_o^2 + p + pB^2} = -\sqrt{p} B_o \pm \sqrt{p B^2}$$

and take the positive root to get

$$\chi = -\sqrt{p} B_o + \sqrt{p} B = \sqrt{p} (B - B_o)$$

As expected for a universal formulation, all three relations are very similar.

Recall that when we began deriving this method, we defined the derivative of the universal variable ($\dot{\chi} = \sqrt{\mu}/r$). Now, we still have a combination of energy and orbit type included in χ . This combination of parameters means the universal-variable method will work for different orbit types.

Selecting an initial guess for χ is very important, because its accuracy determines the number of iterations. Thus, we'll talk a bit about identifying the proper choices. Also note that we can solve the routine for times-of-flight that exceed one orbital period by simply reducing the time to be within one period. This is possible because we consider only two-body motion, and the orbit will retrace itself over time. For real-world analyses including perturbation effects, this approach *doesn't* work.

We've shown that the universal variable is related to the change in eccentric anomaly for elliptical orbits ($\chi = \sqrt{a} \Delta E$). During one period, the satellite completes 360° of change in eccentric anomaly ($\chi = \sqrt{a} 2\pi$). This permits us to set up a ratio using the period, Eq. (2-16):

$$\frac{\chi}{2\pi\sqrt{a}} = \frac{\Delta t}{P}$$

As an aside, this relation becomes an equality for circular orbits. We can see this in Fig. 4-15 and the linear relation for circular orbits. As the eccentricity increases, the curves slowly depart from a linear relation. Using the expression for the period and letting ($\alpha = 1/a$) result in

$$\chi_o \approx \sqrt{\mu} (\Delta t) \alpha \quad (4-70)$$

Check to see if $\alpha = 1.0$ because these cases will yield a first guess that causes numerical instability. (It's too close to the correct answer.) This case represents a satellite with a semimajor axis equal to the radius of the Earth. Although this isn't very likely, you may encounter it during validation checks.

For the initial guess for a parabolic orbit, use $\chi/\sqrt{p} = \Delta B$. Unfortunately, we only have v_o , so the estimate is

$$\chi_o \approx \sqrt{p} \tan\left(\frac{v_o}{2}\right) \quad (4-71)$$

Selecting the hyperbolic initial guess is a little more complicated, but Bate, Mueller, and White (1971, 206–207) provide the template. By Eq. (4-47), if the change in hyperbolic anomaly is large, ψ will be a large negative number. When ψ is negative, evaluate the c_2 function using the hyperbolic cosine expression. You may also equate $\cosh \sqrt{-\psi}$ with its exponential identity, and if $\sqrt{-\psi}$ is a large positive number, $e^{\sqrt{-\psi}}$ will be large compared to one. So,

$$c_2 = \frac{1 - \frac{e^{\sqrt{-\psi}} + e^{-\sqrt{-\psi}}}{2}}{\psi} \approx \frac{-e^{\sqrt{-\psi}}}{2\psi} = \frac{-ae^{\sqrt{-\psi}}}{2\chi^2}$$

Likewise, because $\sqrt{-\psi^3} = \pm \frac{\chi^3}{a\sqrt{-a}}$, c_3 becomes

$$c_3 = \frac{\frac{e^{\sqrt{-\psi}} - e^{-\sqrt{-\psi}}}{2} - \sqrt{-\psi}}{\sqrt{-\psi^3}} \approx \frac{e^{\sqrt{-\psi}}}{2\sqrt{-\psi^3}} = \frac{a\sqrt{-a}e^{\sqrt{-\psi}}}{\pm 2\chi^3}$$

You can immediately determine the signs because, from Fig. 4-5, the c_3 function is positive for all values of ψ . Therefore, if χ is positive, $(t - t_o)$ is positive, and you should use the positive sign—and vice versa. In this case,

$$c_3 \approx \text{SIGN}(t - t_o) \frac{-a\sqrt{-a}e^{\sqrt{-\psi}}}{2\chi^3}$$

Next, substitute these approximate values into the time-of-flight equation, Eq. (4-49), and neglect the last term, $r_o \chi / \sqrt{\mu}$, because it's small compared to terms with $e^{\sqrt{-\psi}}$. You'll now have

$$t - t_o \approx -\frac{\vec{r}_o \cdot \vec{v}_o}{2\mu} a e^{\sqrt{-\psi}} - \text{SIGN}(t - t_o) \left\{ 1 - \frac{r_o}{a} \right\} \frac{a\sqrt{-a}}{2\sqrt{\mu}} e^{\sqrt{-\psi}}$$

Solving for $e^{\sqrt{-\psi}}$ results in

$$e^{\sqrt{-\psi}} \approx \frac{-2\mu(t - t_o)}{a \vec{r}_o \cdot \vec{v}_o + \text{SIGN}(t - t_o) \sqrt{-\mu a} \left(1 - \frac{r_o}{a} \right)}$$

Recognizing that $\sqrt{-\psi} = \pm \chi / \sqrt{-a}$, solve for χ_o and resolve the sign by seeing that χ will be positive when $(t - t_o)$ is positive. Thus,

$$\chi_o \approx \text{SIGN}(t - t_o) \sqrt{-a} \ln \left\{ \frac{-2\mu\alpha(t - t_o)}{\vec{r}_o \cdot \vec{v}_o + \text{SIGN}(t - t_o) \sqrt{-\mu a} (1 - r_o\alpha)} \right\} \quad (4-72)$$

Implementing Kepler's Problem

Algorithm 31, below, uses universal variables and *one formulation* for all orbit types—an overwhelming advantage over using classical orbital elements, despite having to use many more operations. The iteration for this problem may take various forms; however, I'll present the Newton-Raphson iteration because it is very robust for this application.

Using canonical units allows us to drop many mathematical operations involving the gravitational parameter. This speeds up the algorithm while also enabling the code to remain the same, even when new physical constants are introduced. Speed is important for this routine because a program may use it many times. For instance, a typical groundtrack program may call this routine more than 1400 times to calculate one track! Including the extra operations for calculating the gravitational parameter will slow it down a lot. Estimation techniques and processing of observations also require this speed. We make computations more efficient by storing several temporary variables, the dot product of the initial position and velocity vectors, and $\chi_o(1 - \psi c_3)$, χ^2 , and χ^3 . By using α , we can avoid having to calculate the eccentricity to determine the type of orbit in the initial guess. Finally, for parabolic orbits, the initial guess requires we find the semiparameter as h^2/μ .

ALGORITHM 31: *Kepler* ($\vec{r}_o, \vec{v}_o, \Delta t \Rightarrow \vec{r}, \vec{v}$)

$$\xi = \frac{v_o^2}{2} - \frac{\mu}{r_o}$$

$$a = -\frac{\mu}{2\xi} \quad \alpha = \frac{1}{a} = \frac{-v_o^2}{\mu} + \frac{2}{r_o}$$

Circle or Ellipse ($\alpha > 0.000\ 001$): $\chi_o \approx \sqrt{\mu}(\Delta t)\alpha$

Check if $\alpha = 1.0$; if it does, the first guess will be too close to converge.

Parabola ($|\alpha| < 0.000\ 001$):

$$\begin{cases} \hat{h} = \vec{r}_o \times \vec{v}_o \\ p = \frac{h^2}{\mu} \\ \cot(2s) = 3 \sqrt{\frac{\mu}{p^3}} (\Delta t) \\ \tan^3(w) = \tan(s) \end{cases} \quad \chi_o \approx \sqrt{p} 2 \cot(2w)$$

Hyperbola ($\alpha < -0.000\ 001$):

$$\chi_o \approx \text{SIGN}(\Delta t) \sqrt{-a} \ln \left\{ \frac{-2\mu\alpha(\Delta t)}{\vec{r}_o \cdot \vec{v}_o + \text{SIGN}(\Delta t) \sqrt{-\mu a} (1 - r_o \alpha)} \right\}$$

LOOP

$$\psi = \chi_n^2 \alpha$$

Find $c_2 c_3$ ($\psi \Rightarrow c_2, c_3$)

$$r = \chi_n^2 c_2 + \frac{\vec{r}_o \cdot \vec{v}_o}{\sqrt{\mu}} \chi_n (1 - \psi c_3) + r_o (1 - \psi c_2)$$

$$\chi_{n+1} = \chi_n + \frac{\sqrt{\mu} \Delta t - \chi_n^3 c_3 - \frac{\vec{r}_o \cdot \vec{v}_o}{\sqrt{\mu}} \chi_n^2 c_2 - r_o \chi_n (1 - \psi c_3)}{r}$$

$$\chi_n \Leftarrow \chi_{n+1}$$

UNTIL $|\chi_n - \chi_{n-1}| < 1 \times 10^{-6}$

$$f = 1 - \frac{\chi_n^2}{r_o} c_2 \quad g = \Delta t - \frac{\chi_n^3}{\sqrt{\mu}} c_3$$

$$\dot{g} = 1 - \frac{\chi_n^2}{r} c_2 \quad \dot{f} = \Delta t - \frac{\sqrt{\mu}}{r r_o} \chi_n (\psi c_3 - 1)$$

$$\dot{\vec{r}} = f \vec{r}_o + g \vec{v}_o$$

$$\dot{\vec{v}} = \dot{f} \vec{r}_o + \dot{g} \vec{v}_o$$

Check $f \dot{g} - \dot{f} g = 1$ for success.

An example will demonstrate Kepler's problem.

▼ **Example 4-4. Solving Kepler's Problem.**

GIVEN: $\dot{\vec{r}}_{IJK} = 1131.340 \hat{i} - 2282.343 \hat{j} + 6672.423 \hat{k}$ km

$$\dot{\vec{v}}_{IJK} = -5.643\ 05 \hat{i} + 4.303\ 33 \hat{j} + 2.428\ 79 \hat{k}$$
 km/s

$$\Delta t = 40 \text{ min} = 2.974\ 674 \text{ TU}$$

FIND: Position and velocity vectors at future time

First, convert units to find

$$\dot{\vec{r}}_{IJK} = 0.177\ 38 \hat{i} - 0.357\ 84 \hat{j} + 1.046\ 14 \hat{k}$$
 ER

$$\dot{\vec{v}}_{IJK} = -0.713\ 83 \hat{i} + 0.544\ 36 \hat{j} + 0.307\ 23 \hat{k}$$
 ER/TU

Use part of Algorithm 5 to find a .

$$\xi = \frac{v_o^2}{2} - \frac{\mu}{r_o} = -0.442\ 89 \text{ ER}^2/\text{TU}^2$$

$$a = \frac{\mu}{2\xi} = 1.128\ 94 \text{ ER} \quad \alpha = \frac{1}{a} = \frac{-v_o^2}{\mu} + \frac{2}{r_o} = 0.885\ 78 \text{ ER}$$

The orbit is elliptical, ($\alpha > 0.000\ 001$), so approximate the first guess as

$$\chi_o \approx \sqrt{\mu} (\Delta t) \alpha = 2.634\ 916\ 7$$

Iterate to get the values shown in Table 4-4:

TABLE 4-4. Values for Kepler's Problem in Example 4-4. Iteration values are shown. Notice how quickly we arrive at the correct answer.

Iteration	χ_n	ψ	r	χ_{n+1}
1	2.634 916	6.149 802	1.36164	2.640 192
2	2.640 192	6.174 452	1.36192	2.640 192

$$f = 1 - \frac{\chi_o^2}{r_o} c_2 = -0.806\ 637 \quad g = \Delta t - \frac{\chi_o^3}{\sqrt{\mu}} c_3 = 0.726\ 406\ 9$$

$$\dot{g} = 1 - \frac{\chi_o^2}{r} c_2 = -0.509\ 878 \quad \dot{f} = \Delta t - \frac{\sqrt{\mu}}{r r_o} \chi_o (\psi c_3 - 1) = -0.780\ 550$$

$$\dot{\vec{r}}_{IJK} = f \dot{\vec{r}}_o + g \dot{\vec{v}}_o = -0.661\ 612\ 5 \hat{i} + 0.684\ 073\ 9 \hat{j} - 0.620\ 680\ 9 \hat{k}$$
 ER

$$\dot{\vec{r}}_{IJK} = -4219.853 \hat{i} + 4363.116 \hat{j} - 3958.789 \hat{k}$$
 km

$$\dot{\vec{v}}_{IJK} = \dot{f} \dot{\vec{r}}_o + \dot{g} \dot{\vec{v}}_o = 0.466\ 738\ 0 \hat{i} - 0.242\ 445\ 5 \hat{j} - 0.773\ 212\ 6 \hat{k}$$
 ER/TU

$$\dot{\vec{v}}_{IJK} = 3.689\ 736 \hat{i} - 1.916\ 620 \hat{j} - 6.112\ 528 \hat{k}$$
 km/s



4.4 Application: Groundtracks

To determine groundtracks for a satellite's orbit, we combine both of Kepler's routines previously described in this chapter with the conversion of a position vector to its sub-satellite point. This particular combination yields a powerful tool for determining orbit position and location relative to a ground site. **Groundtracks** are the locus of points formed by the points on the Earth directly below a satellite as it travels through its orbit. They represent a way to quickly display a satellite's position over a two- or three-dimensional image. By "seeing" *where* a satellite is at any time, we can obtain useful information. For instance, by displaying more than one satellite at once, we can quickly determine whether they're close to each other. We'll discuss this problem of close approach later.

It's easiest to visualize groundtracks over a three-dimensional Earth, because some of the concepts are difficult to grasp on a flat, two-dimensional map. To view a satellite orbit in space in three dimensions, we can fix the satellite's orbit in inertial space or we can show the satellite's position over the Earth as it would appear over a rotating Earth. Although the first viewing method is most common, it doesn't represent the satellite's *true* ground path as the second method does. Consider Fig. 4-16.

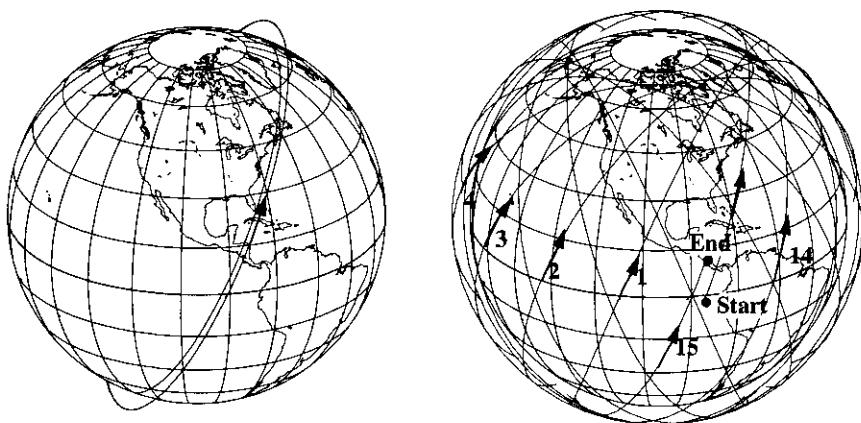


Figure 4-16. Three-D Groundtracks. With the Earth fixed, the satellite's orbital plane is visible, but this figure doesn't represent the satellite as it travels through time. In reality, the Earth moves under the "fixed" orbital plane. When a groundtrack is displayed over the positions it would occupy over time, the result is a "ball-of-yarn." Although this view *does* show the true positions, it's not particularly useful because it contains so much information. I've numbered each revolution for clarity.

Notice that the orbital plane and groundtrack are immediately visible. If we now examine a groundtrack for the same satellite using the second method for three-dimensional groundtracks, we observe a very different pattern. Although this figure isn't particularly

useful and illustrative, it does represent the satellite's *actual path* if we observe it from an imaginary stationary spot above the Earth.

Two-dimensional projections can illustrate the same concepts. Just as with the three-dimensional case, the projection on a fixed Earth suggests that the satellite always remains over one locus of points above the Earth. But this isn't true, so the rotating version of this projection is much more accurate. Figure 4-17 shows the projection on a rotating Earth.

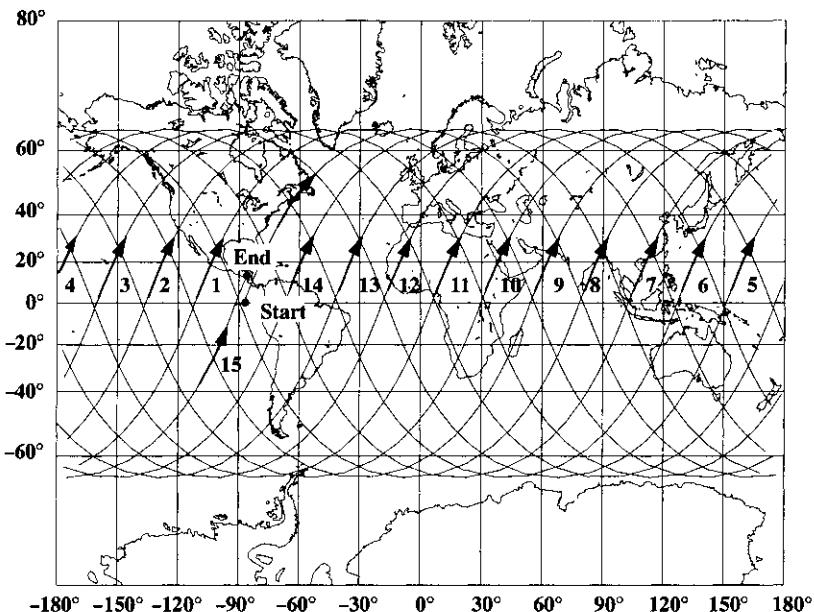


Figure 4-17. Groundtrack of a Satellite in Circular Orbit on a Rotating Earth. This form of the two-dimensional groundtrack is very useful because it shows the actual locations over a rotating Earth. Also notice the mismatch in the starting and ending points after 15 revolutions.

Notice that the groundtrack doesn't *repeat*; that is, the starting and ending points don't coincide. We can adjust the orbital parameters to produce a *repeating groundtrack*, but we won't discuss this until Chap. 10. There, the difference between these two representations will be vitally important. Common terms for these tracks are *passes* and *revolutions (revs)*. A **revolution** is completed each time the satellite makes one trip around the Earth. This definition is somewhat vague without a reference point to determine when the "trip" is complete. We sometimes use perigee, or the ascending node. We'll discuss additional details in Sec. 8.6.1. The satellite in Fig. 4-17 completes about 15 revolutions per day. A **pass** occurs whenever the satellite passes over the same area on the Earth again; in the figure, the end point would be considered *pass 1*.

With the basics of projections identified, we can now explore characteristics of specific groundtracks for certain orbital elements (a, e, i, ω). I won't combine all of these elements because the permutations grow very quickly. The intent is to illustrate some fundamental trends which are helpful in mission analysis.

Figure 4-18 shows some of the groundtrack shapes for satellites in circular orbits. Notice, as the orbital altitude increases (semimajor axis), the satellite travels over less ground. Also, the symmetry about lines of longitude indicates the circular nature of the orbits, and we call this *line symmetry*. Circular orbits also have *hinge symmetry*, which exists for orbits that have equal and identically shaped regions above and below the equator. For some applications, it may be useful to illustrate the same groundtracks on a rotating three-dimensional figure similar to Fig. 4-16.

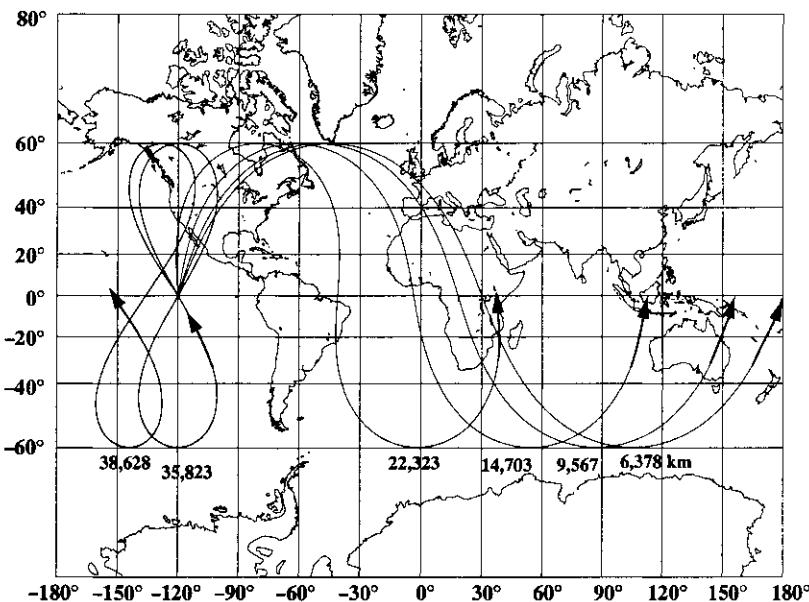


Figure 4-18. Effects of Increasing Altitude for Groundtracks on a Rotating Earth. Notice how the span of the groundtrack shrinks as the altitude increases. All groundtracks shown are for circular orbits. Altitudes are given in kilometers above the Earth's surface.

We can also determine an approximate value for the semimajor axis from the groundtrack. We first find the orbital period and then a , using Eq. (2-16). The change in longitude, $\Delta\lambda$ measured from the initial starting location, in one revolution gives us the satellite's period,

$$\rho = \frac{360^\circ - \Delta\lambda}{15^\circ/\text{sidereal hr}}$$

We must switch the order of the numerator for retrograde orbits.

Looking at changes in eccentricity and argument of perigee, consider Fig. 4-19. Notice that, as the eccentricity increases, the groundtrack becomes more distorted. We can always distinguish circular orbits from elliptical orbits because *all* circular orbits have hinge and line symmetry, whereas elliptical orbits do not. The groundtrack's shape determines the approximate location of perigee. When the satellite is closest to the Earth, it travels across more territory. Notice the $e = 0.6$ case in Fig. 4-19, in which the argument of perigee is 90° (left-hand side). The large, broad, curving region heading to the east shows the area where the satellite is traveling the fastest (covering large longitudinal regions) and hence, is closest to perigee. Remember, we can't determine the exact value for perigee, but we can approximate its location. For the same ($e = 0.6$) case, notice how at apogee the satellite is traveling about the same speed as the Earth and almost begins to go backward. This motion is characterized by the nearly vertical line in the groundtrack. In fact, as orbital altitude increases, the region where the groundtrack turns and begins to travel north again will tighten into a loop. At much higher altitudes, the groundtrack will travel west, indicating that the Earth is traveling "faster" than the satellite.

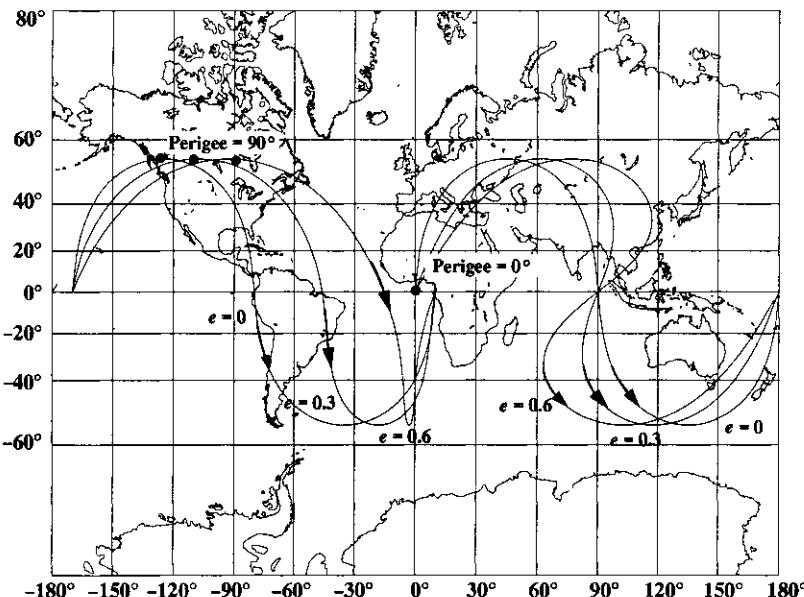


Figure 4-19. Effects of Eccentricity and Perigee Location for Groundtracks on a Rotating Earth. This figure shows two sets of orbits with different perigees. Each set contains three orbits consisting of eccentricity values of 0.0, 0.3, and 0.6. Notice how distorted each groundtrack becomes with increased eccentricity. Perigee locations at $\pm 180^\circ$ result in equal but opposite images.

Changes in inclination aren't as dramatic but do produce some interesting results, as Fig. 4-20 shows. We always determine the inclination based on the maximum latitude achieved during a revolution. Retrograde orbits are determined by subtracting the maximum latitude from 180° .

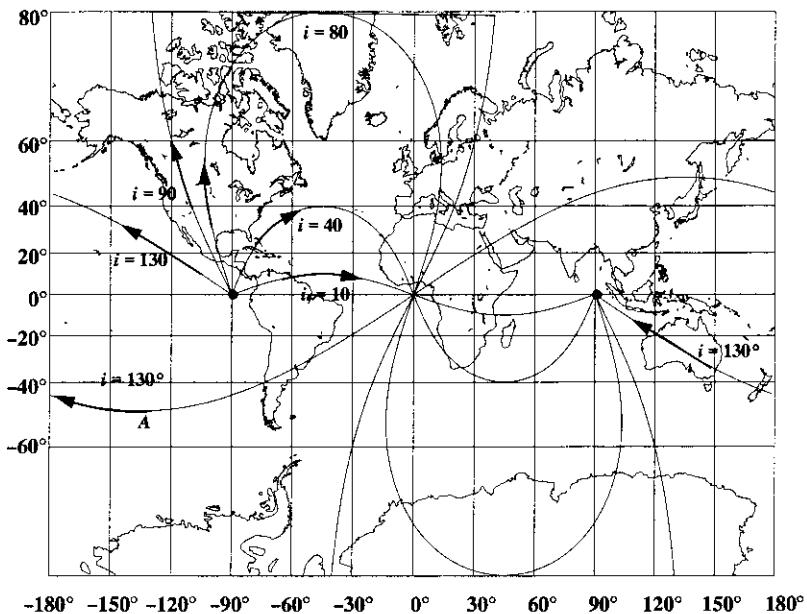


Figure 4-20. Effects of Inclination for Groundtracks on a Rotating Earth. We determine inclination by the maximum latitude on the groundtrack. Notice how the retrograde orbits ($i > 90^\circ$) tend to move to the left. Don't confuse this motion with very high, direct orbits (as in Fig. 4-18 altitude = 38,628 km). The distinction lies in the shape of the corners. Retrograde orbits always have smooth transitions (point A in the figure), whereas very high, direct orbits will have sharp corners.

4.5 Application: Find Time of Flight (*FindTOF*)

While analyzing satellite orbits, we often need to determine the amount of time between two locations on an orbit. Although this problem is related to Kepler's problem, it's slightly different because we already know the two position vectors. At first glance, this problem may even resemble Lambert's problem (Chap. 6); however, because the true anomaly is known, several differences arise. We can solve the time-of-flight using the classical formulations of Kepler's problem (See *Classical Formulas Using f and g Functions*, page 249). Eq. (4-63) to Eq. (4-67) will ultimately be part of any solution. Notice that only Eq. (4-63) does *not* include time. Thus, if we know the change in eccentric, para-

bolic, hyperbolic, or universal variable, we have an immediate solution. This section will focus on the specific case for which we know only the two position vectors, and we must find a and e such that Kepler's equation yields $\Delta t = t - t_0$.

To find a solution, start with Eq. (4-63) and Eq. (4-64) and equate the f functions to get

$$1 - \left(\frac{r}{p} \right) (1 - \cos(\Delta\nu)) = 1 - \frac{a}{r_o} (1 - \cos(\Delta E))$$

Solving for the semimajor axis results in

$$a = \frac{rr_o}{p} \frac{(1 - \cos(\Delta\nu))}{(1 - \cos(\Delta E))} \quad (4-73)$$

Unfortunately, this expression includes additional unknowns—the semiparameter and the change in eccentric anomaly. An expression for the semiparameter is also helpful in the process. Begin with the f functions of Eq. (4-63) and Eq. (4-64):

$$\sqrt{\frac{\mu}{p}} \tan\left(\frac{\Delta\nu}{2}\right) \left(\frac{1 - \cos(\Delta\nu)}{p} - \frac{1}{r} - \frac{1}{r_o} \right) = -\frac{\sqrt{\mu a} \sin(\Delta E)}{r_o r}$$

Substitute Eq. (4-73), simplify, and rearrange to get

$$\frac{1 - \cos(\Delta\nu)}{p} - \frac{r_o + r}{r_o r} = \frac{-\sqrt{a} \sin(\Delta E)}{r_o r} \left(\frac{\sqrt{p} \sin(\Delta\nu)}{1 - \cos(\Delta\nu)} \right)$$

Obtaining common denominators results in

$$\frac{1 - \cos(\Delta\nu)}{p} - \frac{r_o + r}{r_o r} = -\frac{1}{\sqrt{r_o r}} \left(\frac{\sin(\Delta\nu)}{\sqrt{1 - \cos(\Delta\nu)}} \right) \left(\frac{\sin(\Delta E)}{\sqrt{1 - \cos(\Delta E)}} \right)$$

Recognize that you can simplify each parenthetical term by applying appropriate half-angle formulas [Eq. (C-12)]:

$$\frac{1 - \cos(\Delta\nu)}{p} = -\frac{1}{\sqrt{r_o r}} \left\{ \sqrt{2} \cos\left(\frac{\Delta\nu}{2}\right) \sqrt{2} \cos\left(\frac{\Delta E}{2}\right) \right\} + \frac{r_o + r}{r_o r}$$

Finally, find the semiparameter,

$$p = \frac{\{1 - \cos(\Delta\nu)\} r_o r}{r_o + r - 2\sqrt{r_o r} \cos\left(\frac{\Delta\nu}{2}\right) \cos\left(\frac{\Delta E}{2}\right)} \quad (4-74)$$

We now have several options. Fundamentally, if we know the semimajor axis or semiparameter, we can use Eq. (4-63) to determine the f and g functions. Then, we can use

these relations to find the change in anomaly, and finally, the time through the g equations. For this particular problem, we've defined the two position vectors to be on the same orbit; thus, the semiparameter and semimajor axis will be (or should be) known. The parabolic case is difficult because the change in parabolic anomaly is ill-defined. For this reason, I have chosen the formulation of Thorne and Bain (1995), which uses the chord and semi-perimeter (Sec. 6.7.1) to solve the parabolic case. Notice that we only need the two position magnitudes and the change in true anomaly. The complete solution is in Algorithm 32, below:

ALGORITHM 32: FINDTOF ($\vec{r}_o, \vec{r}, p \Rightarrow TOF$)

$$\cos(\Delta\nu) = \frac{\vec{r}_o \cdot \vec{r}}{r_o r}$$

$$k = r_o r (1 - \cos(\Delta\nu))$$

$$l = r_o + r$$

$$m = r_o + r (1 + \cos(\Delta\nu))$$

$$a = \frac{mkp}{(2m-l^2)p^2 + 2klp - k^2}$$

$$f = 1 - \frac{r}{p} (1 - \cos(\Delta\nu))$$

$$g = \frac{r_o r \sin(\Delta\nu)}{\sqrt{\mu p}}$$

IF Elliptical Orbit ($a > 0.0$):

$$\dot{f} = \sqrt{\frac{\mu}{p}} \tan\left(\frac{\Delta\nu}{2}\right) \left\{ \frac{1 - \cos(\Delta\nu)}{p} - \frac{1}{r_o} - \frac{1}{r} \right\}$$

$$\cos(\Delta E) = 1 - \frac{r_o}{a} (1 - f) \quad \sin(\Delta E) = \frac{-r_o r \dot{f}}{\sqrt{\mu a}}$$

$$TOF = g + \sqrt{\frac{a^3}{\mu}} (\Delta E - \sin(\Delta E))$$

IF Parabolic Orbit ($a = \infty$):

$$c = \sqrt{r_o^2 + r^2 - 2r_o r \cos(\Delta\nu)}$$

$$s = \frac{r_o + r + c}{2}$$

$$TOF = \frac{2}{3} \sqrt{\frac{s^3}{2\mu}} \left\{ 1 - \left(\frac{s-c}{s} \right)^{3/2} \right\}$$

IF Hyperbolic Orbit ($a < 0.0$):

$$\cosh(\Delta H) = a - \frac{r_o}{a}(1-f)$$

$$TOF = g + \sqrt{\frac{(-a)^3}{\mu}} (\sinh(\Delta H) - \Delta H)$$

Problems

- Calculate the eccentric anomaly, true anomaly, and universal variable for $M = 235.4^\circ$ and $e = 0.4$. Hint: for χ use $a = 8000$ km.
- A rogue asteroid has just been discovered. Your job is to determine how much time we have to intercept and destroy the object, if $a = -2,797.425\,069$ km, $e = 2.8$, $i = 23^\circ$, $\nu = 249.27^\circ$, and $r = 2,105,124.388$ km.
- Can Algorithm 31 handle multiple-revolution transfers? If so, how; and if not, why not?
- You've just completed an analysis of where the Space Shuttle must be when it performs a critical maneuver. You know the Shuttle is in a circular orbit and has a position vector of

$$\hat{r}_o = 6275.396 \hat{i} + 2007.268 \hat{j} + 1089.857 \hat{k} \text{ km}$$

In 55 minutes, you predict the orbital parameters are

$$a = 1.0470357 \text{ ER}, e = 0.000096, i = 28.5^\circ, M = 278.946\,88^\circ$$

Is your analysis correct?

- A certain groundtrack is noticed with a regular 25° westward displacement on each pass. If the orbit is circular, what is the semimajor axis? How can you be sure?
- You have received two position vectors for a satellite. How much time has elapsed between these two positions? ($p = 6681.571$ km)

$$\hat{r}_o = -2574.9533 \hat{i} + 4267.0671 \hat{j} + 4431.5026 \hat{k} \text{ km}$$

$$\hat{r} = 2700.6738 \hat{i} - 4303.5378 \hat{j} - 4358.2499 \hat{k} \text{ km}$$

7. You've just purchased a new telescope and wish to observe a satellite. First, you must align the telescope with a known star, so you decide to plot the path of the satellite against the star field. Explain in detail how you could do this. Be sure to identify algorithms you need to convert and process the data.
8. We know that propagation techniques must work forward and backward in time. If the change in true anomaly is negative, do the f and g functions in Eq. (4-63) change?

CHAPTER 5 ORBITAL MANEUVERING

- 5.1 Historical Background
- 5.2 Introduction
- 5.3 Coplanar Maneuvers
- 5.4 Noncoplanar Transfers
- 5.5 Combined Maneuvers
- 5.6 Circular Rendezvous
- 5.7 Continuous-Thrust Transfers
- 5.8 Relative Motion

5.1 Historical Background

So far, we've examined the theoretical aspects of determining the orbits of satellites. Bate, Mueller, and White (1971, 151) suggest that Isaac Newton (1642–1727) mentioned orbital maneuvers and satellites in his *Principia* (1687). But it's highly unlikely that he actually envisioned artificial satellites because the ability to launch a satellite, much less to do any maneuver, was several hundred years away. He was more likely referring to the many different orbits which a heavenly body could occupy; this variety is the basis for orbital maneuvers. Artificial satellites have presented some unique challenges and opened up a new chapter in astrodynamics. When we place satellites in orbit, we want a specific type of orbit, which may or may not be achievable directly from launch, given a particular launch site or rocket booster. **Orbital maneuvering** encompasses all orbital changes after insertion required to place a satellite in the orbit we choose.

The early history of maneuvers is sometimes centered on the first experiments with rockets. Konstantin Tsiolkovsky (1857–1935), Herman Oberth (1894–1989), Robert Goddard (1882–1945), Wernher von Braun (1912–1977), and others are part of this history, but they worked with rockets, not satellites, which are the focus of this book. Moreover, actual satellite maneuvering didn't occur until *after* the launch of Sputnik on October 4, 1957! Thus, orbital maneuvering had its roots in the classical formulas and dynamics of astrodynamics from several centuries ago but didn't mature until recently.

The first satellite maneuvers most likely occurred on January 2, 1959, with the Soviet Luna 1 mission to the Moon. Although this was an early attempt to gain "one-upsman-ship" during the Cold War, the feat is still remarkable. Satellites had become a reality only a year or so earlier. Luna 1 missed the Moon by about 6000 km, but coming even this

close required countless maneuvers, including circularizing the initial launch orbit, doing midcourse corrections, and so on. Later, on September 12, 1959, the Soviet Union corrected some of their difficulties and crashed Luna 2 into the lunar surface, near the crater Archimedes. The crash was intentional because retro-rocket technology for a soft landing hadn't matured enough to be installed on a satellite.

After some initial failures at landing on the Moon, the focus of scientists shifted to the Earth's immediate environment. The civilian sector became interested in space very quickly because many missions were uniquely suited to space: weather forecasting, communications, and so on. Maneuvers were important to place each of these satellites, designed for a specific mission and orbit, into the correct orbit.

On March 18, 1965, Alexei A. Leonov (1934–) became the first human to walk in space outside his Voskhod 2 spacecraft. This accomplishment opened up another aspect of maneuvering: relative motion about a primary spacecraft. Although the mathematical basis had been discussed and developed several years earlier, doing an actual spacewalk meant carrying out the theories.

Virgil I. Grissom (1926–1967) and John W. Young (1930–) participated in NASA's first two-man mission. It included the first actual maneuver for a manned spacecraft. On March 23, 1965, their Gemini 3 spacecraft practiced the delicate art of maneuvering to learn about the process. This test was necessary to lay the groundwork for the first rendezvous of two spacecraft—the Gemini 6 (December 15–16, 1965) and the Gemini 7 (December 4–18, 1965). The ability to rendezvous was, and still is, very important, especially considering space-station activities. The increasing frequency of events in worldwide space programs raised the question of a mishap in space, which presented some new technical challenges. How could a spacecraft rendezvous with a damaged spacecraft to rescue the astronauts? Would we rescue astronauts from another country? The initial answers were yes, so a test of rendezvous techniques was in order. Yet a joint test didn't occur until the Apollo-Soyuz test program in July 1975 and again when the Space Shuttle rendezvoused with the space station MIR in 1995. It's sad to reflect that so much time elapsed before nations cooperated, and that the events were so short-lived.

A final note is important. Historically, maneuvers have been treated as instantaneous changes in velocity, or impulsive operations. However, we must calculate *actual* maneuvers by numerical integration because the real world is complex. Burns take place over the finite amount of time required for the maneuver. I won't treat the thrust characteristics and burn duration here, but we would need to do so because real maneuvers aren't impulsive and may last several minutes. I will discuss continuous thrust in general to provide insight into realistic situations.

This chapter describes methods used to establish Earth satellites in both low- and high-altitude orbits and the techniques to maneuver from one orbit to another. I'll also discuss relative motion and rendezvous because they're closely related to maneuvering. The tremendous developments in launch vehicles and the explosion of space applications underscore the importance of this area. Each new mission which attempts to use space for operations requires effective maneuvering to ensure a successful mission.

5.2 Introduction

Two-body analysis of orbital changes usually involves solving the Lambert problem, which examines transfer between position vectors (Sec. 6.7). We'll do a simpler approximation to these more sophisticated routines—coplanar, noncoplanar, and fixed Δv maneuvers. These maneuvers allow us to change various orbital elements. A force applied *in the orbit plane* can change eccentricity, e , semimajor axis, a , and argument of periapsis, ω . In general, a force applied normal to the plane can change inclination, i , and longitude of ascending node, Ω . For all cases, we assume an instantaneous burn. Note that we can also examine the effects of changes on individual orbital elements using equations for the Gaussian variation of parameters, which I'll introduce in Chap. 8. This technique is especially useful to examine the sensitivity of the orbital changes in radial, transverse, and normal directions (RSW), and also to estimate the efficiency of the burns.

A goal of any maneuver, particularly those in which vehicle weight is critical, is efficiency. Efficiency can mean using a minimum amount of fuel (Δv), but if such a maneuver takes too much time, increased requirements for support systems may more than offset the fuel savings. Consequently, the desirable maneuver is the one that accomplishes the mission while requiring as little added fuel (weight) as possible.

Before discussing the required mechanics, let's develop a standardized notation. For clarity, I've identified the positions $r_{initial}$, r_{final} and velocities $v_{initial}$, v_{final} with an orbit (initial, final, and trans) *and* the location in the orbit (a , b , ...). Intermediate orbits have the notation $trans_i$, where i will range from 1 to the number of intermediate orbits. Finally, the change in velocities we must apply are lettered as a , b , c , ... to match the positions in the orbit. Figure 5-1 shows these notations. An example of the nomenclature results in the following definitions for the change in velocity at the second maneuver.

$$\Delta v_b = v_{final} - v_{trans_b}$$

The subscripts at points a and b are necessary even with the initial and final orbits because they may be noncircular. I'll use this notation throughout this section, so it's important to understand it.

5.3 Coplanar Maneuvers

As the name implies, coplanar maneuvers don't change the orbital plane, so the initial and final orbits lie in the same plane. These maneuvers can change the orbit's size and shape (semimajor axis and eccentricity) and the location of the line of apsides (argument of perigee). Coplanar burns are either *tangential* or *nontangential*. All of the historical work in coplanar transfers was restricted to circular orbits for which the velocity vector is always tangential to the orbit. For maximum efficiency of the burn, we must impose an *additional requirement* that the flight-path angle be zero, so tangential burns occur only at apoapsis and periapsis on elliptical orbits. These burns allow us to do three types of coplanar

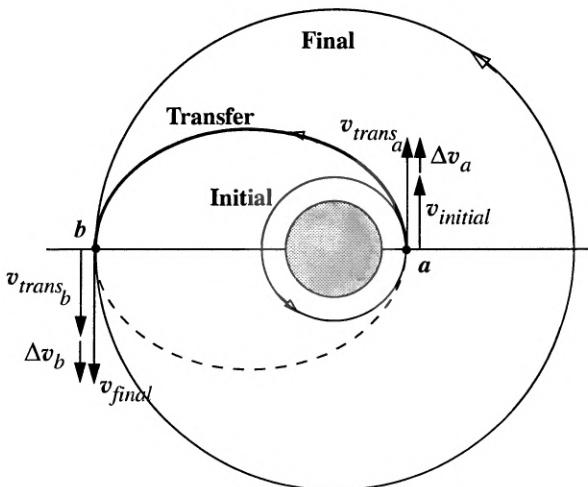


Figure 5-1. Orbit Notation. The letters a, b, \dots represent changes in velocities and locations of the firings. When we use multiple transfer orbits, numbers signify which transfer we're considering.

changes: Hohmann transfers (two tangential burns), one-tangent burns (one tangential burn), and general transfers (two nontangential burns).

Consider the simple tangential transfer of Fig. 5-2. Here, both orbits are tangent at the transfer point. As a result, the velocity vectors are parallel, and we can directly find the required change in velocity:

$$\Delta v_a = v_{final} - v_{initial} \quad (5-1)$$

Although the figure indicates a vector relation, the equation is scalar because of the velocity vector's alignment. Most analyses using these algorithms don't need the vector components in an answer. Notice we can satisfy the requirement for tangential burns with circular, elliptical, parabolic, and hyperbolic orbits, with the only restriction being that the flight-path angle, ϕ_{fpa} , must be 0° . We can determine the direction of firing by the sign of the change in velocity. For instance, the left orbit in Fig. 5-2 has a positive Δv because the velocity is added in the same direction as the original velocity vector. In the other orbit, the change in velocity is applied *opposite* to the direction of motion, and the satellite slows down to the circular orbit as shown. I've written equations in this book so we can determine positive and negative changes in the velocities. Determining the correct orientation for a firing is very important. It relies on accurate knowledge of the satellite's attitude, which we won't explore due to the subject's length. Selecting a motor depends only on the maximum needed change in velocity, but *using* the motor requires a direction. The direction is important for planning and operations.

We can find v_{final} and $v_{initial}$ from the velocity equations [Eq. (2-20), Eq. (2-21), Eq. (2-22) and Eq. (2-23)]:

$$v = \sqrt{2\left(\frac{\mu}{r} + \xi\right)} \quad v = \sqrt{\frac{2\mu}{r} - \frac{\mu}{a}}$$

$$v = \sqrt{\frac{\mu}{r}\left(2 - \frac{1-e^2}{1+e\cos(\nu)}\right)} \quad v_{circle} = \sqrt{\frac{\mu}{r}}$$

We need various equations to provide appropriate expressions for all potential problems. Because we eventually want general solutions, I'll try to express all equations in terms of *position*, r , *eccentricity*, e , and *true anomaly*, ν . Using the semimajor axis is more convenient if the orbits are all circular. If we need the semimajor axis, we can readily determine it and could replace eccentricity and true anomaly as variables.

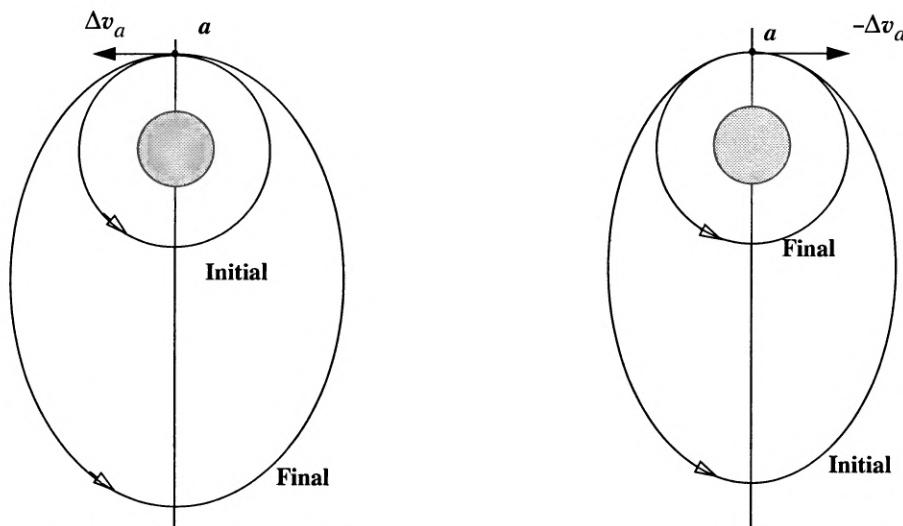


Figure 5-2. Tangential Orbit Transfer ($\phi_{fpa} = 0^\circ$). Tangential burns are defined to occur with a zero flight-path angle. Velocity increases enlarge the semimajor axis (left), whereas decreases cause the orbit to come closer to the Earth (right).

If we perform a tangential Δv_a maneuver to enter a transfer orbit, we may need to consider a nontangential transfer ($\phi_{fpa} \neq 0$) to achieve the final orbit. In this case, the velocity vectors are no longer parallel (Fig. 5-3), although the orbits are still coplanar. From Fig. 5-3, the included angle between the velocity vectors is the difference in the flight-path angles ($\phi_{transb} - \phi_{final}$). We can get the final flight-path-angle from Eq. (2-57). We measure the flight-path angles from the same horizon line (the satellite has the same position immediately before and after the burn) because we've assumed an instantaneous

change in velocity. Determine the velocities v_{transb} and v_{final} using Eq. (2-20), Eq. (2-21), Eq. (2-22), or Eq. (2-23). Make sure you use the proper orbital elements for each case.

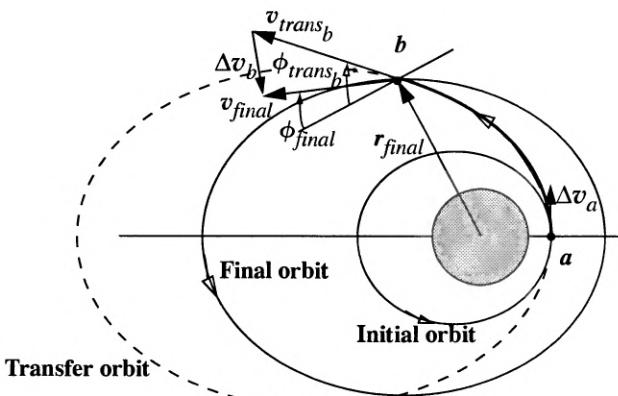


Figure 5-3. Nontangential Orbit Transfer ($\phi_{fpa} \neq 0$). The nontangential transfer, point b , has nonparallel velocity vectors because both orbits have flight-path angles greater than zero.

To determine the change in velocity, examine the figure and notice that if two sides and the included angle of a planar triangle are known, you can find the third side using the law of cosines for plane triangles, Eq. (C-19). For this case, you know the final velocity from parameters of the desired final orbit, and the magnitude and vector relations become

$$\Delta v_b = \sqrt{v_{transb}^2 + v_{final}^2 - 2v_{transb}v_{final}\cos(\phi_{transb} - \phi_{final})} \quad (5-2)$$

$$\vec{\Delta v}_b = \vec{v}_{final} - \vec{v}_{transb}$$

If the two orbits are tangent, you should be able to reduce this problem to the tangential case of Eq. (5-1). Using the (initial – final) convention allows you to determine whether the flight-path angle is increasing or decreasing. This is useful in calculations to determine if the satellite will reenter during the transfer. Again, this maneuver may occur between any of the different types of orbits, with the only difference being the equation used to calculate the initial and final velocity magnitudes.

In Sec. 4.2.6, I've already shown how to determine the flight-path angle as a function of position, eccentricity, and true anomaly. With an understanding of these two types of burns, we may now examine three different types of coplanar orbit changes.

5.3.1 Hohmann and Bi-elliptic Transfers

Walter Hohmann (1880–1945) proposed a theory which suggested the minimum-energy (and therefore most efficient) transfer could be achieved between orbits by using two tan-

gential burns (Hohmann, 1925). Although his original work considered only transfer between circular orbits, other authors have explored transfers between coaxially aligned elliptical orbits and concluded the transfer energy was lowest using two tangential burns (Lawden, 1952, and Thompson, 1986, 70). Remember that ϕ_{fpa} must be zero. This type of transfer is called the *Hohmann transfer*, and the resulting transfer orbit between two circular orbits is elliptical; the transfer between two elliptical orbits may be circular or elliptical depending on the geometry of the initial and final orbits. Notice the necessary condition for tangential burns excludes using parabolic and hyperbolic transfer orbits because of the second tangential burn. Figure 5-4 shows Hohmann transfers from circular and elliptical initial orbits.

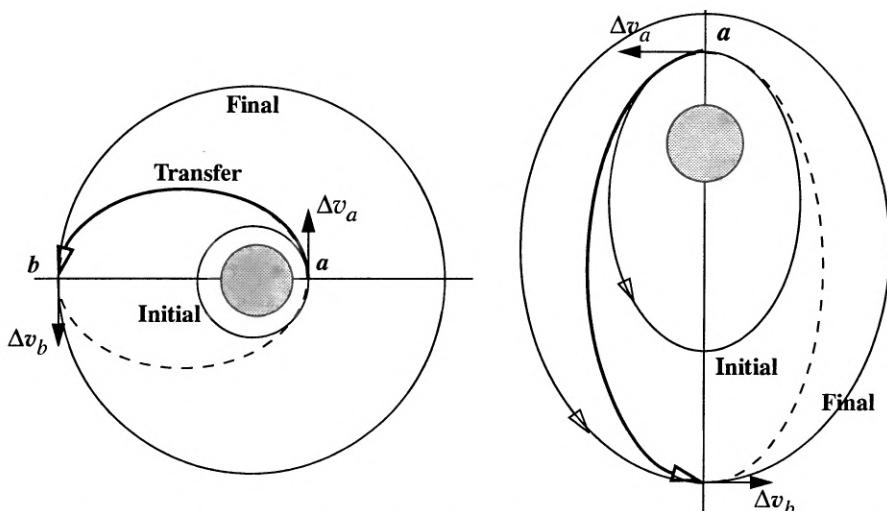


Figure 5-4. Hohmann Transfer. The two-impulse minimum change in velocity maneuver is the Hohmann transfer. Elliptical transfers must start at either apogee or perigee for this maneuver to be a minimum.

Because the radii of apogee and perigee are known, the semimajor axis of the transfer is readily defined, and we discover that the transfer time, τ_{trans} , for the Hohmann transfer is simply half the orbital period of the transfer orbit:

$$a_{trans} = \frac{r_{initial} + r_{final}}{2}$$

$$\tau_{trans} = \frac{\rho_{trans}}{2} = \pi \sqrt{\frac{a_{trans}^3}{\mu}} \quad (5-3)$$

A variant of the Hohmann transfer is a method which actually performs two Hohmann transfers in series. Figure 5-5 shows the *bi-elliptic transfer* as a transfer into the transfer

ellipse, $trans_1$, at point a , followed by a transfer into a second transfer ellipse, $trans_2$, at point b , and a transfer into the final orbit at point c . Note that there is an intermediate circular orbit between the two elliptical transfer orbits. We simplify the middle velocity-change calculation by determining it using the two elliptical orbits, rather than two separate circular orbits. The technique is usually more efficient for distant transfers, as we'll show later. We can also transfer from elliptical orbits, as long as the tangential ($\phi_{fpa} = 0$) condition is satisfied. The intermediate radius at point b (r_b) *must* be a value greater than the final radius, r_{final} . Available thrust and timing requirements usually determine how we select r_b . Sec. 5.3.2 will show several values we can use to find general characteristics of r_b . Finally, the transfer time for the maneuver is now the sum of the two Hohmann-like transfer times. Sec. 5.3.2 shows that in some cases, the bi-elliptic transfer can reduce the total Δv necessary for the transfer.

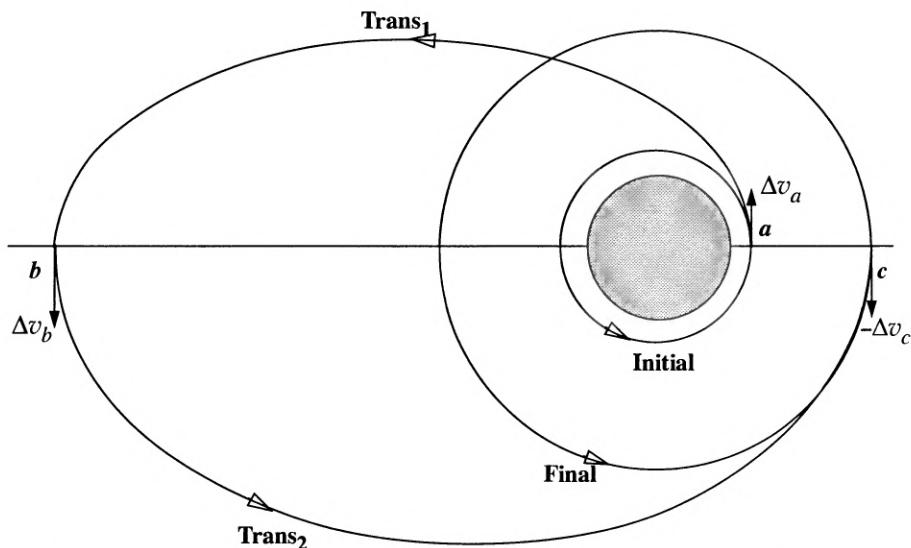


Figure 5-5. Bi-elliptic Transfer. The bi-elliptic transfer employs two Hohmann transfers in succession. This allows us to combine the middle two burns. Using three burns adds complexity but can save Δv in some instances.

Implementing Hohmann-Transfer Solutions

The algorithm for the Hohmann transfer is rather long if we include the potential for transfers between elliptical orbits. The added benefit of including this possibility is marginal at best because many additional options arise for this case. If you choose to examine elliptical transfers, the following items are important. You must

1. Recognize that the initial and final orbits will be coaxially aligned; that is, the semimajor axes will be aligned. This is the only way that you can

achieve two tangential burns and have the true anomalies of each orbit be $\pm 180^\circ$ or 0° . The circular case requires this alignment, but it's especially relevant for the elliptical case.

- Use velocity formulas for elliptical orbits in the formulas. The initial and final orbits will now have varying velocities.

$$v = \sqrt{\frac{2\mu}{r} - \frac{\mu}{a}}$$

- Specify the semimajor axes of the initial and final orbits from the trajectory equation (+ apogee, - perigee).

$$a = r \left(\frac{1 + e \cos(\nu)}{1 - e^2} \right) = \frac{r}{1 \pm e}$$

- Draw a diagram! It's very important to get the right formula for the particular problem being solved.

The process is arduous; thus, I'll present an algorithm only for the "classic" situation—that of circular orbits. The most important thing to remember is the significance of notation and the correct use of various values for the radii and semimajor axes throughout the calculations. Always draw a quick diagram (even with the circular case) to ensure you're using the correct values. This becomes even more important when we reverse the process and the transfer takes place from apoapsis to periapsis on the transfer ellipse.

We improve computational efficiency by using canonical units, so we can drop many mathematical operations involving the gravitational parameter. We can also use a temporary variable for the reciprocal of the radii.

ALGORITHM 33: Hohmann Transfer

$$(r_{initial}, r_{final} \Rightarrow a_{trans}, \tau_{trans}, \Delta v_a, \Delta v_b)$$

$$a_{trans} = \frac{r_{initial} + r_{final}}{2}$$

$$v_{initial} = \sqrt{\frac{\mu}{r_{initial}}} \quad v_{trans_a} = \sqrt{\frac{2\mu}{r_{initial}} - \frac{\mu}{a_{trans}}}$$

$$v_{final} = \sqrt{\frac{\mu}{r_{final}}} \quad v_{trans_b} = \sqrt{\frac{2\mu}{r_{final}} - \frac{\mu}{a_{trans}}}$$

$$\Delta v_a = v_{trans_a} - v_{initial}$$

$$\Delta v_b = v_{final} - v_{trans_b}$$

$$\Delta v = |\Delta v_a| + |\Delta v_b|$$

$$\tau_{trans} = \pi \sqrt{\frac{a_{trans}^3}{\mu}}$$

The bi-elliptic algorithm is similar to the Hohmann, but the notation differs due to the extra required burn. Sketching and carefully examining the required parameters are even more important here. The same extra conditions apply as in the Hohmann transfer if we use elliptical orbits; thus, I'll still present just the circular case.

ALGORITHM 34: *Bi-elliptic Transfer* ($r_{initial}, r_b, r_{final} \Rightarrow a_{trans1}, a_{trans2}, \tau_{trans1}, \tau_{trans2}, \Delta v_a, \Delta v_b, \Delta v_c$)

$$a_{trans1} = \frac{r_{initial} + r_b}{2}$$

$$a_{trans2} = \frac{r_b + r_{final}}{2}$$

$$v_{initial} = \sqrt{\frac{\mu}{r_{initial}}}$$

$$v_{trans1a} = \sqrt{\frac{2\mu}{r_{initial}} - \frac{\mu}{a_{trans1}}}$$

$$v_{trans1b} = \sqrt{\frac{2\mu}{r_b} - \frac{\mu}{a_{trans1}}}$$

$$v_{trans2b} = \sqrt{\frac{2\mu}{r_b} - \frac{\mu}{a_{trans2}}}$$

$$v_{trans2c} = \sqrt{\frac{2\mu}{r_{final}} - \frac{\mu}{a_{trans2}}}$$

$$v_{final} = \sqrt{\frac{\mu}{r_{final}}}$$

$$\Delta v_a = v_{trans1a} - v_{initial}$$

$$\Delta v_b = v_{trans2b} - v_{trans1b}$$

$$\Delta v_c = v_{final} - v_{trans2c}$$

$$\Delta v = |\Delta v_a| + |\Delta v_b| + |\Delta v_c|$$

$$\tau_{trans} = \pi \sqrt{\frac{a_{trans1}^3}{\mu}} + \pi \sqrt{\frac{a_{trans2}^3}{\mu}}$$

An example transfer is useful to compare each method.

▼ Example 5-1: Calculating a Hohmann Transfer.

GIVEN: Initial orbit alt = 191,344.11 km
 Final orbit alt = 35,781,348.57 km

FIND: Δv , time of flight (TOF)

Begin by converting units and determining radii values:

$$r_{initial} = \frac{191,344\ 11 + 6378.1363}{6378.137} = 1.03 \text{ ER}$$

$$r_{final} = \frac{35,781,348\ 57 + 6378.1363}{6378.137} = 6.61 \text{ ER}$$

Next, find the velocities at each position in the transfer. From Eq. (2-23) for a satellite's circular speed,

$$v_{initial} = \sqrt{\frac{\mu}{r_{initial}}} = \sqrt{\frac{1}{1.03}} = 0.985\ 329 \frac{\text{ER}}{\text{TU}}$$

$$v_{final} = \sqrt{\frac{\mu}{r_{final}}} = \sqrt{\frac{1}{6.61}} = 0.388\ 955 \frac{\text{ER}}{\text{TU}}$$

For the transfer orbit,

$$a_{trans} = \frac{r_{initial} + r_{final}}{2} = 3.820 \text{ ER}$$

From Eq. (2-21) for elliptical satellite speed, find the two transfer velocities:

$$v_{trans_a} = \sqrt{\frac{2\mu}{r_{initial}} - \frac{\mu}{a_{trans}}} = \sqrt{\frac{2(1)}{1.03} - \frac{1}{3.82}} = 1.296\ 136 \frac{\text{ER}}{\text{TU}}$$

$$v_{trans_b} = \sqrt{\frac{2\mu}{r_{final}} - \frac{\mu}{a_{trans}}} = \sqrt{\frac{2(1)}{6.61} - \frac{1}{3.82}} = 0.201\ 970 \frac{\text{ER}}{\text{TU}}$$

The total change in velocity is the sum of the two burns. Notice that each velocity is positive, indicating increasing semimajor axis values.

$$\Delta v = \Delta v_a + \Delta v_b = |v_{trans_a} - v_{initial}| + |v_{final} - v_{trans_b}|$$

$$\Delta v = |1.296\ 136 - 0.985\ 329| + |0.388\ 955 - 0.201\ 970| = 0.497\ 792 \frac{\text{ER}}{\text{TU}} = 3.935\ 224 \frac{\text{km}}{\text{s}}$$

Time of flight for a Hohmann transfer is half the period of the transfer orbit:

▲ $\tau_{trans} = \pi \sqrt{\frac{a_{trans}^3}{\mu}} = \pi (3.82)^{3/2} = 23.455\ 512 \text{ TU} = 5.256\ 713 \text{ h}$

▼ Example 5-2. Calculating a Bi-elliptic Transfer.

GIVEN: Initial orbit alt = 191,344 11 km
 Point b alt = 503,873 km (79 ER)
 Final orbit alt = 37,6310 km (59 ER)

FIND: Δv , time of flight (TOF)

From Example 5-1, we know that $r_{initial} = 1.03 \text{ ER}$, $r_b = 80 \text{ ER}$, and $r_{final} = 60 \text{ ER}$. We must find the velocities at each point of the transfer. From Eq. (2-23) for circular satellite speed,

$$v_{initial} = \sqrt{\frac{\mu}{r_{initial}}} = \sqrt{\frac{1}{1.03}} = 0.985\ 329 \frac{\text{ER}}{\text{TU}}$$

$$v_{final} = \sqrt{\frac{\mu}{r_{final}}} = \sqrt{\frac{1}{60.00}} = 0.129\ 099 \frac{\text{ER}}{\text{TU}}$$

For the transfer orbits,

$$a_{trans1} = \frac{r_{initial} + r_b}{2} = \frac{1.03 + 80}{2} = 40.515 \text{ ER}$$

$$a_{trans2} = \frac{r_b + r_{final}}{2} = \frac{80 + 60}{2} = 70.0 \text{ ER}$$

From Eq. (2-21) for a satellite's elliptical speed,

$$v_{trans1_a} = \sqrt{\frac{2\mu}{r_{initial} - a_{trans1}}} = \sqrt{\frac{2(1)}{1.03 - 40.515}} = 1.384\ 581 \frac{\text{ER}}{\text{TU}}$$

$$v_{trans2_c} = \sqrt{\frac{2\mu}{r_{final} - a_{trans2}}} = \sqrt{\frac{2(1)}{60.000 - 70.000}} = 0.138\ 013 \frac{\text{ER}}{\text{TU}}$$

and for the additional burn of the bi-elliptic,

$$v_{trans1_b} = \sqrt{\frac{2\mu}{r_b - a_{trans1}}} = \sqrt{\frac{2(1)}{80.000 - 40.515}} = 0.017\ 826 \frac{\text{ER}}{\text{TU}}$$

$$v_{trans2_b} = \sqrt{\frac{2\mu}{r_b - a_{trans2}}} = \sqrt{\frac{2(1)}{80.000 - 70.000}} = 0.103\ 510 \frac{\text{ER}}{\text{TU}}$$

The total change in velocity is the sum of the three tangential burns. Notice here that the final burn is negative, indicating the satellite must slow down.

$$\Delta v = \Delta v_a + \Delta v_b + \Delta v_c = |v_{trans1_a} - v_{initial}| + |v_{trans2_b} - v_{trans1_b}| + |v_{final} - v_{trans2_c}|$$

$$\Delta v = |1.384\ 581 - 0.985\ 329| + |0.103\ 510 - 0.017\ 826| + |0.129\ 099 - 0.138\ 013|$$

$$\Delta v = 0.493\ 849 \frac{\text{ER}}{\text{TU}} = 3.904\ 057 \frac{\text{km}}{\text{s}}$$

Time of flight for a bi-elliptic transfer is half the period of both the transfer orbits:

$$\tau_{trans} = \pi \sqrt{\frac{a_{trans1}^3}{\mu}} + \pi \sqrt{\frac{a_{trans2}^3}{\mu}} = \pi \sqrt{\frac{40.515^3}{1}} + \pi \sqrt{\frac{70.000^3}{1}}$$



$$\tau_{trans} = 2650.076\ 794 \text{ TU} = 593.919\ 803 \text{ h}$$

5.3.2 Comparing Hohmann and Bi-elliptic Transfers

The Hohmann transfer is the minimum-energy transfer between most—but not all—coplanar orbits. In some cases, the bi-elliptic transfer may use less energy. To determine the regions where each transfer is superior, consider the following development. This development follows Escobal ([1968] 1979, 58–67) and considers *only* circular motion.

For a Hohmann transfer, write the change in velocity Δv_a as follows. Recall the velocity formulas for elliptical and circular orbits [Eq. (2-21) and Eq. (2-23), respectively] and remember that $a_{initial} = r_a$ for circular orbits:

$$\Delta v_a = \sqrt{\frac{2\mu}{a_{initial}}} - \sqrt{\frac{\mu}{a_{trans}}} - \sqrt{\frac{\mu}{a_{initial}}}$$

but because $a_{trans} = \frac{a_{initial} + a_{final}}{2}$,

$$\Delta v_a = \sqrt{\frac{2\mu}{a_{initial}} \left(\frac{a_{final}}{a_{initial} + a_{final}} \right)} - \sqrt{\frac{\mu}{a_{initial}}}$$

Likewise,

$$\Delta v_b = -\sqrt{\frac{2\mu}{a_{final}} \left(\frac{a_{initial}}{a_{initial} + a_{final}} \right)} + \sqrt{\frac{\mu}{a_{final}}}$$

Write the total Δv as the sum of these changes in velocities. Then, scale the result by the magnitude of $v_{initial}$, using this ratio (with *circular* orbits you may also use the semimajor axes):

$$R \equiv r_{final} / r_{initial} = a_{final} / a_{initial} \quad (5-4)$$

Remembering the formula for circular velocity, write

$$\frac{\Delta v_{Hohmann}}{v_{initial}} = \left(1 - \frac{1}{R} \right) \sqrt{\frac{2R}{1+R}} + \sqrt{\frac{1}{R}} - 1 \quad (5-5)$$

Figure 5-6 shows the magnitude of the total velocity change for different values of R . Find the maxima by differentiating Eq. (5-5) with respect to R and setting it equal to zero. From Escobal ([1968] 1979, 60),

$$R^3 - 15R^2 - 9R - 1 = 0 \quad R = 15.581\,76 \quad (5-6)$$

Now consider the limiting bi-elliptic case in which the intermediate transfer maneuver occurs at an infinite distance. The initial Δv_a is simply the limit of Eq. (5-5) as R goes to ∞ , and the final Δv_b is velocity to escape from infinity to the final orbit. To find Δv_b you must use Eq. (5-5) but scale the result by v_{final} because $v_{initial}$ at ∞ is zero:

$$\Delta v_a = (\sqrt{2} - 1) v_{initial}$$

$$\Delta v_b = (\sqrt{2} - 1) v_{final}$$

Write the total Δv as the sum of these and, to make the result general, divide by the magnitude of $v_{initial}$. Again from Escobal ([1968] 1979, 61),

$$\frac{\Delta v_{\infty}}{v_{initial}} = (\sqrt{2} - 1) \left(1 + \sqrt{\frac{1}{R}} \right) \quad (5-7)$$

Figure 5-6 plots the total change in velocity for the bi-elliptic burns, as well as the Hohmann change in velocity. This will allow us to evaluate the efficiency of each method. We find the intersection between the two by equating Eq. (5-5) and Eq. (5-7), so

$$R^3 - (7 + 4\sqrt{2}) R^2 + (3 + 4\sqrt{2}) R - 1 = 0 \quad R = 11.93876 \quad (5-8)$$

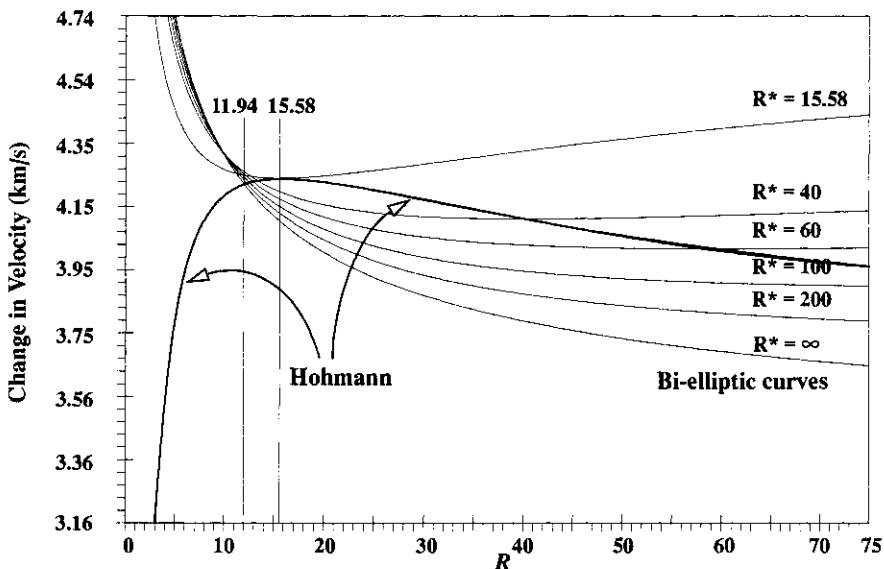


Figure 5-6. Hohmann Transfer—Comparison of Change in Velocity. The change in velocity is shown when using Hohmann and bi-elliptic orbit changes. Be aware of the different notations for the ratios (final / initial) (R for Hohmann and R^* for bi-elliptic).

We can derive the bi-elliptic case similarly, except that the intermediate radius requires a new definition, $R^* = r_b / r_{initial}$ (Escobal, [1968] 1979, 63):

$$\frac{\Delta v_{bi-elliptic}}{v_{initial}} = \sqrt{\frac{2R^*}{1+R^*}} - 1 + \sqrt{\frac{2}{R^*}} \left(\sqrt{\frac{1}{1+\frac{R^*}{R}}} - \sqrt{\frac{1}{1+R^*}} \right) + \sqrt{\frac{1}{R}} \left(\sqrt{\frac{2R^*}{R+R^*}} - 1 \right) \quad (5-9)$$

Figure 5-6 also shows bi-elliptic changes in velocities.

Now, let's draw some conclusions. Using the ratio R defined previously,

1. If $0.0 < R < 11.94$, the Hohmann transfer is superior.
2. If $11.94 < R < 15.58$, we need another test to see which method is best for realistic transfers. Escobal ([1968] 1979, 66) describes and gives equations for these conditions.
3. If $15.58 < R < \infty$, the bi-elliptic method *can be* superior, although we may need a very large intermediate orbit, significantly increasing the transfer time.

Escobal also notes that if the final radius is larger than both the initial and transfer orbits, the Hohmann transfer is always superior in terms of total change in velocity. This is reasonable because all cases in which the final radius is largest are like phasing orbits, which I'll show later are all variants of the basic Hohmann transfer.

The 11.94 and 15.58 ratios are known as the *critical limits* of the transfer problem. We find them by noting the intersections of the Hohmann transfer curve with the bi-elliptic curves ($R^* = \infty$ and $R^* = 15.58$). The equations for Δv and time of flight are the same as the Hohmann transfer, but the two maneuvers need two sets of equations.

5.3.3 Transfers Using the One-Tangent Burn

The major drawback to the Hohmann transfer is the long flight time. The bi-elliptic takes even longer. To reduce time of flight, we must select a trajectory using a shorter path with higher velocities. The fastest possible path would involve a Δv approaching infinity, but that's not practical. Thus, we must select a trajectory that reduces time of flight at the expense of an acceptable increase in Δv . One solution is the one-tangent burn.

As the name implies, a *one-tangent burn* has one tangential burn and one nontangential burn. This method reduces the transfer time of the Hohmann techniques but increases Δv requirements. Figure 5-7 shows the situation. As in the Hohmann transfer, the orbits must be circular or coaxial elliptic. Unlike the Hohmann transfer, however, the one-tangent burn routinely uses any orbit type for the transfer orbit and often uses parabolic and hyperbolic transfers, depending on the available time and change in velocity capability. Note that we must know the transfer orbit's true anomaly (semimajor axis, or eccentricity) to locate the nontangential burn. If we use elliptical initial and final orbits, the transfer true anomaly is identical to the final or initial true anomaly. This is evident in Fig. 5-7 because of the requirement for the ellipses to be coaxially aligned.

A fundamental quantity is the eccentricity of the transfer orbit. Using known values, we can find expressions for the initial and final positions. Although the satellite is shown at periapsis of the transfer orbit in both cases, Eq. (2-38) applies for the radius of periapsis (-) or apoapsis (+).

$$r_{initial} = a_{trans} (1 + e_{trans})$$

$$r_{final} = \frac{a_{trans} (1 - e_{trans}^2)}{1 + e_{trans} \cos(\nu_{trans_b})}$$

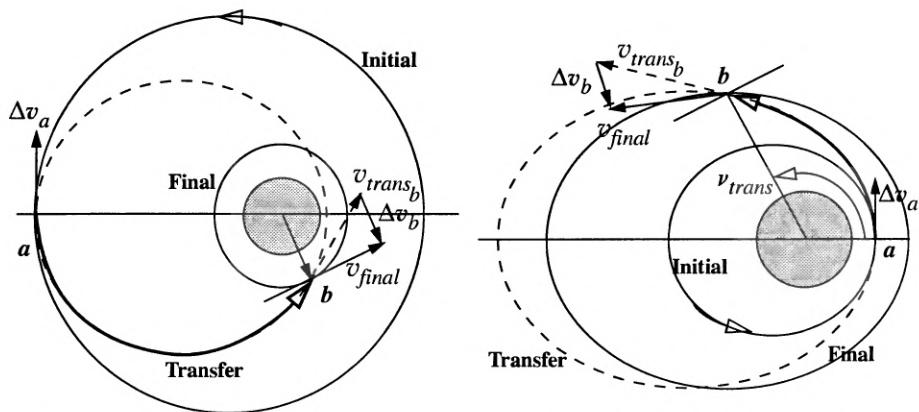


Figure 5-7. One-Tangent Burns. For both circular (left) and elliptical (right) orbit changes, the true anomalies of the transfer and final orbits are equal. The maneuver starts at apoapsis (left) and periapsis (right) of the transfer orbit.

Substituting $r_{initial}$ for $a_{trans}(1 \mp e_{trans})$ in the numerator (periapsis and apoapsis) gives us

$$r_{final} = \frac{r_{initial}(1 \pm e_{trans})}{1 + e_{trans} \cos(\nu_{trans_b})}$$

Cross-multiplying and separating terms of eccentricity results in

$$e_{trans} \left(r_{final} \cos(\nu_{trans_b}) \mp r_{initial} \right) = r_{initial} - r_{final}$$

Using the reciprocal of Eq. (5-4), we see that

$$R^{-1} = \frac{r_{initial}}{r_{final}}$$

and solving for eccentricity yields the desired result:

$$e_{trans} = \frac{R^{-1} - 1}{\cos(\nu_{trans_b}) \mp R^{-1}} \quad \begin{cases} -\text{periapsis} \\ +\text{apoapsis} \end{cases} \quad (5-10)$$

What difference, if any, must take place if the tangential burn occurs at apoapsis and not periapsis? The change in velocity must be in the opposite direction and reformulated because now apoapsis is the first burn location. The values of true anomaly are also limited. By solving for the denominator of Eq. (5-10), we see that the true anomaly must be

less than one revolution and must obey the following restriction to avoid numerical indeterminacy [determine the signs opposite from Eq. (5-10)]:

$$\nu_{trans_b} \neq \pm \cos^{-1}(R^{-1}) \quad \begin{cases} + \text{ periapsis} \\ - \text{ apoapsis} \end{cases}$$

We need the semimajor axis for the transfer ellipse, but we can't calculate it as in Eq. (5-3) because values of periapsis and apoapsis don't exist. Yet we can determine the transfer semimajor axis knowing that the maneuver originates at either apoapsis or periapsis. Solving Eq. (2-38), we see that

$$a_{trans} = \frac{r_{initial}}{1 + e_{trans}}$$

Remember the multiple signs are necessary to enable calculations where the maneuver originates at periapsis or apoapsis (respectively) of the transfer orbit.

Find the initial velocity change using techniques from the Hohmann transfer and the second velocity change using the cosine law [Eq. (5-2)]. Remember, you must know the flight-path angle [Eq. (2-58)].

Finally, calculate the transfer time using Kepler's equation [Eq. (4-7)] because the burn depends on the true anomaly of the transfer orbit where the nontangential transfer occurs. Remember to include the number of times the orbit goes through periapsis, and calculate the eccentric anomaly, E , using Eq. (4-8). Here, the true anomaly takes on values for the transfer at a and b , and you need both the sine and cosine values for a computer solution. If you use only one of these formulas, be sure to check for the correct quadrant. This development changes slightly if the maneuver always starts at periapsis. The values of eccentric anomaly, E_o , will be zero, so the time equation is simpler.

Implementing Solutions for the One-Tangent Burn

We can write the one-tangent burn for elliptical orbits, but I'm presenting only the circular case here for the reasons identified with the Hohmann transfer.

ALGORITHM 35: One-Tangent Burn ($r_{initial}, r_{final}, \nu_{trans_b} \Rightarrow e_{trans}, a_{trans}, \tau_{trans}, \Delta v_a, \Delta v_b$)

$$R^{-1} = \frac{r_{initial}}{r_{final}}$$

$$e_{trans} = \frac{R^{-1} - 1}{\cos(\nu_{trans_b}) + R^{-1}} \quad \begin{cases} - \text{ periapsis} \\ + \text{ apoapsis} \end{cases}$$

$$a_{trans} = \frac{r_{initial}}{1 + e_{trans}}$$

$$v_{initial} = \sqrt{\frac{\mu}{r_{initial}}} \quad v_{trans_a} = \sqrt{\frac{2\mu}{r_{initial}} - \frac{\mu}{a_{trans}}}$$

$$v_{final} = \sqrt{\frac{\mu}{r_{final}}} \quad v_{trans_b} = \sqrt{\frac{2\mu}{r_{final}} - \frac{\mu}{a_{trans}}}$$

$$\Delta v_a = v_{trans_a} - v_{initial}$$

$$\tan(\phi_{trans_b}) = \frac{e_{trans} \sin(v_{trans_b})}{1 + e_{trans} \cos(v_{trans_b})}$$

$$\Delta v_b = \sqrt{v_{trans_b}^2 + v_{final}^2 - 2v_{trans_b}v_{final} \cos(\phi_{trans_b})}$$

$$\Delta v_{otb} = |\Delta v_a| + |\Delta v_b|$$

$$\cos(E) = \frac{e_{trans} + \cos(v_{trans_b})}{1 + e_{trans} \cos(v_{trans_b})}$$

$$\tau_{trans} = \sqrt{\frac{a_{trans}^3}{\mu}} \left\{ 2k\pi + (E - e_{trans} \sin(E)) - (E_o - e_{trans} \sin(E_o)) \right\}$$

Let's look at an example of a one-tangent burn, so we can compare it to other methods.

▼ Example 5-3. Performing a One-tangent Burn.

GIVEN: Initial orbit alt = 191,344 11 km
 Final orbit alt = 35,781,348 57 km
 Transfer point $v_{trans_b} = 160^\circ$

FIND: Δv , time of flight (τ_{trans})

From Example 5-1, we know that the initial and final radii are 1.03 ER and 6.61 ER, respectively. Before determining the total change in velocity, calculate the transfer orbit's eccentricity:

$$R^{-1} = \frac{r_{initial}}{r_{final}} = \frac{1.03}{6.61} = 0.155\ 821$$

$$e_{trans} = \frac{R^{-1} - 1}{\cos(v_{trans_b}) - R^{-1}} = \frac{0.155\ 821 - 1}{\cos(160^\circ) - 0.155\ 821} = 0.770\ 578$$

$$a_{trans} = \frac{r_{initial}}{(1 - e_{trans})} = \frac{1.03}{1 - 0.770\,578} = 4.489\,303 \text{ ER}$$

From the formula for a satellite's circular speed,

$$v_{initial} = \sqrt{\frac{\mu}{r_{initial}}} = \sqrt{\frac{1}{1.03}} = 0.985\,355 \frac{\text{ER}}{\text{TU}}$$

$$v_{final} = \sqrt{\frac{\mu}{r_{final}}} = \sqrt{\frac{1}{6.61}} = 0.388\,961 \frac{\text{ER}}{\text{TU}}$$

From the equation for a satellite's elliptical speed,

$$v_{trans_a} = \sqrt{\frac{2\mu}{r_{initial}} - \frac{\mu}{a_{trans}}} = \sqrt{\frac{2(1)}{1.03} - \frac{1}{4.489\,303}} = 1.311\,144 \frac{\text{ER}}{\text{TU}}$$

$$v_{trans_b} = \sqrt{\frac{2\mu}{r_{final}} - \frac{\mu}{a_{trans}}} = \sqrt{\frac{2(1)}{6.61} - \frac{1}{4.489\,303}} = 0.282\,542 \frac{\text{ER}}{\text{TU}}$$

$$\Delta v_a = v_{trans_a} - v_{initial} = 1.311\,144 - 0.985\,355 = 0.325\,789 \frac{\text{ER}}{\text{TU}}$$

Use the flight-path angle for the nontangential transfer and, from Eq. (2-58) and Eq. (2-59),

$$\tan(\phi_{trans_b}) = \frac{e_{trans} \sin(\nu_{trans_b})}{1 + e_{trans} \cos(\nu_{trans_b})} = \frac{0.770\,578 \sin(160^\circ)}{1 + 0.770\,578 \cos(160^\circ)} \quad \phi_{trans_b} = 43.689\,288^\circ$$

Because the final orbit is circular, the angle change is ϕ_{trans_b} , so

$$\Delta v_b = \sqrt{v_{final}^2 + v_{trans_b}^2 - 2v_{final}v_{trans_b} \cos(\phi_{trans_b})}$$

$$\Delta v_b = \sqrt{(0.388\,961)^2 + (0.282\,542)^2 - 2(0.388\,961)(0.282\,542) \cos(43.689\,288^\circ)} = 0.268\,678 \frac{\text{ER}}{\text{TU}}$$

The total Δv_{otb} is simply the sum of the two burns:

$$\Delta v_{otb} = |\Delta v_a| + |\Delta v_b| = 0.325\,789 + 0.268\,678 = 0.594\,466 \frac{\text{ER}}{\text{TU}} = 4.699\,475 \frac{\text{km}}{\text{s}}$$

Calculate time of flight using Kepler's equation:

$$\tau_{trans} = \sqrt{\frac{a_{trans}^3}{\mu}} (2k\pi + (E - e_{trans} \sin(E)) - (E_o - e_{trans} \sin(E_o)))$$

Because this transfer starts at periapsis, $E_o = 0^\circ$. The transfer doesn't pass perigee, so k must equal zero. From Eq. (4-9),

$$\cos(E) = \frac{e_{trans} + \cos(\nu_{trans_b})}{1 + e_{trans} \cos(\nu_{trans_b})} = \frac{0.770\,578 + \cos(160^\circ)}{1 + 0.770\,578 \cos(160^\circ)}$$

$$E = 127.804\,010^\circ \text{ and } 232.195\,990^\circ$$

Choose $E = 127.804\,010^\circ$ because ν_{trans_b} is in the second quadrant. Substituting the values into the time of flight equation [Eq. (4-6)] results in

$$\tau_{trans} = \sqrt{4.489\,303^3} \left(\frac{127.804\,010}{57.2958} - 0.770\,578 \sin(127.804\,010^\circ) \right)$$

▲ $\tau_{trans} = 15.426\,19 \text{ TU} = 3.457\,228 \text{ h}$

Table 5-1 summarizes the results of the examples and additional calculations. Notice that using the one-tangent burn when going to geosynchronous orbit reduces time of flight 1.799 hours or 34.2%, for an increase in Δv of 0.764 km/s or 19.4% for $\nu_{trans_b} = 160^\circ$. For a transfer to the Moon using $\nu_{trans_b} = 175^\circ$, we save 30.0% of the time while expending about 0.133 km/s of Δv or 3.35%.

TABLE 5-1. Comparison of Coplanar Orbital Transfers. This table presents results of the trade-off between change in velocity and time of flight for different types of orbital changes. Recognize that the initial and final orbits are circular, whereas the transfer orbits are elliptical, with perigee and apogee values as shown. Notice that it's almost as hard to go to geosynchronous as it is to the Moon. You can verify this from Fig. 5-6.

	Initial Alt (km)	Final Alt (km)	ν_{trans_b}	Bi-elliptic Transfer Alt (km)	Δv (km/s)	τ_{trans} (h)
Transfer to Geosynchronous						
Hohmann	191,344.11	35,781.35			3.935	5.256
One-tangent	191,344.11	35,781.35	160°		4.699	3.457
Bi-elliptic	191,344.11	35,781.35		47,836.00	4.076	21.944
Transfer to the Moon						
Hohmann	191,344.11	376,310.00			3.966	118.683
One-tangent	191,344.11	376,310.00	175°		4.099	83.061
Bi-elliptic	191,344.11	376,310.00		503,873.00	3.904	593.919

Table 5-1 also illustrates that the shorter the path taken between two orbits, the shorter the time of flight. The Hohmann transfer is in the middle of the range of times between one-tangent and bi-elliptic burns. Notice also, in transferring to the Moon, the bi-elliptic saves Δv with an *extreme* penalty in time. Although not discussed yet, the time required to wait before a transfer to *rendezvous* with a satellite affects our choice of orbit-change technique. Sec. 5.6 will discuss rendezvous.

These results lead to the question of where we can most efficiently apply a change in velocity. Several choices exist. Consider the energy equation [Eq. (2-10)]. Because the energy depends directly on the orbit, we want the largest change in energy from a given change in velocity. Thus, if we add a specified change in velocity,

$$\xi_{new} = \frac{(v + \Delta v)^2}{2} - \frac{\mu}{r}$$

Let's expand this while separating the original energy:

$$\xi_{new} = \xi_{old} + v\Delta v + \frac{\Delta v^2}{2}$$

Assuming a fixed Δv , the free “variable” in the above equation is the velocity. You'll get the most additional energy, $\Delta\xi$, if you apply the Δv at the point in the orbit where the velocity is greatest. Thus, *to get the most efficient burn, do the maneuver as close to perigee as possible*. This may seem counter-intuitive. Energy change per Δv reflects efficiency. The largest change in energy of orbits for the smallest Δv maneuver is the most efficient burn. The bi-elliptic transfer would seem to use the distant transfer to reduce the change in velocity over the Hohmann transfer. In actuality, for cases in which the bi-elliptic is superior, the initial change in velocity is higher than for the equivalent Hohmann transfer; however, for the second and third burns, it's much less.

We don't always want to burn near perigee. Suppose we want to change the height at periapsis. The most efficient place to change the radius of periapsis is at apoapsis. Consider applying Δv at some other point in the orbit. In this case, both periapsis and apoapsis change. If we apply the Δv tangentially at apoapsis, all of the Δv will go into changing periapsis, and vice versa. For example, in an episode of *Star Trek, The Next Generation*, the crew of the *Starship Enterprise* tried to save a planet from an asteroidal moon which was projected to pass within 500 km of the surface on the next periapsis and destroy the planet. Unfortunately, their solution was to wait until *periapsis* and then apply the Δv (episode #161, “Deja Q”). Besides erroneously referring to the point of closest approach as “perigee,” they should have recognized that, although their Δv application increased apoapsis and the period of the moon, it would still return to the point of application (periapsis) on the next orbit and destroy the planet. This would happen only if the 500 km periapsis passage didn't result in an impact as they had predicted. The correct solution was to apply Δv immediately *before* the moon reached periapsis and destroyed the planet!

5.3.4 General Transfers

The most general transfer is one using two nontangential burns. This transfer employs any orbit type because there are no restrictions for flight-path angle or tangency requirements. To solve this problem in a simple sense is of little value because we need many formulas, figures to keep track of orbital orientations with respect to each other, and knowledge of so many parameters about the orbit. In practice, if you're trying to do this complex a maneuver, you must solve Lambert's problem (Sec. 6.7). You'll need vectors to do so, but by the time you set up a general transfer, you'll have enough information to form these vectors.

5.4 Noncoplanar Transfers

Coplanar maneuvers enable a satellite to alter two orbital elements: semimajor axis, and eccentricity. To do an intercept or rendezvous, we may need to change two other elements:

inclination, i , and the longitude of the ascending node, Ω . This requires applying Δv out of the orbit plane—a *noncoplanar change*. Because noncoplanar maneuvers change only two elements (i and Ω), there are three possibilities for maneuvers (i only, Ω only, and i and Ω). Applying an out-of-plane Δv at a nodal crossing will change only the inclination. An out-of-plane Δv applied at a certain point in an orbit will change only the longitude of ascending node. A Δv applied at *any other point* will change both inclination and longitude of ascending node. Once we've explored general noncoplanar maneuvers, we'll look at combined (coplanar and noncoplanar) maneuvers.

5.4.1 Introduction

We need noncoplanar maneuvers mainly because (1) a launch-site location restricts the initial orbit inclination for the satellite; (2) the direction of the launch can influence the amount of velocity the booster must supply, so certain orientations may be more desirable; and (3) timing constraints may dictate a *launch window* that isn't the best, from which we must make changes. Let's look briefly at each area.

Geometry for Launching a Satellite

The need for inclination-only maneuvers is apparent when considering the physical limitations of direct launches. For a *direct launch*—that is, a launch from the ground *directly* into the desired orbit—the launch-site latitude must be less than or equal to the desired inclination; otherwise, we must change the inclination in orbit. Countries that want equatorial orbits must accept this reality unless they can use launch sites on or close to the equator (the Ariane launch site at Kourou, for instance). Spherical geometry allows us to show this fact mathematically. Consider Fig. 5-8.

By examining the right spherical triangle, we can use Eq. (C-23) to develop an expression for the inclination:

$$\cos(i) = \cos(\phi_{gc}) \sin(\beta)$$

and then solve for the launch azimuth, β :

$$\sin(\beta) = \frac{\cos(i)}{\cos(\phi_{gc})} \quad (5-11)$$

Because $|\sin(\beta)| \leq 1$, the launch latitude, ϕ_{gc} , must be less than or equal to the desired inclination. If we don't meet this requirement, we can't launch the satellite *directly* into the desired orbit, so we may need to change the inclination. The sine law for spherical triangles can also demonstrate this result, and it introduces an extremely useful formula:

$$\sin(\phi_{gc}) = \sin(i) \sin(\omega + \nu) \quad (5-12)$$

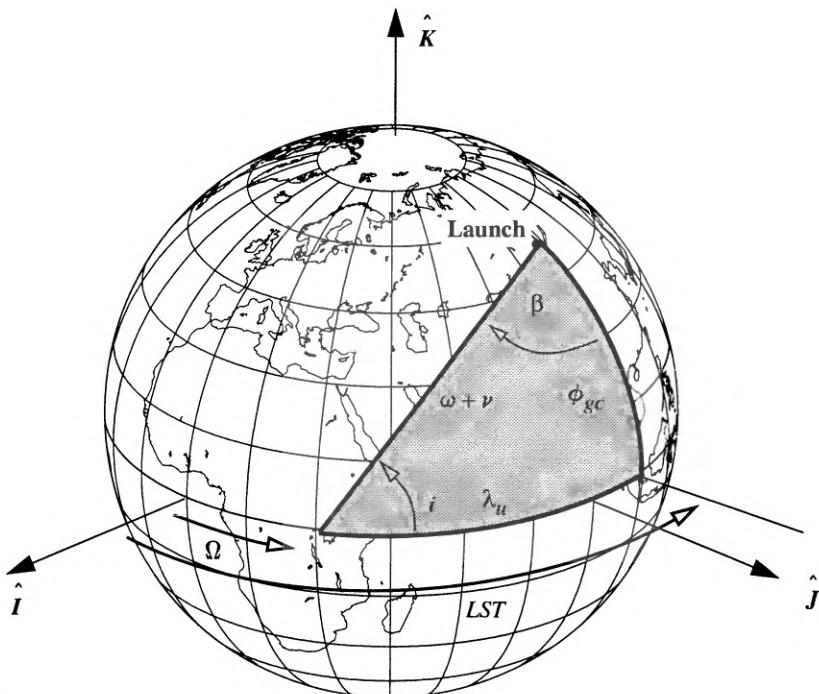


Figure 5-8. Launch-Window Geometry. Direct launches can occur only if the launch site's latitude is less than or equal to the satellite's inclination, i . Determining the launch window involves three basic steps: (1) finding the launch azimuth, β , from the site (I' ve shown an equivalent angle in the triangle because β is measured from north), (2) determining the auxiliary angle, λ_u , and (3) finding the GST (or LST) of the window.

Launch Location

The next factor affecting our need for noncoplanar changes is the launch site's location—and the direction of the launch. Of course, every launch site is on the Earth, so it has an initial velocity. At the equator, this initial velocity is highest; at the poles, it's zero. To determine the effect of the launch site's velocity on the velocity required for a booster, we can use simple angular-velocity relations. The total velocity of the launch site is

$$\vec{v}_L = \vec{\omega}_{\oplus} \times \vec{r}_{site} \quad \text{or} \quad v_L = |\vec{\omega}_{\oplus} \times \vec{r}_{site}| = \omega_{\oplus} r_{site} \cos(\phi_{gc}) \quad (5-13)$$

However, notice all the velocity is eastward if viewed in the SEZ coordinate system. This suggests that launching a satellite from the equator (where the radius is greatest) on a 90° azimuth will best support the mission. We can consider the Earth's effect small but appreciable. Using data from the table on the inside front cover, we find the velocity for a site at the equator is

$$v_L = \omega_{\oplus} r_{site} = 7.292\ 115\ 9 \times 10^{-5} (6,378,136.3) = 465.101\ 14 \text{ m/s} \quad (5-14)$$

For a westward launch from the same site on the equator, the booster must make up this velocity. It amounts to almost 0.9302 km/s, but remember, this example presumes a launch due east on the equator for comparison. Based on this discussion, I've shown several launch sites in Fig. 5-9, along with their recommended firing azimuths (due to safety considerations during launches). Notice how most sites are at low latitudes. Table 5-2 lists data for the launch sites.

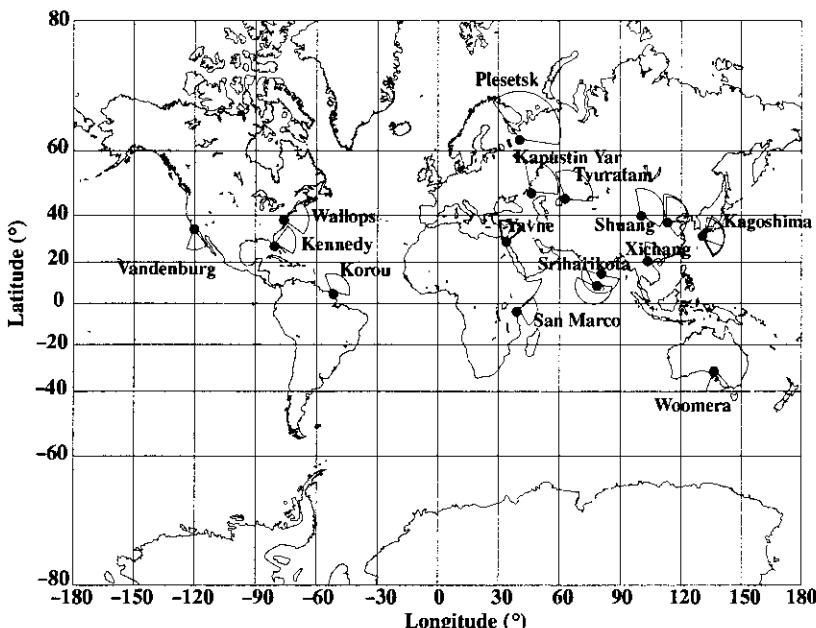


Figure 5-9. Locations for Selected Launch Sites. Direct launches are shown from various launch sites. Note the restrictions so boosters don't hit civilian areas and the relatively uniform low latitudes to take advantage of the Earth's rotation. Nominal 1000 km ranges are shown for each site.

Launch Windows

Finally, launch windows are extremely important to mission planning. A *launch window* is the time during which we may launch a satellite into a particular orbit and still keep desired mission parameters. Typically, mission planners select certain orbital characteristics. Then, they determine which launch sites offer a potential to launch and select one that will best place the satellite into the selected orbit. Basically, this involves selecting the *UT* for launch to achieve the desired orbit's initial nodal location. See Fig. 5-8 to visualize these relationships. From Eq. (5-11), we can determine the launch azimuth. Remember the

TABLE 5-2. Locations of Selected Launch Sites. Minimum and maximum azimuth values are usually determined so launches don't fly over populated areas. [Source: Larson and Wertz (1992, 679–681) and Chiulli (1994)]

Site	Latitude (°)	Longitude (°)	Azimuth Min (°)	Azimuth Max (°)
Vandenberg	34.600 000	-120.600 000	147	201
Cape Kennedy	28.500 000	-80.550 000	37	112
Wallops	37.850 000	-75.466 67	30	125
Kourou	5.200 000	-52.800 000	340	100
San Marco	-2.933 333	40.200 000	50	150
Plesetsk	62.800 000	40.600 000	330	90
Kapustin Yar	48.400 000	45.800 000	350	90
Tyuratam	45.600 000	63.400 000	340	90
Sriharikota	13.700 000	80.250 000	100	290
Shuang-Ch'Eng-Tzu	40.416 667	99.833 333	350	120
Xichang	28.250 000	102.200 000	94	105
Tai-yuan	37.766 667	112.500 000	90	190
Kagoshima	31.233 333	131.083 333	20	150
Woomera	-30.950 000	136.500 000	350	15
Yavne	31.516 667	34.450 000	350	120

sine will always give two answers (the other angle is $180^\circ - \beta$). For Eq. (5-11), the two angles represent ascending ($-90^\circ < u < 90^\circ$) and descending ($90^\circ < u < 270^\circ$) passes. For complete analysis, we should consider both cases. We find the auxiliary angle, λ_u , using the right spherical triangle shown in Fig. 5-8. Thus,

$$\cos(\lambda_u) = \frac{\cos(\beta)}{\sin(i)} \quad (5-15)$$

Two answers also result from this relation (the other angle is $360^\circ - \lambda_u$). Here, the two angles represent direct and retrograde orbits, respectively.

The main question still remaining is when to launch? To answer it, we determine the Greenwich sidereal time, θ_{GST} , when the angles are in the proper position. Notice that from Fig. 5-8, $\lambda_u = \theta_{LST} - \Omega$. For this step, we need the longitude of the launch site, λ , so we can use GST and

$$\theta_{GST} = \Omega + \lambda_u - \lambda$$

We know the latitude and inclination are fixed, and we usually have a tolerance for the longitude of the ascending node ($\pm\Delta\Omega$). This tolerance creates the launch window. Once we determine the range of values for GST , the final step is to find the time of day (UT) for launch. Solving Eq. (1-45) for UT gives us

$$UT = \frac{\theta_{GST} - \theta_{GST0}}{\omega_{\oplus}} \quad (5-16)$$

Recall that θ_{GST0} is the *GST* at 0 h on a particular day. As we substitute particular values of θ_{GST0} for each day into Eq. (5-16), we get a series of potential launch times. Any variability in Ω gives us a window on each day, within which we can achieve mission success.

This process can become quite involved, especially with the multiple angles for β and λ_u . The key to a correct solution is to use a consistent set. The type of orbit (direct or retrograde) determines the correct value of λ_u , so our only choice is whether we do the maneuver at the ascending or descending node. Choosing the node determines the correct value of β in Eq. (5-11). You must use this value in Eq. (5-15) for consistency.

For instance, proceed as follows to determine the correct *GST* at the descending node of a direct orbit. Find the launch azimuth using Eq. (5-11). Select the second angle ($180^\circ - \beta$) because you want the descending node. Next find the auxiliary angle from Eq. (5-15). The orbit is direct; therefore, use the first angle. Handle other cases similarly. Retrograde orbits can be tricky, so you should always draw a diagram to ensure you're using the proper angles.

5.4.2 Inclination-Only Changes

When we want *only* a change in inclination, we must change it at a node, i.e., an equator crossing. This is evident from Fig. 5-10. Notice the only two common points in the old and new orbits occur at the nodes.

Although most analysis focuses on circular orbits, elliptical orbits present a problem when considering the angle through which to change. We can determine this rotation angle in two basic ways: (1) rotate through the desired inclination change or (2) rotate through an auxiliary angle which we must determine based on orbital information. The simplest method is to use the desired inclination change, but that means using the normal component of the velocity vector. This is purely a consequence of the definition for inclination. If we don't use the normal component, we have to use an auxiliary angle. If the velocity vectors were available, a dot product easily gives us a solution:

$$\hat{v}_o \cdot \hat{v} = v_o v \cos(\Theta)$$

To find expressions for the velocity vectors, consider each vector in the *RSW* coordinate system. The initial vector consists only of the magnitude and flight-path angle:

$$\hat{v}_o = v_o \sin(\phi_{fpa}) \hat{R} + v_o \cos(\phi_{fpa}) \hat{S}$$

The second vector must use the same *RSW* system, but because there is an additional change in the inclination,

$$\hat{v} = v \sin(\phi_{fpa}) \hat{R} + v \cos(\phi_{fpa}) \cos(\Delta i) \hat{S} + v \cos(\phi_{fpa}) \sin(\Delta i) \hat{W}$$

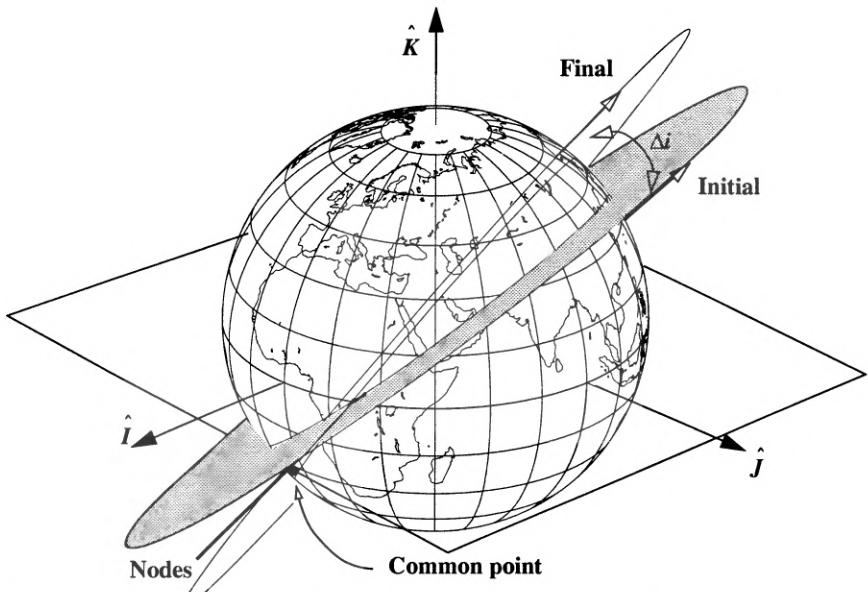


Figure 5-10. Inclination-Only Changes. An inclination-only change must occur at one of the nodal crossings because these are the only two points common to *both* orbits. The angular separation is equal to the change in inclination.

Use these velocity-vector expressions in the dot product and solve for the transfer angle:

$$\cos(\Theta) = \sin^2(\phi_{fpa}) + \cos^2(\phi_{fpa}) \cos(\Delta i) \quad (5-17)$$

The second method uses the fact that the satellite is at the same point in both orbits (true anomalies will be equal). Because the final orbit's size and shape are identical to those of the original, the magnitudes of the initial and final velocity are also equal. This is a consequence of examining a pure inclination change at a node. If the velocity's magnitude doesn't change, the velocity vectors must form an isosceles triangle, which makes the change in velocity simple:

$$\sin\left(\frac{\Delta i}{2}\right) = \frac{\Delta v_{i \text{ only}}}{2v_{initial} \cos(\phi_{fpa})}$$

Solving for the change in velocity gives us

$$\Delta v_{i \text{ only}} = 2v_{initial} \cos(\phi_{fpa}) \sin\left(\frac{\Delta i}{2}\right) \quad (5-18)$$

where we include the subscripts to remind ourselves of the initial values, even though they'll equal the final values. Analysis of Eq. (5-18) shows that, the smaller the value of

$v_{initial}$, the less Δv we need to change planes. To change only inclination, select the node where $v_{initial}$ is less to minimize $\Delta v_{i\ only}$. Recognize that plane changes are expensive. For circular orbits, if $\Delta i = 60^\circ$, $\Delta v_{i\ only} = v_{initial}$! Also note that if you use the auxiliary angle in place of the desired inclination change, the required change in velocity is the same, and

$$\Delta v_{i\ only} = 2v_{initial} \sin\left(\frac{\Theta}{2}\right) \quad (5-19)$$

Comparing Eq. (5-18) and Eq. (5-19), we see that

$$\sin\left(\frac{\Theta}{2}\right) = \sin\left(\frac{\Delta i}{2}\right) \cos(\phi_{fpa})$$

The inclination-only maneuver is a very simple operation. If the change in velocity is important, be sure to check *both* nodes and use the node with the smaller initial velocity. For circular orbits, there will no difference. The flight-path angle is necessary for elliptical orbits.

ALGORITHM 36: Inclination Only (Δi , ϕ_{fpa} , $v_{initial} \Rightarrow \Delta v_{i\ only}$)

$$\Delta v_{i\ only} = 2v_{initial} \cos(\phi_{fpa}) \sin\left(\frac{\Delta i}{2}\right)$$

The following example illustrates a simple inclination change.

▼ **Example 5-4. Calculating Velocity Change to Change Inclination Only.**

GIVEN: $\Delta i = 15.0^\circ$, $v_{initial} = 5.892\ 311$ km/s, or $5.892\ 311 / 7.905\ 365 = 0.745\ 356$ ER/TU,
 $e_{initial} = 0.0$

FIND: $\Delta v_{i\ only}$

Because the orbit is circular, ϕ_{fpa} is zero, so we can solve directly:

$$\Delta v_{i\ only} = 2v_{initial} \cos(\phi_{fpa}) \sin\left(\frac{\Delta i}{2}\right) = 2(0.745\ 356)(1.0)\sin\left(\frac{15.0}{2}\right)$$

$$\Delta v_{i\ only} = 0.194\ 577\ 0 \text{ ER/TU} = 1.538\ 202\ 1 \text{ km/s}$$

Now suppose we have an elliptical orbit with the following parameters:

$$e = 0.3, p = 17,858.7836 \text{ km or } 17,858.7836 / 6378.137 = 2.8 \text{ ER}, \omega = 30^\circ, \nu = 330^\circ$$

Find the semimajor axis:

$$a = \frac{p}{1 - e^2} = \frac{2.8}{1 - 0.3^2} = 3.076\ 923 \text{ ER}$$

Then, find the position using the trajectory equation:

$$r = \frac{p}{1 + e \cos(\nu)} = \frac{2.8}{1 + 0.3 \cos(330)} = 2.222\ 562 \text{ ER}$$

The velocity's magnitude is

$$v = \sqrt{\frac{2\mu}{r} - \frac{\mu}{a}} = \sqrt{\frac{2(1)}{2.222\ 56} - \frac{1}{3.0769}} = 0.758\ 197 \frac{\text{ER}}{\text{TU}}$$

Finally, determine the flight-path angle:

$$\tan(\phi_{fpa}) = \frac{e \sin(\nu)}{1 + e \cos(\nu)} = \frac{0.3 \sin(330)}{1 + 0.3 \cos(330)}$$

$$\phi_{fpa} = -6.7900^\circ$$

This makes sense because the satellite is heading towards perigee—thus, the negative sign. The change in velocity is now

$$\Delta v_{i\ only} = 2v_{initial} \cos(\phi_{fpa}) \sin\left(\frac{\Delta i}{2}\right) = 2(0.758\ 196\ 9) \cos(-6.7900) \sin\left(\frac{15.0}{2}\right)$$

$$\Delta v_{i\ only} = 0.196\ 540\ 8 \text{ ER/TU} = 1.553\ 72 \text{ km/s}$$

Notice that, for elliptical cases, we must check both nodes. The true anomaly for the other node is $\nu = 150^\circ$ ($330^\circ - 180^\circ$). Using the trajectory equation and the velocity equations, find

$$r = \frac{p}{1 + e \cos(\nu)} = \frac{2.8}{1 + 0.3 \cos(150^\circ)} = 3.7828 \text{ ER}$$

$$v = \sqrt{\frac{2\mu}{r} - \frac{\mu}{a}} = \sqrt{\frac{2(1)}{3.7828} - \frac{1}{3.0769}} = 0.451\ 341\ 2 \frac{\text{ER}}{\text{TU}}$$

The flight-path angle is

$$\phi_{fpa} = 11.4558^\circ$$

The positive value is correct because we're headed towards apogee. The velocity change is now

$$\Delta v_{i\ only} = 0.115\ 476\ 4 \text{ ER/TU} = 0.912\ 883\ 5 \text{ km/s}$$

▲ Notice the reduced overall velocity. It's important to check both nodes for elliptical orbits.

5.4.3 Changes in the Longitude of Ascending Node

As discussed earlier, the inclination required for the first part of a space mission can be achieved at launch without waiting for a launch window if that inclination is greater than the launch site's latitude. But we need a launch window to match up the line of nodes. If we have to wait too long for such a window, we can launch the satellite sooner and then apply thrust to change the longitude of ascending node. This is just one possible case which may require a change in node.

We can only do this one-burn analysis for either a circular or an elliptical-polar orbit. In general, elliptical orbits require multiple burns to complete just a change in the node without affecting other orbital elements. For elliptical orbits, a change in the node affects the argument of perigee because it's measured from the node. To avoid this complication, we'll examine circular orbits only.

For circular (or very nearly circular orbits), consider Fig. 5-11. There are only two possible common points of intersection which, unfortunately, are at neither the maximum

(minimum) latitude nor a node. Applying a change in velocity at any point other than the common points will change both inclination and the longitude of ascending node (Sec. 5.4.4). The point of intersection usually occurs *near* the maximum (minimum) latitude and slightly to the east (for direct orbits) for a positive change in Ω .

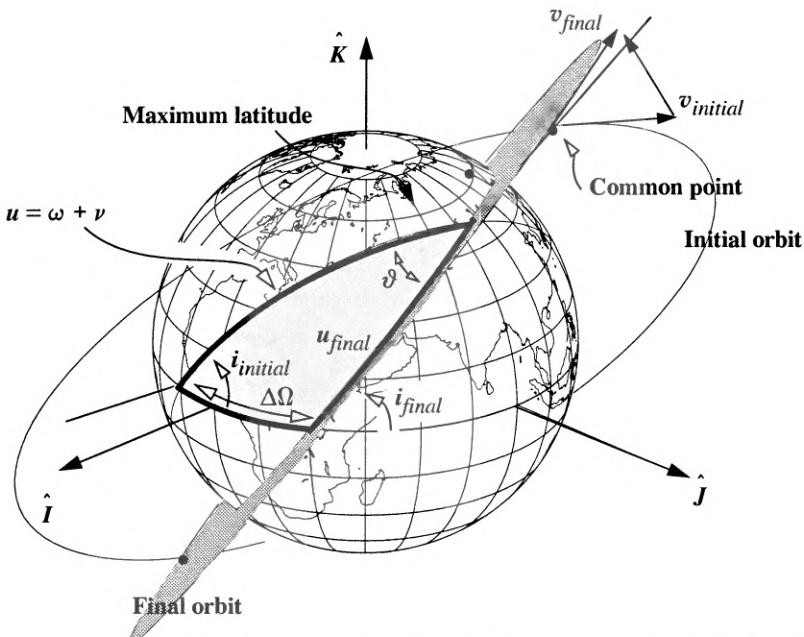


Figure 5-11. Effect of Changing the Longitude of Ascending Node. Circular orbits will always have two common points where the orbit paths cross each other. Because these points occur at different true-anomaly values in an elliptical orbit, we must analyze them separately.

To solve the problem we must determine several quantities: the location ($u = \omega + \nu$) of the common point for the burn; the angle, ϑ , through which we must rotate the orbit plane; the velocity at the new orbit; and the required Δv . Fig. 5-11 shows the spherical triangle used for solution.

Using spherical trigonometry, we can develop a cotangent expression to solve for the location of the burn, the argument of latitude, u :

$$\text{COT}(u_{\text{initial}}) \sin(\Delta\Omega) = -\sin(i_{\text{initial}}) \cot(i_{\text{final}}) + \cos(i_{\text{initial}}) \cos(\Delta\Omega)$$

Unfortunately, computer solutions may have difficulty evaluating the quadrants for this expression. Because the range of values for the argument of latitude is between 0° and $\pm 180^\circ$, we want to use a cosine relation. Eq. (C-26) allows us to find expressions for the

first and last values of argument of latitude. Substituting known values for the spherical triangle gives us the initial argument of latitude:

$$\cos(u_{initial}) \sin(\vartheta) \cos(i_{initial}) = \sin(i_{final}) \cos(\Delta\Omega) - \cos(\vartheta) \sin(i_{initial})$$

Simplifying, because the inclinations are equal, we get

$$\cos(u_{initial}) = \tan(i_{initial}) \frac{\cos(\Delta\Omega) - \cos(\vartheta)}{\sin(\vartheta)} \quad (5-20)$$

We may need the final argument of latitude for certain applications. Let's find it by using spherical-trigonometry formulas on the same spherical triangle:

$$\cos(u_{final}) = \frac{\cos(i_{initial}) \sin(i_{final}) - \sin(i_{initial}) \cos(i_{final}) \cos(\Delta\Omega)}{\sin(\vartheta)}$$

For equal inclinations, this expression simplifies to

$$\cos(u_{final}) = \cos(i_{initial}) \sin(i_{initial}) \left(\frac{1 - \cos(\Delta\Omega)}{\sin(\vartheta)} \right) \quad (5-21)$$

To determine the burn angle ϑ , use Eq. (C-24). With the values from the triangle,

$$\cos(\vartheta) = \cos(i_{initial}) \cos(i_{final}) + \sin(i_{initial}) \sin(i_{final}) \cos(\Delta\Omega)$$

Simplifying the expression, because the inclinations are equal, gives us

$$\cos(\vartheta) = \cos^2(i_{initial}) + \sin^2(i_{initial}) \cos(\Delta\Omega) \quad (5-22)$$

Although the sine law is equally simple, the range of values for ϑ is within $\pm 180^\circ$. Using the cosine formula resolves quadrant problems over the possible range of values. Remember that two answers result from an arccosine evaluation. The second answer represents the other common point in the orbit, occurring in the Southern Hemisphere in Fig. 5-11.

Because the orbit is circular, the velocity is equal at all points ($v_{initial}$ equals v_{final}), and the ϕ_{fpa} is always zero. Therefore, we can find the velocity change by using Eq. (5-18) but replacing Δi with ϑ and eliminating ϕ_{fpa} . Thus,

$$\Delta v_{\Omega \text{ only}_{cir}} = 2v_{initial} \sin\left(\frac{\vartheta}{2}\right) \quad (5-23)$$

Analyzing an elliptical orbit begins the same as for the circular one; we still have an intersection of the old and new orbits at the burn location. Although the altitude is common, the true anomalies aren't, because the argument of perigee is different. The change in ω is caused by the nodal change. Remember that in Fig. 5-11, $u_{initial} \neq u_{final}$. We won't develop options to solve this problem because it's much more appropriate to solve the

Lambert problem in order to find the actual velocity required for the maneuver. I'll show the algorithm and an example for circular orbits only.

ALGORITHM 37: Change in Longitude of Ascending Node—Circular

$(\Delta\Omega, i_{initial}, v_{initial} \Rightarrow \Delta v_{\Omega \text{ only}})$

$$\cos(\vartheta) = \cos^2(i_{initial}) + \sin^2(i_{initial}) \cos(\Delta\Omega)$$

$$\Delta v_{\Omega \text{ only}_{cir}} = 2v_{initial} \sin\left(\frac{\vartheta}{2}\right)$$

An example will illustrate changing the longitude of ascending node for a circular orbit.

▼ **Example 5-5. Finding Velocity Change for Changes in the Longitude of Ascending Node Only.**

GIVEN: $i = 55.0^\circ$, $\Delta\Omega = 45.0^\circ$, $v_{initial} = 5.892\ 311 \text{ km/s} = 5.892\ 311 / 7.905\ 365$
 $= 0.745\ 356 \text{ ER/TU}$, $e = 0.0$

FIND: $\Delta v_{\Omega \text{ only}}$

First, determine the angle through which the velocity is rotated:

$$\cos(\vartheta) = \cos^2(i) + \sin^2(i) \cos(\Delta\Omega) = \cos^2(55^\circ) + \sin^2(55^\circ) \cos(45^\circ)$$

$$\vartheta = 36.537\ 66^\circ$$

Immediately find the change in velocity:

$$\Delta v_{\Omega \text{ only}} = 2v_{initial} \sin\left(\frac{\vartheta}{2}\right) = 2(0.745\ 356) \sin\left(\frac{36.537\ 66}{2}\right)$$

$$\Delta v_{\Omega \text{ only}} = 0.467\ 302\ 3 \text{ ER/TU} = 3.694\ 195 \text{ km/s}$$

If you require the argument of latitude for the common points, use Eq. (5-20) and Eq. (5-21).

▲ $u_{initial} = 103.3647^\circ$ and $u_{final} = 76.6353^\circ$.

5.4.4 Changes to Inclination and Longitude of Ascending Node

At first glance, this change may appear to be similar to the preceding case, but the position of the firing in the orbit differs. If we want to change inclination, i , and node, Ω , the process is more complicated. Again, we'll look only at circular orbits. Discussion of elliptical orbits is complicated by the fact that we really need two burns to get the desired argument of perigee.

Using spherical trigonometry, we can develop expressions for ϑ and i whenever a Δv applies at the intersection of the two orbits, as Fig. 5-12 shows. Algorithm 38 captures two such expressions.

For circular orbits, solution of the main triangle again uses spherical geometry. To solve for the location of the burn, use the cosine law [spherical trigonometry Eq. (C-24)] so you can avoid quadrant checks. Substituting in known values for the triangle, find

$$\cos(u_{initial}) = \frac{\sin(i_{final}) \cos(\Delta\Omega) - \cos(\vartheta) \sin(i_{initial})}{\sin(\vartheta) \cos(i_{initial})} \quad (5-24)$$

You can't simplify this equation, as we did for changes in the longitude of ascending node, because the inclination change is predetermined. Find the final orbit argument of latitude from the development of Eq. (5-21), also with no simplification:

$$\cos(u_{final}) = \frac{\cos(i_{initial}) \sin(i_{final}) - \sin(i_{initial}) \cos(i_{final}) \cos(\Delta\Omega)}{\sin(\vartheta)} \quad (5-25)$$

The separation angle is the angle between the two orbits, which you can find developed in Eq. (5-22). However, the inclinations are *not* the same, so don't simplify the equation.

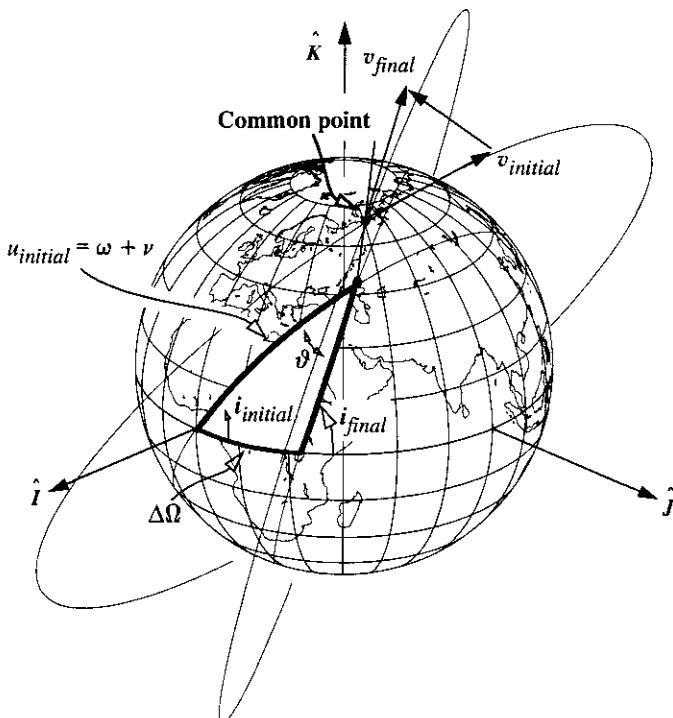


Figure 5-12. Geometry for General Noncoplanar Change. This type of change requires a general solution to find the common point of the two orbits. Note that both the inclination and ascending node are changing.

The Δv will be the same as for the node change discussed previously [Eq. (5-23)]. As before, elliptical orbits cause difficulties because you must correct the argument of perigee. In these cases, it's better to use alternate methods (Lambert, Sec. 6.7).

ALGORITHM 38: Combined Changes to i and Ω —Circular

$$(i_{initial}, i_{final}, \Delta\Omega, v_{initial} \Rightarrow \Delta v_i \text{ and } \Omega)$$

$$\cos(\vartheta) = \cos(i_{initial}) \cos(i_{final}) + \sin(i_{initial}) \sin(i_{final}) \cos(\Delta\Omega)$$

$$\Delta v_{i \text{ and } \Omega} = 2v_{initial} \sin\left(\frac{\vartheta}{2}\right)$$

An example illustrates the process.

▼ Example 5-6. Finding Velocity Changes for a Combined Change to i and Ω .

GIVEN: $i_{initial} = 55^\circ$, $i_{final} = 40^\circ$, $\Delta\Omega = 45^\circ$, $v_{initial} = 5.892\ 311 \text{ km/s} = 5.892\ 311 / 7.905\ 365$
 $= 0.745\ 356 \text{ ER/TU}$, $e = 0.0$

FIND: $\Delta v_i \text{ and } \Omega$

Begin by finding the angle between the velocity vectors of the initial and final orbits:

$$\begin{aligned} \cos(\vartheta) &= \cos(i_{initial}) \cos(i_{final}) + \sin(i_{initial}) \sin(i_{final}) \cos(\Delta\Omega) \\ &= \cos(55^\circ) \cos(40^\circ) + \sin(55^\circ) \sin(40^\circ) \cos(45^\circ) \end{aligned}$$

$$\vartheta = 35.737\ 09^\circ$$

The result is

$$\Delta v_{i \text{ and } \Omega} = 2(0.745\ 356)(1.0) \sin\left(\frac{35.737\ 09}{2}\right) = 0.457\ 401\ 4 \text{ ER/TU} = 3.615\ 925 \text{ km/s}$$

▲ The location of the burn (argument of latitude) is $u_{initial} = 128.9041^\circ$ and $u_{final} = 97.3803^\circ$.

5.5 Combined Maneuvers

Up to this point, we've isolated coplanar and noncoplanar maneuvers from each other. For real-world missions, we often need to change the orbit's inclination and its size and shape. If so, additional concerns arise. One approach could be to use each of the maneuvers discussed thus far in sequence. This would result in separate operations for the plane change and Hohmann transfer. Unfortunately, this particular combination is time-consuming and inefficient. Although multiple burns are usually necessary, they tend to increase complexity because our knowledge of the satellite's attitude is imprecise, and we must account for nonideal burn times, orientation, and thrust profiles. When possible, it's better to combine the transfers into a single maneuver. I'll show shortly that this process actually reduces the overall change in velocity requirements while achieving all the desired objectives. It also reduces the number of separate burns and decreases the time needed to complete the operation.

Notice this section *assumes* transfers which originate from *circular orbits* and occur *only* at the nodal crossing (remember the importance of nodal firings from the discussion

of noncoplanar maneuvers). To analyze more complex transfers, it's better to use the *Lambert* algorithm, which we'll discuss in Sec. 6.7. Although we can also determine elliptical orbits through this analysis, added complications can obscure the result. Therefore, we'll solve Lambert's problem to address elliptical orbits.

Before outlining the solution, let's introduce a few real-world constraints. When we need to change an orbit's plane, we should optimize the inclination change during each of the two burns, based on characteristics of the initial and final orbits. Direct insertions into deep-space orbits (geosynchronous, for example) usually aren't practical with current technology and may not be realistic and economical for many years. Thus, we typically use transfer boosters to place the satellite into the correct orbit after establishing it in low-Earth orbit. The process is termed a *fixed-Δv* transfer if the boosters are solid rockets that provide a fixed amount of change in velocity. Figure 5-13 shows the situation, which we need to address before we try to solve a particular mission with changes in orbital inclination, size, and shape.

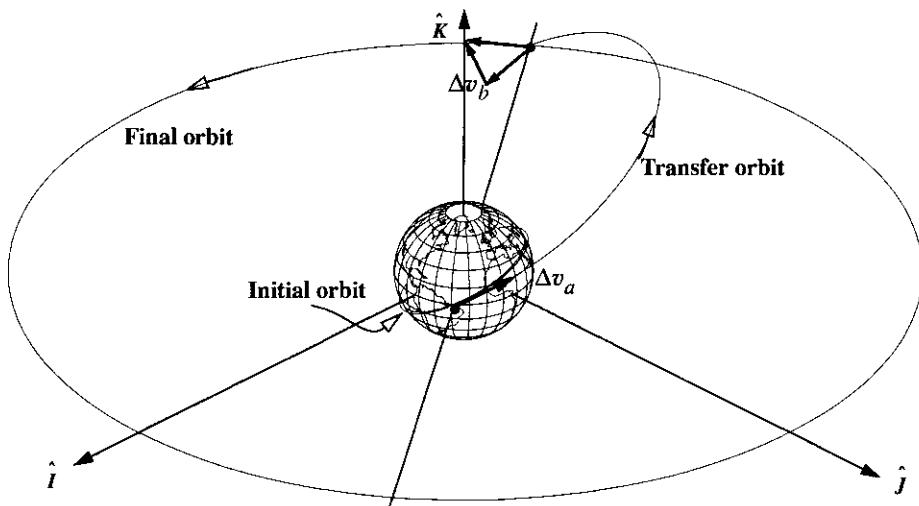


Figure 5-13. Geometry for a Combined Maneuver. A transfer from LEO to GEO is shown to the correct scale. The transfer orbit is just slightly below the initial 28.5° LEO orbit. Notice how close the LEO satellite appears to the Earth when it's drawn to scale.

5.5.1 Minimum-Inclination Maneuvers

From the previous sections, you should see that Hohmann transfers typically don't require a lot of change in velocity whereas even a small change in inclination takes a large Δv . Even if we combine the maneuvers, the plane change still drives the total change in velocity. Because we'll try to incorporate part of the plane change with each burn in the Hohmann-transfer maneuver, we must try to determine the *best* amount to change the inclination with each maneuver, so we keep the entire change in velocity low.

Start by examining the change in the velocity at each burn. Both changes in velocity will be combined maneuvers (changing inclination and altitude), so you need to use a form of Eq. (5-2). The major difference from Eq. (5-2) is that the angular separation will equal the change in inclination. Thus,

$$\Delta v = \sqrt{v_{trans}^2 + v_{initial}^2 - 2v_{trans}v_{initial}\cos(\Delta i)} \quad (5-26)$$

Each velocity contains the Hohmann transfer's contribution, and the change in inclination is the angle through which you rotate the velocity vector. The total change in velocity consists of two changes in velocity. Because you don't know the best values for the change in inclination at each location, identify each using a scaling term, s , that represents a percentage of the initial (or first) inclination change.

$$\begin{aligned} \Delta v &= \Delta v_a + \Delta v_b = \sqrt{v_{initial}^2 + v_{trans_a}^2 - 2v_{initial}v_{trans_a}\cos(\Delta i_{initial})} \\ &\quad + \sqrt{v_{final}^2 + v_{trans_b}^2 - 2v_{final}v_{trans_b}\cos(\Delta i_{final})} \end{aligned}$$

Taking the partial derivative of Δv with respect to s , you will find an expression with square roots in the denominator. Recognize these as Δv_a and Δv_b and solve for $\sin(s\Delta i)$.

$$\sin(s\Delta i) = \frac{\Delta v_a v_{final} v_{trans_b} \sin((1-s)\Delta i)}{\Delta v_b v_{initial} v_{trans_a}}$$

Because $\sin(s\Delta i)$ appears on both sides, you must iterate to find a solution.

An analytical approximation originally from Lisowski (1995) provides an estimate of the inclination change that's accurate to about 0.5° . Begin by squaring the two velocities to remove the square roots:

$$\begin{aligned} \Delta v_a^2 + \Delta v_b^2 &\approx v_{initial}^2 + v_{trans_a}^2 - 2v_{initial}v_{trans_a}\cos(s\Delta i) \\ &\quad v_{final}^2 + v_{trans_b}^2 - 2v_{final}v_{trans_b}\cos((1-s)\Delta i) \end{aligned}$$

Notice this approximation ignores the cross product terms ($2\Delta v_a \Delta v_b$). This allows you to differentiate this expression with respect to s and set the answer to zero:

$$\frac{\partial(\Delta v_a^2 + \Delta v_b^2)}{\partial s} \approx \Delta i 2v_{initial}v_{trans_a} \sin(s\Delta i) - \Delta i 2v_{final}v_{trans_b} \sin((1-s)\Delta i) = 0$$

Collecting terms and rearranging results in

$$\frac{\sin(s\Delta i)}{\cos(s\Delta i)} = \frac{v_{trans_b}v_{final}\sin(\Delta i)}{v_{initial}v_{trans_a} + v_{trans_b}v_{final}\cos(\Delta i)}$$

and finally,

$$s \approx \frac{1}{\Delta i} \tan^{-1} \left[\frac{\sin(\Delta i)}{\frac{v_{initial} v_{trans_a}}{v_{final} v_{trans_b}} + \cos(\Delta i)} \right]$$

You can modify this relation by replacing the velocity magnitudes with position magnitudes. The requirement for circular orbits makes this process much easier because radii and semimajor axis values are equal. Substituting appropriate velocity formulas from Eq. (2-21) and Eq. (2-23) results in

$$\frac{v_{initial} v_{trans_a}}{v_{final} v_{trans_b}} = \frac{\sqrt{\mu} \sqrt{\frac{2\mu}{r_{initial}} - \frac{\mu}{a_{trans}}}}{\sqrt{\frac{\mu}{r_{final}} \sqrt{\frac{2\mu}{r_{final}} - \frac{\mu}{a_{trans}}}}} = \frac{\sqrt{\mu} \sqrt{2\mu} \sqrt{\frac{r_{final}}{r_{initial}(r_{initial} + r_{final})}}}{\sqrt{\mu} \sqrt{2\mu} \sqrt{\frac{r_{initial}}{r_{final}(r_{initial} + r_{final})}}}$$

Using the ratio defined by Eq. (5-4), find that

$$\frac{v_{initial} v_{trans_a}}{v_{final} v_{trans_b}} = \sqrt{\frac{r_{final}^3}{r_{initial}^3}} = \sqrt{R^3}$$

You can now write Algorithm 39, below, but be careful with the units of s .

ALGORITHM 39: Minimum Combined Plane Change ($i_{initial}$, $r_{initial}$, $v_{initial}$,
 v_{trans_a} , v_{trans_b} , r_{final} , v_{final} , $\Delta i \Rightarrow \Delta i_{initial}$, Δi_{final})

$$\left\{ \begin{array}{l} \text{Iterate: } \sin(s\Delta i) = \frac{\Delta v_a v_{final} v_{trans_b} \sin((1-s)\Delta i)}{\Delta v_b v_{initial} v_{trans_a}} \\ \text{or Estimate: } R = \frac{r_{final}}{r_{initial}} \quad s \approx \frac{1}{\Delta i} \tan^{-1} \left[\frac{\sin(\Delta i)}{R^{3/2} + \cos(\Delta i)} \right] \end{array} \right.$$

$$\Delta i_{initial} = s\Delta i \quad \Delta i_{final} = (1-s)\Delta i$$

$$\Delta v_{initial} = \sqrt{v_{initial}^2 + v_{trans_a}^2 - 2v_{initial} v_{trans_a} \cos(\Delta i_{initial})} \quad \Delta v = \Delta v_{initial} + \Delta v_{final}$$

$$\Delta v_{final} = \sqrt{v_{final}^2 + v_{trans_b}^2 - 2v_{final} v_{trans_b} \cos(\Delta i_{final})}$$

Table 5-3 shows results for some characteristic orbits.

Knowing the inclination change we need for each maneuver lets us use a combined maneuver. But first, let's examine how to determine the amount of change in inclination.

TABLE 5-3. Minimum Plane Changes and Δv for Common Orbital Transfers. The first line shows the nonoptimal change in velocity with the entire plane change on the second burn. Notice how each case places about 5% to 10% of the total inclination change into the first maneuver. To reverse the changes (going from higher altitude to lower), simply reverse all quantities. The optimal solutions are from the iteration method.

$r_{initial}$	r_{final}	Δt°	Δi_a°	Δi_b°	Δv_{total} m/s
6671.53 km (1.046 ER)	42,163.95 km (6.6107 ER)	10.0	0.0	10.0	3945.42
		10.0	0.917	9.083	3940.90
6671.53 km (1.046 ER)	26,558.56 km (4.164 ER)	28.5	0.0	28.5	4121.20
		28.5	3.305	25.195	4058.97
6671.53 km (1.046 ER)	42,163.95 km (6.6107 ER)	45.0	0.0	45.0	4674.44
		45.0	2.751	42.249	4637.37

5.5.2 Fixed- Δv Maneuvering

We must consider fixed- Δv maneuvering when planning a satellite mission. Virtually every satellite launched into anything other than a low-Earth orbit must use some form of transfer booster to achieve the final orbit. The Space Shuttle routinely uses spin-stabilized upper stages to transfer payloads to geosynchronous equatorial orbits from the Shuttle's circular, inclined orbit. The maneuvers are very close to the combined maneuvers we're discussing in this section and resemble a Hohmann transfer. These upper stages have a *fixed- Δv* capacity—it can't change because most of the rockets used have solid propellant and can't be turned on and off. We can design solid-rocket motors to have one of a range of Δv capabilities before launch, but for our discussion, we'll consider them as having a *single*, fixed- Δv capability once they're on orbit.

The problem of how to use this Δv most effectively isn't difficult to solve, but it requires several subtasks. You need to know how to compute the firing attitude of the payload vehicle as it crosses a node and how much plane change you get after completing the first Δv . This is really a combined maneuver because the first burn completes part of the noncoplanar change in addition to changing the orbit's size and shape. Remember, you can optimize this value based on calculations from the previous section. The second burn at apogee completes the plane change and circularizes the final orbit. Usually, two motors do these burns because they're solid-rocket motors which can't be restarted.

We've already discussed how to compute the required velocities before and after the first burn of a transfer for a given change in inclination. Now, however, the Δv is a given value, and we want to determine the change in inclination, Δi . We can find the plane change by rearranging Eq. (5-26) for combined plane changes with the known parameters. Notice the inclination change replaces the difference in flight-path angle because we're still assuming circular orbits and an orbital transfer at the nodes:

$$\cos(\Delta i_1) = \frac{v_{initial}^2 + v_{trans_a}^2 - \Delta v^2}{2v_{initial}v_{trans_a}} \quad (5-27)$$

Compare this change in inclination with the *optimum* amount of an inclination change for the first burn ($\Delta i_{initial}$ in Algorithm 39). They usually aren't equal, but we have to use the actual data. The bottom line is that we must examine each case.

All that remains is finding the firing attitude to achieve the required Δi . Referencing Fig. 5-14, we see the **payload angle**, γ , is the angle between the initial velocity vector and the change in velocity vector. This is an exterior angle, as shown in the diagrams. When only magnitudes are known, we must use a series of sign corrections to determine the proper orientation. Rigorous analyses will use the vectors to determine the three-dimensional orientation of the firing.

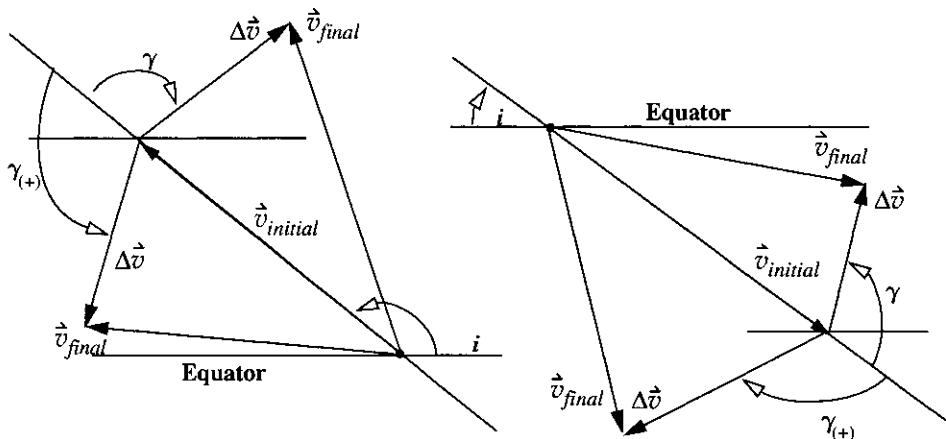


Figure 5-14. Geometry for Fixed- Δv Maneuvers. This figure includes changes to the retrograde ascending-node (left) and direct descending-node (right). I've shown two inclination changes (+ and -) for each case. Determine the proper sign for the payload angle, γ , based on the sign of the inclination change. [Positive γ values are shown with a (+).]

There are different geometries for ascending-node and descending-node firings, as well as for direct and retrograde orbits. The new inclination determines the sign. If the new inclination is larger than the original, γ is positive. Remember, this convention is arbitrary.

From the law of cosines, we can set up a relationship to find the angle γ :

$$\cos(180^\circ - \gamma) = \frac{v_{initial}^2 + \Delta v^2 - v_{final}^2}{2v_{initial}\Delta v}$$

and simplify to get

$$\cos(\gamma) = -\frac{v_{initial}^2 + \Delta v^2 - v_{final}^2}{2v_{initial}\Delta v}$$

The analysis for fixed- Δv maneuvers is often omitted in calculations because it depends on the mission, but the tools presented here allow you to decide whether a particular mission scenario will work.

ALGORITHM 40: *Fixed- Δv Maneuvers* ($v_{initial}$, v_{final} , v_{trans_a} , $\Delta v \Rightarrow \gamma, \Delta i_1$)

$$\cos(\gamma) = -\frac{v_{initial}^2 + \Delta v^2 - v_{final}^2}{2v_{initial}\Delta v}$$

$$\Delta i - \text{decreasing} \quad -180^\circ \leq \gamma \leq 0^\circ$$

$$\Delta i + \text{increasing} \quad 0^\circ < \gamma < 180^\circ$$

$$\cos(\Delta i_1) = \frac{v_{initial}^2 + v_{trans_a}^2 - \Delta v^2}{2v_{initial}v_{trans_a}}$$

Example 5-7 includes an example of fixed- Δv calculations.

Implementing Combined Maneuvers

Now let's explore the complete process to find the change in velocity for a combined maneuver.

1. Determine the best amount to change the inclination with each burn, or calculate the inclination change for a fixed Δv .
2. Compute the changes in velocities and the total change in velocity for the maneuver.

Again, remember we're considering only *circular* orbits. We can analyze elliptical orbits, but the process quickly becomes so specific that any benefit vanishes. It's better simply to solve Lambert's problem (Sec. 6.7). Example 5-7 shows the complete process.

▼ Example 5-7. Performing a Combined Maneuver.

GIVEN: $i_{initial} = 28.5^\circ$, $i_{final} = 0^\circ$, $e_{initial} = e_{final} = 0.0$, $alt_{initial} = 191$ km, $alt_{final} = 35,780$ km

FIND: $\Delta i_{initial}$, Δv_a , Δv_b , τ_{trans}

Find the radii for the orbits along with the transfer semimajor axis:

$$r_{initial} = \frac{6378.1363 + 191}{6378.1363} = 1.029\,946 \text{ ER}$$

$$r_{final} = \frac{6378.1363 + 35,780}{6378.1363} = 6.609\ 909 \text{ ER}$$

$$a_{trans} = \frac{r_{initial} + r_{final}}{2} = \frac{1.029\ 94 + 6.609\ 909}{2} = 3.819\ 927 \text{ ER}$$

First, determine the velocities:

$$v_{initial} = \sqrt{\frac{\mu}{r_{initial}}} = \sqrt{\frac{1}{1.0299}} = 0.985\ 377\ 11 \text{ ER/TU}$$

$$v_{final} = \sqrt{\frac{\mu}{r_{final}}} = \sqrt{\frac{1}{6.6099}} = 0.388\ 957\ 86 \text{ ER/TU}$$

$$v_{trans_a} = \sqrt{\frac{2\mu}{r_{initial}} - \frac{\mu}{a_{trans}}} = \sqrt{\frac{2(1)}{1.0299} - \frac{1}{3.819\ 92}} = 1.296\ 206\ 2 \text{ ER/TU}$$

$$v_{trans_b} = \sqrt{\frac{2\mu}{r_{final}} - \frac{\mu}{a_{trans}}} = \sqrt{\frac{2(1)}{6.6099} - \frac{1}{3.819\ 92}} = 0.201\ 967\ 45 \text{ ER/TU}$$

Then, determine how much to change the inclination with the first burn with Algorithm 39. The ratio $R = 6.417\ 724$.

$$s = \frac{1}{\Delta i} \tan^{-1} \left(\frac{\sin(\Delta i)}{R^{3/2} + \cos(\Delta i)} \right) = \frac{1}{28.5} \tan^{-1} \left(\frac{\sin(28.5^\circ)}{6.417\ 724^{3/2} + \cos(28.5)} \right) = 0.055\ 962^\circ$$

$$\Delta i_1 = s\Delta i = (0.055\ 962)(28.5^\circ) = 1.594\ 920^\circ$$

$$\Delta i_2 = (1-s)\Delta i = (1-0.055\ 962)(28.5^\circ) = 26.905\ 080^\circ$$

Now, find the change in velocities:

$$\begin{aligned} \Delta v_a &= \sqrt{v_{initial}^2 + v_{trans_a}^2 - 2v_{initial}v_{trans_a} \cos(\Delta i_{initial})} = 0.314\ 057 \text{ ER/TU} \\ &= 2469.681 \text{ m/s} \end{aligned}$$

$$\begin{aligned} \Delta v_b &= \sqrt{v_{final}^2 + v_{trans_b}^2 - 2v_{final}v_{trans_b} \cos(\Delta i - \Delta i_{initial})} = 0.227\ 973 \text{ ER/TU} \\ &= 1802.209 \text{ m/s} \end{aligned}$$

The total change in velocity is 4271.8902 m/s. Compare this result with the available change in velocity supplied by a booster (see Jane's, 1994, for instance). Assume that the booster can provide the calculated change in velocity.

Finally, determine the orientation of the burns using Algorithm 40.

$$\cos(\gamma_1) = \frac{v_{initial}^2 + \Delta v^2 - v_{final}^2}{2v_{initial}\Delta v} = \frac{0.985\ 377^2 + 0.314\ 057^2 - 1.296\ 206^2}{2(0.985\ 377)0.314\ 057}$$

$$\cos(\gamma_2) = \frac{v_{initial}^2 + \Delta v^2 - v_{final}^2}{2v_{initial}\Delta v} = \frac{0.201\ 967^2 + 0.227\ 973^2 - 0.388\ 957^2}{2(0.201\ 987)0.227\ 973}$$

▲ $\gamma_1 = 9.4384^\circ$, and $\gamma_2 = 50.5439^\circ$

5.6 Circular Rendezvous

After determining the Δv required to do an orbit maneuver, we must next consider *when* the particular maneuver should occur. Here we'll analyze only circular initial and final orbits with tangential burns and explore both coplanar and noncoplanar transfers. In Sec. 6.7 we'll develop a more general approach using the Lambert problem. These techniques are necessary when considering real-world operations, such as placing satellites into a constellation. A communications firm is now considering placing nearly 800 satellites into orbit. It's highly unlikely that a company could afford 800 individual launch vehicles or that all 800 satellites would reside in *one* orbital plane. Phasing, combined with the noncoplanar maneuvers previously discussed, makes possible a staggered deployment of many satellites from a single launch vehicle.

This section often refers to *target* and *interceptor* spacecraft. The *target* is typically used as a reference point for times and angles, and it is the desired endpoint for the *interceptor* satellite. Finally, we use values for the semimajor axis rather than position magnitudes because all orbits are circular.

5.6.1 Circular Coplanar Phasing

Coplanar transfers consist of three basic types: trailing interceptors, leading interceptors, and interceptors in different orbits but the same plane. In the first two, the interceptor and the target are in the same orbit but are displaced ahead or behind by some angular amount. The third case involves two separate orbits.

The first coplanar case is one in which a target and interceptor are in the same circular orbit with the interceptor trailing the target by an angle, as in Fig. 5-15. A sign convention is important. We'll define a *phase angle*, ϑ , to be measured from the target *to* the interceptor, with the positive sense being in the direction of the target satellite's motion.

Notice we use the *phasing* term whenever the interceptor is only adjusting its position within the *same* orbit. Although this may seem to be a departure from the previous notations for *transfer* orbits, the two are distinct and necessary.

The goal is to find the Δv required and how large the transfer-phasing orbit is in order to intercept within a specified amount of time. We always use the target satellite's period because it's "fixed" in terms of the problem, whereas the interceptor is variable and may use any one of several orbits to rendezvous. Because the satellites are in the same orbit, the angular velocity of the interceptor orbit, ω_{int} , is the same as the target's angular velocity, ω_{tgt} , which you should also recognize as the mean motion, n . Thus,

$$\omega_{tgt} = \sqrt{\frac{\mu}{a_{tgt}^3}} = n \quad (5-28)$$

In the time the interceptor revolves once in the phasing orbit, the target completes one revolution *minus* the initial separation, or ϑ . But if we remember the sign convention for the phase angle, the formula is

$$\tau_{phase} = \Delta t = t_1 - t_0 = \frac{k_{tgt}(2\pi) + \vartheta}{\omega_{tgt}} \quad (5-29)$$

where multiple revolutions of the target satellite ($k_{tgt} = 0, 1, 2, \dots$) are used to obtain solutions.

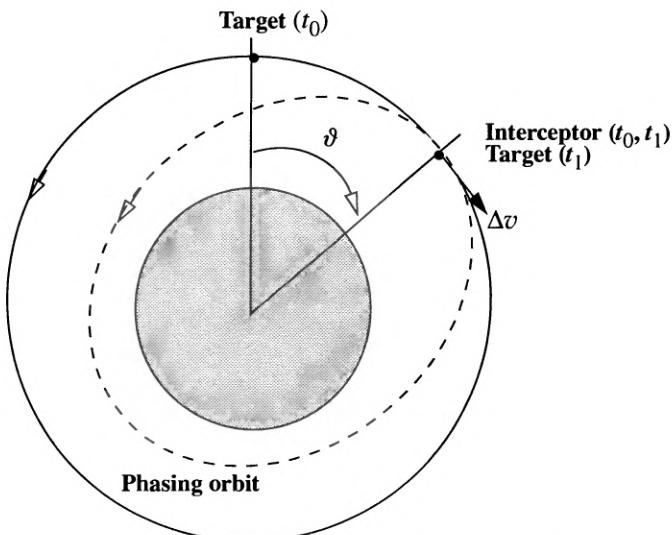


Figure 5-15. Circular Coplanar Rendezvous (Interceptor Trailing Target). Because the interceptor trails the target, the phase angle, ϑ , is negative. To rendezvous, the interceptor must move into a lower orbit with a shorter period to make up for the initial time displacement.

Because the phasing orbit consumes one of the interceptor satellite's periods as it accomplishes the rendezvous, we may solve the time-period equation [Eq. (2-16)] for the semimajor axis:

$$a_{phase} = \left(\mu \left(\frac{\tau_{phase}}{k_{int}(2\pi)} \right)^2 \right)^{1/3} \quad (5-30)$$

Notice a_{phase} must be large enough so the perigee radius is greater than the radius of the Earth. Multiple revolutions of the interceptor ($k_{int} = 0, 1, 2, \dots$) are included so we can adjust the phasing semimajor axis. Don't confuse this quantity with the target revolutions, k_{tgt} . Using the known radius at apogee on the transfer orbit, we find the perigee radius is

$$r_p = 2a_{phase} - r_a$$

The perigee radius must not be less than the Earth's radius, or the intercept will fail.

Now, determine the change in velocity for the maneuver. Because you're examining only circular orbits with tangential burns ($\phi_{fpa} = 0^\circ$), the change in velocity is simply the change in magnitudes of the different orbits. Thus for the first burn, the change is

$$\Delta v = v_{phase} - v_{int}$$

because the satellite must slow down to enter the phasing orbit. For the second burn, the change is $\Delta v = v_{int} - v_{phase}$ because the satellite must speed up to reenter the target orbit. Notice both changes in velocities have the same magnitude but different directions. For the solution in this section, you need only the magnitude. The total Δv is twice the change in velocity of the initial burn because your satellite must reenter the original orbit. Thus,

$$\Delta v = 2|v_{phase} - v_{int}| = 2 \left| \sqrt{\frac{2\mu}{a_{tgt}}} - \sqrt{\frac{\mu}{a_{phase}}} - \sqrt{\frac{\mu}{a_{tgt}}} \right| \quad (5-31)$$

The second coplanar case exists when a target and interceptor are in the same circular orbit, with the interceptor *leading* the target by a phase angle, as in Fig. 5-16. To rendezvous, the target needs to travel one complete orbit *plus* the phase angle, ϑ , in the time the interceptor completes one orbit on the phasing orbit. The time for the phase will be the same as in Eq. (5-29). You can find the semimajor axis of the phasing orbit using Eq. (5-30) and the change in velocity using Eq. (5-31).

If time permits, you can slightly improve Eq. (5-31). The initial parameters of the problem fix the Δv , except for the semimajor axis of the phasing orbit. If you adjust this axis, you can decrease overall Δv .

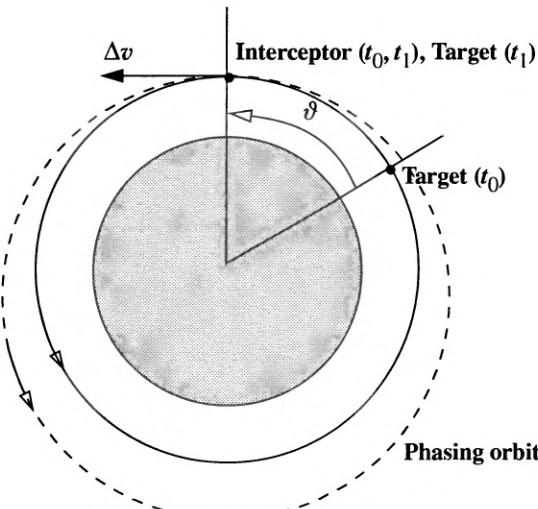


Figure 5-16. Circular Coplanar Rendezvous (Interceptor Leading Target). When the interceptor leads the target, the phase angle is positive, and the phasing orbit must be longer than the target orbit to permit the target satellite to *catch up*.

The basic idea is to adjust the phasing orbit as little as possible and make several revolutions in the phasing orbit until rendezvous occurs. Because the semimajor axis of the

phasing orbit depends on the available time, constraints on the phasing time drive this process. For instance, if time permits, you could attempt two phasing revolutions ($k_{int} = 2$).

In Fig. 5-16 then, if the target were twice as far back, the interceptor could use the same transfer shown; however, the interceptor would stay on the phasing orbit for two revolutions (instead of one) before the intercept. In the limit, as $t_1 - t_0$ approaches ∞ , Δv approaches 0, because the target and interceptor close a tiny amount on each revolution.

The last coplanar case involves a target and interceptor in separate, coplanar, circular orbits, as shown in Fig. 5-17. To minimize the change in velocity, the interceptor will use a Hohmann transfer to rendezvous with the target. Notice now that we're using the *transfer-orbit* term because the interceptor will move into a *different* orbit. If the satellites are in the positions shown, there would be no waiting because the target will travel to the rendezvous point in the same amount of time the interceptor takes to move from its current position to the rendezvous point. The geometry shown requires the interceptor to lead the target by some amount—the *lead angle*, α_L . The *initial phase angle*, ϑ_i , depends on the convention previously defined and may or may not equal the *phase angle*, ϑ , required to start the transfer. The time for the transfer orbit is already defined because we're using the Hohmann transfer (half the transfer period only):

$$\tau_{trans} = \pi \sqrt{\frac{a_{trans}^3}{\mu}}$$

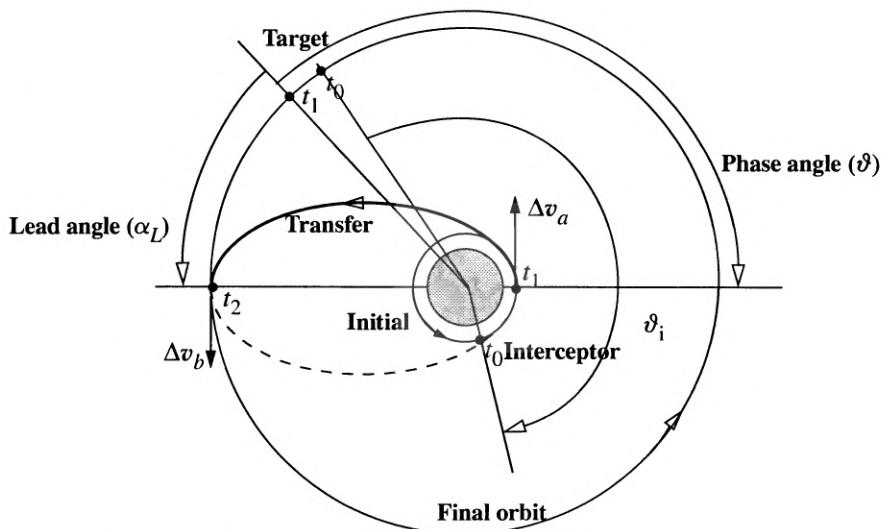


Figure 5-17. Circular Coplanar Rendezvous (Different Orbit). Whenever the satellites are in different orbits, we must calculate a lead angle (and therefore a phase angle). Also, the initial separation of the satellites (at t_0) usually differs from the required phasing for transfer, thus the symbol for the initial phase, ϑ_i . The final intercept occurs at t_2 .

Using the angular velocity of the target and the phasing time, find the lead angle by

$$\alpha_L = \omega_{tgt} \tau_{trans} \quad (5-32)$$

Then, find the required phase angle, ϑ , from the target to the interceptor:

$$\vartheta = \alpha_L - \pi \quad (5-33)$$

The **wait time**, τ_{wait} , or the time until the interceptor and target are in the correct positions, depends on the phase-angle separations (initial and final) and the relative velocities of the two satellites. Quantitatively,

$$\tau_{wait} = \frac{\vartheta - \vartheta_i + 2\pi k}{\omega_{int} - \omega_{tgt}} \quad (5-34)$$

where ϑ_i is the initial phase angle or the phase angle at the start of the problem, ω_{int} is the interceptor's angular velocity, and k is the time during which the relative geometry repeats. This time it's called the **synodic period**. We must include k for the revolutions because the gain in phase angle repeats with intervals of 360° . The quantity $(\vartheta_i - \vartheta)$ may be negative, say -60° . This means the phase angle has just occurred and will recur in 300° , 660° , etc.

Carefully examining Eq. (5-34) yields an interesting insight into the problem. Consider the case of two satellites in similar orbits. Because their average angular velocities are nearly the same, the denominator will be small, and the wait time may be very large. On the other hand, with orbits spaced far apart, the denominator is larger, and the wait time is shorter. Thus, if you want to try to rendezvous with another spacecraft when time is critical, you want your spacecraft to be in a different orbit as opposed to a similar orbit.

Algorithms 41 and 42 summarize these techniques, and a few examples illustrate the possibilities of changes to coplanar orbits.

ALGORITHM 41: Circular Coplanar Phasing (Same Orbits)

$$(a_{tgt}, \vartheta, k_{tgt}, k_{int} \Rightarrow \tau_{phase}, \Delta v, a_{phase})$$

$$\omega_{tgt} = \sqrt{\frac{\mu}{a_{tgt}^3}}$$

$$\tau_{phase} = \frac{2\pi k_{tgt} + \vartheta}{\omega_{tgt}}$$

$$a_{phase} = \left(\mu \left(\frac{\tau_{phase}}{2\pi k_{int}} \right)^2 \right)^{1/3}$$

$$\Delta v = 2 \left| \sqrt{\frac{2\mu}{a_{tgt}}} - \frac{\mu}{a_{phase}} - \sqrt{\frac{\mu}{a_{tgt}}} \right|$$

ALGORITHM 42: Circular Coplanar Phasing (Different Orbits)

$(\vartheta_i, a_{int}, a_{tgt}, k \Rightarrow \tau_{trans}, \Delta v, a_{trans})$

$$\omega_{tgt} = \sqrt{\frac{\mu}{a_{tgt}^3}} \quad \omega_{int} = \sqrt{\frac{\mu}{a_{int}^3}}$$

$$a_{trans} = \frac{a_{int} + a_{tgt}}{2}$$

$$\tau_{trans} = \pi \sqrt{\frac{a_{trans}^3}{\mu}}$$

$$\alpha_L = \omega_{tgt} \tau_{trans}$$

$$\vartheta = \alpha_L - \pi$$

$$\tau_{wait} = \frac{\vartheta - \vartheta_i + 2\pi k}{\omega_{int} - \omega_{tgt}}$$

$$\Delta v = \left| \sqrt{\frac{2\mu}{a_{int}}} - \frac{\mu}{a_{trans}} - \sqrt{\frac{\mu}{a_{int}}} \right| + \left| \sqrt{\frac{2\mu}{a_{tgt}}} - \frac{\mu}{a_{trans}} - \sqrt{\frac{\mu}{a_{tgt}}} \right|$$

▼ Example 5-8. Velocity Change for a Rendezvous in the Same Orbit.

GIVEN: $a_{int} = a_{tgt} = 12,756.274$ km, $\vartheta = 20^\circ$ (target behind interceptor)

FIND: Δv

Use Fig. 5-16 for the general geometry. First, convert the input radii to canonical units:

$$a_{int} = a_{tgt} = (12,756.274)/6378.1363 = 2.0 \text{ ER}$$

$$\omega_{tgt} = \sqrt{\frac{1}{2^3}} = 0.35355 \text{ rad/TU}$$

Let the target satellite complete one more orbit ($k_{tgt} = 1$) because the initial separation is small. Thus,

$$\tau_{phase} = \frac{2\pi k_{tgt} + \vartheta}{\omega_{int}} = \frac{2\pi + 20(\pi/180)}{0.35355} = 18.7588 \text{ TU}$$

The interceptor will complete one revolution on the transfer orbit ($k_{int} = 1$). This value should be larger than the original orbit for this example.

$$a_{phase} = \left(\mu \left(\frac{\tau_{phase}}{2\pi k_{int}} \right)^2 \right)^{1/3} = 2.0734 \text{ ER}$$

Next, find the initial change in velocity for the interceptor to enter the higher orbit:

$$\Delta v_{initial} = \left| \sqrt{\frac{2\mu}{a_{int}}} - \sqrt{\frac{\mu}{a_{phase}}} \right| = 0.719514 - 0.707107 = 0.01241 \text{ ER/TU}$$

The total Δv to rendezvous will be twice this amount, or

$$\Delta v = 2 (\Delta v_{initial}) = 0.02482 \text{ ER/TU} = 196.1790 \text{ m/s}$$

Now evaluate the solution for the case in which the target can complete two revolutions, $k_{tgt} = 2$. The phasing time is

$$\tau_{phase} = \frac{2\pi k_{tgt} + \vartheta}{\omega_{int}} = \frac{2(2\pi) + 20(\pi/180)}{0.35355} = 36.53037 \text{ TU}$$

With such a large increase in time, it's reasonable to let the interceptor revolve twice in the phasing orbit, $k_{int} = 2$. This allows you to find the semimajor axis and the change in velocity.

$$a_{phase} = \left(\mu \left(\frac{\tau_{phase}}{2\pi k_{int}} \right)^2 \right)^{1/3} = 2.0369 \text{ ER}$$

$$\Delta v_{initial} = 0.00637 \text{ ER/TU}$$

$$\Delta v = 2 (\Delta v_{initial}) = 0.01274 \text{ ER/TU} = 100.7251 \text{ m/s}$$

▲ Notice the large reduction in velocity when you use two phasing orbits.

▼ Example 5-9. Velocity Change for a Rendezvous from Different Orbit

GIVEN: $a_{int} = 12,756.274 \text{ km}$, $a_{tgt} = 42,164.17 \text{ km}$, $\vartheta_i = 20^\circ$ (target behind interceptor)

FIND: Δv

First convert the radii to canonical units:

$$a_{int} = 12,756.274 / 6378.1363 = 2 \text{ ER}, a_{tgt} = 42,164.17 / 6378.1363 = 6.61 \text{ ER}$$

Now find the mean motions, using the radii as the semimajor axis because the orbits are circular:

$$\omega_{tgt} = \sqrt{\frac{\mu}{a_{tgt}^3}} = \sqrt{\frac{1}{6.61^3}} = 0.058843 \text{ rad/TU}$$

$$\omega_{int} = \sqrt{\frac{\mu}{a_{int}^3}} = \sqrt{\frac{1}{2^3}} = 0.353553 \text{ rad/TU}$$

The semimajor axis of the phasing orbit is

$$a_{trans} = \frac{a_{int} + a_{tgt}}{2} = \frac{2 + 6.61}{2} = 4.305 \text{ ER}$$

The time required for the transfer is half the period of the Hohmann transfer, or

$$\tau_{trans} = \pi \sqrt{\frac{a_{trans}^3}{\mu}} = \pi \sqrt{\frac{4.305^3}{1}} = 28.06142 \text{ TU} = 377.338 \text{ min}$$

The lead angle the target satellite must have for the transfer to take place is

$$\alpha_L = \omega_{tgt} \tau_{trans} = 0.058\,843 (28.061\,42) = 1.651\,218\,1 \text{ rad} = 94.608\,48^\circ$$

The phasing angle for the target satellite, when it is in the required position, is

$$\vartheta = \alpha_L - \pi = 94.608\,48^\circ - 180^\circ = -85.391\,52^\circ = -1.490\,374\,5$$

From the initial conditions, you know that the target satellite began 20° behind the interceptor, so you can determine the wait time until the transfer can start. Use a positive sign with the target satellite's starting location because it is *behind* the interceptor. Also, let $k = 1$ to get a positive wait time. Use radians for the correct units. The wait time,

$$\tau_{wait} = \frac{\vartheta - \vartheta_i + 2\pi k}{\omega_{int} - \omega_{tgt}} = \frac{-1.490\,374\,5 - 0.349\,065\,85 + 2\pi}{0.353\,553 - 0.058\,843} = 15.078\,37 \text{ TU} = 202.756\,59 \text{ min}$$

This is a relatively long time to wait for a rendezvous opportunity. Imagine if the two orbits had been much closer, and the denominator had thus been much smaller! Finally, determine the total change in velocity using the equations for the Hohmann transfer:

$$\Delta v = \left| \sqrt{\frac{2\mu}{a_{int}}} - \sqrt{\frac{\mu}{a_{int}}} + \sqrt{\frac{2\mu}{a_{tgt}}} - \sqrt{\frac{\mu}{a_{tgt}}} \right|$$

$$\blacktriangle = \left| \sqrt{\frac{2(1)}{2}} - \frac{1}{4.305} - \sqrt{\frac{1}{2}} \right| + \left| \sqrt{\frac{2(1)}{6.61}} - \frac{1}{4.305} - \sqrt{\frac{1}{6.61}} \right| = 0.292\,928\,9 \text{ ER/TU} = 2.315\,71 \text{ km/s}$$

5.6.2 Circular Noncoplanar Phasing

As with the coplanar cases, the noncoplanar cases presented here will assume *circular* orbits and transfer only with a Hohmann transfer. Let's examine the case in which the target and interceptor are in separate, noncoplanar, circular orbits, as shown in Fig. 5-18.

This is the same problem as the coplanar case, except we're restricted to starting the transfer at the relative node between the two orbital planes because we assume the use of a Hohmann transfer. If the final orbit is equatorial, we can start at a node of the initial orbit. If the transfer doesn't start along a line formed by $\hat{h}_{tgt} \times \hat{h}_{int}$, we have to make several unnecessary intermediate plane changes. Unfortunately, it's not always timely to wait for the natural motion of the satellites to place the interceptor at the node and the target at the correct phase angle (or lead angle) necessary for the non-coplanar intercept. Moreover, it may be days and even weeks before the geometry is correct. There are two solutions. First, we can use phasing orbits to help place the interceptor at a node at the proper time. This method is very realistic because we can change the velocity to enter the phasing orbit so it becomes part of the total change in velocity. Thus, no velocity is wasted. The second method involves solving Lambert's problem. This may not be as efficient because a change from an arbitrary location in the orbit may involve unnecessary plane changes. In either case the problem is important to the mission planner, so we should explore all options when time and resources permit.

As suggested, one way to avoid the potentially long wait is to use an intermediate phasing orbit, as in Fig. 5-19. The period of the phasing orbit is precisely designed so the interceptor travels from time t_p and arrives back at the node at the same time the phase

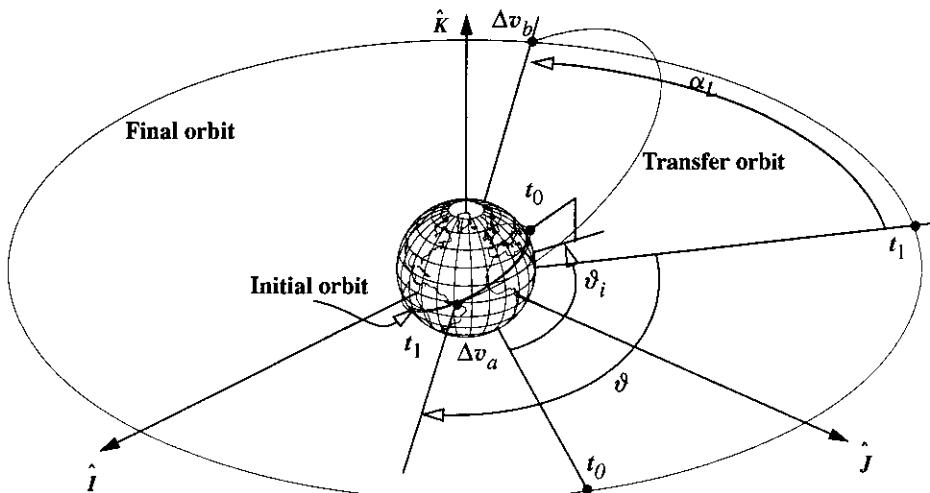


Figure 5-18. Geometry for a Circular, Noncoplanar, Orbit Transfer. This transfer combines the Hohmann transfer from LEO to GEO with the change in inclination plane. The maneuver can't start until the interceptor is at a node, t_1 , and the target, t_1 , is at the required lead angle. Because both satellites begin at t_0 , this event may require an extremely long wait time. Notice the convention used in measuring angles *from* the target satellite. When drawn to scale, the LEO orbit appears almost on the Earth.

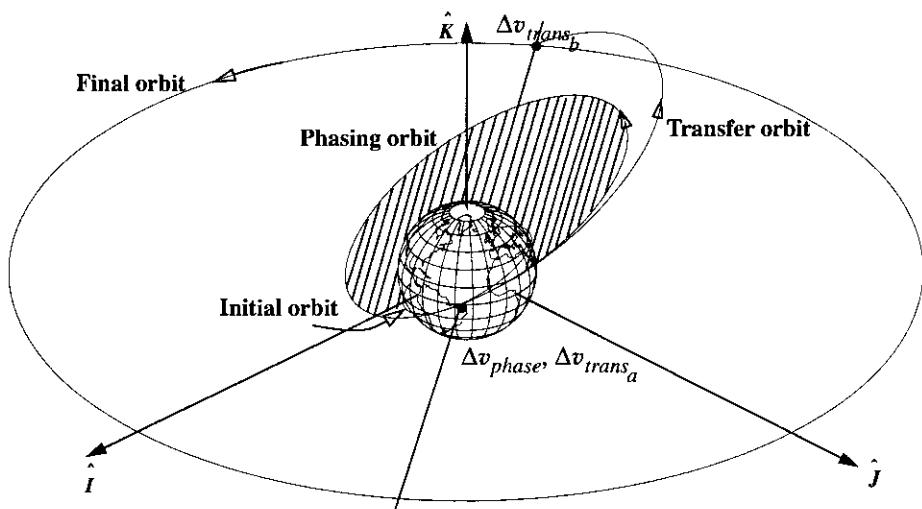


Figure 5-19. Alternate Geometry for a Circular, Noncoplanar, Orbit Transfer. An intermediate phasing orbit is useful to permit a quicker time to rendezvous. The size of the phasing orbit is calculated so the satellite returns to the node in time for the transfer into the final orbit. Both burns (Δv_{phase} and Δv_{trans_a}) occur at the same point.

angle is correct to start the Hohmann-transfer orbit. But we must make sure the phasing orbit doesn't hit the Earth!

The phasing orbit can be in the plane of the initial orbit or in any other plane. Because the most efficient location to do a plane change is at apogee, the phasing orbit could stay in the plane of the initial orbit if the transfer occurs as shown in Fig. 5-19. But remember our discussion of the best amount of plane change when doing combined maneuvers. With a tremendous number of possibilities available, actually selecting one orbit over another will depend on the specific problem.

Initial solution of the problem recasts the first few calculations from the coplanar case. First, determine the transfer orbit's semimajor axis and related coefficients. Find the semi-major axis of the transfer orbit using

$$a_{trans} = \frac{a_{initial} + a_{final}}{2}$$

The time of flight for the transfer is half the orbital period, or

$$\tau_{trans} = \pi \sqrt{\frac{a_{trans}^3}{\mu}}$$

and you can now find the lead angle, α_L , using Eq. (5-32).

At this point, the problem diverges from previous cases. The next step is to determine how long it will take for the interceptor to reach a node, after which you can determine the phasing orbit's characteristics. The initial position of the interceptor must be known and, because it's seldom at a node, you must find the amount of time until the interceptor is at a node. Although the initial phase angle, ϑ_i , may be known, it's somewhat irrelevant; you have to recalculate it after the amount of time has passed for the interceptor to reach a node. The phasing orbit must "fill up" the time it takes for the target to travel from this new position to the lead angle from which the combined maneuver can take place.

Find the angle, $\Delta\vartheta_{int}$, the interceptor must travel until reaching perigee of the transfer orbit by using the initial location data and determining the angle required until it reaches 0° true anomaly. Special orbits may require different orbital elements (u, λ_{true}) to achieve this requirement; the bottom line is that the satellite be in the Earth's equatorial plane if the final orbit is equatorial. This allows you to calculate the amount of time required:

$$\Delta t_{node} = \frac{\Delta\vartheta_{int}}{\omega_{int}} \quad (5-35)$$

Next, find the angular distance the target has traveled from its initial location and determine the new phase angle by adding the target's initial phase angle to the angular distance the target travels during the interceptor's trip to reach a node. Thus,

$$\lambda_{true_1} = \lambda_{true_0} + \omega_{tgt}\Delta t_{node}$$

Determine the new phase angle, ϑ_{new} , based on the new locations of both satellites at time t_1 . This often involves using Eq. (2-55) to find λ_{true} . Now the lead angle is

$$\alpha_{new} = \pi + \vartheta_{new}$$

The phasing orbit must account for the time between the lead angles, so

$$P_{phase} = \frac{\alpha_{new} - \alpha_L + 2\pi k_{tgt}}{\omega_{tgt}} \quad (5-36)$$

Notice the angle is now divided only by the target's angular velocity because the interceptor's orbit is undetermined at this point. Also, you don't use the phase angles, ϑ , because they change throughout the problem, and you want the interceptor to reach a *node* with the *proper* lead angle for starting the final Hohmann transfer. Knowing the phasing orbit's period, calculate the semimajor axis of the phasing orbit using Eq. (5-30).

After calculating a_{phase} , verify ($a_{int} < a_{phase} < a_{tgt}$) to ensure there is no wasted change in velocity. Because the phasing orbit is in the same plane as the initial orbit at the first firing, v_{int} , v_{phase} , and v_{trans} are all tangent and in the same direction. Therefore, the Δv required to achieve the phasing orbit is "free." Notice this is true only for the cotangential case.

Finally, let's calculate the delta velocities for each maneuver: (1) Δv_{phase} to enter the phasing orbit, (2) Δv_{trans1} to enter the transfer orbit, and (3) Δv_{trans2} to complete the transfer to the target orbit. All the Δv s are tangential, so you can use the magnitudes. The complete algorithm is involved but useful to show how you can splice many techniques to arrive at a solution.

ALGORITHM 43: *Noncoplanar Phasing* (ϑ_p , a_{int} , a_{tgt} , k_{tgt} , u_{int} , Ω_{int} , $\lambda_{true,tgt1}$, $\Delta i \Rightarrow \tau_{trans}$, τ_{phase} , Δv_{phase} , Δv_{trans1} , Δv_{trans2} , a_{phase})

$$\omega_{tgt} = \sqrt{\frac{\mu}{a_{tgt}^3}} \quad \omega_{int} = \sqrt{\frac{\mu}{a_{int}^3}}$$

$$a_{trans} = \frac{a_{initial} + a_{final}}{2}$$

$$\tau_{trans} = \pi \sqrt{\frac{a_{trans}^3}{\mu}}$$

$$\alpha_L = \omega_{tgt} \tau_{trans}$$

Find $\Delta\vartheta_{int}$ to reach a node (180° or $360^\circ - u_{int}$)

$$\Delta t_{node} = \frac{\Delta \vartheta_{int}}{\omega_{int}}$$

$$\lambda_{true_{tgt1}} = \lambda_{true_0} + \omega_{tgt} \Delta t_{node}$$

Find λ_{true} for interceptor at t_1 $\lambda_{true_{int1}} = \Omega + u_{int}$

$$\vartheta_{new} = \lambda_{trueintl} - \lambda_{truetgtl}$$

$$\alpha_{new} = \pi + \vartheta_{new}$$

$$\rho_{phase} = \frac{\alpha_{new} - \alpha_L + 2\pi k_{tgt}}{\omega_{tet}}$$

$$a_{phase} = \left(\mu \left(\frac{\rho_{phase}}{k_{int} 2\pi} \right)^2 \right)^{1/3}$$

$$\Delta v_{phase} = \left| \sqrt{\frac{2\mu}{a_{int}}} - \frac{\mu}{a_{phase}} - \sqrt{\frac{\mu}{a_{int}}} \right|$$

$$\Delta v_{trans_1} = \left| \sqrt{\frac{2\mu}{a_{int}} - \frac{\mu}{a_{trans}}} - \sqrt{\frac{2\mu}{a_{int}} - \frac{\mu}{a_{phase}}} \right|$$

$$\Delta v_{trans_2} = \sqrt{\left(\frac{2\mu}{a_{tgt}} - \frac{\mu}{a_{trans}}\right) + \left(\frac{\mu}{a_{tgt}}\right)} - 2\sqrt{\frac{2\mu}{a_{tgt}} - \frac{\mu}{a_{trans}}} \sqrt{\frac{\mu}{a_{tgt}}} \cos(\Delta i)$$

$$\tau_{total} = 2\pi \sqrt{\frac{3}{a_{phase}}} + \tau_{trans} + \Delta t_{node}$$

Because this problem is rather involved, let's look at an example to show the operation.

▼ Example 5-10. Velocity Change for a Rendezvous Using Noncoplanar Orbits.

Consider the following problem of two circular orbits, as illustrated in Fig. 5-20.

Interceptor	Target
$a = 7143.51 \text{ km} = 1.12 \text{ ER}$	$a = 42,159.4855 \text{ km} = 6.61 \text{ ER}$
$i = 28.5^\circ$	$i = 0$
$\Omega = 45^\circ$	$\lambda_{true} = 200^\circ$
$\mu = 15^\circ$	

First, find the desired lead angle, α_f , and some initial quantities:

$$\omega_{tgt} = \sqrt{\frac{\mu}{a_{tgt}^3}} = 0.058\ 843 \text{ rad/TU} \quad a_{trans} = \frac{1.12 + 6.61}{2} = 3.865 \text{ ER}$$

$$\omega_{int} = \sqrt{\frac{\mu}{a_{int}^3}} = 0.843\ 67 \text{ rad/TU} \quad \tau_{trans} = \pi \sqrt{\frac{a_{trans}^3}{\mu}} = 23.871 \text{ TU}$$

$$\alpha_L = \omega_{tgt} \tau_{trans} = 80.481^\circ$$

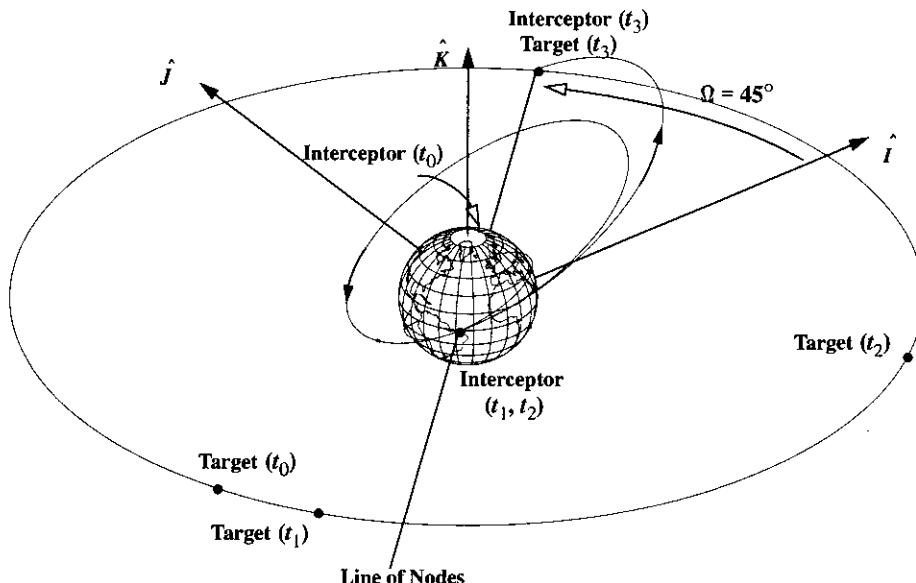


Figure 5-20. Geometry for the Noncoplanar Orbit Transfer in Example 5-10. The interceptor starts in a 28.5° inclined orbit, behind the Earth, whereas the final orbit is equatorial. The phasing orbit is designed to place the interceptor at the node at the proper time for transfer to the final orbit.

To find α_p , recall the transfer must occur at a node. Therefore, you must move the elements forward in time until the interceptor is at the descending (or ascending) node, an angle $\Delta\vartheta_{int}$ of 165° ($180^\circ - 15^\circ$). To determine the time for such a movement, use

$$\Delta t_{node} = \frac{\Delta\vartheta_{int}}{\omega_{int}} = \frac{165^\circ}{0.843\ 67} = 3.4134 \text{ TU}$$

Now, how far has the target moved in this length of time?

$$\omega_{tgt} \Delta t_{node} = (0.058\ 843\ 4) (3.4134) \left(\frac{180}{\pi} \right) = 11.5^\circ$$

This allows you to look at the phase angle when the interceptor is at the node ready to transfer to the target orbit. Therefore, at time t_1 , the target is at $\lambda_{true1} = 211.5^\circ$, and the interceptor is at $\alpha = 180^\circ$. From Fig. 5-20, the new phase angle is

$$\vartheta_{new} = (45^\circ + 180^\circ) - 211.5^\circ = 13.5^\circ$$

The new lead angle is

$$\alpha_{new} = \pi + \vartheta_{new} = 193.5^\circ$$

Determine the period of the phasing orbit, being sure to convert to radians and letting the transfer occur with no additional target revolutions ($k_{tgt} = 0$):

$$P_{phase} = \frac{\alpha_{new} - \alpha_L + 2\pi k_{tgt}}{\omega_{tgt}} = \frac{193.5^\circ - 80.48^\circ}{0.0588434} = 33.5224 \text{ TU}$$

If the interceptor revolves once in a phasing orbit ($k_{int} = 1$),

$$a_{phase} = \left(\mu \left(\frac{P_{phase}}{k_{int} 2\pi} \right)^2 \right)^{1/3} = 3.0533 \text{ ER}$$

Check $a_{int} < a_{phase} < a_{tgt}$. Finally, calculate the velocities required for each maneuver. The initial velocities you know are v_{int} and v_{tgt} :

$$v_{int} = \sqrt{\frac{\mu}{a_{int}}} = 0.94491 \text{ ER/TU} \quad \text{and} \quad v_{tgt} = \sqrt{\frac{\mu}{a_{tgt}}} = 0.38895 \text{ ER/TU}$$

Also, find v_{phase} , v_{trans1} , and v_{trans2} :

$$v_{phase} = \sqrt{\frac{2\mu}{a_{int} - a_{phase}}} = 1.207558 \text{ ER/TU}$$

$$v_{trans1} = \sqrt{\frac{2\mu}{a_{int} - a_{trans}}} = 1.235711 \text{ ER/TU}$$

$$v_{trans2} = \sqrt{\frac{2\mu}{a_{tgt} - a_{trans}}} = 0.209379 \text{ ER/TU}$$

Therefore, to find the Δv s, use

$$\Delta v_{phase} = |v_{phase} - v_{int}| = 0.262647 \text{ ER/TU}$$

$$\Delta v_{trans1} = |v_{trans1} - v_{phase}| = 0.028152 \text{ ER/TU}$$

$$\Delta v_{trans2}^2 = v_{trans2}^2 + v_{tgt}^2 - 2v_{trans2}v_{tgt}\cos(\Delta i) = 0.282760 \text{ ER/TU}$$

Finally, the total Δv is

▲ $\Delta v_{TOTAL} = \Delta v_{phase} + \Delta v_{trans1} + \Delta v_{trans2}, \Delta v_{trans2} = 0.573560 \text{ ER/TU} = 4.53420 \text{ km/s.}$

5.7 Continuous-Thrust Transfers

For most maneuvers discussed so far, we've assumed the use of *impulsive* thrust. Although impulsive models approximate most maneuvers well, some energetic burns may require more time. In addition, we now have the technology to continuously thrust satellites, so we need to analyze certain characteristics of continuous-thrust maneuvers. Although we could use a continuous-thrust maneuver for any of the transfers in the chapter, this section focuses on (1) examining *any* continuous-thrust level to raise or resize an orbit, and (2) exploring low-continuous-thrust maneuvers with which to simultaneously

change orbital inclination and size (combined maneuvers in Sec. 5.5). This next section develops simple graphical and analytical tools to determine the minimum elapsed time and the associated fuel requirements for a continuously thrusting vehicle.

It's beneficial for you to be familiar with *Hamiltonian* formulations, *Lagrange multipliers*, and *optimal-control theory*, because this section departs from much of the analysis we've used to this point. Although a complete description of optimal-control theory would take an entire book, such as Bryson and Ho (1987), I'll try to introduce a few very broad concepts to help transition you into this section.

The fundamental idea of optimal-control theory, as applied to the particular problem of orbital maneuvers, is to form a Hamiltonian and take partial derivatives to determine the optimal solution for a desired set of cost functions. *State equations* (equations of motion) determine the laws and boundaries of a system. These can be natural (gravity), or physical constraints (maximum thruster force). They can also define restrictions such as keep-out zones. *Cost functions*, also known as *penalty functions*, or *performance indices*, assess a weight (cost) to certain conditions. These conditions can be terminal (miss distance from a location) or cumulative (fuel usage). The weights themselves may be a function of time or state conditions.

The Hamiltonian adjoins the state equations to the cost function. This is a rigorous mathematical process with specific mechanisms to form the Hamiltonian. A basic requirement is to reformulate each state equation so it equals zero before combining it with the Lagrange multipliers. We typically have more than one cost function, so the Hamiltonian combines all the cost functions and equations of motion into a single, one-dimensional solution space. We take partial derivatives of the Hamiltonian to determine the minima and maxima of the function. When these are all combined, we have the global solution.

If we knew the solution for each function, we could immediately form the Lagrange multipliers, and the partial derivatives would all equal zero. This rarely happens, but the partial derivatives give us an indication of how to move from our initial guess. The overall process is mathematically intensive and sometimes requires "human" intervention to arrive at the proper solution. That's why a main result from this section is the formation of tables and plots which an operator can use to estimate the answer; without having to implement the complex mathematical relations. Formal control theory includes a wealth of methods to determine the initial guesses, as well as techniques to iterate towards the local extrema. We won't discuss these in detail, but you can find useful information in books such as Bryson and Ho (1987).

5.7.1 Introduction

It's possible to change a satellite's orbital inclination and semimajor axis using low-thrust, many-revolution maneuvers. Alfano (1982) and Wiesel and Alfano (1985) have developed practical techniques to model these changes. The techniques are limited to low-thrust, many-revolution orbit transfers because the mathematical model assumes the thrust is a small perturbation. Alfano and Thorne (1994) show how to assess changes in semimajor axis with any thrust level. The resulting charts enable mission planners and satellite

designers to determine preliminary fuel requirements and transfer times for constant-thrust systems. They also permit propulsion trade-off studies, for which we need only a straight-edge and scientific calculator.

Low-thrust trajectories have attracted a lot of work, and studies showing how to make them optimal and continuous are especially relevant. For example, Edelbaum has published many papers on changes in semimajor axis and inclination, and the *Alfano transfer* is an analytical solution for the nearly optimal case of a continuous, tangential, orbit transfer using low thrust and many revolutions.

The equations of motion for obtaining the best control are in their complete form for the coplanar, circle-to-circle case. The *initial acceleration*, a_i , appears in these equations and is allowed to vary from case to case as a parameter. Changing the spacecraft's initial acceleration always changes the minimum time to raise the orbit. To make the solution data more universal, we must rescale the problem using the initial orbit radius and the gravitational parameter of the central body. We replace the *time of flight*, t_f , with the *total accumulated velocity change*, v_{acc} . Graphs present the results of many such optimal cases, showing the relationship between spacecraft acceleration, propellant mass fraction, ratio of the initial-to-final orbit radii, and the minimum accumulated velocity change. The graphs themselves consist of parametric families of optimal solutions assembled in comprehensive charts so planners of orbital missions can interpolate them. Analytical equations are provided when practical.

The charts and equations cover the range of a spacecraft's acceleration values. Many users of space assets are interested in raising and repositioning orbits in minimum time, so the high-acceleration cases will be increasingly interesting as technology improves for continuous-thrust propulsion.

Equations of Motion

The differential equations defining spacecraft orbit raising use the following assumptions:

1. The force of thrust is constant and always in the plane of motion.
2. The vehicle has a *fixed* propellant mass flow rate.
3. The vehicle's acceleration is due solely to the force of thrust and an inverse-square, central gravitational field that is spherically symmetrical.

We must form the equations of motion as a first-order system. Bryson and Ho (1987, 42–89) give the polar form of the equations of motion as

$$\begin{aligned} \dot{r} &= v_R & \dot{v}_R &= \frac{v_S^2}{r} - \frac{\mu}{r^2} + \frac{a_i \sin(\phi_c)}{1 + \dot{m}t} \\ \dot{v}_S &= -\frac{v_R v_S}{r} + \frac{a_i \cos(\phi_c)}{1 + \dot{m}t} & \dot{\theta}_p &= \frac{v_S}{r} \end{aligned} \quad (5-37)$$

where v_R = radial-velocity component (RSW system)

v_S = transverse velocity component (RSW system)

a_i = initial vehicle acceleration

ϕ_c = in-plane control angle (measured from the velocity vector to the thrust vector with positive values in the radial direction)

θ_p = polar angle denoting the vehicle's location referenced from the start of the maneuver

The *specific propellant mass flow rate*, \dot{m} , is the actual mass flow rate divided by the initial mass; this value is negative when dealing with propellant loss.

Accumulated Velocity Change

We define the *accumulated velocity change*, v_{acc} , as the total velocity imparted by the force of thrust during the elapsed transfer time, t_f , so that

$$v_{acc} = \int_0^{t_f} \frac{a_i}{1 + \dot{m}t} dt \quad (5-38)$$

Eq. (5-38) is a variation of the *ideal rocket equation*, which is more commonly known as

$$\Delta v = g I_{sp} \text{LN} \frac{m_{init}}{m_{init} - m_{prop}}$$

where m_{init} = the vehicle's initial mass

m_{prop} = propellant mass

g = gravity

I_{sp} = specific impulse

Eq. (5-38) models the effects of continuously expending fuel; it also recasts the solution in terms of accumulated velocity change. This formula serves two purposes: it models the effect of fuel depletion and also defines the relationship between time and thrust-induced velocity change through the equation that completes the integration in Eq. (5-38):

$$v_{acc} = \frac{a_i}{\dot{m}} \text{LN} (1 + \dot{m}t_f) \quad (5-39)$$

Recasting the minimum-time solution in terms of v_{acc} allows us to graph compactly the initial acceleration and the *propellant mass fraction*, m_p ,

$$m_p = -\dot{m}t_f$$

Given a_i , m_p and v_{acc} , we can find the specific mass flow rate and final time:

$$\dot{m} = \frac{a_i}{v_{acc}} \text{LN} (1 - m_p) \quad t_f = -\frac{m_p}{\dot{m}} \quad (5-40)$$

For the limiting case in which \dot{m} and m_p approach zero, the total transfer time from Eq. (5-38) is simply

$$t_f = \frac{v_{acc}}{a_i} \quad (5-41)$$

5.7.2 Orbit Raising

The first major continuous-thrust application is the ability to change an orbit's size, which is extremely important for any operation in space. We usually can't insert a payload directly into a geosynchronous orbit because it requires a large booster. A very appealing alternative is to launch the satellite into a low-Earth orbit and have a transfer booster place it into the correct final orbit. Techniques already discussed showed the amount of change in velocity required for an impulsive maneuver to do this. We can use a continuously thrusting maneuver whenever (1) time is available to place the satellite into the final orbit and (2) an efficient motor is available to provide the required change in velocity for the operating period. Mission constraints usually determine the time available to maneuver. Recent improvements in technology have given us low-thrust electrical vehicles and intermediate-thrust, nuclear, transfer vehicles. So we must now determine—with the following techniques—which technology best fits the available funds, equipment, and requirements. As mentioned, we'll consider only changes to an orbit's size because some applications don't work well for mathematical solutions concerning changes to orbital inclination.

Optimal-Control Formula—Orbit Raising

Given an initial and final radius, we want to use as little fuel as possible between circular coplanar orbits; this means minimizing transfer time because we assume \dot{m} is constant. It's the process we explored in the Hohmann transfer, except that we now use nonimpulsive maneuvers. The control formula for this problem follows Bryson and Ho (1987, 42–89), but I've repeated it here for convenience. First, write the *Hamiltonian*, \mathcal{H} , as

$$\mathcal{H} = 1 + \lambda_r \dot{r} + \lambda_{v_R} \dot{v}_R + \lambda_{v_S} \dot{v}_S$$

where you omit the polar angle because the exact initial and final positions aren't specified and θ_p doesn't appear in the equations of motion [Eq. (5-37)]. Capture the behavior of the Lagrange multipliers, called λ *dynamics*, by

$$\begin{aligned}\dot{\lambda}_r &= -\lambda_{v_R} \left(-\frac{v_S^2}{r^2} + \frac{2\mu}{r^3} \right) - \lambda_{v_S} \left(\frac{v_R v_S}{r^2} \right) \\ \dot{\lambda}_{v_R} &= -\lambda_r + \lambda_{v_S} \left(\frac{v_S}{r} \right) \\ \dot{\lambda}_{v_S} &= -\lambda_{v_R} \left(\frac{2v_S}{r} \right) + \lambda_{v_S} \left(\frac{v_R}{r} \right)\end{aligned}$$

The partial of \mathcal{H} with respect to the control parameter (in-plane thrust control angle) must equal zero, so

$$\frac{\partial \mathcal{H}}{\partial \phi_c} = \left(\lambda_{v_R} \cos(\phi_c) - \lambda_{v_S} \sin(\phi_c) \right) \frac{a_i}{1 + \dot{m}t} = 0$$

Because the acceleration term can't be zero, the parenthetical expression establishes the *control law* as

$$\tan(\phi_c) = \frac{-\lambda_{v_R}}{-\lambda_{v_S}}$$

You choose signs in the ratio of Lagrange multipliers that are different from those in Bryson and Ho because they were trying to maximize—whereas you're trying to minimize—time. With the negative signs in the control law, the resulting sine and cosine expressions will guarantee a local minimum for this problem:

$$\sin(\phi_c) = \frac{-\lambda_{v_R}}{\sqrt{\lambda_{v_R}^2 + \lambda_{v_S}^2}} \quad \cos(\phi_c) = \frac{-\lambda_{v_S}}{\sqrt{\lambda_{v_R}^2 + \lambda_{v_S}^2}}$$

Your initial choice of Lagrange multipliers characterizes the complete transfer. To date, the only closed-form solution that exists is for the many-revolution case, assuming intermediate eccentricity is zero and considering only tangential thrust (Edelbaum, 1961 and Alfano, 1982). Recognizing that these multipliers can be scaled without affecting the control law or λ dynamics, initialize λ_r to -1 , and use a numerical search to find the remaining two. This process is time-consuming, but the charts in the next section remove this complexity and give us a quick solution.

Generating and Discussing Charts and Equations

Proper scaling can eliminate dependence on a specific central attracting body and allow global mapping of solutions. The canonical units introduced in Sec. 1.8 offer this type of scaling, but I'll develop canonical-based units that allow you to simply change the gravitational parameter for *any* central body.

$$1 \text{ DU}^* = R \quad 1 \text{ TU}^* = \sqrt{\frac{R^3}{\mu}}$$

Although the canonical units depend on the physical parameters of a given transfer, the equations of motion don't. The gravitational parameter is always $1 \text{ DU}^{*3}/\text{TU}^{*2}$, and the initial values of the (r_1, v_S) array for any circle-to-circle coplanar transfer are simply $\hat{r} = 1\hat{R} \text{ DU}^*$ and $v = 1S \text{ DU}^*/\text{TU}^*$. The final values are $\hat{r} = R\hat{R} \text{ DU}^*$ and $v = \sqrt{1/R}\hat{S} \text{ DU}^*/\text{TU}^*$, where R is the ratio [recall Eq. (5-4)] of the final radial distance to the initial. This section considers only orbit raising ($R > 1.0$).

As mentioned, the initial values of the Lagrange multipliers completely define the transfer. To solve this two-point, boundary-value problem, set λ_r to -1 and use a shooting method (Press et al. 1992, Chap. 17) to find λ_{v_R} , λ_{v_S} , and t_f . Transform minimum-time solutions for various *orbit ratios*, R , and *mass fractions*, m_p , into accumulated velocity change and plot them against initial acceleration to produce Figs. 5-21 through 5-23. The ripples in the curves of Figs. 5-21 and 5-22 reflect a transition region from gravitational-force dominance to thrust dominance; you can use them to define the boundaries of low thrust (flat curve), intermediate thrust (rippled curve), and high thrust (constant upward-sloping curve). Also, the ripples and subsequent upward slope testify to the additional cost of placing the satellite in its final circular orbit.

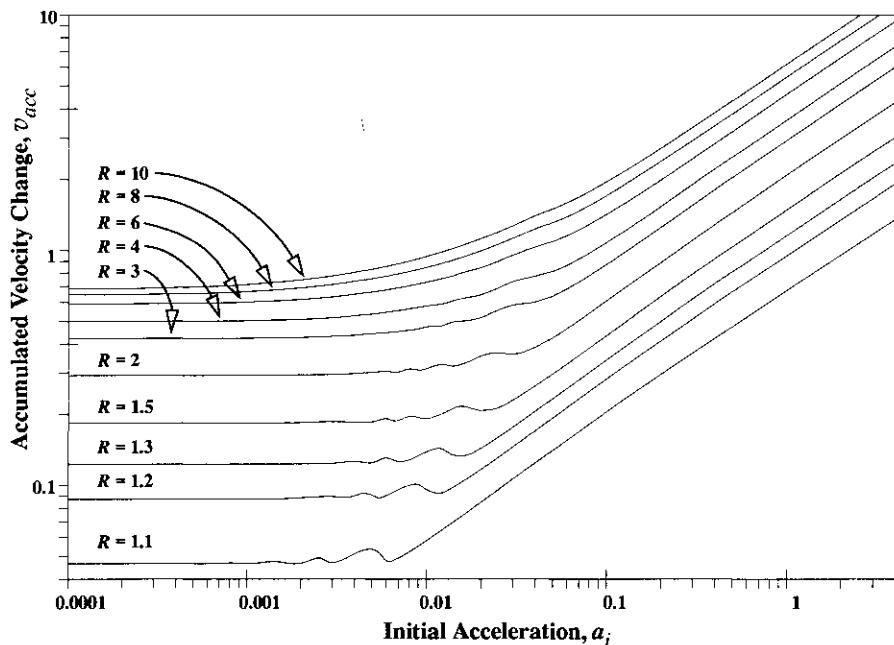


Figure 5-21. v_{acc} Contours for Various Orbit Ratios, R ($m_p = 0.25$). Notice the various non-dimensional initial accelerations with their resulting accumulated velocity changes for different orbit ratios (final to initial). The ripples represent a transition between gravitational and thrust dominance, which induces a slight eccentricity in the orbit.

Figure 5-23 results from taking the optimal solutions for $m_p = 0$. As you can see, not completing a revolution during transfer increases the associated cost (more accumulated velocity). The optimal-steering law causes the eccentricity to increase for the first half of each revolution and then diminish in the latter half. The greater the thrust, the more eccentricity increases. If the total polar angle isn't a multiple of 360° , we must zero out the

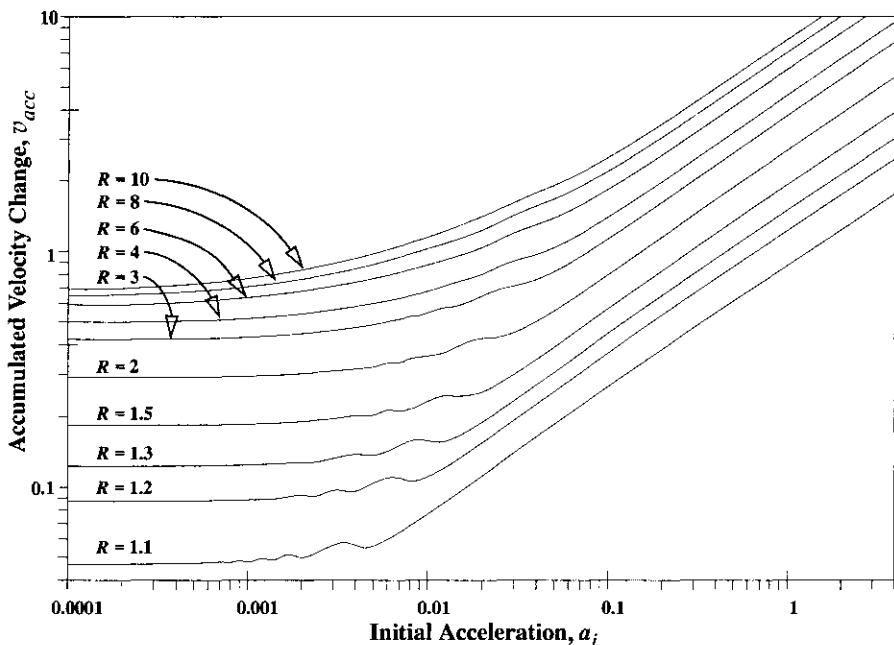


Figure 5-22. v_{acc} Contours for Various Orbit Ratios, R ($m_p = 0.75$). The nondimensional initial accelerations versus velocity change curve is similar to Fig. 5-21, although the ripples occur at lower initial accelerations due to the increased mass loss.

eccentricity to meet the final condition of a circular orbit. The peaks in the curves reflect the added cost of this process. When the satellite revolves more than five times around the central body (to achieve the same final orbit as a single revolution), this effect is negligible because the thrust is much smaller. Therefore, the induced eccentricity is very small and correctable with little additional maneuvering. If the satellite reaches its final state in less than one revolution, the intermediate eccentricity and associated cost grow with the shortness of the transfer arc. The positive slope of the curves reflects these relationships. The cost of recircularizing diminishes as the ratio R increases because the gravitational force is less, making the vehicle's acceleration more effective. Also, when compared to Fig. 5-21, the smaller ripples of curves for increasing mass fraction reflect the increased thrust effectiveness due to mass loss.

According to Alfano (1982), all continuous-thrust, circle-to-circle transfers for which $a_i < 10^{-4}$ ER/TU² are due to the constancy of accumulated velocity change for a given orbit ratio. The equation relating orbit radius to velocity change for coplanar transfers (Alfano, 1985) reduces to

$$\frac{dr}{dv} = 2 \sqrt{\frac{r^3}{\mu}}$$

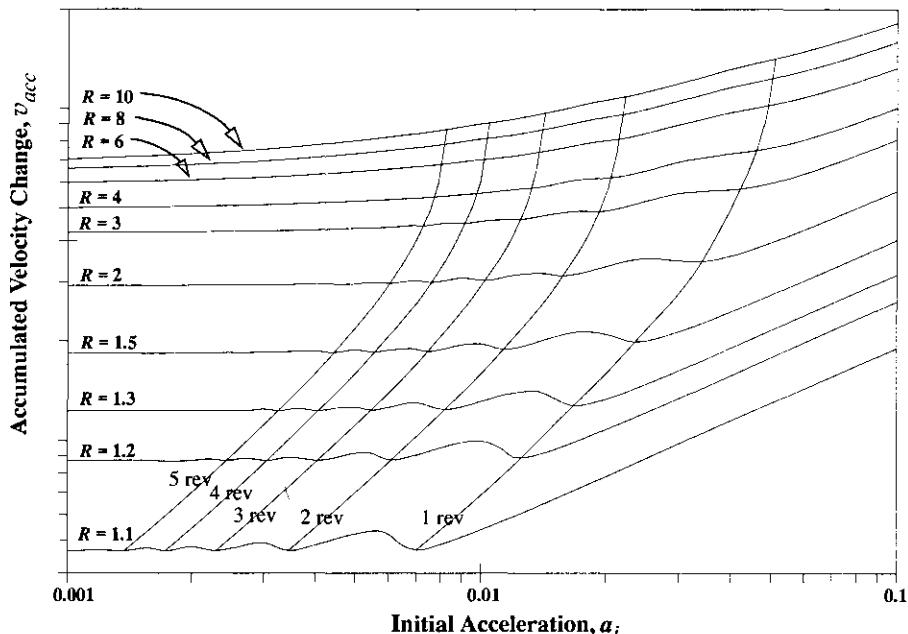


Figure 5-23. v_{acc} Contours for Various Orbit Ratios, R ($m_p = 0$). The revolution lines are superimposed to show the penalty for not completing a revolution during the transfer. As we use more revolutions, the penalty diminishes, but for the first iteration, it can be significant. Mission planners should choose thrust levels which give polar angles that are multiples of 360° .

Using the previously defined units and boundary conditions, we discover the accumulated velocity change for orbit-raising is simply

$$v_{acc} = 1 - \sqrt{\frac{\mu}{R}} \quad (5-42)$$

With this equation we can find the accumulated velocity change given the ratio of the desired orbital change.

Let's analytically determine the constant upward-sloping lines of Figs. 5-21 to 5-22 by examining the limiting high-thrust case; the thrust dwarfs the gravitational force, which you can assume to be negligible. This results in a straight-line, radial trajectory, for which acceleration is directly away from the central body until the *switching time*, t_s ; then it reverses to complete the transfer. The switching time does *not* occur at the half-way point due to fuel loss during the transfer. You can approximate all transfers for which $a_i > 4$ by the equation

$$v_{acc} = 2\sqrt{(R-1)a_i} \quad m_p = 0$$

$$v_{acc} = -\ln(1-m_p) \sqrt{\frac{(R-1)a_i}{2-m_p-2\sqrt{1-m_p}}} \quad m_p > 0$$

Then, find the final time from Eq. (5-40) or Eq. (5-41) and discover the switching time to be

$$t_s = 0.5t_f \quad m_p = 0$$

$$t_s = \left(\frac{1-\sqrt{1-m_p}}{m_p} \right) t_f \quad m_p > 0$$

Implementing Continuous-Thrust Solutions

By convention, I'll always give the input parameters as the array $(r_i, r_f, \mu, a_i, m_p)$. I assume you can determine total accumulated velocity to two significant figures from the transfer charts; I'll carry six significant figures in all computations to reduce round-off error. See the table on the inside front cover for the physical constants.

To use the charts, begin with the input parameters. Scale these values with the canonical units and determine the accumulated velocity change from the appropriate chart. You can determine the mass flow rate and minimum time of flight from Eq. (5-40) and Eq. (5-41) if they're needed. Let's look at an example to illustrate the use of the charts.

▼ Example 5-11. Determining the Time Needed for a Transfer between Earth and Mars.

Alfano and Thorne (1994) use an example from Bryson and Ho (1987, 18) to check the validity of their concept. The case is a transfer between Earth and Mars, for which we assume both planetary orbits are circular and coplanar. The spacecraft starts in a heliocentric orbit free of Earth's gravitational field and finishes free of Mars's field. We don't consider Earth escape and Mars capture. A mass flow rate of -5.8513 kg/day (-12.9 lb/day) is given for a 4535.92 kg (10^4 lb) vehicle instead of the propellant mass fraction we need. Interpolation finds a solution. The specific mass flow rate is $-1.29 \times 10^{-3} \text{ /day}$ or $-1.493.06 \times 10^{-7} \text{ /s}$; as a starting point m_p is set at 0.25.

Find initial data from Table D-3 and Table D-4. The radius of the Earth from the Sun is $1.49598 \times 10^8 \text{ km}$ and that of Mars from the Sun is $2.27939 \times 10^8 \text{ km}$. The gravitational parameter of the Sun is $1.32712 \times 10^{11} \text{ km}^3/\text{s}^2$ and $8.33173 \times 10^{-7} \text{ km/s}^2$. The propellant mass fraction is 0.25.

Convert to standard canonical units. 1.0 AU , 1.52368 AU , $1.0 \text{ AU}^3/\text{TU}_\odot^2$, $0.1405 \text{ AU}/\text{TU}_\odot^2$, 0.25. For this problem, the standard units are identical to the "starred" canonical units because the radius of the Earth, the initial radius, is already unity. From Fig. 5-21, find $v_f = 0.54 \text{ AU}/\text{TU}_\odot$

$$\dot{m} = -0.0748506 / \text{TU}_\odot = -1.49026 \times 10^{-8} \text{ s}$$

Now the transfer time $t_{trans} = 3.33999 \text{ TU}_\odot = 1.67756 \times 10^7 \text{ s}$

This final time solution is adequate because \dot{m} from Eq. (5-40) is within two significant figures of the original. For illustration purposes, let's increase m_p to 0.5 and interpolate to refine the results and better determine the transfer time.

From Fig. 5-21, find $v_f = 0.59 \text{ AU}/\text{TU}_\odot$

$$\dot{m} = -0.165\ 063 / \text{TU}_\odot = -3.286\ 37 \times 10^{-8} / \text{s}$$

$$t_f = 3.029\ 15 \text{ TU}_\odot = 1.521\ 44 \times 10^7 \text{ s}$$

Linear interpolation of \dot{m} produces a final time of $3.339\ 51 \text{ TU}_\odot$, $1.677\ 29 \times 10^7 \text{ s}$, or 194.131 days. This is within two significant figures of Bryson and Ho's solution of 193 days and the exact numerical solution of 192.748 days.



▼ Example 5-12. Transferring from LEO to GEO.

This case involves a coplanar transfer from a circular, low-Earth parking orbit of 1.05 ER to a final circular GEO orbit of 6.61 ER. The vehicle's initial acceleration is $4.0 \times 10^{-6} \text{ m/s}^2$, with a propellant mass fraction of 0.25.

Initial data is 1.05 ER, 6.61 ER, $1.0 \text{ ER}^3/\text{TU}^2$, $4.0 \times 10^{-6} \text{ m/s}^2$, 0.25

Conversions to "new" canonical units yield 1.0 DU^* , $6.295\ 24 \text{ DU}^*$, $1.0 \text{ DU}^{*3}/\text{TU}^{*2}$, $4.500\ 79 \times 10^{-4} \text{ DU}^*/\text{TU}^{*2}$, 0.25

From Eq. (5-42), find $v_f = 0.601\ 44 \text{ DU}^*/\text{TU}^*$

$$\dot{m} = -2.152\ 82 \times 10^{-4} / \text{TU}^* = -2.48 \times 10^{-7} / \text{s}$$

$$t_f = 1.161\ 26 \times 10^3 \text{ TU}^* = 1.008\ 06 \times 10^6 \text{ s}$$

▲ This case involves low thrust (flat curve for Fig. 5-21 to Fig. 5-23, v_{acc} constant for a given orbit ratio R regardless of m_p or a_p) and agrees with the numerical solution.

5.7.3 Low-Thrust, Noncoplanar Transfers

Several authors have studied low-thrust orbital transfers using minimum time. In this section we'll also use two time scales, separating the optimal-control problem into a fast-time-scale problem over one orbit and a slow-time-scale problem over the entire transfer. I'll call the process the *Alfano transfer* (1982) because it represents the first analytical closed-form solution. We'll review Edelbaum's solution of the fast-time-scale problem and then solve the optimal-control problem for the overall transfer. For the slow time scale, we can find the optimal-control law in explicit form. This reduces the slow-time-scale problem to a two-point, boundary-value problem with the semimajor axis and inclination. The solution space for this problem is global, and we can easily obtain explicit requirements for the total velocity change for any desired transfer. The solution assumes constant thrust as the vehicle's mass decreases but is also optimum for any program that uses a slowly varying throttle. To understand the solution, keep in mind the thrust is treated as a perturbation. Chapters 7 and 8 detail the formulas and solution techniques for perturbations. In this application, we simply use the equations.

Short-Time-Scale Problem

Edelbaum (1961) first solved the optimal-control problem over the fast time scale of one orbit, but we'll quickly review it for later discussions. For a low-thrust vehicle in a nearly circular orbit, we use the Lagrange planetary equations for semimajor axis [Eq. (8-12)] and inclination [Eq. (8-24)]. We'll also use *vehicle acceleration*, a_v , and the *thrust-control angle*, ϑ , from the orbital plane. ϑ is sometimes thought of as the *vehicle's pitch angle*.

depending on whether the thruster is on the satellite or has a gimbal mechanism. For small eccentricity, we don't consider any yaw component of the vehicle's acceleration.

Assuming the *control* $\vartheta(\nu)$ as a function of true anomaly, we can try to get the most inclination change over one orbit while still changing the semimajor axis a given amount, Δa . We use the true anomaly recognizing that variations may occur. Because the orbit is generally circular, we could use the argument of latitude. Equatorial orbits would require true longitude, so we use true anomaly as a representative angle. We could also use equinoctial elements to fix these difficulties. Assuming small changes over the orbit leads to

$$\delta \int_0^{2\pi} \left[\frac{a^2 a_v}{\mu} \sin(\vartheta(\nu)) \cos(\nu) + \lambda \left(\frac{2a^3 a_v}{\mu} \cos(\vartheta(\nu)) - \frac{\Delta a}{2\pi} \right) \right] d\nu = 0$$

Optimizing yields the *control law* for the short-time-scale problem, $\vartheta_{SS}(\nu)$:

$$\tan(\vartheta_{SS}(\nu)) = \frac{\cos(\nu)}{\sqrt{\frac{1}{c_v} - 1}} \quad (5-43)$$

Here, anticipating the next section, we've used the *control variable*, c_v , with an operative range of c_v from 0.0 to 1.0:

$$c_v \equiv \frac{1}{4\lambda^2 a^2 + 1}$$

The changes in semimajor axis and inclination per revolution are

$$\begin{aligned} \Delta a_{rev} &= \frac{8a^3 a_v}{\mu} \sqrt{1 - c_v} K(c_v) \\ \Delta i_{rev} &= \frac{4a^2 a_v}{\mu} \left[\frac{1}{\sqrt{c_v}} E(c_v) + \left(\sqrt{c_v} - \frac{1}{\sqrt{c_v}} \right) K(c_v) \right] \end{aligned}$$

K and E are the complete elliptic integrals of the first and second kinds, respectively. Abramowitz (1970, 591–592) gives convenient formulas to approximate these functions.

Several other results must be mentioned. The net change in eccentricity and node per orbit is zero with this control program; therefore, circular orbits stay nearly circular. The control law $\vartheta(\nu)$ varies from a pure in-track acceleration for $c_v = 0$, when only the semimajor axis changes, to a pure out-of-plane thrust profile that switches at the node for $c_v = 1$, when only the inclination changes.

Long-Time-Scale Problem

The previous section showed an approximately optimal way to produce small changes in the orbital elements over one orbit. On the long time scale of the entire transfer, these

expressions can be divided by the Keplerian period [Eq. (2-16)] (short time means the period is less than one revolution) to yield equations of motion on the long time scale. Also, the long time scale must include the effects of depleting fuel, which (for constant thrust) cause the vehicle's acceleration to vary as

$$a_{th}(t) = \frac{a_i}{1 + \dot{m}t}$$

This last effect introduces explicit time dependence into the control Hamiltonian, but we can eliminate it if we introduce the new independent variable τ by placing $d\tau = a_0(t)dt$. This leads to the time transformation already presented in Eq. (5-39), ($v_{acc} = \tau_f$). Physically, v_{acc} is the total accumulated velocity change of the vehicle during the transfer. The transformed equations of motion on the long time scale then become

$$\begin{aligned}\frac{\Delta a}{\Delta \tau} &= \frac{4}{\pi} \sqrt{\frac{a^3}{\mu}} \sqrt{1 - c_v} K(c_v) & P(c_v) &= \sqrt{1 - c_v} K(c_v) \\ \frac{\Delta i}{\Delta \tau} &= \frac{2}{\pi} \sqrt{\frac{a}{\mu}} \left[\frac{1}{\sqrt{c_v}} E(c_v) + \left(\sqrt{c_v} - \frac{1}{\sqrt{c_v}} \right) K(c_v) \right] & R(c_v) &= \frac{1}{c_v} E(c_v) + \left(\sqrt{c_v} - \frac{1}{\sqrt{c_v}} \right) K(c_v)\end{aligned}$$

In this form the equations of motion are free of the values of a_i and \dot{m} for a particular vehicle and are also free of the variable τ .

We can now pose the long-time-scale problem. We wish to transfer from given initial values of a_0 and i_0 to given final values of a_f and i_f using the minimum change in velocity, v_{acc} . At the same time, elapsed time and fuel consumption must be as low as possible. This is a standard problem: optimal control in minimum time (Bryson and Ho, 1987, 76–89). The *control Hamiltonian*, \mathcal{H}_c , is

$$\mathcal{H}_c = 1 + \lambda_a \frac{4}{\pi} \sqrt{\frac{a^3}{\mu}} P(c_v) + \lambda_i \frac{2}{\pi} \sqrt{\frac{a}{\mu}} R(c_v)$$

where I've introduced two Lagrange multipliers. Get the optimal control profile by solving

$$\frac{da}{d\tau} = \frac{\partial \mathcal{H}_c}{\partial \lambda_a} \quad \frac{di}{d\tau} = \frac{\partial \mathcal{H}_c}{\partial \lambda_i}$$

$$\frac{d\lambda_a}{d\tau} = -\frac{\partial \mathcal{H}_c}{\partial a} \quad \frac{d\lambda_i}{d\tau} = -\frac{\partial \mathcal{H}_c}{\partial i}$$

$$\frac{\partial \mathcal{H}_c}{\partial c_v} = 0$$

$$\mathcal{H}_c(\tau_f) = 0$$

The first two equations repeat the equations of motion. Because the control Hamiltonian is independent of inclination, λ_i is a constant. Also, because \mathcal{H}_c is not an explicit function of τ , \mathcal{H}_c itself is constant, and the transversality condition (the beginning and the end conditions are the same) tells us this value of \mathcal{H}_c is zero.

Let's find the explicit control law for this problem. The optimality condition and the statement that $\mathcal{H}_c = 0$ supply two linear equations for λ_a and λ_i . Solve these for the constant value of λ_i and invert the result to yield the control law for the long-term problem in terms of the semimajor axis and the constant value of λ_i :

$$c_v = \varphi^{-1} \left[\frac{\pi}{2\lambda_i \sqrt{a}} \sqrt{\mu} \right] \quad (5-44)$$

The function φ is defined as

$$\vartheta_{LS}(c_v) = \frac{R'(c_v)P(c_v)}{P'(c_v)} - R(c_v)$$

where primes refer to differentiation with respect to v . This function is monotonic, so we can't invert it. Shortly, I'll show you an approximate numerical formula for φ^{-1} .

By discovering the explicit control law, we've reduced the minimum-time control problem to solving the state equations of motion [Eq. (5-37)]. This solution is subject to the control law [Eq. (5-44)], with given boundary conditions. Thus, we've significantly reduced the order compared to that of the original problem. Now, starting from given initial conditions, a_o and i_o , we must determine two unknown constants, λ_i , and τ_f , while meeting the final conditions, a_f and i_f . We can easily survey this boundary-value problem with two variables by using numerical techniques and then globally map the results.

Now, we introduce nondimensional units by setting $a_o = 1$, $i_o = 0^\circ$, and setting the gravitational parameter $\mu = 1$, which—by Kepler's third law—implicitly sets the time unit. Numerically integrating the state equations of motion with the control [Eq. (5-44)] produces Fig. 5-24, in which contours of λ_i and τ_f are plotted on the a_f , i_f plane.

Curves of constant λ_i are also trajectories across the planet. Note that large inclination changes occur at the largest value of semimajor axis in a particular transfer—a behavior paralleling that of impulsive maneuvers. The values of λ_i range from 0° to -90° , while the τ contour interval is 0.1 in the dimensionless units (DU*/TU*). The figure permits simple interpolation of approximate values of the constants τ_f and λ_i to solve any boundary-value problem over a long time scale.

Implementing Low-Thrust Solutions

We have formed the problem for transferring from one circular orbit to another, using low thrust over many revolutions. Under very general conditions, we've also reduced this optimal-control problem to quadratures. Knowing the desired initial and final orbits, we first

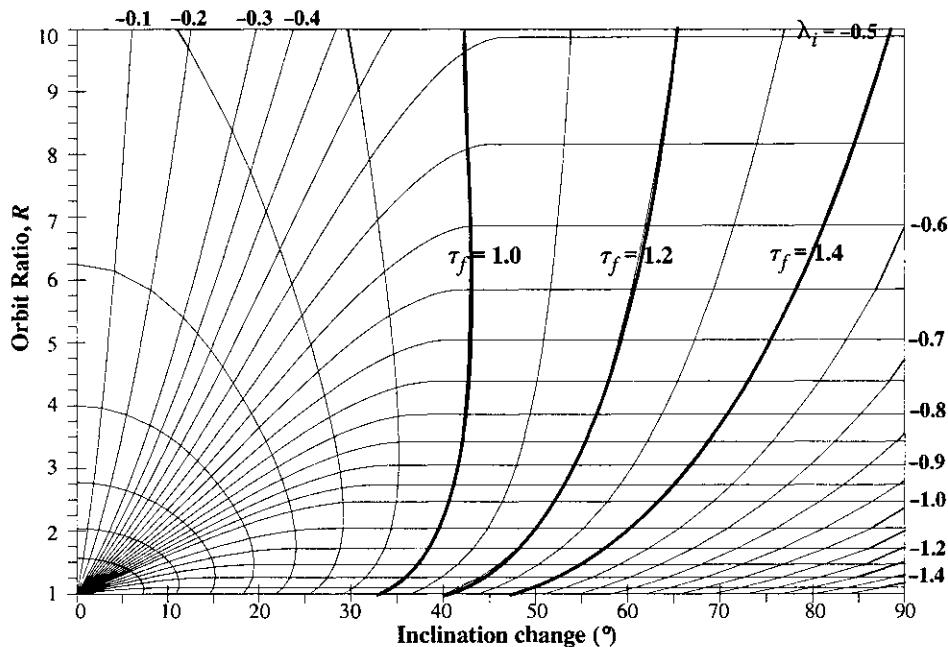


Figure 5-24. Low Thrust, Noncoplanar Transfers using the Alfano Transfer. This chart permits us to estimate changes in semimajor axis and inclination for satellites using continuous low thrust. Contours of λ_i represent trajectories and are labeled with the appropriate value. The contour interval for τ_f is 0.1.

solve the boundary-value problem for the slow time scale. The vehicle then executes a periodic program for pitch angle given by Eq. (5-43), with the slow-control variable, u , from Eq. (5-44). Whenever we meet the basic assumptions of the two time-scale methods, the fast-time variable, v , and the slow-time variable, τ , are independent of one another. As long as at least five or more revolutions elapse, the vehicle arrives in the final orbit when $\tau = \tau_f$. The vehicle need not be able to solve the optimal-control problem. Given the values of λ_i and τ_f for the desired transfer, the vehicle only needs access to the current values of a and v to calculate the control functions.

Separating the two time variables has several advantages for operations. Should a thruster fail, for example, the optimal solution *doesn't change*. Only the τ to t conversion changes as vehicle parameters alter. Should the vehicle need to return to its original orbit after depositing a payload, the best return trajectory will be the simple τ -time reverse of the outbound transfer, as long as more than five revolutions elapse on the return leg. In fact, the time transformation isn't limited to the constant-thrust vehicle we've assumed. Any arbitrary thrust program, $a_v(t)$, used in the time transformation produces the same trajectory results with an optimal τ -time, as long as $a_v(t)$ varies only "slowly" over one orbital period.

We have simulated several transfers numerically by using this control law in the full set of Lagrange planetary equations. The method of two time scales—and our solution—is exact only when an infinite number of revolutions elapse during the transfer.

Numerical Considerations

In this section we detail the closed-form solution to the problem of using low thrust to transfer from circular orbit to circular orbit. The solution is exact for an infinite number of revolutions during the transfer and is simple enough for the onboard control system of an orbit-transfer vehicle to handle it. It can also tolerate changes in thrust programs over a very wide range, including loss of thrusters.

Wiesel and Alfano (1985) provide a method to reproduce the results using a Chebyshev polynomial fit, but it requires a way to easily generate values of the function φ^{-1} . We can best approximate the function $v = \varphi^{-1}(X)$ as a rational function of the variable $Z = X^{-2}$ in the form

$$u = \varphi^{-1}(Z = X^{-2}) \approx \sum_{i=0}^{10} \alpha_i Z^i \sum_{i=0}^{10} \beta_i Z^i$$

The variable Z eliminates a singularity in φ . The coefficients α_i and β_i yield values for φ^{-1} that are accurate to at least seven figures (see Table 5-4). This is a Chebyshev fit, not a Taylor series, so don't truncate it.

TABLE 5-4. Expansion Coefficients for $\varphi^{-1}(Z)$.

i	α_i	β_i
0	0.0	1.0
1	2.467 410 607	0.460 969 883 8
2	-1.907 470 562	13.775 631 532 4
3	35.892 442 177	-69.124 531 667 8
4	-214.672 979 624	279.067 183 250 0
5	947.773 273 608	-397.662 895 213 6
6	-2114.861 134 906	-70.013 993 504 7
7	2271.240 058 672	528.033 426 684 1
8	-1127.457 440 108	-324.930 383 652 0
9	192.953 875 268	20.583 824 517 0
10	8.577 733 773	18.816 537 077 8

5.8 Relative Motion

This section develops and analyzes equations of motion between two close-orbiting satellites. We want to look at the geometry relative to the target satellite and with respect to a ground station. First, I'll develop the equations of motion and then show the solutions needed to analyze two- and three-dimensional graphs of the results. The analysis follows the development of the *Clohessy-Wiltshire* or *CW* equations (Clohessy and Wiltshire, 1960), also known as *Hill's equations* [Hill (1878)].

To find the equations of motion, look first at Fig. 5-25, showing the geometry of the target and interceptor satellites. The target will represent the *primary* satellite throughout this section, and the interceptor will represent the *secondary* satellite. Because both satellites are in orbit, the choice of one as a target over the other is arbitrary and depends on the application. Although we begin with a general formulation, we'll end up with a restriction for circular orbits. Later, we'll compare the results to numerical solutions to help determine the robustness of Hill's equations.

To derive the equations of motion, we use the *RSW* coordinate system introduced in Sec. 1.4. The *R*-axis is colinear with the position vector; *S* is in the *direction* of the velocity vector aligned with the local horizontal; and *W* is normal to the orbit plane. These arrangements are consistent with Kaplan's notation (1976, 108–115) but different from the Clohessy-Wiltshire geometry, which has *R* and *S* switched and *S* in the direction *opposite* to satellite motion. This is really just a rotation about the *W* axis of -90° from the system above. If you intend to use the coordinate system described by Clohessy-Wiltshire, be sure to rederive the equations—several signs change with the different definition. We'll also use an *xyz* notation for the components of the *RSW* system because (1) it's probably the best “standard” convention used in the literature for Hill's equations and (2) using *xyz* reminds us that results are approximate.

We can define the equation of motion for the target satellite as two-body motion, so

$$\ddot{\vec{r}}_{tgt} = -\frac{\mu \vec{r}_{tgt}}{r_{tgt}^3} \quad (5-45)$$

The interceptor will have the same equation; however, if we want a rendezvous, we may need to account for other forces, such as thrusting, drag on lower satellites, solar-radiation pressure on higher satellites, and so on. Thus,

$$\ddot{\vec{r}}_{int} = -\frac{\mu \vec{r}_{int}}{r_{int}^3} + \vec{F}_{thrust} \quad (5-46)$$

From Fig. 5-25, we find the relative range vector from the target to the interceptor, \vec{r}_{rel} , through vector operations:

$$\vec{r}_{rel} = \vec{r}_{int} - \vec{r}_{tgt}$$

Differentiating twice yields

$$\ddot{\vec{r}}_{rel} = \ddot{\vec{r}}_{int} - \ddot{\vec{r}}_{tgt}$$

Substituting the two-body equations of motion—Eq. (5-45) and Eq. (5-46)—gives us

$$\ddot{\vec{r}}_{rel} = -\frac{\mu \vec{r}_{int}}{r_{int}^3} + \vec{F}_{thrust} + \frac{\mu \vec{r}_{tgt}}{r_{tgt}^3} \quad (5-47)$$

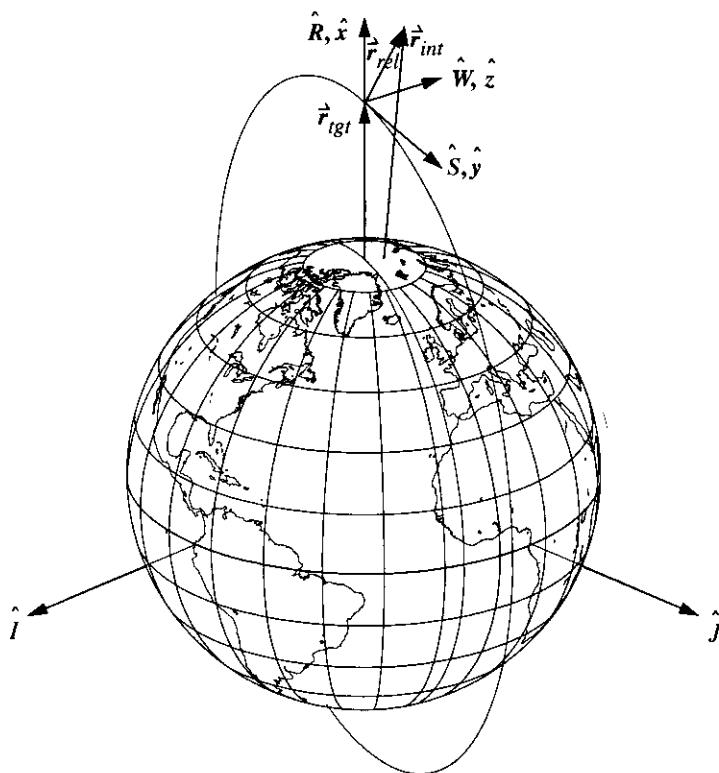


Figure 5-25. Coordinate-System Geometry for Relative Motion. The xyz notation reminds us the results are approximate. Remember that the S (and y) axes will be aligned with the velocity vector only in circular orbits. A major assumption in relative motion is that the target satellite is in a nearly circular orbit.

To be useful, this equation needs a little work. Our choice of coordinate systems makes a difference. I'm using Kaplan's coordinate system. You can change it to match Clohessy-Wiltshire, or any other coordinate system, but you'll need to reformulate the equations of motion from this point on. First, let's solve the relative-range vector equation for the interceptor:

$$\vec{r}_{int} = \vec{r}_{tgt} + \vec{r}_{rel}$$

Next, we need an expression for the interceptor's position vector, divided by the magnitude of this vector cubed. Using the cosine law for sides of an oblique triangle, find

$$\frac{\vec{r}_{int}}{r_{int}^3} = \frac{\vec{r}_{tgt} + \vec{r}_{rel}}{(r_{tgt}^2 + 2\vec{r}_{tgt} \cdot \vec{r}_{rel} + r_{rel}^2)^{3/2}}$$

Notice you find the magnitude in the denominator using the cosine law, with the dot product replacing the cosine of the included angle θ in the expression. Next, simplify by assuming the magnitude of the relative vector, r_{rel}^2 , is small compared to r_{tgt}^2 , and factoring out the r_{tgt}^2 term:

$$\frac{\vec{r}_{int}}{r_{int}^3} = \frac{\vec{r}_{tgt} + \vec{r}_{rel}}{r_{tgt}^3} \left\{ \frac{1}{\left(1 + \frac{2\vec{r}_{tgt} \cdot \vec{r}_{rel}}{r_{tgt}^2} \right)^{3/2}} \right\}$$

Now simplify using a binomial series on the dot-product term. Assume a series of the form,

$$(1+x)^n = 1 + nx + \frac{n(n-1)x^2}{2!} + \dots$$

and let x be the dot-product term of the denominator, so that

$$\frac{\vec{r}_{int}}{r_{int}^3} = \frac{\vec{r}_{tgt} + \vec{r}_{rel}}{r_{tgt}^3} \left\{ 1 - \frac{3}{2} \left(\frac{2\vec{r}_{tgt} \cdot \vec{r}_{rel}}{r_{tgt}^2} \right) + \dots \right\}$$

Now substitute this result into Eq. (5-47):

$$\ddot{\vec{r}}_{rel} = -\mu \left(\frac{\vec{r}_{tgt} + \vec{r}_{rel}}{r_{tgt}^3} \left\{ 1 - \frac{3}{2} \left(\frac{2\vec{r}_{tgt} \cdot \vec{r}_{rel}}{r_{tgt}^2} \right) + \dots \right\} \right) + \dot{\vec{F}}_{thrust} + \frac{\mu \vec{r}_{tgt}}{r_{tgt}^3}$$

Expanding and removing terms of opposite signs and keeping only first-order terms gives you

$$\ddot{\vec{r}}_{rel} = -\frac{\mu}{r_{tgt}^3} \left\{ -\frac{3}{2} \left(\frac{2\vec{r}_{tgt} \cdot \vec{r}_{rel}}{r_{tgt}^2} \right) + \dot{\vec{r}}_{rel} - \frac{3\vec{r}_{rel}}{2} \left(\frac{2\vec{r}_{tgt} \cdot \vec{r}_{rel}}{r_{tgt}^2} \right) \right\} + \dot{\vec{F}}_{thrust}$$

Here it's possible to reduce some terms using another assumption about the product of \vec{r}_{rel} and the dot product. Assume this quantity is small and drop it to simplify the problem:

$$\ddot{\vec{r}}_{rel} = -\frac{\mu}{r_{tgt}^3} \left\{ -\frac{3\dot{\vec{r}}_{tgt}}{2r_{tgt}} \left(\frac{2\dot{\vec{r}}_{tgt} \cdot \vec{r}_{rel}}{r_{tgt}} \right) + \dot{\vec{r}}_{rel} \right\} + \vec{F}_{thrust}$$

I've rearranged this equation to show unit vectors, \hat{R} , in the direction of the target satellite. The unit vector and relative vector dot product yield only the x -component. Thus,

$$\ddot{\vec{r}}_{rel} = -\frac{\mu}{r_{tgt}^3} \{ -3x\hat{R} + \vec{r}_{rel} \} + \vec{F}_{thrust} \quad (5-48)$$

This expression represents the *inertial* acceleration of the interceptor in the target frame. However, because this frame is moving, we must analyze further. In general, because the coordinate system is rotating, we may differentiate the position vector, but remember the coordinate system now changes with time. The solution is to use Eq. (1-30), which relates inertial and relative vectors in a rotating frame. Therefore,

$$\ddot{\vec{r}}_I = \ddot{\vec{r}}_R + \dot{\vec{\omega}}_R \times \vec{r}_R + 2\vec{\omega}_R \times \dot{\vec{r}}_R + \vec{\omega}_R \times (\dot{\vec{\omega}}_R \times \vec{r}_R)$$

Each of the four added terms on the right side of this equation is recognizable. The first term is the acceleration with respect to the rotating coordinate system. The next term is the tangential acceleration due to changing ω , which is zero for a circular orbit. The next is the Coriolis acceleration, and the last is the centripetal acceleration due to the angle between $\vec{\omega}_R$ and \vec{r} . Solving for the relative acceleration gives us

$$\ddot{\vec{r}}_R = \ddot{\vec{r}}_I - \dot{\vec{\omega}}_R \times \vec{r}_R - 2\vec{\omega}_R \times \dot{\vec{r}}_R - \vec{\omega}_R \times (\dot{\vec{\omega}}_R \times \vec{r}_R)$$

Writing this in terms of the vectors in this section results in

$$\begin{aligned} \ddot{\vec{r}}_{rel_R} &= \ddot{\vec{r}}_{rel_I} - \dot{\vec{\omega}}_R \times \vec{r}_{rel} - 2\vec{\omega}_R \times \dot{\vec{r}}_{rel} - \vec{\omega}_R \times (\dot{\vec{\omega}}_R \times \vec{r}_{rel}) \\ \omega &= \sqrt{\frac{\mu}{r_{tgt}^3}} \end{aligned} \quad (5-49)$$

Notice the angular rate is simply the target satellite's mean motion because the orbits are circular. Now, let's determine the cross products for the chosen coordinate system. To simplify the notation, we'll let x , y , and z represent the three components of \vec{r}_{rel} in the target's coordinate system:

$$\dot{\vec{\omega}}_R \times \vec{r}_{rel} = \begin{bmatrix} \hat{R} & \hat{S} & \hat{W} \\ 0 & 0 & \dot{\omega} \\ x & y & z \end{bmatrix} = -\dot{\omega}y\hat{R} + \dot{\omega}x\hat{S}$$

$$\vec{\omega}_R \times \dot{\vec{r}}_{rel} = \begin{bmatrix} \hat{R} & \hat{S} & \hat{W} \\ 0 & 0 & \omega \\ \dot{x} & \dot{y} & \dot{z} \end{bmatrix} = -\omega \dot{y} \hat{R} + \omega \dot{x} \hat{S}$$

$$\begin{aligned} \vec{\omega}_R \times (\vec{\omega}_R \times \dot{\vec{r}}_{rel}) &= \vec{\omega} \times \begin{bmatrix} \hat{R} & \hat{S} & \hat{W} \\ 0 & 0 & \omega \\ x & y & z \end{bmatrix} = \vec{\omega} \times (-\omega y \hat{R} + \omega x \hat{S}) \\ &= \begin{bmatrix} \hat{R} & \hat{S} & \hat{W} \\ 0 & 0 & \omega \\ -\omega y & \omega x & 0 \end{bmatrix} = -\omega^2 x (\hat{R} - \omega^2 y \hat{S}) \end{aligned}$$

Substituting these values and the derived value for the inertial acceleration of the interceptor [Eq. (5-48)] into the relative acceleration [Eq. (5-49)] results in the generalized equations of motion:

$$\ddot{\vec{r}}_{rel_R} = -\frac{\mu}{r_{tgt}^3} \{ \dot{\vec{r}}_{rel} - 3x\hat{R} \} + \vec{F}_{thrust} + \dot{\omega}y\hat{R} - \dot{\omega}x\hat{S} - 2(-\omega \dot{y}\hat{R} + \omega \dot{x}\hat{S}) + \omega^2 x\hat{R} + \omega^2 y\hat{S}$$

If we assume circular motion, the constant in the first term is simply ω^2 and $\dot{\omega} = 0$. Expanding the terms gives us

$$\ddot{\vec{r}}_{rel_R} = -\omega^2 \{ x\hat{R} + y\hat{S} + z\hat{W} - 3x\hat{R} \} + \vec{F}_{thrust} + 2\omega \dot{y}\hat{R} - 2\omega \dot{x}\hat{S} + \omega^2 x\hat{R} + \omega^2 y\hat{S}$$

Now we write each vector component separately and simplify F_{thrust} (as f_i) to aid future analysis. This yields the results known as *Hill's equations* or the *Clohessy-Wiltshire equations*:

$$\left. \begin{aligned} \ddot{x} - 2\omega \dot{y} - 3\omega^2 x &= f_x \\ \ddot{y} + 2\omega \dot{x} &= f_y \\ \ddot{z} + \omega^2 z &= f_z \end{aligned} \right\} \text{near-circular orbits} \quad (5-50)$$

Note the types of acceleration in these equations. In the f_x equation—from left to right—they are total, Coriolis, and centripetal acceleration.

5.8.1 Position Solutions for Nearly Circular Orbits

To solve Hill's equations, let's assume no external forces are imparted by the interceptor satellite—i.e., $F_{thrust} = 0$. This is actually a good assumption because we may consider any Δv impulsive, with the resulting velocity serving as an initial condition to restart the analysis. This assumption *doesn't* allow analysis of continuous low thrust to achieve a rendezvous.

One approach to solve these equations is to use Laplace operators. Starting with Eq. (5-50), first take the derivative of the x equation, without any external forces and with $\dot{\omega}$ assumed to be zero. In this case,

$$\ddot{x} = 3\omega^2x + 2\omega\dot{y}$$

$$\ddot{x} = 3\omega^2\dot{x} + 2\omega\ddot{y}$$

Substitute $\ddot{y} = -2\omega\dot{x}$ into \ddot{x} to get

$$\ddot{x} = 3\omega^2\dot{x} - 4\omega^2\dot{x}$$

$$\ddot{x} + \omega^2\dot{x} = 0$$

Take the Laplace transform so that

$$\mathcal{L}[\ddot{x} + \omega^2\dot{x}] = \{s^3X(s) - s^2x_o - s\dot{x}_o - \ddot{x}_o\} + \omega^2\{sX(s) - x_o\} = 0$$

and expand to get

$$s^3X(s) - s^2x_o - s\dot{x}_o - \ddot{x}_o + sX(s)\omega^2 - x_o\omega^2 = 0$$

Group terms to reduce this expression to

$$sX(s)(s^2 + \omega^2) = (s^2 + \omega^2)x_o + s\dot{x}_o + \ddot{x}_o$$

Now solve for $X(s)$:

$$X(s) = \frac{x_o}{s} + \frac{\dot{x}_o}{(s^2 + \omega^2)} + \frac{\ddot{x}_o}{s(s^2 + \omega^2)}$$

To help convert this expression back to the time domain, expand the last term as a partial fraction. Don't forget to include the single s term in $Bs + C$ to make the solution perfectly general.

$$\frac{\ddot{x}_o}{s(s^2 + \omega^2)} = \frac{A}{s} + \frac{Bs + C}{s^2 + \omega^2}$$

Find the coefficient A by multiplying through by s and letting $s = 0$:

$$A = \frac{\ddot{x}_o}{\omega^2}$$

Find B by putting the right-hand side of the equation over a common denominator. Then, equate both numerators using the value of A above to find

$$B = -\frac{\ddot{x}_o}{\omega^2} \quad C = 0.0$$

Substitute to get

$$X(s) = \frac{x_o}{s} + \frac{\dot{x}_o}{(s^2 + \omega^2)} + \frac{\ddot{x}_o}{s\omega^2} - \frac{s\ddot{x}_o}{\omega^2(s^2 + \omega^2)}$$

Take the inverse Laplace transform to arrive at the time-domain result:

$$x(t) = x_o + \frac{\ddot{x}_o}{\omega^2} + \frac{\dot{x}_o}{\omega} \sin(\omega t) - \frac{\ddot{x}_o}{\omega^2} \cos(\omega t)$$

Substitute \ddot{x}_o from Eq. (5-50) (with no external forces) to produce

$$x(t) = x_o + \frac{3\omega^2 x_o + 2\omega \dot{y}_o}{\omega^2} + \frac{\dot{x}_o}{\omega} \sin(\omega t) - \frac{3\omega^2 x_o + 2\omega \dot{y}_o}{\omega^2} \cos(\omega t)$$

and simplify so that

$$x(t) = x_o + 3x_o + \frac{2\dot{y}_o}{\omega} + \frac{\dot{x}_o}{\omega} \sin(\omega t) - \left(3x_o + \frac{2\dot{y}_o}{\omega} \right) \cos(\omega t)$$

and

$$x(t) = 4x_o + \frac{2\dot{y}_o}{\omega} + \frac{\dot{x}_o}{\omega} \sin(\omega t) - \left(3x_o + \frac{2\dot{y}_o}{\omega} \right) \cos(\omega t)$$

You can find the y solution by integrating the original Hill's equation [Eq. (5-50)] for y with respect to time, but you need \dot{x} to do so. Remembering the initial values are constant, differentiate the $x(t)$ solution above:

$$\dot{x}(t) = \frac{\dot{x}_o}{\omega} \omega \cos(\omega t) + \left(3x_o + \frac{2\dot{y}_o}{\omega} \right) \omega \sin(\omega t)$$

Simplifying gives you

$$\dot{x}(t) = \dot{x}_o \cos(\omega t) + (3\omega x_o + 2\dot{y}_o) \sin(\omega t)$$

Now, substitute this expression into the \ddot{y} equation:

$$\ddot{y} = -2\omega [\dot{x}_o \cos(\omega t) + (3\omega x_o + 2\dot{y}_o) \sin(\omega t)]$$

$$\ddot{y} = -2\omega \dot{x}_o \cos(\omega t) - 2\omega (3\omega x_o + 2\dot{y}_o) \sin(\omega t)$$

Integrate to find \dot{y} and y :

$$\dot{y} = -2\dot{x}_o \sin(\omega t) + 2(3\omega x_o + 2\dot{y}_o) \cos(\omega t) + C$$

and

$$y = \frac{2\dot{x}_o}{\omega} \cos(\omega t) + \left(6x_o + \frac{4\dot{y}_o}{\omega} \right) \sin(\omega t) + Ct + D$$

To determine the constants of integration, substitute $t = 0$ into the \dot{y} and y equations:

$$\dot{y}_o = 6\omega x_o + 4\dot{y}_o + C \quad C = -6\omega x_o - 3\dot{y}_o$$

$$y_o = \frac{2\dot{x}_o}{\omega} + D \quad D = -\frac{2\dot{x}_o}{\omega} + y_o$$

The answer is then

$$y(t) = \frac{2\dot{x}_o}{\omega} \cos(\omega t) + \left(6x_o + \frac{4\dot{y}_o}{\omega} \right) \sin(\omega t) - (6\omega x_o + 3\dot{y}_o)t - \frac{2\dot{x}_o}{\omega} + y_o$$

The solution to the z equation is much simpler because it is uncoupled. This equation is usually identified as a *simple harmonic oscillator*:

$$\ddot{z} + \omega^2 z = 0$$

Using the Laplace transform again gives you

$$s^2 Z(s) - sz_o - \dot{z}_o + \omega^2 Z(s) = 0$$

$$Z(s)(s^2 + \omega^2) = sz_o + \dot{z}_o$$

$$Z(s) = \frac{sz_o}{(s^2 + \omega^2)} + \frac{\dot{z}_o}{(s^2 + \omega^2)}$$

so the solution in the time domain is

$$z(t) = z_o \cos(\omega t) + \frac{\dot{z}_o}{\omega} \sin(\omega t)$$

The presence of sine and cosine terms indicate some type of oscillatory motion.

Shortly, we'll see the resulting motion is somewhat *elliptical*. Although the interceptor doesn't actually "orbit" the target satellite, the instantaneous parameters often result in an elliptical-like motion. The term multiplied by time in the y equation represents the drift of the interceptor with respect to the target and is the reason the interceptor's path isn't truly elliptical—due mainly to small differences in the semimajor axis of the interceptor's and the target's paths. Finally, the last terms in x and y represent the initial displacement of the interceptor's average position from the target (we'll talk about this later). We can differentiate the position equations to obtain the velocity with respect to the target, remembering the initial values do not change.

Summarizing now, let's combine the position and velocity equations into a computational algorithm.

ALGORITHM 44: Hill's Equations ($x_o, y_o, z_o, \dot{x}_o, \dot{y}_o, \dot{z}_o$ of the interceptor,

$\omega_{tgt}, \Delta t \Rightarrow x, y, z, \dot{x}, \dot{y}, \dot{z}$ of the interceptor)

$$\begin{aligned} x(t) &= \frac{\dot{x}_o}{\omega} \sin(\omega t) - \left(3x_o + \frac{2\dot{y}_o}{\omega} \right) \cos(\omega t) + \left(4x_o + \frac{2\dot{y}_o}{\omega} \right) \\ y(t) &= \left(6x_o + \frac{4\dot{y}_o}{\omega} \right) \sin(\omega t) + \frac{2\dot{x}_o}{\omega} \cos(\omega t) - (6\omega x_o + 3\dot{y}_o) t + \left(y_o - \frac{2\dot{x}_o}{\omega} \right) \\ z(t) &= z_o \cos(\omega t) + \frac{\dot{z}_o}{\omega} \sin(\omega t) \end{aligned}$$

$$\dot{x}(t) = \dot{x}_o \cos(\omega t) + (3\omega x_o + 2\dot{y}_o) \sin(\omega t)$$

$$\dot{y}(t) = (6\omega x_o + 4\dot{y}_o) \cos(\omega t) - 2\dot{x}_o \sin(\omega t) - (6\omega x_o + 3\dot{y}_o)$$

$$\dot{z}(t) = -z_o \omega \sin(\omega t) + \dot{z}_o \cos(\omega t)$$

An example shows Hill's equations in operation:

▼ **Example 5-13. Scenario Using the Hubble Space Telescope.**

GIVEN: The Hubble Space Telescope is about to be released from the Space Shuttle, which is in a circular orbit at 590 km altitude. The relative velocity (from the Space Shuttle bay) of the ejection is 0.1 m/s up, 0.04 m/s backwards, and 0.02 m/s to the right.

FIND: Position and velocity of the Hubble Space Telescope after 5 and 20 minutes.

First, convert the given data to the correct sign convention for the relative motion definitions. The initial velocities are then

$$\dot{x}_o = 0.1 \quad \dot{y}_o = -0.04 \quad \dot{z}_o = -0.02 \text{ m/s}$$

Next, find the angular rate of the target (Shuttle). In a 590 km altitude orbit,

$$\omega = \sqrt{\frac{\mu}{r_{int}^3}} = \sqrt{\frac{398,600.5}{(6968.1363)^3}} = 0.001\,085\,4 \text{ rad/s}$$

The initial position of 0.0 simplifies the equations in Algorithm 44 to

$$x(t) = \frac{\dot{x}_o}{\omega} \sin(\omega t) - \frac{2\dot{y}_o}{\omega} \cos(\omega t) + \frac{2\dot{z}_o}{\omega}$$

$$y(t) = \frac{4\dot{x}_o}{\omega} \sin(\omega t) + \frac{2\dot{y}_o}{\omega} \cos(\omega t) - 3\dot{y}_o t - \frac{2\dot{x}_o}{\omega}$$

$$z(t) = \frac{\dot{z}_o}{\omega} \sin(\omega t)$$

$$\dot{x}(t) = \dot{x}_o \cos(\omega t) + 2\dot{y}_o \sin(\omega t)$$

$$\dot{y}(t) = 4\dot{x}_o \cos(\omega t) - 2\dot{y}_o \sin(\omega t) - 3\dot{y}_o$$

$$\dot{z}(t) = \dot{z}_o \cos(\omega t)$$

Substituting the time of 5 minutes (300 s) yields

$$x(5) = -33.345 \text{ m} \quad \dot{x}(5) = -0.12 \text{ m/s}$$

$$y(5) = -1.473 \text{ m} \quad \dot{y}(5) = 0.03 \text{ m/s}$$

$$z(5) = -5.894 \text{ m} \quad \dot{z}(5) = -0.02 \text{ m/s}$$

and, for the case of 20 minutes (1200 s),

$$x(20) = -143.000 \text{ m} \quad \dot{x}(20) = -0.10 \text{ m/s}$$

$$y(20) = 137.279 \text{ m} \quad \dot{y}(20) = 0.27 \text{ m/s}$$

$$z(20) = -17.766 \text{ m} \quad \dot{z}(20) = -0.01 \text{ m/s}$$

▲ Notice how both the position and velocity change dramatically.

5.8.2 Trend Analysis

We can use the equations in Algorithm 44 to get information about the relative motion of the two satellites. A helpful first step is to determine, in general, what happens for each initial condition. As mentioned earlier, the interceptor moves in an *elliptical-like* pattern around the target satellite. For some applications, it's convenient to calculate the size and shape (C and D) and location (x_c, y_c) of this ellipse. Remember, the target is not at the center of the ellipse. In addition, we should balance all of these trends with the accuracy analysis of Sec. 5.8.3. By examining the x or y positions in Algorithm 44, we can introduce some definitions to simplify the results. Define C as

$$C \equiv \sqrt{\left(3x_o + \frac{2\dot{y}_o}{\omega}\right)^2 + \left(\frac{\dot{x}_o}{\omega}\right)^2}$$

We'll see shortly that this is really the magnitude of the oscillation. The x equation is now

$$x(t) = C \left\{ \frac{\dot{x}_o}{\omega} \sin(\omega t) - \frac{3x_o + \frac{2\dot{y}_o}{\omega}}{C} \cos(\omega t) \right\} + 4x_o + \frac{2\dot{y}_o}{\omega}$$

Define the two trigonometric terms:

$$\text{SIN}(\psi_o) \equiv -\frac{3x_o + \frac{2\dot{y}_o}{\omega}}{C} \quad \text{COS}(\psi_o) \equiv \frac{\dot{x}_o}{\omega}$$

Substitute to get

$$x(t) = C \{ \text{COS}(\psi_o) \text{SIN}(\omega t) + \text{SIN}(\psi_o) \text{COS}(\omega t) \} + 4x_o + \frac{2\dot{y}_o}{\omega}$$

Recognizing the trigonometric summation identity [Eq. (C-13)] and performing similar operations on $y(t)$ result in

$$\begin{aligned} x &= x_c + C \text{SIN}(\omega t + \psi_o) \\ y &= y_{c_o} + \dot{y}_c t + 2C \text{COS}(\omega t + \psi_o) \end{aligned} \tag{5-51}$$

I've identified the coordinates for the center of the ellipse as x_c and y_c . Because of the drift term, \dot{y}_c , the *ellipse* is instantaneous and the value for y_c is good only at time $t = 0$; hence, the notation y_{c_o} . The x_c displacement is constant over time and doesn't need to be further distinguished. The initial displacement terms in Algorithm 44 give the center's position:

$$y_{c_o} = y_o - \frac{2\dot{x}_o}{\omega} \quad x_c = 4x_o + \frac{2\dot{y}_o}{\omega}$$

Notice the drift term in the y direction, which means

$$\dot{y}_c = -6\omega x_o - 3\dot{y}_o$$

Solving this expression for \dot{y}_o and substituting into the x_c equation result in

$$x_c = 4x_o + \frac{2(-6\omega x_o - \dot{y}_c)}{3\omega}$$

$$x_c = 4x_o - 4x_o - \frac{2\dot{y}_c}{3\omega} = -\frac{2\dot{y}_c}{3\omega}$$

Let's also use trigonometric substitutions to solve the z equation. Again, begin with definitions for the z oscillations and trigonometric terms:

$$D \equiv \sqrt{\left(\frac{\dot{z}_o}{\omega}\right)^2 + z_o^2}$$

$$\sin(\phi_o) \equiv \frac{\dot{z}_o}{D} \quad \cos(\phi_o) \equiv \frac{z_o}{D}$$

Now, rewrite the general form for the z equation using Eq. (C-13):

$$z = D \cos(\omega t + \phi_o) \quad (5-52)$$

Find the center z_c by substituting $t = 0$ into Algorithm 44, or $z_c = z_o$. Summarize the equations above to get

$$\begin{aligned} y_{c_o} &= y_o - \frac{2\dot{x}_o}{\omega} & C &= \sqrt{\left(3x_o + \frac{2\dot{y}_o}{\omega}\right)^2 + \left(\frac{\dot{x}_o}{\omega}\right)^2} \\ \dot{y}_c &= -6\omega x_o - 3\dot{y}_o & D &= \sqrt{\left(\frac{\dot{z}_o}{\omega}\right)^2 + z_o^2} \\ x_c &= -\frac{2\dot{y}_c}{3\omega} & z_c &= z_o \end{aligned} \quad (5-53)$$

where C = the magnitude of the variation in x (radial position) or the semiminor axis
 $2C$ = the instantaneous magnitude of the y variation or the semimajor axis
 D = the z offset

Also notice the x_c and z_c equations are constant because they depend only on initial conditions. This simply means there is no drift in the x and z directions, unlike y_c , which

includes a drift term (\dot{y}_c). If we write the equations of motion [Eq. (5-51)] in a moving frame, we can separate terms as follows:

$$x - x_c = C \sin(\omega t + \psi_o)$$

$$y - y_{c_o} = 2C \cos(\omega t + \psi_o)$$

Thus, we have equations of motion that represent an ellipse in the x - y plane.

$$\frac{x^2}{C^2} + \frac{y^2}{4C^2} = 1$$

Similar analysis with Eq. (5-52) indicates a circle in the y - z plane. The preceding results permit us to draw an elliptical cylinder of the interceptor's motion (Fig. 5-26).

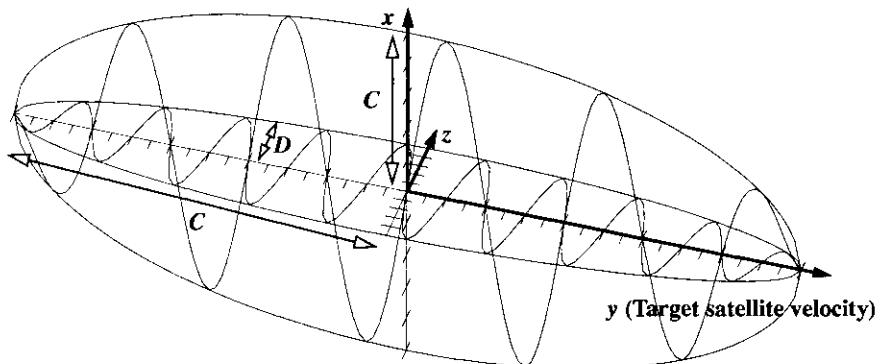


Figure 5-26. Interceptor's Instantaneous Relative Motion about the Target. We can define an elliptical region in which the interceptor “orbits” about the target satellite. (This is similar to a prolate spheroid, discussed in Sec. 1.3.2.) Although it’s not actually an orbit, we can approximate the motion as an ellipse for short time spans. Also recognize that this is really an irregular ellipsoidal region and is exaggerated in this figure. The periodic curves are illustrative and aren’t representative of actual motion.

Remember the drift term will displace this motion either to the right or left, depending on the sign of $(6\omega x_o + 3\dot{y}_o)$. Now we can use Eq. (5-53) to determine some features of this ellipse and examine results using a set of initial conditions. Let’s continue with a circular satellite for the target, in a 222 km altitude orbit ($r_{tgt} = 6378.1363 + 222 \text{ km} = 6600.1363 \text{ km}$). To find how changing initial conditions will affect the size and shape of the ellipse, we begin by finding the angular rate of the target satellite:

$$\omega = \sqrt{\frac{\mu}{r_{tgt}^3}} = \sqrt{\frac{398,600.5}{6600.1363^3}} = 0.001\ 177\ 44 \text{ rad/s}$$

We can examine the x and y directions independently of z because their motion doesn't couple with z . The analysis in Fig. 5-27 comes from a computer program that graphically displayed the results of the interceptor motion with respect to the target satellite. By analyzing this motion, we can generalize to determine the direction and type of motion, depending on the interceptor satellite's initial position and velocity in relation to the target.

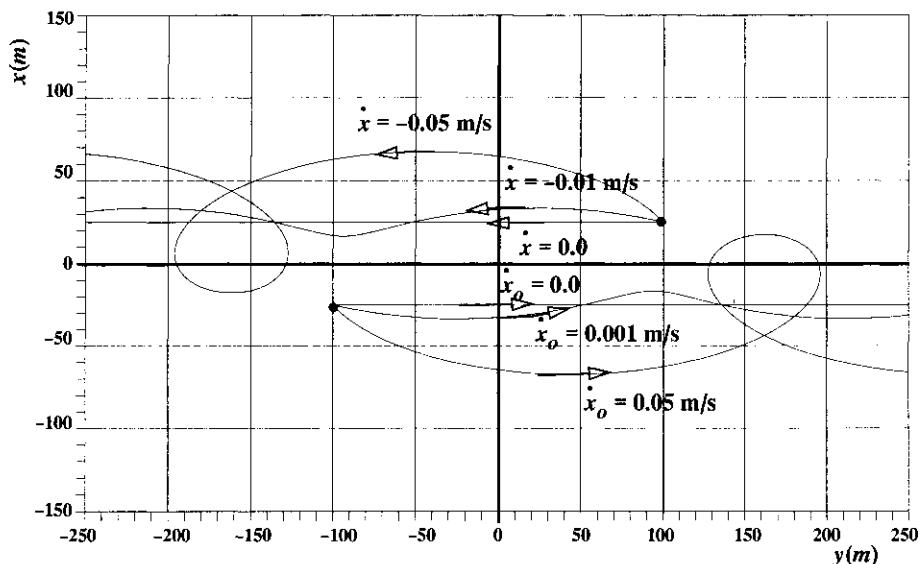


Figure 5-27. Motion of Interceptor for x_o and y_o Displacements, \dot{x}_o fixed and \dot{y}_o and \dot{x}_o Variations. Varying the interceptor's initial conditions gives us several interesting motions, including straight lines produced by $\dot{y}_o = 0.04415 \text{ m/s}$.

The first rule derives from basic orbital mechanics. For any orbit, a larger semimajor axis will result in a longer period and a slower velocity. If we limit our study, for the moment, to orbits with very small eccentricities, the satellite in the higher orbit will appear to move *backward* in relation to the lower because its period is longer and its velocity is slower. Figure 5-27 illustrates the motion of two different interceptor satellites relative to a target and three combinations of initial velocities. The two interceptors are offset from the target by a 25 m altitude difference (one above, one below). Notice each plot of the higher satellite moves backward, whereas each plot of the lower-altitude satellite moves forward.

The straight-line cases depend on giving the interceptor an initial y velocity equal to $-(3/2)x_o\omega$. The plots of the wobbles and loops result from using the same initial conditions and varying the x velocity. A 0.01 m/s increase in \dot{x}_o produces the wobbles, and 0.05 m/s additional x velocity creates the loops. The wobbles and loops which appear in the interceptor's motion result from changing the interceptor's orbit from circular to slightly eccentric, but it needs more velocity to be circular!

We may now explore each component to determine its effect on the motion. Begin by examining the x direction. Notice that changes in x represent a different period from the target because the semimajor axis has changed. In addition, the initial velocity is the same as the target's, but the interceptor is in a higher (or lower) orbit, meaning it's in a slightly eccentric orbit. Using the results from Fig. 5-27, you would expect loops due to the eccentricity and a backward or forward motion depending on the initial x displacement. Figure 5-28 shows various initial x displacements, positive and negative, between 5 and 500 m.

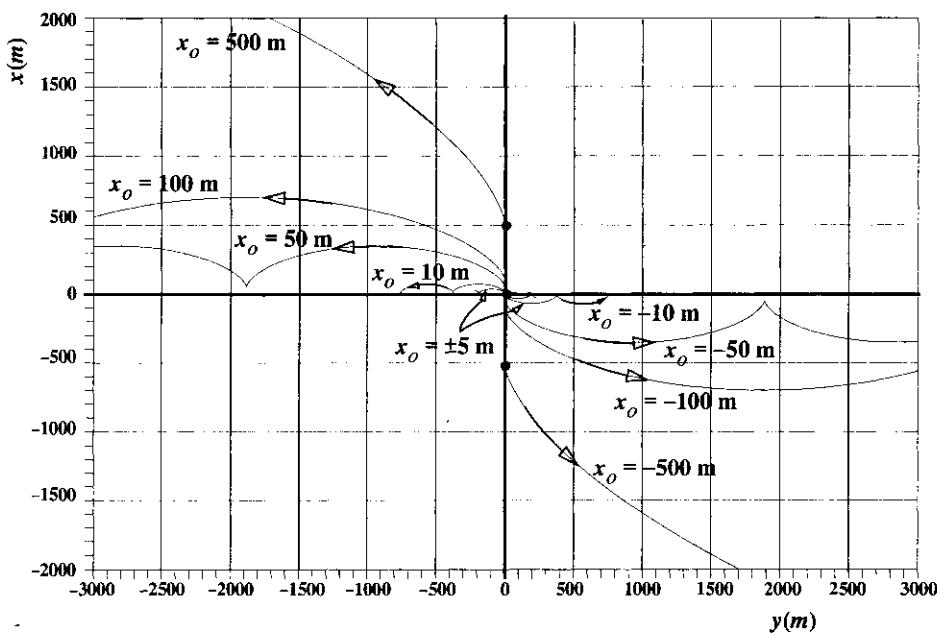


Figure 5-28. Motion of the Interceptor for Various x_o Displacements. Notice how all displacements above the satellite move to the left (lower velocity causes a lagging effect), whereas those below move to the right (higher velocity causes a lead). In addition, as the satellite's initial location gets farther from the target, the motion away from the target increases.

You can also deduce the results of Fig. 5-28 using the formulas given in Eq. (5-53). From these equations, *any nonzero* value of x_o will increase the size of the *ellipse*. C may be zero only if certain combinations of initial conditions are met. Notice that you must differentiate the last two relations to show solutions where you allow x_o and \dot{y}_o , respectively, to vary. Thus,

$$\begin{aligned}\dot{x}_o &= \dot{y}_o = x_o = 0 \\ \text{or} \\ \dot{x}_o &= 0 \quad \dot{y}_o = -\frac{3}{2}x_o\omega \\ \text{or} \\ \dot{x}_o &= 0 \quad x_o = \frac{-2\dot{y}_o}{3\omega}\end{aligned}\tag{5-54}$$

The final case allows a displaced value of x_o to have no *semimajor axis* (C), even though it's displaced from the target satellite. This is the result used to make the straight lines in Fig. 5-27.

As x_o grows from 5 to 500 m from the target in Fig. 5-28 (positive and negative), notice how quickly the size of the ellipse grows! For displacements above the target, the motion is always to the left, whereas negative displacements are always to the right. This is exactly what we predicted. The slight eccentricity of the interceptor's orbit creates the wobbles. Because an elliptical orbit doesn't have a constant velocity throughout the orbit, the interceptor will appear to move faster and slower relative to the target's constant motion; thus, the loops.

Next, look at the effects produced by changes in the y direction. Plots of only y_o displacements aren't very interesting because the interceptor is in the same orbit and doesn't move relative to the target. Any plot is simply a dot at the initial y displacement point.

Turning to changes in velocity, *any nonzero* value of \dot{x}_o will produce an *osculating ellipse* (Fig. 5-29). The osculating nature arises from the fact that the interceptor's orbit will periodically "kiss" or touch the target's orbit. A burn in the x direction won't change the period, but instead will change the eccentricity. Because the change in x velocity occurs at the target satellite, the rest of the interceptor's motion will be toward apogee or perigee depending on the sign of \dot{x}_o (positive or negative, respectively). The two satellites will drift apart for half a period, then back together. Because the period is the same, both satellites will rendezvous every period.

Nonzero values of \dot{y}_o will also produce an *elliptical motion* (Fig. 5-30). Both the mathematics from Eq. (5-53) and physical understanding show this to be true. Any velocity change in the y direction will change the eccentricity and the semimajor axis (period). If velocity is added, the initial point becomes perigee, and the period and eccentricity increase. Because the period has changed, the interceptor will rendezvous with the target satellite's altitude in one period of the *interceptor*. As shown in the figure, the rendezvous point moves away from the target satellite depending on the amount of the change in the period. All increases to velocity will ultimately fall to the left, whereas decreases will all travel to the right in the figure.

The final area to explore in the x and y directions is the effect of combined changes. The easiest effect is with y displacements and either x or y velocity changes. Because the interceptor starts in the same orbit, the plots are exactly the same as Figs. 5-29 and 5-30, except they are displaced to the initial y position. Note that certain values of x velocity,

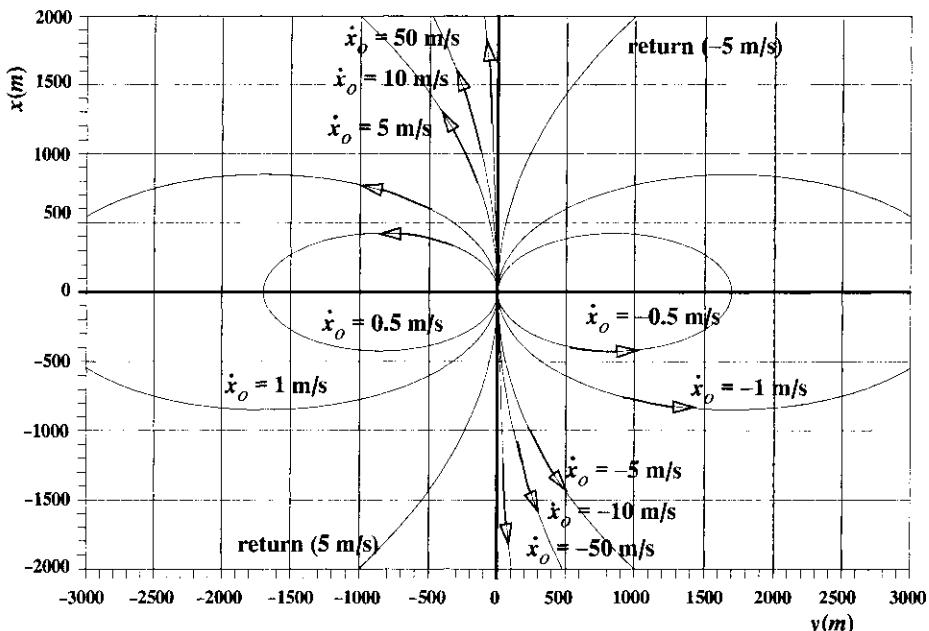


Figure 5-29. Motion of the Interceptor for \dot{x}_o Variations. All motions which result from variations in the x velocity become periodic motions about the target. Notice each path will return to the target in one period.

combined with y displacements, determine the initial center of the ellipse. From Eq. (5-53), it's apparent that \dot{y}_o and \dot{x}_o are the only two variables that affect this center. If

$$y_o = \frac{2\dot{x}_o}{\omega}$$

the initial center of the ellipse will be about the target. Any other value of \dot{x}_o will simply make the ellipse larger because there are no other values to cause a drift. We can visualize this effect by "sliding" the motions in Fig. 5-30 to the left or right (a y displacement), until one of the ellipses is centered about the target satellite (has the correct x velocity).

Coupled changes of x displacement and x velocity are a little more complicated. Figure 5-31 shows various x velocities for an initial 50 m displacement in x . The initial x displacement means the interceptor is in a slightly eccentric orbit and will therefore drift generally to the left or behind the target. The change in x velocity simply magnifies the change in eccentricity. From Eq. (5-53), the initial x displacement means the drift, y_c , will be constant and the initial center of the ellipse, y_{c_o} , will be ahead or behind the target, depending on the sign of x_o .

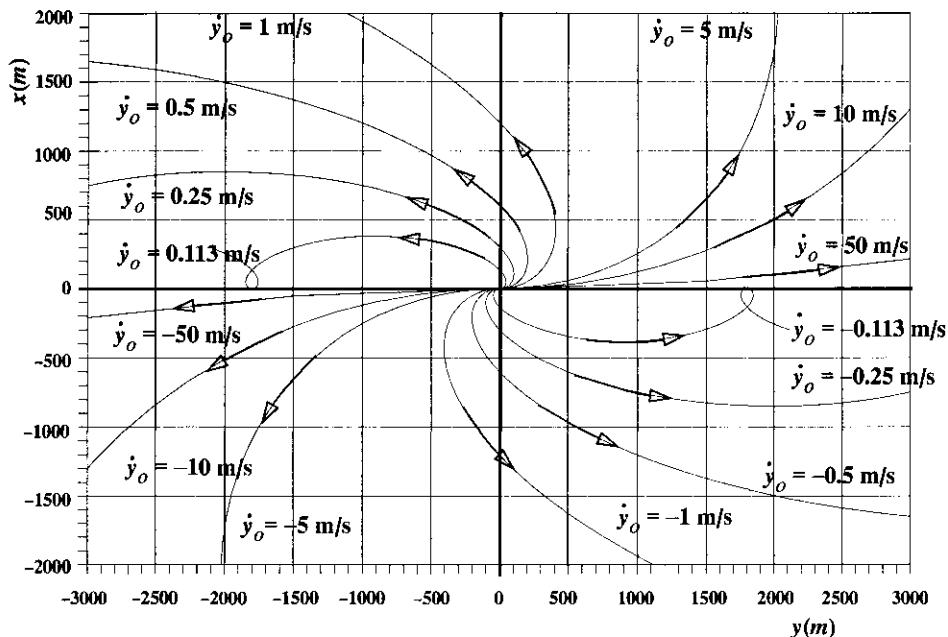


Figure 5-30. Motion of the Interceptor for \dot{y}_o Variations. All positive y velocity changes cause the interceptor to travel left, whereas negative velocities cause motion to the right. Although $\dot{y}_o = 50 \text{ m/s}$ appears to travel to the right, it will curve back around to the left over time.

Note that we'd get an exact mirror image if the initial x displacement had been negative. Also, some of the negative \dot{x}_o velocities appear to be moving in the same direction as the target, but over time, they'll loop back and move to the left, behind the target.

Perhaps most interesting is the combined effect of x displacements and y velocity changes. It turns out that the variation of x_o and \dot{y}_o solely determines the direction the interceptor will drift with respect to the target. As mentioned before, if the initial x displacement is negative, we'd expect all paths to drift ahead of the target (to the right). However, depending on the value of \dot{y}_o , the resulting motion may be either to the right, the left, or in "orbit" about the target. We can solve the \dot{y}_c equation for the initial y velocity (\dot{y}_o) that results in zero drift ($\dot{y}_c = 0$). The example shown in Fig. 5-32 uses a 50 m displacement for x . In this case,

$$\dot{y}_o = -2\omega x_o = -2(0.001\ 177\ 44)(50) = -0.117\ 744 \text{ m/s}$$

Initial y velocities larger than this value ($\dot{y}_o > -2\omega x_o$) will yield negative y drift rates, whereas smaller velocities ($\dot{y}_o < -2\omega x_o$) will result in positive drifts.

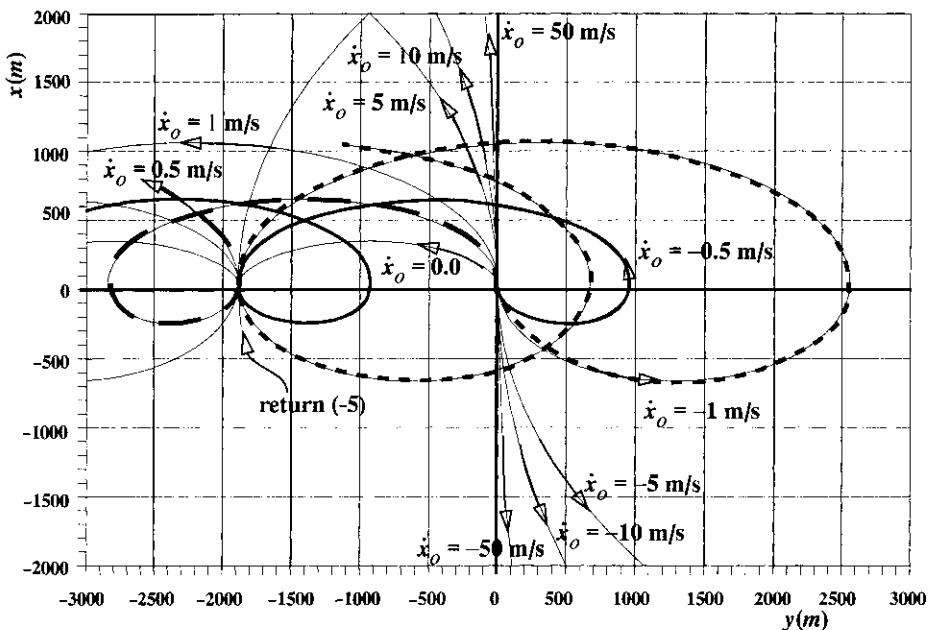


Figure 5-31. Motion of the Interceptor for x_o Displacements and \dot{x}_o Variations. This figure represents the most complex motion of the individual changes. The graph is based on a constant offset of 50 m for x_o , with velocity changes in the x direction as indicated. Notice how many of the curves converge at the same point (near the return). Eventually, all paths will end up drifting behind (to the left).

It's also interesting to note that the size of the *ellipse* is not always symmetrical about the y axis. As the *ellipse* stretches out, the drift term decreases the size of the semiminor axis.

Finally, the analysis for the z direction is relatively simple because there is no coupling with x and y . Examination of D in Eq. (5-53) shows that *any* initial z velocity will increase the amount of the variation in z and that any initial displacement in z will have a corresponding oscillation. We can simply add the results of the z variation to those of x and y . Figure 5-33 shows this combination. The x - y motion is the same as the loops in Fig. 5-27, except for an initial 80 m offset in the z direction. Notice how the x - y motion is unaffected, and the z motion is strictly oscillatory.

Let's summarize trends for initial values:

1. The interceptor will tend to drift *ahead* of the target if it begins below it.
2. The interceptor will drift *behind* the target if it begins above it.
3. The larger the displacement in x_o , the larger the size of the relative-motion plot.

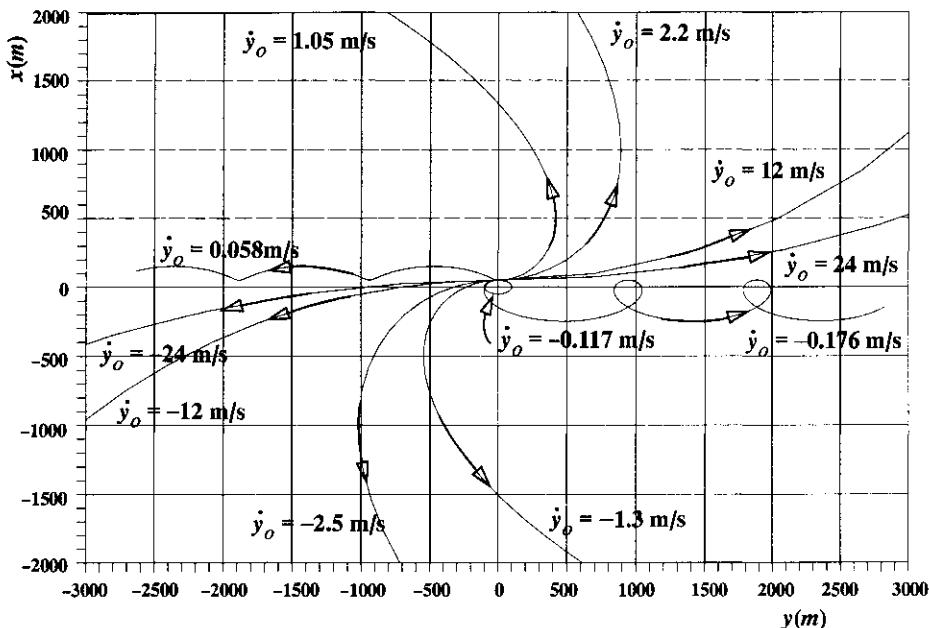


Figure 5-32. Motion of the Interceptor for x_o Displacements and \dot{y}_o Variations. Each of the variations above comes from a 50 m x_o displacement coupled with varying values of y velocity as shown. Notice the special case of an “orbit” around the target satellite at -0.117 m/s.

4. Adjusting the value of \dot{y}_o to $\dot{y}_o = -2\omega x_o$ will stop any drift in the ellipse.
5. Any amount of displacement in y causes identical motions because the interceptor begins in the same orbit.
6. Any z values superimpose an oscillatory motion over the x - y motion.

Another useful technique is to find analytical equations for the initial velocity in each component, especially for rendezvous operations of spacecraft close to one another. To do so, solve Algorithm 44 for position to get the initial velocity. Rearrange the z equation to solve for \dot{z}_o . Then, solve the coupled y and x equations simultaneously to get \dot{y}_o and \dot{x}_o . Writing the equations equal to zero (meaning the interceptor has docked with the target satellite) will yield

$$\frac{\dot{x}_o}{\omega} \sin(\omega t) - \left(3x_o + \frac{2\dot{y}_o}{\omega} \right) \cos(\omega t) + \left(4x_o + \frac{2\dot{y}_o}{\omega} \right) = 0$$

$$\left(6x_o + \frac{4\dot{y}_o}{\omega} \right) \sin(\omega t) + \frac{2\dot{x}_o}{\omega} \cos(\omega t) - (6\omega x_o + 3\dot{y}_o) t + \left(y_o - \frac{2\dot{x}_o}{\omega} \right) = 0$$

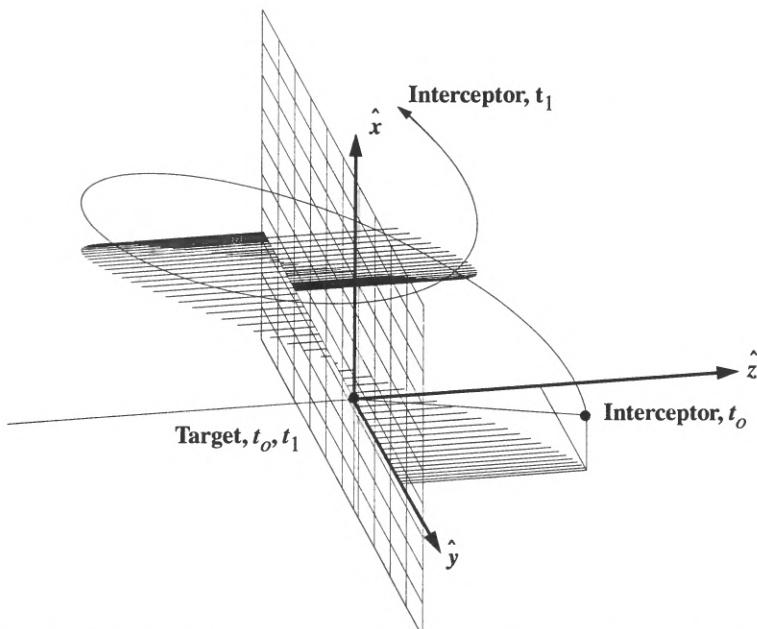


Figure 5-33. Relative Motion of the Interceptor with a z Variation. Adding a z variation to any of the other motions described simply adds a three-dimensional component. In this example, the interceptor goes above and then below the target satellite.

$$z_o \cos(\omega t) + \frac{\dot{z}_o}{\omega} \sin(\omega t) = 0$$

Although not immediately apparent, substituting as follows will allow you to simplify these equations:

$$\begin{aligned} A &= \frac{4}{\omega} \sin(\omega t) - 3t & B &= \frac{2}{\omega} \cos(\omega t) - \frac{2}{\omega} \\ C &= y_o + 6x_o \sin(\omega t) - 6\omega x_o t & D &= \frac{\sin(\omega t)}{\omega} \\ E &= 4x_o - 3x_o \cos(\omega t) \end{aligned}$$

Using these definitions in the equations above results in

$$-B\dot{y}_o + D\dot{x}_o + E = 0$$

$$A\dot{y}_o + B\dot{x}_o + C = 0$$

Multiply the first equation by $(-\frac{B}{D})$ and add to the second equation to get

$$\left(\frac{B^2}{D} + A\right)\dot{y}_o + \left(-\frac{BE}{D} + C\right) = 0$$

Then, solve for

$$\dot{y}_o = \frac{BE - CD}{B^2 + AD}$$

and use this result to find

$$\dot{x}_o = -\frac{E - B\dot{y}_o}{D}$$

To find the expressions, substitute the definitions into the two velocity equations above, so

$$\dot{y}_o = \frac{\left(\frac{2}{\omega}\cos(\omega t) - \frac{2}{\omega}\right)(4x_o - 3x_o \cos(\omega t)) - (y_o + 6x_o \sin(\omega t) - 6\omega x_o t)\left(\frac{\sin(\omega t)}{\omega}\right)}{\left(\frac{2}{\omega}\cos(\omega t) - \frac{2}{\omega}\right)^2 + \left(\frac{4}{\omega}\sin(\omega t) - 3t\right)\left(\frac{\sin(\omega t)}{\omega}\right)}$$

You'll want to obtain similar coefficients in the final answer to compute more efficiently. The terms needed are $(1 - \cos(\omega t))$ and $(4 - 3\cos(\omega t))$. Multiply by ω and factor the 2 from the first term in the numerator and denominator. Expanding the rest of the numerator yields

$$\dot{y}_o = \frac{-2\omega x_o (1 - \cos(\omega t)) (4 - 3\cos(\omega t)) - \left(y_o \omega \sin(\omega t) + 6x_o \omega \sin^2(\omega t) - 6\omega^2 x_o t \sin(\omega t)\right)}{4(1 - \cos(\omega t))^2 + (4\sin(\omega t) - 3\omega t)(\sin(\omega t))}$$

Factoring out an $\omega \sin(\omega t)$ in the numerator, and then a $6x_o$, results in

$$\dot{y}_o = \frac{-2\omega x_o (1 - \cos(\omega t)) (4 - 3\cos(\omega t)) - \left(\omega \sin(\omega t) \{y_o + 6x_o (\sin(\omega t) - \omega t)\}\right)}{4(1 - \cos(\omega t))^2 + (4\sin(\omega t) - 3\omega t)(\sin(\omega t))}$$

Solve the x equation in the same way, using \dot{y}_o :

$$\dot{x}_o = -\frac{(4x_o - 3x_o \cos(\omega t)) - \left(\frac{2}{\omega}\cos(\omega t) - \frac{2}{\omega}\right)\dot{y}_o}{\frac{\sin(\omega t)}{\omega}}$$

Multiply through by ω ; then factor out an ωx_o and a -2 in the numerator to get

$$\dot{x}_o = -\frac{\omega x_o (4 - 3\cos(\omega t)) + 2(1 - \cos(\omega t))\dot{y}_o}{\sin(\omega t)}$$

Because the equation for z isn't coupled with either of the other equations, you can find \dot{z}_o individually:

$$z_o \cos(\omega t) + \frac{\dot{z}_o}{\omega} \sin(\omega t) = 0$$

$$\frac{\dot{z}_o}{\omega} \sin(\omega t) = -z_o \cos(\omega t)$$

$$\dot{z}_o = -z_o \omega \cot(\omega t)$$

Thus, the required velocity values are

$$\begin{aligned}\dot{y}_o &= \frac{(6x_o(\omega t - \sin(\omega t)) - y_o)\omega \sin(\omega t) - 2\omega x_o(4 - 3\cos(\omega t))(1 - \cos(\omega t))}{(4\sin(\omega t) - 3\omega t)\sin(\omega t) + 4(1 - \cos(\omega t))^2} \\ \dot{x}_o &= -\frac{\omega x_o(4 - 3\cos(\omega t)) + 2(1 - \cos(\omega t))\dot{y}_o}{\sin(\omega t)} \\ \dot{z}_o &= -z_o \omega \cot(\omega t)\end{aligned}\quad (5-55)$$

Let's look at an example of this technique.

▼ Example 5-14. Solving for Velocity in Hill's Equation.

GIVEN: A satellite in circular orbit at 590 km altitude. Suppose we continue Example 5-13; at 10 minutes, we decide to retrieve the Hubble Space Telescope.

FIND: Δv to rendezvous with the Shuttle in 5 and 15 minutes.

Examining Eq. (5-55) reveals quantities similar to those in Example 5-13:

$$\omega = 0.001\ 085\ 4\ \text{rad/s}$$

Now, solve for the required velocity to rendezvous [Eq. (5-55)] in 5 and 15 minutes (300 s and 900 s), but be sure to calculate y_o first because you need it to calculate x_o .

At $t = 5$ minutes (300 seconds), At $t = 15$ minutes (900 seconds),

$$\dot{y}_o = 0.0134\ \text{m/s} \quad \dot{y}_o = 0.0752\ \text{m/s}$$

$$\dot{x}_o = 0.2742\ \text{m/s} \quad \dot{x}_o = 0.1356\ \text{m/s}$$

$$\dot{z}_o = 0.3591\ \text{m/s} \quad \dot{z}_o = 0.0082\ \text{m/s}$$

Notice how much the required velocity drops as the allowable time to rendezvous increases.

This problem has a practical application—rescuing astronauts who become detached from the space station. Williams and Baughman (1994) show how these calculations are used for designing self-rescue modules for astronauts.

As we've developed equations in this section, we have relied on assumptions that are worth summarizing:

1. The two satellites are within a few kilometers of each other (see the next section). This allowed a simplified calculation of the magnitude of the interceptor, using only the x component of the relative-position vector.
2. We used a binomial series to find an expression for the position magnitude of the interceptor and assumed a first-order relation.
3. Because we assumed the satellites were sufficiently close (#1 above), we could assume the product of the relative-range vector and the dot product of the target and relative vector was small.
4. The target is in a circular orbit. Although we didn't calculate the appropriate equations for elliptical orbits, we should be able to derive them, given the discussion in this chapter. But we need more math because we took out ω early in the analysis.
5. There are no external forces on the interceptor. Any Δv s are impulsive, and the resulting velocity is an initial condition for the problem. We could also solve for these values by straightforwardly extending the derivation presented here.
6. We must use the target's position vector to obtain a latitude and local side-real time for the interceptor when rotating the relative-range vector. This is necessary because we don't yet know the interceptor's range vector from the center of the Earth.

5.8.3 Extending the Results

The two-dimensional assumption in this section poses an immediate question: how can we extend the results to position and velocity vectors? The answer is difficult to quantify, but we can tackle it using known techniques. The primary condition used in the relative-motion equations is that we examine the difference in the orbits and consider the interceptor's motion with respect to the target satellite. The obvious choice for comparison is to examine the satellites as two independent vehicles orbiting the Earth, which, in fact, is the more realistic scenario for the motion. We detailed techniques for modeling the satellite's motion in Sec. 4.3 while formulating Kepler's problem and will use them for perturbed orbits in Chap. 8. Either method will work. Using vectors allows us to determine quickly and efficiently the "true" relative distances over time. The analysis also provides the means to check the limits of accuracy for the relative-motion equations.

The xyz system we use for relative motion can be misleading in its portrayal of the actual relative differences. The x component is correctly modeled, but the y and z components are more subtle. Consider an interceptor satellite which is 100 m ahead of the target satellite. Figure 5-34 indicates this 100-m displacement actually has a y and an x component due to the shape of the target satellite's orbit.

The immediate question we must solve is how to convert between frames—between the true vectorial positions, and the linearized, or relative Hill's equation, positions.

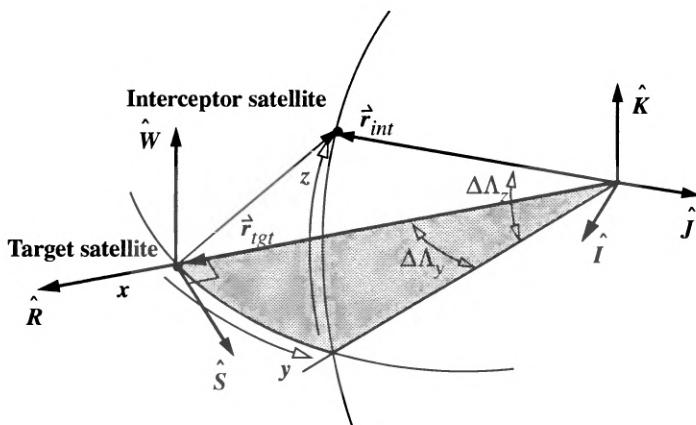


Figure 5-34. Relative Position of Interceptor and Target. This diagram is extremely exaggerated to illustrate the inconsistency in assuming the y and z component displacements transfer directly to position and velocity vectors.

Remember that we must always have the IJK state vector of the target satellite to find an answer.

We'll examine the case for which we know the interceptor's position and velocity in the relative, RSW , frame. To determine the position and velocity vectors in the geocentric, IJK , frame, we must first know the target satellite's geocentric position vector. We determine the interceptor's position and velocity vectors by initializing them to the target satellite's vector in the RSW frame. A series of rotations is then necessary to "add in" the effect of the y and z offsets, as shown in Fig. 5-34. The process is shown below. Note the use of the transformation matrix, which employs the RSW unit vectors (for the target satellite) from Eq. (1-26). I'll use a state vector notation to simplify the equations, so

$$\mathbf{X} = \begin{bmatrix} \mathbf{r} \\ \mathbf{v} \end{bmatrix}$$

IJK to Hill's

Hill's to IJK

$$[\mathbf{X}_{tgt}]_{RSW} \Rightarrow \hat{\mathbf{R}} \hat{\mathbf{S}} \hat{\mathbf{W}}$$

$$[\mathbf{X}_{tgt}]_{IJK} \Rightarrow \hat{\mathbf{R}} \hat{\mathbf{S}} \hat{\mathbf{W}}$$

$$[\mathbf{X}_{int}]_{RSW} = [\hat{\mathbf{R}} : \hat{\mathbf{S}} : \hat{\mathbf{W}}]^T [\mathbf{X}_{int}]_{IJK}$$

$$[\mathbf{X}_{int}]_{RSW} = [\hat{\mathbf{R}} : \hat{\mathbf{S}} : \hat{\mathbf{W}}]^T [\mathbf{X}_{tgt}]_{IJK}$$

$$[X_{tgt}]_{RSW} = [\hat{R} \mid \hat{S} \mid \hat{W}]^T [X_{tgt}]_{IJK}$$

$$\dot{\vec{r}}_{int_R} = \dot{\vec{r}}_{int_R} + \vec{x}$$

$$\dot{\vec{v}}_{int} = \dot{\vec{v}}_{int} + \dot{\vec{v}}_{rel}$$

$$\boxed{\begin{array}{l} \text{IF } |r_{int_w}| > 0.0 \\ \\ \end{array}}$$

$$\Delta\Lambda_z = \text{TAN}^{-1}\left(\frac{r_{int_w}}{r_{tgt}}\right)$$

$$[X_{int}] = \text{ROT2}(-\Delta\Lambda_z) [X_{int}]$$

$$\boxed{\begin{array}{l} \text{IF } |y| > 0.0 \\ \\ \end{array}}$$

$$\Delta\Lambda_y = \text{TAN}^{-1}\left(\frac{y}{r_{tgt}}\right)$$

$$[X_{int}] = \text{ROT2}(-\Delta\Lambda_y) [X_{int}]$$

$$\boxed{\begin{array}{l} \text{IF } |r_{int_s}| > 0.0 \\ \\ \end{array}}$$

$$\Delta\Lambda_y = \text{TAN}^{-1}\left(\frac{r_{int_s}}{r_{tgt}}\right)$$

$$[X_{int}] = \text{ROT3}(\Delta\Lambda_y) [X_{int}]$$

$$\boxed{\begin{array}{l} \text{IF } |z| > 0.0 \\ \\ \end{array}}$$

$$\Delta\Lambda_z = \text{TAN}^{-1}\left(\frac{z}{r_{tgt}}\right)$$

$$[X_{int}] = \text{ROT3}(\Delta\Lambda_z) [X_{int}]$$

$$\dot{\vec{r}}_{int_hill} = \begin{bmatrix} r_{int_I} - r_{tgt_I} \\ \Delta\Lambda_y r_{int} \\ \Delta\Lambda_z r_{int} \end{bmatrix}$$

$$[X_{int}]_{IJK} = [\hat{R} \mid \hat{S} \mid \hat{W}] [X_{int}]_{RSW}$$

$$\dot{\vec{v}}_{int_hill} = \dot{\vec{v}}_{int} - \dot{\vec{v}}_{tgt}$$

The reverse process is similar and still requires a transformation from the *IJK* to the *RSW* coordinate system.

Problems

1. The Space Shuttle ($i = 28.5^\circ$) often uses intermediate orbits to achieve its final parking orbit. Assume the solid rocket motors place the Shuttle at apogee of a 0 km by 110 km altitude orbit. The Shuttle then uses a 100 km by 300 km Hohmann transfer orbit to achieve the final orbit (300 km by 300 km). If the Shuttle remains in each orbit for only $\frac{1}{2}$ revolution, how long after the first burn (to leave the 0 km by 110 km initial orbit) will it take the Shuttle to achieve its final orbit? How much change in velocity is needed? If the Shuttle needs to revisit the Hubble Space Telescope (590 km altitude circular orbit, $i = 28.5^\circ$), how much more velocity does it need?
2. Derive the alternate expression for the change in inclination angle when the application of Δv is not along the normal component.
3. A satellite in a circular orbit has enough fuel to double its semimajor axis through a two-burn Hohmann transfer. How much plane change can the satellite make? (Hint: consider minimum calculations)
4. A satellite is about to be launched from Cape Kennedy into a Molniya orbit. Its orbital elements are $a = 26,500$ km, $e = 0.7$, $i = 63.4^\circ$, $\Omega = 45^\circ$, $\omega = 90^\circ$. The launch booster can place the satellite in a circular orbit at an altitude of 300 km. If a *Payload Assist Module*, PAM-D [from *Jane's* (1994, 265)] weighing 2025 kg can deliver 67.16 kN of thrust for 88.1 seconds, develop a launch-window schedule for orbit insertion between July 30 and September 2, 1999.
5. Find initial costate values of the Lagrange multipliers for the continuous-thrust problem in Example 5-11 [Hint: see Thorne and Hall (1996)].
6. Hint: This problem mainly uses techniques from this chapter but ultimately requires a solution of the Lambert problem discussed in Chap. 6. You are working as a junior engineer for NASA. The U.S. has launched the Space Shuttle for a scientific mission. The orbital elements are: $a = 6563.3696$ km, $e = 0.004\ 806\ 0$, $i = 45.0^\circ$, $\Omega = 95.26\ 277\ 50^\circ$, $\omega = 50.375\ 078\ 0^\circ$, $M = 0.0^\circ$ on the epoch of December 24, 1998, at 0 h UT, Julian date = 2,448,980.500. The Russian space shuttle (Buran) is about to be launched and has two satellites on board: a tracking and relay satellite (Gorizont) and a new rescue module for their MIR space station. An unknown piece of debris hits and cripples the U.S. Shuttle, threatening the lives of the astronauts. Your job is to determine how soon the experimental Russian module can retrieve the astronauts. Several restrictions apply:
 - a. We know the injection longitude of Buran to be 80.0° W, and the elapsed time from launch is 75.667 min for standard launches. This will place Buran at apogee on the initial revolution of a 0 km \times 80 km orbit. In addition, Buran will be launched from Tyuratam, located at 46.0° N, 65.0° E. The orbit will be a standard

circular orbit with $i = 51.6^\circ$. The sequence to achieve orbit is: 1: $(0 \text{ km} \times 80 \text{ km})$, 2: $(80 \times 364 \text{ km})$, 3: $(364 \times 364 \text{ km})$. Buran can remain in orbit for only two days after launch and can't maneuver. It must launch on December 24.

- b. The Gorizont satellite will occupy a geostationary orbit at $43.5^\circ\text{E} \pm 3^\circ$ and must be on station *before* December 26, 1998, 12 h *UTC*. The Gorizont satellite has a fixed Δv motor capable of 2819.4 m/s to start the combined Δv transfer to geosynchronous. Ninety minutes are required to check out the Gorizont immediately before deployment. The Gorizont satellite must use 0 to 6 phasing revolutions to minimize transfer time. We like to start the transfer orbit at the ascending node to allow tracking stations in the Northern Hemisphere to monitor the transfer. Any motor firing must be within sight of a U.S. ground station to enhance detection and initial orbit determination. The Gorizont satellite requires Sun acquisition for alignment at perigee motor firing. The Sun sensor's cone of view has a width of 25° and can be configured to view from 45° to 135° off the satellite's spin axis. The configuration is established before launch and can't be changed in orbit. The satellite is deployed at the node before starting the transfer to allow the orbiter to obtain proper separation and avoid plume impingement. This deployment *should* also occur in sight of a ground station but doesn't have to.
 - c. The rescue module uses the Gorizont tracking satellite for data; thus, the Gorizont must deploy first, and the module can't leave Buran until Gorizont is in place at geosynchronous. The ten minutes from rescue-module removal to motor firing must occur in sight of the Sun and Gorizont. The rescue module's perigee motor for the first transfer Δv can provide up to 3657.6 m/s; the apogee motor can produce 2600 m/s. Time of flight from the rescue-module deployment to rendezvous must not exceed 90 minutes because of limited oxygen supplies (suggested minimum time is 30 min).
7. Section 5.8.3 shows how we can transform results to position and velocity vectors. Analyze the results in Fig. 5-28. How accurate are they? Hint: use a propagation technique from Chap. 7 or Chap. 8 to check your answers.
 8. Using a classical formulation of the Hohmann transfer, is it more difficult to transfer from a low-Earth orbit (200 km circular altitude) to an orbit at the Moon's altitude, or to geosynchronous orbit? Explain your answer.
 9. Find the achievable inclinations for a direct launch from the sites listed in Table 5-2.

CHAPTER 6 INITIAL ORBIT DETERMINATION

- 6.1 Historical Background
- 6.2 Obtaining Data
- 6.3 Observations of Range, Azimuth, and Elevation
- 6.4 Angles-only Observations
- 6.5 Mixed Observations
- 6.6 Three Position Vectors and Time
- 6.7 Two Position Vectors and Time—Lambert’s Problem
- 6.8 Application: Targeting Problem

6.1 Historical Background

It's especially interesting to think how much of today's analysis scientists did before artificial satellites even existed! Indeed, most astrodynamics problems require us to get initial data and develop a state vector. Obtaining physical constants was more difficult in the past, but it remains difficult for high-precision studies today. First, we must find an accurate initial orbit or state vector from which to do future studies. We define the overall topic of ***initial orbit determination*** as the first-time development of orbital elements for a body in motion in space. For this book, our main application is the Earth-orbiting satellite; however, historical applications focused entirely on the motion of moons, planets, and asteroids.

In a sense, the first determinations of initial orbits were performed by scientists who produced astronomical tables of the positions of the planets and stars (ephemerides). Of note is the work of Johann Kepler (1571–1630) and his *Rudolphine Tables* in 1627. Kepler's data was accurate enough to predict the conjunctions of Mercury and the Sun to within five hours on November 7, 1631—a success that placed the Rudolphine tables securely in history.

Newton devised and described in his *Principia* the first practical method of finding the orbit of a body from three observations. The comet of 1680 provided an opportunity to use Newton's new technique. Georg Dörffel suggested a parabolic orbit for the comet but was unable to predict its true path. Newton succeeded using his new method. This was the first true determination of a comet's parabolic orbit from three observations. The most publicized result of applying Newton's method of orbit determination belongs to Edmond Halley (1656–1742), who discovered the first periodic comet—Halley's comet.

Newton's method was difficult to understand, so there were many misapplications in the decades after its original publication. Halley did understand the technique. He wrote Newton in 1687 and suggested Newton use the technique to examine the orbits of other comets. But Halley didn't try his own suggestion until 1695. Among several interesting results, he was particularly interested in the comet of 1682, observed by John Flamsteed (1646–1719). His original work suggested it might be the same comet that appeared in 1531 and 1607, but he needed more observations. By 1696, he had obtained Flamsteed's observations (through Newton) and was now ready to analyze the orbits of all well-observed comets.

In 1702, Halley finished his work. One might wonder about the long delay, but Halley had many other civil duties during this time, including determining longitude while at sea on several ocean voyages as well as diplomatic obligations. Finally, in 1705, Halley published his famous treatise on comets for 1337 to 1698. In his report, Halley says

"Many considerations incline me to believe the comet of 1531 observed by Apianus to have been the same as that described by Kepler and Longomontanus in 1607 and which I again observed when it returned in 1682. All the elements agree,... whence I would venture confidently to predict its return, namely in the year 1758." (Bate, Mueller and White, 1971, 52)

Years later, he changed his words slightly—probably to acknowledge the difficulty of the problem and the numerous unknowns, such as Jupiter's effect on the orbit. In his tables of 1749 and 1752, Halley changed his confident prediction to "it should return again about the year 1758" (Yeomans, 1991, 122–123).

Halley's comet did reappear in late 1758. Modern investigations have chronicled an entire series of earlier appearances of Halley's comet—the earliest in 466 B.C. The comet has appeared since in 1835, 1910, and 1986; its next visit will be in 2061.

Newton's method of determining a parabolic orbit from observations depended on successive graphical approximations. Yeomans (1991, 103) suggests that some controversy exists concerning the accuracy of Newton's method because some of his answers are too good. Whether or not Newton combined calculus or other techniques with his graphical method doesn't matter. The first completely analytical method for solving the same problem was given by Leonhard Euler (1707–1783) in 1744 in his *Theoria Motuum Planetarum et Cometarum (Theory of the Motion of Planets and Comets)*. Euler was a true mathematician and had a grand reputation. His technique required three relatively closely spaced observations. In 1743, Euler also discovered the equation connecting two radius vectors and the subtended chord of the parabola with the interval of time. The resulting infinite series was complex, but accurate.

In his works of 1761 to 1771, Johann H. Lambert (1728–1779) generalized Euler's formulas to include elliptical and hyperbolic orbits. Because Lambert was a geometer at heart, he seldom offered analytical solutions. His insight was often responsible for developments by other scientists. Today, solving for the orbit between two known position vectors is usually known as **Lambert's problem** and is very significant to astrodynamics. A lot of confusion exists in the literature because writers often refer to Lambert's problem as Gauss's problem. Although they're somewhat correct, Gauss's method is actually a solu-

tion Gauss developed for Lambert's problem. To be perfectly correct, we should identify it as Euler's problem because Euler first analyzed it in 1744, but history has favored Lambert's contributions.

Joseph Louis Lagrange (1736–1813) published important works on orbit determination: two in 1778 and one in 1783. He was just 16 when he completed his master's degree and became professor of mathematics at Turin. Just as Halley provided the spark which produced Newton's laws, a paper Halley wrote inclined the young Lagrange towards mathematics. A true mathematician, Lagrange brought a new level of elegance and generality to the subject. Despite his tremendous output and broad mathematical reputation, his recognition in astrodynamics is limited mainly to the stable solutions of the three-body problem, the Lagrange expansion theorem which provides a series representation of Kepler's equation, and the Lagrange planetary equations [Eq. (8-12)].

Pierre-Simon de Laplace (1749–1827) published an entirely new method of orbit determination based only on angles. It didn't require closely spaced observations. The work appeared in his papers, which were published in *Mémoires de l'Academie Royale des Sciences de Paris (Collected Works, Vol. X, 93–146)* in 1780. Laplace coined the term *celestial mechanics* and had a trademark of not using geometrical figures in his works (Bell, 1965, 154). Although there was enough observational data for Laplace's method in the 17th and 18th centuries, researchers couldn't compute the intensely mathematical solution. As we'll see in Sec. 6.4.1, this solution isn't trivial, especially without computers to do the complex calculations.

Karl Friedrich Gauss (1777–1855) was ultimately responsible for developing a practical technique for initial orbit determination. His motivation was the rediscovery of the minor planet Ceres in 1801 and 1802. On January 1, 1801, Giuseppe Piazzi (1746–1826) of Palermo discovered Ceres, although he first thought it was a comet. He continued observations until February 11, 1801, when the Sun obscured it. This short interval of time permitted observation of only about 9° of the orbit. The German astronomer Franz Xaver von Zach (1754–1832) was the editor of *Montaliche Correspondenz*, a monthly journal of scientific correspondence. In June 1801, he published the orbital positions of Ceres. Because Bode's law had suggested the existence of a planet or small asteroid between the orbits of Mars and Jupiter, the initial sighting generated a lot of interest in reobserving Ceres. In the September, 1801 issue of *Montaliche Correspondenz*, Gauss confidently announced his prediction that Ceres would reappear in late 1801 to early 1802, although his answers were quite different from others. On September 29, 1801, Gauss's *Disquisitiones Arithmeticae* (arithmetic researches) was published. This work established Gauss's mathematical ability, but his fame as an astronomer was still lacking. However, his prediction for Ceres generated considerable interest, and the search began. On December 7, 1801, both Zach and Heinrich Wilhelm Olbers (1758–1840, famous for his work dealing with comets) observed Ceres very near to Gauss's prediction. Buhler (1981, 47) suggests that Gauss's original prediction may have been the result of heuristic considerations, as opposed to rigorous mathematical theory. Whatever the case, Gauss had become a famous astronomer.

Gauss's method is just as pertinent today as it was in 1802, but for a different reason. Gauss used the right ascension and declination at three observation times to determine Ceres' orbit. His method is much simpler if the original data consists of two position vectors and the time of flight. We're interested in determining an orbit from two positions and time because the solution applies directly to problems of intercept, rendezvous, and ballistic-missile targeting. I will formally define the problem of orbit determination from two positions and time and give it the name of the original inventor—the *Lambert problem*.

Although not usually considered among the classical astrodynamacists, Josiah Gibbs (1839–1903) deserves our attention. He received his first engineering degree in 1863, and most of his work was in thermodynamics. His work in astrodynamics marks the first significant contribution from an American scientist. Although his technique is really only a geometrical solution, it still applies for initial calculations. Gibbs announced his discovery among papers published in 1889 as an improvement to Gauss's method for determining the position vectors at the first and third times of an observation. Samuel Herrick (1911–1974) developed an important derivative application—a Taylor-series approach to the Gibbs problem. It allows us to solve for very short time spans, which is necessary for modern radar systems and dense observational data. We'll discuss these methods later.

This chapter describes various analytical methods for determining initial orbits. Today, this problem is still relevant because sensor sites produce raw data that we must process into position and velocity vectors before determining an orbit. Although we use differential correction to actually determine the orbit (Chap. 9), we couldn't process data without the individual vectors determined through these techniques. Other methods exist, but you should know at least these basic relations and algorithms.

6.2 Obtaining Data

To see how to obtain data, we first need to examine the physics of the problem and become familiar with the electromagnetic spectrum shown in Fig. 6-1. We derive all data (range, azimuth, elevation) from signals of electromagnetic radiation, which travel near the speed of light in a vacuum, but slow somewhat depending on the medium they travel through. Remember, we can relate wavelength, λ , with frequency, f , and the speed of light, $c = 2.997\,924\,58 \times 10^{10}$ cm/s:

$$\lambda = \frac{c}{f}$$

Numerous variations tend to degrade the performance of any observations. We'll introduce just atmospheric and timing variations.

The atmosphere alters the time in which a signal travels from the satellite to the ground, or vice versa. The ionosphere (See Chap. 7) causes some large disturbances and may distort the signals. To a lesser extent, the troposphere also causes interference. Precise applications must account for these variations. The causes of the interactions aren't exactly known due to complex coupling with the Sun, the planets, and the Earth's gravitational and magnetic fields. Unfortunately, our knowledge of the atmosphere is less than perfect,

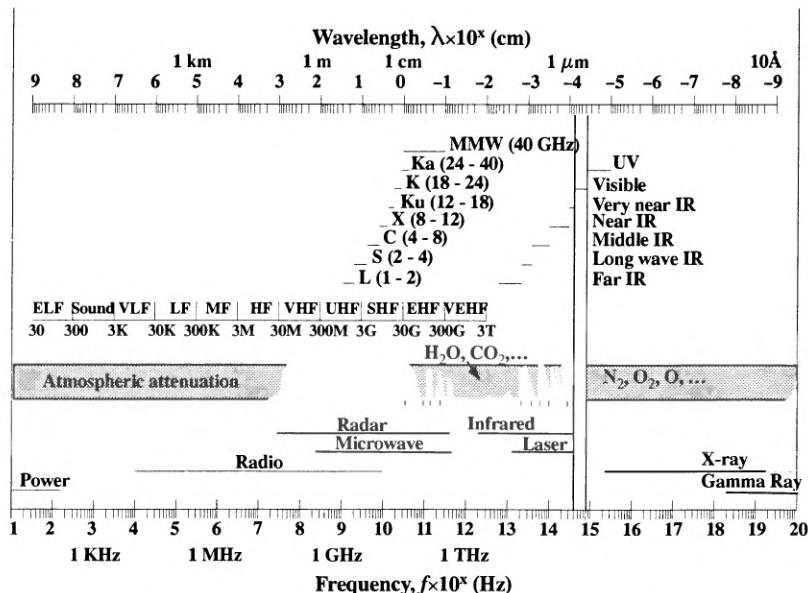


Figure 6-1. Electromagnetic Spectrum. Atmospheric attenuation partially blocks much of the incoming electromagnetic radiation, as the dark bar illustrates. Frequency values are given for various bands next to each letter descriptor. Values are in GHz, THz, and so on, as appropriate. [Hovanessian (1988, 5–7), Chetty (1988, 311), Gammon (1976, 17), Morgen & Gordon (1989, 233)]

making the task difficult. Rain and fog limit some applications, as Hovanessian (1988, 6–9) discusses in greater detail.

We could talk at length about variations in time. First, let's consider the use of standard time, *UTC*. To be precise, all observations should be *UTC*, but how? Special receivers can electronically receive broadcast *UTC*, but how does a site maintain them? Also, the site and satellite must know the exact offsets (if any) in their time systems to prevent additional errors. This is not an easy task!

Sensors receive one-way and two-way data. **One-way** signals are those which we simply measure at the sensor location. **Two-way** signals are those which a sensor receives as reflected or retransmitted waves. Signal strength is important, and retransmitted signals are usually best because there is no loss through reflection. **Retransmitted signals** occur whenever the satellite receives a signal and retransmits a new signal at some later time. Laser tracking has become popular in recent years because the signals can reflect off **corner-cube retro-reflectors**, which reflect the light back parallel to the incoming beam. We'll see later that typical wavelengths for lasers permit very high accuracy measurements (~ 1 mm) in range, compared to the range accuracy (~ 3 m) that's possible from the wavelengths of radar measurements. We can also consider these forms of data as one-way or two-way information, as shown in Fig. 6-2.

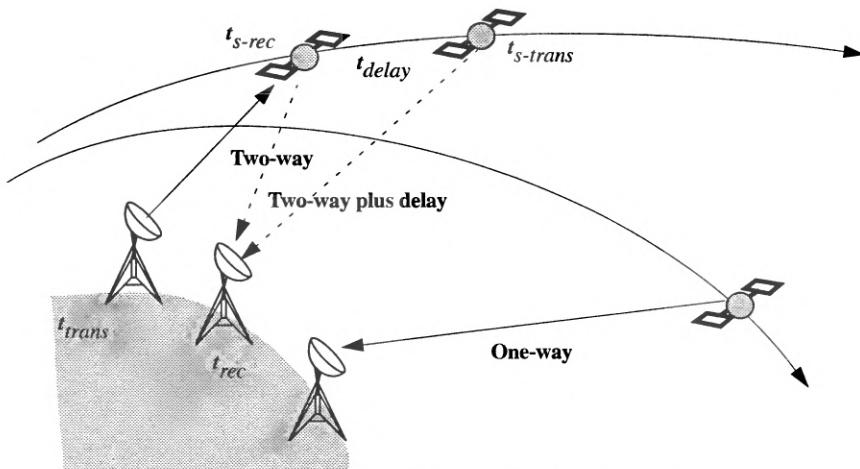


Figure 6-2. Geometry for One-way and Two-way Transmission. When a signal transmits to or from a satellite, we determine precise values of range from the speed of light and the total time (including any delays) for the trip. We must calculate physical and atmospheric distortions during each trip through the atmosphere. Although I've shown the two-way transmissions using two separate stations (*bi-static*) for clarity, the most common case is for the same station to transmit and receive (*mono-static*).

Figure 6-2 suggests we need to get three measurements: total time, direction of the signal, and Doppler shift. We'll discuss **Doppler shift** shortly, but for initial orbit determination, it allows us to determine the range-rate of the satellite by analyzing the frequency shift of the signals. The ultimate goal of most forms of tracking is to obtain range, range rate, and angular information for a satellite. But a particular technique may not produce all types of data. One-way data usually comes from signal analysis, which accounts for physical and atmospheric variations. Two-way data emerges from radars, transponders, and lasers. We'll introduce the two-way process because it highlights the concepts for each type of measurement, but you should consult Hovanessian (1988) or Long et al. (1989, Chap. 7) for more information. The preliminaries discussed here are from these sources.

We determine the time through mathematical operations. Notice the notation identifying signals for transmit, receive, send, and final receive. It's important to have all clocks defined to a known baseline. Depending on the tracking system involved, a transponder delay may or may not be present.

We find the range, ρ , using the total time and the signal speed divided by two. Thus,

$$\rho = \frac{c\Delta t}{2}$$

using the speed of light, c , and the total transfer time, Δt . For two-way systems that retransmit the signal, this expression becomes more complex. With $\Delta t = t_{rec} - t_{trans}$ and t_{s-rec} indicating the time that the satellite receives the signal,

$$\rho = |r_{sat}(t_{s-rec}) - r_{site}(t_{trans})| + |r_{site}(t_{rec}) - r_{sat}(t_{rec} + \Delta t)|$$

For the greatest accuracy, we must iterate on the transfer times to solve this equation. Remember that when the signal is received, the sender has traveled a distance of $v\Delta t$. Long et al. (1989, 7-13 to 7-14) discuss this in much greater detail.

The direction of the signal gives us angular measurements, which we determine partly from the signal, but in actuality, from the physical sensor. The latter process involves determining gimbal angles for the measuring device and may include antenna-mount models and local reference frames. Because determining angular observations requires physical measurements, these observations are usually less accurate than range measurements. In addition, they're a less sensitive measurement because we must scale them with range ($r\Delta\theta$). Hovanessian (1988, 9-11) shows how the signal wavelength can indicate the expected accuracy (to within about 30%). The accuracy of angular measurement is directly related to the **beamwidth**, θ_{BW} , or the signal angle,

$$\theta_{BW} = \frac{\lambda}{D}$$

where λ is the wavelength and D is the aperture diameter. A smaller beamwidth will produce more accurate results for a fixed-size aperture, D . Thus, we'd like shorter wavelengths and larger sensor diameters. The geometry of the pass is also important. Angular measurements are usually much less precise than range and range-rate information. Because the atmosphere causes the signals to "bend," determining the satellite's true direction demands more analysis.

We're all familiar with the Doppler shift of a sound wave from a train horn changing frequency as it passes. This shift alters the frequency of electromagnetic radiation as it travels to and from a satellite. It also affects all forms of data-gathering devices. We determine Doppler shift by comparing the frequency of the signal received either at the satellite or subsequently on the ground to the transmitted frequency, f_o . This lets us evaluate the range rate because the sensor receives the return signal at $f_o + f_d$. In general, the frequency shift, f_d , is

$$f_d = \frac{2\rho}{\lambda}$$

This relationship exists for a system where the frequency doesn't change due to retransmission at the satellite—such as simple reflection in radar and laser systems. For most transponder systems, the frequency is changed for retransmission, so this equation is no longer valid. Long et al. (1989, 7-15 to 7-18) show several ways to determine range rate.

6.2.1 Quantity of Data

An important influence on the accuracy of techniques for initial orbit determination is the quantity and proximity of observations to other data. We usually want as much data as possible for initial orbit determination because we have little or no knowledge of the orbit.

Most sensors can provide observations at far higher frequencies than needed, even considering the requirements of initial orbit determination. We'll see in Chap. 9 that differential-correction techniques must have *sufficiently dense* data to achieve a desired result. This often means hundreds of observations each time we observe a satellite.

Whatever the density, we must also be aware of the proximity of data to other information we process. In cases where the observations are limited to a small arc of the orbit, we use the term *short-arc* to denote the limited part of the orbit we observe. The inverse problem is that of observing a satellite for a *long-arc*, usually over multiple revolutions. There are advantages and disadvantages for each, but we often prefer long-arc cases because they provide more accurate determination of the satellite's orbit. Short-arc determinations result in an approximate estimate of the orbital characteristics and are often better indicators of near-term satellite motion. This chapter will cover techniques to handle both cases.

6.2.2 Types of Data

Seven combinations of basic data types are usually available. Depending on the type of data, we can select different solution methods for determining the initial orbits. Table 6-1 shows representative observation types and corresponding solution methods.

TABLE 6-1. Deciding on an Appropriate Technique for Initial Orbit Determination. This table pairs solution techniques with available observational data. Each data type must include station, satellite location ($\phi_{gd}, \lambda, h_{ellp}$) and time (UTC). Section 9.8 discusses techniques which can process ρ and $\dot{\rho}$ data.

Data Type	Restrictions	Sec./Chap.	Solution Method
$\dot{\rho}$	None	9	Estimation
β, el	3 sets minimum	6.4	<i>Laplace, Gauss, Double-r</i>
ρ, β, el	2 sets minimum	6.3, 6.6, 6.7	<i>SITE-TRACK</i> , then <i>Lambert</i> (2) or <i>GIBBS/HGIBBS</i> (3)
$\rho, \beta, el, \dot{\rho}$	2 or 3 sets minimum	6.3	<i>SITE-TRACK</i>
$\rho, \beta, el, \dot{\rho}, \dot{\beta}, \dot{el}$	None	6.3	<i>SITE-TRACK</i>
α_r, δ_t	3 sets minimum	6.4	<i>Laplace, Gauss, Double-r</i>
ρ^*	6 simultaneous, None	6.5, 9	Trilateration, Estimation

* $\dot{\rho}$ is often included with ρ values.

6.2.3 Example Applications

I'll introduce two satellite systems to illustrate the principles of tracking: TOPEX/Poseidon and TDRS.

TOPEX was launched on August 10, 1992 to take precise sea-level measurements. The radial component of the satellite's position must be known within centimeters. It's in a nearly circular orbit, about 1334 km in altitude, and has a groundtrack which repeats after completing 127 revolutions in 10 days. Cangahuala et al. (1995) give excellent background on the TOPEX satellite and are my main source for the following discussion.

TOPEX contains five systems to determine orbits: laser retro-reflectors, dual Doppler receivers for *Doppler Orbitography and Radio positioning Integrated by Satellite* (DORIS), *Global Positioning System Demonstration Receivers* (GPSDR), TDRS, and single and dual channel altimeters. Thus, TOPEX is very useful for exploring techniques of orbit determination. NASA operates the *Satellite Laser Ranging* (SLR) network to gather laser measurements of the satellite's position. The French Space Agency operates the DORIS system. It uses two signals to remove first-order effects of atmospheric refraction. GPSDR also uses two signals but incorporates the GPS's differential ranging to help process data.

All the data available for TOPEX results in high-precision ephemerides, or *precision orbit ephemerides* (POE). The required radial accuracy was originally 13 cm, but 2 cm is typically achieved. Remember, these are post-processed (see Sec. 9.8) solutions for a satellite ostensibly outside the drag regime.

NASA's TDRS uses many techniques similar to those of TOPEX but has some subtle variations. The system consists of six geostationary satellites which support satellite communications. TDRS's main tracking capability comes from the Bilateral Ranging Transponder System. Marshall et al. (1995) indicate that this system uses S-band telemetry to obtain range and range-rate information for the satellite. He also discusses how to increase the accuracy of TDRS's orbit ephemerides using one-way, two-way, and POE information from TOPEX. This process uses signals from TDRS to determine another satellite's position.

6.3 Observations of Range, Azimuth, and Elevation

Radar and modern laser systems have greatly enhanced our ability to determine initial orbits, especially by giving us range information. The classical discoveries highlighted in Sec. 6.1 resulted mainly from optical observations and included only angular data. Range information was rarely available, and *accurate* range data simply didn't exist. Although we still use techniques to determine angles (see Sec. 6.4), range information allows us to analyze data faster, more simply, and more accurately. In Sec. 9.5.4 we'll discuss the types of data commonly available, but for now, let's assume data coming from a sensor site includes range, azimuth, elevation, and the rates of each measurement. Angle rates are rarely obtained because mechanical drive assemblies usually make them noisy. Sometimes, the inaccuracy of rate information can overwhelm the true values, so we get little useful information for orbit determination. Although not all sensors provide all six pieces of information, some do, so I'll present this method as primary for determining initial orbits.

6.3.1 Application: SITE-TRACK

The process I've labeled **SITE-TRACK** is a common way of determining a satellite's position and velocity vector from sensor data. The data usually consists of range, azimuth, elevation, rates of each variable, times of each observation, and the site location. Sometimes, only subsets of this data may be available, and I'll present other techniques in this chapter to determine the orbit for those cases. For the **SITE-TRACK** problem here, we'll assume *all* the information is available. If the algorithm doesn't have rate information on range, azimuth, and elevation, it still works but determines just the position vector. In this case, skip each operation for the velocity vector.

Although ground-based sensors are addressed here, the process applies equally well for a space-based sensor; however, a few difficulties arise. First, if the "site" is a satellite, it is more difficult to determine its position accurately. Many operational routines also can't process information from space-based sensors because assumptions have been made (and often not documented), such as the inability to input negative elevations.

The general process is to find the sensor site's position and velocity vectors, find the satellite's vectors *relative to the site (SEZ)*, convert the vectors to a common coordinate frame (*ECEF*), and then transform the position and velocity vectors to the inertial frame (*IJK*). Figure 6-3 shows the geometry.

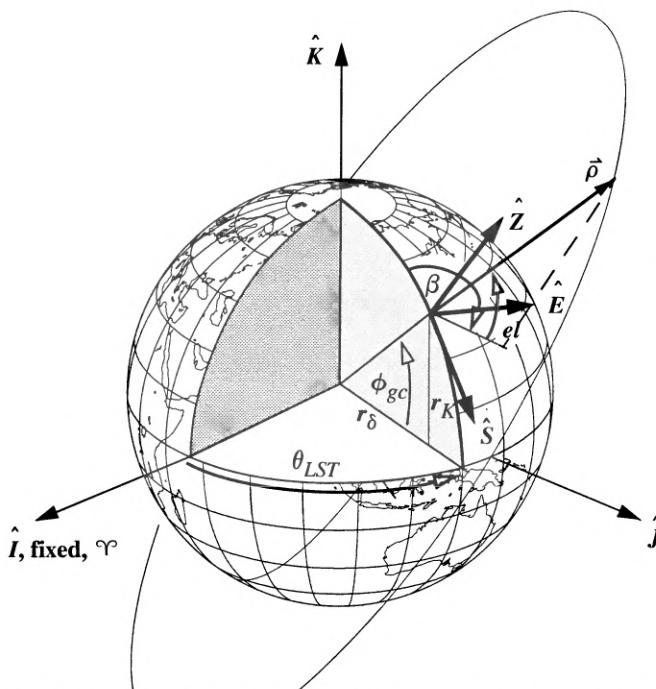


Figure 6-3. Geometry of Site Coordinates. The site's position depends on the r_δ and r_K components. Notice we must know LST and site latitude.

Distinguishing coordinate systems is important for highly accurate studies. The *ECEF* system is an important intermediate step. If we consider rotating coordinate systems (Sec. 1.5.2), the final conversion to *IJK* is actually a conversion to an inertial system that we must specify. The process usually involves accounting for sidereal time, nutation, and precession. Polar motion is often included in the station parameters and location information. I won't address these concepts in this section because *SITE-TRACK* is a general-purpose routine. If your application requires higher fidelity, be sure to include the additional conversions.

The first step is to find the site's position vector. Recalling the discussion in Sec. 1.3.2 about the ellipsoidal model for the Earth's shape, it's prudent to account for the precise shape of the Earth when finding site coordinates. The need for this adjustment grows as the desired accuracy increases. Typically, sensor measurements are accurate from about 10 m to 400 m. Not accounting for the Earth's equatorial bulge results in errors on the order of 15,000 m! We determine the site's position components from Eq. (1-16). Using Fig. 6-3, it's easy to construct the position vector directly in the inertial Geocentric Equatorial System (*IJK*).

Notice, we must know the Local Sidereal Time, *LST* or θ_{LST} . *LST* uniquely locates the site in inertial space at the time of each observation. In practice, we often skip the intermediate step of determining the *LST*, and keep the station parameters and time as the only parameters. Remember the time is usually given as *UTC*. To obtain *LST*, we must convert *UTC* to *UT1* and then determine *LST*. See Algorithm 3 for additional information. We'll use *LST* to simplify the notation and to maintain consistency with other derivations in the book. The site's position vector in the *IJK* frame is

$$\vec{r}_{siteIJK} = \begin{bmatrix} r_\delta \cos(\theta_{LST}) \\ r_\delta \sin(\theta_{LST}) \\ r_K \end{bmatrix} \quad (6-1)$$

If we need the site's velocity vector, we can use Eq. (5-13).

Once we know the site's location, we must find the satellite's position and velocity vectors in the topocentric coordinate system, *SEZ*. In Fig. 6-4, I've enlarged a section of Fig. 6-3 to show the geometry.

Defining *elevation*, *el*, and *azimuth*, *β*, is important. As discussed in Sec. 1.4, we measure elevation from the horizon of the site to the satellite and azimuth clockwise from north. This means azimuth varies from 0 to 360°, whereas elevation varies between -90° and 90°. This convention works for space-based sensors as well.

From Fig. 6-4, we can reconfirm Eq. (3-3) and Eq. (3-4). Notice we now have the site's position vector in the *IJK* coordinate system, whereas the satellite's vectors are in the topocentric *SEZ* system. Transforming the coordinates will complete the process. Because we want to determine the satellite's position and velocity vectors in the *IJK* coordinate system, we'll transform from *SEZ* to *IJK*, using one of the methods discussed in Sec. 1.5. Keeping in mind the rotation notation we've developed, recall Eq. (1-33):

$$\begin{aligned}\hat{\rho}_{IJK} &= [\text{ROT3}(-\theta_{LST})] [\text{ROT2}(-(90^\circ - \phi_{gd}))] \hat{\rho}_{SEZ} \\ \dot{\hat{\rho}}_{IJK} &= [\text{ROT3}(-\theta_{LST})] [\text{ROT2}(-(90^\circ - \phi_{gd}))] \dot{\hat{\rho}}_{SEZ}\end{aligned}$$

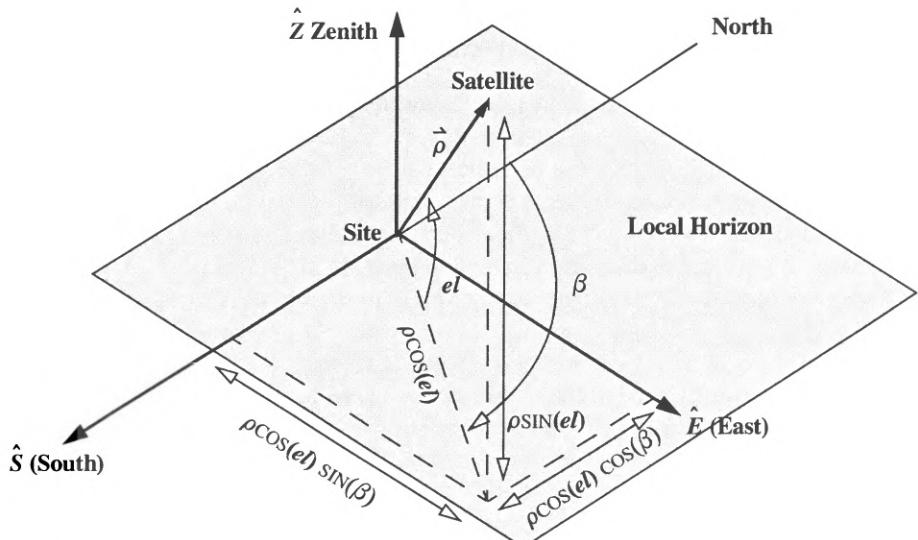


Figure 6-4. Geometry of the Sensor Site. This figure shows the site geometry from a different perspective. Elevation is measured positive up from the local horizon. Notice the positive definition because we measure azimuth from north.

We must reexamine the physical situation to complete the process. Figure 6-5 shows the geometry of the site and satellite. The geocentric position vector is

$$\hat{r}_{IJK} = \hat{\rho}_{IJK} + \hat{r}_{siteIJK} \quad (6-2)$$

We can't simply “add” the velocity vectors. The satellite's velocity, as measured from the site, depends on the site's motion. Therefore, to find the inertial velocity, our procedure must account for the site's relative motion, as well as the measured velocity terms from the radar site. It relies on the general principle of determining a relative velocity with respect to a fixed frame, as described in Sec. 1.5.2. To find the velocity of the object in the inertial frame, we add the velocity of the satellite relative to the moving frame and the velocity of the moving frame. For this particular transformation, the *SEZ* system is the moving frame, and the *IJK* is the fixed or pseudo-inertial frame. If the satellite is at the inertial position \hat{r}_{IJK} , its velocity due to the motion of the reference frame is the cross product $(\hat{\omega}_\oplus \times \hat{r}_{IJK})$. Now we add this correction to the velocity vector determined from

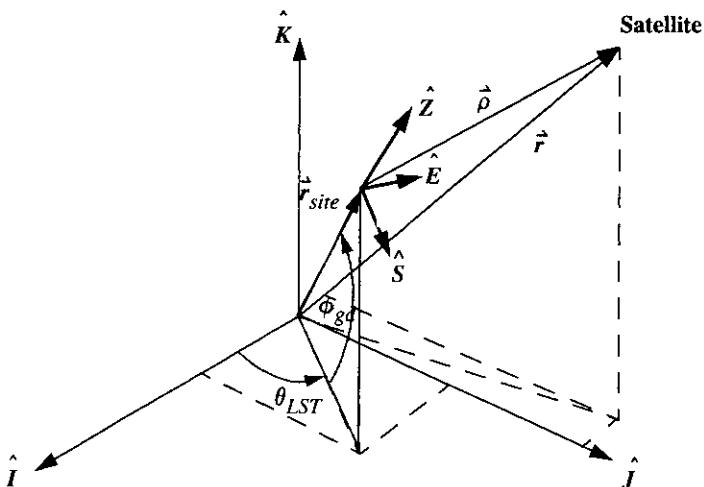


Figure 6-5. Site-to-Satellite Geometry. To complete **SITE-TRACK**, we must use the slant-range vector and the site vector and rotate through geodetic latitude, ϕ_{gd} , and local sidereal time, θ_{LST} .

observations at the site. An alternate expression can use the slant-range vector expressed in the geocentric coordinate system:

$$\vec{v}_{IJK} = \vec{\rho}_{IJK} + \vec{\omega}_{\oplus} \times \vec{r}_{IJK} = \vec{\rho}_{IJK} + \vec{\omega}_{\oplus} \times \vec{\rho}_{IJK} + \vec{v}_{siteIJK} \quad (6-3)$$

Implementing the SITE-TRACK Algorithm

Summarizing, we observe a subtlety in the formulation—we need geodetic latitude. Remember the discussion on geodesy from Chap. 1 (Sec. 1.3.3). When we use latitude and longitude coordinates for a site, they are always referenced to a specific datum. Unfortunately, all geodetic datums aren't consistent—they can differ by translation, rotation, and stretching. These effects are beyond our scope, and because geodetic latitude in any datum is always perpendicular to the ellipsoid, the use of this algorithm remains the same.

To speed calculations, we can save several quantities as temporary variables and use them in multiple locations. For example, saving the Earth's radius divided by the square root in the site coordinates eliminates many transcendental operations. The velocity equation in the *SEZ* frame may also use the components of the position vector (in *SEZ*) to eliminate trigonometric operations. We should also store the trigonometric values in the transformation matrix to *IJK* and use the complete matrix rather than rotation procedures. Finally, using the sine and cosine of the azimuth and elevation angles eliminates more trigonometric variables. Including all these recommendations will yield a robust and fast algorithm.

ALGORITHM 45: SITE-TRACK

$$(\phi_{gd}, \theta_{LST}, h_{ellp}, \rho, \beta, el, \dot{\rho}, \dot{\beta}, \dot{el} \Rightarrow \vec{r}_{IJK}, \vec{v}_{IJK}, \vec{r}_{siteIJK})$$

$$C_{\oplus} = \frac{R_{\oplus}}{\sqrt{1 - e_{\oplus}^2 \sin^2(\phi_{gd})}} \quad S_{\oplus} = C_{\oplus}(1 - e_{\oplus}^2)$$

$$r_{\delta} = (C_{\oplus} + h_{ellp}) \cos(\phi_{gd})$$

$$r_K = (S_{\oplus} + h_{ellp}) \sin(\phi_{gd})$$

$$\vec{r}_{siteIJK} = \begin{bmatrix} r_{\delta} \cos(\theta_{LST}) \\ r_{\delta} \sin(\theta_{LST}) \\ r_K \end{bmatrix}$$

$$\vec{\rho}_{SEZ} = \begin{bmatrix} -\rho \cos(el) \cos(\beta) \\ \rho \cos(el) \sin(\beta) \\ \rho \sin(el) \end{bmatrix}$$

$$\dot{\vec{\rho}}_{SEZ} = \begin{bmatrix} -\dot{\rho} \cos(el) \cos(\beta) + \rho \sin(el) \cos(\beta) \dot{el} + \rho \cos(el) \sin(\beta) \dot{\beta} \\ \dot{\rho} \cos(el) \sin(\beta) - \rho \sin(el) \sin(\beta) \dot{el} + \rho \cos(el) \cos(\beta) \dot{\beta} \\ \dot{\rho} \sin(el) + \rho \cos(el) \dot{el} \end{bmatrix}$$

$$\vec{r}_{IJK} = [\text{ROT3}(-\theta_{LST})] [\text{ROT2}(-(90^\circ - \phi_{gd}))] \vec{\rho}_{SEZ} = [\frac{IJK}{SEZ}] \vec{\rho}_{SEZ}$$

$$\dot{\vec{r}}_{IJK} = [\text{ROT3}(-\theta_{LST})] [\text{ROT2}(-(90^\circ - \phi_{gd}))] \dot{\vec{\rho}}_{SEZ} = [\frac{IJK}{SEZ}] \dot{\vec{\rho}}_{SEZ}$$

$$[\frac{IJK}{SEZ}] = \begin{bmatrix} \sin(\phi_{gd}) \cos(\theta_{LST}) & -\sin(\theta_{LST}) & \cos(\phi_{gd}) \cos(\theta_{LST}) \\ \sin(\phi_{gd}) \sin(\theta_{LST}) & \cos(\theta_{LST}) & \cos(\phi_{gd}) \sin(\theta_{LST}) \\ -\cos(\phi_{gd}) & 0 & \sin(\phi_{gd}) \end{bmatrix}$$

$$\vec{r}_{IJK} = \vec{\rho}_{IJK} + \vec{r}_{siteIJK}$$

$$\vec{v}_{IJK} = \dot{\vec{\rho}}_{IJK} + \vec{\omega}_{\oplus} \times \vec{r}_{IJK}$$

▼ Example 6-1: Using Algorithm 45 to Determine the SITE-TRACK.

GIVEN: $\phi_{gd} = 39.007^\circ$, $\lambda = 104.883^\circ\text{W}$, alt = 2187 m

with an observation of $\rho = 604.68$ km, $\beta = 205.6^\circ$, $el = 30.7^\circ$, $\dot{\rho} = 2.08$ km/s, $\dot{\beta} = 0.15^\circ/\text{s}$, $\dot{el} = 0.17^\circ/\text{s}$ on May 20, 1995 at 03:17:02.00 UT

FIND: \hat{r}_{site} , \hat{r} , \hat{v}

The first task is to determine the site's position vector. The auxiliary terms are

$$C_\oplus = \frac{R_\oplus}{\sqrt{1 - e_\oplus^2 \sin^2(\phi_{gd})}} = 1.001\,328\,7 \text{ ER}$$

$$S_\oplus = C_\oplus \left(1 - e_\oplus^2\right) = 0.994\,625 \text{ ER}$$

The vertical and horizontal components are

$$r_\delta = (C_\oplus + h_{ellp}) \cos(\phi_{gd}) = 0.778\,368 \text{ ER}$$

$$r_K = (S_\oplus + h_{ellp}) \sin(\phi_{gd}) = 0.626\,248 \text{ ER}$$

Next, let's determine the LST. From Algorithm 2 and Algorithm 1, we know

May 20, 1995, 3 hr, 17 min, 2 s $\Rightarrow JD = 2,449,857.636\,829$

$$\theta_{GST} = 286.573\,89^\circ$$

$$\theta_{LST} = 286.573\,89^\circ + (-104.883) = 181.690\,89^\circ$$

The site vector is

$$\hat{r}_{site} = \begin{bmatrix} r_\delta \cos(\theta_{LST}) \\ r_\delta \sin(\theta_{LST}) \\ r_K \end{bmatrix} = \begin{bmatrix} -0.778\,029 \\ -0.022\,967 \\ 0.626\,248 \end{bmatrix} \text{ ER} = \begin{bmatrix} -4962.375\,55 \\ -146.486\,67 \\ 3994.295\,54 \end{bmatrix} \text{ km}$$

The slant-range position and velocity vectors in the SEZ coordinate system are

$$\hat{r}_{SEZ} = \begin{bmatrix} -\rho \cos(el) \cos(\beta) \\ \rho \cos(el) \sin(\beta) \\ \rho \sin(el) \end{bmatrix} = \begin{bmatrix} 0.073\,515 \\ -0.035\,223 \\ 0.048\,402 \end{bmatrix} \text{ ER}$$

$$\overset{\wedge}{\rho}_{SEZ} = \begin{bmatrix} -\dot{\rho} \cos(el) \cos(\beta) + \rho \sin(el) \cos(\beta) \dot{el} + \rho \cos(el) \sin(\beta) \dot{\beta} \\ \dot{\rho} \cos(el) \sin(\beta) - \rho \sin(el) \sin(\beta) \dot{el} + \rho \cos(el) \cos(\beta) \dot{\beta} \\ \dot{\rho} \sin(el) + \rho \cos(el) \dot{el} \end{bmatrix} = \begin{bmatrix} 0.025\,136 \\ -0.202\,971 \\ 0.329\,473 \end{bmatrix} \text{ ER/TU}$$

The transformation matrix from SEZ to IJK is best determined as a complete matrix:

$$\begin{aligned} \begin{bmatrix} IJK \\ SEZ \end{bmatrix} &= \begin{bmatrix} \sin(\phi_{gd}) \cos(\theta_{LST}) & -\sin(\theta_{LST}) & \cos(\phi_{gd}) \cos(\theta_{LST}) \\ \sin(\phi_{gd}) \sin(\theta_{LST}) & \cos(\theta_{LST}) & \cos(\phi_{gd}) \sin(\theta_{LST}) \\ -\cos(\phi_{gd}) & 0 & \sin(\phi_{gd}) \end{bmatrix} \\ &= \begin{bmatrix} -0.629\,141 & 0.029\,507 & -0.776\,731 \\ -0.018\,572 & -0.999\,564 & -0.022\,929 \\ -0.777\,069 & 0 & 0.629\,415 \end{bmatrix} \end{aligned}$$

Doing the matrix operations gives us the position and velocity vectors in the *IJK* frame.

$$\dot{\vec{r}}_{IJK} = \begin{bmatrix} IJK \\ SEZ \end{bmatrix} \dot{\vec{r}}_{SEZ} = \begin{bmatrix} -0.084\ 886 \\ 0.032\ 732 \\ -0.026\ 662 \end{bmatrix} \text{ER}$$

$$\dot{\vec{v}}_{IJK} = \begin{bmatrix} IJK \\ SEZ \end{bmatrix} \dot{\vec{v}}_{SEZ} = \begin{bmatrix} -0.277\ 716 \\ 0.194\ 862 \\ 0.187\ 843 \end{bmatrix} \text{ER/TU}$$

Finally, we must determine the position and velocity vectors in the *IJK* frame:

$$\dot{\vec{r}} = \dot{\vec{r}}_{IJK} + \dot{\vec{r}}_{siteIJK} = \begin{bmatrix} -0.862\ 915 \\ 0.009\ 764 \\ 0.599\ 586 \end{bmatrix} \text{ER} = \begin{bmatrix} -5503.795\ 62 \\ 62.281\ 91 \\ 3824.244\ 80 \end{bmatrix} \text{km}$$

$$\dot{\vec{v}} = \dot{\vec{v}}_{IJK} + \vec{\omega}_{\oplus} \times \dot{\vec{r}}_{IJK} = \begin{bmatrix} -0.278\ 290 \\ 0.144\ 093 \\ 0.187\ 843 \end{bmatrix} \text{ER/TU} = \begin{bmatrix} -2.199\ 987 \\ 1.139\ 111 \\ 1.484\ 966 \end{bmatrix} \text{km/s}$$



6.4 Angles-only Observations

Modifications of the *SITE-TRACK* procedure we've discussed do most of the processing for today's systems, but some measurements need different techniques. Angles-only methods can span this gap in some cases; yet, they're fundamentally deficient because they lack range information. Iteration or more processing overcomes this problem, but not without additional complexity and computation. Thus, we want to use good range measurements whenever possible but must use angles-only for optical sensors, such as the early Baker-Nunn camera or *Ground-Based Electro-Optical Deep Space Sensors* (GEODSS) sites. *Optical sensors* essentially take pictures of regions of the sky. When a satellite "streaks" across the picture (in the Baker-Nunn cameras), its location is compared to known stars in the picture and the angular directions are determined, but without range information. The GEODSS sites operate this way using charge-coupled-device cameras instead of photographs.

The problem still consists of determining the satellite's state vector, which requires six independent quantities. As we've seen so far, these may be the classical orbital elements and their variations—the range, azimuth, elevation, and rate information of the *SITE-TRACK* problem; or the position vectors—in this case, geocentric or topocentric right ascension and declination. Figure 6-6 shows such an example case for which an optical site provides angular data for at least three line-of-sight vectors. We may tend to view the unit vectors as position vectors. They're not. Indeed, our basic difficulty with the angles-only technique is the lack of range information.

We'll explore three techniques in this section: Laplacian, Gaussian, and double *r*-iteration. It's important to know all three methods because each solves one aspect of the problem. The Laplacian method fits only the middle point but is usually close to the other data points. I'm including it only for its historical significance; it works poorly for near-Earth

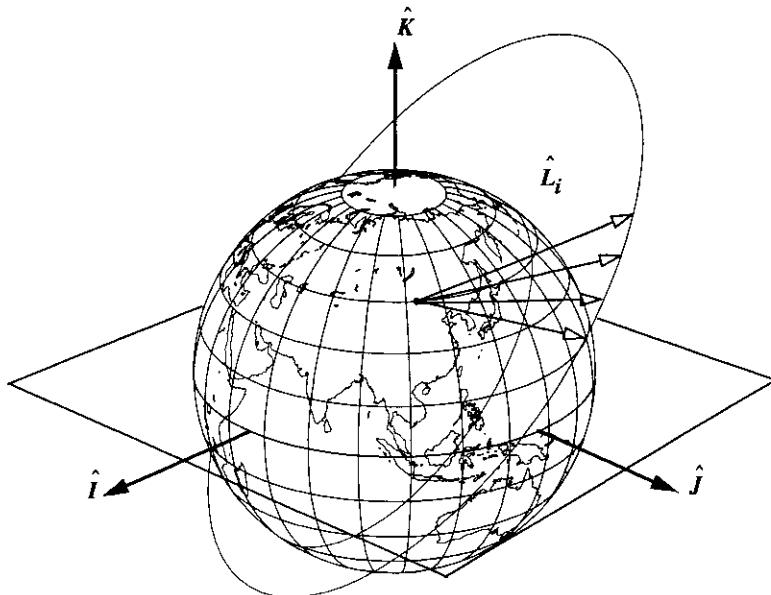


Figure 6-6. Geometry of Angles-only Observations. The underlying principle for the angles-only technique is the use of topocentric angles to form line-of-sight vectors. For objects that are very distant, the distinction for topocentric values diminishes.

satellites. The Gaussian method fits data to *all three* points and is valid for all the data. This approach is better suited for near-Earth satellites, but the formulation limits the spread of data (usually less than 60° apart). Finally, the double r -iteration is effective for large spreads in the data, as in the case of multiple observing locations. This method from Escobal ([1965] 1985, 272–288) is a good approach for general applications. Lane (1991) presents a numerical approach that is valid for geosynchronous and Molniya orbits. It may also apply to general purposes.

Several preliminary remarks are necessary for each process. Each method begins by assuming we've taken the *topocentric* measurements and considered them as *ordered pairs* of right ascension and declination $[(\alpha_1, \delta_1), (\alpha_2, \delta_2), (\alpha_3, \delta_3)]$ and times, t_i . Remember, there's no practical way to measure geocentric values. The most common way to find these angles is to photograph the satellite and determine them from the star background. This technique assumes the star positions are accurately known. Be aware, the stellar right ascensions and declinations as cataloged are *geocentric*, whereas the optical observations for the stars and satellites yield *topocentric* right ascensions and declinations. However, stars are so distant that the difference in the two quantities is negligible. In contrast, topocentric values for satellites in low-Earth orbits can differ greatly from geocentric α and δ because they're nearby (recall Example 3-7.) As an aside, we can also use β and

et, but we must first convert them using Algorithm 17. Finally, it's crucial to use a common reference frame for all processing, which requires us to do precession and nutation calculations for mean and true equinoxes and equators. Typically, the observations are obtained in a true-equator true-equinox frame. The calculations should proceed in a common inertial frame—usually the J2000, but may take place in older frames like B1950. The first step is to form line-of-sight unit vectors, \hat{L} , to the satellite at each observation time. Use Eq. (3-11):

$$\hat{L}_i = \begin{bmatrix} \cos(\delta_{ti}) \cos(\alpha_{ti}) \\ \cos(\delta_{ti}) \sin(\alpha_{ti}) \\ \sin(\delta_{ti}) \end{bmatrix} \quad i = 1 \dots 3 \quad (6-4)$$

If we knew the slant range, ρ , to each satellite and the observation site's location, we could easily write the position vectors. In general, each of these vectors would be

$$\vec{r} = \rho \hat{L} + \vec{r}_{site} \quad (6-5)$$

This relation formalizes our use of the *topocentric* values for right-ascension and declination.

We use Algorithm 45 to obtain the site vectors. As with the observations, it's very important to form these vectors in the proper reference frame. If we dot Eq. (6-5) with itself, we get an expression for the satellite's distance:

$$r = \sqrt{\rho^2 + 2\rho \hat{L} \cdot \vec{r}_{site} + r_{site}^2} \quad (6-6)$$

6.4.1 Laplace's Method

This procedure was actually one of the first methods for determining orbits because it used only optical sightings when accurate range measurements were unavailable. Laplace first proposed this method in the *Mémoires de l'Academie Royale des Sciences de Paris* in 1780 (Escobal, [1965] 1985, 261). He used it to estimate the middle position and velocity vectors in a span of observational data for the orbits of comets and minor planets. The method requires three sets of angular measurements to provide the six independent quantities required for the solution. In the following derivation, notice this method can process data from different sites at different times. Although this may at first seem unlikely because times from different sites are typically hours apart, the solution does converge.

Let's do the preliminary calculations of Eq. (6-4) and Eq. (6-5) first and then use any reduction calculations with the resulting unit and site vectors. Our inability to find the range will complicate the procedure later; however, for now, assume we have at least an estimate of this quantity. Differentiating Eq. (6-5) twice produces the velocity and acceleration for each observation:

$$\begin{aligned}\dot{\hat{r}} &= \dot{\rho}\hat{L} + \rho\dot{\hat{L}} + \dot{\hat{r}}_{site} \\ \ddot{\hat{r}} &= \ddot{\rho}\hat{L} + 2\dot{\rho}\dot{\hat{L}} + \ddot{\rho}\hat{L} + \ddot{\hat{r}}_{site}\end{aligned}\quad (6-7)$$

The process assumes two-body motion [Eq. (2-4)], so we may substitute this relation for the satellite's acceleration above and combine with Eq. (6-5) to get

$$-\frac{\mu}{r^3}(\rho\hat{L} + \dot{\hat{r}}_{site}) = \ddot{\rho}\hat{L} + 2\dot{\rho}\dot{\hat{L}} + \ddot{\rho}\hat{L} + \ddot{\hat{r}}_{site}$$

Rearranging gives us

$$\ddot{\rho}\hat{L} + 2\dot{\rho}\dot{\hat{L}} + \rho\left(\ddot{\hat{L}} + \frac{\mu}{r^3}\hat{L}\right) = -\ddot{\hat{r}}_{site} - \frac{\mu}{r^3}\dot{\hat{r}}_{site} \quad (6-8)$$

All the derivatives in the above equation are unknown; thus, we must continue the analysis. Because each line-of-sight vector is at a particular time, we can use the Lagrange interpolation formula (Escobal [1965] 1985, 252),

$$\hat{r}(t) = \sum_{i=1}^n \hat{r}_i \prod_{k \neq i} \frac{t - t_k}{t_i - t_k} \quad (6-9)$$

to derive an approximate expression for $\hat{L}(t)$ at any time ($n = 3$ for three observations):

$$\hat{L}(t) = \frac{(t - t_2)(t - t_3)}{(t_1 - t_2)(t_1 - t_3)}\hat{L}_1 + \frac{(t - t_1)(t - t_3)}{(t_2 - t_1)(t_2 - t_3)}\hat{L}_2 + \frac{(t - t_1)(t - t_2)}{(t_3 - t_1)(t_3 - t_2)}\hat{L}_3$$

Differentiating results in

$$\begin{aligned}\dot{\hat{L}}(t) &= \frac{2t - t_2 - t_3}{(t_1 - t_2)(t_1 - t_3)}\hat{L}_1 + \frac{2t - t_1 - t_3}{(t_2 - t_1)(t_2 - t_3)}\hat{L}_2 + \frac{2t - t_1 - t_2}{(t_3 - t_1)(t_3 - t_2)}\hat{L}_3 \\ \ddot{\hat{L}}(t) &= \frac{2}{(t_1 - t_2)(t_1 - t_3)}\hat{L}_1 + \frac{2}{(t_2 - t_1)(t_2 - t_3)}\hat{L}_2 + \frac{2}{(t_3 - t_1)(t_3 - t_2)}\hat{L}_3\end{aligned}$$

We can simplify these relations if we assume the middle time to be zero ($t = t_2 = 0$). This assumption is common because it allows us to use the differences in times. It's also important to recognize that, if more observations are available, we can estimate the first and second derivatives much more accurately by using Lagrange's interpolation formula with these additional observations. We can also use least-squares techniques to provide even better estimates. Either of these approaches is also necessary for improved accuracy if the higher derivatives are significant.

Next we need the derivatives of the site vector. To make the procedure completely general, you'll find it convenient to be able to process information from different sites or from a single site. If all the observations occur from one site, form

$$\dot{\vec{r}}_{site_2} = \vec{\omega}_{\oplus} \times \vec{r}_{site_2} \quad \ddot{\vec{r}}_{site_2} = \vec{\omega}_{\oplus} \times \dot{\vec{r}}_{site_2}$$

If the observations are from different sites, use a **Herrick-Gibbs** (Sec. 6.6.2) approach to determine the average site velocity and acceleration vectors. Again, use Lagrange's interpolation formula [Eq. (6-9)], with a shorthand notation for the change in the times (Escobal [1965] 1985, 268).

$$\tau_i = t_i - t_2$$

$$\dot{\vec{r}}_{site_2} = -\frac{\tau_3}{\tau_1(\tau_1 - \tau_3)} \vec{r}_{site_1} - \frac{\tau_3 + \tau_1}{\tau_1 \tau_3} \vec{r}_{site_2} - \frac{\tau_1}{\tau_3(\tau_3 - \tau_1)} \vec{r}_{site_3}$$

Find the acceleration equation by the interpolation formula:

$$\ddot{\vec{r}}_{site_2} = \frac{2}{\tau_1(\tau_1 - \tau_3)} \vec{r}_{site_1} + \frac{2}{\tau_1 \tau_3} \vec{r}_{site_2} + \frac{2}{\tau_3(\tau_3 - \tau_1)} \vec{r}_{site_3}$$

Referring to the acceleration [Eq. (6-8)], you now know all the variables, except the range, ρ , and its derivatives, and the position magnitude, r . As pointed out earlier, this becomes the difficult part of the routine. Begin by assuming you have a value for the position magnitude. The three equations in Eq. (6-8) become

$$\left[\hat{L} + 2\dot{\hat{L}} + \ddot{\hat{L}} + \frac{\mu}{r^3} \hat{L} \right] \begin{bmatrix} \ddot{\rho} \\ \dot{\rho} \\ \rho \end{bmatrix} = - \left[\ddot{\vec{r}}_{site} + \frac{\mu}{r^3} \vec{r}_{site} \right]$$

They can be solved using determinants and Cramer's rule (see Sec. C.2). Find the *determinant*, D , for the system of equations in slant range:

$$D = \left| \hat{L} + 2\dot{\hat{L}} + \ddot{\hat{L}} + \frac{\mu}{r^3} \hat{L} \right| \quad (6-10)$$

Do some column reduction by subtracting μ/r^3 times the first column from the third. This will reduce the complexity slightly. Factoring out the 2 from the second column (given the matrix rules from Sec. C.2) yields

$$D = 2 \left| \hat{L} + \dot{\hat{L}} + \ddot{\hat{L}} \right|$$

Now apply Cramer's rule using Eq. (6-8). Because the third column represents the slant-range component, replace that column with the right-hand side of Eq. (6-8) (factor out the -1 from that column).

$$D\rho = -2 \begin{vmatrix} \hat{L}_1 & \hat{L}_2 & \ddot{\vec{r}}_{site} + \frac{\mu}{r^3} \vec{r}_{site} \end{vmatrix}$$

Then, split this determinant as follows:

$$D\rho = -2 \begin{vmatrix} \hat{L}_1 & \hat{L}_2 & \ddot{\vec{r}}_{site} \end{vmatrix} - 2 \frac{\mu}{r^3} \begin{vmatrix} \hat{L}_1 & \hat{L}_2 & \vec{r}_{site} \end{vmatrix}$$

Define the first determinant as D_1 and the second determinant as D_2 . Solve for the range, ρ , to get

$$\rho = \frac{-2D_1}{D} - \frac{2\mu D_2}{r^3 D} \quad (6-11)$$

Notice the determinant could be zero, which causes numerical difficulties. Escobal ([1965] 1985, 265–266) points out that this will occur only when the site lies on the great circle that encompasses the satellite's orbit. Although this possibility is remote, we still need to consider it when developing computer routines.

Because we have to guess at the slant range, we must iterate to determine the final value for ρ . Escobal ([1965] 1985, 264–265) presents an approach, as do Baker and Makemson (1967, 134). Substituting Eq. (6-11) into Eq. (6-6) results in

$$r_2^2 = \left(\frac{-2D_1}{D} - \frac{2\mu D_2}{r_2^3 D} \right)^2 + 2 \left(\frac{-2D_1}{D} - \frac{2\mu D_2}{r_2^3 D} \right) \hat{L}_2 \cdot \vec{r}_{site_2} + r_{site_2}^2$$

Expanding this equation and letting C equal the dot product will produce an eighth-order polynomial:

$$C = \hat{L}_2 \cdot \vec{r}_{site_2}$$

$$r_2^8 + \left(\frac{4CD_1}{D} - \frac{4D_1^2}{D^2} - r_{site_2}^2 \right) r_2^6 + \mu \left(\frac{4CD_2}{D} - \frac{8D_1D_2}{D^2} \right) r_2^3 - \frac{4\mu^2 D_2}{D^2} = 0 \quad (6-12)$$

Prussing and Conway (1992, 174) state that the correct root(s) is the real root from this equation. However, multiple roots may exist, in which case we must either try each root and compare to a priori data or process more observations to isolate the correct root. Selecting the correct root can be very difficult when designing multi-purpose routines. The final step is to solve Eq. (6-11) for the slant range at the middle time.

Now, repeat this process for the velocity. Use Cramer's rule again starting from Eq. (6-8) and remember that the velocity component was the middle term in the determinant expression, so

$$D\dot{\rho} = - \left| \begin{array}{c} \hat{L} + \ddot{\vec{r}}_{site} + \frac{\mu}{r^3} \vec{r}_{site} \\ \hat{L} \end{array} \right|$$

Split this determinant:

$$D\dot{\rho} = - \left| \begin{array}{c} \hat{L} + \ddot{\vec{r}}_{site} \\ \hat{L} \end{array} \right| - \frac{\mu}{r^3} \left| \begin{array}{c} \hat{L} + \ddot{\vec{r}}_{site} \\ \hat{L} \end{array} \right|$$

Define the first velocity determinant as D_3 and the second determinant as D_4 . Solving for the range rate, $\dot{\rho}$, results in

$$\dot{\rho} = - \frac{D_3}{D} - \frac{\mu D_4}{r^3 D} \quad (6-13)$$

So you can use the velocity determinants and solve for the middle velocity vector in Eq. (6-7):

$$\dot{\vec{v}}_2 = \dot{\rho} \hat{L}_2 + \rho \dot{\hat{L}}_2 + \dot{\vec{r}}_{site_2}$$

Notice this method doesn't require iteration if we find the correct root the first time in Eq. (6-12). The relation doesn't depend on range. Escobal ([1965] 1985, 265) remarks that, for satellites near Earth, we must greatly modify the line-of-sight unit vectors for solution. The resulting process still isn't robust enough to be a reliable approach for routine satellite observations using only sparse data (3–5 points). The method becomes more efficient when combined with statistical techniques discussed in Chap. 9. But it's historically important, and it remains valid for interplanetary observations.

6.4.2 Gaussian Determination

Gauss's method of orbit determination from angles-only data receives mixed reviews from the astrodynamical community. The opinions range from little concern because the method works best for interplanetary studies, to feeling that it's not very accurate for satellite-orbit determination, to reverence for the achievement realized at a time when data was limited. Long et al. (1989, 9-9) suggest it works best when the angular separation between observations is less than about 60° . The method performs remarkably well when the data is separated by 10° or less. This separation translates to observations that are (at most) about five to ten minutes apart for low-Earth satellites. Of course, you should always analyze the orbits you intend to process before establishing how often to collect data. The success of Gauss's method also depends on the method used to determine the f and g series, which we'll discuss below. Whatever the case, when properly formulated, the Gaussian routine is a modestly robust way to determine a satellite's position with angles-only data.

First, let's calculate the unit vectors and site vectors and transform them to a proper coordinate frame. The true starting point of the Gaussian technique relies on assuming the three position vectors lie in a single plane. Of course, virtually *all* real-world data can include points that are out of plane. Because the three vectors lie in the same plane,

$$\vec{c}_1 \vec{r}_1 + \vec{c}_2 \vec{r}_2 + \vec{c}_3 \vec{r}_3 = \vec{0} \quad (6-14)$$

We can get expressions for the coefficients by taking the cross product of the first and third position vectors with Eq. (6-14). Assuming c_2 is not zero, we have

$$\vec{r}_1 \times \vec{r}_3 (c_1) = \vec{r}_2 \times \vec{r}_3 (-c_2) \quad \vec{r}_1 \times \vec{r}_3 (c_3) = \vec{r}_1 \times \vec{r}_2 (-c_2)$$

But from Chap. 4's discussion of the "Classical Formula" on page 249, we know we can apply the f and g functions to form the position vectors, using the middle position and velocity vectors. Thus,

$$\vec{r}_i = f_i \vec{r}_2 + g_i \vec{v}_2, i = 1, 3$$

Eq. (6-14) permits us to rescale the coefficients for convenience. Therefore, by letting c_2 equal -1 , we can solve the vector relations, substitute the f and g values, and find alternate expressions for the coefficients:

$$c_1 = \frac{\vec{r}_2 \times (f_3 \vec{r}_2 + g_3 \vec{v}_2)}{(f_1 \vec{r}_2 + g_1 \vec{v}_2) \times (f_3 \vec{r}_2 + g_3 \vec{v}_2)} = \frac{g_3}{f_1 g_3 - f_3 g_1}$$

$$c_3 = \frac{(f_1 \vec{r}_2 + g_1 \vec{v}_2) \times \vec{r}_2}{(f_1 \vec{r}_2 + g_1 \vec{v}_2) \times (f_3 \vec{r}_2 + g_3 \vec{v}_2)} = \frac{-g_1}{f_1 g_3 - f_3 g_1}$$

If we knew the position and velocity vectors, the f and g functions could be solved to get a final solution [Eq. (4-63)]. Because the velocity is still unknown, we can use the *series* form of the f and g functions (see Chap. 4, "Series Forms," page 254). Remember, we're defining the data times as $\tau_i = t_i - t_2$ and the series coefficient, u , as $u = \mu/r_2^3$.

Next, let's substitute these coefficients into the c coefficient functions to get a modified form of the coefficients:

$$c_1 = \frac{g_3}{f_1 g_3 - f_3 g_1} = \frac{\tau_3}{\tau_3 - \tau_1} + \frac{u \tau_3 ((\tau_3 - \tau_1)^2 - \tau_1^2)}{6(\tau_3 - \tau_1)} = a_1 + a_{1u} u$$

$$c_3 = \frac{-g_1}{f_1 g_3 - f_3 g_1} \approx -\frac{\tau_1}{\tau_3 - \tau_1} - \frac{u \tau_1 ((\tau_3 - \tau_1)^2 - \tau_1^2)}{6(\tau_3 - \tau_1)} = a_3 + a_{3u} u$$

Although we do have expressions to approximate the coefficients, c_i , we haven't discovered expressions that allow us to find the slant-range values. From Eq. (6-5) we know the relation for each observation, including the site vector. Substituting into Eq. (6-14) gives us

$$c_1(\hat{\rho}_1 + \hat{r}_{site_1}) + c_2(\hat{\rho}_2 + \hat{r}_{site_2}) + c_3(\hat{\rho}_3 + \hat{r}_{site_3}) = \hat{0}$$

Separating terms results in

$$c_1\hat{\rho}_1 + c_2\hat{\rho}_2 + c_3\hat{\rho}_3 = -c_1\hat{r}_{site_1} - c_2\hat{r}_{site_2} - c_3\hat{r}_{site_3}$$

Now, recognize that we can introduce the unit vectors with the slant-range vectors [by Eq. (6-5)] and write the next relation in matrix form. This permits matrix solution of simultaneous equations and shows that we *must* solve them simultaneously.

$$\underbrace{[\hat{L}_1 : \hat{L}_2 : \hat{L}_3]}_{\mathbf{L}} \begin{bmatrix} c_1\rho_1 \\ c_2\rho_2 \\ c_3\rho_3 \end{bmatrix} = \begin{bmatrix} \hat{r}_{site_1} & \hat{r}_{site_2} & \hat{r}_{site_3} \end{bmatrix} \begin{bmatrix} -c_1 \\ -c_2 \\ -c_3 \end{bmatrix}$$

A basic objective of this problem is to determine the slant-range values. We can invert the unit-vector matrix, \mathbf{L} , as long as it's not singular, by applying Cramer's rule. Refer to Sec. C.2 for additional information. With unit vectors written as $\hat{L}_i = L_{xi}\hat{I} + L_{yi}\hat{J} + L_{zi}\hat{K}$, the solution is

$$\mathbf{L}^{-1} = \frac{\begin{bmatrix} L_{y2}L_{z3} - L_{y3}L_{z2} & L_{y1}L_{z3} - L_{y3}L_{z1} & L_{y1}L_{z2} - L_{y2}L_{z1} \\ L_{x2}L_{z3} - L_{x3}L_{z2} & L_{x1}L_{z3} - L_{x3}L_{z1} & L_{x1}L_{z2} - L_{x2}L_{z1} \\ L_{x2}L_{y3} - L_{x3}L_{y2} & L_{x1}L_{y3} - L_{x3}L_{y1} & L_{x1}L_{y2} - L_{x2}L_{y1} \end{bmatrix}}{|L|}$$

$$\begin{bmatrix} c_1\rho_1 \\ c_2\rho_2 \\ c_3\rho_3 \end{bmatrix} = \mathbf{L}^{-1} \begin{bmatrix} \hat{r}_{site_1} & \hat{r}_{site_2} & \hat{r}_{site_3} \end{bmatrix} \begin{bmatrix} -c_1 \\ -c_2 \\ -c_3 \end{bmatrix} \quad (6-15)$$

At this point, we can examine the slant-range equation from Eq. (6-15) for the middle time. Although not evident from a programming standpoint, it's useful to keep the c coefficients separated in a form $(a_i + a_{iu}u)$. So let's find the middle slant range (designate terms of the $\mathbf{L}^{-1}\hat{r}_{site}$ matrix as M 's):

$$\rho_2 = M_{21}a_1 - M_{22} + M_{23}a_3 + (M_{21}a_{1u} + M_{23}a_{3u})u = d_1 + d_2u$$

Label the slant-range coefficients (of u) to be d_1 and d_2 , respectively, and don't confuse the gravitational parameter, μ , with the coefficient u from the f and g series. Substituting this expression into Eq. (6-6), gives you an eighth-order polynomial in r_2 . Although the form is similar to Eq. (6-12) in the Laplace solution, it's *not* the same because we've defined the coefficients differently. We still face the difficulty of choosing between each positive, real root(s) in the solution. The solution techniques of Eq. (6-12) apply here, too, so

$$C = \hat{L}_2 \cdot \vec{r}_{site_2}$$

$$r_2^8 - \left(d_1^2 + 2Cd_1 + r_{site_2}^2 \right) r_2^6 - 2\mu(Cd_2 + d_1d_2)r_2^3 - \mu^2 d_2^2 = 0 \quad (6-16)$$

Once we pick an appropriate root, we can update our guess for the f and g series coefficient, u . After recalculating the coefficients, c_i , and solving the full matrix relation, we can find all three position vectors.

At this point, we could stop processing if we don't need high accuracy, but it's often better to continue by using one of several schemes. For example, invoking another technique for determining initial orbits, such as **GIBBS** (Algorithm 48), **Herrick-Gibbs** (Algorithm 49), or **Lambert-Universal** (Algorithm 52), allows us to determine the velocity at the middle time. Then we can reevaluate the f and g series coefficients using the first-derivative terms because, by differentiation,

$$\dot{u} = \frac{-3\mu \dot{r}_2}{r_2^4}$$

We must be sure to use the radial component of the middle velocity vector and not the magnitude of the velocity vector. In addition, Prussing and Conway (1993, 180) state that the f and g *functions* should be used instead of the series coefficients on all preceding iterations. This is reasonable, because the estimated velocity allows us to use the more "exact" functions.

Now, solve the matrix relation again and repeat the process until the slant-range values converge. Final solution of the position vectors yields the desired results. The iteration portion of the algorithm improves the routine's accuracy.

Implementing the Gaussian, Angles-only Algorithm

The overall process may seem complex; however, the key to initial orbit determination is determining the slant-range value after solving the eighth-order relation. Unlike **SITE-TRACK**, this technique seldom finds an exact answer, even though there is enough information. It's best suited for use with the differential-correction techniques we'll discuss in Chap. 9.

ALGORITHM 46: ANGLES-ONLY GAUSS $(\hat{L}_1, \hat{L}_2, \hat{L}_3, JD_1, JD_2,$

$$JD_3, \vec{r}_{site_1}, \vec{r}_{site_2}, \vec{r}_{site_3} \Rightarrow \vec{r}_2, \vec{v}_2)$$

$$\tau_1 = JD_1 - JD_2 \text{ and } \tau_3 = JD_3 - JD_2$$

$$a_1 = \frac{\tau_3}{\tau_3 - \tau_1} \quad a_{1u} = \frac{\tau_3 ((\tau_3 - \tau_1)^2 - \tau_3^2)}{6(\tau_3 - \tau_1)}$$

$$a_3 = -\frac{\tau_1}{\tau_3 - \tau_1} \quad a_{3u} = \frac{\tau_1 ((\tau_3 - \tau_1)^2 - \tau_1^2)}{6(\tau_3 - \tau_1)}$$

$$\mathbf{L}^{-1} = \frac{\begin{bmatrix} L_{y2}L_{z3} - L_{y3}L_{z2} & L_{y1}L_{z3} - L_{y3}L_{z1} & L_{y1}L_{z2} - L_{y2}L_{z1} \\ L_{x2}L_{z3} - L_{x3}L_{z2} & L_{x1}L_{z3} - L_{x3}L_{z1} & L_{x1}L_{z2} - L_{x2}L_{z1} \\ L_{x2}L_{y3} - L_{x3}L_{y2} & L_{x1}L_{y3} - L_{x3}L_{y1} & L_{x1}L_{y2} - L_{x2}L_{y1} \end{bmatrix}}{|L|}$$

$$[\mathbf{M}] = [\mathbf{L}^{-1}] [\mathbf{r}_{site}]$$

$$d_1 = M_{21}a_1 - M_{22} + M_{23}a_3$$

$$d_2 = M_{21}a_{1u} + M_{23}a_{3u}$$

$$C = \hat{L}_2 \cdot \vec{r}_{site_2}$$

$$r_2^8 - \left(d_1^2 + 2Cd_1 + r_{site_2}^2 \right) r_2^6 - 2\mu(Cd_2 + d_1d_2)r_2^3 - \mu^2 d_2^2 = 0$$

Solve for the correct real root, r_2 , and calculate

$$u = \frac{\mu}{r_2^3}$$

$$c_1 = a_1 + a_{1u}u$$

$$c_2 = -1$$

$$c_3 = a_3 + a_{3u}u$$

$$\begin{bmatrix} c_1 \rho_1 \\ c_2 \rho_2 \\ c_3 \rho_3 \end{bmatrix} = [\mathbf{M}] \begin{bmatrix} -c_1 \\ -c_2 \\ -c_3 \end{bmatrix}$$

Iterate to refine the initial estimate of the slant ranges:

LOOP

$$\begin{bmatrix} \hat{\mathbf{r}}_1 \\ \hat{\mathbf{r}}_2 \\ \hat{\mathbf{r}}_3 \end{bmatrix} = \begin{bmatrix} \rho_1 \hat{\mathbf{L}}_1 + \hat{\mathbf{r}}_{site_1} \\ \rho_2 \hat{\mathbf{L}}_2 + \hat{\mathbf{r}}_{site_2} \\ \rho_3 \hat{\mathbf{L}}_3 + \hat{\mathbf{r}}_{site_3} \end{bmatrix}$$

Use **HGIBBS** (small $\Delta\nu$) or **GIBBS** (large $\Delta\nu$) to find $\hat{\mathbf{v}}_2$ Find semiparameter p using **ELORB**

$$f_i = 1 - \left(\frac{r_i}{p} \right) (1 - \cos(\Delta\nu_i)) \quad i = 1, 3$$

$$g_i = \frac{r_i r_2 \sin(\Delta\nu_i)}{\sqrt{\mu p}} \quad i = 1, 3$$

$$c_1 = \frac{g_3}{f_1 g_3 - f_3 g_1} \quad c_3 = \frac{-g_1}{f_1 g_3 - f_3 g_1}$$

Compute and compare new slant-range values

$$\begin{bmatrix} c_1 \rho_1 \\ c_2 \rho_2 \\ c_3 \rho_3 \end{bmatrix} = [\mathbf{M}] \begin{bmatrix} -c_1 \\ -c_2 \\ -c_3 \end{bmatrix}$$

UNTIL slant ranges stop changing

$$\begin{bmatrix} \hat{\mathbf{r}}_1 \\ \hat{\mathbf{r}}_2 \\ \hat{\mathbf{r}}_3 \end{bmatrix} = \begin{bmatrix} \rho_1 \hat{\mathbf{L}}_1 + \hat{\mathbf{r}}_{site_1} \\ \rho_2 \hat{\mathbf{L}}_2 + \hat{\mathbf{r}}_{site_2} \\ \rho_3 \hat{\mathbf{L}}_3 + \hat{\mathbf{r}}_{site_3} \end{bmatrix}$$

6.4.3 Double r -iteration

Escobal ([1965] 1985, 272–288) describes an interesting combination of numerical and dynamical techniques to solve the angles-only problem, so I won't derive the complete algorithm. The algorithm is important because it replaces the Gaussian technique for observations which are far apart. The method can handle observations that are days apart. A complete system of orbit determination requires this routine. Treating the iterations is

the difficult part. Long et al. (1989, 9-19 to 9-52) use four steps to arrive at a solution. The first step bounds the guesses from the available information. The second section, detailed here, is the actual double- r iteration. The iteration techniques use this second portion to determine intermediate guesses, so it's important to have a modular routine that is easily accessible. The third section begins the formal iterative process. It tries to align the times with the estimated values of the orbits. Finally, a type of differential correction determines the answer. I've changed the signs on the site vectors from Escobal to maintain compatibility with Eq. (6-5). Consult Escobal ([1965] 1985, 281) for hyperbolic relations.

ALGORITHM 47: ANGLES-DOUBLE R

$$(\hat{L}_1, \hat{L}_2, \hat{L}_3, JD_1, JD_2, JD_3, \vec{r}_{site_1}, \vec{r}_{site_2}, \vec{r}_{site_3} \Rightarrow \vec{r}_2, \vec{v}_2)$$

$$\tau_1 = JD_1 - JD_2 \text{ and } \tau_3 = JD_3 - JD_2$$

Guess r_1 or let $r_1 = 1.0 \text{ ER}$

Guess r_2 or let $r_2 = 1.1 \text{ ER}$

LOOP

$$\left. \begin{aligned} c_i &= 2\hat{L}_i \cdot \vec{r}_{site_i} \\ \rho_i &= \frac{-c_i + \sqrt{c_i^2 - 4(r_{site_i}^2 - r_i^2)}}{2} \end{aligned} \right\} i = 1, 2$$

$$\vec{r}_i = \rho_i \hat{L}_i + \vec{r}_{site_i} \quad i = 1, 2$$

$$\hat{W} = \frac{\vec{r}_1 \times \vec{r}_2}{|\vec{r}_1||\vec{r}_2|}$$

$$\rho_3 = \frac{\vec{r}_{site_3} \cdot \hat{W}}{\hat{L}_3 \cdot \hat{W}}$$

$$\vec{r}_3 = \rho_3 \hat{L}_3 + \vec{r}_{site_3}$$

FOR $j = 2, 3$ and $k = 1, 2$,

$$\cos(\Delta\nu_{jk}) = \frac{\vec{r}_j \cdot \vec{r}_k}{r_j r_k}$$

$$\sin(\Delta\nu_{jk}) = \frac{|\vec{r}_j \times \vec{r}_k|}{r_j r_k} \quad \sin(\Delta\nu_{jk}) = \sqrt{1 - \cos^2(\Delta\nu_{jk})}$$

IF $\Delta\nu > 180^\circ$

$$\left[\begin{array}{l} c_1 = \frac{r_2 \sin(\Delta\nu_{32})}{r_1 \sin(\Delta\nu_{31})} \quad c_3 = \frac{r_2 \sin(\Delta\nu_{21})}{r_3 \sin(\Delta\nu_{31})} \\ p = \frac{c_1 r_1 + c_3 r_3 - r_2}{c_1 + c_3 - 1} \\ \text{ELSE} \\ c_1 = \frac{r_1 \sin(\Delta\nu_{31})}{r_2 \sin(\Delta\nu_{32})} \quad c_3 = \frac{r_1 \sin(\Delta\nu_{21})}{r_3 \sin(\Delta\nu_{32})} \\ p = \frac{c_3 r_3 + c_1 r_2 - r_1}{-c_1 + c_3 + 1} \\ e \cos(\nu_i) = \frac{p}{r_i} - 1, i = 1, 2, 3 \end{array} \right]$$

IF $\Delta\nu \neq \pi$,

$$e \sin(\nu_2) = \frac{-\cos(\Delta\nu_{21}) \cos(\nu_2) + e \cos(\nu_1)}{\sin(\Delta\nu_{21})}$$

ELSE

$$e \sin(\nu_2) = \frac{\cos(\Delta\nu_{32}) e \cos(\nu_2) - e \cos(\nu_3)}{\sin(\Delta\nu_{31})}$$

$$e = \sqrt{(e \cos(\Delta\nu_2))^2 + (e \sin(\Delta\nu_2))^2}$$

$$a = \frac{p}{1 - e^2} \quad n = \sqrt{\frac{\mu}{a^3}}$$

$$S = \frac{r_2}{p} \sqrt{1 - e^2} e \sin(\Delta\nu_2) \quad C = \frac{r_2}{p} \{ e^2 + e \cos(\Delta\nu_2) \}$$

$$\sin(\Delta E_{32}) = \frac{r_3}{\sqrt{ap}} \sin(\Delta\nu_{32}) - \frac{r_3}{p} (1 - \cos(\Delta\nu_{32})) S$$

$$\cos(\Delta E_{32}) = 1 - \frac{r_2 r_3}{ap} (1 - \cos(\Delta\nu_{32}))$$

$$\sin(\Delta E_{21}) = \frac{r_1}{\sqrt{ap}} \sin(\Delta\nu_{21}) + \frac{r_1}{p} (1 - \cos(\Delta\nu_{21})) S$$

$$\cos(\Delta E_{21}) = 1 - \frac{r_2 r_1}{ap} (1 - \cos(\Delta\nu_{21}))$$

$$\Delta M_{32} = \Delta E_{32} + 2S \sin^2\left(\frac{\Delta E_{32}}{2}\right) - C \sin(\Delta E_{32})$$

$$\Delta M_{12} = -\Delta E_{21} + 2S \sin^2\left(\frac{\Delta E_{21}}{2}\right) + C \sin(\Delta E_{21})$$

$$F_1 = \tau_1 - \frac{\Delta M_{12}}{n} \quad F_2 = \tau_3 - \frac{\Delta M_{32}}{n}$$

Estimate the accuracy of each pass

$$Q = \sqrt{F_1^2 + F_2^2}$$

Repeat all calculations for $F_1(r_1 + \Delta r_1, r_2)$, let $\Delta r_1 \approx 0.04\%$ of r_1

$$\frac{\partial F_1}{\partial r_1} = \frac{F_1(r_1 + \Delta r_1, r_2) - F_1}{\Delta r_1} \quad \frac{\partial F_2}{\partial r_1} = \frac{F_2(r_1 + \Delta r_1, r_2) - F_2}{\Delta r_1}$$

Repeat all calculations for $F_2(r_1, r_2 + \Delta r_2)$, let $\Delta r_2 \approx 0.04\%$ of r_2

$$\frac{\partial F_1}{\partial r_2} = \frac{F_1(r_1, r_2 + \Delta r_2) - F_1}{\Delta r_2} \quad \frac{\partial F_2}{\partial r_2} = \frac{F_2(r_1, r_2 + \Delta r_2) - F_2}{\Delta r_2}$$

$$\Delta = \frac{\partial F_1}{\partial r_1} \left(\frac{\partial F_2}{\partial r_2} \right) - \frac{\partial F_2}{\partial r_1} \left(\frac{\partial F_1}{\partial r_2} \right)$$

$$\Delta_1 = \frac{\partial F_2}{\partial r_2} F_1 - \frac{\partial F_1}{\partial r_2} F_2$$

$$\Delta_2 = \frac{\partial F_1}{\partial r_1} F_2 - \frac{\partial F_2}{\partial r_1} F_1$$

$$\Delta r_1 = -\frac{\Delta_1}{\Delta}$$

$$\Delta r_2 = -\frac{\Delta_2}{\Delta}$$

UNTIL converged

$$f = 1 - \frac{a}{r_2} (1 - \cos(\Delta E_{32}))$$

$$g = \tau_3 - \sqrt{\frac{a^3}{\mu}} (\Delta E_{32} - \sin(\Delta E_{32}))$$

$$\dot{v}_2 = \frac{\dot{r}_3 - f \dot{r}_2}{g}$$

▼ **Example 6-2. Determining Initial Orbits Using Angles Only.**

GIVEN: Three sets of right ascension-declination observations (topocentric) from a station at 40°N, 110°W, alt = 2000 m, on August 20, 1996. (Use observations number 6, 10, and 14.)

Obs. #	h min s	α_t	δ_t
1	08: 05: 00.00	72.763 895 2	-16.729 648 1
2	08: 10: 00.00	79.488 195 6	-6.190 545 1
3	08: 15: 00.00	87.345 686 6	4.192 062 0
4	08: 20: 00.00	96.526 058 6	13.738 363 4
5	08: 25: 00.00	107.059 583 0	21.721 533 4
6	08: 30: 00.00	118.678 087 2	27.577 795 2
7	08: 35: 00.00	130.760 092 4	31.078 542 6
8	08: 40: 00.00	142.496 621 8	32.370 637 5
9	08: 45: 00.00	153.209 474 6	31.865 843 1
10	08: 50: 00.00	162.558 046 5	30.060 500 2
11	08: 55: 00.00	170.516 250 4	27.396 753 6
12	09: 00: 00.00	177.235 675 3	24.207 630 3
13	09: 05: 00.00	182.926 035 4	20.720 368 0
14	09: 10: 00.00	187.790 712 3	17.081 080 2
15	09: 15: 00.00	192.002 956 5	13.380 074 2

FIND: Position and velocity vectors for the observation using the Gaussian method.

First, find the Julian dates, and the change in times between the observations:

$$\tau_1 = JD_1 - JD_2 = 2,450,315.854\ 167 - 2,450,315.868\ 056 \Rightarrow -20.0\ \text{min} = -1.487\ 337\ \text{TU}$$

$$\tau_3 = JD_3 - JD_2 = 2,450,315.868\ 056 - 2,450,315.881\ 944 \Rightarrow 20.0\ \text{min} = 1.487\ 337\ \text{TU}$$

Next, find the line-of-sight unit vector. Remember that each observation produces one unit vector.

$$\hat{L}_i = \begin{bmatrix} \cos(\delta_i) \cos(\alpha_i) \\ \cos(\delta_i) \sin(\alpha_i) \\ \sin(\delta_i) \end{bmatrix}, \mathbf{L} = \begin{bmatrix} -0.4254 & -0.8257 & -0.9471 \\ 0.7776 & 0.2594 & -0.1296 \\ 0.4629 & 0.5009 & 0.2937 \end{bmatrix}$$

Find the matrix of site vectors using Algorithm 45.

$$\mathbf{r}_{site} = \begin{bmatrix} 0.7460 & 0.7589 & 0.7659 \\ -0.1797 & -0.1138 & -0.0471 \\ 0.6396 & 0.6396 & 0.6396 \end{bmatrix}$$

The change in times lets you estimate parameters that lead to finding the middle range magnitude.

$$a_1 = \frac{\tau_3}{\tau_3 - \tau_1}, \quad a_{1u} = \frac{\tau_3 \left((\tau_3 - \tau_1)^2 - \tau_3^2 \right)}{6(\tau_3 - \tau_1)}, \quad a_1 = 0.5 \text{ and } a_{1u} = 0.553\ 043$$

$$a_3 = \frac{\tau_1}{\tau_3 - \tau_1}, \quad a_{3u} = \frac{\tau_1 \left((\tau_3 - \tau_1)^2 - \tau_1^2 \right)}{6(\tau_3 - \tau_1)}, \quad a_3 = 0.5 \text{ and } a_{3u} = 0.553\ 043$$

You also need the inverse of the unit vector:

$$\mathbf{L}^{-1} = \frac{\begin{bmatrix} L_{y2}L_{z3}-L_{y3}L_{z2} & L_{y1}L_{z3}-L_{y3}L_{z1} & L_{y1}L_{z2}-L_{y2}L_{z1} \\ L_{x2}L_{z3}-L_{x3}L_{z2} & L_{x1}L_{z3}-L_{x3}L_{z1} & L_{x1}L_{z2}-L_{x2}L_{z1} \\ L_{x2}L_{y3}-L_{x3}L_{y2} & L_{x1}L_{y3}-L_{x3}L_{y1} & L_{x1}L_{y2}-L_{x2}L_{y1} \end{bmatrix}}{|L|} = \begin{bmatrix} -1.8311 & 3.0089 & -4.5769 \\ 3.7426 & -4.0684 & 10.2726 \\ -3.4964 & 2.1956 & -6.9006 \end{bmatrix}$$

Using matrix operations, form intermediate relations ($\mathbf{M} = \mathbf{L}^{-1} [\mathbf{r}_{site}]$), and ultimately determine parameters for use in the eighth-order equation:

$$d_1 = M_{21}a_1 - M_{22} + M_{23}a_3 = -0.012\,638$$

$$d_2 = M_{21}a_{1u} + M_{23}a_{3u} = 10.906\,609$$

The eighth-order equation is

$$C = \hat{\mathbf{L}}_2 \cdot \dot{\mathbf{r}}_{site_2} = -0.335\,748$$

$$r_2^8 - (d_1^2 + 2Cd_1 + r_{site_2}^2) r_2^6 - 2\mu(Cd_2 + d_1d_2) r_2^3 - \mu^2 d_2^2 = 0$$

$$r_2^8 - 1.006\,517 r_2^6 + 7.599\,428 r_2^3 - 118.954\,139 = 0$$

The real roots for this equation are 1.797 022 and -1.983 452. Use the first root, r_2 , and calculate

$$u = \frac{\mu}{r_2^3} = 0.172\,322$$

This value allows you to find an initial estimate of the slant-range values, which ends the Gaussian part of the solution.

$$\begin{bmatrix} -c_1 \\ -c_2 \\ -c_3 \end{bmatrix} = \begin{bmatrix} -a_1 - a_{1u}u \\ 1 \\ -a_3 - a_{3u}u \end{bmatrix} \text{ and } \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} -0.595\,301 \\ 1.0 \\ -0.595\,301 \end{bmatrix}$$

Determine the initial guess of the slant ranges.

$$\begin{bmatrix} c_1\rho_1 \\ c_2\rho_2 \\ c_3\rho_3 \end{bmatrix} = [\mathbf{M}] \begin{bmatrix} -c_1 \\ -c_2 \\ -c_3 \end{bmatrix} \Rightarrow \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \end{bmatrix} = \begin{bmatrix} 0.880\,161 \\ -1.866\,807 \\ 1.381\,339 \end{bmatrix} \text{ ER}$$

Next, refine the initial estimate. Iterate the following equations:

$$\begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \mathbf{r}_3 \end{bmatrix} = \begin{bmatrix} \rho_1 \hat{\mathbf{L}}_1 + \dot{\mathbf{r}}_{site_1} \\ \rho_2 \hat{\mathbf{L}}_2 + \dot{\mathbf{r}}_{site_2} \\ \rho_3 \hat{\mathbf{L}}_3 + \dot{\mathbf{r}}_{site_3} \end{bmatrix} = \begin{bmatrix} 0.117\,100 & 0.970\,064 & 1.324\,053 \\ -0.078\,197 & 0.370\,475 & 1.574\,681 \\ -1.431\,677 & -0.347\,732 & 1.321\,131 \end{bmatrix} \text{ ER}$$

Because the vectors are "far" apart, use **GIBBS** to find $\dot{\mathbf{v}}_2$. The result is

$$\dot{\mathbf{v}}_2 = -0.583\,968 \hat{\mathbf{i}} - 0.510\,943 \hat{\mathbf{j}} - 0.022\,495 \hat{\mathbf{k}} \text{ ER/TU}$$

Now, find the f and g functions to update your guess.

$$f_i = 1 - \left(\frac{r_i}{p} \right) (1 - \cos(\Delta\nu_i)), f_1 = 0.796\,615, f_3 = -1.380\,055$$

$$g_i = \frac{r_i r_2 \sin(\Delta\nu_i)}{\sqrt{\mu p}}, g_1 = 0.834\,289, g_3 = 1.409\,228$$

$$c_1 = \frac{g_3}{f_1 g_3 - f_3 g_1}, c_3 = \frac{-g_1}{f_1 g_3 - f_3 g_1}, c_1 = 0.619\,719, c_3 = 0.606\,890$$

Compare the new slant-range values to the previous estimate. When the slant-range values stop changing, you have an answer. For this particular problem, after two iterations, you converge to a solution after either the first or second iteration. The last step finds the position vectors at each of the observation times.

$$\begin{bmatrix} \hat{r}_1 \\ \hat{r}_2 \\ \hat{r}_3 \end{bmatrix} = \begin{bmatrix} \rho_1 \hat{L}_1 + \hat{r}_{site_1} \\ \rho_2 \hat{L}_2 + \hat{r}_{site_2} \\ \rho_3 \hat{L}_3 + \hat{r}_{site_3} \end{bmatrix}, \text{ and } \begin{bmatrix} \hat{r}_1 \\ \hat{r}_2 \\ \hat{r}_3 \end{bmatrix} = \begin{bmatrix} 0.025\,297 & 1.137\,897 & 1.423\,968 \\ -1.078\,198 & 0.463\,357 & 1.754\,025 \\ -1.802\,426 & -0.398\,457 & 1.436\,116 \end{bmatrix}$$



6.5 Mixed Observations

It would be nice to imagine that all observations fit the previous models of range-azimuth-elevation and right ascension and declination. Unfortunately, this isn't true. Sensors sometimes report only range, whereas others report only angles, and so on. Escobal ([1965] 1985, Chap. 8) details several techniques that use combinations of available data. Here, we'll examine only range and range-rate processing, as well as range-only processing.

It's important to distinguish between techniques for determining initial orbits and the data available for processing. For instance, *Satellite Laser Ranging* (SLR) is becoming very popular as a highly accurate way to determine orbits. Unfortunately, there is no *direct* technique to process the resulting range-only information without already knowing the orbit. Thus, SLR data isn't a technique; it's an input. The range-only technique we'll introduce is a numerical method to process this data.

The overall topic of mixed observations is closely related to the *differential correction* of orbits (Chap. 9). Specific techniques to process range-only, range and range-rate information, and so forth, are very desirable when viewed with differential correction because they allow us directly to formulate the partial-derivative matrices. If we can't analytically formulate the problem of initial orbit determination, we must modify the differential-correction process to *simulate* the available data. Although that's possible, a direct technique is sometimes better.

6.5.1 Range and Range-Rate Processing

Escobal ([1965] 1985, 309–314) discusses a method of *trilateration* that involves processing at least three simultaneous observations of a satellite from three separate sensors.

Because the data consists only of range and range-rate information, we must have at least six observations for a solution. Simultaneous observations are necessary to form the solution but may be difficult. Of course, space-based sensors *could* provide the necessary observations. Moreover, GPS receivers are beginning to permit satellites to obtain continuous data we can use in orbit-determination schemes. If the observations are simultaneous, the solution is exact, and we don't need to iterate. It's also possible to take closely spaced observations and fit high-order polynomials through the data to interpolate "simultaneous" observations, as long as we have highly accurate observations *close* to a given epoch. As space-surveillance techniques become more sophisticated, incorporating highly accurate laser measurements and GPS data could give us the chance to use these analyses.

To begin the analysis, consider the vector relations of Eq. (6-5) and Eq. (6-6). Recognize that you don't need unit vectors. Solving for the slant range results in

$$\begin{aligned}\hat{\rho}_i &= \hat{r} - \hat{r}_{site_i} \\ \rho_i^2 &= r^2 - 2\hat{r} \cdot \hat{r}_{site_i} + r_{site_i}^2, \quad i = 1 \dots 3\end{aligned}\quad (6-17)$$

The position vector doesn't have subscripts because all observations are simultaneous, and you can eliminate it by using the slant-range magnitude equation. By subtracting Eq. (6-17) with $i = 1$ from Eq. (6-17) with $i = 2, 3$, and leaving the dot product on one side, you get

$$\begin{aligned}\frac{1}{2} [\rho_3^2 - \rho_1^2 - (r_{site3}^2 - r_{site1}^2)] &= \hat{r} \cdot (\hat{r}_{site3} - \hat{r}_{site1}) \\ \frac{1}{2} [\rho_2^2 - \rho_1^2 - (r_{site2}^2 - r_{site1}^2)] &= \hat{r} \cdot (\hat{r}_{site2} - \hat{r}_{site1})\end{aligned}$$

Now, replace the left-hand side with temporary variables (d_3, d_2). Then, separate the dot-product terms:

$$\begin{aligned}d_3 &= (r_{site3I} - r_{site1I}) r_I + (r_{site3J} - r_{site1J}) r_J + (r_{site3K} - r_{site1K}) r_K \\ d_2 &= (r_{site2I} - r_{site1I}) r_I + (r_{site2J} - r_{site1J}) r_J + (r_{site2K} - r_{site1K}) r_K\end{aligned}\quad (6-18)$$

Because the ultimate goal is to determine the position vector, isolate the J and K terms by multiplying Eq. (6-18) by $(r_{site2J} - r_{site1J})$ and $(r_{site3I} - r_{site1I})$, respectively. The result is lengthy, and Escobal defines several new terms to simplify the notation:

$$\begin{aligned}D &= (r_{site3K} - r_{site1K}) (r_{site2J} - r_{site1J}) - (r_{site2K} - r_{site1K}) (r_{site3J} - r_{site1J}) \\ a_1 &= \frac{(r_{site2I} - r_{site1I}) (r_{site3J} - r_{site1J}) - (r_{site3I} - r_{site1I}) (r_{site2J} - r_{site1J})}{D} \\ a_2 &= \frac{d_3 (r_{site2J} - r_{site1J}) - d_2 (r_{site3J} - r_{site1J})}{D}\end{aligned}$$

$$a_3 = \frac{(r_{site2I} - r_{site1I}) (r_{site3K} - r_{site1K}) - (r_{site3I} - r_{site1I}) (r_{site2K} - r_{site1K})}{-D}$$

$$a_4 = \frac{d_3 (r_{site2K} - r_{site1K}) - d_2 (r_{site3K} - r_{site1K})}{-D}$$

The result is

$$\begin{aligned} r_J &= a_3 r_I + a_4 \\ r_K &= a_1 r_I + a_2 \end{aligned} \quad (6-19)$$

We know that $r^2 = r_I^2 + r_J^2 + r_K^2$. If you use this expression with Eq. (6-19) and substitute it into Eq. (6-17), a quadratic equation in r_I results:

$$\begin{aligned} &\left\{ a_1^2 + a_3^2 + 1 \right\} r_I^2 + \left\{ 2(a_1 a_2 + a_3 a_4 - r_{site1I} - a_3 r_{site1J} - a_1 r_{site1K}) \right\} r_I \\ &+ \left\{ a_2^2 + a_4^2 - 2a_4 r_{site1J} - 2a_1 r_{site1K} + r_{site1}^2 - \rho_1^2 \right\} = 0 \end{aligned} \quad (6-20)$$

The quadratic results in two roots. You can reject the root that yields an angle greater than 90° because it implies the satellite is not in the field-of-view. Therefore,

$$\cos(\theta) = \frac{\vec{r}_{site1} \cdot \vec{\rho}_1}{r_{site1} \rho_1}$$

Once r_I is known, you can use Eq. (6-19) to find r_J and r_K , which determine the position vector. Consult Escobal ([1965] 1985, 311–312) for the derivation on range-rate information.

$$E_i = \rho_i \dot{\rho}_i - \dot{r}_{site_i} \cdot \vec{\rho}_i, \quad i = 1, 2, 3$$

$$\mathbf{M} = \begin{bmatrix} \vec{\rho}_1 \\ \vec{\rho}_2 \\ \vec{\rho}_3 \end{bmatrix} \quad \vec{v} = \mathbf{M}^{-1} \begin{bmatrix} E_1 \\ E_2 \\ E_3 \end{bmatrix}$$

6.5.2 Range-only Processing

A method of range-only processing appears in Smith and Huang (1985) and Carter (1996). Its solution is more numerical than analytical; it allows us to process information which is *not* simultaneous and is from a single site. I'll introduce the basic concepts in this section.

In principle, any six independent time-tagged observations of a satellite give us six nonlinear equations which we may attempt to solve to determine the orbital elements at a specified epoch. We must model the observations only as functions of the observation time, the epoch elements, and the epoch time.

To see how this works in the simplest case, imagine an observation whose value at time t is a known function of time and the satellite elements at that time. Ignoring light-time correction, the range to the satellite from a fixed tracking station is an example—*inertial* position of the tracking station depends on t (the Earth rotates!) and the *inertial* position of the satellite depends on the orbital elements at time t . Range is the magnitude of the difference of these two inertial positions.

We may view this observation as given by a function of time, epoch state, and epoch time which is evaluated as follows:

1. Propagate the orbit from epoch time t_0 to time t of the observation, given the epoch state X_0 .
2. Evaluate the observation, given the state at time t , X .

The technique applies if we are given six instances of a particular observation at six different times—for instance, six different time-tagged range observations from a fixed tracking location. It's convenient to abstract the problem even further by thinking of the six observations as a single observation vector (\mathbf{b}). We now have the state as a function of the observations, $X = f(\mathbf{b})$. We can also determine the observations as a function of the state $\mathbf{b} = g(X)$.

Here, X denotes the epoch elements we seek, \mathbf{b} denotes our vector of six observations, and various modeling details (epoch time, observation times, etc.) have been hidden in the definition of f . Clearly, we can apply any numerical technique for solving simultaneous nonlinear equations. The best-known numerical method is the Newton-Raphson iteration [Eq. (4-50)]. In orbit determination, the Newton-Raphson iteration is sometimes called the method of *differential correction* (which we'll explore in Chap. 9), and it generalizes to the setting in which there are (many) more observations than unknowns. Unfortunately, we usually need a very good initial guess for differential correction to converge, so the method is not of much use for deterministic orbit determination. Indeed, deterministic orbit determination is often used as a starter for differential correction.

Such discouraging comments aside, recent progress in numerical solution of nonlinear equations has given us techniques whose convergence properties are more global than Newton-Raphson. The *homotopy continuation method*, first applied to deterministic orbit determination in the NASA-sponsored research of Smith and Huang, is of particular interest. We wish to find X such that $f(X) = \mathbf{b}$. Let X_0 be an initial guess and put $\mathbf{b}_0 = f(X_0)$. Put $\mathbf{b}_1 = \mathbf{b}$ and for each real number λ put

$$\mathbf{b}_\lambda = (1 - \lambda)\mathbf{b}_0 + \lambda\mathbf{b}_1$$

We'll seek a continuous family (X_λ) of solutions to $f(X_\lambda) = \mathbf{b}_\lambda$. Note that any solution corresponding to $\lambda = 1$ is a solution to our original equation.

Differentiating with respect to a parameter s (which is the arc length from \mathbf{b}_0) yields

$$\frac{\partial f}{\partial \mathbf{X}} \frac{d\mathbf{X}}{ds} = (\mathbf{b}_1 - \mathbf{b}_0) \frac{d\lambda}{ds}$$

or in matrix form,

$$\begin{bmatrix} \frac{\partial f}{\partial \mathbf{X}} & (\mathbf{b}_1 - \mathbf{b}_0) \end{bmatrix} \begin{bmatrix} \frac{d\mathbf{X}}{ds} \\ \frac{d\lambda}{ds} \end{bmatrix} = 0$$

Note that this is of the form

$$\tilde{J} \frac{d\tilde{\mathbf{X}}}{ds} = 0$$

where $\tilde{\mathbf{X}}$ is the augmented state

$$\begin{bmatrix} \mathbf{X} \\ \lambda \end{bmatrix}$$

We can propagate this relationship numerically. It's an initial-value problem because we know

$$\tilde{\mathbf{X}}(0) = \begin{bmatrix} \mathbf{X}_0 \\ 0 \end{bmatrix}$$

To continue, it's convenient to have a functional description of the augmented state derivative—that is, to express $\tilde{J} \frac{d\tilde{\mathbf{X}}}{ds} = 0$ in the form

$$\frac{d\tilde{\mathbf{X}}}{ds} = h(\tilde{\mathbf{X}}, s)$$

But establishing this expression is possible. The function h can be any globally defined map which returns a unit vector in the null-space of \tilde{J} .

Solving these equations is the basis for the homotopy method. Consult Smith and Huang's publications (1985, 1986) as well as Hart (1991) for more information.

6.6 Three Position Vectors and Time

The previous methods provide data which we often have to process further in order to determine the initial orbit. In the **SITE-TRACK** discussion, I mentioned that some pieces of observational data may be missing (specifically, range or range-rate, and angle-rate information). In certain cases, we may have data from a sensor site which permits us to develop only position vectors. Although they're useful, they don't allow us to calculate all six orbital elements because several of them depend on the velocity. Thus, we sometimes need methods which let us generate position and velocity information from position-only data. The simplest solutions use three position vectors [and times (*UTC*) for some methods] to develop the required data. We commonly use the **Gibbs** and **Herrick-Gibbs** methods. Both methods are necessary because the Gibbs solution is geometrical, so it fails when the vectors are closely spaced. The **Herrick-Gibbs** method works best when the vectors are relatively close together because it's basically a Taylor-series expansion about the solution.

6.6.1 Gibbs Method

The Gibbs method uses three position vectors to determine the orbit. Solving **GIBBS** relies on knowing the Gauss formulation. Indeed, the first few steps are actually a variant of the original Gaussian solution. We form the Gibbs problem supposing we know three nonzero, coplanar position vectors, which represent *three time-sequential vectors* of a satellite in its orbit. These assumptions are important for a solution. Requiring the vectors to be nonzero simply prevents divide-by-zero operations, as well as needless analysis of orbits that don't exist. The sequential requirement is *very* important because we consider the vectors sequential while forming the solution and take several cross products based on the given order. Changing from a sequential order will give us erroneous results. Finally, requiring the vectors to be coplanar is important, but we can relax it somewhat in practice. Because real-world observations aren't perfect, we may obtain vectors which deviate by a degree or two from each other (in inclination). It's important to define this desired tolerance for the coplanar requirement before solving this problem. To describe the procedure, assume the vectors are coplanar. Figure 6-7 depicts the geometry for the Gibbs problem.

Notice the spacing between the vectors. As mentioned earlier, this procedure is basically vector analysis (this isn't surprising, because Gibbs is considered a pioneer of vector analysis). If the vectors are too close together, numerical errors ensue, as I'll discuss later.

My discussion follows Bate, Mueller, and White (1971, 109–115). The overall procedure is to find a constant (the middle velocity vector) which is common between each of the given vectors. Because the three vectors $\vec{r}_1, \vec{r}_2, \vec{r}_3$ must be coplanar, only two can be independent. Therefore, the third vector must be a linear combination of the first two, and we can find three constants, c_i , to satisfy the relation of Eq. (6-14). Next, we need a method to isolate the coefficients. Start by crossing Eq. (6-14) with \vec{r}_1, \vec{r}_2 , and \vec{r}_3 respectively, recalling that $\vec{r}_i \times \vec{r}_i = 0$:

$$\begin{aligned} \vec{c}_2 \vec{r}_1 \times \vec{r}_2 &= \vec{c}_3 \vec{r}_3 \times \vec{r}_1 \\ \vec{c}_1 \vec{r}_1 \times \vec{r}_2 &= \vec{c}_3 \vec{r}_2 \times \vec{r}_3 \\ \vec{c}_1 \vec{r}_3 \times \vec{r}_1 &= \vec{c}_2 \vec{r}_2 \times \vec{r}_3 \end{aligned} \quad (6-21)$$

Dot Eq. (6-14) with the eccentricity vector:

$$\hat{\vec{e}} \cdot \vec{c}_1 \vec{r}_1 + \hat{\vec{e}} \cdot \vec{c}_2 \vec{r}_2 + \hat{\vec{e}} \cdot \vec{c}_3 \vec{r}_3 = 0 \quad (6-22)$$

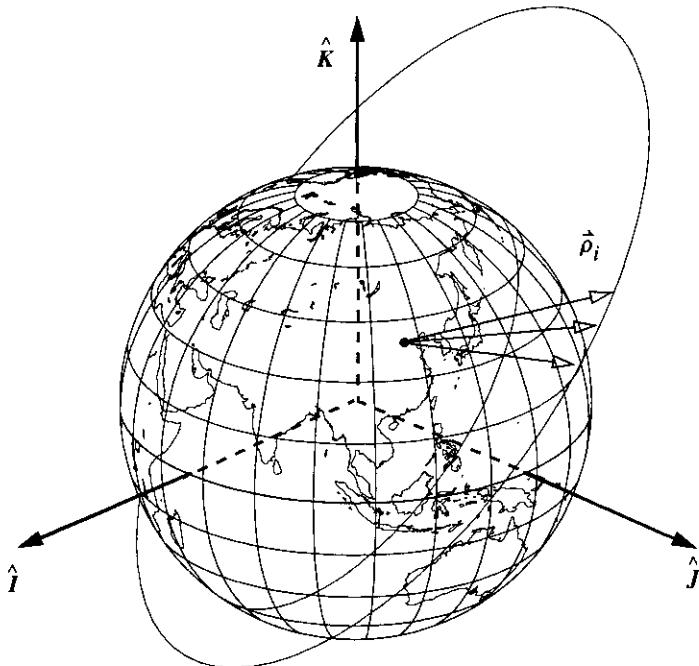


Figure 6-7. Orbit Geometry for the Gibbs Method. To determine the initial orbit using the Gibbs method, we must have at least three sequential observations.

Now, simplify this relation because $\hat{\vec{e}} \cdot \vec{r} = er\cos(\nu)$. Recall the angle between the eccentricity and position vector is the true anomaly. Expressing the trajectory equation as

$$p = r(1 + e\cos(\nu)) = r + re\cos(\nu) = r + \hat{\vec{e}} \cdot \vec{r}$$

produces an intermediate result:

$$\hat{\vec{e}} \cdot \vec{r}_i = p - r_i \quad (6-23)$$

Substituting this expression into Eq. (6-22) results in

$$c_1(p - r_1) + c_2(p - r_2) + c_3(p - r_3) = 0$$

Multiply this result by $\vec{r}_3 \times \vec{r}_1$:

$$\vec{r}_3 \times \vec{r}_1 c_1(p - r_1) + \vec{r}_3 \times \vec{r}_1 c_2(p - r_2) + \vec{r}_3 \times \vec{r}_1 c_3(p - r_3) = \vec{0}$$

Substitute the relations in Eq. (6-21) into the last equation to generate an equation with only the constant c_2 :

$$c_2 \vec{r}_2 \times \vec{r}_3(p - r_1) + c_2 \vec{r}_3 \times \vec{r}_1(p - r_2) + c_2 \vec{r}_1 \times \vec{r}_2(p - r_3) = \vec{0}$$

Dividing c_2 out produces

$$p(\vec{r}_1 \times \vec{r}_2 + \vec{r}_2 \times \vec{r}_3 + \vec{r}_3 \times \vec{r}_1) = r_1(\vec{r}_2 \times \vec{r}_3) + r_2(\vec{r}_3 \times \vec{r}_1) + r_3(\vec{r}_1 \times \vec{r}_2)$$

Define the right-hand side of the equation as \vec{N} and the vector part of the left-hand side as \vec{D} :

$$\begin{aligned} \vec{D} &= \vec{r}_1 \times \vec{r}_2 + \vec{r}_2 \times \vec{r}_3 + \vec{r}_3 \times \vec{r}_1 \\ \vec{N} &= r_1(\vec{r}_2 \times \vec{r}_3) + r_2(\vec{r}_3 \times \vec{r}_1) + r_3(\vec{r}_1 \times \vec{r}_2) = p\vec{D} \end{aligned} \quad (6-24)$$

\vec{D} is a vector perpendicular to the plane formed by the three vectors (Selby, 1975, 557). In addition, \vec{N} is a vector perpendicular to the plane, as long as all three vectors are coplanar. You can check the coplanar condition by testing if $\vec{N} \cdot \vec{D} = N\vec{D}$. The vectors, \vec{N} and \vec{D} , are in the same direction for this case, which is necessary for a solution. If one of the vectors isn't coplanar, the process will break down. Obviously, you need a tolerance because the vectors could be out of plane due to observational errors or perturbing forces that disturb the satellite. About 2° or 3° seems to work well, depending on the accuracy you want for the results.

Next, recall the perifocal coordinate system. The \hat{P} vector pointed towards perigapsis, and the \hat{W} vector pointed in the direction of the angular momentum. Because the \vec{N} vector described above also points in the direction of angular momentum, you can write

$$\hat{P} = \frac{\vec{e}}{|\vec{e}|} \quad \hat{W} = \frac{\vec{N}}{|\vec{N}|}$$

The perifocal coordinate system is orthogonal, so you can define the third vector as

$$\hat{Q} = \hat{W} \times \hat{P} = \frac{\vec{N} \times \vec{e}}{|\vec{N}| |\vec{e}|}$$

Substitute the definition of \vec{N} from Eq. (6-24) and find

$$Ne\hat{Q} = \vec{N} \times \vec{e} = r_1(\vec{r}_2 \times \vec{r}_3) \times \vec{e} + r_2(\vec{r}_3 \times \vec{r}_1) \times \vec{e} + r_3(\vec{r}_1 \times \vec{r}_2) \times \vec{e}$$

Now use the general relationship for triple cross products of vectors [Eq. (C-5)], and rewrite as

$$\begin{aligned} Ne\hat{Q} = & r_1(\hat{r}_2 \cdot \hat{e})\hat{r}_3 - r_1(\hat{r}_3 \cdot \hat{e})\hat{r}_2 + r_2(\hat{r}_3 \cdot \hat{e})\hat{r}_1 - r_2(\hat{r}_1 \cdot \hat{e})\hat{r}_3 \\ & + r_3(\hat{r}_1 \cdot \hat{e})\hat{r}_2 - r_3(\hat{r}_2 \cdot \hat{e})\hat{r}_1 \end{aligned}$$

Again using the relationship in Eq. (6-23) and factoring p from the right-hand side, write

$$Ne\hat{Q} = p[(r_2 - r_3)\hat{r}_1 + (r_3 - r_1)\hat{r}_2 + (r_1 - r_2)\hat{r}_3] = p\hat{S}$$

where \hat{S} is defined by the bracketed quantity above. Notice \hat{Q} and \hat{S} must be in the same direction. Also, it's possible to write $Ne = pS$ because the magnitudes must be equal. If you solve for eccentricity and recognize the magnitudes $N = pD$ [Eq. (6-24)] must be equal, you can write

$$e = \frac{S}{D}$$

The original Gibbs formulation ended here because he had found the eccentricity and semiparameter. Although the process we've described permits us to discover the magnitude of the middle velocity vector, we can also develop an expression that gives the velocity vector directly in terms of D , N , and S . See Bate, Mueller, and White (1971, 109–114) for the original presentation. It's convenient to consider a picture of all the vectors, as in Fig. 6-8.

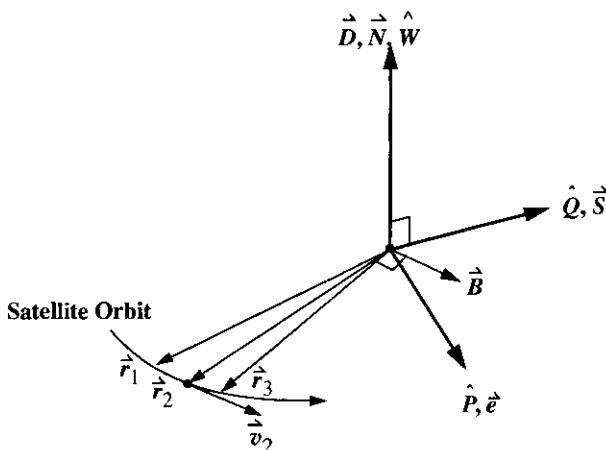


Figure 6-8. Geometry of Vectors for the Gibbs Method. It's useful to visualize the vectors which we use to develop the Gibbs solution. The vectors are not all equal. For instance, P and the eccentricity vector aren't equal, but they're in the same direction. The B vector is normal to \vec{r}_2 and \vec{D} .

From Eq. (2-13), recall the velocity vector at the middle time satisfies the expression

$$\dot{\vec{r}}_2 \times \vec{h} = \mu \left(\frac{\vec{r}_2}{r_2} + \vec{e} \right)$$

Cross the angular momentum vector into the equation above to obtain

$$\vec{h} \times (\dot{\vec{r}}_2 \times \vec{h}) = \mu \left(\frac{\vec{h} \times \vec{r}_2}{r_2} + \vec{h} \times \vec{e} \right)$$

Using the vector identity in Eq. (C-5), simplify the left side of the equation:

$$h^2 \vec{v}_2 = \mu \left(\frac{\vec{h} \times \vec{r}_2}{r_2} + \vec{h} \times \vec{e} \right)$$

You can write $\vec{h} = h\hat{W}$ and $\vec{e} = e\hat{P}$, so

$$\vec{v}_2 = \frac{\mu}{h} \left(\frac{\hat{W} \times \vec{r}_2}{r_2} + e \hat{W} \times \hat{P} \right) = \frac{\mu}{h} \left(\frac{\hat{W} \times \vec{r}_2}{r_2} + e \hat{Q} \right)$$

Using $N = pD$ and Eq. (2-9) gives you

$$h = \sqrt{\frac{N\mu}{D}}$$

Now, let's combine this result with our knowledge that $e = \frac{S}{D}$, \hat{Q} is a unit vector of \hat{S} , and \hat{W} is a unit vector of \hat{D} . This combination results in

$$\vec{v}_2 = \frac{1}{r_2} \sqrt{\frac{\mu}{ND}} \hat{D} \times \vec{r}_2 + \sqrt{\frac{\mu}{ND}} \hat{S}$$

Notice we've introduced notation for the middle velocity vector because this is our end goal. Define the following quantities to streamline the calculations:

$$\vec{B} \equiv \hat{D} \times \vec{r}_2$$

$$L_g \equiv \sqrt{\frac{\mu}{ND}}$$

Don't confuse the intermediate quantity, L_g , with the line-of-sight matrix in the angles-only orbit determination. The substitutions let you express the middle velocity vector as

$$\vec{v}_2 = \frac{L_g}{r_2} \vec{B} + L \hat{S}$$

which completes the Gibbs method of orbit determination, except that the computational algorithm must take into account our assumption of angular separations and coplanar vectors.

The angular-separation values are of interest because the method rests on geometry. Small (geocentric) angles will cause numerical instability and may yield incorrect answers. The method is robust and works “tolerably” with angles as close as about 1° , but it *quickly* degrades with smaller angles. Of course, this depends on the computer we use, as well as the programming, but these are general guidelines. Typical observations from a sensor site may be very closely spaced—a few seconds or fractions of a degree apart. Although we can use 1° as a guide, the Herrick-Gibbs approach (to be discussed shortly) doesn’t suffer from a lower limit, so we prefer it for *very* closely spaced observations. It’s reasonable to assume (from a geometric standpoint) that vectors 50° apart will yield more accurate results than vectors 5° apart. For this reason, we should calculate the exact angles for further examination. Using the dot product formula, let’s find the two separation angles (α_{12} , α_{23} between vectors 1 and 2 and vectors 2 and 3):

$$\cos(\alpha_{12}) = \frac{\hat{r}_1 \cdot \hat{r}_2}{|\hat{r}_1||\hat{r}_2|} \quad \cos(\alpha_{23}) = \frac{\hat{r}_2 \cdot \hat{r}_3}{|\hat{r}_2||\hat{r}_3|} \quad (6-25)$$

In practice, it’s sufficient to choose position vectors at some reasonable spacing in time, rather than calculate the angular separations, but we’ll concentrate on the angular separations here. We can tailor these angles to specific needs as an internal filter for calculations. Quadrant checks are unnecessary as long as the vectors aren’t too widely dispersed because the arccosine function will return answers within the range of 0° to 180° . Three sequential vectors from different sites could have angular separations of greater than 180° .

Let’s also find the coplanar condition from the initial vectors. First, find the normal vectors from the given positions. Let \hat{Z}_{23} be the normal vector from the second and third position vectors.

$$\hat{Z}_{23} = \hat{r}_2 \times \hat{r}_3$$

The vectors are coplanar if \hat{Z}_{23} is perpendicular to \hat{r}_1 , so use a dot product:

$$\hat{Z}_{23} \cdot \hat{r}_1 = 0 \quad (6-26)$$

Your particular tolerance for “zero” determines the accuracy you wish to place on this coplanar requirement. The actual angle is

$$\alpha_{cop} = 90^\circ - \cos^{-1}\left(\frac{\hat{Z}_{23} \cdot \hat{r}_1}{|\hat{Z}_{23}||\hat{r}_1|}\right) \quad (6-27)$$

If the vectors are exactly coplanar, α_{cop} will take on a value of zero. Typical values for real-world data from a sensor site yield results which may vary by as much as 2° or 3° .

Implementing the GIBBS Algorithm

Although the number of operations in this algorithm seems large, most are multiplications, which usually aren't as slow as division and trigonometric operations. We get some efficiency by storing the temporary vectors from the three cross products in the calculation of vector N . The angular checks may not always be necessary, but before you remove the equations, be sure to document the change.

ALGORITHM 48: GIBBS ($\vec{r}_1, \vec{r}_2, \vec{r}_3$ in IJK $\Rightarrow \vec{v}_2$)

$$\hat{Z}_{12} = \vec{r}_1 \times \vec{r}_2 \quad \hat{Z}_{23} = \vec{r}_2 \times \vec{r}_3 \quad \hat{Z}_{31} = \vec{r}_3 \times \vec{r}_1$$

$$\alpha_{cop} = 90^\circ - \cos^{-1} \left(\frac{\hat{Z}_{23} \cdot \vec{r}_1}{|\hat{Z}_{23}| |\vec{r}_1|} \right)$$

$$\cos(\alpha_{12}) = \frac{\vec{r}_1 \cdot \vec{r}_2}{|\vec{r}_1| |\vec{r}_2|} \quad \cos(\alpha_{23}) = \frac{\vec{r}_2 \cdot \vec{r}_3}{|\vec{r}_2| |\vec{r}_3|}$$

$$\vec{N} = \vec{r}_1 \hat{Z}_{23} + \vec{r}_2 \hat{Z}_{31} + \vec{r}_3 \hat{Z}_{12}$$

$$\vec{D} = \hat{Z}_{12} + \hat{Z}_{23} + \hat{Z}_{31}$$

$$\vec{S} = (r_2 - r_3) \vec{r}_1 + (r_3 - r_1) \vec{r}_2 + (r_1 - r_2) \vec{r}_3$$

$$\vec{B} = \vec{D} \times \vec{r}_2$$

$$L_g = \sqrt{\frac{\mu}{ND}}$$

$$\vec{v}_2 = \frac{L_g}{r_2} \vec{B} + L_g \vec{S}$$

▼ Example 6-3. Using Three Position Vectors for GIBBS Orbit Determination.

GIVEN: $\vec{r}_1 = 6378.137 \hat{K}$ km, $\vec{r}_2 = -4464.696 \hat{J} - 5102.509 \hat{K}$ km, and

$$\vec{r}_3 = 5740.323 \hat{J} - 3189.068 \hat{K}$$
 km

FIND: \vec{v}_{IJK}

First, convert to canonical units by dividing by the radius of the Earth, 6378.1363 km:

$$\vec{r}_1 = 1 \hat{K}$$
 ER, $\vec{r}_2 = -0.7 \hat{J} - 0.8 \hat{K}$ ER, $\vec{r}_3 = 0.9 \hat{J} + 0.5 \hat{K}$ ER

Always test to be sure the input vectors are coplanar. In this case,

$$\hat{Z}_{23} = \hat{r}_2 \times \hat{r}_3 = 0.37\hat{I}$$

$$\alpha_{cop} = 90^\circ - \cos^{-1}\left(\frac{\hat{Z}_{23} \cdot \hat{r}_1}{|\hat{Z}_{23}| |\hat{r}_1|}\right) = 0^\circ$$

Next, make sure the vectors have a moderate angular separation:

$$\cos(\alpha_{12}) = \frac{\hat{r}_1 \cdot \hat{r}_2}{|\hat{r}_1| |\hat{r}_2|} \Rightarrow 138.81^\circ \quad \cos(\alpha_{23}) = \frac{\hat{r}_2 \cdot \hat{r}_3}{|\hat{r}_2| |\hat{r}_3|} \Rightarrow 160.24^\circ$$

Having met both conditions, you can proceed. Find several intermediate vectors:

$$\vec{N} = \hat{r}_1 (\hat{Z}_{23}) + \hat{r}_2 (\hat{Z}_{31}) + \hat{r}_3 (\hat{Z}_{12}) = 2.047\hat{I}$$

$$\vec{D} = \hat{Z}_{12} + \hat{Z}_{23} + \hat{Z}_{31} = 1.970\hat{I}$$

$$\hat{S} = (r_2 - r_3) \hat{r}_1 + (r_3 - r_1) \hat{r}_2 + (r_1 - r_2) \hat{r}_3 = -0.077\hat{J} - 0.022\hat{K}$$

$$\hat{B} = \vec{D} \times \hat{r}_2 = 1.576\hat{J} - 1.379\hat{K}$$

Use one more constant to get the final solution:

$$L_g = \sqrt{\frac{\mu}{ND}} = 0.497\ 926$$

so

$$\hat{v}_2 = \frac{L_g}{r_2^2} \hat{B} + L_g \hat{S} = 0.699\ 670\hat{J} - 0.656\ 744\hat{K} \text{ ER/TU} = 5.531\ 148\hat{J} - 5.191\ 806\hat{K} \text{ km/s}$$



6.6.2 Herrick-Gibbs

The immediate question arising from the Gibbs method is what to do when the vectors are too close to one another. In fact, a typical sensor site's observation of a pass by a satellite usually results in hundreds of observations which are *very* close together. For those occasions where the position vectors are very closely spaced, answers from the Gibbs method are unreliable at best. One solution is the **Herrick-Gibbs (HGIBBS)** method, which tries to find the middle velocity vector given three sequential position vectors ($\hat{r}_1, \hat{r}_2, \hat{r}_3$) and their observation times (t_1, t_2, t_3). **HGIBBS** is just a variation of the Gibbs method. The basic premise uses a Taylor-series expansion (thus the requirement for the times of each observation) to obtain an expression for the middle velocity vector. Because this method is only approximate, it's important to note that the Herrick-Gibbs method is *not* as robust as the Gibbs method, and it has a more limited application. Just as with Gibbs, we must determine the actual angle between the vectors so a user can ascertain the accuracy of the calculation. If the vectors are 3° apart, the results may be acceptable, whereas vectors 70° apart would give us erroneous results. To ensure the vectors are close enough together, we use vectors from a *single* pass of the satellite over a particular ground station. Figure 6-9

shows the basic geometry. Notice how near the position vectors are. The closer the position vectors (short-arc observations), the greater the impact of observational errors on the final result.

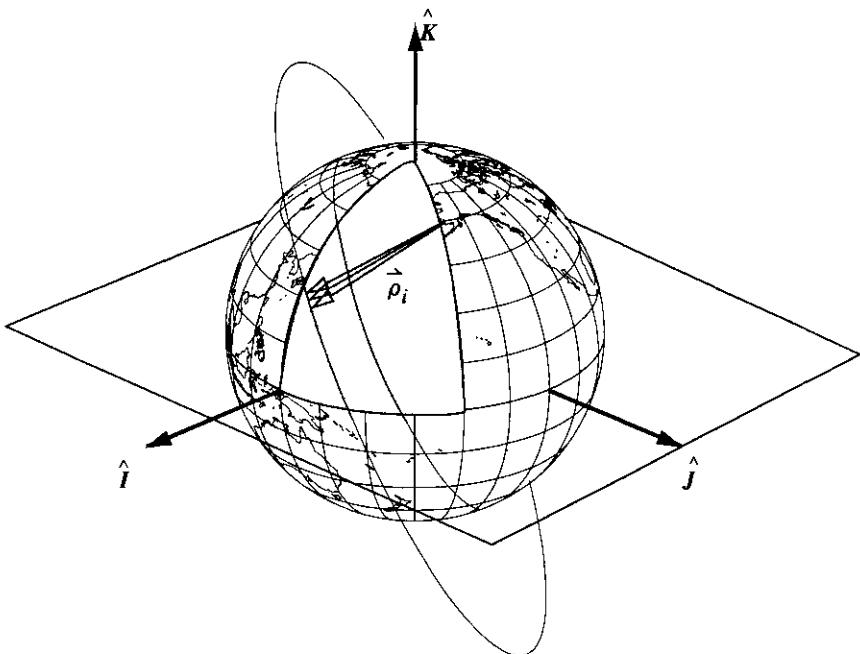


Figure 6-9. Orbit Geometry for the Herrick-Gibbs Method. Orbit determination using the Herrick-Gibbs method works well with closely spaced observations.

As in Gibbs, some preliminary calculations are necessary to ensure the given position vectors are coplanar, or coplanar enough for a particular application. We use Eq. (6-27) to determine the coplanar property. The separation between the vectors is also important, but now we want a *small* separation. Eq. (6-25) permits us to determine these angles.

To begin the procedure, expand the position vector as a Taylor series about the middle time, t_2 . In general, the form of the Taylor series is

$$\vec{r}(t) = \vec{r}_2 + \dot{\vec{r}}_2(t - t_2) + \frac{\ddot{\vec{r}}_2(t - t_2)^2}{2!} + \frac{\dddot{\vec{r}}_2(t - t_2)^3}{3!} + \frac{\ddot{\ddot{\vec{r}}}_2(t - t_2)^4}{4!} + \dots$$

Now, use this form for the two position vectors, \vec{r}_1 , \vec{r}_2 , and simplify the notation for the time differences as $\Delta t_{ij} = t_i - t_j$:

$$\begin{aligned}\overset{\circ}{r}_1 &= \overset{\circ}{r}_2 + \overset{\circ}{r}_2 \Delta t_{12} + \frac{\overset{\circ}{r}_2 \Delta t_{12}^2}{2!} + \frac{\overset{\circ}{r}_2 \Delta t_{12}^3}{3!} + \frac{\overset{\circ}{r}_2 \Delta t_{12}^4}{4!} + \dots \\ \overset{\circ}{r}_3 &= \overset{\circ}{r}_2 + \overset{\circ}{r}_2 \Delta t_{32} + \frac{\overset{\circ}{r}_2 \Delta t_{32}^2}{2!} + \frac{\overset{\circ}{r}_2 \Delta t_{32}^3}{3!} + \frac{\overset{\circ}{r}_2 \Delta t_{32}^4}{4!} + \dots\end{aligned}\quad (6-28)$$

The goal is to find the middle velocity vector. To do so, begin by ignoring all terms higher than fourth order. Next, find expressions for the second through fourth derivatives. The need to extend the analysis to the fourth order will become apparent shortly. First, eliminate the second-order derivative: multiply the first equation by $(-\Delta t_{32}^2)$ and add it to the second equation multiplied by (Δt_{12}^2) to produce

$$\begin{aligned}-\overset{\circ}{r}_1 \Delta t_{32}^2 + \overset{\circ}{r}_3 \Delta t_{12}^2 &= \overset{\circ}{r}_2 (-\Delta t_{32}^2 + \Delta t_{12}^2) + \overset{\circ}{r}_2 (-\Delta t_{32}^2 \Delta t_{12} + \Delta t_{12}^2 \Delta t_{32}) \\ &\quad + \frac{\overset{\circ}{r}_2}{6} (-\Delta t_{32}^2 \Delta t_{12}^3 + \Delta t_{12}^2 \Delta t_{32}^3) \\ &\quad + \frac{\overset{\circ}{r}_2}{24} (-\Delta t_{32}^2 \Delta t_{12}^4 + \Delta t_{12}^2 \Delta t_{32}^4)\end{aligned}$$

To save a lot of effort, expand and simplify the quantities in each of the brackets for the derivative terms. The three quantities take the following three forms:

$$\begin{aligned}\Delta t_{12}^2 \Delta t_{32} - \Delta t_{32}^2 \Delta t_{12} &= \Delta t_{12} \Delta t_{32} \Delta t_{13} \\ \Delta t_{12}^2 \Delta t_{32}^3 - \Delta t_{32}^2 \Delta t_{12}^3 &= \Delta t_{12}^2 \Delta t_{32}^2 \Delta t_{31} \\ \Delta t_{12}^2 \Delta t_{32}^4 - \Delta t_{32}^2 \Delta t_{12}^4 &= \Delta t_{12}^2 \Delta t_{32}^2 \Delta t_{31} \{ \Delta t_{32} + \Delta t_{12} \}\end{aligned}$$

This allows you to solve the velocity vector in terms of the quantities remaining in the equation. Notice the order of the Δt_{31} term instead of Δt_{13} , and the middle velocity, $\overset{\circ}{r}_2$, switched to account for its being on the other side of the equation. Thus,

$$\begin{aligned}\overset{\circ}{r}_2 (\Delta t_{12} \Delta t_{32} \Delta t_{31}) &= \overset{\circ}{r}_1 \Delta t_{32}^2 + \overset{\circ}{r}_2 (-\Delta t_{32}^2 + \Delta t_{12}^2) - \overset{\circ}{r}_3 \Delta t_{12}^2 \\ &\quad + \frac{\overset{\circ}{r}_2}{6} (\Delta t_{12}^2 \Delta t_{32}^2 \Delta t_{31}) + \frac{\overset{\circ}{r}_2}{24} (\Delta t_{12}^2 \Delta t_{32}^2 \Delta t_{31} \{ \Delta t_{32} + \Delta t_{12} \})\end{aligned}\quad (6-29)$$

The problem now is to determine expressions for the third and fourth derivatives because everything else in Eq. (6-29) is known. Using Eq. (6-28), differentiate each equation twice, which provides the starting point to remove the higher derivatives:

$$\overset{\circ}{r}_1 = \overset{\circ}{r}_2 + \overset{\circ}{r}_2 \Delta t_{12} + \frac{\overset{\circ}{r}_2 \Delta t_{12}^2}{2} + \dots$$

$$\ddot{\vec{r}}_3 = \ddot{\vec{r}}_2 + \ddot{\vec{r}}_2 \Delta t_{32} + \frac{\overset{iv}{\vec{r}_2}}{2} \Delta t_{32}^2 + \dots$$

Eliminate the third derivative term by forming $-\Delta t_{32} \overset{iv}{\vec{r}_1} + \Delta t_{12} \overset{iv}{\vec{r}_3}$

$$-\overset{iv}{\vec{r}_1} \Delta t_{32} + \overset{iv}{\vec{r}_3} \Delta t_{12} = \overset{iv}{\vec{r}_2} (-\Delta t_{32} + \Delta t_{12}) + \frac{\overset{iv}{\vec{r}_2}}{2} (-\Delta t_{32} \Delta t_{12}^2 + \Delta t_{12} \Delta t_{32}^2)$$

The two bracketed quantities on the right-hand side make the solution more difficult, but you can simplify them by expansion and reduction. For the first bracket, notice

$$-(t_3 - t_2) + (t_1 - t_2) = -(t_3 - t_1) = -\Delta t_{31} = \Delta t_{13}$$

and for the second, factor out common terms and find

$$\Delta t_{12} \Delta t_{32} (-\Delta t_{12} + \Delta t_{32})$$

However, notice the new bracket above is simply the negative of the expression in the first bracket. Now, substitute both simplifications and find

$$-\overset{iv}{\vec{r}_1} \Delta t_{32} + \overset{iv}{\vec{r}_3} \Delta t_{12} = \overset{iv}{\vec{r}_2} (-\Delta t_{31}) + \frac{\overset{iv}{\vec{r}_2}}{2} (\Delta t_{12} \Delta t_{32} \Delta t_{31})$$

which you can solve to isolate the fourth derivative:

$$\overset{iv}{\vec{r}_2} = \frac{2}{\Delta t_{12} \Delta t_{32} \Delta t_{31}} (-\overset{iv}{\vec{r}_1} \Delta t_{32} + \overset{iv}{\vec{r}_2} \Delta t_{31} + \overset{iv}{\vec{r}_3} \Delta t_{12}) \quad (6-30)$$

Similarly, isolate the third derivative by forming

$$-\Delta t_{32}^2 \overset{iv}{\vec{r}_1} + \Delta t_{12}^2 \overset{iv}{\vec{r}_3}$$

and

$$-\overset{iv}{\vec{r}_1} \Delta t_{32}^2 + \overset{iv}{\vec{r}_3} \Delta t_{12}^2 = \overset{iv}{\vec{r}_2} (-\Delta t_{32}^2 + \Delta t_{12}^2) + \overset{iv}{\vec{r}_2} (-\Delta t_{32}^2 \Delta t_{12} + \Delta t_{12}^2 \Delta t_{32}) \quad (6-31)$$

Again, you must simplify the two brackets on the right-hand side. Beginning with the first bracket, expand each of the terms:

$$\begin{aligned} -\Delta t_{32}^2 + \Delta t_{12}^2 &= -(t_3 - t_2)(t_3 - t_2) + (t_1 - t_2)(t_1 - t_2) \\ &= -(t_3 - t_2 + t_1 - t_1)(t_3 - t_2) + (t_1 - t_2 + t_3 - t_3)(t_1 - t_2) \\ &= -(t_3 - t_1)(t_3 - t_2) \\ &\quad - (t_1 - t_2)(t_3 - t_2) - (t_3 - t_1)(t_1 - t_2) + (t_3 - t_2)(t_1 - t_2) \\ &= -(t_3 - t_1)[(t_3 - t_2) + (t_1 - t_2)] = -\Delta t_{31}(\Delta t_{32} + \Delta t_{12}) \end{aligned}$$

Except for a minus sign, the second bracket is the same as from the derivation of the fourth order:

$$\Delta t_{12}\Delta t_{32}(\Delta t_{12}-\Delta t_{32}) = \Delta t_{12}\Delta t_{32}\Delta t_{13}$$

Substituting the reduced bracketed quantities into Eq. (6-31) gives you

$$-\ddot{\vec{r}}_1\Delta t_{32}^2 + \ddot{\vec{r}}_3\Delta t_{12}^2 = \ddot{\vec{r}}_2(-\Delta t_{31}[\Delta t_{32} + \Delta t_{12}]) + \ddot{\vec{r}}_2(\Delta t_{12}\Delta t_{32}\Delta t_{13})$$

and the final result for the third derivative is

$$\ddot{\vec{r}}_2 = \frac{1}{\Delta t_{12}\Delta t_{32}\Delta t_{13}}(-\ddot{\vec{r}}_1\Delta t_{32}^2 + \ddot{\vec{r}}_2(\Delta t_{31}[\Delta t_{32} + \Delta t_{12}]) + \ddot{\vec{r}}_3\Delta t_{12}^2) \quad (6-32)$$

Now, substitute these two expressions for the derivatives into Eq. (6-29), divide through by $(\Delta t_{12}\Delta t_{32}\Delta t_{31})$, and simplify to get

$$\begin{aligned} \ddot{\vec{r}}_2 &= \frac{1}{\Delta t_{12}\Delta t_{32}\Delta t_{31}}[\dot{\vec{r}}_1\Delta t_{32}^2 + \dot{\vec{r}}_2(-\Delta t_{32}^2 + \Delta t_{12}^2) - \dot{\vec{r}}_3\Delta t_{12}^2] \\ &\quad - \frac{1}{6\Delta t_{31}}[-\ddot{\vec{r}}_1\Delta t_{32}^2 + \ddot{\vec{r}}_2\Delta t_{31}(\Delta t_{32} + \Delta t_{12}) + \ddot{\vec{r}}_3\Delta t_{12}^2] \\ &\quad + \frac{(\Delta t_{32} + \Delta t_{12})}{12\Delta t_{31}}[-\ddot{\vec{r}}_1\Delta t_{32} + \ddot{\vec{r}}_2\Delta t_{31} + \ddot{\vec{r}}_3\Delta t_{12}] \\ &= \frac{\dot{\vec{r}}_1\Delta t_{32}}{\Delta t_{12}\Delta t_{31}} - \frac{\dot{\vec{r}}_2(\Delta t_{32} + \Delta t_{12})}{\Delta t_{12}\Delta t_{32}} - \frac{\dot{\vec{r}}_3\Delta t_{12}}{\Delta t_{32}\Delta t_{31}} + \ddot{\vec{r}}_1\left(\frac{\Delta t_{32}^2}{6\Delta t_{31}} - \frac{\Delta t_{32}(\Delta t_{32} + \Delta t_{12})}{12\Delta t_{31}}\right) \\ &\quad + \ddot{\vec{r}}_2\left(\frac{-\Delta t_{31}(\Delta t_{32} + \Delta t_{12})}{6\Delta t_{31}} + \frac{\Delta t_{31}(\Delta t_{32} + \Delta t_{12})}{12\Delta t_{31}}\right) \\ &\quad + \ddot{\vec{r}}_3\left(\frac{-\Delta t_{12}^2}{6\Delta t_{31}} + \frac{\Delta t_{12}(\Delta t_{32} + \Delta t_{12})}{12\Delta t_{31}}\right) \\ &= \frac{\dot{\vec{r}}_1\Delta t_{32}}{\Delta t_{12}\Delta t_{31}} - \frac{\dot{\vec{r}}_2(\Delta t_{32} + \Delta t_{12})}{\Delta t_{12}\Delta t_{32}} - \frac{\dot{\vec{r}}_3\Delta t_{12}}{\Delta t_{32}\Delta t_{31}} + \ddot{\vec{r}}_1\left(\frac{\Delta t_{32}(2\Delta t_{32} - \Delta t_{32} - \Delta t_{12})}{12\Delta t_{31}}\right) \\ &\quad + \ddot{\vec{r}}_2\left(\frac{-\Delta t_{31}(\Delta t_{32} + \Delta t_{12})}{12\Delta t_{31}}\right) + \ddot{\vec{r}}_3\left(\Delta t_{12}\frac{-2\Delta t_{12} + \Delta t_{32} + \Delta t_{12}}{12\Delta t_{31}}\right) \end{aligned}$$

Eliminate the acceleration vectors by recognizing that you've used two-body motion and

$$\ddot{\vec{r}} = -\frac{\mu}{r^3}\vec{r}$$

Simplifying yields the desired answer—the middle velocity vector, which I've shown in the algorithm. Although this process is lengthy, it's only a Taylor-series approximation. Thus in principle, we can estimate the order of the error by simply finding the next higher derivative and evaluating it. To do so, include fifth-order derivatives of Eq. (6-28).

Unfortunately, when using only three points, we can't eliminate higher-order derivatives, as in the method just presented. Essentially, the result has four unknowns in three equations. The obvious solution would be to include four observations in order to make the solution consistent. The resulting solution is long, tedious, and too complex to compare it to the original method. Time and positions from the fourth observation cause this complexity. Adding a fourth observation suggests a statistical form of processing, which I'll talk more about in Chap. 9.

Implementing the Herrick-Gibbs Method

As with many routines, using canonical variables eliminates a few operations involving the gravitational parameter. Unfortunately, we can't streamline the Herrick-Gibbs method much with temporary variables. In Algorithm 49, below, I've shown the Julian date (*JD*) as inputs for the time argument. This approach consolidates the time information in a single variable, but it may affect accuracy for computer programs due to round off. I recommend the modified Julian date to resolve this difficulty.

ALGORITHM 49: HGIBBS ($\vec{r}_1, \vec{r}_2, \vec{r}_3, JD_1, JD_2, JD_3 \Rightarrow \vec{v}_2$)

$$\begin{aligned} \Delta t_{31} &= JD_3 - JD_1 \\ \Delta t_{32} &= JD_3 - JD_2 \\ \Delta t_{21} &= JD_2 - JD_1 \\ \vec{Z}_{23} &= \vec{r}_2 \times \vec{r}_3 \\ \alpha_{cop} &= 90^\circ - \cos^{-1} \left(\frac{\vec{Z}_{23} \cdot \vec{r}_1}{|\vec{Z}_{23}| |\vec{r}_1|} \right) \\ \cos(\alpha_{12}) &= \frac{\vec{r}_1 \cdot \vec{r}_2}{|\vec{r}_1| |\vec{r}_2|} \quad \cos(\alpha_{23}) = \frac{\vec{r}_2 \cdot \vec{r}_3}{|\vec{r}_2| |\vec{r}_3|} \\ \vec{v}_2 &= -\Delta t_{32} \left(\frac{1}{\Delta t_{21} \Delta t_{31}} + \frac{\mu}{12 r_1^3} \right) \vec{r}_1 + (\Delta t_{32} - \Delta t_{21}) \left(\frac{1}{\Delta t_{21} \Delta t_{32}} + \frac{\mu}{12 r_2^3} \right) \vec{r}_2 \\ &\quad + \Delta t_{21} \left(\frac{1}{\Delta t_{32} \Delta t_{31}} + \frac{\mu}{12 r_3^3} \right) \vec{r}_3 \end{aligned}$$

It's possible to compare the accuracies of the Herrick-Gibbs and Gibbs approaches by running identical test cases with each. Because one performs well with widely spaced

data, whereas the other works better with closely spaced data, we can find an approximate cross-over point between 1° and 5° . Below 1° , **HGIBBS** is superior; above 5° , **GIBBS** is superior.

▼ **Example 6-4. Using Three Position Vectors and Time for Herrick-Gibbs Orbit Determination.**

GIVEN: $\hat{r}_1 = 3419.855\ 64\hat{i} + 6019.826\ 02\hat{j} + 2784.600\ 22\hat{k}$ km $t_1 = 0.0$

$$\hat{r}_2 = 2935.911\ 95\hat{i} + 6326.183\ 24\hat{j} + 2660.595\ 84\hat{k}$$
 km $t_2 = 1 \text{ min } 16.48 \text{ s}$

$$\hat{r}_3 = 2434.952\ 02\hat{i} + 6597.386\ 74\hat{j} + 2521.523\ 11\hat{k}$$
 km $t_3 = 2 \text{ min } 33.04 \text{ s}$

FIND: \hat{v}_2

Convert the position vectors to canonical units:

$$\hat{r}_1 = 0.536\ 184\hat{i} + 0.943\ 822\hat{j} + 0.436\ 585\hat{k}$$
 ER

$$\hat{r}_2 = 0.460\ 308\ 7\hat{i} + 0.991\ 854\ 4\hat{j} + 0.417\ 143\ 1\hat{k}$$
 ER

$$\hat{r}_3 = 0.381\ 765\ 4\hat{i} + 1.034\ 375\hat{j} + 0.395\ 338\hat{k}$$
 ER

As with the Gibbs procedure, it's useful to determine the coplanar condition of the vectors as well as their angular separation. Remember, this is a Taylor-series approximation, and the angles should be relatively close together.

$$\hat{Z}_{23} = \hat{r}_2 \times \hat{r}_3 = -0.039\ 364\hat{i} - 0.022\ 727\hat{j} + 0.097\ 476\hat{k}$$

$$\alpha_{cop} = 90^\circ - \cos^{-1}\left(\frac{\hat{Z}_{23} \cdot \hat{r}_1}{|\hat{Z}_{23}| |\hat{r}_1|}\right) = 0.0^\circ$$

$$\cos(\alpha_{12}) = \frac{\hat{r}_1 \cdot \hat{r}_2}{|\hat{r}_1| |\hat{r}_2|} \Rightarrow 4.5^\circ \quad \cos(\alpha_{23}) = \frac{\hat{r}_2 \cdot \hat{r}_3}{|\hat{r}_2| |\hat{r}_3|} \Rightarrow 4.5^\circ$$

Now, find the middle velocity vector directly:

$$\hat{v}_2 = -0.814\ 845\hat{i} + 0.477\ 857\hat{j} - 0.217\ 648\hat{k}$$
 ER/TU

▲ $= -6.441\ 645\hat{i} + 3.777\ 634\ 3\hat{j} - 1.720\ 587\hat{k}$ km/s

6.7 Two Position Vectors and Time—Lambert's Problem

The problem of two position vectors and the time of flight between them is usually known as *Lambert's problem* because Lambert first formed the solution. In this context, it's an initial orbit-determination technique. Although similar to the Kepler problem discussed in Sec. 4.3, it's considered a technique for determining the initial orbit because the orbit is not yet fully known. For Lambert's problem, two position vectors and the time of flight between each position are known, but the orbit *between* the endpoints is unknown. We can

also classify the problem as a transfer problem (rendezvous and intercept). The solutions I present here are the correct way to analyze the complex maneuvers introduced in Chap. 5. In this section, I'll refer to the technique as initial orbit determination, but remember it can also be used for an orbital-transfer algorithm. In Sec. 6.8, I'll further explore the orbital transfer aspect of Lambert's problem.

Solutions to Lambert's problem abound in the literature, as they did even in Lambert's time shortly after his original formulation in 1761. Examples are Lambert's original geometric formulation, which provides equations to determine the minimum-energy orbit, and the original Gaussian formulation, which gives geometrical insight into the problem. I won't detail Lambert's original method because it requires eight separate cases to provide a complete answer. The universal-variable solution and algorithms by Battin (1987) are elegant and robust algorithms for general purposes. Gooding's algorithm (1990), and Thorne's direct solution (Thorne and Bain, 1995) are more recent developments which appear in the literature. For additional techniques not discussed here, see Escobal ([1965] 1985, 197–227). He shows the Lambert-Euler technique, which iterates on the semimajor axis to arrive at a solution. He also presents solutions that iterate on true anomaly, semiparameter, f and g series, and eccentricity.

This section describes several of the important methods of solution and concludes with an algorithmic solution. Some preliminary concepts are common to each technique. All the techniques require us to know the angle through which the transfer takes place, as Fig. 6-10 illustrates.

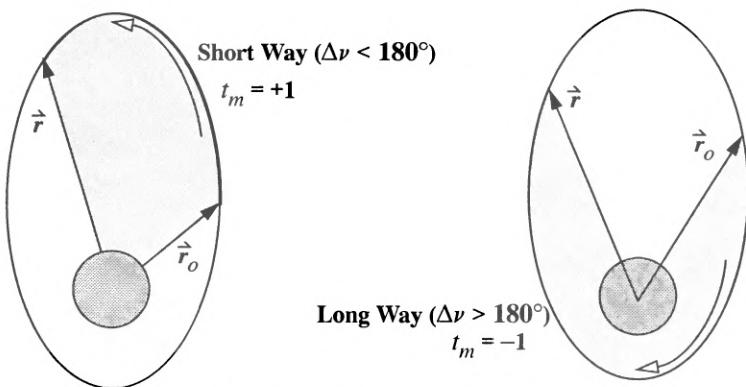


Figure 6-10. Transfer Methods, t_m , for the Lambert problem. Traveling between the two specified points can take the long way or the short way. For the long way, the change in true anomaly exceeds 180° .

We can see right away from Fig. 6-10 that two paths will satisfy the requirement to find an orbit between the two given position vectors. In addition, the two vectors immediately define the orbital plane. If the **transfer method**, t_m , is specified, the problem has one unique solution. For most techniques, if $\Delta\nu$ is 180° , we can't determine a solution because

many answers yield the same final position vector. In addition, the plane of the transfer orbit is not uniquely determined. Multiples of $\Delta\nu = k \cdot 360^\circ$ ($k = 1, 2, 3, \dots$) are possible only if the time is greater than 0 (we must have at least two *independent* position vectors). This includes the case in which the two position vectors are on the same orbit displaced through time, or on different orbits through which we're trying to transfer. The orbits do not exist and are degenerate conics (straight lines) if the time is zero and the vectors aren't the same. Notice we've discussed the transfer as though a satellite is traveling from one orbit to another. But the procedure works equally well on, and was originally formulated for, a single orbit in which two position vectors at different times were known. In determining the transfer angle or the angle between two position observations in the original orbit-determination problem, we can use both the dot and cross products:

$$\begin{aligned}\cos(\Delta\nu) &= \frac{\vec{r}_o \cdot \vec{r}}{r_o r} \\ \sin(\Delta\nu) &= \frac{|\vec{r}_o \times \vec{r}|}{r_o r} \quad \text{or} \quad \sin(\Delta\nu) = t_m \sqrt{1 - \cos^2(\Delta\nu)}\end{aligned}\tag{6-33}$$

In these relations, we often use just the cosine expression, but I've provided the sine expression to resolve quadrants in computer software. Recognize that t_m is (+1) for short-way transfers and (-1) for long-way transfers.

6.7.1 Lambert—Minimum Energy

Lambert's original formulation is important because it geometrically interprets the problem and it provides a way to determine the minimum-energy transfer between the two known endpoints. Note that we need separate cases to program a general-purpose routine using this formulation (Kaplan, 1976, 324). Because his original formulation requires several checks, I'll introduce only the basic concepts and concentrate on the minimum-energy solution. Lambert's basic premise is stated by Battin (1987, 276) as a theorem:

The orbital transfer time depends only upon the semimajor axis, the sum of the distances of the initial and final points of the arc from the center of force, and the length of the chord joining these points.

My analysis follows Prussing and Conway (1993, 62–79) and Kaplan (1976, 308–329). Because Lambert's technique relies on the problem's geometry, consider Fig. 6-11. The central idea of Lambert's problem begins by defining the chord length between the two position vectors, c . Find the chord length using the cosine law for the triangle formed by the two vectors, and the chord, or shortest point between the two endpoints.

$$c = \sqrt{r_o^2 + r^2 - 2r_o r \cos(\Delta\nu)}$$

Define the *semiperimeter*, s , as half the sum of the sides of the triangle created by the position vectors and the chord. Don't confuse this quantity with the semiparameter, p !

$$s = \frac{r_o + r + c}{2} \quad (6-34)$$

From Eq. (1-2), we know that the sum of the distances from the foci to any point on the ellipse equals twice the semimajor axis. For instance, using Fig. 6-11 with F , F' , and r ,

$$2a = r + (2a - r)$$

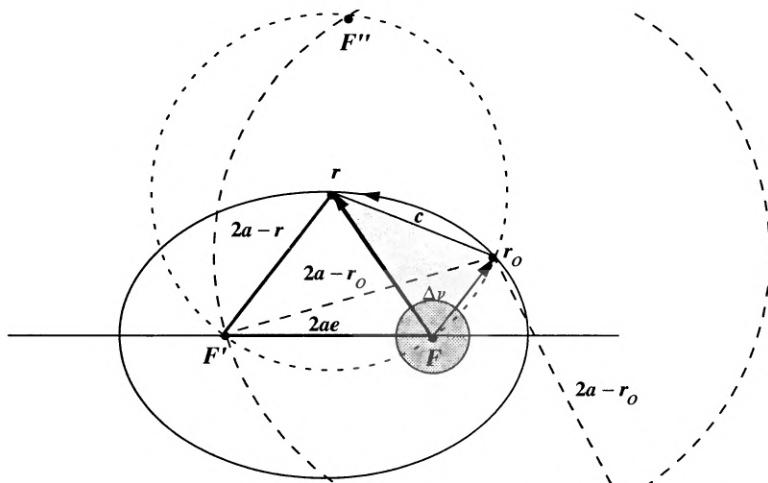


Figure 6-11. Geometry for the Lambert Problem (I). This figure shows how we locate the secondary focus—the intersection of the dashed circles. The chord length, c , is the shortest distance between the two position vectors. The sum of the distances from a focus to any point, r or r_o , is equal to twice the semimajor axis.

We can use this relation to locate the other focus and determine the orbit's semimajor axis. Figure 6-11 shows we can draw circles about each position with radii $2a_i - r$ and $2a_i - r_o$. The intersections of corresponding circles (of equal a_i) mark the possible locations of the other focus. In Fig. 6-11, I've shown the foci for one semimajor axis, along with a corresponding orbit. In Fig. 6-12 I've shown many circles representing different values of a .

Kaplan (1976, 308–319) and Prussing and Conway (1992, 64–67) discuss several useful facts from Fig. 6-12.

1. The intersection of the circles depicts possible locations of the other foci.
2. The locus of secondary foci lies on the two branches of a hyperbola with eccentricity $1/e_F$, where e_F is the minimum-eccentricity orbit passing through the two position vectors, $e_F = (r - r_o)/c$.
3. At each point on the locus of secondary foci positions (F_i), the difference in distances to r_o and to r is $r - r_o/c$.

4. The secondary focus whose distance to the primary is shorter has an orbit with a smaller eccentricity. In Fig. 6-11, the orbit with the focus, F' , has a smaller eccentricity than the alternate point of intersection, F'' .
5. There is only one secondary focus in the minimum-energy solution as shown in Fig. 6-12 where the darker circles just touch.

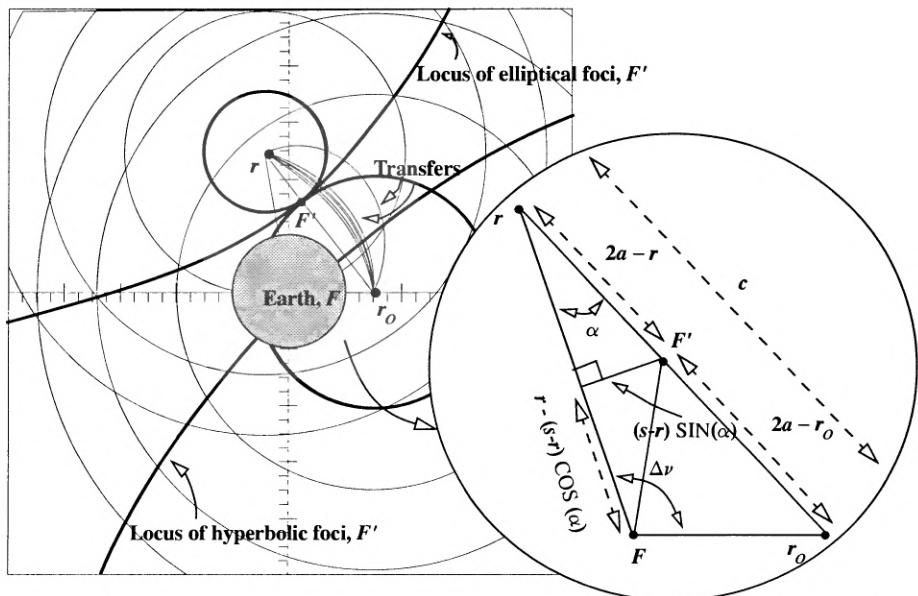


Figure 6-12. Geometry for the Lambert Problem (II). Solving Lambert's problem relies on many geometrical quantities. Be sure to allow for the Earth when viewing representations like this. I've shown transfers between a satellite at 9567.2 km (1.5 ER) and 15,307.5 km (2.4 ER) from the Earth's center. We can use the inset figure to find the transfer semimajor axis. The concentric circles are drawn for elliptical values of a . When the circles (of the same a) touch, half the sum of their radii equals a of the transfer, and the intersection is the location of the second focus, F' .

Let's explore the final point about the minimum energy. It represents the minimum-energy orbit because the chord length equals the sum of the two radii. Semimajor axes that are smaller don't have enough energy ($\xi = -\mu/2a$) to travel between both foci. Thus,

$$2a - r + 2a - r_o = c$$

We can solve for the semimajor axis with minimum energy:

$$a_{min} = \frac{s}{2} = \frac{r_o + r + c}{4} \quad (6-35)$$

Solutions aren't possible with a semimajor axis less than this minimum ($a < a_{min}$). Battin (1987, 258) remarks that the ellipse for minimum-eccentricity transfer is often called the **fundamental ellipse** because the major axis is parallel to the chord.

To finish describing the minimum-energy orbit, we require the semiparameter and the eccentricity. To determine the semiparameter, begin with the distance between the two foci, $2ae$. Kaplan (1976, 312–313) suggests a solution method. If you let the central angle between the final position and the chord be α , the Pythagorean theorem and the small triangle in Fig. 6-12 permit

$$\begin{aligned} 4a_{min}^2 e_{min}^2 &= \{ (s - r) \sin(\alpha) \}^2 + \{ r - (s - r) \cos(\alpha) \}^2 \\ &= (s - r)^2 \{ 1 - \cos^2(\alpha) \} + r^2 + (s - r)^2 \cos^2(\alpha) - 2r(s - r) \cos(\alpha) \quad (6-36) \\ &= s^2 - 2r(s - r) (1 + \cos(\alpha)) \end{aligned}$$

It's useful to determine an expression for the central angle with the cosine law:

$$\begin{aligned} r_o^2 &= r^2 + c^2 - 2rc \cos(\alpha) \\ \cos(\alpha) &= \frac{-r_o^2 + r^2 + c^2}{2rc} \end{aligned}$$

Now consider the expression,

$$2s(s - r_o) - rc$$

and expand it using Eq. (6-34). After reduction,

$$2s(s - r_o) - rc = \frac{-r_o^2 + r^2 + c^2}{2}$$

Therefore, you can write the cosine as

$$\cos(\alpha) = \frac{2s(s - r_o)}{rc} - 1$$

Substituting into the original distance formula [Eq. (6-36)] results in

$$4a_{min}^2 e_{min}^2 = s^2 - \frac{4s}{c} (s - r_o) (s - r)$$

We could solve this for e_{min} , but let's use a different method. From Eq. (1-11), develop $4a^2 e^2 = 4a(a - p)$. Remembering Eq. (6-35), this becomes

$$4a_{min}^2 e_{min}^2 = s^2 - 2sp_{min}$$

Now you can solve for the semiparameter by equating the two previous equations for the distance between the foci (Kaplan, 1976, 313, shows an alternate form using $\Delta\nu$).

$$p_{min} = \frac{2}{c} (s - r_o) (s - r) = \frac{r_o r}{c} (1 - \cos(\Delta\nu))$$

From Eq. (2-43), it's easy to determine the eccentricity of this minimum-energy transfer. Kaplan (1976, 313) shows that the minimum value for the eccentricity is

$$e_{min} = \sqrt{1 - \frac{2p_{min}}{s}}$$

Next, we want to determine the amount of time required for the transfer. Let's introduce several new constants for elliptical and hyperbolic orbits. Prussing and Conway (1993, 68–70) show the development of these relations and extensively discuss how to select the correct quadrant for α and β (1993, 70–75). For general purposes, set β_e to be negative if $\Delta\nu$ [Eq. (6-33)] is greater than 180° . The quadrant of α_e requires the time of flight for a minimum-energy transfer, Eq. (6-38). Let $\alpha_e = 2\pi - \alpha_e$ if the desired time of flight is greater than t_{min} :

$$\begin{aligned} \sin\left(\frac{\alpha_e}{2}\right) &= \sqrt{\frac{r_o + r + c}{4a}} = \sqrt{\frac{s}{2a}} & \sinh\left(\frac{\alpha_h}{2}\right) &= \sqrt{\frac{s}{-2a}} = \sqrt{\frac{r_o + r + c}{-4a}} \\ \sin\left(\frac{\beta_e}{2}\right) &= \sqrt{\frac{r_o + r - c}{4a}} = \sqrt{\frac{s - c}{2a}} & \sinh\left(\frac{\beta_h}{2}\right) &= \sqrt{\frac{s - c}{-2a}} = \sqrt{\frac{r_o + r - c}{-4a}} \end{aligned}$$

Kaplan (1976, 319–322) shows that the general time of flight, sometimes called **Lambert's equation**, is

$$\begin{aligned} t &= \sqrt{\frac{a^3}{\mu}} [\alpha_e - \sin(\alpha_e) - (\beta_e - \sin(\beta_e))] \\ &= \sqrt{\frac{-a^3}{\mu}} [\sinh(\alpha_h) - \alpha_h - (\sinh(\beta_h) - \beta_h)] \end{aligned} \tag{6-37}$$

Find the time of flight for the orbit of a minimum-energy transfer by substituting the appropriate values for a minimum-energy transfer. For minimum values, the hyperbolic relations aren't useful. Here,

$$t_{min} = \sqrt{\frac{a_{min}^3}{\mu}} \left\{ \pi - \beta_e + \sin(\beta_e) \right\} \tag{6-38}$$

The geometry of the minimum-energy transfer can be very useful. The “standard” presentation uses only magnitudes and tries to find the time for a minimum-energy transfer. However, let's determine the velocity vector for the transfer orbit so we can compare with other Lambert techniques. Although the new parameters in Eq. (6-37)

resemble the eccentric and hyperbolic anomalies, they aren't the same. They're also not the same as the auxiliary angle introduced in the derivation of the minimum semiparameter value. However, they are related to the eccentric and hyperbolic anomalies.

$$\Delta E = \alpha_e - \beta_e$$

$$\Delta H = \alpha_h - \beta_h$$

Knowing these values allows us to use Eq. (4-64) and Eq. (4-66) to find the f and g functions, and then the position and velocity vectors. Solving Eq. (4-57) for the initial velocity vector and using Eq. (4-63) produces

$$\dot{\vec{v}}_o = \frac{1}{g} \{ \dot{\vec{r}} - f \vec{r}_o \} = \frac{\sqrt{\mu p}}{r_o r \sin(\Delta\nu)} \{ \dot{\vec{r}} - \left[1 - \frac{r}{p} \{ 1 - \cos(\Delta\nu) \} \right] \vec{r}_o \}$$

Algorithm 50 summarizes the results of the minimum-energy transfers.

ALGORITHM 50: Lambert's Problem—Minimum Energy

$$(\vec{r}_o, \dot{\vec{r}} \Rightarrow a_{min}, e_{min}, t_{min}, \dot{\vec{v}}_o)$$

$$\cos(\Delta\nu) = \frac{\vec{r}_o \cdot \dot{\vec{r}}}{r_o r}$$

$$c = \sqrt{r_o^2 + r^2 - 2r_o r \cos(\Delta\nu)}$$

$$s = \frac{r_o + r + c}{2}$$

$$a_{min} = \frac{s}{2} \quad p_{min} = \frac{r_o r}{c} (1 - \cos(\Delta\nu))$$

$$e_{min} = \sqrt{1 - \frac{2p_{min}}{s}}$$

$$\alpha_e = \pi \quad \sin\left(\frac{\beta_e}{2}\right) = \sqrt{\frac{s-c}{s}} \text{ minimum values}$$

$$t_{min} = \sqrt{\frac{a_{min}^3}{\mu}} \left\{ \pi - \beta_e + \sin(\beta_e) \right\}$$

$$\dot{\vec{v}}_o = \frac{\sqrt{\mu p_{min}}}{r_o r \sin(\Delta\nu)} \left\{ \dot{\vec{r}} - \left[1 - \frac{r}{p_{min}} \{ 1 - \cos(\Delta\nu) \} \right] \vec{r}_o \right\}$$

Thorne's Solution of the Lambert Problem

Thorne and Bain (1995) developed an important and useful series solution to the Lambert problem. It's equally valid for hyperbolic and elliptical transfer times that are less than the time for minimum-energy transfer. As shown earlier, the transfer time is equal to a complicated function of the unknown semimajor axis of the observed orbit. We can solve the series form of the time-of-flight equation for the semimajor axis using series reversion and inversion. The complete mathematical technique is too lengthy for this book, but the result is the only direct, practical series solution known for the Lambert problem. We can apply Thorne's solution without changes to cases of hyperbolic, elliptical, and even colinear position vectors. It also has the advantage of not requiring any initial guess, unlike the other iterative methods. Finally, Thorne's solution for the semimajor axis contains the transfer time explicitly, so we can take derivatives of the series for optimizing orbital problems—minimum time, minimum fuel. I'll introduce the relevant equations. Refer to Thorne (1989) for more information.

To find the unknown semimajor axis, a , first define the quantity $T = (\Delta t/t_p - 1)$ as a nondimensional time parameter, where Δt is the desired flight time and t_p is the known parabolic flight time between the two given position vectors.

$$t_p = \frac{2}{3} \sqrt{\frac{s^3}{\mu}} \left\{ 1 - \left(\frac{s-c}{s} \right)^{\frac{3}{2}} \right\} \quad (6-39)$$

After some algebraic manipulation, you may then express Lambert's Time Function as (see also Battin, 1987, 305):

$$T = \sum_{i=1}^{\infty} \frac{\left(1 - \left(\frac{s-c}{s} \right)^{i+\frac{3}{2}} \right)}{\left(1 - \left(\frac{s-c}{s} \right)^{\frac{3}{2}} \right)} \frac{\left(\frac{1}{2} \right)_i \left(\frac{3}{2} \right)_i \left(\frac{s}{2a} \right)^i}{\left(\frac{5}{2} \right)_i i!} = \sum_{i=1}^{\infty} A_i \left(\frac{s}{2a} \right)^i \quad (6-40)$$

The terms $(x)_i$ are Pochhammer symbols, which are defined by $(x)_i = x(x+1)\dots(x+i-1)$. The geometric parameters s and c are the semiperimeter and chord, as defined previously in this book. This form of the transfer time equation has been shown to be equivalent to other power series expressions given in Battin (1987, 213–215). Once you've calculated the A_i terms to the desired order from Eq. (6-40), you may use them to find the B_i coefficients of the following series:

$$a = \left(\frac{s}{2} \right) \sum_{i=1}^{\infty} B_i T^{(i-2)} \quad (6-41)$$

Use this series solution only when the time parameter lies in the interval $(-1 < T < 1)$. If T is exactly zero, the path is parabolic, and the orbit geometry is already known. If $T <$

0.0, the transfer is hyperbolic, and if $T > 0.0$, it's elliptical. You can treat the B_i coefficients as a vector, where $B = QA$, and find the elements of Q from the following recursive expressions:

$$\begin{aligned} Q_{(1,1)} &= A_1^{-1} \\ Q_{(i,j)} &= \sum_{k=1}^{i-1} Q_{(i-k,j-1)} \quad (i = 2, 3, 4, \dots), 1 < j \leq i \\ Q_{(i,1)} &= \sum_{k=1}^{i-1} \left(\frac{-1}{A_1} \right) Q_{(i,k+1)} A_{(k+1)} \end{aligned} \quad (6-42)$$

Be very careful to keep track of the subscripts in the above equations. Their purpose is to populate the lower triangular matrix, Q , by working downward from the upper left element. The B_i coefficients are

$$B_i = \sum_{k=1}^{i-1} Q_{(i-1,k)} A_{(k+1)} \quad (6-43)$$

Finally, the semimajor axis, a , is given by Eq. (6-41), and the problem is solved. Generally, the closer T is to zero, the more accurate a will be.

Because of the leading inverse term in the series solution, there is a singularity when $T = 0$ which corresponds to the parabolic case. Because we define the parabola with an infinite semimajor axis, the behavior of the series solution is qualitatively correct. To examine the accuracy of the solution near $T = 0$, suppose that an object passes through two points located 180° apart, each at one distance unit from the attracting center (1 ER). The parabolic transfer time for this case is exactly $4/3$ time units. Suppose further that the observed time of flight is 1.333 333 4 time units, and we wish to know the value of the semimajor axis. This particular flight time demonstrates the near-parabolic case. Using these values, T is equal to 0.000 000 1. Table 6-2 shows the process of convergence.

To check the accuracy of the converged value, we can substitute it into the Lambert-Euler time-of-flight equation. The resulting time of flight is 1.333 333 4, as desired from the original problem statement.

6.7.2 Lambert—Gaussian Solution

The original method Gauss proposed to solve Lambert's problem relied on geometric considerations and was limited to elliptical transfers. Of course, this isn't surprising because he formed this procedure while trying to rediscover Ceres a year after its last observation. The method was originally proposed in 1809. Bate, Mueller, and White (1971, 258–264) indicate that the complete derivation is long. To bridge the gap between Lambert's for-

TABLE 6-2. Thorne's Solution. This table shows the process of convergence, where the first column is the index, the second column is the series term, and the third column is the partial sum of the series which represents the semimajor axis.

Index	Series Term	Partial Sum, a
1	$5.999\,999\,999\,983\,59 \times 10^6$	5,999,999.999 983 6
2	$5.357\,142\,857\,142\,86 \times 10^{-1}$	6,000,000.535 697 9
3	$1.003\,873\,771\,733\,66 \times 10^{-8}$	6,000,000.535 697 9
4	$-5.464\,171\,599\,405\,57 \times 10^{-17}$	6,000,000.535 697 9
5	$8.449\,953\,356\,474\,02 \times 10^{-25}$	6,000,000.535 697 9
6	$-1.779\,550\,598\,415\,56 \times 10^{-32}$	6,000,000.535 697 9
7	$4.392\,517\,075\,146\,11 \times 10^{-40}$	6,000,000.535 697 9
8	$-1.197\,083\,329\,047\,31 \times 10^{-47}$	6,000,000.535 697 9
9	$3.494\,455\,569\,673\,83 \times 10^{-55}$	6,000,000.535 697 9
10	$-1.073\,554\,071\,522\,98 \times 10^{-62}$	6,000,000.535 697 9

mula and Gauss's historical work, consult Escobal ([1965] 1985, 188–197) and Moulton ([1914] 1970, 237–249) for more complete derivations. Bate, Mueller, and White also show the extensions for hyperbolic orbits. Don't confuse this technique with the Gaussian technique I've presented for dealing with angles-only data, even though the two are related. Indeed, Herrick (1971, 247) remarks that

The process [Gaussian angles-only] is iterated until [the vectors] are accurate and consistent; the ratios of sector to triangle within each iteration are themselves calculated by processes that involve subiterations. In the end the ratio of sector to triangle based upon the now accurate first and third positions is used in the determination of a selected set of elements and derived constants.

Gauss's original method relied on the area the satellite sweeps out during the transfer. Figure 6-13 shows the geometry.

From Kepler's laws, we know that area is swept out at a constant rate,

$$dt = \frac{2}{h} dA$$

which makes the total area

$$A = \frac{1}{2} \sqrt{\mu p} \Delta t$$

where Δt represents the time of flight between the two position vectors. We find the area of the triangle formed by the two position vectors and the connecting chord as

$$A_{\Delta} = \frac{1}{2} r_o r \sin(\Delta\nu)$$

The ratio, y , of the area the satellite sweeps out to the triangular area bounded by the two position vectors is an important relation:

$$y = \frac{\sqrt{\mu p} \Delta t}{r_o r \sin(\Delta\nu)} \quad (6-44)$$

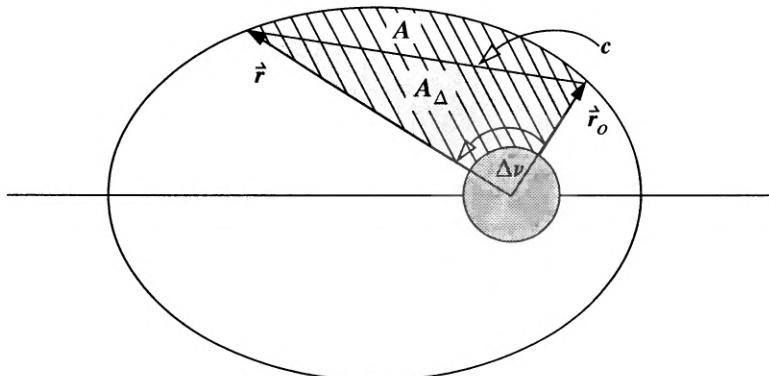


Figure 6-13. Area Swept Out by a Satellite During Transfer. Gauss's method relies on the area swept out by the satellite during its orbit. Two regions are useful—the triangular shaded region, A_{Δ} , and the total area, A , swept out by the satellite.

Gauss's method is based on obtaining two independent equations relating y and the change in eccentric anomaly (ΔE). Typically, we use a trial value of $y = 1$ in the first equation and then solve for ΔE . Although this approach works well for transfers in which the position vectors are relatively close together, it converges slowly. Whenever the vectors are far apart, it fails. Gauss was fortunate that he tried to determine the location of Ceres a year later. It was in about the same place, making the angle between the two position vectors relatively small.

Using Eq. (4-74) gives us an alternate expression for the semiparameter:

$$p = \frac{2r_o r \sin^2\left(\frac{\Delta\nu}{2}\right)}{r_o + r - 2\sqrt{r_o r} \cos\left(\frac{\Delta\nu}{2}\right) \cos\left(\frac{\Delta E}{2}\right)}$$

Now substitute this expression into the sector relation [Eq. (6-44)] and square the result:

$$y^2 = \frac{\mu \Delta t^2 \operatorname{SEC}^2\left(\frac{\Delta \nu}{2}\right)}{2r_o r \left(r_o + r - 2\sqrt{r_o r} \cos\left(\frac{\Delta \nu}{2}\right) \cos\left(\frac{\Delta E}{2}\right)\right)}$$

This is usually known as Gauss's first equation. Using several temporary variables simplifies the notation, but their formation isn't obvious:

$$t \equiv \frac{r_o + r}{4\sqrt{r_o r} \cos\left(\frac{\Delta \nu}{2}\right)} - \frac{1}{2} \quad m \equiv \frac{\mu \Delta t^2}{\left\{2\sqrt{r_o r} \cos\left(\frac{\Delta \nu}{2}\right)\right\}^3} \quad x_1 \equiv \sin^2\left(\frac{\Delta E}{4}\right)$$

The final result is the common form of the *first equation of Gauss* (note that some sources use a different formula for m , requiring it to be squared):

$$y^2 = \frac{m}{t + x_1} \quad (6-45)$$

The next goal is to arrive at an independent equation relating the change in eccentric anomaly and y . With a temporary variable,

$$x_2 = \frac{\Delta E - \sin(\Delta E)}{\sin^3\left(\frac{\Delta E}{2}\right)}$$

The presence of trigonometric terms can present some computational difficulty, so we choose an approach by Moulton ([1914] 1970, 242). This method converges for vectors that are less than 90° apart:

$$x_2 = \frac{4}{3} \left(1 + \frac{6x_1}{5} + \frac{6(8)x_1^2}{5(7)} + \frac{6(8)(10)x_1^3}{5(7)(9)} + \dots \right)$$

Escobal ([1965] 1985, 192–194) shows how to develop the *second equation of Gauss*:

$$y^3 - y^2 = mx_2 = y^2(y - 1) \quad (6-46)$$

At this point, traditional solutions take one of two paths. The first approach simply solves the equations and proceeds to solution. If we substitute Eq. (6-45) for y^2 into Eq. (6-46) and solve for the sector ratio,

$$y = (t + x_1)x_2 + 1$$

The second path is the more traditional approach, which uses continued fractions to avoid trigonometric substitutions. Alternatively, we could evaluate x_2 using a continued fraction, ξ , but this requires us to solve a third-order equation. Thus,

$$y^3 - y^2 - hy - h = 0$$

$$h = \frac{m}{\frac{5}{6} + 1 + \xi}$$

We use Eq. (4-73) and Eq. (4-74) and appropriate end conditions to yield an expression for the semimajor axis if we need it.

$$a = \frac{r_o + r - 2\sqrt{r_o r} \cos\left(\frac{\Delta\nu}{2}\right) \cos\left(\frac{\Delta E}{2}\right) \sin(\Delta E)}{2 \sin^2\left(\frac{\Delta E}{2}\right)} \quad p = \frac{y^2 r_o^2 r^2 \sin^2(\Delta\nu)}{\mu \Delta t^2}$$

Escobal ([1965] 1985, 195) shows the semiparameter result using the sector area y . Once we find y , we find ΔE by solving Gauss's first equation. It's important that ΔE be in the same half-plane as $\Delta\nu$ if we intend to use it separately. For most cases, we can treat $\cos(\Delta E/2)$ as a single variable and avoid quadrant checks.

Algorithm 51 uses the first approach because computing trigonometric terms is realistic today. The extension for parabolic and hyperbolic orbits is from Bate, Mueller, and White (1971, 263–264); it affects only our determination of the eccentric anomaly. Computation is more efficient if we store the denominator of m to use in calculating t and p . A transfer method term is input allowing us to choose either the short- or long-way trajectories— $t_m = +$ (Short Way) or $-$ (Long Way).

ALGORITHM 51: Lambert—Gaussian Solution ($\vec{r}_o, \vec{r}, \Delta t, t_m \Rightarrow \vec{v}_o, \vec{v}$)

$$\cos(\Delta\nu) = \frac{\vec{r}_o \cdot \vec{r}}{r_o r}$$

$$\sin(\Delta\nu) = \frac{|\vec{r}_o \times \vec{r}|}{r_o r} \quad \text{or} \quad \sin(\Delta\nu) = t_m \sqrt{1 - \cos^2(\Delta\nu)}$$

$$t \equiv \frac{r_o + r}{4\sqrt{r_o r} \cos\left(\frac{\Delta\nu}{2}\right)} - \frac{1}{2} \quad m \equiv \frac{\mu \Delta t^2}{\left\{2\sqrt{r_o r} \cos\left(\frac{\Delta\nu}{2}\right)\right\}^3}$$

Guess $y = 1$

LOOP

$$x_1 = \frac{m}{y^2} - t$$

$$x_2 = \frac{4}{3} \left(1 + \frac{6x_1}{5} + \frac{6(8)x_1^2}{5(7)} + \frac{6(8)(10)x_1^3}{5(7)(9)} + \dots \right)$$

$$y = 1 + x_2(t + x_1)$$

UNTIL y stops changing

IF eccentric orbits,

$$\cos\left(\frac{\Delta E}{2}\right) = 1 - 2x_1$$

$$p = \frac{r_o r (1 - \cos(\Delta\nu))}{r_o + r - 2\sqrt{r_o r} \cos\left(\frac{\Delta\nu}{2}\right) \cos\left(\frac{\Delta E}{2}\right)}$$

IF hyperbolic orbits,

$$\cosh\left(\frac{\Delta H}{2}\right) = 1 - 2x_1$$

$$p = \frac{r_o r (1 - \cos(\Delta\nu))}{r_o + r - 2\sqrt{r_o r} \cos\left(\frac{\Delta\nu}{2}\right) \cosh\left(\frac{\Delta H}{2}\right)}$$

$$f = 1 - \frac{r}{p} (1 - \cos(\Delta\nu)) \quad g = \frac{r r_o \sin(\Delta\nu)}{\sqrt{\mu p}}$$

$$\dot{f} = \sqrt{\frac{1}{p}} \tan\left(\frac{(\Delta\nu)}{2}\right) \left(\frac{1 - \cos(\Delta\nu)}{p} - \frac{1}{r} - \frac{1}{r_o} \right)$$

$$\dot{g} = 1 - \frac{r_o}{p} (1 - \cos(\Delta\nu))$$

$$\dot{\vec{v}_o} = \frac{\dot{\vec{r}} - f \vec{r}_o}{g} \quad \dot{\vec{v}} = \frac{\dot{\vec{g}} \vec{r} - \vec{r}_o}{g}$$

6.7.3 Lambert—Universal Variables

As noted in the previous section, Gauss's original formula is limited. It works well for closely spaced vectors but is limited by the type of orbit transfer and the spread between the vectors. As we showed in Sec. 4.3 about solving Kepler's problem, a method using universal variables allows a simple solution for *any* type of transfer orbit. The universal-variable method for the Lambert problem is particularly important because it can apply to many intercept and rendezvous problems.

The process parallels the development of universal variables to solve Kepler's problem. Remember several points when you look at the final solution. The *universal variable*, χ , relates energy, ξ , and angular momentum, h ($\chi = \sqrt{a} \Delta E$). The variable, y , relates to the area swept out during the transfer, as shown in Gauss's original method, and the variable,

ψ , represents the square root of the distance traveled between the two known position vectors ($\psi \Delta E^2$).

Bate, Mueller, and White (1971, 193–198) present Newton-Raphson's iterative scheme to find the universal variables. A bisection technique sacrifices a little time but appears to be much more robust for a wide variety of orbits. In this case, the object is to find the value of ψ that corresponds to the given change in time. We accomplish this by bounding the correct value of ψ and picking a trial value of ψ halfway between these bounds. Subsequent iterations determine the upper and lower bounds and successively readjust them until the interval for ψ is tight enough to locate the correct value of ψ .

Recall the solution of Kepler's equation from Sec. 4.3. The universal-variable solution contained f and g expressions which permitted us to determine the future position and velocity vectors [Eq. (4-67)]:

$$f = 1 - \frac{\chi_o^2}{r_o} c_2 \quad g = t - \frac{\chi_o^3}{\sqrt{\mu}} c_3$$

$$\dot{g} = 1 - \frac{\chi_o^2}{r} c_2 \quad \dot{f} = \frac{\sqrt{\mu}}{rr_o} \chi_o (\psi c_3 - 1)$$

We also developed expressions for the f and g functions if the change in true anomaly were known (notice an equivalent expression replaces the tangent half angle) [Eq. (4-63)]:

$$f = 1 - \frac{r}{p} (1 - \cos(\Delta\nu)) \quad g = \frac{rr_o \sin(\Delta\nu)}{\sqrt{\mu p}}$$

$$\dot{f} = \sqrt{\frac{\mu}{p}} \left(\frac{1 - \cos(\Delta\nu)}{\sin(\Delta\nu)} \right) \left(\frac{1 - \cos(\Delta\nu)}{p} - \frac{1}{r} - \frac{1}{r_o} \right) \quad \dot{g} = 1 - \frac{r_o}{p} (1 - \cos(\Delta\nu))$$

Because solving the Lambert problem involves determining the orbit that will fit between two vectors (the change in true anomaly is known), it's reasonable to proceed by solving for the universal variable from the f equation in both relations. Thus,

$$\chi_o = \sqrt{\frac{rr_o (1 - \cos(\Delta\nu))}{pc_2}} \quad (6-47)$$

If we substitute this result into the \dot{f} expression and equate it with \dot{f} for the universal formulation,

$$\left(\frac{1 - \cos(\Delta\nu)}{\sin(\Delta\nu)} \right) \left(\frac{1 - \cos(\Delta\nu)}{p} - \frac{1}{r} - \frac{1}{r_o} \right) = \sqrt{\frac{1 - \cos(\Delta\nu)}{rr_o c_2}} (\psi c_3 - 1)$$

Let's simplify the expression by multiplying by the magnitudes of the position vectors. Although not apparent yet, this will allow us to define a new constant:

$$\begin{aligned} & \left(\frac{1 - \cos(\Delta\nu)}{\sin(\Delta\nu)} \right) \left(\frac{rr_o(1 - \cos(\Delta\nu))}{p} \right) + \left(\frac{rr_o(1 - \cos(\Delta\nu))}{\sin(\Delta\nu)} \right) \left(-\frac{1}{r_o} - \frac{1}{r} \right) \\ &= \frac{rr_o\sqrt{1 - \cos(\Delta\nu)}}{\sqrt{rr_o}\sqrt{c_2}} (\psi c_3 - 1) \end{aligned}$$

Dividing through by the first factor of the first term yields a relationship between the universal variable and the change in true anomaly:

$$\frac{rr_o(1 - \cos(\Delta\nu))}{p} = r_o + r + \frac{\sqrt{rr_o}\sin(\Delta\nu)}{\sqrt{1 - \cos(\Delta\nu)}} \frac{(\psi c_3 - 1)}{\sqrt{c_2}} \quad (6-48)$$

Now, define two new variables:

$$y \equiv \frac{rr_o(1 - \cos(\Delta\nu))}{p} \quad A \equiv \frac{\sqrt{rr_o}\sin(\Delta\nu)}{\sqrt{1 - \cos(\Delta\nu)}} \quad (6-49)$$

and write Eq. (6-48) as

$$y = r_o + r + \frac{A(\psi c_3 - 1)}{\sqrt{c_2}}$$

From Eq. (6-47) and the definition in Eq. (6-49), the universal variable becomes:

$$\chi_o = \sqrt{\frac{y}{c_2}}$$

Now you can solve the g function equation for time,

$$t = \frac{\chi_o^3}{\sqrt{\mu}} c_3 + \frac{rr_o \sin(\Delta\nu)}{\sqrt{\mu p}}$$

and, when you include the new variables,

$$t = \frac{\chi_o^3 c_3}{\sqrt{\mu}} + \frac{A \sqrt{y}}{\sqrt{\mu}}$$

As with the Kepler problem, we now have the universal variables in terms of known initial conditions. It's useful to convert the f and g functions of Eq. (4-63) to a form using the newly defined variables:

$$f = 1 - \frac{y}{r_o} \quad \dot{g} = 1 - \frac{y}{r}$$

$$g = A \sqrt{\frac{y}{\mu}}$$

We can't determine the rate of change of f as easily, but we *can* find an expression because we know that $f\dot{g} - \dot{f}g = 1$. Thus,

$$\dot{f} = \frac{1-f\dot{g}}{g} = \frac{\sqrt{\mu y}(-r-r_o+y)}{r_o r A}$$

Let's set up an iterative procedure to find the variables and then determine the final vectors. Recall the solution to Kepler's problem in which the second velocity vector was a function of the f and g functions and the two position vectors:

$$\vec{r} = f\vec{r}_o + g\vec{v}_o$$

If we solve this expression for the initial velocity vector, we'll get half the solution:

$$\vec{v}_o = \frac{\vec{r} - f\vec{r}_o}{g}$$

Now, let's determine the final velocity vector using the velocity-update equation if we've calculated the rate of f .

$$\vec{v} = \dot{f}\vec{r}_o + \dot{g}\vec{v}_o$$

We can get an alternate solution by recalling the identity for the f and g functions, $f\dot{g} - \dot{f}g = 1$. Then, the final velocity is

$$\vec{v} = \frac{f\dot{g} - 1}{g} \vec{r}_o + \dot{g} \frac{\vec{r} - f\vec{r}_o}{g}$$

or

$$\vec{v} = \frac{\dot{g}\vec{r} - \vec{r}_o}{g}$$

Implementing the Lambert—Universal Variable Algorithm

Solving Lambert's problem with this technique is difficult because the iteration isn't always well-behaved. As mentioned earlier, a Newtonian iteration works fine on most problems but fails to converge on many difficult (hyperbolic) orbits. By examining Fig. 6-14, we can see that the curves vary significantly for some orbits. In particular, longer times of flight and hyperbolic orbits occur in "steep" and "flat" portions of the curves, respectively. For this reason, we choose a simpler, slightly longer approach which doesn't suffer

the same limitations. The bisection technique works well on all orbit types and is only about 5% slower. The initial limits are chosen to let the most common solutions converge. If you'll be encountering very eccentric (hyperbolic) orbits, you should expand the lower negative limit. Setting the initial value of ψ_o to zero is purely arbitrary. Canonical units are very useful because they eliminate all operations with the gravitational parameter and replace them with unity values. This replacement saves a lot of time. The transfer-method term is input, allowing us to choose either the short- or long-way trajectories— $t_m = +$ (Short Way) or $-$ (Long Way). Finally, we can easily solve for multiple revolutions by adjusting the initial bounds.

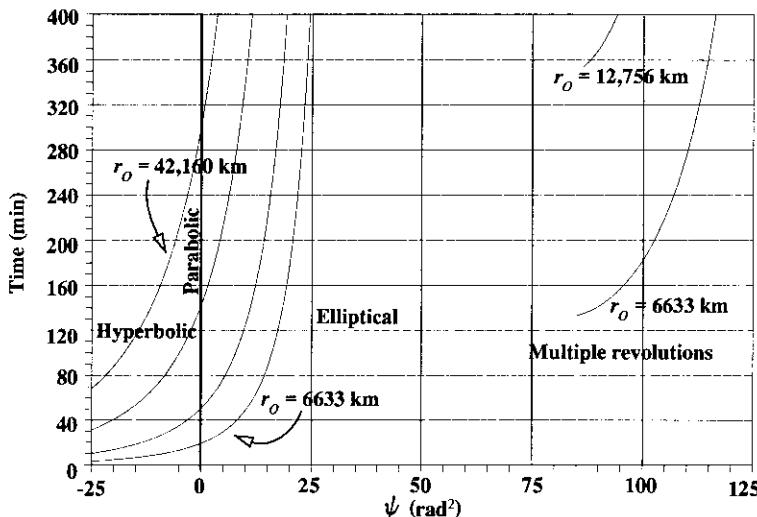


Figure 6-14. Plot of Time versus ψ . Several different transfers are shown with their equivalent values of ψ . Hyperbolic transfers have negative values of ψ , and elliptical orbits have positive values of ψ . Parabolic orbits occur when $\psi = 0$. As the initial orbit position, r_o , gets farther from the Earth, the curves shift up because in a short amount of time, the transfers are almost all hyperbolic. Each final position vector is about 170° from the initial position.

ALGORITHM 52: Lambert—Universal Variables ($\vec{r}_o, \vec{r}, \Delta t_o, t_m \Rightarrow \vec{v}_o, \vec{v}$)

$$\cos(\Delta\nu) = \frac{\vec{r}_o \cdot \vec{r}}{|\vec{r}_o||\vec{r}|}$$

$$\sin(\Delta\nu) = \frac{|\vec{r}_o \times \vec{r}|}{|\vec{r}_o||\vec{r}|} \quad \text{or} \quad \sin(\Delta\nu) = t_m \sqrt{1 - \cos^2(\Delta\nu)}$$

$$A = t_m \sqrt{r r_o (1 + \cos(\Delta\nu))}$$

If $A = 0.0$, we can't calculate the orbit.

$$\psi_n = 0.0, \text{ therefore } c_2 = \frac{1}{2} \text{ and } c_3 = \frac{1}{6}$$

$$\psi_{up} = 4\pi^2 \text{ and } \psi_{low} = -4\pi$$

LOOP

$$y_n = r_o + r + \frac{A(\psi_n c_3 - 1)}{\sqrt{c_2}}$$

IF $A > 0.0$ and $y < 0.0$ THEN

readjust ψ_{low} until $y > 0.0$

$$\chi_o = \sqrt{\frac{y_n}{c_2}}$$

$$\Delta t = \frac{\chi_o^3 c_3 + A \sqrt{y_n}}{\sqrt{\mu}}$$

[IF $\Delta t \leq \Delta t_o$
 reset $\psi_{low} \leftarrow \psi_n$
 ELSE
 reset $\psi_{up} \leftarrow \psi_n$

$$\psi_{n+1} = \frac{\psi_{up} + \psi_{low}}{2}$$

Find $c_2 c_3 (\psi_{n+1} \Rightarrow c_2, c_3)$

$$\psi_n \leftarrow \psi_{n+1}$$

Check if the first guess is too close

UNTIL $|\Delta t - \Delta t_o| < 1 \times 10^{-6}$

$$f = 1 - \frac{y_n}{r_o} \quad \dot{g} = 1 - \frac{y_n}{r} \quad g = A \sqrt{\frac{y_n}{\mu}}$$

$$\dot{\vec{v}}_o = \frac{\dot{\vec{r}} - f \vec{r}_o}{g} \quad \dot{\vec{v}} = \frac{\dot{g} \vec{r} - \vec{r}_o}{g}$$

6.7.4 Lambert Solution—Battin Method

Battin (1987, 325–342) challenges us to adopt a technique he developed which is robust, doesn't suffer from the 180°-transfer difficulty of most Lambert routines, and is very appropriate to this discussion of the problem. Battin has described the somewhat lengthy derivation in several papers and in his book. Thus I'll briefly introduce it here and present an algorithm for computation.

Battin's method has its roots in Gauss's original technique, so much of the initial development resembles the Lambert and Gaussian formulations. He extensively uses continued fractions because they guarantee convergence. Battin suggests refinements to the continued-fraction equations to make them converge better. I've listed formulas to determine the coefficients so you can easily append additional terms into a particular solution. The method also requires us to solve a cubic equation. Although analytical techniques exist which determine all real and imaginary roots, Battin shows an intriguing development which automatically yields the largest correct, real root. It uses continued fractions.

Battin introduces the *parabolic mean point radius*, r_{op} , as the arithmetic mean between the semimajor axis of the fundamental ellipse and the line segment from the primary focus to the point arithmetic mean location on the chord (1987, 269). This element is very useful in deriving the method. Consult Battin (1987, 267–271) for additional information.

Note that the variable y in Algorithm 53 is *not* the same as the sector-to-triangle ratio used in the Gaussian development. Relations from the original Gaussian technique determine the eccentric and hyperbolic anomalies. The sine and cosine expressions for half angles will not resolve the quadrant, although you can use either for solution. We use the f and g functions, Eq. (4-64) or Eq. (4-66), to find the resulting vectors. The algorithm uses a transfer-method term that allows us to choose either the short- or long-way trajectories— $t_m = +$ (Short Way) or $-$ (Long Way).

ALGORITHM 53: Lambert—Battin Method ($\vec{r}_o, \vec{r}, \Delta t, t_m \Rightarrow \vec{v}_o, \vec{v}$)

$$\cos(\Delta\nu) = \frac{\vec{r}_o \cdot \vec{r}}{r_o r}$$

$$\sin(\Delta\nu) = \frac{|\vec{r}_o \times \vec{r}|}{r_o r} \quad \text{or} \quad \sin(\Delta\nu) = t_m \sqrt{1 - \cos^2(\Delta\nu)}$$

$$c = \sqrt{r_o^2 + r^2 - 2r_o r \cos(\Delta\nu)}$$

$$s = \frac{r_o + r + c}{2}$$

$$\epsilon = \frac{r - r_o}{r_o}$$

$$\tan^2(2w) = \frac{\frac{\epsilon^2}{4}}{\sqrt{\frac{r}{r_o}} + \frac{r}{r_o} \left(2 + \sqrt{\frac{r}{r_o}} \right)}$$

$$r_{op} = \sqrt{r_o r} \left\{ \cos^2 \left(\frac{\Delta\nu}{4} \right) + \tan^2(2w) \right\}$$

$$t = \frac{\sin^2 \left(\frac{\Delta\nu}{4} \right) + \tan^2(2w)}{\sin^2 \left(\frac{\Delta\nu}{4} \right) + \tan^2(2w) + \cos \left(\frac{\Delta\nu}{2} \right)} \quad 0^\circ < \Delta\nu < 180^\circ$$

$$t = \frac{\cos^2 \left(\frac{\Delta\nu}{4} \right) + \tan^2(2w) - \cos \left(\frac{\Delta\nu}{2} \right)}{\cos^2 \left(\frac{\Delta\nu}{4} \right) + \tan^2(2w)} \quad 180^\circ < \Delta\nu < 360^\circ$$

$$m = \frac{\mu \Delta t^2}{8r_{op}^3}$$

Let $x = t$ if the orbit is elliptical, 0.0 otherwise

LOOP

$$\eta = \frac{x}{(\sqrt{1+x}+1)^2}$$

$$\xi(x) = \cfrac{8(\sqrt{1+x}+1)}{3 + \cfrac{1}{9 + \cfrac{\bar{7}\eta}{5 + \eta + \cfrac{16}{1 + \cfrac{63\eta}{1 + \cfrac{c_\eta\eta}{1 + \dots}}}}}}$$

$$c_\eta = \frac{n^2}{(2n)^2 - 1} \quad n = 4, 5\dots$$

$$h_1 = \frac{(t+x)^2 (1 + \xi(x) (1 + 3x))}{(1 + 2x + t) (3 + x (1 + 4\xi(x)))}$$

$$h_2 = \frac{m (1 - (x-t) (\xi(x)))}{(1 + 2x + t) (3 + x (1 + 4\xi(x)))}$$

Solve Cubic ($y^3 - y^2 - h_1 y - h^2 = 0$) or use:

$$B = \frac{27h_2}{4(1+h_1)^3} \quad U = \frac{-B}{2(\sqrt{1+B}+1)}$$

IF $n = (1, 3, 5, \dots)$

$$K(U) = \frac{\frac{1}{3}}{1 - \frac{\frac{8}{27}U}{1 - \frac{c_U U}{1 - \dots}}}$$

IF $n = (0, 2, 4, \dots)$

$$c_U = \frac{2(3n+1)(6n-1)}{9(4n-1)(4n+1)}$$

$$c_U = \frac{2(3n+2)(6n+1)}{9(4n+1)(4n+3)}$$

$$y = \frac{1+h_1}{3} \left(2 + \frac{\sqrt{1+B}}{1-2U(K(U))^2} \right)$$

$$x = \sqrt{\left(\frac{1-t}{2} \right)^2 + \frac{m}{y^2}} - \frac{1+t}{2}$$

UNTIL x stops changing

$$a = \frac{\mu \Delta t^2}{16 r_{op}^2 x y^2}$$

IF $a > 0.0$

$$\sin\left(\frac{\beta_e}{2}\right) = \sqrt{\frac{s-c}{2a}}$$

IF $\Delta\nu > \pi$

$$\beta_e = -\beta_e$$

$$a_{min} = \frac{s}{2}$$

$$t_{min} = \sqrt{\frac{a_{min}^3}{\mu}} \{ \pi - \beta_e + \sin(\beta_e) \}$$

$$\sin\left(\frac{\alpha_e}{2}\right) = \sqrt{\frac{s}{2a}}$$

IF $\Delta t > t_{min}$

$$\alpha_e = 2\pi - \alpha_e$$

$$\Delta E = \alpha_e - \beta_e$$

$$\left. \begin{array}{l}
 f = 1 - \frac{a}{r_o} (1 - \cos(\Delta E)) \quad g = \Delta t - \sqrt{\frac{a^3}{\mu}} (\Delta E - \sin(\Delta E)) \\
 \dot{g} = 1 - \frac{a}{r} (1 - \cos(\Delta E)) \\
 \text{IF } a < 0.0 \\
 \sinh\left(\frac{\alpha_h}{2}\right) = \sqrt{\frac{s}{-2a}} \\
 \sinh\left(\frac{\beta_h}{2}\right) = \sqrt{\frac{s-c}{-2a}} \\
 \Delta H = \alpha_h - \beta_h \\
 f = 1 - \frac{a}{r_o} (1 - \cosh(\Delta H)) \quad g = \Delta t - \sqrt{\frac{-a^3}{\mu}} (\sinh(\Delta H) - \Delta H) \\
 \dot{g} = 1 - \frac{a}{r} (1 - \cosh(\Delta H)) \\
 \dot{v}_o = \frac{\dot{r} - \dot{f}r_o}{g} \quad \dot{v} = \frac{\dot{g}r - \dot{r}_o}{g}
 \end{array} \right. \quad \dots$$

▼ Example 6-5. Solving Lambert's Problem.

GIVEN: $\dot{r}_o = 15,945.34 \hat{i}$ km $\Delta t = 76.0$ min

$\dot{r} = 12,214.83899 \hat{i} + 10,249.46731 \hat{j}$ km $t_m = \text{short way}$

FIND: \dot{v}_o, \dot{v} using each technique

First, convert to canonical units by dividing by the Earth's radius:

$$\dot{r}_o = 2.5 \hat{i} \text{ ER}, \dot{r} = 1.915111 \hat{i} + 1.606969 \hat{j} \text{ ER}, \Delta t = 5.6519 \text{ TU}$$

Find several initial parameters for solution:

$$\psi_n = 0.0, \text{ therefore } c_2 = \frac{1}{2} \text{ and } c_3 = \frac{1}{6}$$

$$\psi_{up} = 4\pi^2 \text{ and } \psi_{low} = -4\pi$$

$$\left. \begin{array}{l}
 \cos(\Delta\nu) = \frac{\dot{r}_o \cdot \dot{r}}{r_o r} \\
 \sin(\Delta\nu) = \frac{|\dot{r}_o \times \dot{r}|}{r_o r} \quad \text{or} \quad \sin(\Delta\nu) = t_m \sqrt{1 - \cos^2(\Delta\nu)} \\
 \end{array} \right\} \Delta\nu = 40^\circ$$

Now check to see if the orbit is possible. A must not be zero.

$$A = t_m \sqrt{rr_o(1 + \cos(\Delta\nu))} = 3.322$$

Table 6-3 shows the results of the iteration loop.

TABLE 6-3. Results of Example Problem for the Lambert-Universal Variable. Iteration values are shown. Although the process converges slowly at the end, we achieve the correct answer. Initial values of $\psi_{low} = 0$, $\psi_{high} = 4\pi^2$.

Iteration	y_n	x_o	Δt	ψ_{n+1}
1	7.845 86	11.059 22	92.696 71	19.739 21
2	5.000 00	4.967 29	19.847 16	9.869 60
3	2.913 81	2.992 49	9.156 43	4.934 80
10	1.738 69	2.081 70	5.701 62	2.583 06
17	1.719 50	2.066 89	5.651 77	2.546 62
25	1.719 54	2.066 93	5.651 88	2.546 70

The final answers are

$$\hat{v}_o = 0.260\ 445\ 0 \hat{i} + 0.368\ 858\ 9 \hat{j} \text{ ER/TU} = 2.058\ 913 \hat{i} + 2.915\ 965 \hat{j} \text{ km/s}$$

$$\hat{v} = -0.436\ 610\ 4 \hat{i} + 0.115\ 151\ 5 \hat{j} \text{ ER/TU} = -3.451\ 565 \hat{i} + 0.910\ 315 \hat{j} \text{ km/s}$$

Using the minimum-energy solution, we find that

$$\hat{v}_o = 0.258\ 990 \hat{i} + 0.369\ 876 \hat{j} \text{ ER/TU} = 2.047\ 409 \hat{i} + 2.924\ 003 \hat{j} \text{ km/s}$$

$$\hat{v} = -0.436\ 149 \hat{i} + 0.116\ 866 \hat{j} \text{ ER/TU} = -3.447\ 919 \hat{i} + 0.923\ 867 \hat{j} \text{ km/s}$$

Remember that the minimum energy approach determines a different time that corresponds to that particular orbit. For this case, the minimum time is 75.6708 min, and the semimajor axis is 10,699.451 km. The proximity of the solution to the other Lambert techniques is only illustrative due to the change in time that was used (76 min). A different time would produce significantly different results for the Lambert solutions, but the same minimum solution.

The original Gaussian technique finds

$$\hat{v}_o = 0.260\ 446 \hat{i} + 0.368\ 858 \hat{j} \text{ ER/TU} = 2.058\ 925 \hat{i} + 2.915\ 956 \hat{j} \text{ km/s}$$

$$\hat{v} = -0.436\ 611 \hat{i} + 0.115\ 150 \hat{j} \text{ ER/TU} = -3.451\ 569 \hat{i} + 0.910\ 301 \hat{j} \text{ km/s}$$

Battin's method requires much fewer iterations (5) and arrives at the same answer:

$$\hat{v}_o = 0.260\ 445\ 0 \hat{i} + 0.368\ 858\ 9 \hat{j} \text{ ER/TU} = 2.058\ 913 \hat{i} + 2.915\ 965 \hat{j} \text{ km/s}$$

$$\hat{v} = -0.436\ 610\ 4 \hat{i} + 0.115\ 151\ 5 \hat{j} \text{ ER/TU} = -3.451\ 565 \hat{i} + 0.910\ 315 \hat{j} \text{ km/s}$$

▲ Numerical analysis indicates that the universal-variable and Battin approaches provide the most accurate answer, but this doesn't mean they will always be best.

6.8 Application: Targeting Problem

Solving Lambert's problem allows us to examine several useful situations. For initial orbit determination, these techniques yield velocity vectors from sets of observed position vectors. The orbit should be the same unless we've observed the satellite during a maneuver. The problem can also be viewed as two positions in one orbit, referred to as a transfer orbit, used to plan a maneuver of a satellite between two separate positions. Another application is the case in which the position vectors represent places in two *separate* orbits. This is also a technique for initial orbit determination because we're trying to establish an orbit for the transfer. This case is the solution alluded to in Chap. 5 for the complex problems of orbital maneuvering. We call this problem intercept and rendezvous or, more generally, the *targeting problem*.

Let's apply our solution of Lambert's problem to the important issues of intercept and rendezvous. These two missions both begin with a maneuver; however, the intercept may or may not have a second maneuver depending on whether it's a fly-by or a rendezvous. This section explores various combinations of maneuvers while trying to quantify (1) the minimum change in velocity (Δv) to intercept or rendezvous, and (2) the minimum time to intercept or rendezvous with a specified change in velocity. These maneuvers are classified together as the *targeting problem* and are further distinguished by the number of propulsive burns. *Intercept* maneuvers usually employ a single impulse, whereas *rendezvous* maneuvers use two burns.

A special focus of this section is to analyze the transfer between orbits and the timing necessary to do it. Chapter 5 discussed many cases in which the analysis became too complex to use with simplified techniques. Solving Lambert's problem allows us to tackle these tougher cases. It also lets us explore minimum-fuel and minimum-time contestants for complex and general problems. Most literature on the targeting problem deals with simple transfers between orbits. These transfers are unconstrained by time and therefore much easier to analyze, but they don't represent realistic targeting maneuvers. Here, we'll start with the positions of the interceptor and target satellite at an epoch time and explore ways to find the minimum velocity and the time at which a specified velocity change should occur. Let's first list the assumptions:

1. **Two-body dynamics defines all motion.** This is a considerable simplification but is necessary because the classical techniques to do targeting calculations *all* use two-body dynamics. Baker (1967, 198–208) does show expressions to include some secular effects of central-body variations, but the theory doesn't apply to all cases. This assumption obviously degrades if we try transfers between near-Earth orbits using long times of flight (*TOF*). Recognize that the time and nearness to Earth are key to the accuracy of this assumption. Near-Earth orbits using large *TOFs* to do targeting maneuvers react much more to perturbing forces of the nonspherical Earth and atmospheric effects. Higher-altitude orbits with long *TOFs* change significantly because of third-body forces and solar-radiation pressure.

2. **All maneuvers employ impulsive burns.** Although actual burns require a finite amount of time to impart the actual change in velocity, this is still a reasonable assumption in most cases. In fact, as long as the burn is a “small” fraction of the orbital period, this assumption is valid.
3. **Phasing orbits are not considered** directly because they add complexity and increase the number of separate burns which must occur during the transfer. Simply matching the times for the maneuver can dominate the computations. Phasing orbits place additional calculations at each step of the process and are beyond the scope of this section. Using phasing orbits may be the best solution to targeting problems whenever much time must pass before the maneuver is complete. That’s because a direct transfer involves expending a lot of energy to meet the target at a specified point.
4. **The section of the transfer orbit we use must not pass through the Earth.** Several calculations can ensure this is true, but we’ll make it an assumption. Of course, if the satellite hits the Earth while on the transfer, the analysis is pointless.

Fig. 6-15 shows the overall geometry. We know the initial position and velocity vectors of the interceptor, \mathbf{r}_{int} and \mathbf{v}_{int} , and the target, \mathbf{r}_{tgt} and \mathbf{v}_{tgt} . We also know either the desired *TOF* to complete the maneuver or the maximum change in velocity of the interceptor. Methods of solving the targeting problems differ depending on what else is known.

The first step of the solution is to consider the target’s motion. The target must be propagated from its initial position to where it will be *after the TOF of the transfer and before any targeting calculations may occur*. Figure 6-15 shows this symbolically as the movement from \mathbf{r}_{tgt} and \mathbf{v}_{tgt} , to \mathbf{r}_{tgt_b} and \mathbf{v}_{tgt_b} after *TOF*, which is known as Kepler’s problem (Sec. 4.3).

It’s interesting to note that, because the Kepler algorithm merely propagates the satellite’s position and velocity vectors, the *TOF* may be either positive or negative. Solutions are possible for all cases except where the *TOF* is very close to zero. Computer programs should be able to handle cases in which the *TOF* is 0.0 because the satellite is still in the same place. A tolerance of about 0.0001 minutes for this case helps us avoid computational difficulties. Using the universal-variable algorithm (Algorithm 52) or the Battin method (Algorithm 53) allows us to solve all possible orbits.

The propagation part of the targeting problem permits relatively easy use of perturbation methods because we can handle the propagation through various well-known techniques, including secular updates of J_2 (**PKEPLER**), Cowell’s formulation, an analytical method, and so on (see Chaps. 7 and 8 for details). Including perturbations for the target transfer may even be advisable for simulations, in which the target will move for a long time before the actual targeting maneuver takes place.

The second phase of the problem deals with actually transferring the interceptor from its initial position to that of the target satellite after it has moved through the *TOF*. This problem is known in the classical literature as Lambert’s problem. Various solution techniques *all* use two-body dynamics. This section uses the universal-variable form of solu-

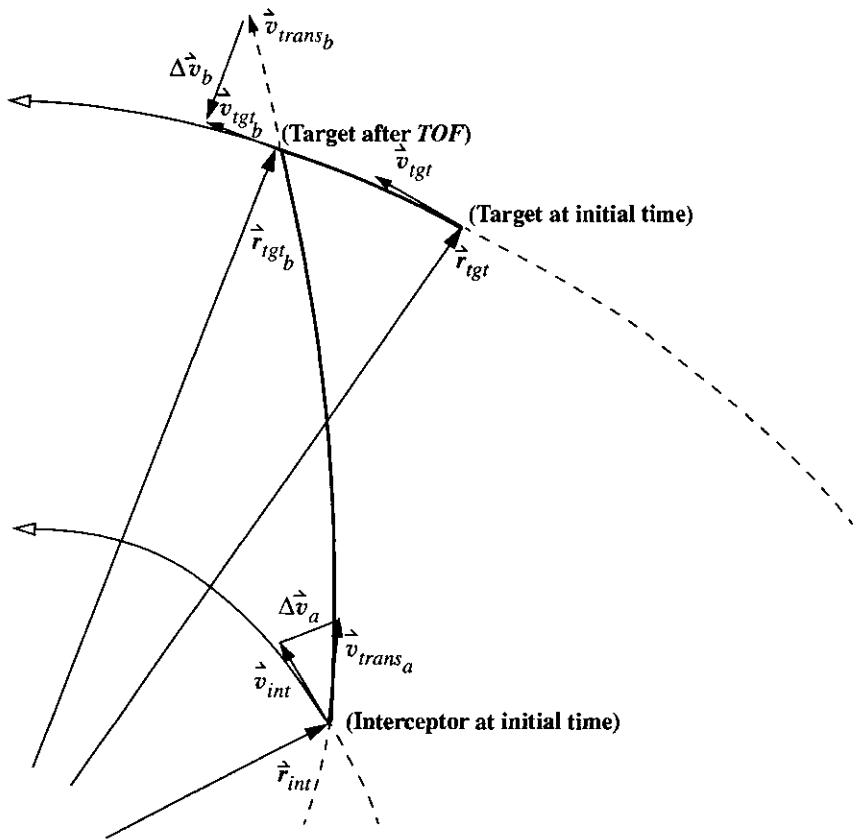


Figure 6-15. Geometry for Intercept Maneuvers. For the targeting maneuver to take place, we must account for the target's motion during the transfer.

tion because we're considering all orbit types in the evaluation. Recall that this method tries to find a transfer between two known position vectors.

The short way to the target is almost always used because of time constraints. The long way can save more fuel in many circumstances, so we should consider it when trying to determine the smallest change in velocity. As with the solution for Kepler's problem, we use universal variables. A solution for all conic sections is especially important in this phase because of the problem's geometry. Figure 6-16 shows that, as we increase the TOF for transfer, the transfer trajectory can change through each type of conic section.

In Algorithm 52, the first calculation finds $\cos(\Delta\nu)$, or the cosine of the angle between the two position vectors. This value then adds to 1.0, and the algorithm proceeds. Notice if the vectors are exactly 180° apart, the addition yields 0.0, and we can't solve the problem because there's no distinction between the long and short ways. To solve this special case,

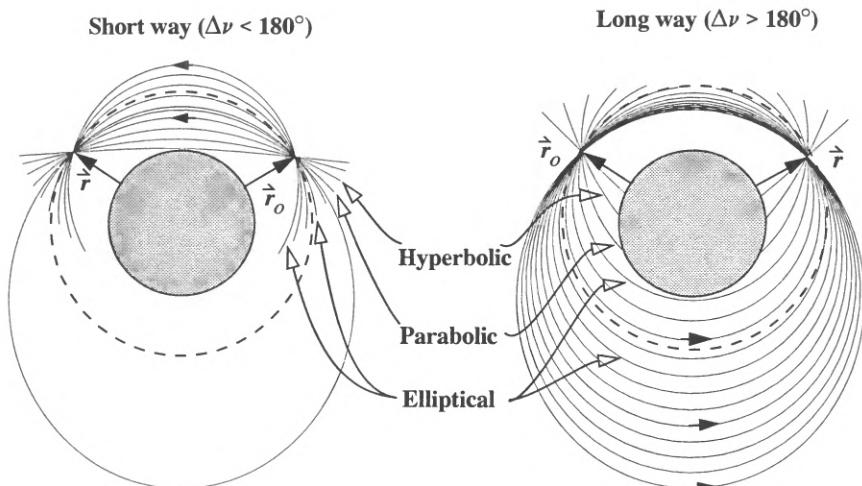


Figure 6-16. Varying Time of Flight for Intercept. As the time of flight increases for the transfer to the target (short way left, long way right), the transfer orbit becomes less eccentric until it reaches a minimum at $e = 0$. The circular orbit (dashed circle) is the point at which the long way and short way are equal, although their change in velocity is not. The eccentricity begins to increase after the circular orbit. Note the position of the initial vector for each case.

we use the Battin, Thorne, or r -iteration technique discussed previously (see Escobal [1965] 1985, 213–218). The 180° case also represents a Hohmann-like transfer, which we can solve using simple two-body dynamics. In addition, the code contains an iteration loop; the algorithm's success depends on the maximum number of iterations. If it doesn't converge within the specified iterations, extremely hyperbolic orbits are usually the cause.

Before trying to work through various results, we must check all solutions to ensure the transfer trajectory doesn't intersect the Earth. Figure 6-17 shows the orbits that could strike the Earth. Remember, the transfer orbit may intersect the Earth as long as the satellite isn't *on* the transfer orbit during those times.

We can use several ways to determine if the trajectory will hit the Earth. Although the usual method involves calculating the flight-path angle and the radius of perigee, r_p , it's computationally faster first to check the dot products of the position and velocity vectors and to calculate the perigee radius *only if necessary*. Recall the properties of the flight-path angle, ϕ_{fpa} , for an orbit. ϕ_{fpa} is positive as the satellite travels from perigee to apogee and negative on the return. The dot product of the position and velocity vectors gives this change of sign without the trigonometric calculations required to calculate ϕ_{fpa} . The position vector of the interceptor and the transfer velocity will give the sign at the start of the transfer, and the target's propagated position vector and the transfer velocity at the end of

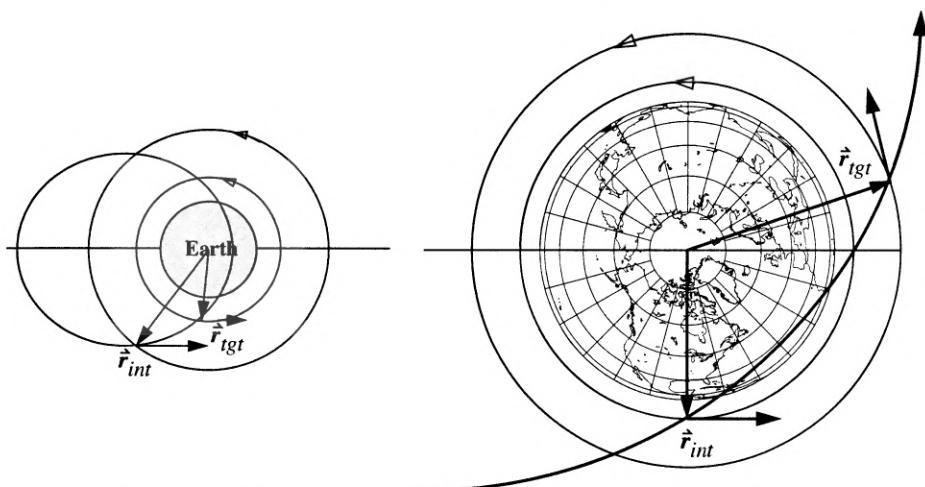


Figure 6-17. Geometry of Orbits that Can Hit the Earth. The orbit transfer on the left is valid because the transfer orbit intersects the Earth after the transfer. The transfer on the right would be unacceptable.

the maneuver will give the final sign. If both signs are positive, the satellites can't hit the Earth because the transfer occurs as they're headed away from the Earth. If both signs are negative, as shown in Fig. 6-17, the transfer is still possible because the target satellite is *in orbit* at the end of the transfer, and therefore it's not on an impact trajectory. In fact, the only case we need to check is when the initial sign is negative and the final is positive, indicating the transfer orbit's perigee occurred *during* the maneuver. This particular case occurs if $\vec{r}_{int} \cdot \vec{v}_{trans_a} < 0$ (satellite initially headed towards perigee) and $\vec{r}_{tgt} \cdot \vec{v}_{trans_b} > 0$ (satellite finally headed towards apogee). If so, we must calculate the radius of perigee.

ALGORITHM 54: Hit Earth ($\vec{r}_{int}, \vec{r}_{tgt}, \vec{v}_{trans_a}, \vec{v}_{trans_b}, a \Rightarrow r_p$, collision)

$$\text{IF } ((\vec{r}_{int} \cdot \vec{v}_{trans_a}) > 0.0) \text{ and } ((\vec{r}_{tgt} \cdot \vec{v}_{trans_b}) < 0.0)$$

$$\xi = \frac{\vec{v}_{trans_a}^2}{2} - \frac{\mu}{r_{int}} = -\frac{\mu}{2a}$$

$$h_t = \left| \vec{r}_{int} \times \vec{v}_{trans_a} \right|$$

$$p = \frac{h_t^2}{\mu}$$

$$e = \sqrt{\frac{a-p}{a}}$$

$$r_p = a(1-e)$$

IF $r_p \leq 6378.136$ km (1.0 ER)
 Collision
 ELSE
 No Collision

With the basics identified, we can form an algorithm to use for each scenario we develop. Algorithm 55 shows propagations using only two-body dynamics. Although this isn't the most accurate means of determining future positions, a basic assumption of the Lambert problem was the two-body equations of motion for the actual transfer. For longer propagation times, we can use more accurate propagation algorithms, including the effects of perturbations.

ALGORITHM 55: $\text{Target}(\vec{r}_{int}, \vec{r}_{tgt}, \vec{v}_{int}, \vec{v}_{tgt}, \Delta t \Rightarrow \vec{\Delta v}_a, \vec{\Delta v}_b)$

Propagate target through the desired time-of-flight:

PKEPLER ($\vec{r}_{tgt}, \vec{v}_{tgt}, \Delta t \Rightarrow \vec{r}_{tgt_b}, \vec{v}_{tgt_b}$)

Determine transfer orbit using Lambert solution:

Lambert—Universal ($\vec{r}_{int}, \vec{r}_{tgt_b}, \Delta t, t_m \Rightarrow \vec{v}_{trans_a}, \vec{v}_{trans_b}$)

$$\vec{\Delta v}_a = \vec{v}_{trans_a} - \vec{v}_{int}$$

$$\vec{\Delta v}_b = \vec{v}_{tgt_b} - \vec{v}_{trans_b}$$

Check for Earth impact:

Hit Earth ($\vec{r}_{int}, \vec{r}_{tgt}, \vec{v}_{trans_a}, \vec{v}_{trans_b}, a \Rightarrow r_p, \text{collision}$)

To gain insight into the two steps for solving this problem, first examine the solution and some characteristics of the Lambert problem. Recall Fig. 6-16, which showed two satellite position vectors and the range of trajectories between them. These represent solutions of Lambert's problem using the universal-variable routine for various TOFs. To fully examine the range of potential solutions, it's useful to look at an extended problem. We'll vary time and examine two cases: (1) a fixed target and interceptor, and (2) a moving target and interceptor. The fixed case isn't realistic, but it lays a foundation to compare with

moving methods. It also provides a reference to the simplified techniques in Chap. 5. Figure 6-18 shows the change in velocity between two specified position vectors for the short- and long-way trajectories. The initial conditions are:

INITIAL CONDITIONS FOR FIXED-TARGET PROBLEM

$$\text{INTERCEPTOR} \quad \dot{r}_{IJK} = -6518.1083 \hat{i} - 2403.8479 \hat{j} - 22.1722 \hat{k} \text{ km/s}$$

$$\dot{v}_{IJK} = 2.604\,057 \hat{i} - 7.105\,717 \hat{j} - 0.263\,218 \hat{k} \text{ km/s}$$

$$\text{TARGET} \quad \dot{r}_{IJK} = 6697.4756 \hat{i} + 1794.5832 \hat{j} + 0.0 \hat{k} \text{ km/s}$$

$$\dot{v}_{IJK} = -1.962\,373 \hat{i} + 7.323\,674 \hat{j} + 0.0 \hat{k} \text{ km/s}$$

I've kept the scale for each figure constant to give you a relative sense of equality. The long-way total Δv is usually greater than 30 km/s and doesn't appear on the figures. The initial positions of the interceptor and the target are about 175° apart.

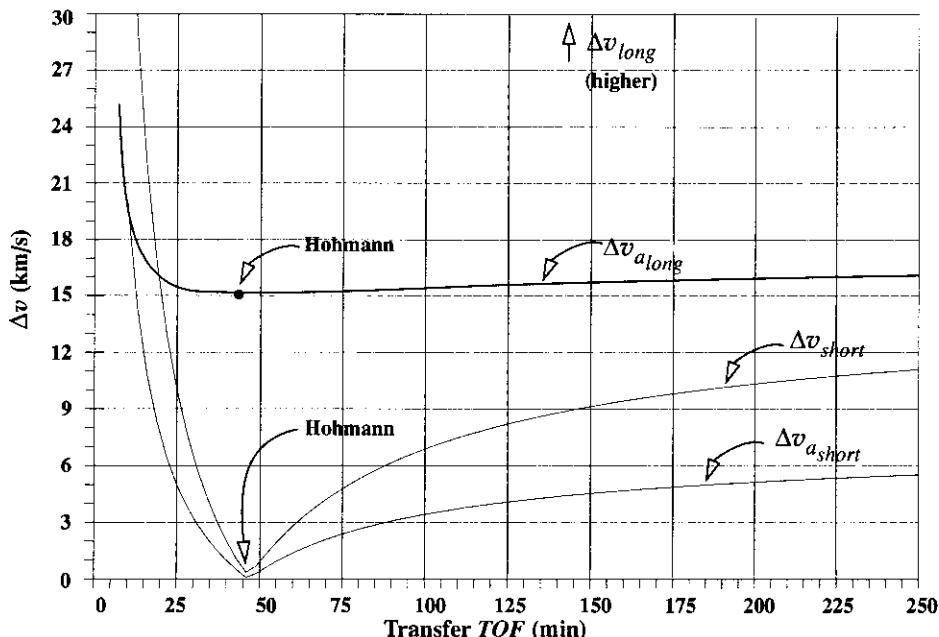


Figure 6-18. Short- and Long-Way Δv for Varying TOF, Target Fixed. As the time of flight increases, the change in velocity reaches a minimum at the point where the transfer mimics a Hohmann transfer. First and total change in velocities (Δv_a , Δv) are shown for both short and long transfers.

As time increases, the velocity decreases to a local minimum. The minimum change in velocity for Fig. 6-18 occurs at 46.68 minutes on the short-way trajectory; $\Delta v_a = 1.88$ m/s, and $\Delta v_b = 265$ m/s. The step-size was adjusted for the interval around 46 minutes to determine this specific value. This same minimum velocity may also be calculated analytically using Escobal's method. Refer to Escobal ([1968] 1979, 81–90).

Because the target is about 175° ahead of the interceptor and the target position is fixed, notice how the short way (Fig. 6-18) produces a dramatically lower change in velocity, whereas the long way requires a relatively high Δv for all transfer times. It turns out the short-way trajectory at 46.68 minutes is similar to a Hohmann transfer. The values for the combined plane change (2°) in this Hohmann-like transfer are about $\Delta v_a = 2.98$ m/s and $\Delta v_b = 262$ m/s. The long way to the target never achieves this value because it is always having to reverse its velocity and thus has much larger values. In general, to get a minimum change in velocity, you want the transfer velocity to be in the *same* direction as your initial velocity.

The second case (moving target and interceptor) considers timing constraints in the computational process. In the real world, the target moves as the interceptor completes the transfer. Thus the static calculations of the fixed-vector case don't apply. The solution is to use the complete targeting algorithm, combining the solutions of the Kepler and Lambert problems. Given the desired *TOF* for the transfer, each case will produce a change in velocity. However, just as with the case of the fixed-position vector, to determine the minimum velocity, we must examine a series of time-of-flight calculations. Because the problem has two parts—the Kepler and Lambert solutions—this starts to become mathematically intensive. For a given range of *TOFs*, the change in velocity will vary, usually through several orders of magnitude, and each set of transfers will contain representative orbits from each type of conic section. We developed the example using one satellite in a circular orbit and one in a near-circular orbit, starting at an epoch where they are very close together. The interceptor vector is changed slightly.

INITIAL CONDITIONS FOR MOVING TARGETS

$$\text{INTERCEPTOR } \dot{\vec{r}}_{IJK} = 5328.7862 \hat{i} + 4436.1273 \hat{j} + 101.4720 \hat{k} \text{ km}$$

$$\dot{\vec{v}}_{IJK} = -4.864\,779 \hat{i} + 5.816\,486 \hat{j} + 0.240\,163 \hat{k} \text{ km/s}$$

$$\text{TARGET } \dot{\vec{r}}_{IJK} = 6697.4756 \hat{i} + 1794.5831 \hat{j} + 0.0000 \hat{k} \text{ km}$$

$$\dot{\vec{v}}_{IJK} = -1.962\,372 \hat{i} + 7.323\,674 \hat{j} + 0.000\,000 \hat{k} \text{ km/s}$$

Figure 6-19 shows the short way and Fig. 6-20 shows the long way. Notice the minimum velocity on each *cycle* (50, 150, 250 minutes on the short way) tends to increase as the *TOF* increases. At first glance, one would think the minimum velocity will remain the same because, after one period, the satellites will be in roughly the same orientation, except that the target will be a little ahead of the interceptor. It turns out the target *is* ahead in time by one period, plus a little bit. To do the transfer, the interceptor must go into a

larger orbit to “waste time” and then return for the final intercept or rendezvous. This is where the concept of phasing revolutions may become important when planning the transfer. It appears that, as the *TOF* increases, the maneuver solved by the Lambert problem becomes more and more inefficient. That’s because the solution trajectory tries to take in *all* the time allotted, rather than phasing and then doing the complete maneuver. Accounting for this time and finding the proper phasing orbit significantly complicates solving for the minimum change in velocity. I’ll discuss this idea later, as it essentially increases the unknowns.

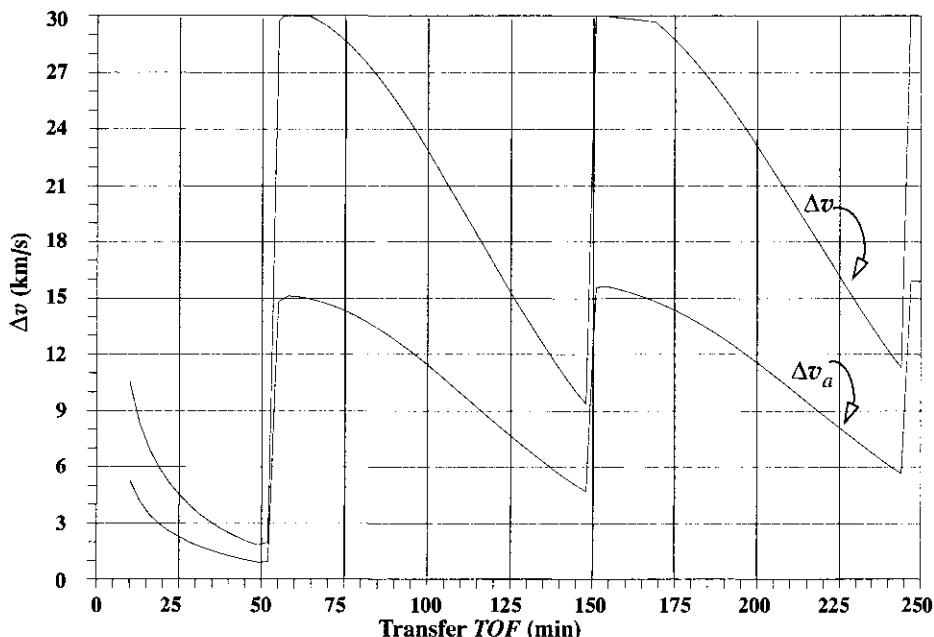


Figure 6-19. Δv for Short Way with Varying *TOF*, Target Moving. As the time of flight increases, the change in velocity decreases initially, but with more revolutions, the change in velocity tends to rise.

Next, each graph shows a large jump discontinuity near 50 minutes, or half the orbital period of the satellites. Because this graph is created from data for the transfer going the short way, the jump represents the point at which the satellite must essentially reverse its velocity vector to maintain a short-way transfer. In order to find the minimum velocity over the range of time specified for analysis, we’d like to examine a curve representing the minimum of both the long and the short ways to the target—a relatively easy task that requires only a few lines of code. Figure 6-21 shows two possible target locations for developing this test.

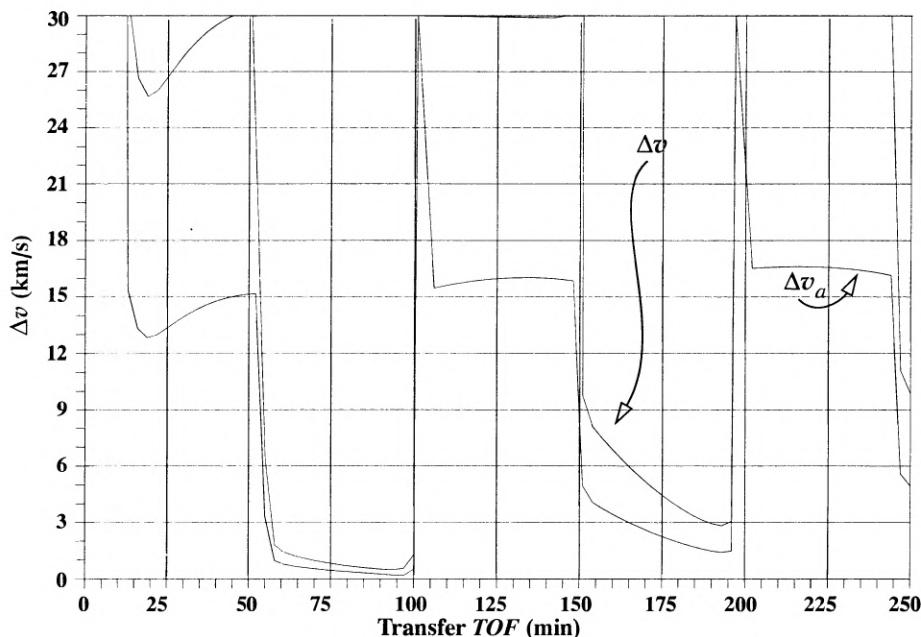


Figure 6-20. Δv for Long Way with Varying TOF, Target Moving. As the time of flight increases, the change in velocity varies significantly. The minimum still occurs at the point where the transfer is a Hohmann transfer.

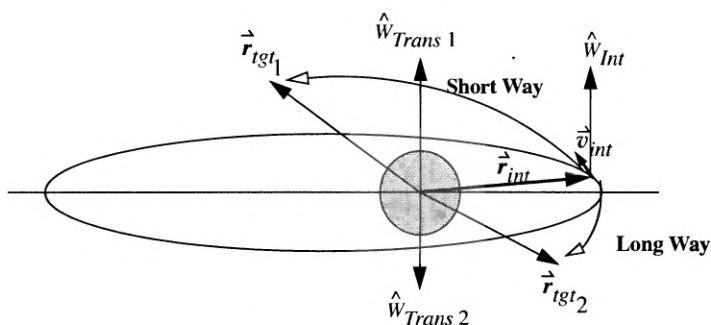


Figure 6-21. Orbital Geometry Showing Transfer Direction. We can determine the “shortest” way to the target by using cross products. The short-way trajectory is used when the normal unit vector from the interceptor, \hat{W}_{Int} , is in the same direction as the transfer unit vector, $\hat{W}_{Trans\ 1}$. The long way results with $\hat{W}_{Trans\ 2}$.

The basic idea relies on the interceptor's direction and the angle between the two position vectors. Using the cross product of the satellite's two state vectors, we can find the normal vector for the transfer orbit ($Tran_n = \dot{r}_{int} \times \dot{r}_{tgt}$) and for the interceptor orbit prior to the burn ($Int_n = \dot{r}_{int} \times v_{int}$). Then,

IF $Int_n \cdot Tran_n < 0$, use long way

IF $Int_n \cdot Tran_n > 0$, use short way

If we're trying to determine the minimum change in velocity, this test will ensure that the interceptor will go the more efficient way to the target; it switches back and forth between the long and short ways as we change the time-of-flight. Several test cases show this test won't *always* produce the minimum change in velocity, but it usually works correctly. To use a more robust method, we should run this check at the start of each half revolution of the target satellite's period. Because the operations are all multiplications, the time penalty isn't terribly large for this extra calculation. The result of this addition yields a change in velocity graph for the combination of long- and short-way transfers, shown in Fig. 6-22.

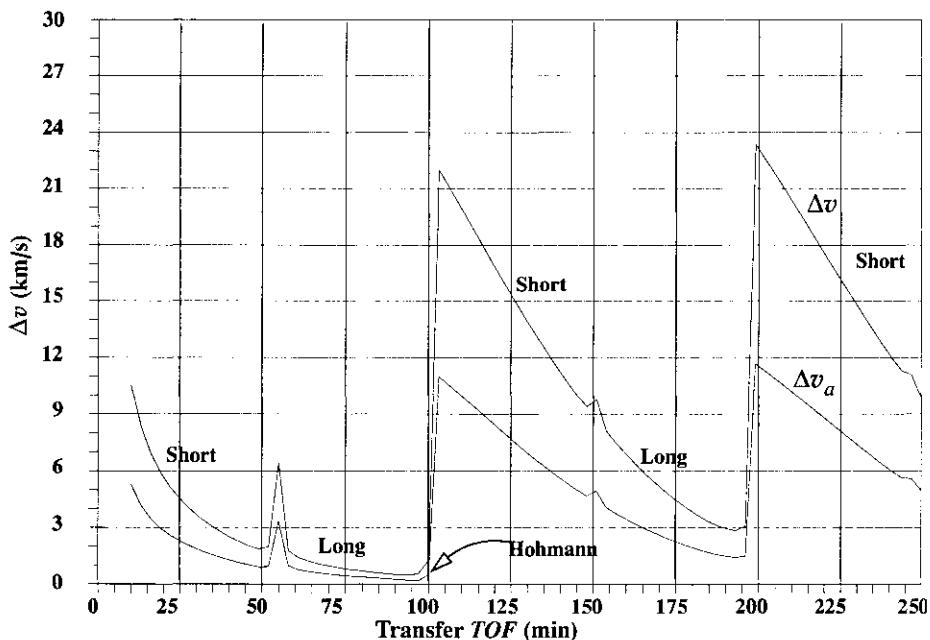


Figure 6-22. Δv for Short and Long Way with Varying TOF, Target Moving. The combination graph for the change in velocity takes the best of both transfers. Remember that the spikes occur because the inclinations differ ($i_{init} = 2^\circ$, $i_{final} = 0^\circ$), and the nodes are different ($\Omega_{init} = 15^\circ$, $\Omega_{final} = \text{undefined}$).

The spikes along the curves are a phenomenon of the inclinations and the longitudes of the ascending node of the two orbits, as well as combining the long- and short-way transfers. The long- and short-way transfers don't overlap exactly, and even with very small step sizes, small spikes can occur depending on the individual inputs. More often, the spikes occur because the transfer orbit contains part of a plane-change maneuver (Sec. 5.4.2). Plane changes result whenever the inclinations of the two orbits or the longitudes of the ascending node are different. The spikes can be very large in cases of different inclinations and may occur throughout several time-of-flight values. In cases where the nodal values differ, the spikes are sometimes smaller, and they often occur as Δv values near 170° to 180° , although this depends on the initial locations of the orbital planes. I've used orbits which illustrate some of these phenomena. You should be careful when analyzing orbital transfers to correctly determine the magnitude of these effects.

The process still isn't complete because, to find the global minimum velocity, we must move both the target and the interceptor from the epoch time. Most of the literature doesn't address this problem at all because locating the global minimum requires a lot of computation. Figure 6-23 shows results for the minimum change in velocities as the target and the interceptor move. The effect is to create a topography of change in velocities for the transfer. Notice how similar this surface (of total Δv) is to the simple, 2-D version shown in Fig. 6-22. For clarity, Fig. 6-23 doesn't show the interim Δv_a . The delay time is the length of time from the epoch before the maneuver starts.

A question arises here: "How do we know when to fire?" The graph *looks* as though each change in velocity is the same with respect to the previous line. Indeed, for Fig. 6-23, the two satellites are in sufficiently close orbits so the change in velocity curve is almost identical for different delay times before starting the transfer. The data for this case is the same as for Fig. 6-20 and Fig. 6-22. If we were to look at only a two-dimensional (2-D) plot (frontal view of the topography) of this data, the curves are almost indistinguishable from the single case of Fig. 6-22.

The preceding figures show that the initial separation between the satellites must be *relatively* large for us to see any differences in the change in velocity graphs. If the particular mission results in satellites being relatively close together, the analysis is simpler because we can examine it in 2-D space only. Timing requirements must be met and become especially important if the analysis is 2-D because we must *add* the interceptor's wait time to the transfer time in order to determine when the rendezvous will be complete. Finally, if the orbits are relatively close to each other, it's unlikely the intercept or rendezvous trajectory would hit the Earth, except in cases involving very short times of flight and high changes in velocity.

Now we can examine specific test cases to determine how sensitive the shape of the graph is to varying input conditions. Figure 6-24 shows a case with "minimum" separation. Initial vectors are

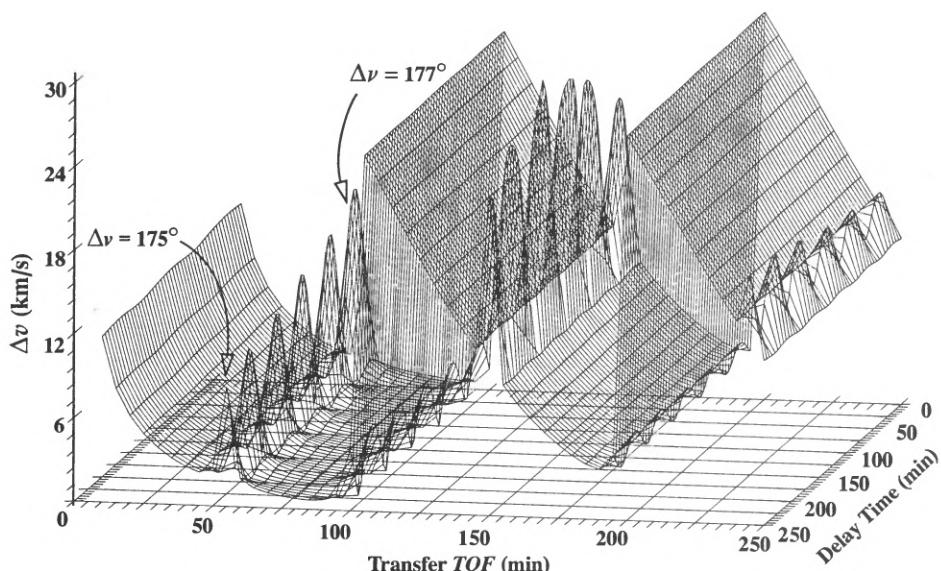


Figure 6-23. Location of Global Minimum. Determining the global minimum is difficult when a delay time is permitted. Even when the curve is rather well-behaved, the three-dimensional graph has too much information for accurate determinations. Notice how the inclination and node differences create large spikes near the $\Delta\nu = 180^\circ$ region.

“MINIMUM” TEST CASE

$$\text{INTERCEPTOR} \quad \dot{r}_{IJK} = 4965.2226 \hat{i} - 1504.1795 \hat{j} + 4617.2886 \hat{k} \text{ km}$$

$$\dot{v}_{IJK} = 3.760\,556 \hat{i} - 3.882\,606 \hat{j} - 5.310\,687 \hat{k} \text{ km/s}$$

$$\text{TARGET} \quad \dot{r}_{IJK} = 5055.2385 \hat{i} - 1599.4069 \hat{j} + 4483.9878 \hat{k} \text{ km}$$

$$\dot{v}_{IJK} = 3.612\,200 \hat{i} - 3.836\,100 \hat{j} - 5.443\,000 \hat{k} \text{ km/s}$$

Notice the similarity between the subsequent curves. The minimum Δv was 26.8 m/s and occurred after the interceptor waited for 0 minutes; the time of flight for the transfer took 90 minutes. Because it's difficult to see variations on a large scale, I've enlarged the first section of Fig. 6-24 to produce Fig. 6-25.

Notice how the graph appears smooth, even with the enlarged scale. We don't need to see other parts of the graph because they all result in higher change in velocities, as discussed earlier.

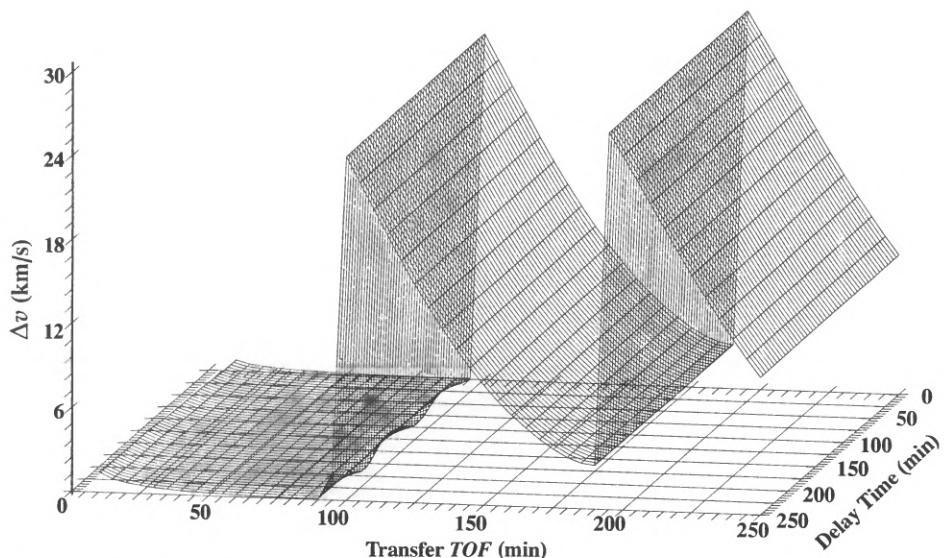


Figure 6-24. Location of Global Minimum (Min Case). For this case, determining the global minimum appears to be simple. But don't forget that a change in velocity of 30 km/s is unrealistic! In this case, the inclinations and nodes are very similar; therefore, we don't see any spikes.

A “maximum” separation results in a different topography (Fig. 6-26). Again, for clarity, I've shown only the first section of this topography. Initial vectors were changed as follows:

“MAXIMUM” TEST CASE

$$\text{INTERCEPTOR } \dot{\vec{r}}_{IJK} = -6175.1034 \hat{i} + 2757.0706 \hat{j} + 1626.6556 \hat{k} \text{ km}$$

$$\dot{\vec{v}}_{IJK} = 2.376\,641 \hat{i} + 1.139\,677 \hat{j} + 7.078\,097 \hat{k} \text{ km/s}$$

$$\text{TARGET } \dot{\vec{r}}_{IJK} = -6078.007289 \hat{i} + 2796.641859 \hat{j} + 1890.7135 \hat{k} \text{ km}$$

$$\dot{\vec{v}}_{IJK} = 2.654\,700 \hat{i} + 1.018\,600 \hat{j} + 7.015\,400 \hat{k} \text{ km/s}$$

The delay time now looms larger in determining the minimum change of velocity. Notice the large variation from the previous figure as the interceptor's *TOF* changes. Also, note that the minimum change in velocity still occurs within one period of the target satellite from the start. The minimum Δv was 18.1 m/s and occurred after the interceptor waited for 95 minutes; the *TOF* for the transfer was 85 minutes.

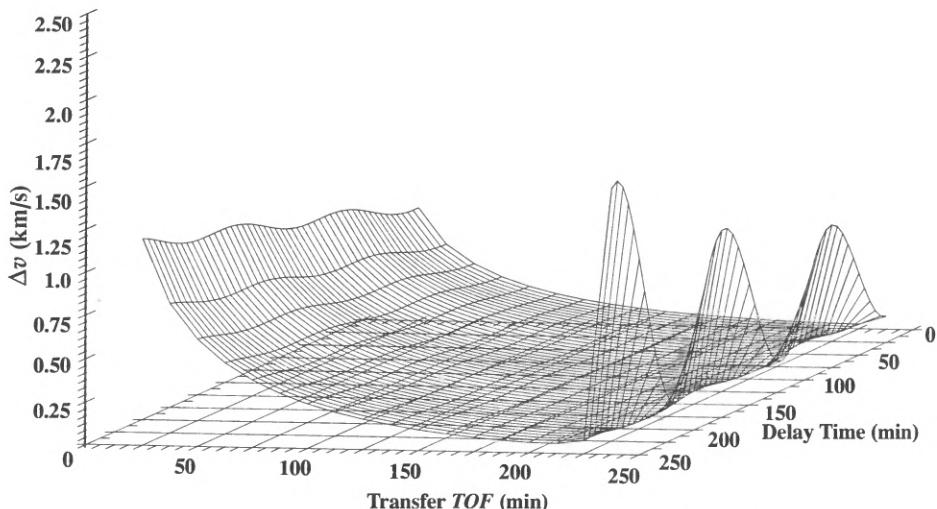


Figure 6-25. Location of Global Minimum (Min Case). With a larger scale, determining the global minimum is simpler, but notice the right-hand edge of the graph. Recall how quickly the curves rose to 30 km/s.

The tremendous change from the previous figure is another reason to use a universal approach to solve this problem. Determining the overall minimum is *very* difficult.

We could run more test cases to examine different initial state vectors for the epoch; however, the problem appears to be very sensitive to these initial parameters. For coplanar transfers using the given epoch-state vectors, the solutions don't appear to vary much from the examples shown. Developing the initial vectors *could* result in large variances. Specifically, the timing and location of the burns can change the shape of the topography. If all burns occur at perigee, the largest amount of specific mechanical energy will transfer to the new orbit. The interceptor satellite will still touch the target orbit at one point, and depending on the timing, this could provide a point for the minimum change in velocity. If all the burns occur at apogee of the current orbit, the interceptor's final orbit *won't* have any points coincident with the target orbit, and the resulting change in velocities is usually higher than those for the burns occurring at perigee. Finally, the initial displacement between the two orbits appears to affect the change in velocity. When the satellites are very close together, the change in velocity seems to be smaller than when the two satellites begin a little farther apart. A rough guide to use for this determination is about 1° to 2° of initial separation. The explanation lies in the earlier analogy to the Hohmann-like transfer; where the satellites are less than a few degrees away from each other, the timing opportunities offer a greater chance of coming close to a Hohmann-like (180°) transfer, and therefore, to a minimum change in velocity. When the satellites are farther apart, the timing requirements make this Hohmann-like transfer difficult.

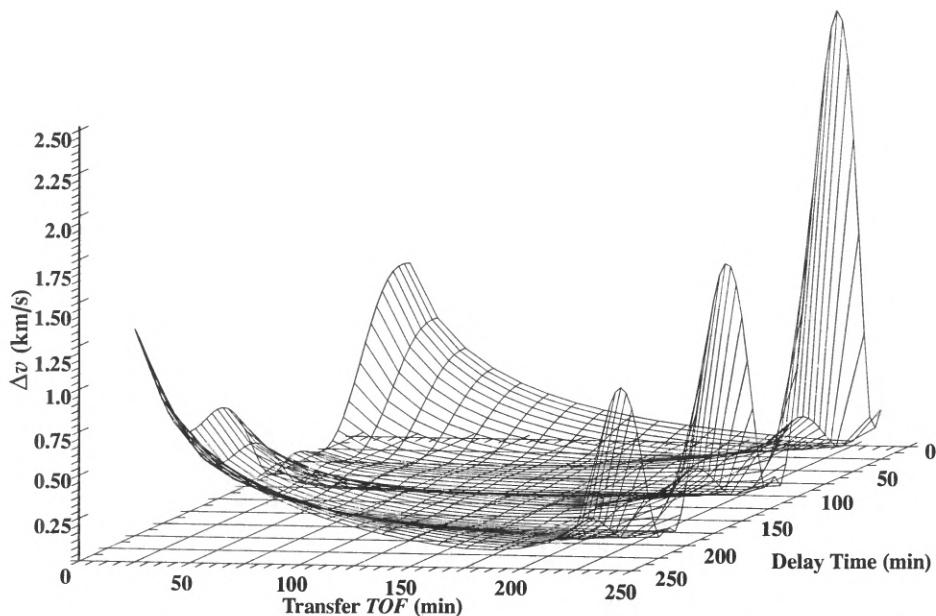


Figure 6-26. Location of Global Minimum (Max Case). Even with a larger scale, the determination of the global minimum is difficult for this “maximum” case. Again notice the right-hand edge of the graph. Recall how quickly the curve rose to 30 km/s. The inclinations and nodes are equal in this example.

The best strategy to rendezvous with the satellite’s original time *and* position is to have the transfer take place as soon as possible, preferably within one period of the target satellite. Many cases probably allow for the transfer at future times using complicated phasing orbits and several revolutions. However, this approach increases the risk and chance for error at each maneuver. The brute-force approach is to do the transfer quickly and with a minimum number of burns.

Problems

1. Gauss tried to locate Ceres with the following data (Montenbruck (1994, 245)):

Sept. 5, 1805 24:16:5 $\alpha = 6^\circ 23' 57.54''$ $\delta = 22^\circ 21' 27.08''$

Jan. 17, 1806 22:09:5 $\alpha = 6^\circ 45' 14.69''$ $\delta = 30^\circ 21' 24.20''$

May 23, 1806 20:39:9 $\alpha = 6^\circ 07' 44.60''$ $\delta = 28^\circ 02' 47.04''$

What are the orbital elements? (Hint: use Laplace's technique)

2. You're trying to get a highly accurate, circular orbit (about 1 m radial position) at 1300 km altitude. If you use an angles-only technique that gives *perfect* results, what combination of frequency (electromagnetic radiation) and apertures should you select to achieve your desired accuracy?
3. Use an angles-only technique to determine the orbit of the following satellite in Earth orbit beginning on May 14, 1996:

h	min	s	α_t (°)	δ_t (°)	ϕ_{rd} (°)	λ (°)	h_{ellp} (km)
23	30	0.00	59.809 765 1	5.785 077 3	-34.618 645	31.054 427	903.899 838
23	40	0.00	61.442 593 3	0.963 851 6	0.000 000	22.016 802	895.718 864
23	50	0.00	62.602 131 9	-3.791 553 3	34.618 645	12.979 177	903.899 838
24	10	0.00	63.838 841 0	-7.293 504 6	68.073 414	-7.157 751	918.883 650
24	30	0.00	65.745 694 8	-8.219 761 7	72.521 101	-133.955 230	924.309 096
24	50	0.00	68.801 384 8	-5.353 928 7	39.708 368	-160.146 385	917.248 548
0	10	0.00	72.857 869 6	1.049 054 9	5.280 043	-169.644 991	910.429 439
0	30	0.00	76.849 032 0	7.842 378 9	-29.252 549	-178.318 964	914.409 694
0	50	0.00	79.699 182 4	11.494 804 5	-63.001 471	165.601 569	923.087 600
1	10	0.00	81.299 487 6	11.250 311 9	-76.834 806	46.741 203	922.076 632
1	30	0.00	82.225 627 7	8.163 782 0	-45.063 420	8.959 525	908.584 733
1	50	0.00	83.148 682 5	3.554 594 9	-10.600 064	-1.284 745	896.563 207
2	10	0.00	84.652 559 4	-1.379 516 9	24.094 841	-9.783 934	899.905 334
2	30	0.00	87.261 216 4	-5.439 902 6	58.178 493	-23.462 257	914.684 420
2	50	0.00	91.367 508 1	-7.253 447 9	79.740 979	-125.023 999	924.234 003

What's unusual about this data? Should you modify the techniques at all? If so, how?

4. You receive two angular measurements of azimuth and elevation, plus one measurement of range and range-rate. Can you estimate the orbit? If so, how?
5. Use the following data to determine the position of a newly discovered asteroid. The data is from AMOS during a preliminary collection on asteroids. Be careful of the units because this problem is solved with AU's. Are parallax and reduction techniques necessary? Hint: see Table 9-1 for the

location of AMOS, and be sure to use the transformations in Chap. 3 to preprocess the data.

Year	Month	Days (UTC)	α_t (°)	δ_t (°)
1995	March	17.435 38	11 00 57.79	5 31 19.7
1995	March	17.465 68	11 00 55.91	5 31 29.0
1995	March	23.421 39	10 55 44.79	6 00 32.7
1995	March	23.451 91	10 55 43.19	6 00 41.0
1995	March	29.358 81	10 51 16.34	6 25 27.7
1995	March	29.422 50	10 51 13.57	6 25 42.2
1995	March	29.457 95	10 51 12.03	6 25 50.1

6. You're in charge of a large orbit-determination system that's been operating for many years. Your primary sensors are old and operate at 4 GHz (C-band with 10 m dishes). A new sensor is planned for next year but is 24 GHz (Ka-band) with the same size sensor dish. Funding difficulties prevent the sensor from being finished. Will your expected accuracy improve more than 50% so you can present a case to complete the sensor?
7. Program Example 5-10 and explore options to decrease the change in velocity. How does your result compare to the solution of Example 5-10?
8. Escobal ([1968] 1979, 81–90) presents a technique to determine the minimum values of time and change in velocity for the targeting problem. Although the solution involves solving a quartic (see page 865), program the technique and compare your results to the examples in the application section (Sec. 6.8).
9. Use Table 9-5 for the first, middle, and last observations of a satellite. What is its orbit? Compare to an answer using only the first three observations. Explain any differences.
10. For the solution of Lambert's problem using universal variables (Algorithm 52), we know from Fig. 4-5 that c_2 can equal zero. Is this a problem we must be careful of for computer implementations?
11. We know that Algorithm 52 can accommodate multiple revolution transfers by adjusting the iteration parameters. Does anything else need to be fixed? Does the same process apply to Algorithm 53?
12. Investigate the claim that the Gaussian angles-only method works best for observations that are less than about 10° apart. Use the data in Example 6-2 with points 6, 8, 10 and 6, 7, 8 to test your hypothesis.
13. Problem 6 in Chap. 5 explains the timing sequence to complete a rescue of the Space Shuttle astronauts. Imagine you're in a company that wants to design a rescue satellite that could assist both MIR and the Space Shuttle. Use intercept and rendezvous techniques with the Lambert solutions to determine minimum Δv and minimum time to "service" both orbits. What

orbit would you recommend? Why? (Hint: Table 8-4 has relevant orbital data for the Space Shuttle and Mir.)

CHAPTER 7 SPECIAL PERTURBATION TECHNIQUES

- 7.1 Historical Background
- 7.2 Introduction to Perturbations
- 7.3 Encke's Formulation
- 7.4 Cowell's Formulation
- 7.5 Numerical Integration Methods
- 7.6 Disturbing Forces
- 7.7 Forming Numerical Solutions
- 7.8 Practical Considerations

7.1 Historical Background

The fundamental problem in perturbation analysis is orbit propagation. Unlike the two-body analysis of Algorithm 30, the most accurate way to analyze perturbations is numerical. We often discuss the sheer numerical complexity of these processes, but to the early pioneers in celestial mechanics, the complexity often signaled a dead end. Of particular importance to the early scientists was their limited ability to calculate large numbers. Indeed, calculations weren't reasonable until John Napier (1550–1617) invented logarithms in 1614. Incidentally, Napier (and later Jorst Borgi) worked on logarithms to help solve problems in astronomy.

As techniques for larger numbers grew, so did the complexity of the theories. It quickly became apparent that the complexity of these theories could easily outpace advances in computing technology. The problem then became how to efficiently evaluate complex theories for computational applications, which even today calls for accurate, reliable interpolation schemes.

We'll see shortly that solution techniques for the perturbation problem fall into three broad categories: analytical, numerical, and semianalytical. The analytical approaches were generally developed first because the other methods required computational horsepower that didn't exist at the time. (We'll discuss analytical techniques in Chap. 8.) But today the modern computer has erased this limitation and opened a whole new era in analyzing perturbations.

The end of the 18th century brought significant developments in perturbation theory particularly in the modeling of the Earth's gravitational field. Laplace discovered a very useful analysis tool—the potential function. However, Laplace's most enduring legacy may be the phrase “it is easy to see that...” (Bell, 1986, 177). He didn't care to do detailed

derivations when the results were obvious (after the fact). Four volumes of his monumental work in celestial dynamics, *Méchanique céleste* (Celestial Mechanics) were published in 1799 and 1802–1805. The fifth and final volume appeared from 1823 to 1825. It contained scant derivations but included his significant contribution of the potential function. We'll explore several potential functions in this chapter.

Adrian Marie Legendre (1752–1833), a French mathematician and professor, did his main work in elliptic functions and number theory. The Legendre functions are solutions of the differential equations arising from his studies of the attraction of spheroids (Boyce and DiPrima, 1977, 83). He published his discovery in 1783 in the first of four memoirs on spheroids. They're an integral part of solving the gravitational-potential problem.

In 1849, Sir George Gabriel Stokes (1819–1903), a professor at the University of Cambridge, published an extremely important formula which allowed accurate modeling of the Earth's shape (Caputo, 1967, 62). Essentially, Stokes found that we can determine the shape of a geoid if we know the surrounding gravity anomalies. This discovery permitted practical evaluation of the Earth's gravity field.

Johann Franz Encke (1791–1865) presented a formulation for computing orbits whenever the perturbations were small with respect to the two-body motion. This method was very popular for decades. He numerically integrated the differences of the osculating orbit from a reference orbit. By integrating the small differences between the orbits, rather than the complete attraction and specific forces involved, he got enough accuracy with limited computing abilities. I'll only briefly introduce this method because computing technology has overcome the need to use such approximations.

Numerical techniques didn't receive much attention until the 20th century because computing ability was limited. The *concept* of numerical integration was well understood, but it didn't receive much attention in astrodynamics until Phillip Herbert Cowell (1870–1949) used a numerical technique to determine the orbit of Jupiter's eighth satellite. Cowell and Crommelin also used a numerical procedure to predict the return of Halley's comet between 1759 and 1910. Their results were published in the "Appendix" to the *Greenwich Observations* 1909 as an "Investigation of the Motion of Halley's Comet." This method has been "rediscovered" many times and continues to gain favor in technical circles as computers become faster and more powerful. In modern times, it's common in astrodynamics to use *Cowell's formulation* to set up the equations of motion for numerical integration.

These are just a few examples of *perturbation analysis*—the study of the effects of disturbing forces on the underlying (two-body) motion. Today, the study of perturbations sometimes includes the effects of small *desired* changes to the orbit's motion, such as thrust. We must incorporate perturbation effects to solve real-world problems of orbital motion. This chapter begins our transition from the idealized theoretical foundations discussed so far to real-world problems. This chapter explores numerical solutions; Chap. 8 discusses analytical techniques.

7.2 Introduction to Perturbations

Studying and modeling perturbations are extremely important to astrodynamics. Although most of the problems have been known for a long time, the literature only recently has begun to present (with increasing exactness) the precise methods needed to satisfactorily study these problems. Leon Blitzer (1970) developed the *Handbook of Orbital Perturbations* and has graciously allowed me to use it as a basis for this chapter and Chap. 8. Blitzer was one of the first scientists to begin detailed analysis of the perturbative effects on satellites near Earth, publishing several papers in the mid-1950s before satellites had become routine. I've also incorporated personal discussions with (listed alphabetically) Cefola, Fonte, Gabor, Junkins, and McClain to support this section. A complete list of their works would take many pages! Szebehely (1989, 13–22) lists other major contributors to the field of perturbations and astrodynamics. Lambeck (1988, 654–709) includes extensive references on perturbations. Consult these sources for more detail.

Perturbations are deviations from a normal, idealized, or unperturbed motion. We tend to view the universe as highly regular and predictable. Yet accurate observational data often reveals unexplained irregularities of motion superimposed on the more regular average or mean motions of the celestial bodies. Real-world problems don't conform exactly to the assumptions made for the ideal two-body and restricted three-body problems. Therefore, these idealized solutions only approximate the observed motion. The actual motion will vary from the theoretical two-body path due to perturbations caused by other bodies (such as the Sun and Moon) and additional forces not considered in Keplerian motion (such as a nonspherical central body and drag). Notice we use the term *central body* to show these perturbation techniques apply about *any* central body. For missions to other planets, the Earth may not be the central body. Indeed, the formulas work equally well for the Earth, Mars, Saturn, and so on—assuming, of course, that we have the data constants and coefficients for each central body.

We're very familiar with perturbations in almost every area of life. Seldom does anything go exactly as planned; rather, it is perturbed by unpredictable circumstances. An excellent example is an aircraft encountering a wind gust. All directional controls and power are kept constant; yet, the path changes abruptly from the predicted course based on previous measurements. In this example, the wind gust is the perturbation. Fortunately, most of the perturbations considered in orbital mechanics are much more predictable and analytically tractable than a wind gust. The usual perturbations for the orbit problem are the asphericity of the central body, atmospheric drag and lift, the presence of other attracting bodies, effects of solar-radiation pressure, thrust, magnetic fields, solid-Earth tides, ocean tides, Earth re-radiation, and relativistic effects. I'll show how to analyze many of these later.

Don't get the idea that perturbations are always small—they can be comparable to the primary attracting force (two-body gravitation). Whenever they're this large, they cease to be perturbations and change the nature of the problem. Examples are atmospheric drag during reentry and the third body in the restricted three-body problem. In fact, many interplanetary missions would miss their target entirely if the perturbing effect of other

attracting bodies weren't taken into account! Ignoring the effects of the central body's oblateness on any satellite keeps us from accurately predicting its position over a long time. Without analyzing perturbations, we couldn't explain or accurately predict the orbit of the Moon about the Earth.

We'll look at three main approaches to examine the effect of perturbations: *special perturbation techniques* (using numerical methods—Sec. 7.5), *general perturbation techniques* (using analytical methods in Chap. 8), and *semianalytical techniques* (Chap. 8). **Special perturbation techniques** numerically integrate the equations of motion including all necessary perturbing accelerations. Because numerical integration is involved, we can think of numerical formulations as producing a *specific*, or *special*, answer that is valid only for the given data (initial conditions and force-model parameters). Although numerical methods usually give very accurate results and often establish the "truth" in analyses, they suffer from their specificity, which keeps us from using them in a different problem. Thus, new data means new integration, which can add hours and days of computing times. NASA began the first complex numerical integrations during the late 1960s and early 1970s. Personal computers are now able to compute fast enough to start doing complex perturbation analyses through numerical techniques. However, numerical methods suffer from errors that build up with round-off and truncation due to fixed computer wordlength, which causes numerical solutions to degrade as the propagation interval lengthens.

Let's look at an example to illustrate the strengths and weaknesses of a numerical method. Start with the equation of motion in Example 7-1, which is simple, first order, and defined with a given initial condition. Find the value of this system where time, $t = 0.5$ s, and $\omega = 1$ rad/s. Obviously, this example permits a direct solution if we recognize the integral as $(\text{SIN}(t))$.

▼ Example 7-1: Using Direct Integration.

GIVEN: $\dot{y}(t) = \cos(\omega t)$

$$y(t = 0) = y_0 = 0$$

FIND: $y(t = 0.5)$

Remembering to evaluate the constant of integration, which is zero here, use direct integration:

$$y(t = 0.5) = \int_0^{0.5} \cos(t) dt = \text{SIN}(0.5) - 0$$

▲ Be sure to use radians so that $y(0.5) = 0.479\ 425\ 538\ 604\ 203$.

Unfortunately, most systems don't have such simple equations of motion, and often the integrals are very complex, as with the orbit problem.

Now let's use a numerical technique to solve the problem. A Taylor series is the underlying principle. For simplicity, use an Euler integrator (first-order derivative of the Taylor series) to propagate the system from $t = 0$ to $t = 0.5$ seconds. The general form of a Taylor series expansion is

$$y(t) = y(t_o) + \dot{y}(t_o)(t - t_o) + \frac{\ddot{y}(t_o)(t - t_o)^2}{2!} + \frac{\ddot{y}(t_o)(t - t_o)^3}{3!} + \dots$$

where the step size ($\Delta t = t - t_o$) is the difference in time. If we retain only the first-order term, the expansion becomes an Euler integrator and allows us to solve the problem. Remember that Δt must be small to permit the removal of higher-order terms.

▼ Example 7-2. Using a Numerical Technique.

GIVEN: $\dot{y}(t) = \cos(\omega t)$

$$y(t=0) = y_o = 0$$

FIND: $y(t = 0.5)$

For each trial, $y(t) = y(t_o) + \dot{y}(t_o) \Delta t$, where Δt is the step size ($t - t_o$) and $\dot{y} = \cos(t)$. Increment the time as $t = t_o + \Delta t$. Begin by assuming a step size of $\Delta t = 0.5$.

Let $\Delta t = 0.5$ and solve (to first order):

$$y(t = 0.5) = y(t_o) + \cos(t_o)(t - t_o) = 0.0 + \cos(0.0)(0.5 - 0) = 0.5$$

Notice how far off from the correct answer this result is. Now try a step size of 0.1 (Let $\Delta t = 0.1$).

$$y(t_{k+1} = 0.1) = y(t_k) + \cos(t_k)(t_{k+1} - t_k) = 0.0 + \cos(0.0)(0.1 - 0) = 0.1$$

$$t_k = t_{k+1}$$

$$y(t_{k+1} = 0.2) = y(t_k) + \cos(t_k)(t_{k+1} - t_k) = 0.1 + \cos(0.1)(0.2 - 0.1) = 0.199\,500\,4$$

$$t_k = t_{k+1}$$

$$y(t_{k+1} = 0.3) = y(t_k) + \cos(t_k)(t_{k+1} - t_k) = 0.199\,500\,4 + \cos(0.2)(0.3 - 0.2) = 0.297\,507\,1$$

$$t_k = t_{k+1}$$

$$y(t_{k+1} = 0.4) = y(t_k) + \cos(t_k)(t_{k+1} - t_k) = 0.297\,507\,1 + \cos(0.3)(0.4 - 0.3) = 0.393\,040\,7$$

$$t_k = t_{k+1}$$

$$\blacktriangleleft y(t_{k+1} = 0.5) = y(t_k) + \cos(t_k)(t_{k+1} - t_k) = 0.393\,040\,7 + \cos(0.4)(0.5 - 0.4) = \underline{\underline{0.485\,146\,8}}$$

Notice several important features. First, each subsequent equation uses the *previous* value for the function (0.0, 0.1, 0.199 500 4). This stepwise process is a basic characteristic of all numerical techniques and is the reason the results are valid for only one set of initial conditions. If we change the original differential equation, we must recompute the entire process. In fact, let's suppose we want to determine the value of the function at $t = 0.48$. We could use interpolation, which is sometimes as accurate as the numerical integrator, or we could start at $t = 0.0$ and proceed with a new step size to determine the final answer. Notice how the step size is constant throughout this example. Step-size changes could easily be input, and the example of changing to determine the result at $t = 0.48$ wouldn't be difficult. But some integrators used with the orbit problem employ fixed step sizes. Certain integrators can overcome this difficulty, but many implementations prevent an easy change. Note the improvement with the smaller step size. The obvious conclusion from these two cases is that we should use a step size of, say, 0.000 000 01, to obtain an answer of 0.479 425 539 216 302, right? Well, maybe. This choice is more accurate than

the analytical method (see Sec. 8.2), but it uses too much computational time. At some point, the roundoff (and truncation) errors between operations become larger than the error from the choice of step size. The problem of determining an appropriate step size can be very real and is a challenge in any numerical process.

Besides selecting an integration technique, we must also consider the forces acting on the satellite. The forces causing the perturbative effects on the satellite are either *conservative* or *nonconservative*. The total energy (kinetic and potential) for conservative-force systems is constant, whereas nonconservative systems may lose or gain energy, mainly through friction or external sources. Central-body and third-body gravitational effects are examples of conservative forces; solar-radiation pressure, thrust, and drag are examples of nonconservative forces. Next, we'll see that we can determine accelerations for conservative forces as the gradient of a potential function.

It's important to know about *gradients*, *accelerations (specific forces)*, and *functions*. A *gradient* is really a directional derivative which gives the rate of change of a *scalar function* in a particular direction (Kreyszig, 1983, 387). It's a vector quantity and the *del operator*, ∇ , designates the gradient process. The gradient gives an acceleration if the scalar function is a potential function related to potential energy, such as the potential function of a central body's gravity field. I distinguish a potential function as the negative of the potential energy (see Sec. 2.2.3). Two conventions are "standard" in this area because many schools of thought have evolved over the last few decades. Brouwer and Clemence (1961), Battin (1987), Long et al. (1989), and others express one of the two main approaches, in which the acceleration is the negative gradient of the potential function. This implies that positive work is done as the potential decreases. The other approach, used mainly by the geophysical community, holds an acceleration to be the positive gradient of the potential function [Lambeck (1988), Kaula (1966), Moritz and Mueller (1987), Kaplan (1976), Roy (1988), and others]. Of course, both methods use potential functions that differ only by a minus sign; therefore, the results are identical! We'll follow the second method and place the sign change between the potential energy and the potential function. I'll also refer to the potential function as simply the *potential* but will always say *potential energy* to avoid confusion.

The distinction between a specific force (used interchangeably with acceleration in this chapter) and a potential is important because analysis of perturbations typically uses both concepts. It's common to analyze perturbations using a *disturbing function* and a *disturbing force*. The *disturbing force* simply expresses (in some coordinate system) the specific force (acceleration) that is perturbing the satellite's orbit. Nonconservative forces, such as the perturbing effects of drag and solar-radiation pressure, are usually modeled as a specific force. *Disturbing functions* are simply the difference between perturbed and unperturbed potential functions. They model conservative forces that perturb the orbit, such as the central body's nonsphericity and third-body attractions.

A *potential function* is one way to mathematically characterize a conservative force, such as the gravitational potential of a *spherical* central body ($U_{2\text{-body}} = \mu/r$). Some people distinguish a disturbing function from a disturbing potential by a minus sign. As mentioned earlier, considering the two to be equal is just as correct, as long as we maintain the

correct sign convention. We'll soon develop a potential function for an *aspherical* central body, U (sometimes referred to as the *anomalous potential*). As you may have guessed, the form is much more complicated than the spherical one. It includes the spherical potential ($U_{2\text{-body}}$) as the first term. The term *geopotential* is often used for this aspherical potential when the central body is the Earth. Using the prefixes helps eliminate any possibility of confusion.

The existence of the spherical and aspherical potentials enables us to form a ***disturbing-potential function***, R , defined as the difference between the two potentials ($R = U - U_{2\text{-body}}$). The disturbing-potential function represents *only* the perturbing effect of the nonspherical Earth (in this case), so we must carefully distinguish it when working problems. The acceleration from the spherical central body's potential, $\nabla U_{2\text{-body}}$, yields the *two-body* acceleration; the gradient of the aspherical central body's potential, ∇U , yields the *complete* acceleration. The acceleration from the disturbing function, ∇R , will yield *only* the acceleration of the perturbing effect, which will equal the difference between the two-body and complete accelerations. We follow this order in defining R because it leaves the disturbing part of the potential as a positive quantity. The accelerations from the spherical potentials and the disturbing potential may differ by orders of magnitude, so be careful!

An example will help us understand the gradient vector operator as it applies to a scalar potential function. Examine the potential of the spherical central body identified previously. The del operator acts by taking the partial derivatives in each of the respective axes. In cartesian coordinates,

$$U_{2\text{-body}} = \frac{\mu}{r}$$

$$\nabla U_{2\text{-body}} = \frac{\partial U_{2\text{-body}}}{\partial x} \hat{I} + \frac{\partial U_{2\text{-body}}}{\partial y} \hat{J} + \frac{\partial U_{2\text{-body}}}{\partial z} \hat{K}$$

where IJK denote unit vectors along the axes of the coordinate system and xyz are the components of the position vector. With a position vector $\vec{r} = x\hat{I} + y\hat{J} + z\hat{K}$ having a magnitude of $r = \sqrt{x^2 + y^2 + z^2}$, the acceleration is

$$\nabla U_{2\text{-body}} = -\frac{1}{2} \frac{\mu}{(x^2 + y^2 + z^2)^{3/2}} [2x\hat{I} + 2y\hat{J} + 2z\hat{K}] = -\frac{\mu}{r^3} \vec{r}$$

Notice the result, as expected, is the two-body acceleration for a satellite. Thus we can write the two-body equation (Eq. (2-4)) very compactly:

$$\ddot{\vec{r}}_{2\text{-body}} = \nabla U_{2\text{-body}} \quad (7-1)$$

If we add conservative perturbative forces to the simple Newtonian attraction, the equation of motion becomes

$$\ddot{\vec{r}} - \nabla U_{2\text{-body}} = \nabla R \quad \text{or} \quad \ddot{\vec{r}} = \nabla U \quad (7-2)$$

The form of the disturbing-potential function, R , will depend on the particular type of perturbing source (aspherical central body, third-body attractions). If there are multiple perturbing forces, the disturbing functions due to each are added to give the total disturbing function, R .

7.3 Encke's Formulation

As mentioned in the introduction, Encke's method isn't very popular today because computers permit a complete numerical solution to many problems. Still, it's important to introduce the method because it is historically important and it provides a computational advantage. Refer to Plummer (1918, 222–224), as well as Brouwer and Clemence (1961, 176–185), for classical presentations and details on the formulation of Encke's method. Kaplan (1976, 345–348) presents a concise algorithm.

Encke's method begins with an osculating orbit. Instead of integrating all the forces on the satellite, he integrates just the difference between the two-body acceleration and the perturbed acceleration. The perturbations to the orbit are integrated in cartesian elements. Because this technique integrates only the perturbations, the magnitudes are much smaller and the computational requirements can be much less, especially for interplanetary trajectories. Figure 7-1 shows the geometry.

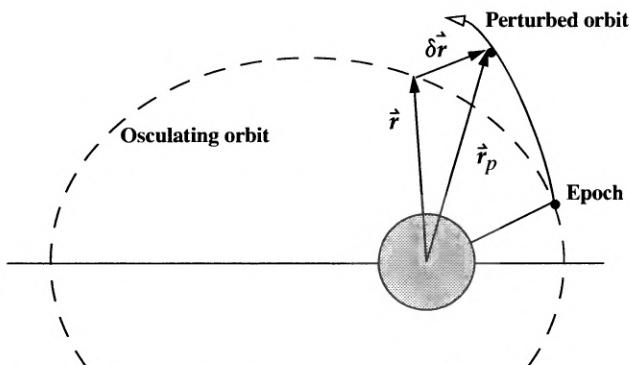


Figure 7-1. Encke's Method. The main principle for Encke's method is the difference between the osculating (two-body) orbit and the actual perturbed motion, shown as $\delta\vec{r}$. By integrating this difference, the method can save many computations.

We know the two-body and perturbed accelerations:

$$\ddot{\vec{r}} = -\frac{\mu}{r^3}\vec{r} \quad \ddot{\vec{r}}_p = -\frac{\mu}{r_p^3}\vec{r}_p + \vec{a}_p$$

This means the difference between the two accelerations is

$$\ddot{\delta r} = \dot{\vec{a}}_p + \frac{\mu}{r^3} \left\{ \left(1 - \frac{r^3}{r_p^3} \right) \dot{\vec{r}}_p - \dot{\delta r} \right\}$$

Note the appearance of the difference of two nearly equal quantities. Kaplan (1976, 347–348) shows how we can treat this parameter and form a solution using Algorithm 56, below. Notice we use two-body relations to move the osculating orbit forward in time.

ALGORITHM 56: *Encke* ($\dot{\vec{r}}_o, \dot{\vec{v}}_o, t_{start}, t_{stop}, \Delta t \Rightarrow \dot{\vec{r}}, \dot{\vec{v}}$)

$$\dot{\delta r} = \dot{\vec{0}} \quad \epsilon = \frac{\dot{\vec{r}}_o \cdot \dot{\delta r}}{r_p^2} \quad f = \frac{1}{\epsilon} \left\{ 1 - \frac{1}{(1-2\epsilon)^{3/2}} \right\}$$

LOOP

KEPLER ($\dot{\vec{r}}_o, \dot{\vec{v}}_o, \Delta t \Rightarrow \dot{\vec{r}}, \dot{\vec{v}}$)

$$\epsilon = \frac{\dot{\vec{r}} \cdot \dot{\delta r}}{r_p^2} \quad f = \frac{1}{\epsilon} \left\{ 1 - \frac{1}{(1-2\epsilon)^{3/2}} \right\}$$

$$\ddot{\delta r} = \dot{\vec{a}}_p + \frac{\mu}{r^3} \left\{ f\epsilon \dot{\vec{r}}_p - \dot{\delta r} \right\}$$

$$\left[\begin{array}{l} \text{IF } \frac{\delta r}{r_p} > \text{tolerance} \\ \quad \dot{\vec{r}} = \dot{\vec{r}}_p \quad \dot{\vec{v}} = \dot{\vec{v}}_p \\ \text{ELSE} \\ \quad \dot{\vec{r}}_p = \dot{\vec{r}} + \dot{\delta r} \quad \dot{\vec{v}}_p = \dot{\vec{v}} + \dot{\delta v} \end{array} \right]$$

UNTIL $t > t_{stop}$

7.4 Cowell's Formulation

As mentioned earlier, numerical methods are enjoying ever-increasing popularity over some analytical methods as computers become faster. One of the most attractive characteristics for numerical techniques is that we can incorporate any arbitrary disturbing acceleration. As mentioned in the introduction to this chapter, the two-body assumption neglects the effects of real-world perturbations. One way to account for these effects is to add the perturbing accelerations to the two-body equation in order to produce a more accurate equation of motion, as shown in Eq. (7-3):

$$\ddot{\vec{r}} = -\frac{\mu}{r^3} \vec{r} + \vec{a}_{perturbed} \quad (7-3)$$

Here, $\vec{a}_{perturbed}$ is the total acceleration caused by other forces acting on the satellite, shown as an additive effect to the original equation. The specific form of $\vec{a}_{perturbed}$ will depend on the number and type of perturbing sources. We'll examine perturbing effects of a nonspherical central body, atmospheric drag, third-body effects (Sun, Moon, and planets), solar-radiation pressure, thrust, and smaller effects of tides. We won't explore other perturbing forces such as plate tectonic motion and geomagnetism. Fortunately, the form of Eq. (7-3) permits us to add each effect linearly. Eq. (7-3) models all complex interactions *among* the effects but leaves us with several difficulties. For example, the particular expressions for a perturbing acceleration can be very complex and computationally intensive. Obtaining the necessary data to use with each model may also limit us, such as in determining the gravitational coefficients for Pluto, a planet whose equatorial radius we don't know precisely. Calculating the Earth's atmospheric density is extremely difficult mainly because we don't fully understand the Sun's internal mechanics and the precise interaction of solar radiation with the molecules in the Earth's upper atmosphere. The Earth's geomagnetic field is also complex.

Equation (7-3) has come to be known as *Cowell's formulation*. I distinguish between *Cowell's method* and *Cowell's formulation*. The preceding discussion is based on Cowell's formulation, in which we formulate second-order differential equations of motion and numerically integrate them. *Cowell's method* is a technique using the *calculus of finite differences* to accomplish the integration. It's preserved in many places, including Brower and Clemence (1961, 167–175).

Many techniques for numerical integration have been developed to operate on various first- and second-order differential equations. It's common to reform the three second-order differential equations of motion into six first-order differential equations because a broader class of integration methods usually applies. This first-order system is known as a variation of Cowell's formulation.

$$\bar{\vec{X}} = \begin{bmatrix} r_I \\ r_J \\ r_K \\ v_I \\ v_J \\ v_K \end{bmatrix} \quad \dot{\bar{\vec{X}}} = \begin{bmatrix} v_I \\ v_J \\ v_K \\ -\frac{\mu r_I}{r^3} + a_{p_I} \\ -\frac{\mu r_J}{r^3} + a_{p_J} \\ -\frac{\mu r_K}{r^3} - a_{p_K} \end{bmatrix}$$

7.5 Numerical Integration Methods

The equations of motion used in astrodynamics aren't trivial to solve because they're usually coupled systems of equations—first-order or second-order, nonlinear, differential equations which have resisted direct solution over the last 300 years. But modern computers allow us to use numerical techniques. In this section, we'll discuss both *single-step* and *multi-step* methods for astrodynamics problems. Many applications can use the simple single-step methods. Multi-step methods are also used on these types of problems, particularly for highly accurate, longer-term studies. Each method comes in various orders, relating to the order of the Taylor series expansion. A multi-step method is usually more accurate and efficient than a single-step method of the same order, but the problem of assembling back values can be a disadvantage. Sadler (1974) presents important information about numerical integration and notation used in the equations.

Single-step methods combine the state at one time with rates at several other times, based on the single-state value at time, t_o . The rates are readily obtained from the equations of motion and allow us to determine the state at succeeding times, $t_o + h$. For many applications, a fourth-order *Runge-Kutta* (RK4, single-step) is sufficient. Runge-Kutta techniques are often preferable because they don't require a sequence of back values to start the integrator, which can sometimes increase the computer run-time. However, Runge-Kutta methods evaluate the function at several intermediate points, and these values are used only once.

Predictor-corrector methods, also called *multi-step* techniques, can sometimes improve the accuracy of the integration, but require us to maintain a series of back values for use in the algorithm. Consequently, they aren't self-starting, despite their efficiency and accuracy. **Multi-step** methods do an initial estimate (predictor) using previous estimates of the function's rate of change, and a second series of calculations (corrector) use the estimated answer to further refine the result. The Adams-Bashforth-Moulton integrator is widely used in highly accurate programs. The Gauss-Jackson integrator also has a long history of operational use. Roy (1988, 230–237) discusses the advantages of the Gauss-Jackson method and briefly summarizes the required equations. He also suggests that Herrick (1972) arrived at the same conclusion: for near-circular, near-Earth orbits, Gauss-Jackson is about one order of magnitude more efficient than a fourth-order Runge-Kutta. However, for higher eccentricities, thrusting, and high-drag orbits, the Runge-Kutta methods are usually competitive and less complex. The predictor-corrector methods are usually more accurate, but this improvement is achieved at the expense of complexity.

Burlisch-Stoer represents a newer alternative that has even fewer calls to the force-model equations (Pohlen and Titus, 1995). This method reduces computer runtime, but it appears to be unreliable as a general-purpose integrator. It works best in quadruple precision when FORTRAN is used, whereas Adams-Bashforth-Moulton and Gauss-Jackson better propagate roundoff error and are very reliable.

Most numerical integrators actually derive from the simplest form of numerical integration—the *Taylor series* integrator. Brook Taylor (1685–1731) published this theorem in 1715 in *Methodus incrementorum directa et inversa*. Recall the form of a Taylor series:

$$y(t) = y(t_o) + \dot{y}(t_o)(t - t_o) + \frac{\ddot{y}(t_o)(t - t_o)^2}{2!} + \frac{\dddot{y}(t_o)(t - t_o)^3}{3!} + \dots \quad (7-4)$$

The difficulty here is two-fold: how do you include an infinite number of terms (or, if you don't, where do you truncate the series?), and how do you calculate all the derivatives, especially with extremely complex functions? The simplest answer to these questions is the basis for an **Euler integrator** (after Leonhard Euler), which is simply a first-order Taylor series:

$$y(t) \cong y(t_o) + \dot{y}(t_o)(t - t_o) \quad (7-5)$$

Notice you need only the first derivative. The Euler integrator actually works by attempting to determine the slope (first derivative) of the function at the starting point. The subtle part of this scheme is that you must carefully choose the *step size*, defined here as $(t - t_o)$. You can probably picture a situation in which a function changes drastically just after the point you've chosen to determine the slope. By examining terms only through the first order, you've implicitly *assumed* your step-size is small enough to handle variations caused by the neglected higher-order derivatives.

Runge-Kutta Methods

Perhaps the most well-known numerical integrators are the Runge-Kutta methods originally presented by Carl Runge (1856–1927) in 1895, and Wilhelm Kutta (1867–1944) in 1901, which actually derive from a Taylor series. They differ from traditional Taylor series integrators because, instead of having to derive application-specific formulas for the higher derivative terms, we can form the approximation by simply using the slope at different points within the interval over which we'll integrate. The general equations for the classical fourth-order Runge-Kutta are

$$\begin{aligned} \dot{y}_1 &= f(t_o, y_o) & \dot{y}_2 &= f\left(t_o + \frac{h}{2}, y_o + \frac{h}{2}\dot{y}_1\right) \\ \dot{y}_3 &= f\left(t_o + \frac{h}{2}, y_o + \frac{h}{2}\dot{y}_2\right) & \dot{y}_4 &= f(t_o + h, y_o + h\dot{y}_3) \\ y(t) &= y(t_o) + \frac{h}{6}(\dot{y}_1 + 2\dot{y}_2 + 2\dot{y}_3 + \dot{y}_4) + O(h^5) \end{aligned} \quad (7-6)$$

Notice that we need only the first derivative, but we require four evaluations of the equations of motion to move forward one step in time—a characteristic of single-step methods. For the satellite problem, we often formulate the equations of motion as a first-order system with initial conditions. Using position and velocity vectors,

$$f(t, y) = \begin{bmatrix} \dot{r} \\ \dot{v} \end{bmatrix} = \begin{bmatrix} \ddot{r} \\ \ddot{v} \end{bmatrix} \quad y(t_o) = \begin{bmatrix} r_o \\ v_o \end{bmatrix} \quad (7-7)$$

Evaluating the step size is important. In general, we can determine a step size by examining the difference between the two functions and dividing by the number of steps in the evaluation.

The fourth-order terminology comes from the fact that our integrator is matched to a fourth-order Taylor series expansion about the initial value $y(t_0)$. We can relate the higher-order time derivatives to the first derivative at other times. The way the other times are chosen forms the particular integration algorithm. Eq. (7-6) shows the values for a fourth-order Runge-Kutta. This expression is really an Euler-type extrapolation using a weighted average of four rates (\dot{y}_i) in the interval $[t_0, t_0 + h]$. Figure 7-2 illustrates the different evaluations of the slope. It can be shown that the truncation error for a fourth-order Runge-Kutta is $O(h^5)$, which is the same as a fourth-order Taylor series. In general, there are several *orders* (number of intermediate evaluations) of numerical-integration routines.

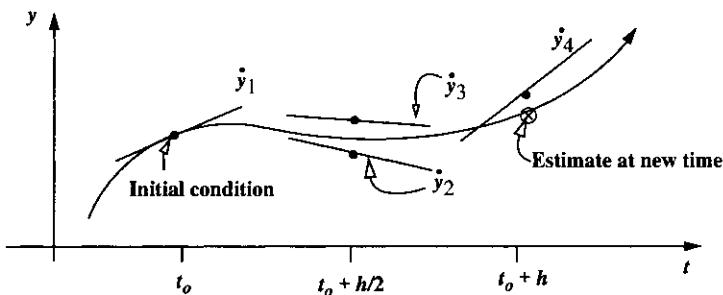


Figure 7-2. The Runge-Kutta Method. This algorithm works by evaluating the derivative (slope) at several different points (four for a fourth-order method) along the estimated trajectory. It's limited because all calculations depend on estimates of the function value. The intermediate estimates aid in finding the final estimate at the new time.

Several other forms derive from this basic method. There are eighth, twelfth, and other order Runge-Kutta methods, which operate as above, but with more terms. Consult Fehlberg (1968 and 1969) and Der (1995) for extensive treatments. They show the general iterative Runge-Kutta formulas, as well as an example six-stage, fifth-order, RK56, evaluation (i -order, j -evaluations per stage).

$$y(t) = y_0 + h \sum_{i=0}^{j-1} b_i \dot{y}_i + O(h^{n+1}) \quad n = 0, 1, 2, \dots, n-1$$

$$\dot{y}_o = f(t_n, y_n) \quad c_{io} = p_i - \sum_{j=1}^{i-1} c_{ij}$$

$$\dot{y}_i = f\left(t_n + p_i h, y_n + h \sum_{j=0}^{i-1} c_{ij} \dot{y}_j\right) \quad i = 1, 2, \dots, j-1$$

The constants b_i , c_{ij} , and p_i must be determined separately. The actual process is lengthy and includes some user-chosen variables: $p_1, p_2, p_5 = 1$, $b_1 = 0$.

The *Runge-Kutta-Fehlberg* method, sometimes called an *embedded Runge Kutta*, uses a variable step size [Fehlberg (1968 and 1969)]. It greatly enhances the algorithm by adjusting h to keep local truncation errors within some tolerances. The method performs two evaluations of the equations of motion at each step and compares the answers. If the answers are reasonably close, they're accepted. The step size is changed if too many, or too few, digits are accurate. The final answer uses the initial values with a fourth- or fifth-order Runge-Kutta. This feature is especially nice on highly eccentric orbits like the Molniya. These orbits are best computed using variable step sizes to maintain some given level of accuracy. Without this variable step size, we waste a lot of time near apoapsis, when the integration is taking too small a step. Likewise, the integrator may not be using a small enough step size at periapsis, where the satellite is traveling very fast. A superior approach uses an integrator with a fixed step size and a variable time parameter, s , which varies as a function of the satellite's position (time-regularized Cowell, for example). We use Eq. (7-7) as the first-order equations of motion for satellite problems. The Runge-Kutta-Fehlberg algorithm employs the following values:

$$k_1 = hf(t_o, y_o)$$

$$k_2 = hf\left(t_o + \frac{1}{4}h, y_o + \frac{1}{4}k_1\right)$$

$$k_3 = hf\left(t_o + \frac{3}{8}h, y_o + \frac{3}{32}k_1 + \frac{9}{32}k_2\right)$$

$$k_4 = hf\left(t_o + \frac{12}{13}h, y_o + \frac{1932}{2197}k_1 - \frac{7200}{2197}k_2 + \frac{7296}{2197}k_3\right)$$

$$k_5 = hf\left(t_o + h, y_o + \frac{439}{216}k_1 - 8k_2 + \frac{3680}{513}k_3 - \frac{845}{4104}k_4\right)$$

$$k_6 = hf\left(t_o + \frac{1}{2}h, y_o - \frac{8}{27}k_1 + 2k_2 - \frac{3544}{2565}k_3 + \frac{1859}{4104}k_4 - \frac{11}{40}k_5\right)$$

The two approximations use these values with a fourth-order and fifth-order Runge-Kutta algorithm. Their difference is

$$err = \frac{1}{360}k_1 - \frac{128}{4275}k_3 - \frac{2197}{75240}k_4 + \frac{1}{50}k_5 + \frac{2}{55}k_6$$

If the absolute value of the error is greater than some tolerance, say 1.0×10^{-5} , we determine whether to double or halve the step size based on the parameter, s .

$$s \equiv 0.8408 \left[\frac{1 \times 10^{-8}h}{err} \right]^{1/4}$$

Using predetermined minimum, $h_{min} = h/64$, and maximum, $h_{max} = h/64$, values for the step size, we determine any changes to the step size.

IF $s < 0.75$ and $h > 2h_{min}$ THEN $h = h/2$

IF $s > 1.5$ and $2h < h_{max}$ THEN $h = 2h$

Otherwise,

$$y = y_o + \frac{25}{216}k_1 + \frac{1408}{2565}k_3 + \frac{2197}{4104}k_4 - \frac{1}{5}k_5$$

Adams-Basforth-Moulton

This multi-step method estimates the state over time using previously determined back values of the solution. Although multi-step methods perform only one evaluation for each step forward (compared to four for the fourth-order Runge-Kutta), they usually have a predictor and a corrector formula, so they really require two evaluations per step. Because these methods require back values, they aren't self-starting. You can usually use a Runge-Kutta technique to supply the initial starting conditions, but you should match the order of both routines (e.g., eighth-order Runge-Kutta and an eighth-order Adams-Basforth-Moulton). Consult Maury and Segal (1969), Long et al. (1989, 6-2 to 6-20), Matthews (1987, 434–437), and Press et al. (1992, 740–744) for more information.

Most presentations start with sum and difference tables to evaluate each step. Herrick (1972, 3-17, 245-302) and Sadler (1974) describe this traditional approach. Montenbruck and Pfleger (1994, 90–91) show how to derive multistep methods, specifically Adams-Basforth's fourth-order method proposed in 1883. Integrating both sides of the first-order equation, [$\dot{y}(t_{n+1}) = f(t, y(t))$], with initial condition [$y(0) = y(t_n)$], we get

$$y(t_{n+1}) = y(t_n) + \int_{t_n}^{t_{n+1}} f(t, y(t)) dt$$

Because we don't know the state at the future time, $y(t)$, we can use an interpolating polynomial. Choosing a cubic polynomial, we get the four points by evaluating the function at the current time, plus the last three times. Thus,

$$\begin{aligned} p(t) &= \alpha_o + \alpha_1\tau + \alpha_2\tau^2 + \alpha_3\tau^3 & \tau &= \frac{\Delta t}{h} \\ \alpha_o &= f(y(t_n)) = f_n & \alpha_1 &= \frac{1}{6}(-2f_{n-3} + 9f_{n-2} - 18f_{n-1} + 11f_n) \\ \alpha_2 &= \frac{1}{6}(-3f_{n-3} + 12f_{n-2} - 15f_{n-1} + 6f_n) & \alpha_3 &= \frac{1}{6}(-f_{n-3} + 3f_{n-2} - 3f_{n-1} + f_n) \end{aligned}$$

Substituting the polynomial results in

$$y_{n+1} = y_n + \int_{t_n}^{t_{n+1}} p(t) dt$$

and the *predictor* formula for the fourth-order (*Adams-Basforth ordinate form*) is

$$y_{n+1}^P = y_n + \frac{h}{24} \{ 55\dot{y}_n - 59\dot{y}_{n-1} + 37\dot{y}_{n-2} - 9\dot{y}_{n-3} \}$$

We'll discuss differences between *summed* and *ordinate* notations shortly. We find the *corrector* part similarly but use the estimated point from the predictor to find \dot{y}_{n+1}^P and discard the $n-3$ value. The fourth-order result (*Adams-Moulton ordinate form*) is

$$y_{n+1} = y_n + \frac{h}{24} \{ 9\dot{y}_{n+1}^P + 19\dot{y}_n - 5\dot{y}_{n-1} + \dot{y}_{n-2} \}$$

Having predicted and corrected values lets us compare the accuracy and repeat the correction process if necessary. As we increase the order of the polynomial, accuracy usually rises. However, complexity also rises, and certain methods produce unstable results. Eighth- and twelfth-order polynomials are common for many highly-accurate formulations. Maury and Segal (1969) show the derivation and give coefficients for orders 1 to 15 of Adams-Basforth-Moulton integrators.

It's important to examine notation. For satellite orbit determination, we need $\dot{\vec{r}}$ when \vec{r} is known. If we model non-conservative forces like drag, we also need $\ddot{\vec{r}}$. Therefore, I'll change the notation from y to show the use of position vectors and their derivatives.

Maury and Segal (1969) state that *summed forms* can improve the accuracy of the integrations. This process uses *back differences*, $\nabla_{i-1}(\dot{\vec{r}}_n)$, and *summed back differences*, S_n^{II} , and S_n^{I} . These values come from summation tables that are usually presented with the derivation of numerical integration routines. The notation typically used with these tables can be very confusing. Sadler (1974, 5) presents a concise explanation. I'll use parentheses to denote that the difference is a function of the state.

$$\begin{aligned} \nabla(\dot{\vec{r}}_n) &= \dot{\vec{r}}_n - \dot{\vec{r}}_{n-1} & \nabla_2(\dot{\vec{r}}_n) &= \dot{\vec{r}}_n - 2\dot{\vec{r}}_{n-1} + \dot{\vec{r}}_{n-2} \\ \nabla_n(\dot{\vec{r}}_n) &= \sum_{i=0}^j (-1)^i \binom{j}{i} \dot{\vec{r}}_{n-i} \end{aligned}$$

The *summed differences* are a way to add the back differences without having to maintain the individual back differences. Maury and Segal (1969) show how to determine the initial and recurring values. The summed back differences at each step are

$$S_n^{\text{I}} = S_{n-1}^{\text{I}} + \dot{\vec{r}}_n \quad S_n^{\text{II}} = S_{n-1}^{\text{II}} + S_{n-1}^{\text{I}} + \dot{\vec{r}}_n = S_n^{\text{I}} + \dot{\vec{r}}_n$$

and the initial values are found from the corrector relations (Adams-Moulton and Cowell), and the previous function evaluations.

$$\dot{\vec{S}}_{n+1}^I = \frac{1}{h} \dot{\vec{r}}_{n+1} - \left\{ \frac{1}{2} \dot{\vec{r}}_{n+1} + \alpha_{AM2} \nabla (\dot{\vec{r}}_{n+1}) + \alpha_{AM3} \nabla_2 (\dot{\vec{r}}_{n+1}) + \alpha_{AM4} \nabla_2 (\dot{\vec{r}}_{n+1}) + \dots \right\}$$

$$\dot{\vec{S}}_{n+1}^{II} = \frac{1}{h} \dot{\vec{r}}_{n+1} - \left\{ \frac{1}{12} \dot{\vec{r}}_{n+1} + \alpha_{C2} \nabla (\dot{\vec{r}}_{n+1}) + \alpha_{C3} \nabla_2 (\dot{\vec{r}}_{n+1}) + \alpha_{C4} \nabla_2 (\dot{\vec{r}}_{n+1}) + \dots \right\}$$

Unfortunately, extensive tables aren't efficient for computer implementations, so I'll proceed to alternate forms. Maury and Segal (1969) show that some forms allow us to use only the back values. The resulting formulas are called **ordinate** forms and they avoid our maintaining extensive tables of back differences for the function. I won't list these formulas because the logical extension is to find a **summed-ordinate** form that has improved truncation properties and uses only back values. These are the forms that are most useful for astrodynamics. For the Adams-Bashforth-Moulton numerical integrator,

PREDICT - ADAMS-BASHFORTH SUMMED-ORDINATE FORM (order j)

$$\dot{\vec{r}}_{n+1}^P = h \left\{ \alpha_{ABo} \dot{\vec{S}}_n^I + \sum_{i=0}^j \beta_{ABi} \dot{\vec{r}}_{n-i} \right\}$$

CORRECT - ADAMS-MOULTON SUMMED-ORDINATE FORM (order j)

$$\dot{\vec{r}}_{n+1} = h \left\{ \alpha_{ABo} \dot{\vec{S}}_{n+1}^I + \sum_{i=0}^j \beta_{AMI} \dot{\vec{r}}_{n+1-i} \right\}$$

Notice that this technique evaluates the derivative of the predicted state before doing the correction. For orbit determination with non-conservative systems, we must use the equations once to find velocity, and again to find position. We then form the predicted state evaluation $\dot{\vec{r}}_{n+1}^P$ using those two results, $\dot{\vec{r}}_{n+1}^P$ and $\dot{\vec{r}}_{n+1}^P$. This predicted value is used when $j = 0$ in the corrector phase. This isn't very efficient. Varying the step size is more difficult than with single-step methods because equal spacing is implicit in the formulation. This difficulty can be overcome with simple interpolation routines and isn't difficult whenever we have to double the step-size (Matthews, 1987, 346–437).

Sum-squared (Gauss-Jackson)

The sum-squared method is another multi-step method. It differs from the Adams-Bashforth-Moulton method because it's designed especially for second-order systems. Maury and Segal (1969) present an excellent summary including tables of coefficients. Herrick (1972, 12–17, 260–275), Long et al. (1989, 6-2 to 6-20), and Baker (1967, 452–453) also introduce the formulation. This method is called by many names. I'll conform to the nomenclature presented by Maury and Segal (1969) and use the *Störmer predictor*, and a *Cowell corrector*. Remember that the formulation is for second order systems—if we

include drag, we must use the Adams-Bashforth-Moulton method, or another first-order formulation because the equations of motion depend on position and velocity vectors.

The same forms exist for the Gauss-Jackson routine as before—summed, ordinate, and summed-ordinate. For computer implementations, we prefer the summed-ordinate formulation. The method uses first and second, summed back-differences shown in the Adams-Bashforth-Moulton method. We use the predicted state in the corrector when $j = 0$.

PREDICT - STÖRMER SUMMED-ORDINATE FORM (order j)

$$\hat{r}_{n+1}^p = h \left\{ \alpha_{So} \hat{S}_n^{\text{II}} + \alpha_{S1} \hat{S}_n^{\text{I}} + \sum_{i=0}^j \beta_{Si} \hat{r}_{n-i} \right\}$$

CORRECT - COWELL SUMMED-ORDINATE FORM (order j)

$$\hat{r}_{n+1} = h \left\{ \alpha_{So} \hat{S}_{n+1}^{\text{II}} + \{ \alpha_{So} + \alpha_{S1} \} \hat{S}_{n+1}^{\text{I}} + \sum_{i=0}^j \beta_{Ci} \hat{r}_{n+1-i} \right\}$$

7.5.1 Implementing an Integrator and Determining Step Size

The advantages of a numerical approach may seem to exclude other forms, but this approach doesn't reveal as much of the problem's qualitative behavior. Also remember the inherent difficulties of computer runtime, and, perhaps less discussed—choice of integrator and step size. The *type* of orbit integrator can drastically affect performance and accuracy. We could spend considerable time discussing how to select an integrator for a particular problem, so I've introduced only some of the more popular techniques.

The first-order equations in Eq. (7-7) allow us to easily implement a single-step method. Multi-step techniques require additional explanation. Suppose we want to numerically integrate a complex perturbation model like the one described in Sec. 7.7.2. We'll need both the Adams-Bashforth-Moulton, and Gauss Jackson routines for a complete solution. First, we use the Adams-Bashforth and Störmer integrators to find the velocity and position, respectively, from the initial back values of the acceleration. Next, we use the predicted position and velocity vectors to determine the predicted acceleration at the new time. We then use the Adams-Moulton and Cowell integrators to find the velocity and position, respectively, using the predicted acceleration, and the remaining back values. Finally, if the predicted and corrected values are close enough, the answer is accepted, otherwise, we repeat the corrector step using the new corrected values as inputs.

Table 7-1 lists coefficients needed to form different orders of numerical integrators. Consult Maury and Segal (1969) for complete tables to order 16.

Finding initial conditions for multi-step methods is difficult. However, we can minimize errors by matching the order of all integrators. For instance, if we use an eighth-order, Gauss-Jackson, multi-step integrator, we should use an eighth-order Runge-Kutta to determine the back values required by the algorithm.

TABLE 7-1. Numerical Integration Coefficients. I've listed sample coefficients for the Adams-Basforth-Moulton (AB and AM) and Störmer-Cowell (S and C) integrators. The common denominators, D , assist computer implementation for the summed ordinate forms (β_i). The non-summed forms (α_i) may be truncated while the summed ordinate forms are designed for a finite number of terms. (Maury and Segal, 1969)

	Non-summed Forms				Summed Ordinate Forms			
	α_{ABi}	α_{AMi}	α_{Si}	α_{Ci}	β_{ABi}	β_{AMi}	β_{Si}	β_{Ci}
D					3,628,880	3,628,880	3,628,880	3,628,880
0	1	1	1	1	10,468,447	1,070,017	2,137,435	229,124
1	1/2	-1/2	0	-1	-32,656,759	1,904,811	-6,996,434	304,443
2	5/12	-1/12	1/12	1/12	59,220,027	-2,696,283	13,508,829	-580,962
3	3/8	-1/24	1/12	0	-77,035,235	2,899,075	-6,555,520	677,885
4	251/720	-19/720	19/240	-1/240	60,933,245	-2,134,045	13,080,805	-516,840
5	95/288	-3/160	3/40	-1/240	-30,239,397	1,012,293	-6,485,130	249,861
6	19,087/ 60,480	-863/ 60,480	863/ 12,096	-221/ 60,480	8,594,089	-278,921	1,841,539	-68,658
7	5,257/ 17,280	-275/ 24,192	275/ 4,032	-19/ 6,048	-1,070,017	33,953	-229,124	8,547
8	1,070,017/ 3,628,800	-33,953/ 3,628,800	33,953/ 518,400	-9,829/ 3,628,800				

Once you've chosen an integrator, you'll need to determine the correct step size for methods using fixed step sizes (shortly, we'll consider variable step-sizes). Recall the example in Sec. 7.2 demonstrating special perturbation techniques. Although the numerical approach was eventually able to arrive at a correct answer, the computing process was long, and adjusting the step size was tedious.

Experience has shown that about 100 steps per revolution are satisfactory for standard applications. For eccentric orbits, however, this rule of thumb doesn't hold for fixed-step integrators. Consider a circular orbit with a period of about 100 minutes. The simple solution suggests using a *Shannon-Sampling Theorem* (Phillips and Nagle, 1984, 78) to determine step size. Shannon's theory states that a function with a given frequency, f , is uniquely determined by sampling points which are $1/(2f)$ apart. This is also the *Nyquist critical frequency*, $2f$. Numerical-integration techniques often calculate at five times this rate, so they often encounter a one-tenth factor. Using $1/10$ times the smallest period of oscillation seems reasonable. Yet, for an orbit with a 100-minute period using a 50×50 gravity field, the cyclical oscillations occur every two minutes ($100/50$). A step size of 12 seconds may be far too small! Even a typical step size of one minute will yield acceptable, moderate accuracy but will take 100 steps to propagate one revolution. Data for Landsat 4 and Landsat 6 (McClain, 1990), which are in basically circular orbits (~800 km), indicates that a one-minute step size yields about a 47 m error. A three-minute step size produces

about a 936 m error (average). That's unacceptable for one revolution. Now consider an elliptical orbit with an eccentricity of about 0.7 and a period of one day. The same one- to three-minute step size is much too small near apogee because the satellite moves very slowly through gradually changing perturbational forces. Near perigee, however, this same step size may be too large to track the transition between drag and the central-body.

We can determine a fixed step size for each class of satellites. For example, a satellite in LEO requires a much smaller step size than one in GEO. The particular choice of step size depends on the most rapidly varying component in the function we're integrating and how accurately we want to track it. For most orbital problems, this component is the highest frequency in the disturbing function we want to track.

Variable step sizes are often used for general-purpose models, or whenever we encounter many kinds of functions. Although they're accurate, they're not usually very efficient for the orbital problem. That's because many orbits are circular, or nearly circular, and a method using fixed step sizes is fast and accurate. With variable step sizes, we spend a lot of time finding a correct value, usually through trial and error. The resulting values are typically very close to those from approaches using a fixed step size, so the additional calculations actually hinder the performance for nearly circular orbits. Variable step sizes are available for most popular integrators from Runge-Kutta to the Adams-Basforth and Gauss-Jackson methods. Unfortunately, one integrator isn't the best for *all* orbit types. The complexity of the integrating routine may be an additional hindrance because multi-step methods usually require equally spaced values to form the polynomials. Adjusting the step size can require us to determine additional past values if the new step size doesn't match existing derivatives. For a discussion of other issues, see Sec. 7.8.1.

Another well-known technique is *regularization*—any approach that eliminates $1/r$ singularities of the motion and replaces time as the independent variable in the integration. Szebehely (1967, 93–116) investigated potential applications for the restricted three-body problem. Stiefel and Scheifele (1971, 1–16, 19–20, 77–87) and others cover all applications. One approach uses an auxiliary variable, s , to replace time, so $s = \Delta t / r$. This particular scheme adjusts the new “time” variable(s) as the orbit's radius oscillates, which strongly improves its performance. It's possible to show that s is proportional to the eccentric anomaly for an unperturbed orbit.

A very popular technique is known as *time-regularized Cowell*, which causes a controlled variation of the time step by integrating with fixed step sizes in another variable, such as the eccentric anomaly or true anomaly. Remember,

$$ndt = \frac{r}{a} dE = \frac{r^2}{a\sqrt{1-e^2}} d\nu$$

Selecting the best alternate variable to use depends on the perturbation. One approach adopts the time transformation (actually a Sundmann transformation) of

$$dt = \left(\frac{r}{a}\right)^{3/2} ds \quad (7-8)$$

and then transforms the equations of motion so they're calculated with respect to the new variable, s :

$$\frac{da}{dt} = \frac{da}{ds} \frac{ds}{dt} \Rightarrow \frac{da}{ds} = \frac{da}{dt} \frac{dt}{ds}$$

Constant s , the new step size, corresponds to a changing Δt , which produces the smallest time step at perigee and the largest time step at apogee. The price for this automatic time-step variation is that we must solve a seventh differential equation [$ds/dt = (r/a)^{3/2}$ from Eq. (7-8)] to obtain $s = s(t)$. Because we want the satellite state to be a function of time and because we no longer have equally time-spaced data, this can cause some extra work to obtain interpolated values of the state. Even so, the accuracy and efficiency of a time-regularized Cowell make it very attractive for eccentric orbits.

7.6 Disturbing Forces

To numerically integrate Cowell's formulation, we must have representations for each perturbing acceleration. These can be analytical formulas, or even tabular representations (e.g., tables of atmospheric density). This section derives the more common disturbing forces, including the accelerations resulting from the central body, drag, third body, solar-radiation pressure, and other smaller forces.

7.6.1 Gravity Field of a Central Body

Given that the potential for a spherical central body (the gradient) will yield the acceleration, we must examine how to form a potential for the perturbing effects of a nonspherical central body. The perturbing effects are mathematically described by an *aspherical-potential function*, U , which we'll develop following the method of Roy (1988, 171–177), Kaplan (1976, 273–282), and Fitzpatrick (1970, 288–308).

Deriving the Aspherical Potential

Because the aspherical-potential function provides the means to determine the gravitational attraction on a satellite, it's logical to assume that a beginning point for the derivation would include the gravitational constant, G , and the masses involved. If we examine an infinite number of masses, m_Q , at points Q , the potential per unit mass is the summation of the potential due to each of these points, and the acceleration is ∇U . Therefore,

$$U = G \sum_{Q=1}^{\infty} \frac{m_Q}{\rho_Q}$$

Figure 7-3 shows the geometry.

I'll use a subscript with the mass, m_{\oplus} , to ensure it's not confused with the index m used later in this derivation. The change in potential due to an individual element of mass, dm_{\oplus} , is

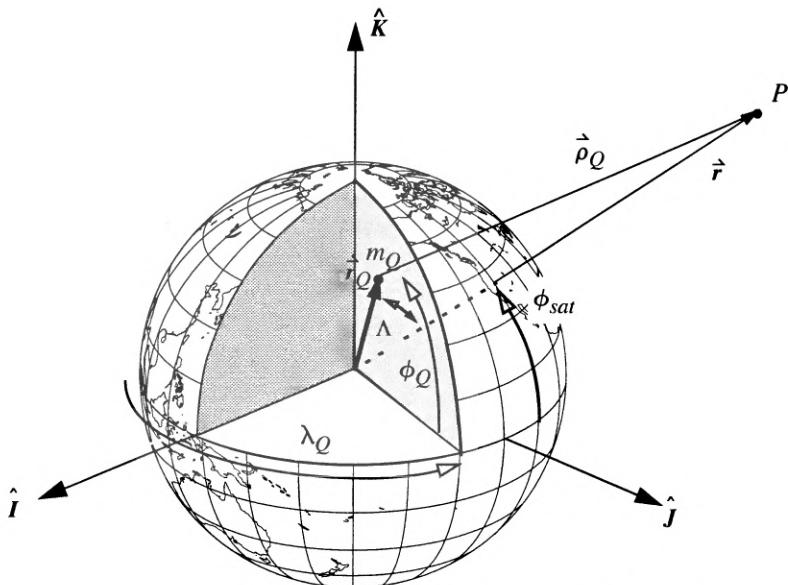


Figure 7-3. Deriving the Gravitational Potential. We consider each point in the Earth, m_Q , when determining the gravitational potential. Remember that the latitude is a geocentric value. Λ is the angle between the vectors r_Q and r .

$$dU = G \frac{dm_{\oplus}}{\rho_Q}$$

In reality, a large but finite number of very small masses are within a body of fixed size; thus, integrating over the entire body gives us the total potential

$$U = G \int_{body} \frac{1}{\rho_Q} dm_{\oplus}$$

The geocentric distance of the point, P , where we're evaluating the potential, r , and the distance from an infinitesimal mass, m_Q , to the center of the Earth, r_Q , are

$$r = \sqrt{x^2 + y^2 + z^2} \quad r_Q = \sqrt{\xi^2 + \eta^2 + \zeta^2}$$

Let's express the slant range, ρ_Q , (using the law of cosines) for the infinitesimal mass to the external point as

$$\rho_Q^2 = r^2 + r_Q^2 - 2rr_Q \cos(\Lambda)$$

We determine the angle between the two position vectors, Λ , by using the dot product:

$$\cos(\Lambda) = \frac{x\xi + y\eta + z\zeta}{rr_Q} = \frac{\vec{r} \cdot \vec{r}_Q}{rr_Q}$$

The slant range is now

$$r_Q = r \sqrt{1 - 2\frac{r_Q}{r} \cos(\Lambda) + \left(\frac{r_Q}{r}\right)^2} = r \sqrt{1 - 2\alpha \cos(\Lambda) + \alpha^2}$$

$$\alpha = \frac{r_Q}{r}$$

The potential (of the body) then becomes

$$U = G \int_{body} \frac{dm_\oplus}{r \sqrt{1 - 2\alpha \cos(\Lambda) + \alpha^2}}$$

The arbitrary parameter, α , will always be less than 1.0 for a point, P , outside the central body ($r > R_\oplus$), and the absolute value of the quantity, $\cos(\Lambda)$, will always be less than or equal to 1.0. Thus, we can use the binomial theorem to expand the denominator of the potential in a series, showing the general form in the middle:

$$\frac{1}{\sqrt{1 - 2\alpha \cos(\Lambda) + \alpha^2}} = \frac{1}{\sqrt{1+x}} = \sum_{t=0}^{\infty} \alpha^t P_t[\cos(\Lambda)]$$

This expression is the definition of *Legendre polynomials*. I'll use brackets to distinguish the argument, $\cos(\Lambda)$. Setting $\gamma = \cos(\Lambda)$, a form of Rodriguez's formula gives the *conventional* Legendre polynomials (Lambeck, 1988, 10):

$$P_t[\gamma] = \frac{1}{2^t t!} \frac{d^t (\gamma^2 - 1)^t}{d\gamma^t}$$

$$P_t[\gamma] = \frac{1}{2^t} \sum_{j=0}^t \frac{(-1)^j (2t-2j)!}{j! (t-j)! (t-2j)!} \gamma^{t-2j} \quad (7-9)$$

The first six Legendre polynomials are

$$P_0[\gamma] = 1$$

$$P_3[\gamma] = \frac{1}{2} (5\gamma^3 - 3\gamma)$$

$$P_1[\gamma] = \gamma$$

$$P_4[\gamma] = \frac{1}{8} (35\gamma^4 - 30\gamma^2 + 3)$$

$$P_2[\gamma] = \frac{1}{2} (3\gamma^2 - 1)$$

$$P_5[\gamma] = \frac{1}{8} (63\gamma^5 - 70\gamma^3 + 15\gamma)$$

Notice I've factored the leading coefficients of the Legendre polynomials from those usually in the literature to make them computationally efficient. We'll explore recurrence relations to find additional polynomials in Eq. (7-52).

I'll talk more about computing the gravitational acceleration in Chap. 8 (page 530). For now, let's write the potential as

$$U = \frac{G}{r} \int_{body} \sum_{l=0}^{\infty} \alpha^l P_l [\cos(\Lambda)] dm_{\oplus} \quad (7-10)$$

To continue the analysis we can use two main methods: MacCullagh's (1855) geometric method and a technique for spherical-harmonic expansion. I'll introduce MacCullagh's technique because it eventually leads to a formula for moment of inertia. It's useful for certain perturbation analyses and in forming spherical-harmonic coefficients of other planets and moons. It also gives a more intuitive feel for gravity and inertia. The spherical-harmonic method may be easier to follow for an introduction to this topic.

With MacCullagh's approach, we express the potential in Eq. (7-10) as

$$U = U_0 + U_1 + U_2 + \dots$$

This permits us to evaluate each component. Let's find the first term using the first Legendre polynomial, P_0 , so that

$$U_0 = \frac{G}{r} \int dm_{\oplus} = \frac{Gm_{\oplus}}{r} = \frac{\mu}{r}$$

This is the two-body potential. Indeed, if we don't need to account for any departure from spherical symmetry (mass anomalies), the potential reduces to the above expression.

The second term is slightly more difficult. If we use the Legendre polynomial and substitutions identified,

$$\begin{aligned} U_1 &= \frac{G}{r} \int \cos(\Lambda) \alpha dm_{\oplus} = \frac{G}{r} \int \frac{x\xi + y\eta + z\zeta}{r^2} dm_{\oplus} \\ &= \frac{G}{r^3} \left(x \int \xi dm_{\oplus} + y \int \eta dm_{\oplus} + z \int \zeta dm_{\oplus} \right) \end{aligned}$$

So each integral (for ξ , η , and ζ) represents the component of the potential to account for a displacement of the center of mass from the origin of the coordinate frame. Even if the origin coincides with the geometric center of the body, it may not coincide with the center of mass. However, if the origin coincides with the center of mass,

$$\int \xi dm_{\oplus} = \int \eta dm_{\oplus} = \int \zeta dm_{\oplus} = 0$$

This is the definition of the center of mass; thus, $U_1 = 0$.

The third term in the expansion is

$$U_2 = \frac{1}{2} \frac{G}{r^3} \int (3\gamma^2 - 1) r_Q^2 dm_{\oplus}$$

Remembering that $\gamma^2 = \cos^2(\Lambda)$, we can substitute using Eq. (C-8).

$$U_2 = \frac{G}{2r^3} \int 2r_Q^2 dm_{\oplus} - \frac{G}{2r^3} \int 3r_Q^2 \sin^2(\Lambda) dm_{\oplus}$$

The left-hand term represents the *moments of inertia* about the three coordinate axes:

$$\left. \begin{aligned} A &\equiv \int (\eta^2 + \xi^2) dm_{\oplus} \\ B &\equiv \int (\xi^2 + \zeta^2) dm_{\oplus} \\ C &\equiv \int (\zeta^2 + \eta^2) dm_{\oplus} \end{aligned} \right\} \quad A + B + C = \int 2r_Q^2 dm_{\oplus}$$

The remaining term is the *polar moment of inertia*, I , that refers to any point r_Q :

$$I \equiv \int r_Q^2 \sin^2(\Lambda) dm_{\oplus}$$

These definitions allow us to write $U_2 = \frac{G}{2r^3} (A + B + C - 3I)$. Substituting results of the three cases discussed so far, and assuming the coordinate frame's origin is at the center of mass, gives us the aspherical potential

$$U = \frac{Gm_{\oplus}}{r} + \frac{G}{2r^3} (A + B + C - 3I) \quad (7-11)$$

This is *MacCullagh's formula* (1855). Evaluating the integrals defining A , B , C , and I presents a practical problem because it requires detailed knowledge of the Earth's mass distribution. Making assumptions about the Earth's size and density distribution allows some progress. Some of the simplest assumptions are the ellipsoidal Earth (oblate spheroid) with uniform density or the prolate spheroid and triaxial ellipsoid. These assumptions aren't very realistic for representing the Earth's potential, but they're a first step if other knowledge is absent.

The geometric approach examines Eq. (7-10). The potential in the form of Eq. (7-10) isn't terribly useful because we can't directly find the angle, Λ . As an aside, this angle becomes important in Chap. 10, so we'll call it the *ground range*, or the *total range angle*, Λ . Spherical trigonometry allows us to develop an equation for this angle. Remember that the satellite's latitude, ϕ_{sat} , will always be a *geocentric* value. The cosine law of spherical trigonometry yields

$$\begin{aligned}\cos(\Lambda) &= \cos(90^\circ - \phi_Q) \cos(90^\circ - \phi_{sat}) \\ &\quad + \sin(90^\circ - \phi_Q) \sin(90^\circ - \phi_{sat}) \cos(\lambda_Q - \lambda_{sat})\end{aligned}$$

Reduction gives us

$$\cos(\Lambda) = \sin(\phi_Q) \sin(\phi_{sat}) + \cos(\phi_Q) \cos(\phi_{sat}) \cos(\lambda_Q - \lambda_{sat}) \quad (7-12)$$

The *addition theorem* of spherical harmonics or the *decomposition formula* provides a way to substitute the expressions for the Λ into Eq. (7-10) (Hobson, [1931] 1965, 141):

$$\begin{aligned}P_t[\cos(\Lambda)] &= P_t[\sin(\phi_Q)] P_t[\sin(\phi_{sat})] \\ &\quad + 2 \sum_{m=1}^t \frac{(l-m)!}{(l+m)!} \{A_{lm} A'_{lm} + B_{lm} B'_{lm}\}\end{aligned} \quad (7-13)$$

where

$$\begin{aligned}A_{lm} &= P_{lm}[\sin(\phi_Q)] \cos(m\lambda_Q) & A'_{lm} &= P_{lm}[\sin(\phi_{sat})] \cos(m\lambda_{sat}) \\ B_{lm} &= P_{lm}[\sin(\phi_Q)] \sin(m\lambda_Q) & B'_{lm} &= P_{lm}[\sin(\phi_{sat})] \sin(m\lambda_{sat})\end{aligned}$$

Now, let's introduce the *associated Legendre functions*, P_{lm} . They have the general form given by Lambeck (1988, 10):

$$P_{l,m}[\gamma] = \frac{1}{2^l l!} (1-\gamma^2)^{m/2} \frac{d^{l+m}}{d\gamma^{l+m}} (\gamma^2 - 1)^l$$

or

$$P_{l,m}[\gamma] = (1-\gamma^2)^{m/2} \frac{d^m}{d\gamma^m} P_l[\gamma] \quad (7-14)$$

Table 7-2 shows a few values. Notice that for zero order ($m = 0$), the associated Legendre functions are simply the conventional Legendre polynomials. Practical computation of these functions is handled by recursion, discussed in Sec. 7.7.2.

Now let's make an important substitution to Eq. (7-13). By separating all the terms that are independent of the satellite's location, we can arrive at a solution that isolates terms which depend only on the central body and those which relate the satellite's position. Define

$$\begin{aligned}C'_{lm} &= \int_{body} r_Q^l 2 \frac{(l-m)!}{(l+m)!} P_{lm}[\sin(\phi_Q)] \cos(m\lambda_Q) dm_\oplus \\ S'_{lm} &= \int_{body} r_Q^l 2 \frac{(l-m)!}{(l+m)!} P_{lm}[\sin(\phi_Q)] \sin(m\lambda_Q) dm_\oplus\end{aligned} \quad (7-15)$$

TABLE 7-2. Associated Legendre Functions. This table gives a few sample values for the associated Legendre function, $P_{l,m}[\sin(\phi_{sat})]$. Remember that for astrodynamics we need the geocentric latitude. (Lambeck, 1988)

$P_{0,0}$	1	$P_{3,2}$	$15\cos^2(\phi_{sat})\sin(\phi_{sat})$
$P_{1,0}$	$\sin(\phi_{sat})$	$P_{3,3}$	$15\cos^3(\phi_{sat})$
$P_{1,1}$	$\cos(\phi_{sat})$	$P_{4,0}$	$\frac{1}{8}\left\{35\sin^4(\phi_{sat}) - 30\sin^2(\phi_{sat}) + 3\right\}$
$P_{2,0}$	$\frac{1}{2}\left\{3\sin^2(\phi_{sat}) - 1\right\}$	$P_{4,1}$	$\frac{5}{2}\cos(\phi_{sat})\left\{7\sin^3(\phi_{sat}) - 3\sin(\phi_{sat})\right\}$
$P_{2,1}$	$3\sin(\phi_{sat})\cos(\phi_{sat})$	$P_{4,2}$	$\frac{15}{2}\cos^2(\phi_{sat})\left\{7\sin^2(\phi_{sat}) - 1\right\}$
$P_{2,2}$	$3\cos^2(\phi_{sat})$	$P_{4,3}$	$105\cos^3(\phi_{sat})\sin(\phi_{sat})$
$P_{3,0}$	$\frac{1}{2}\left\{5\sin^3(\phi_{sat}) - 3\sin(\phi_{sat})\right\}$	$P_{4,4}$	$105\cos^4(\phi_{sat})$
$P_{3,1}$	$\frac{1}{2}\cos(\phi_{sat})\left\{15\sin^2(\phi_{sat}) - 3\right\}$		

The coefficients of Eq. (7-15) represent the mathematical modeling for the Earth's shape using spherical harmonics. The special case for the *zonal harmonics* is

$$C'_{10} = \int_{\text{body}} r_Q^l P_l[\sin(\phi_Q)] dm_{\oplus}$$

which uses the conventional Legendre polynomials. Note that S'_{10} is zero. When we substitute these values into the potential in Eq. (7-10),

$$\begin{aligned} U &= \frac{G}{r} \sum_{l=0}^{\infty} \frac{P_l[\sin(\phi_{sat})]}{r^l} C'_{10} \\ &\quad + \frac{G}{r} \sum_{l=1}^{\infty} \sum_{m=1}^l \frac{P_{lm}[\sin(\phi_{sat})]}{r^l} \{ C'_{lm} \cos(m\lambda_{sat}) + S'_{lm} \sin(m\lambda_{sat}) \} \end{aligned}$$

Finally, a unit analysis suggests we remove units from the result. This leads to the familiar C and S coefficients. R_{\oplus} and m_{\oplus} are the Earth's mean equatorial radius and mass, respectively. We can normalize the gravitational coefficients (Lambeck, 1988, 11):

$$C'_{lm} = C_{lm} R_{\oplus}^l m_{\oplus} \quad S'_{lm} = S_{lm} R_{\oplus}^l m_{\oplus}$$

Now, let's use the definition of the gravitational parameter and find

$$\begin{aligned} U = & \frac{\mu}{r} \sum_{l=0}^{\infty} P_l [\sin(\phi_{sat})] \left(\frac{R_{\oplus}}{r} \right)^l C_{l0} \\ & + \frac{\mu}{r} \sum_{l=1}^{\infty} \sum_{m=1}^l P_{lm} [\sin(\phi_{sat})] \left(\frac{R_{\oplus}}{r} \right)^l \{ C_{lm} \cos(m\lambda_{sat}) + S_{lm} \sin(m\lambda_{sat}) \} \end{aligned} \quad (7-16)$$

I'll present several of the common notations for the aspherical potential. The International Astronomical Union (1961) has adopted these forms, which you can find in Lambek (1988, 11), Kaplan (1976, 281), and others. First, a double summation encompasses the associated Legendre polynomials, P_{lm} , and the harmonic coefficients, and $S_{l0} = 0$.

$$U = \frac{\mu}{r} \sum_{l=0}^{\infty} \sum_{m=0}^l \left(\frac{R_{\oplus}}{r} \right)^l P_{lm} [\sin(\phi_{sat})] \left\{ C_{lm} \cos(m\lambda_{sat}) + S_{lm} \sin(m\lambda_{sat}) \right\}$$

Notice that both summation indices begin at zero. Recall, as part of the derivation, if the center of the coordinate system coincides with the attracting body's center of mass, the coefficients C_{10} , C_{11} , and S_{10} are all zero ($l = 1$ in MacCullagh's approach). This leads to another very common form of this relation, which separates the spherical potential and therefore requires us to adjust the summation limits. The 0^{th} term is within the μ/r term, and the first-degree term is zero.

$$U = \frac{\mu}{r} \left[1 + \sum_{l=2}^{\infty} \sum_{m=0}^l \left(\frac{R_{\oplus}}{r} \right)^l P_{lm} [\sin(\phi_{sat})] \left\{ C_{lm} \cos(m\lambda_{sat}) + S_{lm} \sin(m\lambda_{sat}) \right\} \right] \quad (7-17)$$

We can also separate the zonal and tesseral terms from Eq. (7-17) by recalling that $J_l = -C_{l0}$. The minus sign for the zonal harmonic coefficients is purely conventional, as described in Chap. 1 (Sec. 1.3). Thus,

$$\begin{aligned} U = & \frac{\mu}{r} \left[1 - \sum_{l=2}^{\infty} J_l \left(\frac{R_{\oplus}}{r} \right)^l P_l [\sin(\phi_{sat})] \right. \\ & \left. + \sum_{l=2}^{\infty} \sum_{m=1}^l \left(\frac{R_{\oplus}}{r} \right)^l P_{lm} [\sin(\phi_{sat})] \left\{ C_{lm} \cos(m\lambda_{sat}) + S_{lm} \sin(m\lambda_{sat}) \right\} \right] \end{aligned} \quad (7-18)$$

This form has several important features. The *conventional* Legendre polynomials are merely special cases of the associated Legendre functions. Eq. (7-18) also depends on the C and S coefficients. As we discussed with MacCullagh's formula, we can resort to simplified approximations to find these values, but with the same minimal results. Fortunately, the analysis of satellite motion allows us to empirically determine the C and S coefficients from observations. This is an important method in developing gravitational fields today.

We often use normalized coefficients because the conventional gravitational coefficients become very small as the degree and order (l, m) get large. This is a consequence of the factorial ratios appearing in Eq. (7-15). Computer truncations may introduce errors. It's common to normalize in order to correct this deficiency. Often, we must convert these values to conventional notation for use in certain programs. The following formula permits the transformations used in this book. The overall transformation, Π_{lm} , is often used to abbreviate the relation. The overbar denotes normalized variables.

$$\begin{aligned}\overline{\Pi}_{lm} &= \sqrt{\frac{(l+m)!}{(l-m)!k(2l+1)}} \\ \overline{S}_{lm} &= \Pi_{lm} S_{lm} \quad \overline{C}_{lm} = \Pi_{lm} C_{lm} \\ k &= 1 \text{ if } m = 0 \\ k &= 2 \text{ if } m \neq 0\end{aligned}\tag{7-19}$$

Also, due to the size of the coefficients, Π_{lm} is usually determined with the maximum precision available on the computer. When using normalized coefficients, you must also normalize the Legendre polynomials (conventional and associated) because the product must remain the same. Therefore,

$$\overline{C}_{lm} \overline{P}_{lm} = C_{lm} P_{lm} \quad \overline{S}_{lm} \overline{P}_{lm} = S_{lm} P_{lm}$$

Don't confuse this process with removing units as described earlier. To normalize, use

$$\overline{C}_{lm} = \frac{C_{lm}}{\overline{\Pi}_{lm}} \quad \overline{S}_{lm} = \frac{S_{lm}}{\overline{\Pi}_{lm}} \quad \overline{P}_{lm} = \overline{\Pi}_{lm} P_{lm}$$

where the overbar denotes normalized values, and $\overline{\Pi}_{lm}$ is the normalizing scheme from Eq. (7-19). An example will show how normalizing works.

▼ Example 7-3. Normalizing Gravitational Coefficients.

GIVEN: Gravitational coefficients in Table D-1 and Table D-2.

FIND: Normalized values for (4,0), (50,1), (70,61).

For $l = 4$ and $m = 0$,

$$\begin{aligned}\bar{C}_{lm} &= \sqrt{\frac{(l+m)!}{(l-m)!k(2l+1)}} C_{lm} = \sqrt{\frac{(4+0)!}{(4-0)!1(2(4)+1)}} 0.162\,042\,999 \times 10^{-5} \\ &= 5.401\,433\,3 \times 10^{-7}\end{aligned}$$

Notice that $k = 1$ and S don't exist for zonal harmonics ($m = 0$).

For $l = 50$ and $m = 1$,

$$\begin{aligned}\bar{C}_{lm} &= \sqrt{\frac{(l+m)!}{(l-m)!k(2l+1)}} C_{lm} = \sqrt{\frac{(50+1)!}{(50-1)!2(2(50)+1)}} 1.019\,057\,862\,7 \times 10^{-9} \\ &= 3.620\,705\,899\,9 \times 10^{-9} \\ \bar{S}_{lm} &= \sqrt{\frac{(l+m)!}{(l-m)!k(2l+1)}} S_{lm} = \sqrt{\frac{(50+1)!}{(50-1)!2(2(50)+1)}} (-5.248\,002\,373 \times 10^{-10}) \\ &= -1.864\,611\,799 \times 10^{-9}\end{aligned}$$

Notice now that $k = 2$ because m isn't zero.

For $l = 70$ and $m = 61$,

$$\begin{aligned}\bar{C}_{lm} &= \sqrt{\frac{(l+m)!}{(l-m)!k(2l+1)}} C_{lm} = \sqrt{\frac{(70+61)!}{(70-61)!2(2(70)+1)}} 1.113\,999\,743\,1 \times 10^{-116} \\ &= 1.013\,587\,499\,9 \times 10^{-9} \\ \bar{S}_{lm} &= \sqrt{\frac{(l+m)!}{(l-m)!k(2l+1)}} S_{lm} = \sqrt{\frac{(70+61)!}{(70-61)!2(2(70)+1)}} (-2.893\,882\,910\,2 \times 10^{-116}) \\ &= -2.633\,037\,899\,9 \times 10^{-9}\end{aligned}$$

▲ Notice the values' similar orders of magnitude after normalizing.

Spherical Harmonics

The trigonometric argument of the Legendre polynomials in Eq. (7-17) constitutes surface **spherical harmonics**, for they are periodic on the surface of a unit sphere. When the surface's spherical harmonics are divided by r^{l+1} , they're usually called solid spherical harmonics. The Sturm-Liouville theorem says that the solid spherical harmonics are eigenfunctions that constitute an independent basis for the gravitational model. In essence, they are a Fourier series. The indices l and m determine lines on the sphere along which the functions vanish. These spherical harmonics further break down into three types of terms—zonal, sectorial, and tesseral harmonics—which I'll introduce in this section. Later in this chapter, we'll see how they affect orbital elements.

Zonal harmonics are defined by zeroth order ($m = 0$), where the dependence of the potential on longitude vanishes and the field is symmetrical about the polar axis. These are simply bands of latitude. For any $P_l[\sin(\phi_{gc})]$, there are l circles of latitude along which P_l is zero, and hence $(l + 1)$ zones in which the function is alternately increasing (+) and decreasing (-). Figure 7-4 shows several depictions of the zonal harmonics. J_2 is by far the strongest perturbation due to the Earth's shape. As the table on the inside front cover shows, J_2 is almost 1000 times larger than the next largest coefficient (J_3). Each of the boundaries of the zonal harmonics represents a root of the Legendre polynomial. They are a transition between positive and negative values.

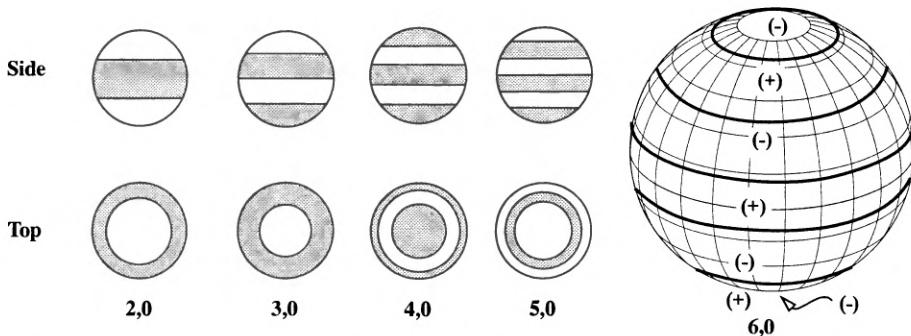


Figure 7-4. **Zonal Harmonics.** J_2 accounts for most of the Earth's gravitational departure from a perfect sphere. This band (and others) reflects the Earth's oblateness. The third harmonic appears similar to the J_2 from the top but is reversed for the bottom view.

Sectorial harmonics occur when $l = m$ and represent bands of longitude. Figure 7-5 shows some sectorial harmonics. The polynomials $P_{ll}[\sin(\phi_{sat})]$ are zero only at the poles ($\phi_{sat} = \pm 90^\circ$). On the other hand, the terms [$\cos(\lambda)$ and $\sin(\lambda)$] are zero for $2l$ different values of λ . Hence the lines along which the functions [$\cos(\lambda)$ and $\sin(\lambda)$] times $P_{ll}[\sin(\phi_{sat})]$ vanish are meridians of longitude, which divide the sphere into $2l$ “orange-slice” sectors. Each sector represents l positive (+) and l negative (-) mass concentrations.

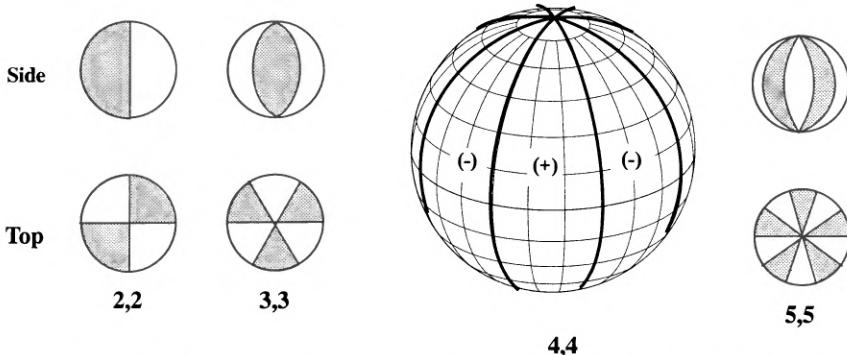
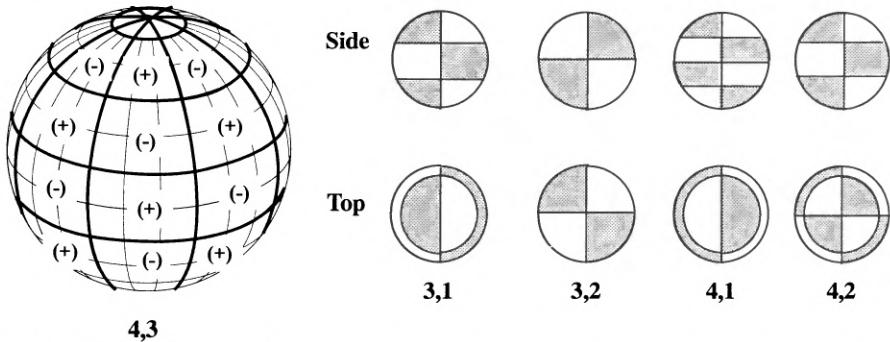


Figure 7-5. **Sectorial Harmonics.** Sectorial harmonics take into account the extra mass distribution in longitudinal regions.

For cases in which $l \neq m \neq 0$, the physical depiction is a specific region on the Earth. These functions [$\cos(m\lambda)$ and $\sin(m\lambda)$] times $P_{lm}[\sin(\phi_{sat})]$ are referred to as **tesseral**

harmonics (because they look like tiles), for the sphere is divided up into a checkerboard array. The number of circles of latitude along which $P_{lm}[\sin(\phi_{sat})]$ is zero is $(l - m)$, whereas the terms [$\cos(m\lambda)$ and $\sin(m\lambda)$] vanish along $2m$ meridians of longitude. Figure 7-6 shows the tesseral harmonics. Notice the similarities to, and differences from, the sectorial harmonics. Some texts refer to the sectorials as a subset of the tesserales. That's correct, but we can also keep them separate to distinguish their physical differences, as I do in this book.



4,3

Figure 7-6. Tesseral Harmonics. Tesseral harmonics attempt to model specific regions on the Earth which depart from a perfect sphere.

Partial Derivatives

Now that we've developed a potential function for the central-body effects, we must take partial derivatives to determine the acceleration we can use in numerical propagations. These partial derivatives are also necessary for the variational equations we'll use in Chap. 9. To determine the acceleration, we use the gradient operation. We'll change the notation slightly to reinforce the fact that we're using the satellite's geocentric latitude, $\phi_{gc} = \phi_{sat}$, and longitude, $\lambda = \lambda_{sat}$. From Long et al. (1989, 4-11 to 4-16), using matrix differentiation with $\dot{\mathbf{r}} = r_I \hat{\mathbf{i}} + r_J \hat{\mathbf{j}} + r_K \hat{\mathbf{k}}$ in the IJK frame, we can directly determine the partial derivatives of the aspherical-potential function (including the two-body attraction), but we must be careful when differentiating the Legendre functions. Because the potential function is expressed in spherical coordinates (r, ϕ_{gc}, λ) , we have to apply the chain rule. The acceleration is

$$\ddot{\mathbf{a}} = \frac{\partial U}{\partial r} \left(\frac{\partial \mathbf{r}}{\partial r} \right)^T + \frac{\partial U}{\partial \phi_{gc}} \left(\frac{\partial \mathbf{r}}{\partial \phi_{gc}} \right)^T + \frac{\partial U}{\partial \lambda} \left(\frac{\partial \mathbf{r}}{\partial \lambda} \right)^T \quad (7-20)$$

Now, using the nonspherical portion of Eq. (7-17),

$$\begin{aligned}
 \frac{\partial U}{\partial r} &= -\frac{\mu}{r^2} \sum_{l=2}^{\infty} \sum_{m=0}^l \left(\frac{R_{\oplus}}{r}\right)^l (l+1) P_{lm} [\sin(\phi_{gc})] \{ C_{lm} \cos(m\lambda) + S_{lm} \sin(m\lambda) \} \\
 \frac{\partial U}{\partial \phi_{gc}} &= \frac{\mu}{r} \sum_{l=2}^{\infty} \sum_{m=0}^l \left(\frac{R_{\oplus}}{r}\right)^l \{ P_{l,m+1} [\sin(\phi_{gc})] - m \tan(\phi_{gc}) P_{lm} [\sin(\phi_{gc})] \} \\
 &\quad \times \{ C_{lm} \cos(m\lambda) + S_{lm} \sin(m\lambda) \} \\
 \frac{\partial U}{\partial \lambda} &= \frac{\mu}{r} \sum_{l=2}^{\infty} \sum_{m=0}^l \left(\frac{R_{\oplus}}{r}\right)^l m P_{lm} [\sin(\phi_{gc})] \{ S_{lm} \cos(m\lambda) - C_{lm} \sin(m\lambda) \}
 \end{aligned} \tag{7-21}$$

The derivatives of the position vector also proceed directly as

$$\begin{aligned}
 \frac{\partial \vec{r}}{\partial \vec{r}} &= \frac{\vec{r}^T}{r} \\
 \frac{\partial \vec{\phi}_{gc}}{\partial \vec{r}} &= \frac{1}{\sqrt{r_I^2 + r_J^2}} \left(-\frac{\vec{r}^T \vec{r}_K}{r^2} + \frac{\partial \vec{r}_K}{\partial \vec{r}} \right) \\
 \frac{\partial \vec{\lambda}}{\partial \vec{r}} &= \frac{1}{r_I^2 + r_J^2} \left(r_I \frac{\partial \vec{r}_J}{\partial \vec{r}} - r_J \frac{\partial \vec{r}_I}{\partial \vec{r}} \right)
 \end{aligned} \tag{7-22}$$

Notice the three partial derivatives in Eq. (7-22) are unit vectors. Putting the acceleration terms together and breaking out individual components make Eq. (7-20)

$$\begin{aligned}
 a_I &= \left\{ \frac{1}{r} \frac{\partial U}{\partial r} - \frac{r_K}{r^2 \sqrt{r_I^2 + r_J^2}} \frac{\partial U}{\partial \phi_{gc}} \right\} \vec{r}_I - \left\{ \frac{1}{r_I^2 + r_J^2} \frac{\partial U}{\partial \lambda} \right\} \vec{r}_J \\
 a_J &= \left\{ \frac{1}{r} \frac{\partial U}{\partial r} - \frac{r_K}{r^2 \sqrt{r_I^2 + r_J^2}} \frac{\partial U}{\partial \phi_{gc}} \right\} \vec{r}_J + \left\{ \frac{1}{r_I^2 + r_J^2} \frac{\partial U}{\partial \lambda} \right\} \vec{r}_I \\
 a_K &= \frac{1}{r} \frac{\partial U}{\partial r} \vec{r}_K + \frac{\sqrt{r_I^2 + r_J^2} \frac{\partial U}{\partial \phi_{gc}}}{r^2} \vec{r}_{\phi_{gc}}
 \end{aligned} \tag{7-23}$$

7.6.2 Atmospheric Drag

Next to the oblateness of the Earth, atmospheric drag most strongly influences the motion of a satellite *near Earth*; in fact, during the last few revolutions of the satellite's life, drag effects can be more important than those from the Earth's oblateness. For more distant satellites, third-body effects and solar-radiation pressure dominate more than oblateness and drag. The Sun's interaction with the upper atmosphere and the Earth's magnetic field

influence drag variations and uncertainty the most. Many references describe the effects and various analytical procedures to model the drag effect [see King-Hele (1987) for example], but models are still incomplete. The study of drag usually divides into three areas: (1) determining orbits under the influence of drag, (2) estimating a satellite's lifetime, and (3) determining physical properties of the upper atmosphere. We'll introduce each.

The cause of drag is the atmospheric density, which retards the satellite's motion. Calculating density is extremely complex for real-world problems. Drag is a nonconservative perturbation because total energy isn't conserved. In fact, all velocity-dependent perturbations are nonconservative. Drag changes mainly the semimajor axis and eccentricity of the orbit. There are periodic effects in the other orbital elements and some coupling aspects with the aspherical potential. We'll examine these ideas shortly.

Studying astrodynamics is most difficult in the atmosphere. To rigorously model the effects of atmospheric perturbation, we must know molecular chemistry, thermodynamics, aerodynamics, hypersonics, meteorology, electromagnetics, and planetary sciences, as well as orbital mechanics. Accurately determining atmospheric properties is crucial to determining drag on satellites. We must understand both the troposphere and the ionosphere to predict the behavior of tracking and communications systems. Investigations of aerobraking and satellite tethers also require accurate atmospheric models. The main idea is to have a good enough model to account for the atmospheric density while analyzing the effects of drag on an orbit.

This book considers various atmospheric models which allow you to use varying levels of sophistication in simulation programs. Some of the basic information in this section is from Jablonski (1991, 1992), who has allowed me to reproduce sections of her papers concerning general atmospheric concepts and the Soviet Cosmos model.

Acceleration Due to Drag

The basic equation for aerodynamic drag combines several factors; I'm showing it here as a specific force or acceleration:

$$\vec{a}_{drag} = -\frac{1}{2} \frac{c_D A}{m} \rho v_{rel}^2 \frac{\vec{v}_{rel}}{|\vec{v}_{rel}|} \quad (7-24)$$

The ***coefficient of drag***, c_D , is a dimensionless value which reflects the satellite's susceptibility to drag forces. The drag coefficient for satellites in the upper atmosphere is often considered to be approximately $c_D \sim 2$. Spheres have $c_D \sim 1.0$. The drag coefficient is configuration-specific and is seldom approximated to more than three significant digits. The ***atmospheric density***, ρ , indicates how dense the atmosphere is at a given altitude and is perhaps the most difficult parameter to determine. Another difficult parameter to estimate is the ***cross-sectional area***, A , defined to be the area which is normal to the satellite's velocity vector. For high-precision studies, we must use attitude determination to help calculate A . For a tumbling reentry vehicle, such as Skylab in 1980, it's nearly impossible to

know the attitude accurately, so A is inherently uncertain. We also need the satellite's mass, m , along with the relative-velocity vector. For some satellites, we can't assume mass is constant. Recognize this is *not* the velocity vector typically found in the state vector. This velocity vector is relative to the atmosphere, hence the subscript. In actuality, the Earth's atmosphere has a mean motion due to the Earth's rotation, and the winds are superimposed on this mean motion. Notice also that the force of drag opposes the velocity vector at all times. This is a primary use for the *NTW* coordinate system. For a nonspherical satellite, we must also consider companion aerodynamic forces such as lift and side forces.

We usually call $m/c_D A$ the **ballistic coefficient**, BC ; it's another measure of the satellite's susceptibility to drag effects. With this definition, a low BC means drag will affect the satellite a lot—and vice versa. These quantities have *many* definitions, so it's very important to understand which one is in use.

The velocity vector relative to the rotating atmosphere is

$$\vec{v}_{rel} = \frac{d\vec{r}}{dt} - \vec{\omega}_{\oplus} \times \vec{r} = \begin{bmatrix} \frac{dx}{dt} + \omega_{\oplus} y \\ \frac{dy}{dt} - \omega_{\oplus} x \\ \frac{dz}{dt} \end{bmatrix}$$

Escobal ([1965] 1985, 60) gives a more general expression including wind variations:

$$\vec{v}_{rel} = \begin{bmatrix} \frac{dx}{dt} + \omega_{\oplus} y + v_w \{ \cos(\alpha) \sin(\delta) \cos(\beta_w) + \sin(\alpha) \sin(\beta_w) \} \\ \frac{dy}{dt} - \omega_{\oplus} x + v_w \{ \sin(\alpha) \sin(\delta) \cos(\beta_w) + \cos(\alpha) \sin(\beta_w) \} \\ \frac{dz}{dt} - v_w \{ \cos(\delta) \cos(\beta_w) \} \end{bmatrix}$$

which requires the wind's speed, v_w , and azimuth, β_w , and the satellite's right ascension and declination. But many applications don't use this expression because the additional information usually isn't available. Rather, the satellite's specific orientation and shape are determined to help determine the satellite's effective cross-sectional area.

Magnetic Field Models

The effect of the magnetic variations of the Earth and Sun is useful with calculations of atmospheric density because it's believed that magnetic variations are related to fluctuations in atmospheric density. The Earth's magnetic field also has a measurable (but small) influence on satellite operations. The effect appears in four ways: (1) the charged particles from any magnetic disturbance cause ionization in the upper atmosphere, thereby affect-

ing the density and, subsequently, the drag; (2) the charges on the particles can actually alter the attractive forces experienced by the satellite; (3) the ionization interferes with satellite tracking and communication, and (4) variations in the magnetic field of the charged particles can interfere with onboard electromagnets that impose torques and carry out slow attitude maneuvers. The attractive force on the satellite is very small and is almost always ignored. But the effect of ionization on atmospheric density is important, so we'll discuss it further.

The magnetic field's strength varies with the environment around the Earth, and the mean magnetic field is usually modeled with a low degree-and-order spherical-harmonic expansion (exactly analogous to the gravitational models). Direct collisions of the solar wind with air particles interacting with the Earth's geomagnetic field heat the atmosphere. Geomagnetic activity must be measured to determine the heat generated. The commonly used geomagnetic ***planetary index***, k_p , is a quasi-logarithmic, worldwide (hence the title) average of geomagnetic activity below the auroral zones. Twelve stations record values of k measured every three hours. Once latitude corrections are applied, the average k_p is formed. The quasi-logarithmic identifier is used because values range from low activity ($k_p = 0.0$) to extreme geomagnetic activity ($k_p = 9.0$) and are quoted to the nearest third of an integer. The Jacchia-Roberts model is an atmospheric model that uses k_p as an input. The geomagnetic ***planetary amplitude***, a_p , is a linear equivalent of the k_p index, designed to minimize differences at 50° latitude. It's also known as the ***3-hourly index*** because eight values of a_p are averaged to create the ***daily planetary amplitude***, A_p , shown in Fig. 7-7. The effects of drag resulting from magnetic disturbances are noticeable for satellites at altitudes between 300 km and 1000 km.

Planetary geomagnetic indices (k_p and a_p) are compiled using measurements from twelve observatories which lie between 48° N and 63° S latitudes; three of these are in the United Kingdom, two in Canada, three in the USA, and the remaining four in New Zealand, Australia, Sweden, and Denmark. The most accepted compilation of the measurements from these observatories is from the Institut für Geophysik at Göttingen University, Germany.

The daily planetary amplitude is in gamma units, where

$$\text{one gamma} = 10^{-9} \text{ Tesla} = 10^{-9} \text{ kg}\cdot\text{s} / \text{m} \quad (7-25)$$

The range of values for A_p is from 0 to 400; values greater than 100 are rare, and values of 10–20 are average. The daily planetary amplitude tends to follow the 11-year cycle of Sun spots, although consistently large maxima of A_p are in the declining phase of each 11-year cycle. There is also a secondary semi-annual cycle due to the variable position of the solar wind with respect to the Earth's magnetosphere. This cycle is just as variable and hard to predict as the Sun-spot cycle. Variations of A_p from the Sun-spot and semi-annual cycles are mainly due to solar flares, coronal holes, disappearing solar filaments, and the solar-wind environment near the Earth (Fraser-Smith, 1972; Nostrand, 1984). Intense geomagnetic activity at the auroral zones affects the shape of the atmosphere and makes atmospheric density depend on latitude (de Lafontaine, 1986).

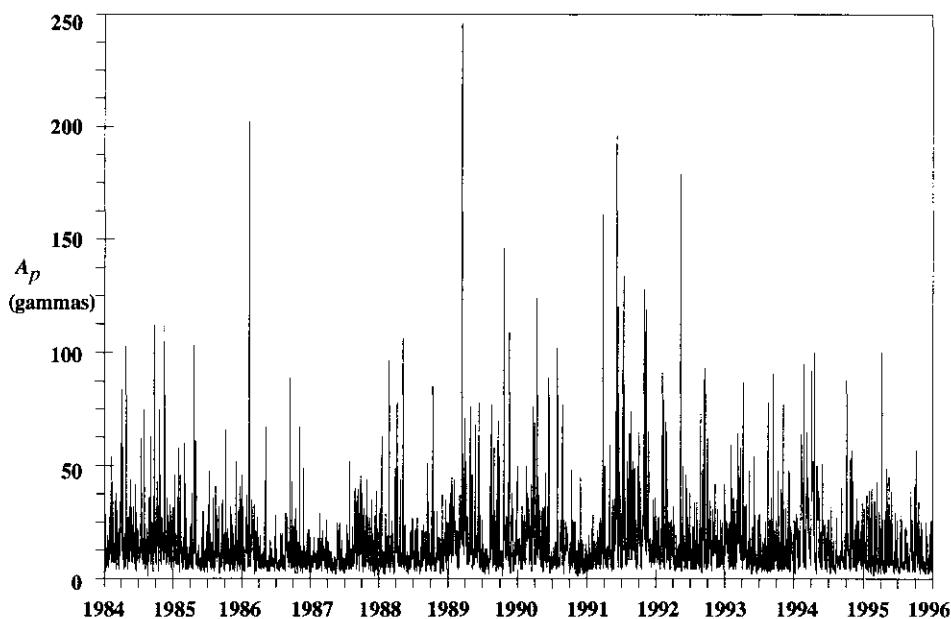


Figure 7-7. Geomagnetic Planetary Index. Notice the erratic behavior at almost all times. Accurate prediction is especially challenging!

Converting between k_p and a_p indices is defined by a standardized process (Jursa, 1985, 4-29). The 3-hourly index, a_p , is approximately related to the planetary index, k_p , by (Wertz, 1978, 122),

$$a_p = \text{EXP}\left(\frac{k_p + 1.6}{1.75}\right) \quad (7-26)$$

To get the field strength of the magnetic field, \vec{B} , we use a more complicated formula. We need the magnetic field only if we want to calculate the perturbative effects of the charged particles on the satellite. It isn't needed for analyzing atmospheric drag. Notice the use again of a spherical harmonic representation of the magnetic-potential function, V . Let's use

$$\vec{B} = \nabla V$$

$$V = R_{\oplus} \sum_{l=1}^k \left(\frac{R_{\oplus}}{r}\right)^{l+1} \sum_{m=0}^l \bar{P}_{lm} [\cos(90^\circ - \phi_{gc})] \left\{ \bar{g}_{lm} \cos(m\lambda) + \bar{h}_{lm} \sin(m\lambda) \right\} \quad (7-27)$$

where the variables are as defined with the gravitational potential. The bars represent normalized parameters [like Eq. (7-19)], which we'll discuss shortly. The *Schmidt functions*, $\bar{P}_{lm} [\cos(90^\circ - \phi_{gc})]$, combine the normalization with the Legendre functions. Note that we need more coefficients if we consider magnetic currents above the Earth as well as those arising only from the Earth. The values in Table 7-3 approximate the *Schmidt coefficients*, \bar{g}_{lm} and \bar{h}_{lm} . See the *International Geomagnetic Reference Field* (IGRF) for current data. We use these coefficients for about five years. Derivatives of the coefficients permit accurate extrapolations.

It's common to normalize as with the gravitational coefficients, but the magnetic-field calculation only partly normalizes, and we refer to the resulting parameters as *Gaussian coefficients*, in honor of Karl Friedrich Gauss.

$$\begin{aligned}\Pi_{lm}^{Schmidt} &= \sqrt{\frac{(2 - \delta_{0,m})}{(l+m)!} \frac{(l-m)!}{(2l-1)!!}} \\ \bar{g}_{lm} &= \Pi_{lm}^{Schmidt} g_{lm} \quad \bar{h}_{lm} = \Pi_{lm}^{Schmidt} h_{lm} \\ \delta_{lm} &= 1 \text{ if } m = l \\ \delta_{lm} &= 0 \text{ if } m \neq l\end{aligned}$$

Notice the Kronecker delta, δ_{lm} , and the double summation in the formulation. Don't confuse this process with normalizing the gravitational coefficients as shown in Eq. (7-19). Also, the difference in normalizing schemes is arbitrary and results from different traditions for geophysics and magnetism. Remember to normalize the Legendre functions if you use normalized coefficients, so the operation ultimately cancels out. You need $(l-m)!$ in both the numerator and denominator to scale the numbers and avoid potential numerical difficulties. Finally, l is greater than 1, and m is greater than 0.

As with the gravitational coefficients, a recursion scheme is possible to compute the geomagnetic potential. Wertz (1978, 781) shows the process:

$$\begin{aligned}\Pi_{0,0}^{Schmidt} &= 1 \\ \Pi_{l,0}^{Schmidt} &= \Pi_{l-1,0}^{Schmidt} \frac{2l-1}{l} \quad l = 1, 2, \dots \\ \Pi_{l,m}^{Schmidt} &= \Pi_{l,m-1}^{Schmidt} \sqrt{\frac{(l-m+1)(\delta_{l,m}+1)}{l+m}} \quad m = 1, 2, \dots\end{aligned}$$

Notice these values are independent of satellite parameters, so we can calculate them once and place them in an external file.

Atmospheric Models

To gain an idea of the importance of modeling atmospheric parameters for astrodynamics, consider Fig. 7-8. The complexity of the atmosphere is apparent from the sheer number of

TABLE 7-3. Gaussian Geomagnetic Coefficients. Example Schmidt coefficients are listed for an 8×8 field. Consider these values valid for 1980–1985. [Jursa (1985, 4-28)].

l	m	$\bar{g}_{l,m}$	$\bar{h}_{l,m}$	l	m	$\bar{g}_{l,m}$	$\bar{h}_{l,m}$	l	m	$\bar{g}_{l,m}$	$\bar{h}_{l,m}$	l	m	$\bar{g}_{l,m}$	$\bar{h}_{l,m}$
l	m	dg/dt	dh/dt	l	m	dg/dt	dh/dt	l	m	dg/dt	dh/dt	l	m	dg/dt	dh/dt
1	0	-29 988		4	2	398	-257	6	2	42	93	7	6	11	-23
1	1	-1957	5606	4	3	-419	53	6	3	-192	71	7	7	-2	-10
2	0	-1997		4	4	199	-298	6	4	4	-43	8	0	20	
2	1	3028	-2129	5	0	-219		6	5	14	-2	8	1	7	7
2	2	1662	-199	5	1	357	46	6	6	-108	17	8	2	1	-18
3	0	1279		5	2	261	149	7	0	70		8	3	-11	4
3	1	-2181	-335	5	3	-74	-150	7	1	-59	-83	8	4	-7	-22
3	2	1251	271	5	4	-162	-78	7	2	2	-28	8	5	4	9
3	3	833	-252	5	5	-48	92	7	3	20	-5	8	6	3	16
4	0	938		6	0	49		7	4	-13	16	8	7	7	-13
4	1	783	212	6	1	65	-15	7	5	1	18	8	8	-1	-15
<hr/>															
1	0	22.4	--	4	2	-8.2	1.6	6	2	3.4	-1.4	7	6	0.1	-0.1
1	1	11.3	-15.9	4	3	-1.8	2.9	6	3	0.8	0.0	7	7	0.0	1.1
2	0	-18.3	--	4	4	-5.0	0.4	6	4	0.8	-1.6	8	0	0.8	--
2	1	3.2	-12.7	5	0	1.5	--	6	5	0.3	0.5	8	1	-0.2	-0.1
2	2	7.0	-25.2	5	1	0.4	1.8	6	6	-0.1	0.0	8	2	-0.3	-0.7
3	0	0.0	--	5	2	-0.8	-0.4	7	0	-1.0	--	8	3	0.3	0.0
3	1	-6.5	0.2	5	3	-3.3	0.0	7	1	-0.8	-0.4	8	4	-0.8	-0.8
3	2	-0.7	2.7	5	4	0.2	1.3	7	2	0.4	0.4	8	5	-0.2	0.2
3	3	1.0	-7.9	5	5	1.4	2.1	7	3	0.5	0.2	8	6	0.7	0.2
4	0	-1.4	--	6	0	0.4	--	7	4	1.6	1.4	8	7	-0.3	-1.1
4	1	-1.4	4.6	6	1	0.0	-0.5	7	5	0.1	-0.5	8	8	1.2	0.8

regimes. In addition, although values are shown for temperature and altitude, they all change over time and are very difficult to predict.

Overall, atmospheric models are either *static* or *time-varying*. Of course, the *static* models are the simplest to use because we assume all the atmospheric parameters remain constant. Yet, some factors affect even static models:

1. **Latitudinal variations:** These effects are easiest to visualize with a satellite in a circular, inclined orbit. Passing over the Earth's equatorial bulge effectively changes the actual altitude and, therefore, the density. This changes the drag at virtually every point on the orbit.

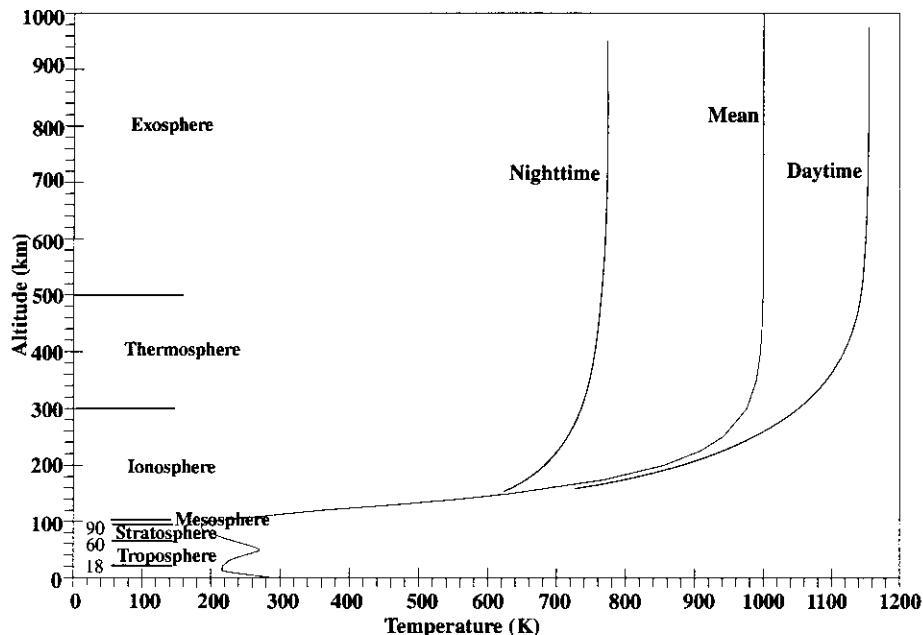


Figure 7-8. Earth's Atmosphere. The Earth's atmosphere is very complex. Notice how rapidly the temperature changes. The bands shown above are general areas where the atmosphere has similar properties. Use this figure carefully—values can change drastically over time.

2. **Longitudinal variations:** Although we consider these effects mainly with time-varying models [specifically the diurnal (daily) effect], they can also influence static models. Consider the difficulty in determining a completely symmetrical atmosphere to conform to every region on the Earth. Mountain ranges in the Himalayas are very different from the vast expanses of the open ocean at the same latitude but different longitudes.

Additional difficulties appear in the *time-varying* models as we consider more real-world effects and drastically increase computations to model them. For example, we must model

1. **Diurnal variations:** These variations occur every day as the Earth rotates. A pronounced bulge (and therefore a density maximum) lags the general direction of the Sun, where the atmosphere is warmest. The bulge centers on meridians where the local time is 2–2:30 P.M. It centers on the equator at the equinoxes but moves to higher latitudes depending on the Sun's declination, which varies primarily due to the obliquity of the ecliptic. Because of the diurnal bulge, the atmospheric density depends on latitude, local time, and time of

year (de Lafontaine, 1986). A minimum value also occurs approximately opposite the bulge at 4:00 A.M. or so each day.

2. **27-day solar-rotation cycle:** This effect comes from the Sun's rotational period and causes a fluctuation which is correlated with the solar decimetric-wavelength radio flux. Irregular changes to the solar flux are related to the growth and decay of active solar regions. These active regions not only have many different patterns of growth, stability, and decay but also cause an uncertain cyclical pattern because they rotate with the Sun over its 27-day period. These patterns are difficult to predict, and it's also difficult to determine how they heat the Earth's atmosphere. This uncertainty affects the accuracy of solar-flux predictions (Nostrand, 1984).
3. **11-year cycle of Sun spots:** The Sun-spot cycle strongly varies the amount of incoming solar radiation. It's actually a 22-year cycle if we include all variations. At the height of the cycle, this effect can create a larger disturbance through solar-radiation pressure than through drag, depending on the satellite's altitude.
4. **Semi-annual variations:** These variations last about six months and are related to the varying distance of the Earth from the Sun and the Sun's latitude during the year. They're usually small.
5. **Cyclical variations:** An 11-year cycle roughly parallels but lags a few years behind the Sun-spot cycle. The minimum of this cycle isn't halfway between the two maxima because the recovery from a maxima is about 6–7 years. The magnitude of the actual peak of the 11-year cycle varies from cycle to cycle. The exact cause is unknown but likely is related to Sun-spot activity.
6. **Rotating atmosphere:** The atmosphere rotates with the Earth to some extent, causing additional time-varying changes in the density. The velocity is usually larger closer to the Earth's surface due to increased friction.
7. **Winds:** Accounting for atmospheric weather patterns and the effect of atmospheric winds on the density is very difficult and extremely complex. They cause temperature variations and therefore changes in density. The effect of actual weather systems is less well-known. Much work remains if we are to fully understand the dynamics of the upper atmosphere.
8. **Magnetic-storm variations:** Usually, fluctuations in the Earth's magnetic field slightly affect the atmosphere, but the effect can be large when geomagnetic activity increases.
9. **Irregular short-periodic variations:** These effects are also small and are often associated with transient geomagnetic disturbances. Included in this section are random solar flares and other very small effects, such as variation in hydrogen currents within the atmosphere.
10. **Tides:** also cause very small variations in the atmospheric density.

The density of the upper atmosphere changes because of a complex interaction between three basic parameters: the nature of the atmosphere's molecular structure, the incident solar flux, and geomagnetic (auroral) interactions. Molecular composition strongly affects the density, so we'll look at it shortly. **Solar flux**, or the incident radiation arriving from the Sun, affects the atmospheric density through nearly instantaneous heating from *Extreme Ultraviolet Radiation (EUV or F_{EUV})*. Geomagnetic activity affects the atmosphere through delayed heating of atmospheric particles from collisions with charged energetic particles from the Sun. Together, these effects increase atmospheric density at higher altitudes by increasing particle collisions. The level of solar flux and geomagnetic activity are difficult to predict but are very important to precise models of the atmosphere.

The gravitational attraction of molecules in the atmosphere mainly determine its pressure and density. The development of both the static and time-varying models relies on a few basic hydrostatic principles which model atmospheric effects. The ideal-gas law relates the *absolute pressure*, p_o , the *mean molecular mass of all atmospheric constituents*, M , the *acceleration due to gravity*, g_o , the *universal gas constant*, R , and the *absolute temperature*, T (Kelvin):

$$\rho = \frac{p_o M}{g_o R T} \quad (7-28)$$

The linkage with temperature is important because it causes much of the difficulty in determining an exact model for the density. The Earth's rotation exposes the atmosphere to the Sun, and the resulting solar heating affects density.

The second important relation is the hydrostatic equation, which relates the change in pressure, Δp , to the density, gravity, and change in altitude, Δh :

$$\Delta p = -\rho g \Delta h \quad (7-29)$$

I won't develop the density equations and associated relations here; see Baker and Makemson (1967, 210–213) and Escobal ([1968] 1979, 18–25) for further information.

Because the Sun's overall activity determines most effects on the atmosphere, the terms *solar maxima* and *solar minima* are often used to refer to its total output. During periods of solar minima, the incoming radiation is less than normal and only slightly influences most satellite orbits. But periods of solar maxima can produce large perturbations in many satellites, and they are now unpredictable.

The contribution of solar flux to atmospheric density is mainly from incoming solar radiation. **Solar flux** (the *Extreme Ultra-Violet, EUV*, radiation that heats the upper atmosphere, F_{EUV}) is impossible to measure at the Earth's surface because the atmosphere doesn't allow transmission of *EUV* radiation. There are no space-borne systems to measure *EUV* flux nor atmospheric-density models to use these space-based measurements if they become available. However, scientists have determined that both *EUV* and incoming solar radiation with a wavelength of 10.7 cm, $F_{10.7}$, originate in the same layers of the Sun's chromosphere and corona. Because the Earth's atmosphere is transparent to the $F_{10.7}$ radiation, we can infer the F_{EUV} measurements from the Earth-based measurements

of 10.7 cm-length radio waves. Regular measurements of $F_{10.7}$ exist from 1940. It's measured in *Solar Flux Units*, SFU, where

$$1 \text{ SFU} = 1 \times 10^{-22} \frac{\text{watt}}{\text{m}^2 \text{ Hz}} \quad (7-30)$$

Typical values range from less than 70 to more than 300 SFU (Nostrand, 1984). The most commonly accepted measurement of $F_{10.7}$ is distributed daily by the National Oceanic and Atmospheric Administration (NOAA) at the National Geophysical Data Center in Boulder, Colorado. Measurements are routinely made at the Algonquin Radio Observatory in Ottawa, Ontario, Canada (Solar). *Monthly values* are averaged to produce mean values denoted with a bar, $\bar{F}_{10.7}$. Figure 7-9 shows the values of $F_{10.7}$ over a period of years. Schatten and Sofia (1987) and Schatten (1988) have developed a monthly estimate of $F_{10.7}$ and a_p until 2015. Figure 7-10 shows some of those predictions with some current data.

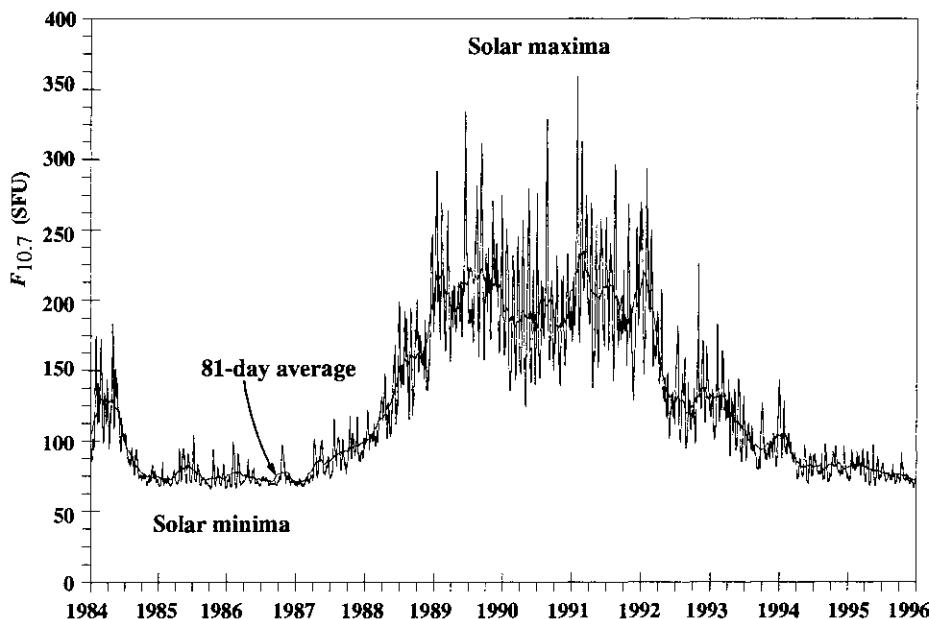


Figure 7-9. Values of Solar Flux. I've shown several years of $F_{10.7}$ values. Remember, $F_{10.7}$ is a radiation that originates at the same place as EUV, but it can't be measured on the Earth's surface. Notice the erratic behavior but a predominant rise during solar max. I've drawn the 81-day average, which we often use in density formulas. It's also very difficult to model.

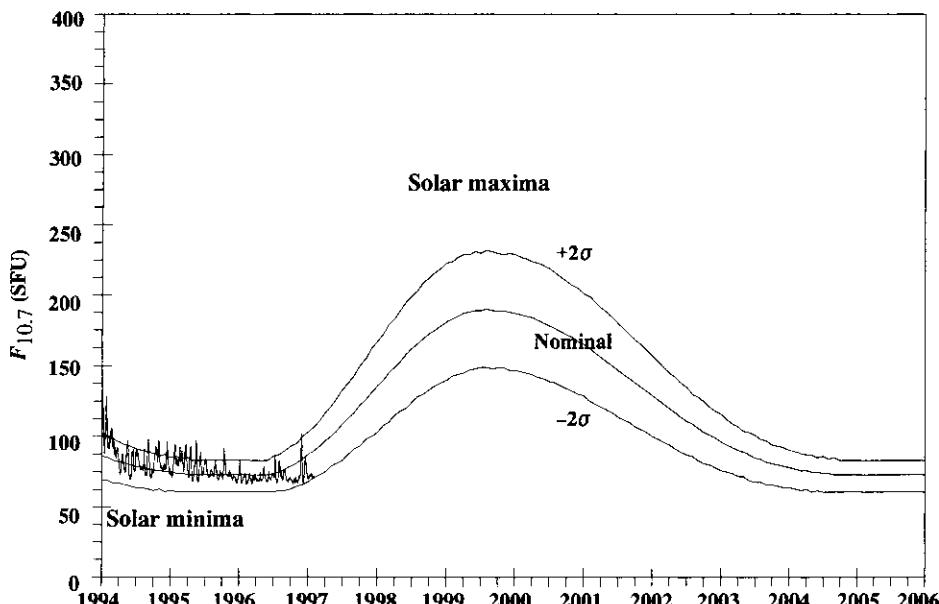


Figure 7-10. Schatten Predicted $F_{10.7}$. Actual data is shown with three Schatten predictions for $F_{10.7}$. When we compare this to actual data in Fig. 7-9, we don't see the individual variations, but the 81-day average is much closer to the nominal prediction. Schatten (1988) also has predictions for early and late occurrences of the solar maxima.

Model Atmospheres

Numerous density models have been developed over the past few decades from two main approaches: 1) combining conservation laws and atmospheric-constituent models into a physical model; 2) using simplified physical concepts developed from in-situ measurements and satellite-tracking data. This book presents both static and time-varying models to find density. Marcos et al. (1993) discuss the development history of many of the representative atmospheric models, as shown in Fig. 7-11.

We should know various models because different applications require differing levels of accuracy. We could always use a time-varying model, like the Jacchia-Roberts atmosphere, but the computational requirements might be too severe. Models of this fidelity represent the most complete and accurate data, but they also demand the highest computational power. On the other hand, the simple exponential model may be too inaccurate for some applications.

Exponential Model (0–1000 km)

This simple, static model assumes the density of the atmosphere decays exponentially from the surface. It also assumes a spherically symmetrical distribution of particles, in which the density, ρ , varies exponentially according to

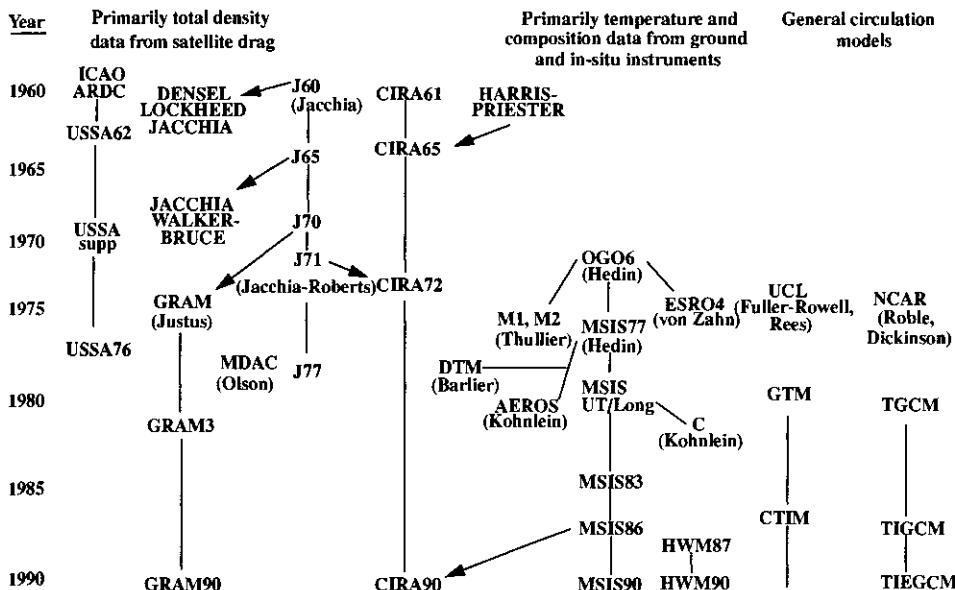


Figure 7-11. Development of Atmosphere Models. Notice the variety of models. Flow of information among the three overall categories is limited (Marcos, et al., 1993, 20). The main models in use today are the Standard Atmosphere, USSA76; variations of the Jacchia-Roberts, J71, J77, and GRAM90; COSPAR International Reference Atmosphere, CIRA90; Mass Spectrometer Incoherent Scatter, MSIS90; Drag Temperature Model, DTM and the general circulation models.

$$\rho = \rho_o \exp \left[-\frac{h_{ellp} - h_o}{H} \right] \quad (7-31)$$

where a *reference density*, ρ_o , is used with the *reference altitude*, h_o , the *actual altitude*, h_{ellp} , and the *scale height*, H . Table 7-4 gives values and descriptions. Although this model approximates much of the atmosphere, it lacks the accuracy and sophistication needed for highly accurate studies. To fix this deficiency, we'd first try to segregate the data into closely related bands. One's choice of bands determines the accuracy; within limits, this approach works for many applications. Extensive tabulated data on atmospheric parameters (density, pressure, temperature, scale height) are in the *U.S. Standard Atmosphere* (1976), together with procedures for calculating values at specified hours of local time, date, and location.

▼ Example 7-4. Calculating Atmospheric Density.

GIVEN: $r = 7125.3489$ km

FIND: ρ

TABLE 7-4. Exponential Atmospheric Model. Although a very simple approach, this method yields moderate results for general studies. Source: Wertz, 1978, 820, which uses the *U.S. Standard Atmosphere* (1976) for 0 km, CIRA-72 for 25–500 km, and CIRA-72 with $T_\infty = 1000$ K for 500–1000 km. The scale heights have been adjusted to maintain a piecewise-continuous formulation of the density.

Altitude h_{ellp} (km)	Base Altitude h_o (km)	Nominal Density ρ_o (kg/m ³)	Scale Height H (km)	Altitude h_{ellp} (km)	Base Altitude h_o (km)	Nominal Density ρ_o (kg/m ³)	Scale Height H (km)
0–25	0	1.225	7.249	150–180	150	2.070×10^{-9}	22.523
25–30	25	3.899×10^{-2}	6.349	180–200	180	5.464×10^{-10}	29.740
30–40	30	1.774×10^{-2}	6.682	200–250	200	2.789×10^{-10}	37.105
40–50	40	3.972×10^{-3}	7.554	250–300	250	7.248×10^{-11}	45.546
50–60	50	1.057×10^{-3}	8.382	300–350	300	2.418×10^{-11}	53.628
60–70	60	3.206×10^{-4}	7.714	350–400	350	9.158×10^{-12}	53.298
70–80	70	8.770×10^{-5}	6.549	400–450	400	3.725×10^{-12}	58.515
80–90	80	1.905×10^{-5}	5.799	450–500	450	1.585×10^{-12}	60.828
90–100	90	3.396×10^{-6}	5.382	500–600	500	6.967×10^{-13}	63.822
100–110	100	5.297×10^{-7}	5.877	600–700	600	1.454×10^{-13}	71.835
110–120	110	9.661×10^{-8}	7.263	700–800	700	3.614×10^{-14}	88.667
120–130	120	2.438×10^{-8}	9.473	800–900	800	1.170×10^{-14}	124.64
130–140	130	8.484×10^{-9}	12.636	900–1000	900	5.245×10^{-15}	181.05
140–150	140	3.845×10^{-9}	16.149	1000–	1000	3.019×10^{-15}	268.00

Eq. (7-31) requires knowledge of the actual altitude, found by subtracting the Earth's radius (6378.137 km) from the satellite's given radius ($h_{ellp} = 747.2119$ km). Now, if we use values from Table 7-4, Eq. (7-31) becomes

$$\rho = 3.614 \times 10^{-14} \exp \left[-\frac{747.2119 - 700}{88.667} \right] = 2.1219854 \times 10^{-14} \frac{\text{kg}}{\text{m}^3}$$

▲ Note that the units in the exponential cancel (all are km), and the result is less than the base value at 700 km, as we would expect.

We can also determine pressure through a similar formula that approximates pressure at any altitude. The formula is

$$p = p_{SL} \exp^{-\frac{h_{ellp}}{H}}$$

where the variables are similar to the density formula. The *scale height*, H , is 7.0104 km, and the *height above the ellipsoid* is h_{ellp} . The sea-level pressure, p_{SL} , which is 1.0133 N/m², lets us find the pressure, p , above sea level, usually in N/m².

Standard Atmosphere (0–1000 km)

The Standard Atmosphere has been published twice (1962 and 1976) in an attempt to provide a standard reference model. It is an ideal, steady-state model of the Earth's atmosphere at a latitude of 45°N during moderate solar activity. The U.S. Committee on Extension to the Standard Atmosphere (COESA) has adopted two definitions that show its intent:

A hypothetical vertical distribution of atmospheric temperature, pressure and density which, by international agreement, is roughly representative of year-round, midlatitude conditions. Typical usages are as a basis for pressure altimeter calibrations, aircraft performance calculations, aircraft and rocket design, ballistic tables, and meteorological diagrams. The air is assumed to obey the perfect gas law and the hydrostatic equation which, taken together, relate temperature, pressure and density with geopotential. Only one standard atmosphere should be specified at a particular time and this standard atmosphere must not be subjected to amendment except at intervals of many years. (US Standard, 1976, xiv)

Satellites operating at much higher altitudes required the Committee to adopt another goal for the Standard Atmosphere:

This atmosphere shall also be considered to rotate with the Earth, and be an average over the diurnal cycle, semi-annual variation, and the range of conditions from active to quiet geomagnetic, and of active to quiet sunspot conditions. Above the tropopause (about 110 km) generalized forms of the hydrostatic equations apply. (US Standard, 1976, xiv).

CIRA (25–2500 km)

The Committee on Space Research (COSPAR) of the International Council of Scientific Unions periodically determines an atmospheric model. The current version is the COSPAR International Reference Atmosphere (CIRA-90) model. The first model was produced in 1965 (CIRA-65), which was basically a new model for altitudes from 30–300 km, based on Champion (1963) and the Harris-Priester (1962) model from 120–800 km. CIRA-72 included mean values from 25–500 km, and Jacchia (1971) prepared models from 110–2000 km. The model is a semi-theoretical technique, but it does have some free variables. Data comes mainly from measurements of satellite drag and ground-based measurements.

Harris-Priester (120–2000 km)

This model is basically static, but several tables account for the densities we may observe within the solar cycle. Interpolation then determines the density at a particular time. This model is recommended for comparing propagation algorithms because it will give fairly accurate results and is computationally efficient. The model was part of CIRA-65, but it

continues to be useful separately for simulations requiring an estimate of density during different regions of the solar cycle.

The Harris-Priester model determines the physical properties of the upper atmosphere by averaging several of the predominant variations. They include approximations for fluxes from the extreme ultraviolet but average the semiannual and seasonal latitudinal variations. Harris-Priester doesn't model the effects of the 27-day solar rotation. Long et al. (1989, 4-57) lists "discrete values of density for the maximum- and minimum-density profiles [shown in Table 7-5]... [These values] correspond to the mean solar activity and are stored in tabular form as $\rho_{min}(h_i)$ and $\rho_{max}(h_i)$, respectively. Different maximum and minimum profiles correspond to different levels of solar activity."

Jacchia-Roberts (70–700 km)

Jacchia's model is probably the most popular of the complex atmospheric models used in astrodynamics. It contains analytical expressions for determining exospheric temperature as a function of position, time, solar activity, and geomagnetic activity. With a computed temperature in hand, we get density from empirically determined temperature profiles or from the diffusion equation.

One advantage of the Jacchia model over CIRA-65 is that it applies over all latitudes. NASA used the Jacchia model in preference to CIRA-65 but employed the diffusion equation instead of Jacchia's numerical approximation to eliminate extensive tables. Jacchia developed this approach in his 1970 models of the upper atmosphere (Jacchia, 1970). These empirically defined models represent atmospheric density as a function of exospheric temperature and altitude. Jacchia divided the upper atmosphere into two areas: 90–125 km and above 125 km. He determined this difference by assuming that mixing dominates between 90 and 100 km, with a fixed-boundary atmospheric condition at 90 km. In addition, Jacchia assumed diffusive equilibrium above 100 km. This led to using a low-altitude temperature profile in the diffusion differential equation between 100 and 125 km and a high-altitude temperature profile above 125 km. Jacchia solved these differential equations by numerically integrating them over various constant values of exospheric temperature. As a point of interest, the procedure routinely uses an 81-day average of solar parameters. At first glance, this may seem arbitrary, but it's necessary to average out differences caused by the solar-rotation cycle (27 days), so we have a three-period cycle. Jacchia tabulated these results for use in atmospheric-drag simulations through interpolation (Jacchia, 1970). If we analyze historical data, the 81-day average is easy to use, but future studies need some estimate such as the reasonably accurate Schatten approximation.

Roberts (1971) recognized that tabular determination of atmospheric drag and numerical integration to calculate partial derivatives for drag is computationally intensive. So, in 1971, he analytically evaluated the 1970 Jacchia models. Roberts uses partial fractions to integrate values between 90 km and 125 km. For altitudes above 125 km, he introduces a different asymptotic function than the one Jacchia introduced to achieve an integrable form. With this substitution, Roberts closely approximates Jacchia's results above 125 km. Although he based his analytical equations on the 1970 Jacchia model (before Jacchia

TABLE 7-5. Tables of Density by Altitude. The Harris-Priester model uses data tables of this type. Long, et al. (1989, 4-58)

Height (km)	Minimum Density (kg/m ³)	Maximum Density (kg/m ³)	Height (km)	Minimum Density (kg/m ³)	Maximum Density (kg/m ³)
100	4.974×10^{-7}	4.974×10^{-7}	420	1.558×10^{-12}	5.684×10^{-12}
120	2.490×10^{-8}	2.490×10^{-8}	440	1.091×10^{-12}	4.355×10^{-12}
130	8.377×10^{-9}	8.710×10^{-9}	460	7.701×10^{-13}	3.362×10^{-12}
140	3.899×10^{-9}	4.059×10^{-9}	480	5.474×10^{-13}	2.612×10^{-12}
150	2.122×10^{-9}	2.215×10^{-9}	500	3.916×10^{-13}	2.042×10^{-12}
160	1.263×10^{-9}	1.344×10^{-9}	520	2.819×10^{-13}	1.605×10^{-12}
170	8.008×10^{-10}	8.758×10^{-10}	540	2.042×10^{-13}	1.267×10^{-12}
180	5.283×10^{-10}	6.010×10^{-10}	560	1.488×10^{-13}	1.005×10^{-12}
190	3.617×10^{-10}	4.297×10^{-10}	580	1.092×10^{-13}	7.997×10^{-13}
200	2.557×10^{-10}	3.162×10^{-10}	600	8.070×10^{-14}	6.390×10^{-13}
210	1.839×10^{-10}	2.396×10^{-10}	620	6.012×10^{-14}	5.123×10^{-13}
220	1.341×10^{-10}	1.853×10^{-10}	640	4.519×10^{-14}	4.121×10^{-13}
230	9.949×10^{-11}	1.455×10^{-10}	660	3.430×10^{-14}	3.325×10^{-13}
240	7.488×10^{-11}	1.157×10^{-10}	680	2.620×10^{-14}	2.691×10^{-13}
250	5.709×10^{-11}	9.308×10^{-11}	700	2.043×10^{-14}	3.325×10^{-13}
260	4.403×10^{-11}	7.555×10^{-11}	720	1.607×10^{-14}	1.779×10^{-13}
280	2.697×10^{-11}	5.095×10^{-11}	760	1.036×10^{-14}	1.190×10^{-13}
290	2.139×10^{-11}	4.226×10^{-11}	780	8.496×10^{-15}	9.776×10^{-14}
300	1.708×10^{-11}	3.526×10^{-11}	800	7.069×10^{-15}	8.059×10^{-14}
320	1.099×10^{-11}	2.511×10^{-11}	840	4.680×10^{-15}	5.741×10^{-14}
340	7.214×10^{-12}	1.819×10^{-11}	880	3.200×10^{-15}	4.210×10^{-14}
360	4.824×10^{-12}	1.337×10^{-11}	920	2.210×10^{-15}	3.130×10^{-14}
380	3.274×10^{-12}	9.955×10^{-12}	960	1.560×10^{-15}	2.360×10^{-14}
400	2.249×10^{-12}	7.492×10^{-12}	1,000	1.150×10^{-15}	1.810×10^{-14}

published his 1971 models), we can still use his corrections with Jacchia's later model. Draper Laboratory uses this approach and has corrected an error in the $W(v)$ function in Equation 12 of Roberts' 1971 paper (Long et al. 1989, 4-51). According to Long et al. (1989, 4-36), the Jacchia-Roberts density deviates no more than 6.7% from associated Jacchia values. Sec. B.1 of the Appendixes details the formulas for the Jacchia-Roberts atmo-

sphere. Oza and Freierstag (1995) suggest that the Jacchia-Roberts model is very responsive to changes in geomagnetic activity. Partial derivatives are also available for the Jacchia-Roberts model in Long et al. (1989, 4-54 to 4-57).

DTM (200–1200 km) and MSIS (0–2000 km)

These two additional models are based on air-glow temperatures (DTM — Barlier, et al., 1978) and incoherent radar scatter (MSIS — Hedin, et al., 1977), respectively. Barlier uses spherical harmonics to incorporate data on satellite drag from over two complete solar cycles, and significant observational data based on the Thuillier et al. (1977) model of global exospheric temperature. An advantage is the smaller amount of code required for execution (compared to Jacchia-Roberts, for instance). Although both models were developed after the Jacchia-Roberts models and incorporate data which was unavailable, Oza and Freierstag (1995) note that a modified form of Jacchia-Roberts still can out-perform these models in some applications. They suggest that the DTM model is more sensitive to changes in the solar flux activity. They also have found that the 81-day $F_{10.7}$ average should be centered on the day of interest, rather than at the end of the cycle.

Soviet Cosmos (160–600 km)

A Soviet density model (Voiskovskii, 1973, Boelitz, 1996, and Amelina et al. 1996) is an analytical method to obtain atmospheric density in an aspherical upper atmosphere from observations of Soviet Cosmos satellites. The elegant part of this algorithm is that it can turn factors affecting atmospheric density *on* or *off* by simply omitting certain k factors. (For example, if diurnal variations aren't needed, set k_2 to 1.) The variability in the Soviet model gives a 1σ prediction error similar to the position error of actual data calculated with a Harris-Priester density model (Carter et al., 1987). The Soviet model differed only 5–10% (Volkov, 1982) from the Jacchia density model of 1971.

The model emerges empirically from observations of the Cosmos satellites' orbital motion. It includes the dependence of the density on solar flux and geomagnetic activity as well as the diurnal and semi-annual density variations. This model is valid for satellites at altitudes of 160–600 km. Appendix B further discusses this model.

7.6.3 Third-Body Perturbations

In Chap. 2, we examined multiple bodies and their effects on satellites from the standpoint of determining orbital motion. We now turn to the *perturbing* effects that other bodies have on satellites. Third bodies, such as the Sun or Moon, mainly affect satellites above low-Earth orbits. Their effects become noticeable about when the effects of drag begin to diminish. Because the main cause of perturbations from a third body (Sun and Moon) is the gravitational attraction, which is conservative, it's reasonable to use a disturbing-function solution. We can also develop the forces. I'll present both forms in this section so you can use them in different systems.

Accelerations Due to Third Body

I developed the equations of motion for a three-body system in Eq. (2-27). (Let the third body be 3 and assume the mass of the satellite is negligible.)

$$\ddot{\vec{r}}_{\oplus sat} = -\frac{Gm_{\oplus} \vec{r}_{\oplus sat}}{r_{\oplus sat}^3} + Gm_3 \left(\frac{\vec{r}_{sat3}}{r_{sat3}^3} - \frac{\vec{r}_{\oplus 3}}{r_{\oplus 3}^3} \right) \quad (7-32)$$

At this point, we can numerically integrate to determine the third-body's effect on the satellite; however, careful examination of Eq. (7-32) reveals a potential difficulty. Suppose the disturbing third body is the Sun. The distance from the satellite to the Sun (in km) and the distance from the Earth to the Sun (also in km) are very similar, and in the final term of the acceleration, the cube of each of these distances is in the denominator. Even with canonical units, each value will be very small; their numerically computed difference is even smaller and may introduce errors during a simulation. Long et al. (1989, 4-8) say this numerical difficulty is not a problem for Earth satellites, but it can be a problem around other central bodies and may produce significant errors for the Moon. Geyling and Westerman (1971, 113) give us a solution using a Taylor series expansion and neglect small terms because $r_{\oplus sat} \ll r_{\oplus 3}$. The final result is

$$\ddot{\vec{r}}_{\oplus sat} = -\frac{Gm_{\oplus} \vec{r}_{\oplus sat}}{r_{\oplus sat}^3} - \frac{Gm_3}{r_{\oplus 3}^3} \left(\vec{r}_{\oplus sat} - 3\vec{r}_{\oplus 3} \frac{\vec{r}_{\oplus sat} \cdot \vec{r}_{\oplus 3}}{r_{\oplus 3}^2} \right) \quad (7-33)$$

Although this form is numerically stable, it's truncated and therefore less accurate. When the third body is the Moon, the distances from Earth to satellite and Earth to Moon are much closer than that from Earth to Sun. We may need more terms from the Taylor series to preserve the accuracy. Neta (1996) shows the next term in the series:

$$-15 \left(\frac{\vec{r}_{\oplus sat} \cdot \vec{r}_{\oplus 3}}{r_{\oplus 3}^2} \right)^2 \vec{r}_{\oplus 3}$$

We can also express the modified form of the equations of motion using Legendre functions [Long et al. (1989, 4-8 to 4-9)]. If we operate on the direct term [left-hand term of the bracketed relation in Eq. (7-32)], we can find the magnitude of r_{sat3} using the cosine law plus the sides of the triangle formed by the Sun, satellite, and center of the inertial coordinate system.

Let's designate ζ as the angle between the third body and satellite as seen from the Earth. If so,

$$r_{sat3}^2 = r_{\oplus sat}^2 + r_{\oplus 3}^2 - 2r_{\oplus sat}r_{\oplus 3}\cos(\zeta)$$

Letting

$$h = \frac{r_{\oplus sat}}{r_{\oplus 3}} \quad B = \sum_{j=1}^{\infty} P_j [\cos(\zeta)] h^j$$

we have a Legendre function of

$$\frac{1}{r_{sat3}} = \frac{1}{r_{\oplus 3}} \{ P_0 [\cos(\zeta)] + P_1 [\cos(\zeta)] h + \dots \} = \frac{1+B}{r_{\oplus 3}}$$

We can substitute this into Eq. (7-32) and obtain the alternate form for the acceleration. We can neglect the satellite's mass, so

$$\ddot{\vec{r}}_{\oplus sat} = -\frac{Gm_{\oplus} \vec{r}_{\oplus sat}}{r_{\oplus sat}^3} + Gm_3 \left\{ \vec{r}_{sat3} \frac{(1+B)^3}{r_{\oplus 3}^3} - \frac{\vec{r}_{\oplus 3}}{r_{\oplus 3}^3} \right\}$$

By expanding $(1+B)^3$ and recognizing that $\vec{r}_{sat3} - \vec{r}_{\oplus 3} = -\dot{\vec{r}}_{\oplus sat}$, we have

$$\ddot{\vec{r}}_{\oplus sat} = -\frac{Gm_{\oplus} \vec{r}_{\oplus sat}}{r_{\oplus sat}^3} - Gm_3 \frac{-\dot{\vec{r}}_{sat3} (3B + 3B^2 + B^3) + \vec{r}_{\oplus sat}}{r_{\oplus 3}^3} \quad (7-34)$$

Now, let's generalize this expression to multiple disturbing third bodies, $k = 1, 2, \dots, n$. Thus,

$$\ddot{\vec{r}}_{\oplus sat} = -\frac{Gm_{\oplus} \vec{r}_{\oplus sat}}{r_{\oplus sat}^3} - \sum_{k=1}^n \frac{Gm_k}{r_{\oplus k}^3} (\vec{r}_{\oplus sat} - \beta_k \vec{r}_{satk})$$

where $\beta_k = 3B_k + 3B_k^2 + B_k^3$ because $B_k = B(\zeta_k)$. This form eliminates the numerical difficulty in Eq. (7-32) when $r_{\oplus k} \approx r_{satk}$, but only at the expense of introducing an infinite series in the formulation. Fortunately, this series converges rapidly for those satellites where $r_{\oplus sat} \ll r_{\oplus 3}$, which is the case for most satellite missions. Thus a few terms suffice for most practical applications.

Partial Derivatives

We require partial derivatives of the third-body acceleration for orbit-determination and estimation techniques in Chap. 9. Begin with Eq. (7-32) and assume the Sun is the third body.

Consider first the partial derivative with respect to position $\vec{r}_{\oplus sat}$:

$$\begin{aligned} \frac{\partial \ddot{\vec{r}}}{\partial \vec{r}_{\oplus sat}} &= -Gm_{\oplus} \left[\frac{\partial (1/r_{\oplus sat}^3)}{\partial \vec{r}_{\oplus sat}} \vec{r}_{\oplus sat} + \frac{1}{r_{\oplus sat}^3} \frac{\partial \vec{r}_{\oplus sat}}{\partial \vec{r}_{\oplus sat}} \right] \\ &\quad + Gm_{\odot} \left[\frac{\partial (1/r_{sat\odot}^3)}{\partial \vec{r}_{\oplus sat}} \vec{r}_{sat\odot} + \frac{1}{r_{sat\odot}^3} \frac{\partial \vec{r}_{sat\odot}}{\partial \vec{r}_{\oplus sat}} \right] \end{aligned} \quad (7-35)$$

Note that the indirect term ($\dot{r}_{\oplus\odot}/r_{\oplus\odot}^3$) in Eq. (7-32) is independent of the satellite's position and velocity; thus, the partial derivative vanishes. Clearly, we'll need to use the identities $\partial r/\partial \dot{r} = \dot{r}/r$ and $\partial \dot{r}/\partial \dot{r} = I$, where I is the 3×3 identity matrix. Also note that because the acceleration and position vectors both have three components, we expect a 3×3 matrix as the result. Furthermore, in view of the identity, the partial derivative in the first term of Eq. (7-35) will produce a result proportional to $\dot{r}_{\oplus\text{sat}}$ resulting in the apparent product of two vectors $\dot{r}_{\oplus\text{sat}} \dot{r}_{\oplus\text{sat}}^T$. This isn't a dot product. Rather, it's the exterior product. If we assume the usual convention of a column representation for a vector, then the exterior product of two vectors is defined as the product of a $n \times 1$ column vector with $1 \times m$ row vector, which results in an $n \times m$ matrix. We use this notation to represent our result. Evaluation of the partial derivatives in Eq. (7-35) yields

$$\frac{\partial(1/r_{\oplus\text{sat}}^3)}{\partial \dot{r}_{\oplus\text{sat}}} = \frac{-3}{r_{\oplus\text{sat}}^4} \frac{\partial r_{\oplus\text{sat}}}{\partial \dot{r}_{\oplus\text{sat}}} = \frac{-3}{r_{\oplus\text{sat}}^5} \dot{r}_{\oplus\text{sat}}$$

$$\frac{\partial(1/r_{\text{sat}\odot}^3)}{\partial \dot{r}_{\oplus\text{sat}}} = \frac{-3}{r_{\text{sat}\odot}^4} \frac{\partial r_{\text{sat}\odot}}{\partial \dot{r}_{\oplus\text{sat}}}$$

Further evaluation and simplification yield the final result. Notice that, because the acceleration is independent of the velocity, the partial derivatives are zero. I've shown the general form using "k" disturbing bodies.

$$\begin{aligned} \frac{\ddot{r}_{\oplus\text{sat}}}{\dot{r}_{\oplus\text{sat}}} &= \left\{ -\frac{Gm_{\oplus}}{r_{\oplus\text{sat}}^3} - G \sum_{i=1}^k m_i \frac{1}{r_{\text{sat}_i}^3} \right\} I \\ &\quad + 3 \left(\frac{Gm_{\oplus} \dot{r}_{\oplus\text{sat}} \dot{r}_{\oplus\text{sat}}^T}{r_{\oplus\text{sat}}^5} + G \sum_{i=1}^k m_i \frac{\dot{r}_{\text{sat}_i} \dot{r}_{\text{sat}_i}^T}{r_{\text{sat}_i}^5} \right) \\ \frac{\ddot{r}_{\oplus\text{sat}}}{\dot{r}_{\oplus\text{sat}}} &= 0 \end{aligned}$$

7.6.4 Solar-Radiation Pressure

Like drag, solar-radiation pressure is a nonconservative perturbation, but it becomes important at higher altitudes. One of the more difficult aspects of analyzing solar radiation is accurately modeling and predicting the solar cycles and variations. During periods of intense solar storms, this effect may be much larger than all the other perturbations (depending on the altitude); at times of low activity, the effect may be negligible. The

same difficulties arise for cross-sectional area as for drag; however, solar-radiation pressure also requires us to determine the shadowing effect on the spacecraft itself.

To examine the overall effect of solar radiation, we must first develop expressions for the specific force (acceleration) and how it is measured. Because the incoming radiation from the Sun causes a force on the satellite, the apparent size of the satellite that faces the Sun is crucial in accurately determining the amount of acceleration. The pressure is simply the force divided by the incident area exposed to the Sun. This means that the pressure distribution is very critical, and this depends on the satellite's shape and composition (a balloon versus a cannonball, for instance). Incorporating the mass then permits us to determine the acceleration. However, this entire process involves determining the Sun's precise location; the correct orbital attitude; the exact value of the solar-radiation pressure; the effective, time-varying, cross-sectional area exposed to the incoming radiation; and the correct and usually time-varying coefficients to model the satellite's reflectivity. I've mentioned the first two problems in calculating drag and third-body effects.

To arrive at the solar-radiation pressure, begin with the intensity of the energy of the incoming radiation from the Sun. Baker (1967, 186) says that 8×10^{17} photons / cm^2 with $\lambda_{avg} = 5560 \text{ \AA}$ reach the Earth with a frequency, f , at the mean Earth-Sun distance. Each photon has a certain amount of energy, hf , where Planck's constant ($h = 6.6256 \times 10^{-34} \text{ J}\cdot\text{s}$) is used. A *solar-radiation constant*—often called the *intensity*, or the solar flux—is

$$SF = 1353 \text{ W/m}^2 \quad (7-36)$$

Most analyses proceed with this value for the solar constant because determining the actual frequencies and energy is very difficult and varies over time. Many of the same factors affecting atmospheric-density calculations (27-day, diurnal, etc.) also affect the magnitude of solar-radiation pressure. Wertz (1978, 130) gives a time-varying formula instead of the constant value, accounting for the variation over a year:

$$SF = \frac{1358}{1.004 + 0.0334 \cos(D_{aphelion})} \text{ W/m}^2 \quad (7-37)$$

where $D_{aphelion}$ is the days from when the Earth is at aphelion. This aphelion point usually occurs about July 4, although it does vary. For instance, in 1996, it was July 6.

We must still determine the pressure, or amount of momentum, imparted. Using Einstein's law relating energy with mass, $E = mc^2$, and with c being the speed of light, we can solve for the momentum, mc , and find

$$mc = \frac{E}{c}$$

which allows us to find the *solar pressure*, p_{SR} , or the change in momentum—the force per area:

$$p_{SR} = \frac{1353 \text{ W} \cdot \text{m}^2}{3 \times 10^8 \text{ m/s}} = 4.51 \times 10^{-6} \frac{\text{W} \cdot \text{s}}{\text{m}^3} = 4.51 \times 10^{-6} \frac{\text{N}}{\text{m}^2}$$

Remember that this value will change if we use more precise values of the solar-radiation intensity, as is possible with Eq. (7-37).

Now we can develop expressions for the force. Using the *reflectivity*, c_R , the *solar-radiation pressure*, p_{SR} , and the *exposed area to the Sun*, A_\odot , let's define

$$F_{SR} = p_{SR} c_R A_\odot \quad (7-38)$$

Notice the symbol for the area facing the Sun has a subscript because it's usually *not* the same as the projected frontal area used in calculations of drag. The *reflectivity*, c_R , is a value between 0.0 and 2.0, which indicates how the satellite reflects incoming radiation. A value of 0.0 means the object is translucent to incoming radiation. No force is transmitted, but there may be some refraction. A value of 1.0 means that all radiation is absorbed, and all the force is transmitted (i.e., it's a black body). Finally, 2.0 indicates that all the radiation is reflected and twice the force is transmitted to the satellite (flat mirror perpendicular to the light source). It turns out that determining c_R properly is extremely difficult. It changes over time and is virtually impossible to predict. For this reason, it's almost always a solution parameter, determined in differential correction (just like the drag coefficient).

Remember that there are two kinds of reflection—*specular* and *diffuse*. We'll consider only specular reflection because analyzing diffuse reflection requires us to have complex three-dimensional models of the satellite. Now consider solar radiation incident on a specular, reflecting ($c_R > 1$), surface element, A_\odot , whose normal makes a *solar-incidence angle*, ϕ_{SR} , with the Sun line (Fig. 7-12).

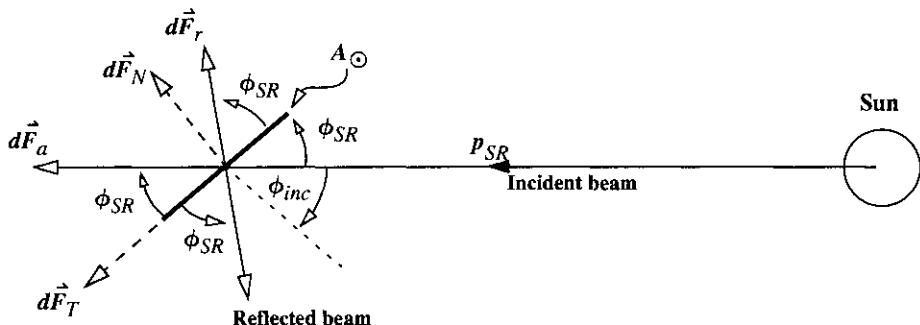


Figure 7-12. Incident Solar Radiation. The incident radiation produces one reflected beam plus reflected, $d\vec{F}_r$, and absorbed, $d\vec{F}_a$, forces. Remember that the incidence, ϕ_{inc} , angles are the same. Therefore the reflected angles, ϕ_{SR} , are also equal. We can also examine the system by looking at the normal, $d\vec{F}_N$, and tangential, $d\vec{F}_T$, force components. The satellite area that's exposed to the Sun is depicted as just a flat area here.

The reflection process is two-fold: absorption based on the reflectivity (c_R), followed by specular reflection. Figure 7-12 permits us to find expressions for the forces imparted by absorption and reflection. Remember that the absorbed and reflected forces aren't nor-

mal, so they don't combine the same as the normal and tangential expressions. The area is always the exposed area to the Sun.

$$\begin{aligned} d\vec{F}_a &= p_{SR} A_\odot \hat{r}_\odot \cos(\phi_{SR}) & d\vec{F}_r &= c_R p_{SR} A_\odot \hat{r}_\odot \cos(\phi_{SR}) \\ d\vec{F}_N &= (d\vec{F}_a + d\vec{F}_r) \cos(\phi_{SR}) = (1 + c_R) p_{SR} A_\odot \hat{r}_\odot \cos^2(\phi_{SR}) \\ d\vec{F}_T &= (d\vec{F}_a - d\vec{F}_r) \sin(\phi_{SR}) = (1 - c_R) p_{SR} A_\odot \hat{r}_\odot \sin(\phi_{SR}) \cos(\phi_{SR}) \end{aligned}$$

We may now begin analyzing the equations of motion. Recognize that we may use any of the formulas presented thus far: Eq. (7-38) or the previous equations for the reflected and normal forces. It's common to assume the surface maintains a constant attitude perpendicular to the Sun, thereby setting $\phi_{SR} = 0^\circ$. Though not very realistic, it does give us rough estimates for the problem of solar radiation. Newton's second law permits us to determine the acceleration with the given force:

$$a_{radiation} = \frac{F}{m} = \frac{p_{SR} c_R A_\odot}{m}$$

Because the direction is always away from the Sun, a unit vector to the Sun from the satellite with a negative sign yields the correct direction. The acceleration is now

$$\dot{\vec{a}}_{radiation} = -\frac{p_{SR} c_R A_\odot}{m} \frac{\hat{r}_{\odot sat}}{|\hat{r}_{\odot sat}|} \quad (7-39)$$

This expression is commonly used for numerical simulations.

Shadow Analysis

Unfortunately, Eq. (7-39) isn't the final solution. It provides a way to determine the acceleration, but the acceleration isn't constant. Most satellites (including geosynchronous) experience periodic eclipses behind the Earth. Of course when the satellite is eclipsed, it's not exposed to solar-radiation pressure. When we need extreme accuracy, we must develop models that turn the solar-radiation calculations "on" and "off," as appropriate, to account for these periods of inactivity. Rather than using a binary switching function, we can analyze the penumbra entry and exit times to increase the accuracy. Escobal ([1965] 1985, 155-162) and Long et al. (1978, 3-28-3-29) show the required formulas.

Begin by examining the geometry of the problem. Figure 7-13 shows the geometry for the Earth shadow. Notice we've assumed the Sun is infinitely far from the Earth, so the light rays are parallel, producing a *cylindrical* Earth shadow of radius R_\oplus .

For the special case when the satellite is at either an entry or exit point, the angular separation with the Sun, ζ , is found using the right triangle included by the sine of $(\zeta - 90^\circ)$. It's more convenient to have the cosine because it ranges from 0° to $\pm 180^\circ$. The dot prod-

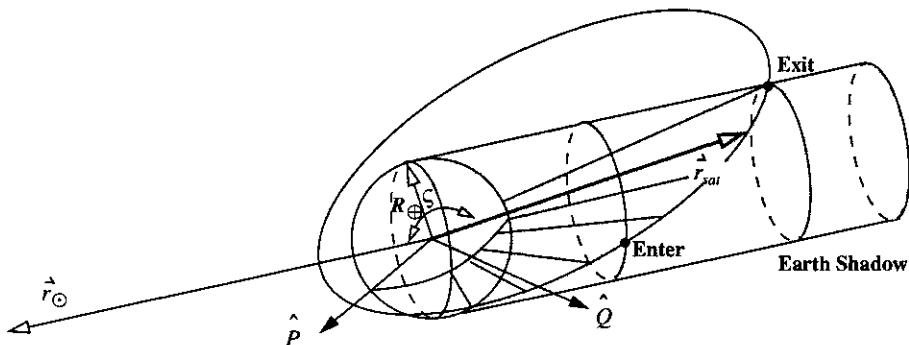


Figure 7-13. Entry-Exit Geometry for Solar Radiation. The geometry of the entry and exit points for a satellite depends on the position of the Sun and the satellite. Here we assume that the shadow is a cylindrical cone beginning at a point on the Earth that's perpendicular to the Sun, $\zeta = 90^\circ$.

uct of the Sun vector and the satellite vector provides a means to determine the angular separation at any time:

$$\cos(\zeta) = \frac{\hat{r}_\odot \cdot \hat{r}_{sat}}{r_\odot r_{sat}} \quad (7-40)$$

Ultimately, we need to find the time at which the satellite enters and exits the Earth's shadow. From Chap. 4, this solution involves obtaining the true anomaly at each time. Unfortunately, Eq. (7-40) doesn't include direct information on the satellite's location within its orbit (we probably don't have the satellite's velocity vector at each time to find v). But recall the development of Eq. (4-57) for the f and g functions:

$$\hat{r}_{sat} = x_w \hat{P} + y_w \hat{Q} = r_{sat} \cos(v) \hat{P} + r_{sat} \sin(v) \hat{Q}$$

Notice the presence of the true anomaly and thus a means to relate the satellite's position to the determination of location within the orbit. Thus, we can change Eq. (7-40) to include the true anomaly:

$$\cos(\zeta) = \frac{r_{sat} \cos(v) \hat{r}_\odot \cdot \hat{P} + r_{sat} \sin(v) \hat{r}_\odot \cdot \hat{Q}}{r_\odot r_{sat}}$$

Now let's define temporary parameters to simplify notation:

$$\beta_1 = \frac{\hat{r}_\odot \cdot \hat{P}}{r_\odot} \quad \beta_2 = \frac{\hat{r}_\odot \cdot \hat{Q}}{r_\odot}$$

and write Eq. (7-40) as

$$\cos(\zeta) = \beta_1 \cos(\nu) + \beta_2 \sin(\nu) \quad (7-41)$$

Eq. (7-40) and Eq. (7-41) represent two approaches to determining the angular separation between the Sun and the satellite. If we square both relations, equate, and use the trajectory equation [Eq. (2-14)] for r_{sat} , we find the *shadow function*,

$$S = R_\oplus^2 (1 + e \cos(\nu))^2 + p^2 \{ \beta_1 \cos(\nu) + \beta_2 \sin(\nu) \}^2 - p^2 \quad (7-42)$$

We can show that, for elliptical orbits, the shadow function vanishes only if

$$1 - \left(\frac{R_\oplus}{a(1-e)} \right)^2 < \beta_1^2 < 1 - \left(\frac{R_\oplus}{a(1+e)} \right)^2$$

Unfortunately, the result contains both sine and cosine values of the true anomaly. By converting to cosine terms, we obtain a fourth-order equation in the cosine of the true anomaly (Escobal [1965] 1985, 158).

$$\begin{aligned} \alpha &= \frac{R_\oplus}{p} \\ \alpha_1 &= \alpha^4 e^4 - 2\alpha^2 (\beta_2^2 - \beta_1^2) e^2 + (\beta_1^2 + \beta_2^2)^2 \\ \alpha_2 &= \alpha^4 4e^3 - 4\alpha^2 (\beta_2^2 - \beta_1^2) e \\ \alpha_3 &= \alpha^4 6e^2 - 2\alpha^2 (\beta_2^2 - \beta_1^2) - 2\alpha^2 (1 - \beta_2^2) e^2 \\ &\quad + 2(\beta_2^2 - \beta_1^2) (1 - \beta_2^2) - 4\beta_2^2 \beta_1^2 \\ \alpha_4 &= \alpha^4 4e - 4\alpha^2 (1 - \beta_2^2) e \\ \alpha_5 &= \alpha^4 - 2\alpha^2 (1 - \beta_2^2) + (1 - \beta_2^2)^2 \\ S &= \alpha_1 \cos^4(\nu) + \alpha_2 \cos^3(\nu) + \alpha_3 \cos^2(\nu) + \alpha_4 \cos(\nu) + \alpha_5 \end{aligned} \quad (7-43)$$

Solution of the quartic is analytical and will yield the desired values of true anomaly for entry and exit. We know that the shadow occurs when ζ is greater than 90° , thus $\cos(\zeta)$ in Eq. (7-41) must be less than zero. We can determine if it's an entry or an exit using Eq. (7-42). For example, if the shadow function, S , is changing from negative to positive, it's an entry. Numerical computations show that, typically, only one real root will satisfy the requirements for our problem. If there is more than one root, we need additional logic.

Eq. (7-42) can be solved numerically. This way, we don't introduce spurious solutions. If we let $x = \cos(\nu)$ and set $S = 0$ for our solution, Eq. (7-42) becomes

$$Ax^2 + Bx + Cx\sqrt{1-x^2} + D = 0$$

We could use a Newton-Raphson technique or Halley's method. Using Halley's method,

$$x_{n+1} = x_n + \delta_n = x_n - \frac{f(x_n)}{f'(x_n) - \frac{f(x_n)f''(x_n)}{2f'(x_n)}}$$

Convergence is determined by checking

$$|f(x_n)| \leq \text{tolerance1}$$

$$|x_{n+1} - x_n| \leq \text{tolerance2}$$

Because we chose $x = \cos(\nu)$, we know the solution must be between -1.0 and 1.0 . The initial guess could easily be zero because it's the midpoint for the interval. Other numerical methods exist for determining the roots of this equation. Escobal's method is attractive because it guarantees all solutions. Newton-Raphson iterations only converge to a root.

We can solve the case for an ellipsoidal Earth in various ways. First, we could rigorously determine the sublatitude point at each time using Algorithm 23, but that's inefficient. Escobal ([1965] 1985, 159–160) elects to determine an approximate value of the sublatitude point and replace R_\oplus in Eq. (7-42) with this value.

Umbra-penumbra corrections require us to change the angular-separation angle, ζ . Now we use Newton's method to find the exact values of true anomaly, given the initial estimates of the calculation. Lundberg (1995) suggests that variations resulting from this correction are usually about one minute.

The Earth's motion is important to the process for two reasons. First, the true-anomaly values will yield times at which entry and exit conditions occur, thereby allowing proper adjustments to calculations of solar-radiation pressure. Second, the times will permit an iteration to use more exact locations of the Sun and satellite throughout the series of equations. The time is found using relations of Chap. 4 on Kepler's equation [Eq. (4-7) and Eq. (4-38)].

To determine these periods, we can use several methods ranging from assuming cylindrical shadows behind the Earth to fully accounting for umbral and penumbral regions and associated variations in the incoming solar radiation. The **SIGHT** and **LIGHT** algorithms discussed in Sec. 3.9 are a quick way to determine if the satellite is in view of the Sun; they're usually enough for all but the most complex studies. See Sec. 8.7.1 (**KEPPLER**), which discusses viewing constraints on satellites, for more insight into geometries of eclipses.

In forming Eq. (7-39), we made several assumptions concerning satellite attitude and parameters. Satellites have complex geometries. Some surfaces reflect diffusely; others reflect specularly and with changing aspect toward the Sun. Hence, at any moment, the satellite will experience a net force, *not* along the Sun-satellite vector, plus a net torque. This implies the possibility of a solar sail, in which solar-radiation pressure becomes the “wind” to move the satellite in a “low-thrust-like” maneuver. Because we’re interested in orbital motion, we ignore the torque and make the simplifying assumption that the disturbing acceleration is given by Eq. (7-39), where A_\odot is some average effective cross-section that implicitly incorporates c_R .

7.6.5 Other Perturbations

Many other forces affect a satellite's orbit, but most are very small and are usually neglected. Let's talk about them briefly. For example, thrust is a large perturbation, but unlike the other forms we've discussed, it's planned. Also, tides are becoming more important as computational resources expand and allow us to consider them.

Thrust

We can consider thrust as an acceleration that induces a perturbation to the orbit. We saw this in the continuous-thrust discussion in Chap. 5. Thrust may quickly produce significant changes in any element of the satellite's orbit. Literature on this subject is extensive, and I've included some discussion in Chap. 5. The simplest formulation uses the acceleration caused by the motor's action.

Simply speaking, thrust is a function of the motor's mass-flow rate and the specific impulse of the fuel. Both thrust and mass-flow rate change over time and must be modeled, including the time dependencies. The acceleration is simply the time-varying thrust, divided by the vehicle's time-varying mass:

$$\begin{aligned} F_{thrust} &= I_{sp} \frac{dm}{dt} \\ \ddot{a}_{thrust} &= \frac{\vec{F}_{thrust}(t)}{m_c} \\ \Delta v &= g I_{sp} \text{LN} \frac{m_{init}}{m_{init} - m_{prop}} \end{aligned} \tag{7-44}$$

Now, find the vehicle's current mass, m_c , using

$$m_c = m_o - \int_{t_o}^{t_c} \frac{dm}{dt} (t) dt \tag{7-45}$$

Next, consider the real-world effects of thrust. Actual motor firings are finite in length and don't occur instantaneously. In addition, the thrust profile is never ideal in that ignition and tail-off characteristics diverge from a commanded behavior, and variations in the mass flow rate or specific impulse introduce time-varying thrust anomalies. Typically, throttling controlled by real-time feedback can compensate for these nonideal effects. McClain (1993) describes ways to account for ramped thrust.

Impulsive thrust is sometimes addressed in numerical operations. For short runs—say about five minutes—this approach works and introduces little error in the final solution. The technique uses the numerical-propagation routine to advance the satellite to the time

of the maneuver, and the change in velocity is added to the velocity vector. The numerical propagation continues using the new velocity vector.

Tides

Discussions of tides usually include solid-Earth and ocean tides. Although the two forms are different, they stem from many of the same sources. See Jursa (1985, 23-24 to 23-29), Kaula (1969), and Casotto (1989) for more information.

Solid-Earth tides are the deformations of the Earth due to perturbing forces. Internal forces result from the Earth's interior structure and involve complex models of the motions for the liquid and solid properties of the matter within the Earth. These forces are only just beginning to be explored.

Many authors have described the Earth's tidal potential caused by external sources such as the Sun and Moon. We've seen that the orbits for the Sun and Moon are complex, and this requires us to consider many different periods in the analysis. The periods are related to the Earth's diurnal motion, the Sun's rotation, the Moon's motions, and so on. Each period has a repetitive frequency, so it's common to model the tidal potential as a harmonic expression, similar to the Earth's aspherical potential. The number of harmonics required to accurately model each of the forces varies from about 13 for six months, to more than 500 for periods of several years and longer.

We can't directly observe the tidal potential, but we can derive it from measurements of the vertical and horizontal components of gravity. It's common to analyze the potential as a function of time, which involves determining the equilibrium behavior of a solid, spherical, oceanless Earth. The three parameters used here are the Love numbers, h , k , and l .

Ocean tides. The oceans cause a large change in mass distribution, which we must account for in highly accurate applications. They are also the main reason for the differences between tidal theory and observational measurements. The ocean tides affect: (1) the direct attraction of the water, (2) the physical bending of the Earth's crust due to the mass of the water, and (3) the distortion of the geoid surface resulting from the previous effects. The perturbation effect of the oceans has been understood conceptually for quite some time; however, only recently has the high-speed computer permitted realistic analysis of the perturbation. Developing correct tidal models is complicated because the resonance frequencies of many of the tidal basins, ocean basins, and so forth, are nearly the same as the overall tidal motion. In addition, swells and troughs associated with large currents (like the Gulf Stream) introduce systematic departures from the ideal predictions of tidal models.

The main problem remaining today is the accurate global measurement of tides using tide gauges. Correctly placing these gauges is crucial, but experience has shown that almost any finite set of locations requires undesirable trade-offs. A second major problem is our limited understanding of the world's oceans, especially the current effects and interactions at the coasts.

7.7 Forming Numerical Solutions

Special perturbations (used interchangeably with numerical methods and Cowell's formulation) are the simplest and most straightforward of all the perturbation methods. An overwhelming advantage is the fact that the solution contains all secular and periodic variations introduced by the perturbing forces. Although the modern computer has eliminated many of the computational constraints, obtaining the result usually takes longer than analytical methods, but produces more accurate results. The solution rests with solving the perturbed equation of motion [Eq. (7-3)] for several perturbing forces. To get the total result, add each of the accelerations and numerically integrate. Remember, the acceleration labeled nonspherical refers only to the perturbing acceleration caused by the nonspherical portion of the mass distribution of the central body, so don't forget to include the two-body component.

$$\ddot{\vec{a}} = \frac{\mu}{r^3} \vec{r} + \vec{a}_{\text{nonspherical}} + \vec{a}_{\text{drag}} + \vec{a}_{\text{3-body}} + \vec{a}_{\text{SR}} \quad (7-46)$$

Although this equation looks innocuous, its solution continues to challenge the greatest intellects. To account for all the major perturbations (central body, drag, third body, solar radiation, thrust, other), we must use expressions of the accelerations developed earlier in this chapter. Remember that the brackets for P_{lm} denote the argument of the Legendre polynomial:

$$U = \frac{\mu}{r} \left[1 + \sum_{l=2}^{\infty} \sum_{m=0}^l \left(\frac{R_{\oplus}}{r} \right)^l P_{lm} [\sin(\phi_{gc})] \left\{ C_{lm} \cos(m\lambda) + S_{lm} \sin(m\lambda) \right\} \right]$$

$$U_{\text{2-body}} = \frac{\mu}{r}$$

$$\vec{a}_{\text{nonspherical}} = \nabla(U - U_{\text{2-body}}) = \nabla R$$

$$\vec{a}_{\text{drag}} = -\frac{1}{2} \rho \frac{c_D A}{m} v_{\text{rel}}^2 \frac{\vec{v}_{\text{rel}}}{|\vec{v}_{\text{rel}}|}$$

Letting the subscript "3" represent any third body, we write the disturbing acceleration as

$$\vec{a}_{\text{3-body}} = -\frac{G(m_{\oplus} + m_{\text{sat}})}{r_{\oplus\text{sat}}^3} \vec{r}_{\oplus\text{sat}} + Gm_3 \left(\frac{\vec{r}_{\text{sat}3}}{r_{\text{sat}3}^3} - \frac{\vec{r}_{\oplus3}}{r_{\oplus3}^3} \right)$$

Acceleration from solar-radiation pressure uses the area exposed to the Sun, A_{\odot} .

$$\vec{a}_{\text{SR}} = -\frac{p_{\text{SR}} c_R A_{\odot}}{m} \frac{\vec{r}_{\odot\text{sat}}}{|\vec{r}_{\odot\text{sat}}|}$$

We can also numerically integrate the variation of parameter equations in Chap. 8 [Eq. (8-12), Eq. (8-14), and Eq. (8-24)]. This process is really a special perturbation solution, but I've kept it with the discussion in Chap. 8 because it focuses on orbital elements, rather than position and velocity vectors. Depending on your application, the immediate availability of orbital elements throughout a program simulation may be very beneficial.

I'll examine two approaches for numerical solutions. The simplified approach is useful for academic applications and it illustrates the process. The technique is quite easy to adapt to the Runge-Kutta methods discussed in Sec. 7.5. The complex method is a brief sketch of the steps required to fully implement Eq. (7-46). Actually forming a complex numerical integration routine can take on several forms depending on the number of desired force models.

7.7.1 Application: Simplified Acceleration Model

Quite often a simplified version of each acceleration is useful for applications needing moderate accuracy. In these cases we model each perturbing force but without rigorous accuracy. We'll look at each force and the necessary approximations. The central-body acceleration is simplified by excluding all tesseral terms (and therefore sectorial terms) in the aspherical potential and examining only the first few zonal harmonics. Although **PKE-PLER** (Algorithm 57) does this analytically (for first-order secular effects), numerical integration provides the means to obtain the secular *and* periodic contributions of the gravitational harmonics.

We know that accurately determining several quantities (A_{\odot} , p_{SR} , c_D , c_R) is extremely complex. High-fidelity formulations for each perturbation are computationally intensive. Often, we don't require the utmost accuracy, so lower-fidelity models will suffice. We'll examine an approach that results in a simplified approach for medium-accuracy requirements.

From the aspherical potential [Eq. (7-16)], we use the gradient to determine the necessary accelerations resulting from the central body. The gradient operation produces acceleration components along each axis, so we'll show each component separately. This is actually a special case for J_2 of the general partial derivative shown on page 496. Begin with the disturbing function (subscripts indicate the zonal harmonic under consideration):

$$R_2 = -\frac{3J_2\mu}{2r} \left(\frac{R_{\oplus}}{r} \right)^2 \left(\sin^2(\phi_{gc}) - \frac{1}{3} \right)$$

Taking the position vector in spherical coordinates, vector r in the *IJK* system, and ϕ_{gc} , determine the accelerations for each axis. To illustrate the process, consider just the *I* component of the acceleration vector. Distribute the common term and recognize ($\sin(\phi_{gc}) = r_K/r$) to obtain

$$R_2 = -\frac{3J_2\mu R_{\oplus}^2}{2r^3} \left(\frac{r_K^2}{r^2} \right) + \frac{J_2\mu R_{\oplus}^2}{2r^3}$$

Differentiate to get

$$\frac{\partial R_2}{\partial r_I} = -\frac{3J_2\mu R_{\oplus}^2 r_K^2}{2} \left(-\frac{5(2r_I)}{2r^7} \right) + \frac{J_2\mu R_{\oplus}^2}{2} \left(-\frac{3(2r_I)}{2r^5} \right)$$

Simplify to produce

$$\frac{\partial R_2}{\partial r_I} = -\frac{3J_2\mu R_{\oplus}^2 r_I}{2r^5} \left(-\frac{5r_K^2}{r^2} + 1 \right)$$

You can handle the remaining axes similarly; however, the K component is slightly different because r_K appears explicitly in the disturbing function. The results are

$$\begin{aligned} a_I &= \frac{\partial R_2}{\partial r_I} = -\frac{3J_2\mu R_{\oplus}^2 r_I}{2r^5} \left(1 - \frac{5r_K^2}{r^2} \right) \\ a_J &= \frac{\partial R_2}{\partial r_J} = -\frac{3J_2\mu R_{\oplus}^2 r_J}{2r^5} \left(1 - \frac{5r_K^2}{r^2} \right) \\ a_K &= \frac{\partial R_2}{\partial r_K} = -\frac{3J_2\mu R_{\oplus}^2 r_K}{2r^5} \left(3 - \frac{5r_K^2}{r^2} \right) \end{aligned} \quad (7-47)$$

The perturbations resulting from the third harmonic are developed from the disturbing potential resulting when $l=3$, so

$$R_3 = -\frac{J_3\mu}{2r} \left(\frac{R_{\oplus}}{r} \right)^3 \left\{ 5\sin^3(\phi_{gc}) - 3\sin(\phi_{gc}) \right\}$$

Similar to J_2 , the accelerations for J_3 are

$$\begin{aligned} a_I &= \frac{\partial R_3}{\partial r_I} = -\frac{5J_3\mu R_{\oplus}^3 r_I}{2r^7} \left(3r_K - \frac{7r_K^3}{r^2} \right) \\ a_J &= \frac{\partial R_3}{\partial r_J} = -\frac{5J_3\mu R_{\oplus}^3 r_J}{2r^7} \left(3r_K - \frac{7r_K^3}{r^2} \right) \\ a_K &= \frac{\partial R_3}{\partial r_K} = -\frac{5J_3\mu R_{\oplus}^3}{2r^7} \left(6r_K^2 - \frac{7r_K^4}{r^2} - \frac{3}{5}r^2 \right) \end{aligned} \quad (7-48)$$

Ordinarily, the K component would include r_K in the denominator of the last term, which could cause numerical difficulties. Indeed, *all* satellite orbits pass through the equatorial plane during each revolution—momentarily, as in inclined orbits, or continuously, as in equatorial orbits. The solution is simply to multiply through by r_K to yield the form shown.

Express the perturbations resulting from the fourth harmonic, $l=4$, as

$$R_4 = -\frac{\mu J_4}{8r} \left(\frac{R_{\oplus}}{r} \right)^4 \left\{ 35 \sin^4(\phi_{gc}) - 30 \sin^2(\phi_{gc}) + 3 \right\}$$

and the accelerations as

$$\begin{aligned} a_I &= \frac{\partial R_4}{\partial r_I} = \frac{15 J_4 \mu R_{\oplus}^4 r_I}{8r^7} \left(1 - \frac{14r_K^2}{r^2} + \frac{21r_K^4}{r^4} \right) \\ a_J &= \frac{\partial R_4}{\partial r_J} = \frac{15 J_4 \mu R_{\oplus}^4 r_J}{8r^7} \left(1 - \frac{14r_K^2}{r^2} + \frac{21r_K^4}{r^4} \right) \\ a_K &= \frac{\partial R_4}{\partial r_K} = \frac{15 J_4 \mu R_{\oplus}^4 r_K}{8r^7} \left(5 - \frac{70r_K^2}{3r^2} + \frac{21r_K^4}{r^4} \right) \end{aligned} \quad (7-49)$$

Escobal ([1965] 1985, 50–51) shows the partials for the zonal harmonics (J_5 and J_6):

$$\begin{aligned} a_I &= \frac{\partial R_5}{\partial r_I} = \frac{3 J_5 \mu R_{\oplus}^5 r_I r_k}{8r^9} \left(35 - 210 \frac{r_k^2}{r^2} + 231 \frac{r_k^4}{r^4} \right) \\ a_J &= \frac{\partial R_5}{\partial r_J} = \frac{3 J_5 \mu R_{\oplus}^5 r_J r_k}{8r^9} \left(35 - 210 \frac{r_k^2}{r^2} + 231 \frac{r_k^4}{r^4} \right) \\ a_K &= \frac{\partial R_5}{\partial r_K} = \frac{3 J_5 \mu R_{\oplus}^5 r_k^2}{8r^9} \left(105 - 315 \frac{r_k^2}{r^2} + 231 \frac{r_k^4}{r^4} \right) + \frac{15 J_5 \mu R_{\oplus}^5}{8r^7} \end{aligned} \quad (7-50)$$

$$\begin{aligned} a_I &= \frac{\partial R_6}{\partial r_I} = -\frac{J_6 \mu R_{\oplus}^6 r_I}{16r^9} \left(35 - 945 \frac{r_k^2}{r^2} + 3465 \frac{r_k^4}{r^4} - 3003 \frac{r_k^6}{r^6} \right) \\ a_J &= \frac{\partial R_6}{\partial r_J} = -\frac{J_6 \mu R_{\oplus}^6 r_J}{16r^9} \left(35 - 945 \frac{r_k^2}{r^2} + 3465 \frac{r_k^4}{r^4} - 3003 \frac{r_k^6}{r^6} \right) \\ a_K &= \frac{\partial R_6}{\partial r_K} = -\frac{J_6 \mu R_{\oplus}^6}{16r^9} \left(245 - 2205 \frac{r_k^2}{r^2} + 4851 \frac{r_k^4}{r^4} - 3003 \frac{r_k^6}{r^6} \right) \end{aligned} \quad (7-51)$$

We can simplify drag with three major assumptions: using an exponential atmosphere; ignoring the complex interactions with the winds in the relative-velocity calculations (and hence, the rotating atmosphere); and assuming v_{rel} is the same as the velocity of the satellite. The result is an acceleration which computes quickly but can sacrifice considerable accuracy.

$$\vec{a}_{drag} = -\frac{1}{2} \rho \frac{c_D A}{m} \vec{v} \vec{v}$$

Third-body perturbations depend foremost on the locations of the other bodies (typically the Sun or Moon). Planetary-ephemeris files provide accurate (numerically generated) positions of each body's location, which programs can interpolate during operation. Several alternatives exist for the resulting problems in computer storage. First, we can calculate the position vectors of each body from approximate formulas presented in Sec. 3.5 and Sec. 3.6. A second simplification, for runs of only a few hours of simulation time, could be to store the position vector at the middle time. That's possible because the Sun and Moon will change only a small amount in "short" simulation times.

For example, consider a one-day simulation for both the Sun and the Moon. The geocentric position vector for the Sun changes about 1° (averaged throughout the year), whereas the Moon moves about 12.85° per day. Given these types of results, we may need to redetermine the acceleration at each point along a one-day simulation. On the other hand, a one-month simulation involving only the Sun may require us to rectify far less often. This gives rise to a phenomenon known as *weak time dependence*, in that the calculations will reflect slight variations as the program updates positions of the Sun and Moon. Fonte (1995) shows that, for systems assuming an orbital period on the order of one day (GEO), the weak time dependence of the influence of the lunar motion on the satellite can contribute up to 200 m of error.

A third and final simplifier for drag, useful whenever long predictions are necessary, is a polynomial-fit technique. With this approach, we fit the planetary-ephemeris files with polynomials before the simulation begins. As the simulation runs, the polynomial coefficients are used to generate the position vectors at each required time. This method adds the benefit of reducing storage space while enabling relatively accurate reconstruction of the original planetary-ephemeris files. We often use Chebyshev polynomials to provide a compact, efficient, and accurate representation of the ephemerides.

Solar radiation presents additional challenges to a simplified result for two main reasons. First, the available data for reliably forecasting future solar activity is extremely limited. Second, the method for determining the regular eclipse interval is complex to implement and may ultimately require us to know the Sun's precise location (see Sec. 3.5 and Sec. 7.6.4). This knowledge presents a chance to share data between perturbation sections (third-body and solar-radiation) but doesn't reduce the computational load of the overall solution. Simplified simulations usually don't address shadowing—a fact that slightly degrades accuracy.

7.7.2 Application: Complex Acceleration Model

Forming a complete, highly accurate, special perturbation method is an important activity. With care, patience, and discipline, you can develop an extremely flexible and accurate tool. The basic equation is Eq. (7-46) just as for the simplified approach. The difference is that we rigorously determine each acceleration.

To start, use the aspherical potential in calculating high-precision ephemerides. Recall the form of the aspherical potential from Eq. (7-17). Implementing a model based on the infinite series of the nonspherical gravitational model requires certain decisions. First,

coefficients for a given model are available to a given degree and order (41×41 , 70×70). Computer runtime may keep you from incorporating an entire field. If you use a subset of the field, say a 50×50 subset of a 70×70 field, you will introduce some (very small) differences into the solution.

The coordinate frame is also important for highly accurate studies. You would usually integrate the equations of motion in an inertial frame, like the *ECI* frame of J2000, which would allow you to avoid the complexity of having to account for rotating coordinates as discussed in Chap. 1. But remember that the satellite sees the “true” orientation of the Earth over time. The natural choice for coordinate systems is the true-of-date (with polar motion) system when you *evaluate* the accelerations. Therefore, you must use a coordinate frame that’s consistent with where the satellite is at an instant of time. If you use a system employing a true equator and true equinox of date, you must remember to integrate the equations of motion in the same frame. This may not be desirable for longer simulations because the true-of-date system continually moves. Another option is to evaluate the equations in a true of date with polar motion frame and convert the accelerations to the mean of J2000 frame before integrating.

Now consider the calculations for the associated Legendre functions. The equation [Eq. (7-14)] to determine these functions is sufficient for a few hand calculations but is terribly inefficient for computers. Recursion is an efficient way to compute orthogonal polynomials and other functions. Recursions replace intensive calculations with simpler ones. For instance, Eq. (7-52) uses simple addition and multiplications instead of more costly trigonometric terms. You should see from Eq. (7-14) that subsequent values depend on previously computed values. Long et al. (1989, 4–13) give specific relations:

$$\begin{aligned} P_{t,0}[\alpha] &= \frac{(2t-1)\alpha P_{t-1,0}[\alpha] - (t-1)P_{t-2,0}[\alpha]}{t} & t \geq 2 \\ P_{t,m}[\alpha] &= P_{t-2,m}[\alpha] + (2t-1)\cos(\phi_{gc})P_{t-1,m-1}[\alpha] & m \neq 0, \quad m < t \\ P_{t,t}[\alpha] &= (2t-1)\cos(\phi_{gc})P_{t-1,t-1}[\alpha] & m \neq 0 \end{aligned} \quad (7-52)$$

with the starting values of $(P_{1,1}[\alpha])$ is from Eq. (7-52) when $t = 1$

$$\begin{aligned} P_{0,0}[\alpha] &= 1 \\ P_{1,0}[\alpha] &= \alpha = \sin(\phi_{gc}) \\ P_{1,1}[\alpha] &= \cos(\phi_{gc}) \end{aligned}$$

For computer implementation, remember that m represents the order of the derivative. If m is larger than t , the derivative value is zero because the t th derivative is always one. This affects only $P_{t-2,m}[\alpha]$. We sum t from 2 to the order desired, and then m goes from 0 to t . We’ll see later that resonance (when the Earth’s rotation and the satellite’s motion are commensurate) can cause large deviations in a satellite’s orbit. Thus, we often want to determine this effect. Unfortunately, only a few gravitational terms usually contribute to this effect. Finding all the Legendre polynomials is very inefficient. In these cases, we can

use the recursion relation [Eq. (7-52)] to find a smaller subset of the gravity field (say 12×12) and then use Eq. (7-14) to find the individual terms responsible for the resonance.

We always address the topic of *stability* whenever we use a recursion algorithm. *Stability* is determined by the particular structure of the recurrence relation and the precision of the computation. For example, suppose the recurrence is given by the difference of two nearly equal quantities that are the same to three significant digits. If our precision is ten digits, the first application of the recursion loses five digits of accuracy. Repeated application of this process quickly produces useless results. Now suppose the recurrence is given by a summation rather than a difference. We retain the precision with only a possibility of losing the final digit due to round-off error. The process would corrupt the final results, but only after thousands of applications of the recurrence relation. These examples don't indicate the relations for the associated Legendre functions but illustrate the concept of stability. A clear indication of an unstable recurrence relation is the appearance of small divisors. Bessel functions appear as coefficients in the Fourier series representation of Kepler's equation (see Sec. 4.2.5). The recurrence relation has eccentricity in the denominator of one term. When e is small, this single term dominates all others, effectively eliminating their information content.

It's useful to obtain expressions for the longitude terms:

$$\begin{aligned}\sin(m\lambda) &= 2\cos(\lambda)\sin\{(m-1)\lambda\} - \sin\{(m-2)\lambda\} \\ \cos(m\lambda) &= 2\cos(\lambda)\cos\{(m-1)\lambda\} - \cos\{(m-2)\lambda\} \\ m\tan(\phi_{gc}) &= (m-1)\tan(\phi_{gc}) + \tan(\phi_{gc})\end{aligned}\quad (7-53)$$

Treating drag in highly accurate programs remains difficult, despite the deceptively simple form of the equation of motion. This is because accurate density calculations require a complex atmospheric model. Typically, the more complete atmospheric models require extensive knowledge about the incoming solar radiation and the atmosphere's molecular composition. Although we know these inputs with good precision from observational measurements, forecasting the data through solar cycles and other variations is much less accurate. Nevertheless, real studies for mission analysis and planning typically warrant the more sophisticated model, despite the associated computational burden. For example, an 81-day running average of $F_{10.7}$ is reasonable for fitting past data; however, it becomes a serious limitation and potential source of error when considering mission *planning*. Indeed, the 81-day running average becomes less realistic the greater the extrapolation into the future. In some of these cases, it might be better to use a model such as the Harris-Priester atmosphere (page 511), which attempts to model certain tendencies during discrete intervals of the solar cycle. Schatten's (1988, 1993) prediction estimates for atmospheric parameters may also be of use in reasonably estimating the future. We can still miss the solar activity when selecting a table value but can greatly reduce variations in the real world.

An important formula, even for high-accuracy simulations, is the relative-velocity calculation:

$$\vec{v}_{rel} = \vec{v}_{ECI} - \vec{\omega} \times \vec{r}_{ECI} + \vec{v}_{wind}$$

For ease of implementation, all the vectors should be in an inertial coordinate system (I show *ECI* here and recommend it). Accurate drag analysis also requires us to know the shape of the spacecraft and its time-varying orientation, or *attitude*. Values for the coefficient of drag for various shapes are available in books describing moments of inertia. We often use a simple sphere to represent the spacecraft's shape. The sphere works reasonably well in many cases, except when the satellite is very large, small, or complex. But highly accurate studies show that even slight changes in c_D (1.98 vs. 2.00, for instance) can strongly change the satellite's orbit (up to hundreds of kilometers); thus, we almost always solve for the drag parameter (c_D) or mean-motion rates (\dot{n} , \ddot{n}) during orbit determination. The estimate c_D helps account for the difference in shape from the assumed model. This is only an issue for extremely accurate studies involving large, complex spacecraft with time-varying attitudes. *Macro models* help in that they try to piece together well-known shapes to approximate the satellite's true shape. Getting the satellite's precise orientation is equally difficult and, in many cases, may actually involve commanding and controlling the satellite to receive health and status information from it.

Treatment of third-body accelerations is virtually the same as for the simplified approach, but we usually use a polynomial recovery technique to recreate the planetary ephemerides as closely as possible. Of course, we must use planetary ephemerides (see Sec. 3.7) consistent with the inertial frame we use for integrating the equations of motion.

Lastly, the equations in Sec. 7.6.4 allow us to model solar-radiation pressure. Difficulties arise that are similar to those for the drag accelerations. Remember, now the effective cross section is the satellite's area (A_{\odot}) that is exposed to the Sun and *not* the component normal to the velocity vector. The solar-ephemeris data required to calculate solar-radiation pressure is the same as that for some atmospheric-drag models and for the third body. The acceleration of solar radiation is also very sensitive to the time in which the Earth eclipses the satellite. A cylindrical shadow model is common (see page 617). But more precise determination requires us to analyze the times when the satellite is in the umbra and penumbra regions, uses an elliptical-Earth model, and updates the location of the Sun and Moon to determine precise entrance and exit times.

7.8 Practical Considerations

Overall, propagating orbits with high accuracy is complicated. There are many techniques, and people often select one method over another for the wrong reason. Computer source code is closely guarded and usually evolves as a conglomeration of many different techniques. Thus, even if software were available, making it work still takes a lot of effort. Development from scratch is another option; yet it also has many drawbacks. Its main disadvantages are trying to prepare for all input types, coupling with a differential-correction program, and the availability of suitable test cases for validation. Because we discuss differential-correction in Chap. 9, we'll concentrate on propagator-related issues in this section.

7.8.1 Validating the Propagator

Perhaps the largest misconception in evaluating methods of propagation is what represents *truth*. Determining the true trajectory is closely related to differential correction and involves the discussions of Sec. 9.5.4 and Sec. 9.8. It usually considers either existing or real-world data. Although each type of data has advantages, both also have important difficulties we must address. We must consider *both* types to completely verify a propagation routine.

First, let's introduce some terms that allow us to discuss how to validate propagation algorithms. When we compare to existing data, it's often simulated data from a program, although it may consist of prior real-world measurements. Because any propagator can produce the exact location of the satellite at certain instants of time, we usually use an *ephemeris*, or a set of position and velocity vectors over time. This ensures we get the average behavior of the propagation routine. We refer to tabular values of position and velocity over time for several satellites as *ephemerides*. Highly accurate ephemerides that come from high-fidelity numerical techniques are called *Precision Orbit Ephemerides* (POE). POEs are available for a few satellites such as GFZ, TOPEX, LAGEOS, ETALON 1 and 2, GPS, TDRSS, and a few others.

When we use existing data to validate a propagator, there are two options. If we use the same system, for instance during a program upgrade, we can use existing test cases and benchmarks to check our results. Although this allows us to check the new implementation, any errors in the existing test cases could migrate to the new system, so this should only be used as an initial verification. A necessary extension seeks an independent system to verify results. This technique overcomes the “familiarity” of the original program, and allows an independent evaluation of the propagation routine. An excellent example of independent benchmarks are the sample orbits documented in Metzinger (1993). This document gives various orbits and their expected output from several different computing platforms. The propagation algorithm used by this study is a highly accurate numerical routine with complete force models, and many sources suggest it is accurate to within ± 5 m of the absolute truth. Because this document has been updated over many years, it contains a wealth of information for any validation effort.

Simulated data can solve some of the problems encountered using actual observations, but it still doesn't model all perturbational forces. Using simulated data allows us to develop consistency for orbit determination because we can know exactly how the underlying orbits correspond to the measurements. We can also develop continuous data streams from which we can estimate the orbit. Remember, this is usually *continuous* data for the entire period; it doesn't represent the real world, but it's enough for initial analyses.

The second primary means of validation involves tests with real-world data and satellites. The advent of modern laser-tracking systems permits very accurate analysis of how well a propagation routine models the real world. On first thought, actual observations should be the most accurate means to determine the *truth*, but as shown in the estimation section (Sec. 9.8), this process isn't perfect. In fact, a review of Chap. 9 could lead us to think the truth is unachievable. Fortunately, this isn't the case. As a practical matter, many

satellite orbits (TOPEX, for instance) are known to centimeter accuracy in *post-processing*. The random-measurement error is often negligible when we use our most accurate observations (i.e., laser ranging).

7.8.2 Physical Data and Sources

It's also important to collect physical data for use with any type of perturbation analysis. Data is what we determine and measure empirically or what we determine experimentally based on known physical models. Of course, the most accurate data exists for past events because we can retrieve actual measurements. Also, we often need predicted values of unavailable current or future data to plan satellite missions. This section lists sources for obtaining physical data (basic defining parameters, aspherical-potential coefficients, exact pole positions, and so on); atmospheric parameters (various solar indices and affected parameters); third-body positions; and timing data.

Gravitational Parameters

It's hard to determine correct gravitational coefficients, and having to resolve multiple data sources and satellite classes makes the process even more difficult. Vetter (1994) concisely recounts the development of most commonly used gravitational models, whose use rests on two important facts:

1. We must use the complete theory and associated constants to obtain documented accuracy because mixing constants from different theories creates errors. This complete set includes the four fundamental defining parameters in the table on the inside front cover, the speed of light, masses of the planets, a sea-surface model, etc. *We must use a complete and consistent set of physical constants (as in that table) for any high-accuracy study.*
2. The gravitational coefficients are entirely accurate only *for the particular orbit classes used in their development*. This point is extremely important because constants derived from a single class of satellite orbits will work very well for that same class, but they may diverge considerably for different classes. Thus, we need to use a model derived from many satellite orbits.

In this light, Vetter (1994) recommends the JGM-2 (Nerem et al., 1994) and the newer JGM-3 models because they're derived from many orbit classes. A new effort is underway among the University of Texas at Austin, the Defense Mapping Agency, Ohio State University, and the Goddard Spaceflight Center. The model uses data from 30 satellites and from surface gravity measurements; it will be a complete 360×360 field. These most accurate models are available from the main developers at the University of Texas at Austin and the Goddard Spaceflight Center. See Appendix D.5, "Data Sources," for the complete address. Any of these gravity fields may be *tuned* for a particular orbit, although tuning is uncommon because it's complex.

Related to the Earth's gravitational model is the precise location of the North Pole. Recall from page 82 that the pole precesses and nutates slightly over time. Highly accurate calculations must include corrections for polar motion. The data usually arrives in the form of x_p and y_p displacements from the *Celestial Ephemeris Pole (CEP)*, defined as the axis of rotation which is normal to the true equator of the Earth. The x_p and y_p coordinates represent displacements from the 0° and 270° longitude meridians, respectively. These coordinates are available from several sources. The U.S. Naval Observatory provides access by computer modem to obtain timing and pole-position data, but you need to send requests for the Automated Data Service and the IERS position bulletins.

Atmospheric and Solar Data

The most accepted compilation of the measurements of planetary geomagnetic indices (k_p and a_p) is created by the Institute für Geophysik, at Göttingen University, Germany (Geomagnetic). The actual measurements occur at the Algonquin Radio Observatory in Ottawa, Ontario, Canada. The most commonly accepted measurement of $F_{10.7}$ is distributed daily by the National Oceanographic and Atmospheric Administration (NOAA) at the National Geophysical Data Center in Boulder, Colorado. It's worth noting that NOAA publishes values for (a_p , k_p , and $F_{10.7}$) at weekly intervals and includes 45-day predictions for each value. High-fidelity programs will find NOAA to be a unique source for accurate measurements and predictions. Appendix D.5, "Data Sources," gives their address. For studies discussing the accuracy of predictions, see Eller et al. (1994), Jablonski (1991), and others.

Planetary Ephemerides

The main source for all planetary and lunar ephemerides is the Jet Propulsion Laboratory (JPL). Data is available for the planets (e.g., Development Ephemeris, DE-245) and the Moon (e.g., Lunar Ephemeris, LE-245). Specific details on forming these ephemerides is in Standish (1990, 252–271). JPL creates these ephemerides by numerical integrations accounting for the influence of the planets.

Time

To convert between time systems, we need to know the difference between atomic time and coordinated universal time ($\Delta AT = TAI - UTC$), as well as the difference between universal time and coordinated time ($\Delta UT1 = UT1 - UTC$). Other users may have a particular epoch, such as January 1, 1970, and would therefore require timing coefficients to enable recovery of positions at that particular time. We must be sure to have access to the correct time associated with *all* observational equipment. The broadcast of universal time by the United States National Bureau of Standards (WWV in Fort Collins, Colorado, and WWVH in Hawaii) is integral to this process. Special receivers take the WWV and WWVH signals and place them in electronic systems (computers). See Appendix D.5, "Data Sources," for additional information.

TABLE 7-6. Available Planetary and Lunar Ephemerides. The table shows available ephemerides and when they apply. Lunar ephemerides (LE-###) are produced with each development ephemerides.

Designation	Basis	Year	Start	Stop	Comments
DE-102	B1950	1977	Apr 16, 1410	Dec 22, 3002	Longest span
DE-118	B1950	1980	Dec 9, 1599	Mar 31, 2169	Previous standard
DE-130	B1950	1987	Dec 4, 1899	Jan 2, 2050	Voyager Neptune
DE-200	J2000	1982	Dec 9, 1599	Mar 31, 2169	from DE118
DE-202	J2000	1987	Dec 4, 1899	Jan 2, 2050	from DE130
DE-234	J2000	1991	May 27, 1969	Apr 6, 1997	Mars Observer
DE-245	J2000	1993	Jun 16, 1750	Aug 1, 2051	Most accurate
DE-403	IERS	1995	Apr 29, 1599	Feb 1, 2200	Includes nutations & librations
DE-404	IERS	1995	Dec 2, -3002	Nov 14, 3000	from DE-403

Problems

1. I've taken the convention of Kaula and used the positive gradient for this book, whereas other authors adopt the acceleration as a *negative* of the potential. Am I justified when examining the choice of c in the development of ξ and the discussion of gradients in Sec. 7.2? How are potential energy, potential functions, and accelerations related?
2. Program the Jacchia-Roberts model discussed in Appendix B, and compare the results to the exponential model for daily values in the months of January, 1986, and May, 1996. Explain differences and similarities. Hint: you'll need to obtain data using information in Sec. D.5.
3. Develop coefficients for an eighth-order Runge-Kutta, Adams-Bashforth, and Gauss-Jackson numerical integrator. Test the resulting propagators on a Molniya orbit.
4. Derive the form of third-body acceleration in Eq. (7-33). Also show that Eq. (7-33) is a simplified form of Eq. (7-34).
5. Compare computing times for Encke's method and the simplified numerical technique discussed previously using a GPS satellite and a one-week simulation in May, 1996. Use the same force models with Encke's method as in the simplified approach. Do your answers differ for TOPEX? (Hint, it's in a roughly circular, 1300 km altitude, 66.0° inclined orbit) Additional credit: obtain the precision orbit ephemerides for TOPEX and analyze which method is more accurate. With Encke's method, was the rule of thumb for rectification points accurate?

6. Calculate entry and exit times for a satellite for one week (January 17 to January 25, 1997) given the following information.

$$\overset{\rightharpoonup}{r} = -26,175.1034 \hat{i} + 12,757.0706 \hat{j} + 14,626.6556 \hat{k} \text{ km}$$

$$\overset{\rightharpoonup}{v} = 2.376\,641 \hat{i} + 0.139\,677 \hat{j} + 2.078\,097 \hat{k} \text{ km/s}$$

What's the percentage increase for the solar-radiation pressure if we ignore the periods when the satellite is in the Earth's shadow?

7. A satellite is about to launch for Earth Resources Technology. It has a radar that can map out crop resources, accurate to about 15 m. Someone forgot to examine the astrodynamical needs for orbit determination, so you've been asked to help. If the planned orbital elements are $a = 6800$ km, $e = 0.02$, $i = 28.5^\circ$, what forces should you include in your analysis?
8. Calculate a one week ephemeris for the TOPEX satellite using the simplified and complex numerical propagation techniques discussed earlier. The orbital data is as follows:

$$\overset{\rightharpoonup}{r} = -4499.361\,363 \hat{i} + 3484.518\,039 \hat{j} + 5210.807\,611 \hat{k} \text{ km}$$

$$\overset{\rightharpoonup}{v} = 0.781\,833\,189 \hat{i} - 5.609\,928\,185 \hat{j} + 4.422\,626\,058 \hat{k} \text{ km/s}$$

Epoch = August 21, 1995, 00:20:00.00 UTC

Research and locate the POE information for TOPEX after the week of interest and compare your answers.

CHAPTER 8 GENERAL PERTURBATION TECHNIQUES

- 8.1 Historical Background
- 8.2 Introduction
- 8.3 Variation of Parameters
- 8.4 Hamilton's Formulation
- 8.5 Disturbing-Potential Formulations
- 8.6 Linearized Perturbations and Effects
- 8.7 Forming Analytical Solutions
- 8.8 Sernianalytical Solutions
- 8.9 Practical Considerations
- 8.10 Summary of Perturbation Effects

8.1 Historical Background

Although I'm discussing perturbations relatively late in this book, their effects were under study *before* the two-body equation was fully developed (see Pannekoek, 1989). Indeed, Kepler wrestled with variations in the Martian orbit caused by perturbations.

The Moon has also been prominent in the study of perturbations. It was a natural choice, but its difficult orbit challenged scientists seeking solutions for perturbed motion. Isaac Newton (c. 1690) laid the basis for determining the Moon's orbit with his law of gravitation. He derived much of his *Principia* from studying the Moon's motion. Alexis Claude Clairaut (1713–1765) continued the study, along with Jean le Rond d'Alembert (1717–1783) and Leonhard Euler (1707–1783). Newton explained most of the variations in the Moon's orbit, except the motion of the perigee. In 1749 Clairaut found that the second-order perturbation terms removed discrepancies between the observed and theoretical values which Newton hadn't treated. Then, about a century later, the full explanation was found in one of Newton's unpublished manuscripts.

Pierre Simon de Laplace (1749–1827) brought his special form of mathematical elegance to the solution of the Moon's motion. First, he used the true longitude (λ) as the independent variable in the equations of motion. Then, he explained the secular acceleration of λ —it depends on the eccentricity of the Earth's orbit, which we know changes over time. John Adams showed this approximation was within about $6''$ of observed values. Roy (1988, 297) suggests this discrepancy is related to tidal friction. As is often the case, the mathematical masters provided the answer hundreds of years ago.

Peter A. Hansen (1795–1874) restructured much of Laplace's work and developed tables that were used from 1862 to about 1922. He was originally a watchmaker, but his real interest was in astronomy, and he was involved in constructing several observatories. His two main works describing the Moon's motion were *Fundamenta* (1838) and *Darlegung* (1862–1864).

Simon Newcomb (1835–1909) was keenly interested in mathematics. He completed an arithmetic book at age seven, but he had little formal education. He was mostly self-taught and was always interested in the function of the American Ephemeris and Nautical Almanac Office. In 1861, he entered the Navy as a mathematician and was assigned to the Naval Observatory in Washington, D.C., where he made observations with the new 26" telescope (remember, this was 1861!). He eventually became the director in 1877 and managed the American Nautical Almanac Office. After that, he worked hard to provide precise ephemerides for navigation. He made some empirical corrections to Hansen's tables, and his modified results were used after 1883.

Charles E. Delaunay (1816–1872) is credited with developing the most complete algebraic solution for the Moon's motion (until the age of modern computers). He published his results in *La théorie du mouvement de la lune* (*The Theory of Lunar Motion*) during 1860–1867. It was accurate enough to predict the Moon's motion to one radius over a period of 20 years. Despite all the intense mathematics, however, the method isn't as accurate as other approximate relations.

The ability to model the Moon's motion remained somewhat elusive as measuring devices became more accurate. George W. Hill (1838–1914), who worked at the U.S. Naval Observatory while Newcomb was its director, significantly improved lunar theory by introducing the theory of *infinite matrices*. He also refined theories for the motions of Jupiter and Saturn. Ernest William Brown (1866–1938) followed the improvements of Hill's theory. A British mathematician, Brown spent most of his life studying the motion of the Moon. He published very accurate tables in 1908. They replaced Hansen's tables in about 1923 and were used for decades afterward. He spent summer vacations at Cambridge University and, much as Newton produced his tremendous work while on vacation, Brown produced a lunar theory consisting of more than 1500 terms during these rests. The theory remained the same until Wallace Eckert (1902–1971) and H. F. Smith Jr. used early computers to expand it by 6000 terms. Gutzwiller and Schmidt (1986) present a summary of the work of Hill, Brown, and Eckert. Andre Deprit (1926–) reproduced Delaunay's lunar theory with computers and symbolic manipulation. He found several errors in Delaunay's theory and therefore improved its accuracy. Deprit was awarded a medal by the National Academy of Science for this work. He, along with A. Rom and J. M. A. Danby, developed the computer program, *Mechanized Algebraic Operations* (MAO), to perform this analysis, as well as others.

Battin (1987, 471) suggests that Leonhard Euler (1707–1783) was the first to present the mathematical method usually known as the *variation of parameters* (VOP). The first practical orbit-determination technique was an application of VOP proposed by Joseph Louis Lagrange (1736–1813) in 1782. He formed what we know today as the Lagrange planetary equations (Bell, 1965, 174), also called the variation of parameters method.

Battin (1987, 471–473) relates how John Couch Adams (1819–1892), an Englishman, and Urbain-Jean-Joseph Le Verrier (1811–1877), a Frenchman, independently used the variation of parameters method to predict Neptune's existence. That VOP was used in the solution is important, but the manner in which the results were received warrants additional explanation. In 1845, Adams had proposed the existence and location of Neptune, which at the time was undiscovered. Unfortunately, his superiors weren't interested. Le Verrier arrived at similar numerical results, and he published them on June 1, 1846. The director of the Greenwich observatory, Sir George Biddell Airy (1801–1892), whom Adams had contacted with his results, noticed the similarity to Le Verrier's article. He suggested they try to locate the planet. Unfortunately, the suggestion went to James Challis, the Director at Cambridge Observatory, who had observed Neptune on July 9, 1846, but misinterpreted the results. Le Verrier had enlisted the support of the German astronomer Johann Gottfried Galle (1812–1910) to observe Neptune. He observed the new planet on September 23, 1846. Battin (1987, 473) adds that

The reader can imagine the controversy between the English and the French over who deserved the credit. But justice prevailed and when the battle subsided, it was universally agreed that both Adams and LeVerrier would share equally in the glory.

Modern analyses of perturbations center on events of the late 20th century. In particular, the launch of the Sputnik I satellite by the Soviet Union on October 4, 1957 sparked tremendous interest in space. Although the pioneers had laid the foundation for many of the required analyses, tremendous technical gaps existed when applying these results to the modern small satellite. Indeed, much of the analysis had been formulated for celestial objects and distant planets. Satellites orbiting near Earth presented some new challenges.

The launching of Sputnik concerned the Western nations because they were ill-prepared to meet the perceived threat of an Eastern nation with space capability. Indeed, 1959 brought the first major efforts to develop adequate theories for controlling and operating satellite systems. Two works stand even today as remarkable achievements with a long track record of operational use, so they deserve special mention here.

On October 30, 1959, Yoshihide Kozai (1928–) published his work at the Smithsonian Astrophysical Observatory and the Harvard College Observatory in Cambridge, Massachusetts. His preface states his assumptions:

In the present paper perturbations of six orbital elements of a close Earth orbiting satellite moving in the gravitational field of the Earth without air-resistance are derived as functions of mean elements and time. No assumptions are made about the order of magnitude of eccentricity and inclination. However, it is assumed that the density distribution of the Earth is symmetrical with respect to the axis of rotation, that the coefficient of the second harmonic of the potential is a small quantity of the first order and that those of the third and fourth harmonics are of second order. The results include periodic perturbations of the first order and secular perturbations of up to second order. (Kozai, 1959, 367)

That singularities exist in this theory (and in many others) is unfortunate. That they remain prevalent today simply results from choices made several decades ago. Kozai's approach had remarkable insight and provided the basis for the first operational, analytical

approaches to determining satellite orbits. But because it didn't include drag, the results were very limited, especially because most of the early satellites operated almost entirely within the drag envelope (the atmospheric region that strongly affects a satellite's orbit).

That same year, Dirk Brouwer (1902–1966) was working at the Yale University Observatory and was under contract with the U.S. Air Force to develop a satellite theory for military planners and operators. He published the results in the very same issue of the *Astronomical Journal* in which Kozai's results appeared, as well as in a technical paper for the U.S. Air Force's Research and Development Command. Brouwer (1959, 378) notes in the preface that his approach

give[s] the solution of the main problem for a spheroidal Earth with potential limited to the principal term and the second harmonic which contains the small factor [J_2]. The solution is developed in powers of [J_2] in canonical variables by a method which is basically the same as that used in treating a different problem by von Zeipel (1916). The periodic terms are divided in two classes: the short-periodic terms contain the mean anomaly in their arguments; the arguments of the long-periodic terms are multiples of the mean argument of perigee....

The results are obtained in closed form...The solution does not apply to orbits near the critical inclination.

Brouwer also gives contributions due to J_3 , J_4 , and J_5 in the same paper. Although Brouwer's ideas were very similar to those of Kozai, he used different methods. This particular theory took hold very quickly and was modified many times. In 1961, Brouwer and Hori extended the original work to include the effects of drag. The Air Force changed and improved the basic theory over the next two decades. Unfortunately, detailed documentation was often missing, and sensor sites acquired many variations and releases of the basic theories. In the late 1970s, the Air Force conducted a survey to determine what theories existed within the government. They published the results in *Spacetrack Report #3* (Hoots and Roehrich, 1980) to solve the lack of configuration control.

But new theories and applications continued to develop. As the computational throughput of machines grew, the complexity of the theories increased. During the mid 1960s and 1970s, several different contributors developed satellite theories based on the VOP formulation—loosely based on the perturbation technique known as the *method of averaging*. One development took place at Goddard Spaceflight Center. This theory was an averaged orbit generator used to propagate a satellite's long period and secular motion for specific mission applications and analysis. The research and development version of the *Goddard Trajectory Determination System*, GTDS, used this approach. It laid the foundation for important technical advances.

The next decade saw a unique semianalytical theory from a team of scientists led by Paul Cefola (1942–). Remarkably, one of the important technical inspirations of their *Draper Semianalytical Satellite Theory* (DSST) came from the work of Hansen in 1855. (Cefola found Hansen's epic work on expansions for modeling elliptical motion in the Museum of Comparative Zoology because technical journals weren't varied enough to include them in the mid-1800s.) Over the years, continual improvements have made DSST well suited for long-term, highly accurate analyses of satellite motion. Recent stud-

ies (Barker et al., 1995) have shown this particular approach to be *considerably* more accurate than existing analytical theories with comparable (or better) computing speed. Although I won't completely describe and derive the theory here, I'll give you a foundation for consulting mathematical documents.

This chapter offers some of the more useful and well-known analytical solutions for the effects of perturbations on orbits, so you can determine when and how to apply them to a specific problem. The semianalytical theory is a modern tool that allows us to predict well into the future with exceptional accuracy.

8.2 Introduction

This chapter develops techniques that allow us to examine the effect of perturbations on the *orbital elements*. The numerical techniques of Chap. 7 yield very high accuracy, but they don't tell us much about the qualitative behavior of the satellite's orbit because the output is usually position and velocity *vectors*. Expressing the orbit's behavior in terms of orbital elements has an advantage. Although position and velocity vectors undergo larger changes with time, most of the Keplerian elements undergo small variations in perturbed motion. It's much easier to see the effects of perturbations in slowly changing variables which reflect the size, shape, and orientation of the orbit over time. As usual, accuracy is especially important for mission analysis and operations. I'll discuss general perturbation and semianalytical techniques in this chapter.

General perturbation techniques replace the original equations of motion with an analytical approximation that captures the essential character of the motion over some limited time interval and which also permits analytical integration. Such methods rely on series expansions of the perturbing accelerations. In practice, we truncate the resulting expressions to allow simpler expressions in the theory. This trade-off speeds up computation but decreases accuracy. Unlike numerical techniques, analytical methods produce approximate, or "general" results that hold for some limited time interval and accept any initial input conditions. The quality of the solution degrades over time, but remember that the numerical solution also degrades—at different rates and for different reasons. Analytical techniques are generally more difficult to develop than numerical techniques, but they often lead to a better understanding of the perturbation source.

Semianalytical techniques combine the best features of numerical and analytical methods to get the best mix of accuracy and efficiency. The result can be a very accurate, relatively fast algorithm which applies to most situations. But semianalytical techniques vary widely. We choose a semianalytical technique mainly for its ability to handle varying applications, its documentation, and the fidelity and number of force models it includes. Most semianalytical techniques have improved accuracy and computational efficiency, but the availability of documentation (including very structured computer code) and flexibility are often important discriminators. In this book, I'll consider a technique semianalytical if it's not *entirely* analytical or numerical.

Let's continue Example 7-1 (page 468) to show how we form an analytical solution. An analytical technique assumes that the integral of our equation of motion is unknown but

that the series representing the equations *is* integrable. In this case, we know we can represent ($\cos(t)$) as an infinite series. We also know the integral of each term. Using only the first three terms (although the exact solution contains an infinite number of terms), the following example illustrates the analytical technique.

▼ **Example 8-1: Using an Analytical Solution.**

GIVEN: $\ddot{y}(t) = \cos(\omega t)$ and $\omega = 1$ rad/s

$$y(t=0) = y_o = 0$$

FIND: $y(t = 0.5 \text{ s})$

Break the “unknown” term into a series that *is* integrable.

$$\cos(t) = 1 - \frac{t^2}{2!} + \frac{t^4}{4!} + \dots$$

$$y(t) = \int_0^{0.5} \left(1 - \frac{t^2}{2!} + \frac{t^4}{4!} + \dots \right) dt$$

Let's just use the first three terms:

$$y(t=0.5) = t - \frac{t^3}{3!} + \frac{t^5}{5!} = 0.5 - 0.020\,833\,33 + 0.000\,260\,42$$

▲ $y(0.5) = 0.479\,427\,\underline{086}$

The underlined digits represent those which are in error from the true (direct) solution. Although the answer is very close to the direct solution (with three terms), consider if we had used only two terms; our solution would be $0.479\,\underline{166}\,67$. Also, some functions (most of which seem to appear in orbital mechanics!) require many more terms for solution. The accuracy depends directly on where we truncate the series and the length of the time interval we wish to apply the solution. Determining the point of truncation is difficult and may not be the same for different aspects of a larger overall problem, or for different applications. There is a subtle point in the example. The expansion parameter, t , is relatively large compared to the usual expansion parameter in orbital theories (eccentricity, J_2 , or $1/r$). It's difficult to show how truncating a series with a small parameter affects solution accuracy. This general area of study involves *asymptotic expansions*, or the behavior of divergent series, which we'll discuss in Sec. 8.2.1.

The results are clear. We prefer direct integration where possible. The analytical method is accurate and yields a quick solution; however, the series truncations may be difficult depending on the equations of motion. The numerical method is very accurate with the correct step size, but this determination may be tricky. The semianalytical technique would combine the analytical and numerical approaches. In general, the orbit-determination problem has few if any equations which may be integrated analytically. We can treat the rest by *either* method, depending on the particular use and the desired results.

Because we're examining the effect of perturbations on the orbital elements, we must characterize how they vary over time. Perturbations on orbital motion result in *secular* and *periodic* changes. Figure 8-1 shows an example of each of these effects.

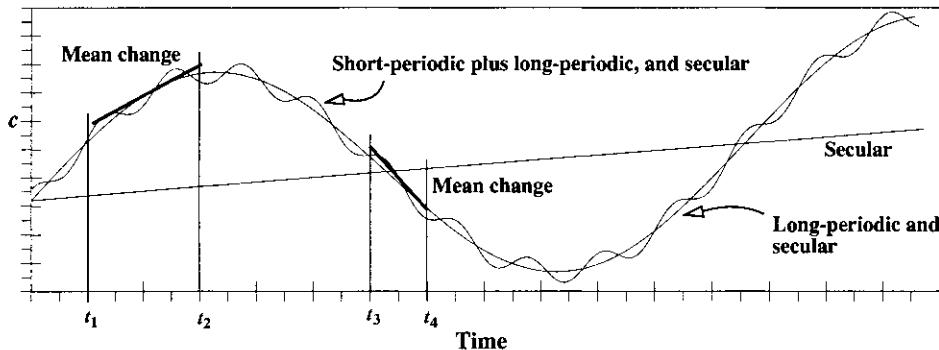


Figure 8-1. Effect of Perturbation Forces on Orbital Elements. The change in any orbital element, c , is illustrated. The straight line shows secular effects. The large oscillating line shows the secular plus long-periodic effects, and the small oscillatory line, which combines all three, shows the short-periodic effects. It's important to understand which elements are averaged when using mean elements. The two examples I've shown are greatly exaggerated but representative of possible errors.

Secular changes in a particular element vary linearly over time, or in some cases, proportionally to some power of time, such as a quadratic. The important point is that secular terms grow with time, and errors in secular terms produce unbounded error growth. Secular terms are the primary contributor to the degradation of analytical theories over long time intervals. Although the dominant perturbing force for the Earth, J_2 , results in all three types of effects, we can do a first-order approximation and approximate the main variations. We can also develop some higher-order solutions. **Periodic** changes are either *short*- or *long*-periodic, depending on the length of time required for an effect to repeat. Because so many definitions exist in the literature and in practice, I'll define each type.

Short-periodic effects typically repeat on the order of the satellite's period or less. **Long-periodic** effects have cycles considerably longer than one orbital period—typically one or two orders of magnitude longer. These long-periodic effects are often seen in the motion of the node and perigee and can last from a few weeks to a month or more. This means a short-periodic effect for a satellite at an altitude of 400 km could vary with periods up to about 100 minutes, whereas a short-periodic effect for a geosynchronous satellite would be up to about 24 hours. Also, short-periodic variations occur when a fast variable (true anomaly, for instance) is present in the contributing perturbational effect. These definitions become especially important when forming analytical and semianalytical theories. Because these short-periodic terms dictate the step-size of any numerical sim-

ulation, we sometimes use averaging techniques to strip off the short-periodic contributions so the remaining secular and long-periodic variations can be numerically integrated with larger time steps—the essence of semianalytical theories.

We must also distinguish certain orbital elements as either *fast* or *slow* variables, depending on their relative rate of change. ***Fast variables*** change a lot during one orbital revolution, even in the absence of perturbations. Examples are the mean, true, and eccentric anomalies, which all change 360° , or the cartesian coordinates, which also change dramatically. ***Slow variables*** (semimajor axis, eccentricity, inclination, node, argument of perigee) change very little during one orbital revolution. Perturbations cause these changes. Without perturbations, all the slow elements would remain constant. Fast variables would continue to change.

We can describe the perturbed motion of a satellite by an ordered series of position and velocity vectors. Consequently, at each point in time, we can use these vectors to find the orbital elements using two-body techniques in Chap. 2. These ***osculating elements*** are defined at any instant in time by the corresponding position and velocity vectors. “Osculate” comes from a Latin word meaning “to kiss.” Thus, the osculating orbit kisses the trajectory at the prescribed instant. We define an ***osculating ellipse*** as the two-body orbit the satellite would follow if the perturbing forces were suddenly removed at that instant. Therefore, *each* point on the trajectory has a corresponding set of osculating elements. Osculating elements are the true time-varying orbital elements, and they include all periodic (long- and short-periodic) and secular effects. They represent the high-precision trajectory and are useful for highly accurate simulations, including real-time pointing and tracking operations.

In contrast, ***mean elements*** are “averaged” over some selected time (or an appropriate angle such as true anomaly), so they’re relatively smoothly varying and don’t chase the short-periodic variations. Notice that mean elements depend on some unspecified averaging interval of the time; the true, eccentric, or mean anomaly; or the longitude. Because there are many kinds of mean elements, it’s important to understand how they’re defined and used. Mean elements are most useful for long-range mission planning because they approximate the satellite’s long-term behavior.

The secular and periodic phenomena would suggest a trigonometric series, more specifically a Fourier sum, as a natural representation for the osculating orbital elements. Escobal ([1965] 1985, 362) shows an example series of the form

$$c = c_o + \dot{c}_1(t - t_o) + K_1 \cos(2\omega) + K_2 \sin(2\nu + \omega) + K_3 \cos(2\nu) \quad (8-1)$$

where K_1 , K_2 , and K_3 are constants. Typically the coefficients, K_i , can be expressed as sums and products of polynomials in a , e , and i . In these cases, Eq. (8-1) is called a ***Poisson Series***. I’ve expanded the equation to show each type of influence from the effect of perturbations: initial value, c_o ; secular change, \dot{c}_1 ; long-periodic, 2ω ; short-periodic, 2ν ; and mixed-periodic, $2\nu + \omega$. Realize that the short-periodic term depends mainly on the fast variable, ν .

Figure 8-1 depicts the decomposition of arbitrary time-varying motion. The behavior of the orbital element depends on the specific time interval. Consider the periods from t_1 to t_2 and t_3 to t_4 . The secular rates will probably be about the same, but the long-periodic rate will vary significantly and the slope sign will even change! We remove particular frequencies from the motion to approximate an average behavior over time. We use the terms single- and double-averaged elements to characterize the resulting motion over the averaging interval. **Single-averaged** elements result from removing the high-frequency short-periodic motions. Figure 8-1 shows an ideal mean rate of change of the osculating elements over two intervals. Another common technique removes the long-periodic and short-periodic variations to find a mean motion. It's called **double-averaging**. There are numerous ways to select the interval over which we determine an average rate—be sure to understand which method you're using.

8.2.1 The Method of Perturbations

Don't confuse the Method of Perturbations with the concept of perturbing forces. The latter describes small forces, relative to the attraction of the central body's point mass, which cause small deviations from the unperturbed motion caused by this central force. Nor should we confuse the method with the VOP equations of motion, which are differential equations describing the effect (i.e., rates) due to the small perturbing forces on the "constants" of the unperturbed solution. These differential equations must still be solved. The **Method of Perturbations** describes a class of mathematical techniques for generating analytical solutions which describe the motion of a satellite subject to disturbing forces. We use the term "analytical solution" somewhat loosely. The solution isn't a rigorous solution of the equations of motion. Rather, it's an analytical approximation to the solution which describes the motion over some definite time interval. Also, the method applies not only to the VOP equations but also to Hamilton's equations and other forms of the equations of motion.

The underlying concept of the Method of Perturbations is that one or more small disturbing forces cause small deviations from the known solution to the unperturbed problem. These small perturbing forces can be associated with small parameters which characterize the magnitude of the disturbing forces, such as

$$\frac{dc_i}{dt} = \epsilon_1 f_i + \epsilon_2 g_i$$

$$\epsilon_1 \ll 1$$

$$\epsilon_2 \ll 1$$

and an approximation to the true solution can be developed as a power series in these small parameters:

$$\begin{aligned}
 c = c_o + \epsilon_1 \alpha_1(t) + \frac{\epsilon_1^2}{2!} \alpha_2(t) + \dots \\
 + \epsilon_2 \beta_1(t) + \frac{\epsilon_2^2}{2!} \beta_2(t) + \dots \\
 + \frac{\epsilon_1 \epsilon_2}{2!} \gamma_2(t) + \dots
 \end{aligned} \tag{8-2}$$

Note that the *order* of the theory is defined as the highest power of the expansion parameter retained in the series expansion. It's common to identify an analytical perturbation theory as "first order," "second order," and so on. The main root of these expressions is expansion using the Taylor series, which is the core of *all* perturbation solutions. Geyling and Westerman (1971, 125–129) concisely discuss the factors that distinguish the order of the theory. The results are similar to selecting terms from the series approach of the f and g functions [Eq. (4-68)].

Examples of small parameters are J_2, J_3, C_{lm}, S_{lm} . Other examples include the ratio of mean semimajor axes of the satellite and third-body orbits (a/a_3) for third-body perturbations, scaled density for atmospheric drag, scaled thrust acceleration for maneuvers, and scaled solar flux for solar-radiation pressure. We don't have to scale these quantities for practical applications, but from a theoretical viewpoint, we want the small parameters to be dimensionless as well.

The small parameters present may not all be of the same order of magnitude—for example, $J_2 \approx 10^{-3}$, $J_4 \approx 10^{-6}$, etc. Consequently, we should be careful to select the appropriate degree of each small parameter so all numerical contributions of a given order of magnitude are included. Hence, J_2^2 terms are often included in solutions when other zonal harmonic terms are included because $J_2 \approx O(J_2^2)$; that is, the other zonal harmonic coefficients, J_3, J_4, J_5 are roughly of order J_2^2 .

Note from Eq. (8-2) that the perturbations at first order simply add, whereas at second and higher orders, mixed or coupled terms [e.g., $(\epsilon_1 \epsilon_2 / 2!) \gamma_2(t)$] also appear. The pure terms ($\epsilon_1, \epsilon_1^2, \epsilon_2, \epsilon_2^2, \dots$) describe the direct effect of the perturbing forces on the motion, whereas the coupled terms ($\epsilon_1 \epsilon_2, \epsilon_1^2 \epsilon_2, \epsilon_1 \epsilon_2^2, \dots$) describe the effect of one perturbation on the component of satellite motion caused by the other perturbations, and vice versa.

Several techniques exist for developing these power-series representations, but they're difficult. An example of an ad hoc technique is the method of successive substitutions. In this approach, we solve the VOP equations representing the disturbed motion by assuming all elements appearing in the right-hand sides are constant except for the fast variable, which varies according to pure two-body motion. The initial conditions are used as the constant values. We then integrate the equations analytically, providing corrections, Δa , to the initial values. These corrections are functions of the initial conditions and the time. We add them to the initial values, $c_1 = c_o + \Delta c$, and substitute the result back into the right-hand side of the VOP equations, which are integrated once again. In principle, we may

repeat the process again and again; however, the complexity grows rapidly, which is readily apparent from the second substitution.

Other techniques use Taylor series expansions of the right-hand side of the VOP equations, and some canonical methods have used Lie-series expansions of the Hamiltonian. The complexity generally grows geometrically with the order; thus, higher-order theories are quite complicated, so only a limited number of theories exist beyond second order. Deprit (1981) provided a notable exception by using computerized symbolic manipulations to develop a solution to $O(J_2^4)$ for the **Main Problem** of artificial-satellite theory, which is the common designation of the two-body + J_2 problem.

These series representations are formal because, generally, they're not uniformly convergent. But they're useful for developing approximate solutions over some fixed time interval, provided ϵ is small enough. The foundation for this approach is the *Theory of Asymptotic Expansions (Les méthodes nouvelles de la méchanique céleste)* originated by Poincaré (1892) to justify the techniques of Euler, Lagrange, and other practitioners of celestial mechanics.

With this general information in hand, we can begin exploring the various forms of the perturbation equations for each type of disturbing effect. Each section derives the equations of motion and generally describes the effects on the orbital elements.

8.3 Variation of Parameters

I'll present several analytical solutions to introduce their common notations and processes. Most of these solutions rely on the *variation of parameters* (VOP) form of the equations of motion originally developed by Euler and improved by Lagrange (1873). We call the overall process *variation of parameters* (VOP) because the orbital elements (the constant parameters in the two-body equations) are changing. Lagrange and Gauss both developed VOP methods to analyze perturbations—Lagrange's technique applies to conservative accelerations, whereas Gauss's technique also works for nonconservative accelerations.

The VOP method is a formulation of the equations of motion that are well suited to weakly perturbed, dynamical systems. The concept is based on the premise that we can use the solution for the unperturbed system to represent the solution of the perturbed system, provided that we can generalize the constants in the solution to be time-varying parameters. The unperturbed system is the two-body system discussed in Chap. 2, and it represents a collection of formulas that provide the position and velocity vectors at a desired time. Remember, these formulas depend only on the six orbital elements and time. In principle, however, we could use any set of constants of the unperturbed motion, including the initial position and velocity vectors. Time is related to the equations of motion through the conversions of mean, eccentric, and true anomaly, as discussed in Chap. 4.

Now consider adding a large perturbing force. In the case of the three-body problem, this changes the overall nature of the problem, and we can no longer determine exact solutions—we must resort to numerical solutions or possibly analytical solutions in special cases. If the perturbations are small, like those in Encke's formulation, we can correct the

unperturbed solution to capture the behavior of the perturbed problem. The situation becomes one in which we know the unperturbed solution, and we only need to form equations of motion that let us represent the time-varying change of the orbital elements—the essence of VOP.

The VOP equations of motion are a system of first-order differential equations that describe the rates of change for the time-varying elements,

$$\frac{d\vec{c}}{dt} = f(\vec{c}, t) \quad (8-3)$$

Strictly speaking, these are no longer the “two-body” elements because they aren’t constant. They’re the osculating elements, which are time-varying and are associated with an osculating orbit at a particular instant of time. Before developing the specific forms of the VOP equations from the original equations of motion for the perturbed system, let’s obtain the relationship between the osculating-element rates and the disturbing acceleration. We need to review certain results for the unperturbed and perturbed systems. In general, we’ll use the following notation for both systems:

$$\vec{r} = \vec{x}(a, e, i, \Omega, \omega, M, t) = \vec{x}(c_1, c_2, c_3, c_4, c_5, c_6, t) = \vec{x}(\vec{c}, t)$$

$$\vec{v} = \dot{\vec{x}}(a, e, i, \Omega, \omega, M, t) = \dot{\vec{x}}(c_1, c_2, c_3, c_4, c_5, c_6, t) = \dot{\vec{x}}(\vec{c}, t)$$

The six constants, \vec{c} , indicate we can use *any* set of orbital elements (or more generally, any six independent constants of the motion) in the formulation. In this instance, I’m using a “vector” that’s larger than three components.

We can also express the equations of motion for both systems (respectively) as follows.

$$\ddot{\vec{x}}(\vec{c}, t) + \frac{\mu \vec{x}(\vec{c}, t)}{|\vec{x}(\vec{c}, t)|^3} = 0 \quad \ddot{\vec{x}}(\vec{c}, t) + \frac{\mu \vec{x}(\vec{c}, t)}{|\vec{x}(\vec{c}, t)|^3} = \vec{a}_{pert} \quad (8-4)$$

where \vec{a}_{pert} is the disturbing acceleration. Finally, we know that the velocity is the derivative of the position vector, so we can again write unperturbed and perturbed equations, respectively. Note that the unperturbed case is a simplified version of the perturbed case in which $dc_i/dt = 0$.

$$\begin{aligned} \dot{\vec{x}}(\vec{c}, t) &= \frac{d\vec{x}(\vec{c}, t)}{dt} = \frac{\partial \vec{x}(\vec{c}, t)}{\partial t} \\ \dot{\vec{x}}(\vec{c}, t) &= \frac{d\vec{x}(\vec{c}, t)}{dt} = \frac{\partial \vec{x}(\vec{c}, t)}{\partial t} + \sum_{i=1}^6 \frac{\partial \vec{x}(\vec{c}, t)}{\partial c_i} \frac{dc_i}{dt} \end{aligned} \quad (8-5)$$

We now differentiate the position vector in the perturbed case [Eq. (8-4)] twice to obtain an expression for the acceleration. But we impose a constraint to guarantee that our

parameters exhibit the osculating characteristic—each position and velocity vector defines an osculating ellipse. The first derivative is given in Eq. (8-5)b, and we impose the following constraint:

$$\sum_{i=1}^6 \frac{\partial \dot{\vec{x}}(\vec{c}, t)}{\partial c_i} \frac{dc_i}{dt} \equiv 0 \quad (8-6)$$

Geyling and Westerman (1971, 158) refer to Eq. (8-6) as the *condition of osculation* because it defines the condition for the osculating orbit. Remember that the osculating orbit corresponds to a two-body orbit at an instant of time. This result doesn't mean the total derivative of each element is zero. It implies that the *sum* of expressions is zero. Substituting this value into the expression for perturbed velocity, Eq. (8-5)b, gives the same result as the unperturbed solution. A second derivative of the position vector, still using the condition of osculation, gives us

$$\frac{d^2 \vec{x}(\vec{c}, t)}{dt^2} = \frac{\partial^2 \vec{x}(\vec{c}, t)}{\partial t^2} + \sum_{i=1}^6 \frac{\partial \dot{\vec{x}}(\vec{c}, t)}{\partial c_i} \frac{dc_i}{dt}$$

We can now find the osculating element rates in terms of the original equations of motion by substituting into Eq. (8-4)b:

$$\frac{\partial^2 \vec{x}(\vec{c}, t)}{\partial t^2} + \sum_{i=1}^6 \frac{\partial \dot{\vec{x}}(\vec{c}, t)}{\partial c_i} \frac{dc_i}{dt} + \frac{\mu \vec{x}(\vec{c}, t)}{|\vec{x}(\vec{c}, t)|^3} = \vec{a}_{pert} \quad (8-7)$$

But we know that the unperturbed equations of motion [Eq. (8-4)] hold for any time and can be written as

$$\frac{\partial^2 \vec{x}(\vec{c}, t)}{\partial t^2} + \frac{\mu \vec{x}(\vec{c}, t)}{|\vec{x}(\vec{c}, t)|^3} = 0$$

so Eq. (8-7) simplifies to

$$\sum_{i=1}^6 \frac{\partial \dot{\vec{x}}(\vec{c}, t)}{\partial c_i} \frac{dc_i}{dt} = \vec{a}_{pert} \quad (8-8)$$

This expression provides an explicit relation between the osculating-element rates and the disturbing acceleration. However, it's incomplete because it represents only three equations (\vec{a}_{pert}) in the six element rates (dc_i/dt), i.e., three equations in six unknowns. The remaining three equations come from the condition of osculation in Eq. (8-6). Using matrix notation, the complete system of equations gives us

$$\begin{bmatrix} \sum_{i=1}^6 \frac{\partial \dot{\vec{x}}(\vec{c}, t)}{\partial c_i} \frac{dc_i}{dt} \\ \sum_{i=1}^6 \frac{\partial \dot{\vec{x}}(\vec{c}, t)}{\partial \dot{c}_i} \frac{dc_i}{dt} \end{bmatrix} = \begin{bmatrix} \dot{\vec{0}} \\ \dot{\vec{a}}_{pert} \end{bmatrix} \quad (8-9)$$

This expression is not well suited for computations because it isn't of the form of Eq. (8-3). We can now examine two approaches for transforming Eq. (8-9) into a more desirable form—Lagrangian and Gaussian VOP.

8.3.1 Lagrangian VOP (Conservative Effects)

The general theory for finding the rates of change of the osculating elements is known as the *Lagrange planetary equations of motion*, or simply the *Lagrangian VOP*, and is attributed to Lagrange because he was the first person to obtain equations in the form of Eq. (8-3) for all six orbital elements. He was concerned with the small disturbances on planetary motion about the Sun due to the gravitational attraction of the planets. He chose to model the disturbing acceleration due to this conservative perturbation as the gradient of a potential function. Consequently, we'll modify Eq. (8-9) by replacing the disturbing acceleration with the gradient of the disturbing potential function, which is the starting point for the discussion. Many references present the formulation: Kaula (1966, 25–30), McClain (1978, 11–23), Roy (1988, 180–181), Brouwer and Clemence (1961, 273–284), and Taff (1985, 302–307). Here, I'll combine and unify these approaches.

The solution begins by taking the dot product of the bottom vector in Eq. (8-9) with $\left(\frac{\partial \dot{\vec{x}}(\vec{c}, t)}{\partial c_k} \right)$ and the top vector in Eq. (8-9) with $\left(-\frac{\partial \dot{\vec{x}}(\vec{c}, t)}{\partial c_k} \right)$, where $k = 1, 2, \dots, 6$ to account for each element. The result is

$$\sum_{i=1}^6 \frac{dc_i}{dt} \left(\frac{\partial \dot{\vec{x}}(\vec{c}, t)}{\partial c_k} \cdot \frac{\partial \dot{\vec{x}}(\vec{c}, t)}{\partial c_i} - \frac{\partial \dot{\vec{x}}(\vec{c}, t)}{\partial c_i} \cdot \frac{\partial \dot{\vec{x}}(\vec{c}, t)}{\partial c_k} \right) = \frac{\partial R}{\partial c_k}$$

Now, let's introduce Lagrange brackets as a definition and for simplification. Kaula (1966, 26–29), as well as Taff and Brouwer and Clemence, describe these brackets fully. Note that the Lagrange bracket is the difference of two dot products and is therefore a scalar quantity:

$$[c_i, c_k] = \frac{\partial \dot{\vec{x}}(\vec{c}, t)}{\partial c_i} \cdot \frac{\partial \dot{\vec{x}}(\vec{c}, t)}{\partial c_k} - \frac{\partial \dot{\vec{x}}(\vec{c}, t)}{\partial c_k} \cdot \frac{\partial \dot{\vec{x}}(\vec{c}, t)}{\partial c_i} \quad (8-10)$$

After performing the matrix operations in Eq. (8-9) and invoking the Lagrange bracket notation, we can rewrite the matrix equation for the element rates.

$$\begin{bmatrix} [c_1, c_1] & \dots & [c_1, c_6] \\ \vdots & & \\ [c_6, c_1] & & [c_6, c_6] \end{bmatrix} \begin{bmatrix} \dot{c}_i \\ \frac{dc_i}{dt} \end{bmatrix} = \begin{bmatrix} \frac{\partial R}{\partial \dot{x}} \end{bmatrix}$$

We can express the time rate of change of each orbital element in matrix notation (let the matrix of Lagrange brackets be L):

$$\dot{\dot{c}}_i = L^{-1} \frac{\partial R}{\partial \dot{c}_i} \quad (8-11)$$

This is a general form of the Lagrange planetary equations.

An important characteristic of Lagrange brackets is that they're time-independent, or constant. From Taff (1985, 306), or Kaula (1966, 27),

$$\frac{d [c_p, c_k]}{dt} \equiv 0$$

Some identities are obvious from the definition [Eq. (8-10)] and they reduce the number of Lagrange brackets we must evaluate to 15:

$$\begin{aligned} [c_p, c_i] &= 0 \\ [c_p, c_k] &= -[c_k, c_i] \end{aligned}$$

Evaluating the remaining Lagrange brackets isn't easy. It requires the partial derivatives

$$\frac{\partial \dot{x}}{\partial c_i} \quad \text{and} \quad \frac{\partial \dot{\dot{x}}}{\partial c_i} \quad \text{where } i = 1 \dots 6$$

Consequently, we need the position and velocity expressed as functions of the elements. Also, because the brackets are independent of time, we are free to evaluate them (and thus the partial derivatives) anywhere in the orbit. Evaluation at perigee is convenient because it provides simplified expressions.

Following Kaula (1966, 27,28) we obtain the explicit dependance of \dot{x} , $\dot{\dot{x}}$ on the orbital elements using the transformation from the geocentric (IJK) to the apsidal (PQW) frame. Let \dot{x} , $\dot{\dot{x}}$ and \dot{r}_{PQW} , \dot{v}_{PQW} represent the position and velocity vectors in the geocentric and apsidal frames, respectively. From Eq. (2-65) and Eq. (2-67), the position and velocity vectors in the PQW frame are

$$\dot{\vec{r}}_{PQW} = \begin{bmatrix} r \cos(\nu) \\ r \sin(\nu) \\ 0 \end{bmatrix} \quad \dot{\vec{v}}_{PQW} = \begin{bmatrix} -na \sin(\nu) \\ \sqrt{1-e^2} \\ \frac{na(e + \cos(\nu))}{\sqrt{1-e^2}} \\ 0 \end{bmatrix}$$

Remember from Eq. (2-14) that

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

Eq. (1-35) allows us to transform between the *IJK* and *PQW* frames. If we let \mathbf{R} represent the transformation matrix from *PQW* to *IJK*,

$$\dot{\vec{x}} = \mathbf{R}_{[IJK] \over [PQW]} \dot{\vec{r}}_{PQW} \quad \dot{\vec{x}} = \mathbf{R}_{[IJK] \over [PQW]} \dot{\vec{v}}_{PQW}$$

The previous relations show an important result— \mathbf{R} is a function of only Ω , ω , and i , whereas $\dot{\vec{r}}_{PQW}$ and $\dot{\vec{v}}_{PQW}$ are functions of a , e , and ν . Consequently,

$$\frac{\partial(\dot{\vec{x}}, \dot{\vec{x}})}{\partial(\Omega, \omega, i)} = \frac{\partial \mathbf{R}_{[IJK] \over [PQW]}}{\partial(\Omega, \omega, i)} (\dot{\vec{r}}_{PQW}, \dot{\vec{v}}_{PQW}) \quad \frac{\partial(\dot{\vec{x}}, \dot{\vec{x}})}{\partial(a, e, M)} = \mathbf{R}_{[IJK] \over [PQW]} \frac{\partial(\dot{\vec{r}}_{PQW}, \dot{\vec{v}}_{PQW})}{\partial(a, e, M)}$$

We can also determine the *PQW* position and velocity vectors at perigee.

$$\dot{\vec{r}}_{PQW} = \begin{bmatrix} a(1-e) \\ 0 \\ 0 \end{bmatrix} \quad \dot{\vec{v}}_{PQW} = \begin{bmatrix} 0 \\ na\sqrt{1+e} \\ 0 \end{bmatrix}$$

Three cases arise in the solution:

1. If both c_i and $c_k \in \{\Omega, i, \text{ or } \omega\}$, then

$$\frac{\partial \dot{\vec{x}}}{\partial c_i} = \begin{bmatrix} \frac{\partial \dot{x}_1}{\partial c_i} \\ \frac{\partial \dot{x}_2}{\partial c_i} \\ \frac{\partial \dot{x}_3}{\partial c_i} \end{bmatrix} \quad \frac{\partial \dot{\vec{x}}}{\partial c_k} = \begin{bmatrix} \frac{\partial \dot{x}_1}{\partial c_k} \\ \frac{\partial \dot{x}_2}{\partial c_k} \\ \frac{\partial \dot{x}_3}{\partial c_k} \end{bmatrix}$$

where at perigee,

$$\frac{\partial \dot{x}_j}{\partial c_i} = \sum_{l=1}^3 \frac{\partial R_{jl}}{\partial c_i} r_{PQW} = \frac{\partial R_{j1}}{\partial c_i} a (1 - e)$$

$$\frac{\partial \dot{x}_j}{\partial c_k} = \sum_{i=1}^3 R_{ji} v_{PQW} = \frac{\partial R_{j2}}{\partial c_k} n a \sqrt{\frac{1+e}{1-e}}$$

The indices j and l designate the matrix elements' row and column of the R matrix. The dot product is then

$$\frac{\partial \dot{x}}{\partial c_i} \cdot \frac{\partial \dot{x}}{\partial c_k} = \sum_{j=1}^3 \frac{\partial \dot{x}_j}{\partial c_i} \frac{\partial \dot{x}_j}{\partial c_k} = \sum_{j=1}^3 n a^2 \sqrt{1-e^2} \frac{\partial R_{j1}}{\partial c_i} \frac{\partial R_{j2}}{\partial c_k}$$

Similarly,

$$\frac{\partial \dot{x}}{\partial c_i} \cdot \frac{\partial \dot{x}}{\partial c_k} = \sum_{j=1}^3 \frac{\partial \dot{x}_j}{\partial c_i} \frac{\partial \dot{x}_j}{\partial c_k} = \sum_{l=1}^3 n a^2 \sqrt{1-e^2} \frac{\partial R_{j2}}{\partial c_i} \frac{\partial R_{j1}}{\partial c_k}$$

and we have

$$[c_i, c_k] = n a \sqrt{1-e^2} \sum_{j=1}^3 \left[\left\{ \frac{\partial R_{j1}}{\partial c_i} \frac{\partial R_{j2}}{\partial c_k} - \frac{\partial R_{j2}}{\partial c_k} \frac{\partial R_{j1}}{\partial c_i} \right\} \right]$$

2. If $c_i \in \{\Omega, i, \text{ or } \omega\}$ and $c_k \in \{a, e, M\}$, similar operations yield

$$[c_i, c_k] = \sum_{j=1}^3 \left[a (1-e) \frac{\partial R_{j1}}{\partial c_i} \left\{ R_{j1} \frac{\partial \dot{v}_P}{\partial c_k} + R_{j2} \frac{\partial \dot{v}_Q}{\partial c_k} \right\} - \frac{\sqrt{1-e^2} n a}{1-e} \frac{\partial R_{j2}}{\partial c_i} \left\{ R_{j1} \frac{\partial r_P}{\partial c_k} + R_{j2} \frac{\partial r_Q}{\partial c_k} \right\} \right]$$

3. If both c_i and $c_k \in \{a, e, M\}$, similar operations yield

$$\begin{aligned} [c_i, c_k] = & \sum_{j=1}^3 \left[R_{j1}^2 \left\{ \frac{\partial r_P}{\partial c_i} \frac{\partial \dot{v}_P}{\partial c_k} - \frac{\partial r_P}{\partial c_k} \frac{\partial \dot{v}_P}{\partial c_i} \right\} + R_{j1} R_{j2} \left\{ \frac{\partial r_P}{\partial c_i} \frac{\partial \dot{v}_Q}{\partial c_k} - \frac{\partial r_P}{\partial c_k} \frac{\partial \dot{v}_Q}{\partial c_i} \right\} \right. \\ & \left. + R_{j1} R_{j2} \left\{ \frac{\partial r_Q}{\partial c_i} \frac{\partial \dot{v}_P}{\partial c_k} - \frac{\partial r_Q}{\partial c_k} \frac{\partial \dot{v}_P}{\partial c_i} \right\} + R_{j2}^2 \left\{ \frac{\partial r_Q}{\partial c_i} \frac{\partial \dot{v}_Q}{\partial c_k} - \frac{\partial r_Q}{\partial c_k} \frac{\partial \dot{v}_Q}{\partial c_i} \right\} \right] \end{aligned}$$

Let's examine an example from Kaula (1966, 28). He gives the form of the Lagrange bracket of

$$[\Omega, i] = \sum_{j=1}^3 \left(\frac{\partial R_{j1}}{\partial \Omega} \cdot \frac{\partial R_{j2}}{\partial i} - \frac{\partial R_{j2}}{\partial \Omega} \cdot \frac{\partial R_{j1}}{\partial i} \right) n a^2 \sqrt{1-e^2}$$

We can get the partial derivatives by differentiating each component of the rotation matrix from $PQW \Rightarrow IJK$, Eq. (1-35):

$$\begin{aligned} [\Omega, i] = & \left[\left\{ -\sin(\Omega) \cos(\omega) - \cos(\Omega) \cos(i) \sin(\omega) \right\} \sin(\Omega) \sin(i) \cos(\omega) \right. \\ & - \left\{ \cos(\Omega) \cos(\omega) - \sin(\Omega) \cos(i) \sin(\omega) \right\} \cos(\Omega) \sin(i) \cos(\omega) \\ & - \left. \left\{ \sin(\Omega) \sin(\omega) - \cos(\Omega) \cos(i) \cos(\omega) \right\} \sin(\Omega) \sin(i) \sin(\omega) \right. \\ & \left. - \left\{ \cos(\Omega) \sin(\omega) + \sin(\Omega) \cos(i) \cos(\omega) \right\} \cos(\Omega) \sin(i) \sin(\omega) \right] n a^2 \sqrt{1-e^2} \end{aligned}$$

After “simplifying” (the second terms in each brace cancel), we have

$$[\Omega, i] = -n a^2 \sqrt{1-e^2} \sin(i)$$

There are only twelve nonzero results (Kaula, 1966, 29), although six values differ only by a sign because $[\Omega, i] = -[\Omega, i]$, and so on. Refer to Fitzpatrick (1970, 179–189) for a complete discussion, including Lagrange brackets using alternate element sets. Also note that Herrick (1972, 125–152) introduces a grave notation and an associated perturbative differential process that avoids computing and inverting the Lagrange brackets. The remaining unique Lagrange brackets expressed in classical elements are

$$[\Omega, a] = \frac{n a \sqrt{1-e^2}}{2} \cos(i) = [\omega, a] \cos(i)$$

$$[\Omega, e] = \frac{-n a^2 e}{\sqrt{1-e^2}} \cos(i) = [\omega, e] \cos(i)$$

$$[\omega, a] = \frac{n a \sqrt{1-e^2}}{2}$$

$$[\omega, e] = \frac{-n a^2 e}{\sqrt{1-e^2}}$$

$$[a, M_o] = \frac{-n a}{2}$$

We can now solve Eq. (8-11) to find the equations for the variation of parameters. The result is called the *Lagrange planetary equations*.

$$\begin{aligned}\frac{da}{dt} &= \frac{2}{na} \frac{\partial R}{\partial M_o} \\ \frac{de}{dt} &= \frac{1-e^2}{na^2 e} \frac{\partial R}{\partial M_o} - \frac{\sqrt{1-e^2}}{na^2 e} \frac{\partial R}{\partial \omega} \\ \frac{di}{dt} &= \frac{1}{na^2 \sqrt{1-e^2} \sin(i)} \left\{ \cos(i) \frac{\partial R}{\partial \omega} - \frac{\partial R}{\partial \Omega} \right\} \\ \frac{d\omega}{dt} &= \frac{\sqrt{1-e^2}}{na^2 e} \frac{\partial R}{\partial e} - \frac{\cot(i)}{na^2 \sqrt{1-e^2}} \frac{\partial R}{\partial i} \\ \frac{d\Omega}{dt} &= \frac{1}{na^2 \sqrt{1-e^2} \sin(i)} \frac{\partial R}{\partial i} \\ \frac{dM_o}{dt} &= -\frac{1-e^2}{na^2 e} \frac{\partial R}{\partial e} - \frac{2}{na} \frac{\partial R}{\partial a}\end{aligned}\tag{8-12}$$

An alternate form of the Lagrange planetary equations is

$$\frac{dc_i}{dt} = -\sum_{k=1}^6 (c_i, c_k) \frac{\partial R}{\partial c_k}, \text{ where } i = 1..6$$

Here, the (c_i, c_k) are the Poisson brackets, which are intimately tied to the Lagrange brackets. Note that L^{-1} is related to the negative matrix of *Poisson brackets*, $-P$, because $LP^T = I$. They are also time independent. Poisson brackets have slowly replaced the Lagrange brackets in many modern formulations. For instance, methods based on Lie series and Lie-Deprit transformations use Poisson brackets because they are better suited for computer implementations.

The practical application of the VOP equations of motion requires that we select an element set. The standard, classical orbital elements have singularity problems for vanishing inclination and eccentricity values. So we need to carefully select elements and the range of application of the resulting theory before developing the analytical model. Developing these models takes a lot of work.

Notice the Lagrange planetary equations provide the time rate of change of the orbital elements; thus, only the disturbing function, R , is needed because this represents the *perturbing* part of the potential (central body, third body, ...). Using the complete aspherical potential, U , would give erroneous results because it includes the spherical potential, which is responsible for two-body motion. If we express the disturbing function as a function of the classical orbital elements, R (a, e, i, Ω, ω , and the initial value of mean anomaly, M_o), we have the expressions in Eq. (8-12). Kaula's form of the potential [Eq. (8-28)] is well suited for use in Eq. (8-12).

Eq. (8-12) uses M_o , but remember that the final mean anomaly is related to the initial value by $M = M_o + n\Delta t$. It's often convenient to include the two-body component of the motion in the VOP equations so we don't have to compute it separately. We sometimes desire a direct expression for \dot{M} . To arrive at this, let $\sigma = -nT$, where T is the time of perigee passage. The mean anomaly is now

$$M = n(t - T) = nt + \sigma \quad (8-13)$$

Differentiating, we have

$$\dot{M} = n + nt + \dot{\sigma} = n + \frac{dn}{da}\dot{a} + \dot{\sigma}$$

Eq. (8-12) gives us the time rate of change for a , and $dn/da = -3n/2a$ from Eq. (2-18). However, from Eq. (8-12), notice that $\partial R/\partial a$ has a contribution from a and an indirect contribution for Eq. (8-13). Thus

$$\frac{\partial R}{\partial a} = \left(\frac{\partial R}{\partial a} \right)_M + \left(-\frac{3n}{2a}t \right) \frac{\partial R}{\partial M}$$

where the M subscript indicates the derivative only with respect to a . Because the partial with respect to e doesn't suffer from this distinction, we can write

$$\dot{M} = n - \frac{3nat}{2a} - \frac{1-e^2}{na^2 e} \frac{\partial R}{\partial e} - \frac{2}{na} \left(\frac{\partial R}{\partial a} \right)_M + \frac{3t}{a^2} \frac{\partial R}{\partial M}$$

If we substitute

$$\dot{a} = \frac{2}{na} \frac{\partial R}{\partial M}$$

the terms with time appearing explicitly cancel, and we have

$$\dot{M} = n - \frac{1-e^2}{na^2 e} \frac{\partial R}{\partial e} - \frac{2}{na} \left(\frac{\partial R}{\partial a} \right)_M$$

The equation dM/dt now represents the complete mean anomaly rate. Remember, two-body motion doesn't affect the other elements. The mean anomaly grows linearly with time much faster than the other elements and is therefore characterized as a fast variable. We can choose either a "fast" element such as M , which changes significantly in two-body motion, or a "slow" variable, such as the time of periapsis passage, which changes only slowly.

Also notice the presence of e and $\sin(i)$ in the denominators of Eq. (8-12). These parameters can become small, or they can vanish. Thus, this form of the VOP is called *singular*, and it isn't well suited for small e or i . Several other element sets exist that avoid the problems of small divisors in the VOP equations. Taff (1985, 308–311) lists several other forms for Eq. (8-12) depending on the final element, and McClain (1978, 21–23) lists forms with the equinoctial orbital elements that are an example of a nonsingular set:

$$B = \sqrt{1 - h_e^2 - k_e^2} \quad C = 1 + p_e^2 + q_e^2$$

$$\begin{aligned}\frac{da}{dt} &= \frac{2}{na} \frac{\partial R}{\partial \lambda_M} \\ \frac{dh_e}{dt} &= \frac{B}{na^2} \frac{\partial R}{\partial k_e} - \frac{h_e B}{na^2(1+B)} \frac{\partial R}{\partial \lambda_M} + \frac{kC}{2na^2B} \left\{ p_e \frac{\partial R}{\partial p_e} + q_e \frac{\partial R}{\partial q_e} \right\} \\ \frac{dk_e}{dt} &= -\frac{B}{na^2} \frac{\partial R}{\partial h_e} - \frac{k_e B}{na^2(1+B)} \frac{\partial R}{\partial \lambda_M} - \frac{h_e C}{2na^2B} \left\{ p_e \frac{\partial R}{\partial p_e} + q_e \frac{\partial R}{\partial q_e} \right\} \\ \frac{dp_e}{dt} &= -\frac{p_e C}{2na^2B} \left\{ k_e \frac{\partial R}{\partial h_e} - h_e \frac{\partial R}{\partial k_e} + \frac{\partial R}{\partial \lambda_M} \right\} + \frac{f_r C^2}{4na^2B} \frac{\partial R}{\partial q_e} \\ \frac{dq_e}{dt} &= -\frac{q_e C}{2na^2B} \left\{ k_e \frac{\partial R}{\partial h_e} - h_e \frac{\partial R}{\partial k_e} + \frac{\partial R}{\partial \lambda_M} \right\} - \frac{f_r C^2}{4na^2B} \frac{\partial R}{\partial p_e} \\ \frac{d\lambda_M}{dt} &= n - \frac{2}{na} \frac{\partial R}{\partial a} + \frac{B}{na^2(1+B)} \left\{ h_e \frac{\partial R}{\partial h_e} + k_e \frac{\partial R}{\partial k_e} \right\} + \frac{C}{2na^2B} \left\{ p_e \frac{\partial R}{\partial p_e} + q_e \frac{\partial R}{\partial q_e} \right\}\end{aligned}\tag{8-14}$$

Another form uses direction cosines of the \hat{K} axis, or the third-body position vector, with respect to the equinoctial frame (Cefola, 1974, McClain, 1978, 23).

8.3.2 Gaussian VOP (Nonconservative and Conservative Effects)

For many applications, it's convenient to express the rates of change of the elements explicitly in terms of the disturbing forces. Again, be aware that although the discussion mentions forces, they are actually acceleration (*specific forces*) to match the units in the equations. Gauss's form is advantageous for nonconservative forces because it's expressed directly from the disturbing acceleration. But it works equally well for conservative forces because the forces are simply gradients of the potential functions. It's also easy to visualize this representation because we're familiar with the concept of a force.

Many sources present Gauss's VOP; some, such as Brouwer and Clemence (1961, 302–306), are much shorter. We'll follow Bate, Mueller, and White (1971, 396–407) for our derivation. This process obtains the VOP equations by direct methods and is an alternate approach to the development of the Lagrangian equations.

We can obtain a general form of the Gaussian VOP equations from Eq. (8-9) expressing it as two separate equations and not using the matrix notation:

$$\sum_{i=1}^6 \frac{\partial \dot{\vec{x}}(\vec{c}, t)}{\partial c_i} \frac{dc_i}{dt} = \vec{0} \quad \sum_{i=1}^6 \frac{\partial \dot{\vec{x}}(\vec{c}, t)}{\partial c_i} \frac{dc_i}{dt} = \vec{a}_{pert}$$

Taking the dot product of the first equation with $\frac{\partial c_j}{\partial \dot{x}}$ and the second by $\frac{\partial c_j}{\partial \ddot{x}}$ and adding the two relations result in

$$\sum_{i=1}^6 \left(\frac{\partial c_j}{\partial \dot{x}} \cdot \frac{\partial \dot{x}}{\partial c_i} + \frac{\partial c_j}{\partial \ddot{x}} \cdot \frac{\partial \ddot{x}}{\partial c_i} \right) \frac{dc_i}{dt} = \frac{\partial c_j}{\partial \dot{x}} \cdot \dot{\vec{a}}_{pert} \quad (8-15)$$

Because the elements depend only on position and velocity and are mutually independent, the quantity inside the parenthesis reduces to the Kronecker δ function, $\delta_{j,i} = 1$ for $j = i$ and $\delta_{j,i} = 0$ for $j \neq i$. Therefore, Eq. (8-15) becomes

$$\sum_{i=1}^6 \delta_{j,i} \frac{dc_i}{dt} = \frac{\partial c_j}{\partial \dot{x}} \cdot \dot{\vec{a}}_{pert}$$

or more simply,

$$\frac{dc_j}{dt} = \frac{\partial c_j}{\partial \dot{x}} \cdot \dot{\vec{a}}_{pert}$$

Thus Gauss's form of the VOP requires the partial derivatives of the elements with respect to the velocity. We must determine these for particular element sets. Be careful when evaluating the dot product because the velocity and the partial derivative must be in the same frame as the acceleration. See Chap. 1 for additional information on coordinate systems (Sec. 1.4).

Gauss chose to develop the equations in the *RSW* system. I'll next present one geometrical derivation of the VOP equations. Summarizing from Sec. 1.5, let the components of the disturbing force (per unit mass) be \hat{R} along the radius vector, \hat{S} perpendicular to the \hat{R} axis in the orbit plane in the *direction* of satellite motion, and \hat{W} normal to the orbit plane, such that the positive directions of \hat{R} , \hat{S} , and \hat{W} form a right-hand set of axes. The \hat{W} axis points in the instantaneous direction of the angular-momentum vector. The disturbing (specific) force becomes $\vec{F} = F_R \hat{R} + F_S \hat{S} + F_W \hat{W}$.

The position vector is simply $\vec{r}_{RSW} = r \hat{R}$. Adopting a polar representation for the radial distance [$r = r(\nu)$] and differentiating (chain rule) result in a very useful expression:

$$\dot{\vec{v}}_{RSW} = \dot{r} \hat{R} + r \dot{\nu} \hat{S} = \dot{\nu} \left(\frac{dr}{d\nu} \hat{R} + r \hat{S} \right)$$

First, we'll derive da/dt . Let's begin with the specific mechanical energy [Eq. (2-11)] and solve for the semimajor axis:

$$\xi = -\frac{\mu}{2a} \quad \text{or} \quad a = -\frac{\mu}{2\xi}$$

In two-body motion, ξ is a constant. But now, we'll allow it to change based on the specific perturbing force. The time rate-of-change of the energy results from the work done by the perturbing force and the distance traveled. Thus,

$$\frac{d\xi}{dt} = \frac{\vec{F} \cdot \vec{v}}{m} = \dot{\nu} \left(\frac{dr}{d\nu} F_R + r F_S \right)$$

Notice, if the disturbing force alters the semimajor axis, the orbit's energy must change. If the energy changes, the force is considered nonconservative. Solve for the semimajor axis and take the derivative of the energy equation:

$$\frac{da}{dt} = \frac{da}{d\xi} \frac{d\xi}{dt} = \frac{\mu}{2\xi^2} \frac{d\xi}{dt}$$

To find $(d\xi/dt)$, you need intermediate expressions for $\dot{\nu}$ and $dr/d\nu$. Differentiate the trajectory equation [Eq. (2-14)] to get

$$\frac{dr}{d\nu} = \frac{r e \sin(\nu)}{1 + e \cos(\nu)}$$

From the angular momentum relations in Eq. (2-8) and Eq. (2-9), and the relation of mean motion [Eq. (2-41)],

$$h = r^2 \dot{\nu} = \sqrt{\mu p} = \sqrt{\mu a (1 - e^2)} = n a^2 \sqrt{1 - e^2} \quad (8-16)$$

Equate the final two terms to get a relation that's useful throughout the development:

$$\sqrt{\mu a} = n a^2$$

It's apparent from Eq. (8-16) that

$$\dot{\nu} = \frac{n a^2}{r^2} \sqrt{1 - e^2} \quad (8-17)$$

Substitute the expressions as

$$\frac{da}{dt} = \frac{\mu}{2\xi^2} \left(\frac{n a^2}{r^2} \sqrt{1 - e^2} \right) \left(\frac{r e \sin(\nu)}{1 + e \cos(\nu)} F_R + r F_S \right)$$

or

$$\frac{da}{dt} = \frac{2e \sin(\nu)}{n \sqrt{1 - e^2}} F_R + \frac{2a \sqrt{1 - e^2}}{nr} F_S \quad (8-18)$$

To derive the remaining element variations, you'll need the time rate-of-change of angular momentum. Don't confuse this with two-body motion, for which angular momen-

tum is constant; in perturbed motion, $dh/dt \neq 0$. This rate is a result of the torque of the perturbing (specific) forces acting on the system. So,

$$\frac{d\hat{h}}{dt} = \dot{\hat{r}} \times \hat{F} = \dot{\hat{r}} \times (F_R \hat{R} + F_s \hat{S} + F_w \hat{W}) = r F_s \hat{W} - r F_w \hat{S}$$

Another way to represent the change in angular momentum is to differentiate the vector itself ($\dot{\hat{h}} = h \dot{\hat{W}}$) in the RSW system. If α represents the angular change in the angular momentum, the rate of change of the angular-momentum vector will have a component in the W direction and a tangential component in the S direction. Thus,

$$\frac{d\hat{h}}{dt} = \dot{h} \hat{W} + h \frac{d\alpha}{dt} \hat{S}$$

The change in angular momentum must be in the $S-W$ plane because the perturbing force applies to r and the momentum change results from $\dot{\hat{r}} \times \hat{F}/m$. By comparing the W component from both expressions for dh/dt , you'll find that

$$\frac{dh}{dt} = \dot{h} = r F_s \quad (8-19)$$

As with the radial rate (\dot{r}), remember that $\dot{h} \neq dh/dt$. To find \dot{e} , recall the expression for the eccentricity [Eq. (2-43)],

$$e = \sqrt{1 - \frac{p}{a}} = \sqrt{1 - \frac{h^2}{\mu a}}$$

Differentiate to get

$$\begin{aligned} \frac{de}{dt} &= \frac{1}{2} \left(1 - \frac{h^2}{\mu a} \right)^{-1/2} \left[\frac{-2h}{\mu a} \frac{dh}{dt} + (-h^2) \mu (-1) (\mu a)^{-2} \frac{da}{dt} \right] \\ &= \frac{1}{2e} \left[\frac{-2h}{\mu a} \frac{dh}{dt} + \frac{h^2}{\mu a^2} \frac{da}{dt} \right] \\ &= \left(-\frac{h}{\mu ae} \frac{dh}{dt} + \frac{h^2}{2\mu a^2 e} \frac{da}{dt} \right) \end{aligned}$$

Substitute (dh/dt) and (da/dt) into the equation above:

$$\frac{de}{dt} = -\frac{h}{\mu ae} r F_s + \frac{h^2}{2\mu a^2 e} \left(\frac{2e \sin(\nu)}{n\sqrt{1-e^2}} F_R + \frac{2a\sqrt{1-e^2}}{nr} F_S \right)$$

Now, use $(h = \sqrt{\mu a (1-e^2)})$ to obtain

$$\frac{de}{dt} = -\frac{\sqrt{1-e^2}}{\sqrt{\mu ae}} (r F_s) + \frac{(1-e^2) \sin(\nu)}{an\sqrt{1-e^2}} F_R + \frac{(1-e^2) \sqrt{1-e^2}}{enr} F_S$$

but $\sqrt{\mu a} = na^2$, so

$$\frac{de}{dt} = -\frac{\sqrt{1-e^2}}{na^2e} rF_S + \frac{\sqrt{1-e^2}\sin(\nu)}{na} F_R + \frac{\sqrt{1-e^2}}{na^2e} \left(\frac{(1-e^2)a^2}{r} F_S \right)$$

and the final result is

$$\frac{de}{dt} = \frac{\sqrt{1-e^2}\sin(\nu)}{na} F_R + \frac{\sqrt{1-e^2}}{na^2e} \left[\frac{a^2(1-e^2)}{r} - r \right] F_S \quad (8-20)$$

We often transform the equation by replacing r :

$$\frac{de}{dt} = \frac{\sqrt{1-e^2}}{na} \left[\sin(\nu) F_R + \left(\cos(\nu) + \frac{e + \cos(\nu)}{1 + e \cos(\nu)} \right) F_S \right]$$

The simplest approach to deriving di/dt is analytical. From Eq. (2-45), recall that

$$\cos(i) = \frac{\hat{h} \cdot \hat{K}}{\|\hat{h}\| \|\hat{K}\|} = \frac{\hat{h} \cdot \hat{K}}{\hat{h}}$$

Differentiating with respect to time to determine the rate of change gives us

$$-\sin(i) \frac{di}{dt} = \frac{\hat{h} \left(\frac{d\hat{h}}{dt} \cdot \hat{K} \right) - (\hat{h} \cdot \hat{K}) \frac{dh}{dt}}{h^2}$$

Using $(d\hat{h}/dt)$ and dh/dt from the development of Eq. (8-19) results in

$$-\sin(i) \frac{di}{dt} = \frac{h \left(rF_S \hat{W} - rF_W \hat{S} \right) \cdot \hat{K} - h \cos(i) rF_S}{h^2}$$

But $\hat{W} \cdot \hat{K} = \cos(i)$ and $\hat{S} \cdot \hat{K} = \sin(i) \cos(u)$. Remember, u is the argument of latitude. Thus,

$$-\sin(i) \frac{di}{dt} = -\frac{rF_W \sin(i) \cos(u)}{na^2 \sqrt{1-e^2}}$$

and

$$\frac{di}{dt} = \frac{r \cos(u) F_W}{na^2 \sqrt{1-e^2}} \quad (8-21)$$

Notice that changes in inclination result only from forces in the \hat{W} direction. In-plane disturbances don't affect inclination.

It's also easiest to use a direct formulation in deriving $d\Omega/dt$. If you use Eq. (2-46) and Eq. (2-47),

$$\cos(\Omega) = \frac{\hat{I} \cdot (\hat{K} \times \hat{h})}{|\hat{K} \times \hat{h}|}$$

Differentiating gives you

$$\begin{aligned} -\sin(\Omega) \frac{d\Omega}{dt} &= \frac{\hat{I} \cdot \left(\hat{K} \times \frac{dh}{dt} \right) |\hat{K} \times \hat{h}| - \hat{I} \cdot (\hat{K} \times \hat{h}) \frac{d}{dt} |\hat{K} \times \hat{h}|}{|\hat{K} \times \hat{h}|^2} \\ &= \frac{\hat{I} \cdot \left[\hat{K} \times \left(rF_s \hat{W} - rF_w \hat{S} \right) \right] h \sin(i) - h \cos(\Omega) \sin(i) \left(\frac{dh}{dt} \sin(i) + h \cos(i) \frac{di}{dt} \right)}{h^2 \sin^2(i)} \end{aligned}$$

Determine the vector cross products using the transformation matrix for PQW to IJK [Eq. (1-35)] and the alternate formulation, Eq. (1-25). First, do the cross product; then, select the dot product from the correct term. Consider $\hat{I} \cdot \hat{K} \times \hat{W}$ first. The cross product result is $-W_2$, W_1 , 0. Taking the dot product results in

$$\hat{I} \cdot \hat{K} \times \hat{W} = \cos(\Omega) \sin(i)$$

Similarly, $\hat{I} \cdot \hat{K} \times \hat{S} = \hat{I} \times \hat{K} \cdot \hat{S} = -\hat{J} \cdot \hat{S} = \sin(\Omega) \sin(u) - \cos(\Omega) \cos(u) \cos(i)$

Substituting for di/dt and dh/dt gives you

$$\frac{d\Omega}{dt} = \frac{rF_w \sin(u)}{h \sin(i)}$$

In terms of orbital elements, use Eq. (8-16) and Eq. (4-5). Just as with di/dt , notice that only the W component of the force perturbs Ω . Therefore,

$$\frac{d\Omega}{dt} = \frac{r \sin(u) F_w}{na^2 \sqrt{1 - e^2} \sin(i)} \quad (8-22)$$

Next, obtain the argument of perigee rate. At this point, it's useful to recall the argument of latitude, u [Eq. (2-52)] as

$$\frac{(\hat{K} \times \hat{h}) \cdot \hat{r}}{|\hat{K} \times \hat{h}|} = r \cos(u)$$

Differentiating gives you

$$\frac{|\hat{K} \times \hat{h}| \left(\hat{K} \times \frac{d\hat{h}}{dt} \cdot \hat{r} \right) - (\hat{K} \times \hat{h} \cdot \hat{r}) \frac{d}{dt} |\hat{K} \times \hat{h}|}{|\hat{K} \times \hat{h}|^2} = -r \sin(u) \left(\frac{d\omega}{dt} + \frac{d\nu}{dt} \right)$$

and solving for the rate of change of the argument of perigee results in

$$\frac{d\omega}{dt} = \frac{-|\hat{K} \times \hat{h}| \left(\hat{K} \times \frac{d\hat{h}}{dt} \cdot \hat{r} \right) + (\hat{K} \times \hat{h} \cdot \hat{r}) \frac{d}{dt} |\hat{K} \times \hat{h}| - \frac{d\nu}{dt} |\hat{K} \times \hat{h}|^2 r \sin(u)}{|\hat{K} \times \hat{h}|^2 r \sin(u)}$$

Remember that the $d\nu/dt$ term is *not* the same as in the two-body problem [Eq. (8-17)] because now perturbations are influencing the orbit. We know several relations,

$$\hat{K} \times \hat{W} \cdot \hat{r} = r \sin(i) \cos(u)$$

$$\hat{K} \times \hat{S} \cdot \hat{r} = r \hat{S} \times \hat{R} \cdot \hat{K} = -r \hat{W} \cdot \hat{K} = -r \cos(i)$$

$$\hat{K} \times \hat{h} \cdot \hat{r} = rh \sin(i) \cos(u)$$

and have already developed dh/dt [from Eq. (8-19)] and di/dt [Eq. (8-21)]. Therefore, you can find

$$\begin{aligned} \frac{d\omega}{dt} &= \left\{ \frac{1}{h^2 \sin^2(i) r \sin(u)} \right\} \left\{ -h \sin(i) \left[\hat{K} \times (r F_s \hat{W} - r F_w \hat{S}) \cdot \hat{r} \right] \right. \\ &\quad \left. + (\hat{K} \times \hat{h} \cdot \hat{r}) \left[\frac{dh}{dt} \sin(i) + h \cos(i) \frac{di}{dt} \right] - \frac{d\nu}{dt} h^2 r \sin^2(i) \sin(u) \right\} \end{aligned}$$

Now you must find an expression for $d\nu/dt$. Perturbations causing motion of the periapsis also affect the true anomaly. A form of the trajectory equation using the angular momentum yields

$$r(1 + e \cos(\nu)) = \frac{h^2}{\mu}$$

You must remember to differentiate the terms which perturbations affect (e , ν , and h):

$$r \left(\frac{de}{dt} \cos(\nu) - e \sin(\nu) \frac{d\nu}{dt} \right) = \frac{2h}{\mu} \frac{dh}{dt}$$

$$re \sin(\nu) \frac{d\nu}{dt} = r \cos(\nu) \frac{de}{dt} - \frac{2h}{\mu} \frac{dh}{dt}$$

At this point, you solve for $d\nu/dt$ and substitute for dh/dt [Eq. (8-19)] and de/dt [Eq. (8-20)], which gives you

$$\frac{d\nu}{dt} = \frac{\cot(\nu)}{e} \left[\frac{\sqrt{1-e^2}}{na} \sin(\nu) F_R + \frac{\sqrt{1-e^2}}{na^2 e} \left(\frac{a^2(1-e^2)}{r} - r \right) F_S \right] - \frac{2h}{\mu r e \sin(\nu)} r F_S$$

Collecting terms of F_S results in

$$\frac{d\nu}{dt} = \frac{\sqrt{1-e^2}}{nae} \cos(\nu) F_R - \frac{\sqrt{1-e^2}}{nae} \frac{2+e \cos(\nu)}{1+e \cos(\nu)} F_S$$

As an aside, there are some subtle issues here. Notice that if there are no perturbing forces ($F_R = F_S = 0$), the rate of change vanishes. The true anomaly does change, even for unperturbed motion. Thus, we write the total change in true anomaly as the summation of a two-body and perturbed contribution. Now, let's substitute the perturbed rate of change of the true anomaly plus Eq. (8-19) and Eq. (8-21), into $d\omega/dt$:

$$\begin{aligned} \frac{d\omega}{dt} = & \frac{1}{h^2 r \sin^2(i) \sin(u)} \left\{ -h \sin(i) [r^2 \sin(i) \cos(u) F_S + r^2 \cos(i) F_W] \right. \\ & + rh \sin(i) \cos(u) \left[r \sin(i) F_S + \frac{r \cos(i) \cos(u)}{na^2 \sqrt{1-e^2}} F_W \right] \\ & \left. - h^2 r \sin^2(i) \sin(u) \frac{\sqrt{1-e^2}}{nae} \left[\cos(\nu) F_R - \frac{2+e \cos(\nu)}{1+e \cos(\nu)} F_S \right] \right\} \end{aligned}$$

and simplify to get

$$\begin{aligned} \frac{d\omega}{dt} = & \frac{\sqrt{1-e^2}}{nae} \left[-\cos(\nu) F_R + \frac{2+e \cos(\nu)}{1+e \cos(\nu)} \sin(\nu) F_S \right] \\ & - \frac{r \cot(i) \sin(u)}{h} F_W \end{aligned} \tag{8-23}$$

You can now see that the change in ω due to the in-plane perturbations (F_R and F_S) is directly related to the change in the true anomaly due to these perturbations.

To develop the time rate-of-change of the mean anomaly, begin with the general definition of Kepler's equation [Eq. (4-7), $M_o = M - n(t - t_o)$]. Differentiating results in

$$\frac{dM_o}{dt} = \frac{dE}{dt} - \frac{de}{dt} \sin(E) - e \cos(E) \frac{dE}{dt} - \frac{dn}{dt} (t - t_o)$$

You can find the rate of change of the mean motion using Eq. (2-17) (See also page 604)

$$\frac{dn}{dt} = -\frac{3n}{2a} da$$

We know de/dt , Eq. (8-20) and $\frac{dE}{dt} = \frac{1}{r\sqrt{a}}\mu$ from Chap. 4, and you can get $\sin(E)$ and $\cos(E)$ from the expressions in Eq. (4-8) and Eq. (4-9). The result (for elliptical orbits only) is given by Bate, Mueller, and White (1971, 406).

$$\frac{dM_o}{dt} = -\frac{1}{na}\left(\frac{2r}{a} - \frac{1-e^2}{e}\cos(\nu)\right)F_R - \frac{(1-e^2)}{nae}\left(1 + \frac{r}{a(1-e^2)}\right)\sin(\nu)F_S - t\frac{dn}{dt}$$

You can simplify this equation as I've shown in Eq. (8-24).

Now, summarize the formulas for the *Gaussian form* of the VOP equations using the disturbing force with specific force components resolved in the RSW system:

$$\begin{aligned} \frac{da}{dt} &= \frac{2}{n\sqrt{1-e^2}} \left\{ e\sin(\nu)F_R + \frac{p}{r}F_S \right\} \\ \frac{de}{dt} &= \frac{\sqrt{1-e^2}}{na} \left\{ \sin(\nu)F_R + \left(\cos(\nu) + \frac{e+\cos(\nu)}{1+e\cos(\nu)} \right)F_S \right\} \\ \frac{di}{dt} &= \frac{r\cos(u)}{na^2\sqrt{1-e^2}}F_w \\ \frac{d\Omega}{dt} &= \frac{r\sin(u)}{na^2\sqrt{1-e^2}\sin(i)}F_w \\ \frac{d\omega}{dt} &= \frac{\sqrt{1-e^2}}{nae} \left\{ -\cos(\nu)F_R + \sin(\nu)\left(1 + \frac{r}{p}\right)F_S \right\} - \frac{r\cot(i)\sin(u)}{h}F_w \\ \frac{dM_o}{dt} &= \frac{1}{na^2e} \left\{ (p\cos(\nu) - 2er)F_R - (p+r)\sin(\nu)F_S \right\} \end{aligned} \quad (8-24)$$

Similar to Eq. (8-24), we can find alternate expressions for the mean anomaly. Taff (1985, 315) lists one expression.

$$\frac{d\sigma}{dt} = -\frac{1}{na} \left\{ \frac{2r}{a} - \frac{(1-e^2)\cos(\nu)}{e} \right\} F_R - \frac{(1-e^2)\sin(\nu)}{nae} \left(1 + \frac{r}{p}\right) F_S$$

These VOP equations in classical orbital elements have some limitations. First, they are limited to eccentricities less than 1.0 because of the presence of $\sqrt{1-e^2}$. Also note that they suffer from the same singularities as Lagrange's form of the VOP equations because the singularities are due to the particular element set, not how the disturbing forces are characterized. The rate of change of Ω has $\sin(i)$ in the denominator. This causes the equation to be indeterminate for vanishing inclinations. A similar problem exists for ω with small values of eccentricity. Thus, this particular set of equations isn't recommended for small values of eccentricity or inclination.

Placing a limit on "small" values of eccentricity or inclination depends on the integration method we're using. For many of the integration techniques discussed in Sec. 7.5, we

can easily process eccentricity and $\sin(i)$ values on the order of 0.001 over reasonable time spans. Of course, this assumes we're using the best precision available on a particular computing platform. On the other hand, general perturbation techniques depend on the existence of a small parameter, ϵ , in the equations of motion. This small parameter is critical in developing the solution because it takes the form of the first few terms in a series.

$$a = a_0 + \epsilon a_1 + \epsilon^2 a_2 + \dots$$

Because the small parameter is much less than 1.0, it assures us that the first few terms will be enough for a reasonably accurate solution, at least for a given time interval. These series usually don't converge, but they're good approximations over some time interval for a small enough ϵ . A small divisor in the equations of motion (e.g. e) causes the solution to behave as

$$a = a_0 + \frac{\epsilon}{e} b_1 + \frac{\epsilon^2}{e^2} b_2 + \dots$$

Consequently, for eccentricities on the order of about 0.001 ($e \approx J_2 \approx 0.001$), the quantity J_2/e is about 1.0, and the series no longer behaves as a series in a small parameter, and the assumption of general perturbations and resulting solutions are no longer valid. Special formulas using other element sets handle nearly circular and parabolic orbits as shown in Taff (1985, 314–316). Each set may apply to different scenarios.

As an aside, if the disturbing function, R , is known, we can use the Gaussian form of the equations by forming each force component. If we can express the disturbing function in spherical coordinates as $R = R(r, u, i)$, with the argument of latitude ($u = \omega + \nu$), the acceleration components of the disturbing force are

$$F_R = \frac{\partial R}{\partial r} \quad F_S = \frac{1}{r} \frac{\partial R}{\partial u} \quad F_W = \frac{1}{r \sin(u)} \frac{\partial R}{\partial i} \quad (8-25)$$

Using the aspherical potential means we can illustrate the disturbing force's components for the zonal harmonics in Eq. (7-18) or Eq. (8-45) with the aid of Eq. (8-25). We can write the primed Legendre polynomials as combinations of other polynomials. Noting from Eq. (5-12) that $\sin(\phi_{gc}) = \sin(i) \sin(u)$, we have (for $m = 0$)

$$\begin{aligned} F_R &= \frac{\mu}{r^2} \sum_{l=2}^{\infty} J_l (l+1) \left(\frac{R_{\oplus}}{r} \right)^l P'_l [\sin(i) \sin(u)] \\ F_S &= -\frac{\mu}{r^2} \sin(i) \cos(u) \sum_{l=2}^{\infty} J_l \left(\frac{R_{\oplus}}{r} \right)^l P'_l [\sin(i) \sin(u)] \\ F_W &= -\frac{\mu}{r^2} \cos(i) \sum_{l=2}^{\infty} J_l \left(\frac{R_{\oplus}}{r} \right)^l P'_l [\sin(i) \sin(u)] \end{aligned} \quad (8-26)$$

where the differential is (use $\alpha = \sin(i) \sin(u)$ for clarity)

$$(1 - \alpha^2) P'_l[\alpha] = (1 - \alpha^2) \frac{dP_l[\alpha]}{d\alpha} = -l\alpha P_l[\alpha] + lP_{l-1}[\alpha]$$

8.4 Hamilton's Formulation

William Rowen Hamilton (1805–1865) developed the *canonical form*, which has come to be known as the *Hamiltonian form*, of the equations of motion. Hamilton showed that if the parameters used to represent the motion were canonical, the equations of motion for a conservative force became particularly simple. The importance of canonical elements is that the Lagrange brackets and also the Poisson brackets become identically zero or one. Thus, if the Lagrange brackets are $[\alpha_i, \alpha_j]$ and we use canonical variables, $[\alpha_i, \alpha_j]$ equals one if $i = j$ and equals zero elsewhere. The same process applies to Poisson brackets.

The concept of canonical variables is embedded in the generalized approach to classical mechanics, which originates with Lagrange's work in analytical dynamics and his use of generalized coordinates and momenta. Two examples of canonical variables are the standard position and velocity vectors, and the Delaunay elements.

The Hamiltonian for the equations of motion differs from the Lagrange planetary equations because the element rates depend not only on the disturbing function but also on the total energy of the disturbed system. The total energy is the sum of the potential and kinetic energy. Letting \mathcal{H} be the total energy—which is also the Hamiltonian—we have

$$\mathcal{H} = KE - U = \frac{1}{2}mv^2 - R - U_{2-body}$$

where we've used the kinetic energy and the disturbing potential, R . Remember, $U = U_{2-body} + R$. Notice that Eq. (8-11) takes the form where the matrix of Lagrange brackets is diagonal and the partial derivatives of the disturbing function are replaced by partial derivatives of the Hamiltonian. Also remember that the c_k are a canonical set of variables.

Computational considerations also warrant a brief discussion of canonical variables. To understand why we use canonical transformations, imagine an observer watching another person on a moving train throwing a ball up in the air. As the observer stands beside the tracks, the motion of the ball is rather complex, including horizontal and vertical components. If the observer were able to move onto the train, the motion of the ball would appear simple because the nonlinearities of the train's motion are “removed” and the vertical motion is directly observable. In the same way, canonical transformations permit us to remove some complicated perturbation motions from the analysis and thereby isolate only the “simple” remainder.

Using a canonical set of elements, the Delaunay variables (L_d , h , H_d , M , ω , Ω), described in Eq. (2-63), we have a Hamiltonian,

$$\mathcal{H} = \frac{\mu^2}{2L_d^2} + R(L_d, h, H_d, M, \omega, \Omega)$$

The canonical equations corresponding to Eq. (8-12) take the particularly simple form

$$\begin{aligned}\dot{L}_d &= \frac{\partial \mathcal{H}}{\partial M} & \dot{M} &= -\frac{\partial \mathcal{H}}{\partial L_d} \\ \dot{h} &= \frac{\partial \mathcal{H}}{\partial \omega} & \dot{\omega} &= -\frac{\partial \mathcal{H}}{\partial h} \\ \dot{H}_d &= \frac{\partial \mathcal{H}}{\partial \Omega} & \dot{\Omega} &= -\frac{\partial \mathcal{H}}{\partial H_d}\end{aligned}\quad (8-27)$$

Note that (L_d, M) , (h, ω) and (H_d, Ω) are canonically conjugate variables (momenta and coordinates). Refer to Taff (1985, 308–311) for varieties of the function of the variational equations. It's possible to extend Hamilton's canonical form of the equations of motion for nonconservative perturbations.

The canonical form is extremely important because of the concept of canonical transformations. Any transformation that maps the canonical set (a_k, b_k) to the set (α_k, β_k) is called a canonical transformation, as long as the form of the equations of motion remains the same. Letting $T(a_k, b_k) = (\alpha_k, \beta_k)$ designate the transformation from one set of variables to another implies a transformation of the Hamiltonian [$\mathcal{H}(a_k, b_k) \Rightarrow \mathcal{H}^*(\alpha_k, \beta_k)$]. However, because the Hamiltonian represents the total energy of the system, for conservative systems, the two Hamiltonians must be equal even though their functional forms and arguments are different [$\mathcal{H}(a_k, b_k) = \mathcal{H}^*(\alpha_k, \beta_k)$] and

$$\dot{\alpha}_k = \frac{\partial}{\partial \beta_k} \mathcal{H}^* \quad \dot{\beta}_k = -\frac{\partial}{\partial \alpha_k} \mathcal{H}^*$$

The importance of canonical transformations is obvious if a transformation can be found which yields a new set of canonical variables or elements that are constants or simple functions of time, such as the classical orbital elements. Remember that the VOP equations for a two-body problem reduce simply to

$$\dot{a} = \dot{e} = \dot{i} = \dot{\Omega} = \dot{\omega} = 0 \quad \dot{M} = n$$

and the corresponding canonical Delaunay variables yield

$$\dot{L}_d = \dot{h} = \dot{H}_d = 0 \quad \dot{\Omega} = k_\Omega \quad \dot{\omega} = k_\omega \quad \dot{M} = n$$

where k_Ω and k_ω are constant rates. Despite the powerful advantage of the Hamiltonian approach, two or more successive canonical transformations are often required. For judiciously chosen transformations, the equations of the transformed system may be easily solved to yield expressions for the transformed variables, which we transform back to the preceding set of variables. We repeat the process until we recover the original variables. Thus, the complexity of the original problem is now distributed among the transformations, and the final form of the differential equations is simple. Not every perturbation

problem permits this type of approach, so generating the correct canonical transformation requires experience and insight.

The most desirable canonical transformations eliminate all the time-varying coordinates in the final Hamiltonian. To illustrate this process, let's use the zonal harmonic disturbing function. But we don't need to use it explicitly. Recall that, when expressed in classical orbital elements, the zonal disturbing function depends on just a , e , i , ω , and M . The longitude of the ascending node is an *ignorable coordinate* because it doesn't appear in the disturbing function. This is also true for a disturbing function expressed in Delaunay variables. The generalized momentum corresponding to ignorable coordinates is constant, so the search for canonical transformations is really a search for additional ignorable coordinates. If we express the disturbing function using the canonical Delaunay variables—which contain ω , Ω , and M — Ω remains an ignorable coordinate. Thus, the Hamiltonian depends on L_d , h , H_d , ω , and M . This first canonical transformation yields a transformed Hamiltonian with transformed variables designated by primes:

$$\mathcal{H}(L_d, h, H_d, \underline{\omega}, M) = \mathcal{H}^*(L'_d, h', H'_d, \underline{\omega'}, \underline{\underline{\omega}})$$

The underlines represent the ignorable coordinates. The next transformation yields the new Hamiltonian \mathcal{H}^{**} with new double-primed variables:

$$\mathcal{H}^*(L'_d, h', H'_d, \underline{\omega'}, \underline{\underline{\omega}}) = \mathcal{H}^{**}(L''_d, h'', H''_d, \underline{\underline{\omega}}, \underline{\underline{\Omega}})$$

The equations of motion for the doubly transformed Hamiltonian become

$$\begin{aligned} \dot{L}''_d &= \frac{\partial \mathcal{H}^{**}}{\partial M''} = 0 & \dot{M}'' &= -\frac{\partial \mathcal{H}^{**}}{\partial L''_d} = k_1 \\ \dot{h}'' &= \frac{\partial \mathcal{H}^{**}}{\partial \omega''} = 0 & \dot{\omega}'' &= -\frac{\partial \mathcal{H}^{**}}{\partial h''} = k_2 \\ \dot{H}''_d &= \frac{\partial \mathcal{H}^{**}}{\partial \Omega''} = 0 & \dot{\Omega}'' &= -\frac{\partial \mathcal{H}^{**}}{\partial H''_d} = k_3 \end{aligned}$$

Clearly, k_1 , k_2 , k_3 are constant because \mathcal{H}^{**} is constant and L''_d , h'' , and H''_d are constants. This follows from the vanishing rates above. Thus,

$$\begin{array}{ll} L''_d = \text{const}_1 & M'' = k_1 \Delta t + M''_o \\ h'' = \text{const}_2 & \omega'' = k_2 \Delta t + \omega''_o \\ H''_d = \text{const}_3 & \Omega'' = k_3 \Delta t + \Omega''_o \end{array}$$

We needed two canonical transformations; the second is inverted to give the single-primed elements as functions of the double-primed elements and the time. The first transformation is inverted to yield the original elements as functions of the single-primed elements.

This process is the basis of various theories for artificial satellites. Two examples include the work of Brouwer, discussed later, and the more recent work of Deprit (1969). The main difference in this class of analytical theories is the technique used to determine the canonical transformations. The other difference is the choice of the original set of canonical variables. Determining these canonical transformations isn't easy. In practice, the technique often depends on the theory of asymptotic expansions, which also provides the foundation for perturbation methods. Details of this process are beyond the current scope.

8.5 Disturbing-Potential Formulations

We can form potential functions for several major perturbing forces. Conservative forces are well suited to potential formulations. I'll present formulas for the forces of central body, third body, solar-radiation pressure, and tides.

8.5.1 Gravity Potential in Terms of the Satellite's Orbital Elements

Although the aspherical potential at point P in Fig. 7-3 is often developed in spherical coordinates (ϕ_{gc}, λ, r) of point P , it's also useful to represent the potential in terms of classical $(a, e, i, \Omega, \omega, M)$, equinoctial, or other orbital elements that specify the satellite at P . The transformation does little to improve computations, but it is convenient for the Lagrange planetary equations, and it permits us to analyze the effects on the orbital elements caused by specific perturbations. The partial derivatives of this form can be used to evaluate numerical, analytical, and semianalytical methods.

Consider a satellite in orbit about the Earth. Its position coordinates (distance, r , argument of latitude, $u = \omega + \nu$, and inclination, i) are uniquely related to osculating elements. Thus, we should be able to express the aspherical potential at the satellite's location in terms of classical orbital elements by a change of variable. Several authors discuss the development, notably Kaula (1966, 30–37), whose name is usually attributed to the technique, and Lambeck (1988, 235–236). Using Fig. 8-2, first transform to orbit-plane coordinates (u, i) by rotating the axes such that Greenwich moves into the node, followed by a rotation about the line of nodes through the inclination ($\text{ROT1}[i] \text{ ROT3}[\Omega]$). This eliminates ϕ_{sat} in favor of u, i, Ω , and θ_{GST} . Recognize that θ_{GST} equals $\omega_{\oplus}(t - t_o)$, as shown in Fig. 8-2. The next step transforms r and u to semimajor axis and mean anomaly using Eq. (4-51). Because we're interested only in the effect of the perturbations on the orbital elements, let's just transform the *disturbing* potential. Kaula (1966, 34–38) gives the final result:

$$U = \frac{\mu}{r} + \sum_{l=2}^{\infty} \sum_{m=0}^l \sum_{p=0}^l \sum_{q=-\infty}^{\infty} \frac{\mu}{a} \left(\frac{R_{\oplus}}{a} \right)^l F_{lmp}(i) G_{lpq}(e) S_{lmpq}(\omega, M, \Omega, \theta_{GST}) \quad (8-28)$$

The S functions given by Kaula (1966, 37) contain the gravitational coefficients:

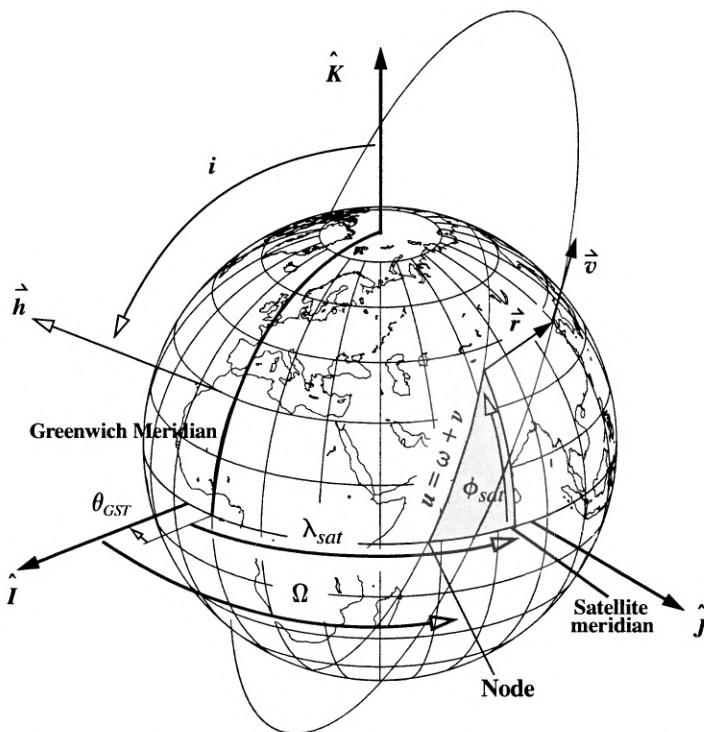


Figure 8-2. Potential Transformation. Converting the potential relies on the relation of the classical orbital elements and spherical coordinates (ϕ_{sat} , λ_{sat} , r).

$$S_{lmpq}(\omega, M, \Omega, \theta_{GST}) = \begin{cases} C_{lm} \cos(\Theta_{lmpq}) + S_{lm} \sin(\Theta_{lmpq}) & \text{if } (l-m) \text{ even} \\ -S_{lm} \cos(\Theta_{lmpq}) + C_{lm} \sin(\Theta_{lmpq}) & \text{if } (l-m) \text{ odd} \end{cases} \quad (8-29)$$

$$\Theta_{lmpq} = (l-2p)\omega + (l-2p+q)M + m(\Omega - \theta_{GST})$$

The F and G functions are simply expressions of the inclination and eccentricity, respectively:

$$F_{lmp}(i) = \sum_{t=0}^{\min(p, k)} \left\{ \frac{(2l-2t)!}{t!(l-t)!(l-m-2t)!} 2^{2l-2t} \sin^{l-m-2t}(i) \right. \\ \times \sum_{s=0}^m \binom{m}{s} \cos^s(i) \sum_c \binom{l-m-2t+s}{c} \binom{m-s}{p-t-c} (-1)^{c-k} \left. \right\} \quad (8-30)$$

$$\times \sum_{s=0}^m \binom{m}{s} \cos^s(i) \sum_c \binom{l-m-2t+s}{c} \binom{m-s}{p-t-c} (-1)^{c-k} \left\} \right.$$

Here, k is the integer part of $(l - m)/2$, and c is summed for all values for which the coefficients are not equal to zero. The eccentricity functions, G_{lpq} , are often called the **Hansen coefficients**. For $q = 2p - l$, we can use

$$G_{lpq}(e) = \frac{1}{(1-e^2)^{l-\frac{1}{2}}} \sum_{d=0}^{p'-1} \binom{l-1}{2d+l-2p'} \binom{2d+l-2p'}{d} \left(\frac{e}{2}\right)^{2d+l-2p'} \quad (8-31)$$

This expression comes from Kaula (1966, 36-37), which shows more formulas. Note here that $p' = p$ if $p \leq l/2$, and $p' = l - p$ if $p \geq l/2$. Kaula (1966), Cayley (1861), and others express the inclination, $F_{lmq}(i)$, and eccentricity functions $G_{lpq}(e)$. Table 8-1 shows some values. Notice that the eccentricity functions are of the order of the $|q|$ power of eccentricity whenever only the magnitude is a concern.

8.5.2 Third-Body Potential in Terms of the Satellite's Orbital Elements

We can also transform the potential for a third body and express it in terms of orbital elements. Although the development is very similar to Kaula's development of the gravity potential, there are several important differences. We begin with the expressions from Chap. 7 which express the third-body potential in terms of the Legendre polynomials.

$$R_{3-body} = \frac{\mu_3}{r_3} \sum_{l=2}^{\infty} \left(\frac{r}{r_3}\right)^l P_l [\cos(\zeta)]$$

Here, r is the geocentric distance, and μ_3 and r_3 are the third-body gravitational parameters and geocentric distance, respectively. We've already introduced the angle between the position vectors to the satellite and the Sun, ζ , in Chap. 7.

We use the addition theorem with which we developed the gravitational potential in Chap. 7, Eq. (7-13), to obtain

$$P_l[\cos(\zeta)] = \sum_{m=0}^l k_m \frac{(l-m)!}{(l+m)!} P_{lm} [\sin(\phi)] P_{lm} [\sin(\phi_3)] \cos(m(\alpha - \alpha_3))$$

The index, k_m , assumes values of 1 for $m = 0$, and 2 for $m \neq 0$, and α, δ and α_3, δ_3 are the right ascension and declination of the satellite and the third body respectively. Now define

$$A_{lm} = \frac{\mu_3}{r_3^{l+1}} k_m \frac{(l-m)!}{(l+m)!} P_{lm} [\sin(\delta_3)] \cos(m\alpha_3)$$

$$B_{lm} = \frac{\mu_3}{r_3^{l+1}} k_m \frac{(l-m)!}{(l+m)!} P_{lm} [\sin(\delta_3)] \sin(m\alpha_3)$$

TABLE 8-1. Inclination and Eccentricity Functions. For more complete lists, see Cayley (1861) and Kaula (1966, 34–38). Note that each $F(i)$ is a finite series, whereas the $G(e)$ expressions are infinite series. The multiple columns for the eccentricity functions denote identical terms.

l	m	p	Inclination Function, $F(i)$			l	p	q	Eccentricity Function, $G(e)$		
2	0	0	$-(3/8)\sin^2(i)$			2	1	-2	$(9/4)e^2 + (7/4)e^4 + \dots$		
2	0	1	$(3/4)\sin^2(i) - 1/2$			2	1	-1	$(3/2)e + (27/16)e^3 + \dots$		
2	0	2	F_{200}			2	1	0	$(1 - e^2)^{-3/2}$		
2	1	0	$(3/4)\sin(i)(1 + \cos(i))$			2	0	-2	0		
2	1	1	$-(3/2)\sin(i)\cos(i)$			2	0	-1	$-(1/2)e + (1/16)e^3 + \dots$		
2	1	2	$-(3/4)\sin(i)(1 - \cos(i))$			2	0	0	$1 - (5/2)e^2 + (13/16)e^4 + \dots$		
2	2	0	$(3/4)(1 + \cos(i))^2$			2	0	1	$(7/2)e - (123/16)e^3 + \dots$		
2	2	1	$(3/2)\sin^2(i)$			2	0	2	$(17/2)e^2 - (115/6)e^4 + \dots$		
2	2	2	$(3/4)(1 - \cos(i))^2$			3	1	-2	$(11/8)e^2 + (49/16)e^4 + \dots$		
3	0	0	$-(5/16)\sin^3(i)$			3	1	-1	$e(1 - e^2)^{-5/2}$		
3	0	1	$(15/16)\sin^3(i) - (3/4)\sin(i)$			3	1	0	$1 + 2e^2 + (239/64)e^4 + \dots$		
3	0	2	$-F_{301}$			3	1	1	$3e + (11/4)e^3 + \dots$		
3	0	3	$-F_{300}$			3	1	2	$(53/8)e^2 + (39/16)e^4 + \dots$		
3	1	0	$-(15/16)\sin^2(i)(1 + \cos(i))$			3	0	-2	$(1/8)e^2 + (1/48)e^4 + \dots$		
3	1	1	$(15/16)\sin^2(i)(1 + 3\cos(i)) - (3/4)(1 + \cos(i))$			3	0	-1	$-e + (5/4)e^3 + \dots$		
3	1	2	$(15/16)\sin^2(i)(1 - 3\cos(i)) - (3/4)(1 - \cos(i))$			3	0	0	$1 - 6e^2 + (423/64)e^4 + \dots$		
3	1	3	$-(15/16)\sin^2(i)(1 - \cos(i))$			3	0	1	$5e - 22e^3 + \dots$		
3	2	0	$(15/8)\sin(i)(1 + \cos(i))^2$			3	0	2	$(127/8)e^2 - (3065/48)e^4 + \dots$		
3	2	1	$(15/8)\sin(i)(1 - 2\cos(i) - 3\cos^2(i))$								
3	2	2	$-(15/8)\sin(i)(1 + 2\cos(i) - 3\cos^2(i))$								
3	2	3	$-(15/8)\sin(i)(1 - \cos(i))^2$								
3	3	0	$(15/8)(1 + \cos(i))^3$								
3	3	1	$(45/8)\sin^2(i)(1 + \cos(i))$								
3	3	2	$(45/8)\sin^2(i)(1 - \cos(i))$								
3	3	3	$(15/8)(1 - \cos(i))^3$								

Notice that we group the parameters relating to the third body into A_{lm} and B_{lm} as we did in the development of the Earth's gravitational potential. We now have

$$R_{3-body} = \sum_{l=2}^{\infty} r^l P_{lm} [\sin(\delta)] \left\{ A_{lm} \cos(m\alpha) + B_{lm} \sin(m\alpha) \right\}$$

This expression assumes we express the satellite's position in coordinates relative to the Earth's equator and the vernal equinox. To express the potential in terms of the orbital elements, we must rotate these spherical harmonics into the satellite's orbital plane. This rotation introduces the inclination, argument of perigee, and node because $\sin(\delta) = \sin(i)$ $\sin(\omega + \nu)$. We also introduce an inclination function F_{tmp} defined in Eq. (8-30). The result is

$$R_{3-body} = \sum_{l=2}^{\infty} r^l \sum_{m=0}^l \sum_{p=0}^l F_{tmp}(i) \left\{ \begin{aligned} & \left[\begin{matrix} A_{lm} \\ -B_{lm} \end{matrix} \right]_{t-m \text{ odd}}^{t-m \text{ even}} \cos [(l-2p) u + m\Omega] \\ & + \left[\begin{matrix} B_{lm} \\ A_{lm} \end{matrix} \right]_{t-m \text{ odd}}^{t-m \text{ even}} \sin [(l-2p) u + m\Omega] \end{aligned} \right\} \quad (8-32)$$

We often wish to express the parameters that depend on the third body in terms of the third body's orbital elements. To do so, we must rotate to the third body's orbital plane referenced to the third body's perigee.

$$\begin{aligned} A_{lm} &= \frac{\mu_3}{r_3^{l+1}} k_m \frac{(l-m)!}{(l+m)!} \sum_{h=0}^l F_{lmh}(i_3) \left[\begin{matrix} \cos \\ \sin \end{matrix} \right]_{t-m \text{ odd}}^{t-m \text{ even}} [(l-2h) u_3 + m\Omega_3] \\ B_{lm} &= \frac{\mu_3}{r_3^{l+1}} k_m \frac{(l-m)!}{(l+m)!} \sum_{h=0}^l F_{lmh}(i_3) \left[\begin{matrix} \sin \\ \cos \end{matrix} \right]_{t-m \text{ odd}}^{t-m \text{ even}} [(l-2h) u_3 + m\Omega_3] \end{aligned} \quad (8-33)$$

Substituting Eq. (8-33) into Eq. (8-32) and simplifying yields

$$R_{3-body} = \frac{\mu_3}{r_3} \sum_{l=2}^{\infty} \left(\frac{r}{r^3} \right)^l \sum_{m=0}^l k_m \frac{(l-m)!}{(l+m)!} \sum_{p=0}^l F_{tmp}(i) \sum_{h=0}^l F_{lmh}(i_3) \cos [(l-2p) (\omega + \nu) - (l-2h) (\omega_3 + \nu_3) + m (\Omega - \Omega_3)]$$

Next, we introduce the eccentricity and mean anomaly because the rotations are related to time. This requires a Fourier series expansion of the product

$$\left(\frac{r}{a} \right)^l \cos [(l-2p) (\omega + \nu) - (l-2h) (\omega_3 + \nu_3) + m (\Omega - \Omega_3)] = \sum_{q=-\infty}^{\infty} H_{lpq}(e) \Theta_{3-body sat}$$

$$\Theta_{3-body sat} = \cos (l-2p) \omega + (l-2p+q) M - (l-2h) (\omega_3 + \nu_3) + m (\Omega - \Omega_3)$$

The new eccentricity function of the satellite is also one of the class of functions known as *Hansen coefficients*. Usually, they are given as infinite power series in e^2 or in β^2 ,

where $\beta = e / (1 + \sqrt{1 - e^2})$. They differ from the $G_{l,pq}$ eccentricity functions introduced with the geopotential conversion because the $H_{l,hq'}$ functions result from a Fourier expansion of functions of the form $(r/a)^l \cos(m\nu)$, in contrast to $G_{l,pq}$ which result from expansions of the form $(a/r)^l \cos(p\nu)$. For the special case where $l - 2p + q = 0$, there are only secular and long-periodic contributions, and the eccentricity function takes a form of a finite series summation in β :

$$H_{l,p,2p-l}(e) = \frac{-\beta^{l-2p'}}{(1+\beta^2)^{l+1}} \binom{2l+1-2p'}{l-2p'} \sum_{k=0} \binom{l+1}{k} \binom{2p'+1}{k} \frac{\beta^{2k}}{\binom{l-2p'+k}{k}}$$

where $p' = p$ for $p \leq l/2$ and $p' = l - p$ for $p \geq l/2$. Similarly, we can expand the product

$$\begin{aligned} \left(\frac{a}{r_3}\right)^{l+1} \cos(l-2p)\omega + (l-2p+q)M - (l-2h)(\omega_3 + \nu_3) + m(\Omega - \Omega_3) \\ = \sum_{q=-\infty}^{\infty} G_{l+1,hq'}(e_3) \cos(\Theta_{3-body}) \end{aligned}$$

$$\Theta_{3-body} = (l-2p)\omega + (l-2p+q)M - (l-2h)\omega_3 - (l-2h+q')M_3 + m(\Omega - \Omega_3)$$

The eccentricity function is the same as in the development of the Kaula potential in the last section. Substituting the expression in yields the final form:

$$\begin{aligned} R_{3-body} = & \frac{\mu_3}{a_3} \sum_{l=2}^{\infty} \left(\frac{a}{a_3}\right)^l \sum_{m=0}^l k_m \frac{(l-m)!}{(l+m)!} \sum_{p=0}^l F_{tmp}(i) \sum_{h=0}^l F_{tmh}(i_3) \times \\ & \sum_{q=-\infty}^{\infty} \sum_{q'=-\infty}^{\infty} H_{l,pq}(e) G_{l+1,hq'}(e_3) \cos(\Theta_{3-body}) \end{aligned} \quad (8-34)$$

This expression is very similar to Eq. (8-28) except for two additional summations, the second inclination function, and the second type of eccentricity function.

This formulation is advantageous for isolating secular contributions and for treating resonance caused by a commensurability with the satellite's and the third-body's motions. However, if these goals aren't desired, the added complexity can be avoided by not introducing the second inclination function (for the third body's inclination) and the Fourier series expansion of the third-body's eccentricity and mean anomaly. The disturbing function retains the functions, A_{lm} , B_{lm} , which depend on the third body and which must be evaluated using the third-body's position vector. This form is often used for low-altitude satellites because the Moon moves only a short distance over one revolution of the satellite and the A_{lm} , B_{lm} functions don't need to be updated as often.

8.5.3 Tidal-Motion Potential in terms of the Satellite's Orbital Elements

In astrodynamics, we analyze two main types of tides: solid-Earth and oceanic. Fortunately, there is little coupling between the effects, and we can use a similar model for each. Kaula (1969) shows the general formula:

$$U_{tides} = \sum_{l=2}^{\infty} \sum_{m=0}^l \sum_{h=0}^l \sum_{q'=-\infty}^{\infty} k_l[\phi_{sat}, \lambda] r^l B_{lm} C_{lmhq'} P_{lm} \sin [\phi_{sat}] S_{lmhq'}(\omega, M, \Omega, \theta_{GST})$$

$$S_{lmhq'}(\omega, M, \Omega, \theta_{GST}) = \begin{cases} \cos(\Theta_{tides}) & \text{if } (l-m) \text{ even} \\ \sin(\Theta_{tides}) & \text{if } (l-m) \text{ odd} \end{cases} \quad (8-35)$$

$$\Theta_{tides} = (l-2h)\omega + (l-2h+q')M - \epsilon_{tmpq}[\phi_{sat}, \lambda] + m\Omega - m(\lambda + \theta_{GST})$$

The B and C functions are from the third-body potential [Eq. (8-34)]. They include the third-body inclination and eccentricity functions.

$$B_{lm} = \mu_3 k_m \frac{(l-m)!}{(l+m)!}$$

$$C_{lmhq'} = \left(\frac{1}{a_3}\right)^{l+1} F_{lmh}(i_3) G_{lhq'}(e_3)$$

We also introduce the *Love numbers*, $k_l[\phi_{sat}, \lambda]$, and the *phase lag*, $\epsilon_{tmpq}[\phi_{sat}, \lambda]$. A. Love introduced the **Love numbers** in 1909 as a way to describe the behavior of a solid, spherical, oceanless Earth in response to tidal motion. The **phase lag** represents long-term changes in the motion of the Sun and Moon and their effect on the tides. For astrodynamics, we usually assume these quantities don't vary with longitude. I won't derive additional formulations because they become complex very rapidly. You should consult Kaula (1969), Jursa (1985, 23-24 to 23-29), or Casotto (1989) for additional information.

8.6 Linearized Perturbations and Effects

Using the VOP technique, we can analyze the effects of perturbations on specific orbital elements. This is very useful in mission planning and analysis, so I'll present formulas and effects for each of the perturbing forces we've discussed so far. Each section presents various equations describing the effects of the perturbation on each orbital element. Then, I discuss some of the more dominant secular and periodic effects. Each section summarizes the dominant effects for that perturbing force.

As mentioned earlier, we want any theory to model as many perturbing forces as possible. Most operational analytical theories are limited to central body and drag. Analytical expressions for third-body and solar-radiation forces are far less common, mainly because their effects are much smaller for many orbits. Also, whenever accuracy requires us to use

effects of third bodies and solar-radiation pressure, numerical integration is usually just as easy for all the perturbing forces.

Finally, remember that much of the literature discusses only certain aspects of analytical theories. Many analytical theories are fairly accurate, but they're very complex. The equations I present are intended to give you a flavor of the process we follow when trying to extend or modify an existing theory. On the other hand, the insights for each perturbing force are intended for general application.

8.6.1 Central-Body Analysis

It shouldn't come as a surprise that the central body usually has the largest perturbative effect on the orbit of a satellite near Earth. After all, gravity is the only factor considered in two-body dynamics. But some of the motions can be quite unexpected. The first zonal harmonic, J_2 , receives a lot of attention. Some theories assume the Earth's oblateness (J_2 in particular) is the *only* perturbing effect on a satellite. Unfortunately, as we've seen, the aspherical-potential function is often modeled with an infinite series. Fortunately, only a few hundred terms are needed to establish highly precise orbits. The remaining harmonics beyond J_2 also affect satellite motion, especially those with degree and order less than about 30. In addition, under certain circumstances (resonance for instance), some tesseral terms may contribute much larger effects than we might expect. Coupling between central-body effects and other perturbations can also significantly affect certain orbits.

Equations for the Effects of Zonal Harmonics

Zonal harmonics cause secular motion in three orbital elements: node, argument of perigee, and mean anomaly. The derivation for secular effects begins with the aspherical potential [Eq. (7-18)]. This section follows Escobal ([1965] 1985, 363–371) and Roy (1988, 278–285). Remember that the aspherical potential includes all the gravitational effects, but because we're trying to develop equations for the *perturbed* results only, we must use the disturbing potential, R . Recall the general form of the disturbing function ($R = U - U_{2\text{-body}}$) and the potential given in Eq. (7-18). Let's consider the case where $l = 2$, thus representing the effects of a 2×2 gravity field, and subtract the spherical potential. Remember that we're using the satellite subpoint, ϕ_{sat} , λ_{sat} , and replacing terms with geocentric values ϕ_{gc} , λ . Thus,

$$R = -\frac{\mu}{r} \left[J_2 \left(\frac{R_\oplus}{r} \right)^2 P_2 [\sin(\phi_{gc})] + \left(\frac{R_\oplus}{r} \right)^2 \left\{ P_{21} [\sin(\phi_{gc})][C_{21} \cos(\lambda) + S_{21} \sin(\lambda)] + P_{22} [\sin(\phi_{gc})][C_{22} \cos(2\lambda) + S_{22} \sin(2\lambda)] \right\} \right]$$

The Legendre polynomial for $l = 2$ is

$$P_2[\gamma] = \frac{3}{2} \left(\gamma^2 - \frac{1}{3} \right)$$

If we ignore all periodic effects (tesseral and sectorial), the remaining part of the potential represents just J_2 . The potential becomes

$$R = -\frac{\mu J_2}{r} \left(\frac{R_\oplus}{r} \right)^2 \frac{3}{2} \left(\sin^2(\phi_{gc}) - \frac{1}{3} \right)$$

We want to express this function in terms of orbital elements. Using Eq. (5-12), we can relate inclination, latitude, and position in the orbit [$\sin(\phi_{gc}) = \sin(i) \sin(\omega + \nu)$]. Now,

$$R = -\frac{3\mu J_2}{2r} \left(\frac{R_\oplus}{r} \right)^2 \left(\sin^2(i) \sin^2(\omega + \nu) - \frac{1}{3} \right)$$

Let's use a trigonometric identity from Eq. (C-13) to find

$$R = -\frac{3\mu J_2}{2r} \left(\frac{R_\oplus}{r} \right)^2 \left(\frac{\sin^2(i)}{2} - \frac{\sin^2(i) \cos(2(\omega + \nu))}{2} - \frac{1}{3} \right)$$

Remember that our goal is to isolate the secular contribution, so we ignore the long- and short-periodic terms (those containing ω and ν) and write

$$\tilde{R} = -\frac{3\mu J_2}{2r} \left(\frac{R_\oplus}{r} \right)^2 \left(\frac{\sin^2(i)}{2} - \frac{1}{3} \right)$$

The tilde overbar indicates the disturbing function is changed by ignoring the appearance of ω and ν . Because $\mu = n^2 a^3$, let's separate out the term $(a/r)^3$ and find

$$\tilde{R} = -\frac{3}{2} n^2 R_\oplus^2 J_2 \left(\frac{a}{r} \right)^3 \left(\frac{\sin^2(i)}{2} - \frac{1}{3} \right)$$

Remember, we want only secular rates of change. The radial distance, r , changes periodically, i.e., over each revolution. Thus, we need to replace $(a/r)^3$ by its average over one period. Averaging the effects over one period is the same as averaging over one revolution of mean anomaly ($0 \leq M \leq 2\pi$), which gives us (Tisserand, 1889)

$$\begin{aligned} \overline{\left(\frac{a}{r} \right)^3} &= \frac{1}{2\pi} \int_0^{2\pi} \left(\frac{a}{r} \right)^3 dM = \frac{1}{2\pi} \int_0^{2\pi} \left(\frac{a}{r} \right)^3 r \frac{dr}{dE} dE \\ &= \frac{1}{2\pi} \int_0^{2\pi} \left(\frac{a}{r} \right)^3 \frac{r^2}{a^2 \sqrt{1-e^2}} d\nu = \frac{1}{2\pi (1-e^2)^{3/2}} \int_0^{2\pi} (1+e \cos(\nu)) d\nu \end{aligned}$$

where we transform variables $dM = \frac{r}{a}dE$ and $dE = \frac{r}{a\sqrt{1-e^2}}d\nu$. Thus,

$$\left(\frac{a}{r}\right)^3 = \frac{1}{(1-e^2)^{3/2}}$$

Replacing $(a/r)^3$ in the disturbing function by its average is effectively the same as averaging \tilde{R} over one revolution. This average, R_{avg} , is

$$R_{avg} = -\frac{3}{2}n^2R_\oplus^2J_2\left(\frac{1}{(1-e^2)^{3/2}}\right)\left(\frac{\sin^2(i)}{2} - \frac{1}{3}\right) \quad (8-36)$$

We can get the same result using Kaula's form of the potential, Eq. (8-28). In this case, $m = 0$ because of zonal harmonics, $q = 0$, and $p = 1/2$. The Kaula argument, Θ , is also zero. This leaves only F_{201} and G_{210} , and $J_2 = -C_{2,0}$, so we find the potential immediately!

$$R_{avg} = \frac{\mu\left(\frac{R_\oplus}{a}\right)^2}{a}F_{201}G_{210}\{C_{20}\cos(\Theta_{2010}) + S_{20}\sin(\Theta_{2010})\}$$

$$R_{avg} = -\frac{\mu J_2}{a}\left(\frac{R_\oplus}{a}\right)^2\left(\frac{3}{4}\sin^2(i) - \frac{1}{2}\right)\left(\frac{1}{(1-e^2)^{3/2}}\right)$$

The formula is reconciled with Eq. (8-36) by substituting $n^2a^3 = \mu$.

Now apply the Lagrange VOP equations with a disturbing function, $R = R_{avg}$. Remember, you don't have to use the R_{avg} if you want a more complete formulation. Look at

$$\frac{d\Omega}{dt} = \frac{1}{na^2\sqrt{1-e^2}\sin(i)} \frac{\partial R}{\partial i}$$

From Eq. (8-36) above, do the partial differentiation to get

$$\frac{\partial R_{avg}}{\partial i} = -\frac{3}{2}n^2R_\oplus^2J_2\left(\frac{1}{(1-e^2)^{3/2}}\right)\left(\frac{2\sin(i)\cos(i)}{2}\right) = -\frac{3}{2}n^2R_\oplus^2J_2\frac{\sin(i)\cos(i)}{(1-e^2)^{3/2}}$$

Combine this expression with the Lagrange VOP to find the secular rate of change,

$$\frac{d\Omega}{dt} = \frac{1}{na^2}\left(-\frac{3}{2}\right)n^2R_\oplus^2J_2\left(\frac{1}{(1-e^2)^2}\right)\cos(i) = -\frac{3nR_\oplus^2J_2}{2p^2}\cos(i)$$

The answer is

$$\dot{\Omega}_{sec} = -\frac{3nR_\oplus^2J_2}{2p^2}\cos(i) \quad (8-37)$$

Notice the units of the parameter, p , cancel because of the R_\oplus^2 in the numerator—an important consideration when using this equation. Programs typically have units of rad/TU

because this calculation is buried deep within a program. If you need an output, try deg/day for reasonable values. You can use the *mean* value for mean motion, but it isn't necessary. See Sec. 8.2 for more discussion on osculating and mean values.

Incorporating higher-order terms improves the accuracy of the first-order approximation; however, these terms quickly become complex. We should include the second-order J_2^2 contributions if other zonal terms are included because J_2^2 is of the same order as J_4 or J_6 ($0.001\ 082\ 63^2 = 0.000\ 001\ 2 \approx J_4$).

Integration of Eq. (8-37) provides the secular contribution to the node. This is sort of a "mean" node because the short-periodic and long-periodic contributions have been ignored. This raises the question of whether we use mean or osculating elements to evaluate the right-hand side of Eq. (8-37). Ideally, we should use mean elements for a consistent theory. However, for a first-order theory, it's immaterial whether we use mean or osculating elements on the right-hand sides of the equations because differences are on the order of J_2 , and the resulting errors will be on the order of J_2^2 . By virtue of our first-order theory, we've already decided to ignore contributions of this order.

In the following equations, all elements on the right-hand sides are assumed to have their ascending-node values. This is consistent with mean elements defined by the average over the argument of latitude ($u = \omega + \nu$) over the interval $[u \pm \pi]$. It's also consistent with the ad hoc process used to average R for the first-order analysis of the node rate. Merson (1961) has presented up to J_6 terms. King-Hele, Cook, and Rees (1963) extend $\dot{\Omega}_{sec}$ through J_{14} , whereas Kozai (1962) includes J_2^3 and the coupling terms $J_2 J_4$.

If you're familiar with these expressions, you'll note several differences. I modify the equations because (1) the equations should be computationally efficient, so I've used a lot of grouping and common denominators; and (2) the equations must be in a similar notation and form, so we can compare theories (Kozai, Brouwer, etc.) and rapidly evaluate their differences.

Using the first six zonal harmonics, we can obtain an expression for the secular effects on the longitude of the ascending node. From Merson (1961), the longitude of the ascending node rate is

$$\begin{aligned} \dot{\Omega}_{sec} = & -\frac{3J_2R_\oplus^2n\cos(i)}{2p^2} + \frac{3J_2^2R_\oplus^4n\cos(i)}{32p^4} \left\{ 12 - 4e^2 - (80 + 15e^2)\sin^2(i) \right\} \\ & + \frac{15J_4R_\oplus^4n\cos(i)}{32p^4} \left\{ 8 + 12e^2 - (14 + 21e^2)\sin^2(i) \right\} \\ & - \frac{105}{1024} \frac{J_6R_\oplus^6n\cos(i)}{p^6} \left\{ 64 + 160e^2 + 120e^4 - (288 + 720e^2 + 540e^4)\sin^2(i) \right. \\ & \quad \left. + (264 + 660e^2 + 495e^4)\sin^4(i) \right\} \end{aligned} \quad (8-38)$$

The derivation for apsidal rotation is simpler, in that we can use the intermediate results from deriving the nodal regression rate. Let's begin with the reduced form of the disturbing function, Eq. (8-36). Applying the Lagrange VOP equation, Eq. (8-12), with $R = R_{avg}$ results in

$$\frac{d\omega}{dt} = \frac{\sqrt{1-e^2}}{na^2e} \frac{\partial R}{\partial e} - \frac{\cot(i)}{na^2\sqrt{1-e^2}} \frac{\partial R}{\partial i}$$

From the derivation of Eq. (8-37), we know $\partial R/\partial i$. Let's examine $\partial R/\partial e$:

$$\begin{aligned}\frac{\partial R_{avg}}{\partial e} &= -\frac{3}{2}n^2R_\oplus^2J_2\left(\frac{\sin^2(i)}{2}-\frac{1}{3}\right)(1-e^2)^{-5/2}\left(-\frac{3}{2}\right)(-2e) \\ &= -\frac{9}{2}n^2R_\oplus^2J_2\frac{e}{(1-e^2)^{5/2}}\left(\frac{\sin^2(i)}{2}-\frac{1}{3}\right)\end{aligned}$$

Substituting into the $d\omega/dt$ equation above gives us

$$\begin{aligned}\frac{d\omega}{dt} &= \frac{\sqrt{1-e^2}}{na^2e}\left[-\frac{9}{2}n^2R_\oplus^2J_2\frac{e}{(1-e^2)^{5/2}}\left(\frac{\sin^2(i)}{2}-\frac{1}{3}\right)\right] \\ &\quad - \frac{\cot(i)}{na^2\sqrt{1-e^2}}\left[-\frac{3}{2}n^2R_\oplus^2J_2\left(\frac{1}{(1-e^2)^{3/2}}\right)\sin(i)\cos(i)\right]\end{aligned}$$

Collecting terms

$$\begin{aligned}\frac{d\omega}{dt} &= -\frac{9nR_\oplus^2J_2}{2a^2(1-e^2)^2}\left(\frac{\sin^2(i)}{2}-\frac{1}{3}\right) \\ &\quad + \frac{3nR_\oplus^2J_2}{2a^2(1-e^2)^2}[\cos^2(i)]\end{aligned}$$

and using the definition for the semiparameter results in

$$\frac{d\omega}{dt} = \frac{3}{4} \frac{nR_\oplus^2J_2}{p^2} \left\{ -3\sin^2(i) + 2 + 2\cos^2(i) \right\}$$

Now, we'll use the trigonometric identity for $[\cos^2(i)]$ and simplify:

$$\dot{\omega}_{sec} = \frac{3nR_\oplus^2J_2}{4p^2} \left\{ 4 - 5\sin^2(i) \right\} \quad (8-39)$$

We can also derive a more complex expression for the argument of perigee, including the first six zonal harmonics:

$$\begin{aligned}
 \dot{\omega}_{sec} = & \frac{3nJ_2R_{\oplus}^2}{4p^2} \left\{ 4 - 5\sin^2(i) \right\} \\
 & + \frac{9nJ_2^2R_{\oplus}^4}{384p^4} \left\{ 56e^2 + (760 - 36e^2)\sin^2(i) - (890 + 45e^2)\sin^4(i) \right\} \\
 & - \frac{15J_4R_{\oplus}^4n}{128p^4} \left\{ 64 + 72e^2 - (248 + 252e^2)\sin^2(i) + (196 + 189e^2)\sin^4(i) \right\} \quad (8-40) \\
 & + \frac{105J_6R_{\oplus}^6n}{2048p^6} \left\{ 256 + 960e^2 + 320e^4 - (2048 + 6880e^2 + 2160e^4)\sin^2(i) \right. \\
 & \left. + (4128 + 13080e^2 + 3960e^4)\sin^4(i) - (2376 + 14520e^2 + 2145e^4)\sin^6(i) \right\}
 \end{aligned}$$

Deriving the mean anomaly introduces some subtle differences. As with the previous derivations, let's begin with the Lagrange planetary equations [Eq. (8-12)]:

$$\frac{dM_o}{dt} = -\frac{1-e^2}{na^2e} \frac{\partial R}{\partial e} - \frac{2}{na} \frac{\partial R}{\partial a}$$

We know the partial derivative with respect to eccentricity from deriving the argument of perigee, and the partial derivative with respect to a is

$$\frac{\partial R_{avg}}{\partial a} = \frac{R_{avg}}{n^2} 2n \frac{dn}{da} = \frac{-3R_{avg}}{a}$$

for the $i = 2$ case. Therefore, combining gives us

$$\begin{aligned}
 \frac{dM_o}{dt} = & -\frac{1-e^2}{na^2e} \left\{ \frac{9}{2} n^2 R_{\oplus}^2 J_2 \frac{e}{(1-e^2)^{5/2}} \left(\frac{\sin^2(i)}{2} - \frac{1}{3} \right) \right\} \\
 & - \frac{2\mu R_{\oplus}^2 J_2}{na} \left(\frac{1}{(1-e^2)^{3/2}} \right) \left(\frac{\sin^2(i)}{2} - \frac{1}{3} \right)
 \end{aligned}$$

Simplifying yields

$$\begin{aligned}
 \frac{dM_o}{dt} = & \frac{9nR_{\oplus}^2 J_2}{2a^2(1-e^2)^{3/2}} \left(\frac{\sin^2(i)}{2} - \frac{1}{3} \right) - \frac{9nR_{\oplus}^2 J_2}{a^2(1-e^2)^{3/2}} \left(\frac{\sin^2(i)}{2} - \frac{1}{3} \right) \\
 \dot{M}_o = & \frac{3nR_{\oplus}^2 J_2 \sqrt{1-e^2}}{4p^2} \left\{ 3\sin^2(i) - 2 \right\} \quad (8-41)
 \end{aligned}$$

The mean anomaly's rate of change due to the first six secular zonal harmonics is

$$\begin{aligned}
 \dot{M}_o = & \frac{3nR_{\oplus}^2 J_2 \sqrt{1-e^2}}{4p^2} \left\{ 3\sin^2(i) - 2 \right\} \\
 & + \frac{3nR_{\oplus}^4 J_2^2}{512p^4 \sqrt{1-e^2}} \left\{ 320e^2 - 280e^4 + (1600 - 1568e^2 + 328e^4) \sin^2(i) \right. \\
 & \left. + (-2096 + 1072e^2 + 79e^4) \sin^4(i) \right\} \\
 & - \frac{45J_4 R_{\oplus}^4 e^2 n \sqrt{1-e^2}}{128p^4} \left\{ -8 + 40\sin(i) - 35\sin^2(i) \right\} \\
 & + \frac{35J_6 R_{\oplus}^6 n \sqrt{1-e^2}}{2048p^6} \left\{ -128 + 320e^2 + 240e^4 + (1344 - 3360e^2 - 2520e^4) \sin(i) \right. \\
 & \left. + (-1512 + 3780e^2 + 2835e^4) \sin^2(i) - (-1848 + 4620e^2 + 3465e^4) \sin^4(i) \right\} \tag{8-42}
 \end{aligned}$$

Formulas for long-periodic motion vary significantly among the methods and are complex, so I won't present specific formulas.

Short-periodic variations are often less complex than long-periodic variations. Indeed, the following short-periodic variations are identical in Kozai and Brouwer (see Sec. 8.7.2 and Sec. 8.7.3).

$$\begin{aligned}
 \Delta a_{SP} = & \frac{J_2 R_{\oplus}}{a} \left[\left(\frac{a}{r} \right)^3 - \frac{1}{(1-e^2)^{3/2}} \right. \\
 & \left. + \left\{ -\left(\frac{a}{r} \right)^3 + \frac{1}{(1-e^2)^{3/2}} + \left(\frac{a}{r} \right)^3 \cos(2\omega + 2\nu) \right\} \frac{3\sin^2(i)}{2} \right] \\
 \Delta e_{SP} = & \frac{J_2 R_{\oplus}^2}{4} \left[\frac{-2}{a^2 e \sqrt{1-e^2}} + \frac{2a(1-e^2)}{er^3} + \left\{ \frac{3}{a^2 e \sqrt{1-e^2}} - \frac{3a(1-e^2)}{er^3} \right. \right. \\
 & \left. - \frac{3(1-e^2) \cos(\nu + 2\omega)}{p^2} - \frac{3 \cos(2\nu + 2\omega)}{a^2 e (1-e^2)} \right. \\
 & \left. + \frac{3a(1-e^2) \cos(2\nu + 2\omega)}{er^3} - \frac{(1-e^2) \cos(3\nu + 2\omega)}{p^2} \right\} \sin^2(i) \left. \right]
 \end{aligned}$$

$$\Delta i_{SP} = \frac{J_2 R_\oplus^2 \sin(2i)}{8p^2} \left\{ 3 \cos(2\omega + 2\nu) + 3e \cos(2\omega + \nu) + e \cos(2\omega + 3\nu) \right\}$$

$$\begin{aligned} \Delta \Omega_{SP} = & -\frac{J_2 R_\oplus^2 \cos(i)}{4p^2} \left[6(\nu - M + e \sin(\nu)) \right. \\ & \left. - 3 \sin(2\omega + 2\nu) - 3e \sin(2\omega + \nu) - e \sin(2\omega + 3\nu) \right] \end{aligned}$$

$$\begin{aligned} \Delta \omega_{SP} = & \frac{3J_2 R_\oplus^2}{2p^2} \left[\left(2 - \frac{5}{2} \sin^2(i) \right) (\nu - M + e \sin(\nu)) \right. \\ & + \left(1 - \frac{3}{2} \sin^2(i) \right) \left\{ \frac{1}{e} \left(1 - \frac{1}{4} e^2 \right) \sin(\nu) + \frac{1}{2} \sin(2\nu) + \frac{e}{12} \sin(3\nu) \right\} \\ & - \frac{1}{e} \left\{ \frac{1}{4} \sin^2(i) + \left(\frac{1}{2} - \frac{15}{16} \sin^2(i) \right) e^2 \right\} \sin(\nu + 2\omega) \\ & + \frac{e}{16} \sin^2(i) \sin(\nu - 2\omega) - \frac{1}{2} \left(1 - \frac{5}{2} \sin^2(i) \right) \sin(2\nu + 2\omega) \\ & + \frac{1}{e} \left\{ \frac{7}{12} \sin^2(i) - \frac{1}{6} \left(1 - \frac{19}{8} \sin^2(i) \right) e^2 \right\} \sin(3\nu + 2\omega) \\ & \left. + \frac{3}{8} \sin^2(i) \sin(4\nu + 2\omega) + \frac{e}{16} \sin^2(i) \sin(5\nu + 2\omega) \right] \end{aligned}$$

$$\begin{aligned} \Delta M_{SP} = & \frac{3J_2 R_\oplus^2 \sqrt{1-e^2}}{2ep^2} \left[- \left(1 - \frac{3}{2} \sin^2(i) \right) \left\{ \left(1 - \frac{e^2}{4} \right) \sin(\nu) + \frac{e}{2} \sin(2\nu) + \frac{e^2}{12} \sin(3\nu) \right\} \right. \\ & + \sin^2(i) \left\{ \frac{1}{4} \left(1 + \frac{5}{4} e^2 \right) \sin(\nu + 2\omega) \right. \\ & - \frac{e^2}{16} \sin(\nu - 2\omega) - \frac{7}{12} \left(1 - \frac{e^2}{28} \right) \sin(3\nu + 2\omega) \\ & \left. \left. - \frac{3e}{8} \sin(4\nu + 2\omega) - \frac{e^2}{16} \sin(5\nu + 2\omega) \right\} \right] \end{aligned}$$

Effects of the Central Body

We can examine the effects caused by the central body physically and mathematically. Let's briefly introduce both methods before describing the individual effects.

The magnitude of the zonal-harmonic effects suggests a physical interpretation to aid understanding. If we include only the terms for the zonal harmonics, the relevant part of the aspherical-potential function from Eq. (7-18) is

$$U_{zonal} = \frac{\mu}{r} \left[1 - \sum_{l=2}^{\infty} J_l \left(\frac{R_{\oplus}}{r} \right)^l P_l [\sin(\phi_{sat})] \right]$$

with values for the zonal harmonics presented in Table D-1. Remember that the latitude is a geocentric value. Now we can examine individual orbital elements using this relation. We know that the central-body disturbing force is *conservative* (mass and energy remain constant), and the total energy (kinetic and potential) is a constant of the motion. The potential energy adds to the kinetic energy for the total energy, ξ , but recall our choice of constants in Chap. 2 for two-body motion (Sec. 2.2.3). We chose $c = 0$; thus, all potential energy for satellites is negative, unless they're on escape trajectories. For this reason, we must add a negative sign to U_{zonal} to obtain

$$\xi = KE + PE = \frac{v^2}{2} - \frac{\mu}{r} \left[1 - \sum_{l=2}^{\infty} J_l \left(\frac{R_{\oplus}}{r} \right)^l P_l [\sin(\phi_{sat})] \right] = \text{constant}$$

Using $\xi = -\mu/2a$ for the energy expression and solving for $1/a$ gives us

$$\frac{1}{a} = -\frac{2}{\mu} \left[\frac{v^2}{2} - \frac{\mu}{r} + \frac{\mu}{r} \sum_{l=2}^{\infty} J_l \left(\frac{R_{\oplus}}{r} \right)^l P_l [\sin(\phi_{sat})] \right]$$

The two-body semimajor axis, a , was defined in terms of the energy as

$$\xi_2 = -\frac{\mu}{2a} = \frac{v^2}{2} - \frac{\mu}{r}$$

Now, we'll substitute the expression for the spherical two-body energy, ξ_2 , into the first energy equation to obtain a formula that gives us insight into the qualitative effects of the disturbing force:

$$\frac{1}{a} = -\frac{2\xi_2}{\mu} - \frac{2}{r} \sum_{l=2}^{\infty} J_l \left(\frac{R_{\oplus}}{r} \right)^l P_l [\sin(\phi_{sat})]$$

For repeating orbits, the position magnitude, r , is periodic as it travels from apogee to perigee and back, and P_l is always periodic (it's a spherical harmonic). Hence, the variations in the semimajor axis are purely *periodic*.

Because the gravity field is axially symmetrical for zonal harmonics, the component of the angular momentum about the polar axis must be a constant of the motion. That is,

$$h \cos(i) = \text{constant} \quad (8-43)$$

If there was a secular change in inclination, at some time the orbit would be polar ($i = 90^\circ$). In this case, $h \cos(i) = 0$, which violates Eq. (8-43). Thus, inclination can't change secularly, and any variations must be purely *periodic*.

From axial symmetry, it also follows that, if at some instant the velocity vector lies in a plane through the polar axis ($i = 90^\circ$), no force will be tending to pull the satellite out of this plane. Hence, the polar orbit lies in a fixed plane ($i = 90^\circ$ and $\dot{\Omega} = 0$). We can also write Eq. (8-43) as

$$h \cos(i) = \sqrt{\mu a(1-e^2)} \cos(i) = \text{constant} \quad (8-44)$$

Because changes in semimajor axis and inclination can only be periodic, Eq. (8-44) dictates that effects on eccentricity must also be *periodic*. Consistent with the previous discussion, there are no secular variations for semimajor axis, eccentricity, and inclination. Thus,

$$\left(\frac{da}{dt}\right)_{\text{sec}} = \left(\frac{de}{dt}\right)_{\text{sec}} = \left(\frac{di}{dt}\right)_{\text{sec}} = 0$$

We can also examine the general nature of the central-body effects using a mathematical approach. This method introduces perturbations caused by the aspherical central body using the transformed potential, Eq. (8-28), which better enables us to categorize the specific effects. According to Fig. 8-1, only three effects require discussion: secular, short-periodic, and long-periodic. But we can distinguish two other important types of perturbations: *m-dailies* and resonance. The starting point for each category is the transformed disturbing-potential function in Eq. (8-28). It's written here with the *S*-function explicitly defined because the analysis proceeds from this point:

$$R = \frac{\mu}{a} \sum_{l=2}^{\infty} \sum_{m=0}^l \left(\frac{R_{\oplus}}{a}\right)^l \sum_{p=0}^l F_{tmp}(i) \sum_{q=-\infty}^{\infty} G_{lpq}(e) \begin{cases} C_{lm} \cos(\Theta) + S_{lm} \sin(\Theta) \\ -S_{lm} \cos(\Theta) + C_{lm} \sin(\Theta) \end{cases} \begin{cases} \text{if } (l-m) \text{ even} \\ \text{if } (l-m) \text{ odd} \end{cases} \quad (8-45)$$

$$\Theta = \Theta_{tmpq} = (l-2p)\omega + (l-2p+q)M + m(\Omega - \theta_{GST})$$

To get insight into the effects of an aspherical Earth's harmonic coefficients, we must examine the trigonometric term in the *S*-function of Eq. (8-29), Θ_{tmpq} , which we'll call the **Kaula gravitational argument**. Notice that both the sine and cosine functions of this argument appear. The terms which influence the values of Θ_{tmpq} are the argument of perigee, ω ; the *order*, m , which will *always* be lower-case so it's not confused with the mean anomaly, M ; the Greenwich Sidereal Time, θ_{GST} ; the node, Ω ; and the indices l , p , and q . A particular set of values can make Θ_{tmpq} behave in a certain manner. Each of these values characterizes a type of perturbation effect. Table 8-2 shows the requirements for each category.

TABLE 8-2. Categories of Central-Body Perturbations. The general behavioral characteristics of perturbations are shown, together with indices and gravitational coefficients that cause them. Notice, *all* coefficients cause short-periodic variations. The p and q indices must always be integer values.

Category	Conditions	p	q	Period P	Comments
Zonal Harmonics ($m = 0$)					
Secular	$l - 2p + q = 0$ $l - 2p = 0$	$p = l/2$	0	∞	Even zonals
Long periodic	$l - 2p + q = 0$ $l - 2p \neq 0$	p	$q = 2p - l$	$\approx \frac{2\pi}{\dot{\omega}}$	Longer periods
Short periodic	$l - 2p + q \neq 0$ $l - 2p = 0$	$p = l/2$	q	$\approx \frac{2\pi}{\dot{M}}$	Even zonals
LP/SP/Beat	$l - 2p + q \neq 0$ $l - 2p \neq 0$	p	q	$\approx \frac{2\pi}{\dot{M}}, \frac{2\pi}{\dot{\omega}}$	two frequencies
Sectorial and Tesseral Harmonics ($m \neq 0$)					
m -daily	$l - 2p + q = 0$ $l - 2p = 0$	$p = l/2$	$q = 0$	$\approx \frac{2\pi}{m\omega_{\oplus}}$	Even degree (l) all orders (m)
LP/MD/Beat	$l - 2p + q = 0$ $l - 2p \neq 0$	p	$q = 2p - l$		
LP/MD/SP/Beat Linear combination	$l - 2p + q \neq 0$ $l - 2p \neq 0$	p	q		
Resonance	$l - 2p \neq 0$	p	0	\approx weeks to years	repeat ground-track
	$l - 2p = 0$	$p = l/2$	q	\approx weeks to years	anomalistic resonance

In addition to knowing the categories of variations, we must estimate the period of oscillation. Some effects will vary rapidly (and usually small in magnitude), whereas others vary slowly (and usually have larger magnitudes). Determining the oscillation period rests on the rate of change of Θ_{lmpq} . Here, we have

$$\dot{\Theta}_{lmpq} = (l - 2p) \dot{\omega} + (l - 2p + q) \dot{M} + m (\dot{\Omega} - \dot{\theta}_{GST}) \quad (8-46)$$

Now we can recognize \dot{M} is dominated by the mean motion (n) and take $\dot{\theta}_{GST}$ as the Earth's angular velocity, ω_{\oplus} . We developed sample values for the secular rates of change earlier [$\dot{\omega}$ in Eq. (8-39), $\dot{\Omega}$ in Eq. (8-37), and \dot{M} in Eq. (8-41)]. The equations permit ini-

tial analysis of the effects of each element on the trigonometric term. For a typical satellite in low-Earth orbit,

$$\dot{\Omega} = 5^\circ \text{ to } 10^\circ / \text{day}, \dot{\omega} = 5^\circ \text{ to } 10^\circ / \text{day} \quad (8-47)$$

$$\dot{M}_{2\text{-body}} = n = 13 \text{ to } 15 \text{ revs / day} = 4700^\circ \text{ to } 5400^\circ / \text{day}$$

$$\dot{\theta}_{GST} = \omega_{\oplus} \equiv 361^\circ / \text{day}$$

These values are important in evaluating the dominant effects on the satellite. It's important to point out that secular, then resonant, and long-periodic terms produce ever smaller effects on the motion. Consider a modified element rate of

$$\dot{c} = c_o + \alpha t + \beta \cos(\omega_{\oplus} t) + \gamma \cos(\dot{\omega} t) + \delta \cos(nt)$$

Integrating this equation while assuming only time is variable yields

$$c = c_o + \alpha t + \beta \frac{\sin(\omega_{\oplus} t)}{\omega_{\oplus}} + \frac{\gamma \sin(\dot{\omega} t)}{\dot{\omega}} + \frac{\delta \sin(nt)}{n}$$

The secular term αt grows without bound at a rate of α . The periodic terms oscillate with a resonance period of $2\pi/\omega_{\oplus}$, the long period, $2\pi/\dot{\omega}$, and the short period, $2\pi/n$. The short-periodic variations let us create a **fundamental period** as a baseline for all oscillations.

$$\rho = 2\pi/n$$

The amplitudes are proportional to $1/\omega_{\oplus}$, $1/\dot{\omega}$, and $1/n$, so the smaller the rate, the greater the amplitude. Thus, long-periodic terms produce larger oscillations than short-periodic terms, and deep resonance produces even larger amplitude effects than the long-periodic terms. Shallow resonance effects are usually smaller relative to the deep resonance, and possibly smaller than the long-periodic amplitudes, depending on the particular orbit.

Secular Effects from Zonal Harmonics

The first-order secular effects on an Earth satellite resulting from the nonspherical gravity field come from the even zonal (axially symmetrical) harmonics of the Earth's gravity field. The governing equation for examining the secular effects is Eq. (8-45), with $m = 0$ and l even. Therefore,

1. $m = 0$ by definition, so $m(\Omega - \theta_{GST}) = 0$.
2. Secular effects require $(l - 2p) = 0$ and $(l - 2p + q) = 0$. Thus $q = 0$, $p = l/2$, so l must always be even. This shows that all secular variations result only from even zonal harmonics.
3. $\Theta_{lmpl} = 0$ for secular contribution

Any other combination of indices will produce only periodic effects.

The simplest way to evaluate the perturbations is to plug the potential into the Lagrange planetary equations and examine the effect. This is precisely the technique we used to develop analytical solutions in Sec. 8.3.1.

We've seen the presence of secular effects in geometrical and mathematical examples. This phenomenon has a simple physical explanation—the added attraction of the Earth's equatorial bulge introduces a force component toward the equator. Consequently, the resultant acceleration will cause the satellite to reach the equator (node) short of the crossing point for a spherical Earth (see Fig. 8-3). This phenomenon causes regression of the node (for direct orbits) or a counter rotation of the orbital plane around the polar axis (Fig. 8-3). The arrows show the direction of force due to equatorial bulge. An expression for the node's secular rate of change appears in Eq. (8-37).

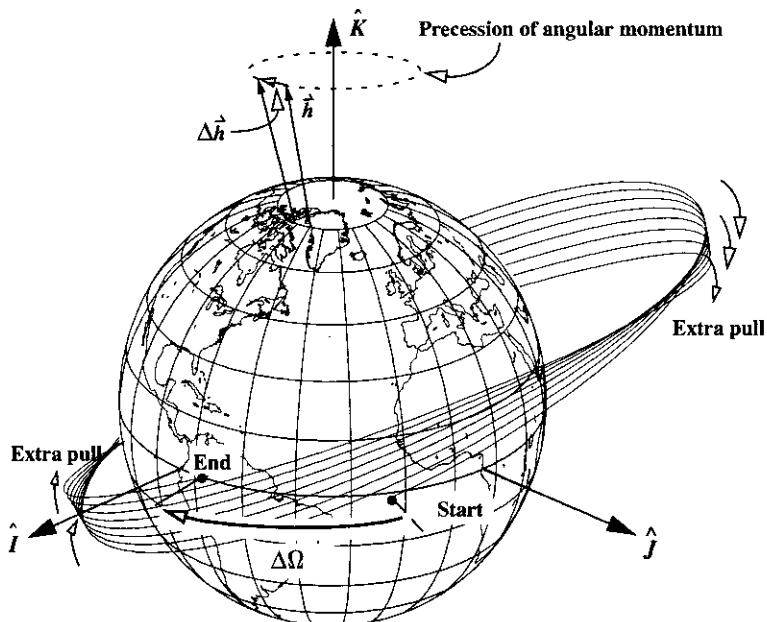


Figure 8-3. How Nodal Regression Works. The effect of a J_2 perturbation that is 20 times the real value of J_2 shows the precession of the longitude of ascending node. The perturbation manifests itself through a change in the angular-momentum vector, and the node regresses much like a precessing top. For this example, the correct interpretation is to say that the satellite's line of nodes experiences a rotation of $50/8^\circ$ per revolution, or 50° per unit of time. The figure is also distorted because there is some apsidal rotation with an inclination not at the critical inclination (63.4°).

We can also see nodal regression as a gyroscopic phenomenon. The torque on the orbit due to the equatorial bulge is about the line of nodes and tends to turn the orbital plane

toward the equator. Because the change in angular momentum, $\Delta\vec{h}$, is parallel to the torque, the precession is parallel to the line of nodes and perpendicular to the angular momentum (Fig. 8-3). The orbit behaves like a spinning top undergoing a torque and precesses as shown. Recall that the rate of change of the angular-momentum vector of a precessing top aligns with the applied-torque vector; thus, in the figure above, the angular momentum moves to the left because the torque effect of the equatorial bulge is in the negative J_1 -direction (at the instant shown, t_{end}).

Figure 8-4 shows the amount of nodal regression for various orbits. Remember that the figure represents only the *secular effects* of J_2 . Although J_2 effects dominate the motion of the node, the other gravitational harmonics can add a lot to these effects. Indeed, when the effect from J_2 is small (around 90° inclination), the periodic effects from the sectorial and tesseral coefficients can be much larger than the secular effect shown above. Analytical solutions use equations such as Eq. (8-37) to determine the change in the node over time:

$$\Omega = \Omega_o + \dot{\Omega}\Delta t$$

where Ω_o is the initial value of the node. As we saw with Eq. (8-37) and Eq. (8-38), several expressions exist for $\dot{\Omega}$. There is a zero point whenever inclination is exactly 90° and the secular effect of J_2 is zero.

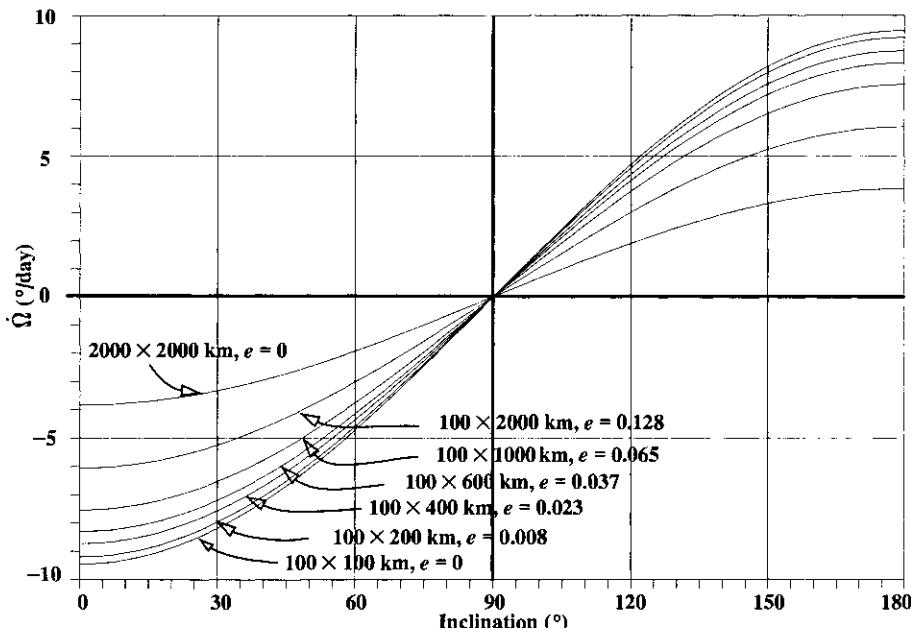


Figure 8-4. Daily Nodal Regression in °/day. For the eccentric orbits, I used a perigee altitude of 100 km with the apogee values as indicated. The 100×2000 km orbit shows that the perturbing effect for each of the *eccentric* orbits would be smaller if the orbit were circular at that apogee altitude (2000×2000 km).

The next major effect of the secular variations caused by J_2 deals with apsidal motion, as Fig. 8-5 shows. Again, it's useful to visualize this effect by examining Fig. 8-6, which I created using Eq. (8-39) and which shows the rate of change for the location of perigee.

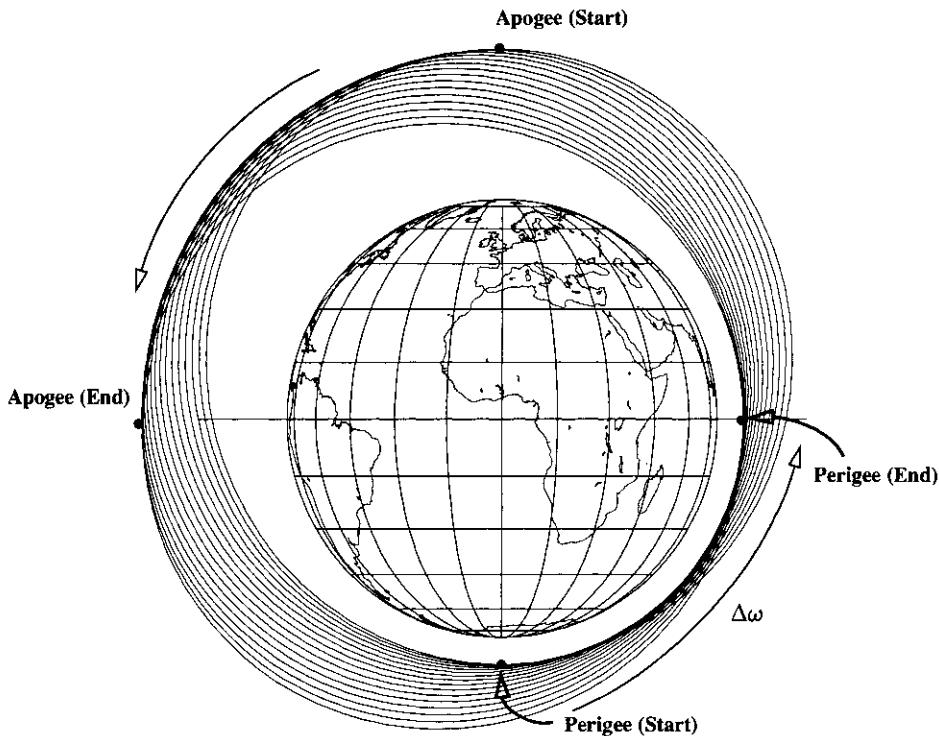


Figure 8-5. How Apsidal Rotation Works. We can see the effect of apsidal rotation after the J_2 perturbation effect increases by a factor of 40 for a polar orbit (to eliminate the nodal regression). Notice how perigee (and apogee) locations change dramatically in a few revolutions.

Notice that certain values of inclination can cause $\dot{\omega}$ to vanish, creating a critical inclination. The **critical inclination**, i_ω , occurs at values of 63.4° ($\sin(i) = 2/\sqrt{5}$) and 116.6° . The literature commonly refers to the critical inclination of 63.4° because it can be used to reduce the effect of J_2 perturbations. Be aware that, when we consider other perturbations such as J_3 , the critical inclination changes slightly.

The entire concept of the critical inclination is a celebrated controversy within the astrodynamics community that traces its origins to the work of A. A. Orlov in 1953. His work developed the **Main Problem**—the problem of a theory for artificial satellites

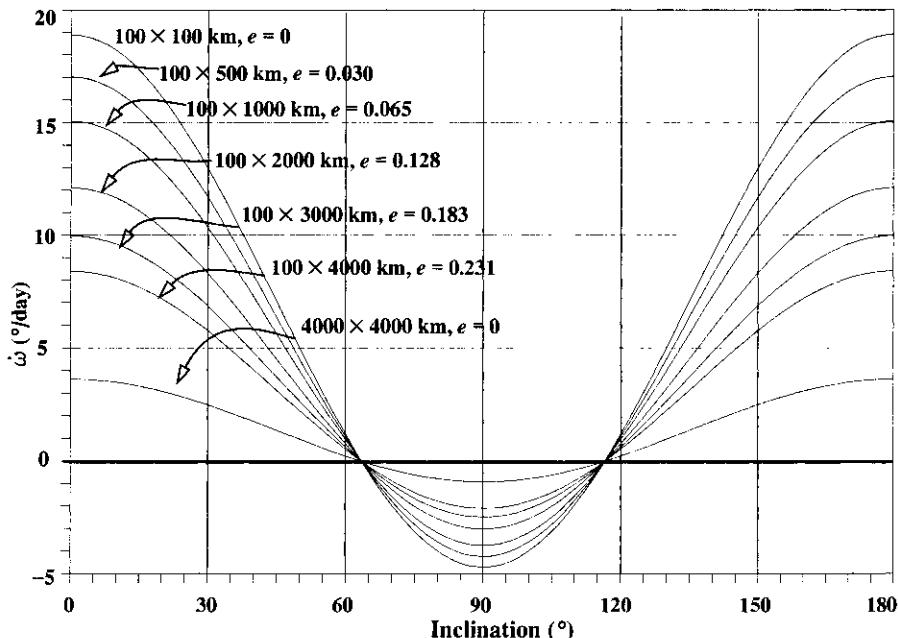


Figure 8-6. Daily Apsidal Regression in °/day. As with nodal regression, circular orbits at an altitude (say, 4000×4000) would have a smaller daily change than an eccentric orbit with apogee at the same altitude (100×4000 km).

defined by two-body motion and perturbed by the J_2 nonspherical term. It was observed that Hamiltonian solutions were plagued by the appearance of a small divisor in the second canonical transformation. This caused the analytical solution to break down and become invalid when the inclination was near this critical inclination of about 63.4° . The controversy centered on whether or not this singularity represented a basic characteristic of the problem, or if it was an artifact of the mathematics used in the approximation. After nearly 30 years of controversy, Deprit and his colleagues demonstrated that this problem represents an essential characteristic of the stability of satellite motion in the Main Problem. The critical inclination is associated with an unstable equilibrium in the phase space of the Main Problem. It's called a **bifurcation point**, and it represents a concrete example of chaos theory in dynamics. Further treatment is beyond our scope, so consult Deprit and Coffey (1980) for additional information.

The remaining element with a secular rate to receive attention is the mean anomaly, M . The secular rate of change for the mean anomaly, independent of the mean motion, caused by J_2 is given in Eq. (8-41). The other secular contribution is caused by the mean motion, which depends on the semimajor axis. J_2 doesn't contribute any secular or long-periodic

motion in a because \dot{a} depends only on terms containing M (only $\partial R/\partial M$ is required). J_2 does contribute short-periodic variations and changes the mean value to be different from the unperturbed two-body value as recognized by Kozai, because of the additional mass in the disturbing potential. We'll explore this in greater detail in Sec. 10.4.2.

Long-Periodic Effects from Zonal Harmonics

Long-periodic variations are especially important in planning lengthy missions. Remember that the governing equation is Eq. (8-45), where $\Theta_{LP} = (l - 2p)\omega$. Long-periodic effects result from terms with indices of $(l - 2p + q) = 0$, $l - 2p \neq 0$, and $m = 0$. Thus, we can see that long-periodic effects are driven by the apsidal rotation rate. A typical period for the fundamental long-periodic effect on a near-Earth satellite is 20 to 60 days. Harmonics of the fundamental period will also be observed at $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$, of this period. The actual value of the fundamental period for long-periodic effects from zonal harmonics is $2\pi/\omega$ and it depends on the inclination [see Eq. (8-39)]. Note that larger or smaller values of the fundamental period can be obtained depending on the value of i .

The magnitude of these effects are usually larger than the short-periodic variations from the zonal harmonics. In fact, the magnitude of J_2 long-periodic effects can be very large. Consider the long-periodic change in a satellite's apogee location. Over an extended period of time, a site will see both perigee and apogee—a difference of $2ae$. This can be several hundred kilometers! The remaining zonal harmonics produce oscillations that are much smaller in magnitude (by a factor of about 1000).

Short-Periodic Effects from Zonal Harmonics

Short-periodic variations cause oscillations in the elements with a fundamental period equal to the Keplerian period ($2\pi/n$). Excluding J_2 , amplitudes of these oscillations are usually the smallest of all periodic contributions. J_2 produces effects for a near-Earth satellite position on the order of 8–10 km. The effect diminishes as the altitude increases, reducing the effect to a km or two for geosynchronous satellites. Short-periodic effects due to other zonal harmonics are very much smaller, and are usually less than 50–100 m.

We consider the short-periodic oscillations to be pure because they're caused by those terms in the zonal harmonic's disturbing potential function that satisfy $l - 2p = 0$. Thus, $l = 2p$ because $m = 0$ for zonal harmonics, and only even zonal harmonics cause the short-periodic contributions. The effects add to the fundamental period with additional harmonics having periods of $2\pi/qn$, where q is the integer index.

Long-Periodic/Short-Periodic Beat Period from Zonal Harmonics

This component of the motion is caused by the interaction of short-periodic and long-periodic variations. This interaction creates a beat period characterized by a high-frequency (short-periodic) oscillation with an amplitude that oscillates at the long-periodic frequency. The result is a high-frequency oscillation contained in a long-periodic envelope, as shown in Fig. 8-7. The magnitudes of these contributions obviously depend on time and

vary from the size of the short-periodic variations up to the size of the long-periodic contributions. Remember that the SP variations are quite small, except for those due to J_2 .

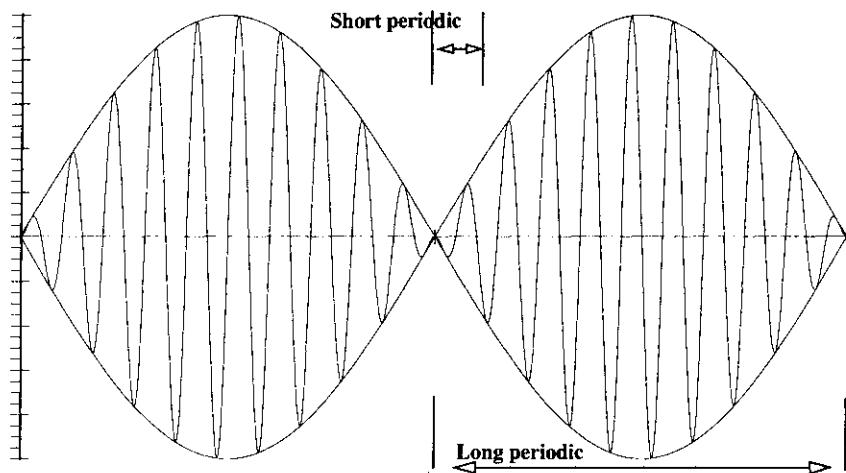


Figure 8-7. Beat-Period Coupling. This figure shows the coupling of short-periodic (SP) and long-periodic (LP) variations for LP/SP beat variations due to zonal harmonics. Over time, a beat phenomenon occurs as a result of combining the two rates. This figure can also represent the coupling of long-periodic and m -daily periods for long-periodic and m -daily beat periods contributed by sectorial and tesseral harmonics.

m-Daily Effects from Sectorial and Tesseral Harmonics

The m -daily effects are oscillations in the orbital elements caused by tesseral and sectorial terms in the potential that satisfy $t - 2p = 0$ and $t - 2p + q = 0$. Consequently, only even values of t apply, $p = t/2$, and $q = 2p - t$. To gain insight into m -daily effects, recall the transformed aspherical potential, written here in Kaula's form, Eq. (8-45), with

$$\Theta_{m\text{-daily}} = m(\Omega - \theta_{GST})$$

Periods are roughly fractional multiples of a day and are therefore often larger than the orbital period. Usually the nodal rate is small compared to the Earth's rotation rate ($\omega_{\oplus} \gg \dot{\Omega}$), and with Eq. (8-46), the period is approximately

$$P_{m\text{-daily}} = \frac{2\pi}{m|\dot{\Omega} - \omega_{\oplus}|} \approx \frac{2\pi}{m\omega_{\oplus}}$$

Only geopotential terms of even degree (l) and all orders ($m \leq l$) contribute. Terms of degree l and order $m = 1$ contribute oscillations of 24-hour periods, whereas $m = 2$ terms contribute 12-hour oscillations, and so on. These fractional multiples of 24 are called harmonics of the fundamental 24-hour period. Higher harmonics (e.g., $m = 14$) can contribute

oscillations with periods roughly the same as the basic short-periodic effects for near-Earth satellites. For a satellite making about 14 revolutions per day, the period is about 100 minutes ($1440/14$), and the period of oscillation is almost the same. Thus, there is little distinction between the short-periodic variations from the zonal harmonics and m -daily periods as we examine higher-degree tesseral terms. Typically, these m -daily variations cause effects of about 500 m to 1 to 2 km, with periods of about one day or fractions of a day.

Long-Periodic/M-Daily Beat Periods from Sectorial and Tesseral Harmonics

This component of the motion comes from the terms in Kaula's expanded potential that satisfy $q = 2p - l$. The source is the interaction between the long-periodic and m -daily frequencies. The beat-period envelope has the same qualitative behavior as in the case of LP/SP effects from zonal harmonics, but the short-periodic frequency inside the envelope is replaced by an m -daily oscillation. The magnitude of the effect varies over time from roughly one kilometer to the size of the long-periodic contribution. See Fig. 8-7 for a qualitative diagram.

Long-Periodic/M-Daily/Short-Periodic Linear Combination Terms from Sectorial and Tesseral Harmonics

These terms are the most common (numerous) terms in Kaula's expanded geopotential function. They represent the interaction between all three fundamental periods in the geopotential. The effect is a "hyper-beat-period" phenomenon, which differs from the beat-period phenomenon in the behavior of the beat-period envelope. In this case, the envelope isn't a simple long-periodic oscillation; it's another envelope with a time-varying, periodic amplitude that contains the short-periodic variations. This would be a changing long-periodic envelope in Fig. 8-7. The size of the effect spans the spectrum of short-periodic/long-periodic effects, with corresponding amplitudes.

Resonance Effects from Sectorial and Tesseral Harmonics

In those special cases when the orbital period matches the Earth's rotational period, the satellite's ground-trace will be a closed or repeating path. The repeating ground-track means the satellite will periodically encounter the same configuration of gravitational forces, a condition which leads to resonance. There are near-repeat conditions, and satellites sometimes repeat their ground tracks after multiple days. Many missions are designed to have a near-repeat groundtrack. An example is TOPEX, which repeats its groundtrack every 127 days.

Resonance is the appearance of unusual long-periodic contributions to the motion caused by a satellite's exposure to the Earth in a repeating orientation. The dominant terms are the tesseral harmonic coefficients in the disturbing potential. It occurs because two frequencies in the disturbing acceleration—the Earth's rotation and the orbital period—are nearly commensurate. The frequencies are commensurate if their ratio is the ratio of two integers. Examples include satellites flying repeat-groundtrack orbits, such as geosynchro-

nous satellites, and satellites at lower altitudes such as Landsat (~700 km), RADARSAT (~800 km), and TOPEX (~1300 km). This causes the satellite to "see," or sample a particular aspect of, the Earth's gravity field over and over, causing dynamical effects that build over time. Without this preferential sampling, the effect doesn't arise because other parts of the Earth's gravity field contribute equally, disrupting any long-term buildup in the motion. The result is a variation in the motion with very long periods. Although these periods may sometimes be relatively short (weeks) for *shallow resonance*, they are often months or even years for *deep resonance*. The amplitudes range from meters to tens or hundreds of kilometers. Combinations having shorter periods (amplitudes of about 300 m to 500 m) result from higher-order resonance whenever the satellite's motion is commensurate to higher-order terms in the geopotential.

Geopotential resonance exists in two special cases where the rate of the Kaula gravitation argument is nearly zero, i.e., $\dot{\Theta}_{lmpq} \approx 0$. ***Repeat ground-track resonance*** is contributed by terms which satisfy the condition $q = 0$, or

$$\dot{\Theta}_{lmpq} \approx 0 \Rightarrow (l - 2p)(\dot{\omega} + \dot{M}) \approx m(\dot{\theta}_{GST} - \dot{\Omega}) = m(\omega_{\oplus} - \dot{\Omega})$$

or

$$\frac{\dot{\omega} + \dot{M}}{\omega_{\oplus} - \dot{\Omega}} \approx \frac{P}{Q}$$

where P, Q are integers, and $Q \neq 0$. Those values of l, m, p which satisfy the condition

$$\frac{P}{Q} = \frac{m}{l - 2p}$$

or $m = jP, Q = j(l - 2p)$, $j = 1, 2, 3, \dots$ identify those terms in the Kaula potential with argument rates that nearly vanish. Such arguments are called critical arguments and are the source of resonance oscillations with very long periods. ***Anomalistic resonance*** is contributed by terms which satisfy $l - 2p = 0, q \neq 0$. It's caused only by tesseral harmonics of even degree because of the constraint $l - 2p = 0$ for cases where the mean motion is commensurate with the Earth's rotation rate.

$$\dot{\Theta}_{lmpq} \approx 0 \Rightarrow q\dot{M} \approx m(\omega_{\oplus} - \dot{\Omega})$$

or

$$\frac{\dot{M}}{\omega_{\oplus} - \dot{\Omega}} \approx \frac{P}{Q}$$

where

$$\frac{P}{Q} = \frac{m}{q}$$

or $m = jP, Q = jq$, $j = 1, 2, 3, \dots$

Note that this type of resonance is not very significant for nearly circular orbits unless q is quite small—i.e., $q = \pm 1, \pm 2$ —because the coefficients of these terms contain eccentricity functions $G_{lm\bar{p}q} \sim Oe^{|q|}$ on the order of $e^{|q|}$ and thus nearly vanish. This causes what otherwise might be very large amplitudes to be no larger than those associated with the commonplace terms.

The **resonance parameter**, ϵ_Θ , is a measure of the “strength” of the resonance. For resonance of repeat ground tracks,

$$\epsilon_\Theta = Q(\dot{\omega} + \dot{M}) - P(\omega_\oplus - \dot{\Omega}) \quad (8-48)$$

Exact resonance ($\epsilon_\Theta = 0$) corresponds to the condition $\dot{\Theta} = 0$, which represents an ideal case. In practical applications, ϵ_Θ isn’t zero, but it is small. If $|\epsilon_\Theta|$ is very small, the resonance is characterized as **deep resonance**, which causes “strong” effects, or larger-amplitude oscillations with very long periods. For larger values of $|\epsilon_\Theta|$, the resonance is characterized as **shallow resonance** with variations that have smaller amplitudes and shorter periods.

The fundamental resonance period is given by

$$P_{res} = \frac{2\pi}{|\epsilon_\Theta|}$$

This resonance causes periodic variations in the motion of this period and its harmonics with periods P_{res}/k $k = 2, 3, \dots$. The contributions of the higher harmonics usually decrease as k increases because shorter periodic contributions have smaller amplitudes, all else being equal.

The ratio of the angular rates can always be put in the ratio of two integers because the rates have a finite decimal representation and the ratio of two such quantities can always be expressed as a rational number. However, this usually results in the ratio of two large integers, which doesn’t identify a resonance contribution of real significance. A harmonic term must have a significant amplitude to contribute. Resonance increases the amplitude of the harmonic by producing a small divisor (ϵ_Θ) in the coefficient of the trigonometric term. Remember, for resonance, $\dot{\Theta} \approx \epsilon_\Theta$ in Eq. (8-48). If we integrate the rate $\cos(\epsilon_\Theta t)$ we get

$$\frac{\sin(\epsilon_\Theta t)}{\epsilon_\Theta} + \text{const.}$$

But large values of P and Q will be associated with large values of l and m —the degree and order of the tesseral-harmonic term. These higher-order harmonics for degree and order have smaller gravitational coefficients, $C_{lm}S_{lm}$, making the contributions less significant. In fact, P and Q may be so large that they determine unrealistically large values of l and m or, at least, values for which the coefficients $C_{lm}S_{lm}$ aren’t determined. This doesn’t

mean resonance isn't a real issue. There may be smaller values of P and Q that produce a ϵ_{Θ} small enough to significantly influence the motion. Consider the example of a satellite that makes roughly 14.5 revs/day. Possible values for P_i/Q_i include 14/1, 15/1, and 29/2. Of course, the third pair is exact but again it depends on the smaller high-order terms in the geopotential. The rigorous approach is to compute the coefficient of each term, including the small divisor, ϵ_{Θ} , and compare these coefficients. But this isn't always convenient. An alternative is to pick out candidate pairs by evaluating the resonance parameter ϵ_{Θ} for each pair P_i, Q_i and choose the pairs with smaller values for ϵ_{Θ} . (There may be more than one pair you wish to include.)

There is no concrete rule to determine the minimum size of ϵ_{Θ} . The best approach is to do numerical simulations in order to evaluate the various contributors. These often lead to rules of thumb. For example, if you wish to track shallow-resonance contributions, include all potentially resonant terms with periods that are one or two orders of magnitude greater than that of the short-periodic value (orbital period). Similarly, for deep resonance, select any terms with periods equal to or greater than the long-periodic interval—possibly one or two orders of magnitude greater, depending on your application. For deep resonance, however, the rates are small, and it will take time to observe the effect. These selections are arbitrary and should depend on the accuracy and propagation interval of your application.

Next, let's consider how we determine the integers P and Q . First, more than one pair, $P_i Q_i$, may satisfy the conditions. This assumes we've reduced P_i and Q_i to the lowest form because multiples of P and Q represent higher harmonics of the fundamental period, i.e., $kP/kQ = P/Q$. The process involves expanding the ratio in a continued fraction

$$\frac{\dot{\omega} + \dot{M}}{\omega_{\oplus} - \dot{\Omega}} = x = \alpha_o + \cfrac{1}{\alpha_1 + \cfrac{1}{\alpha_2 + \cfrac{1}{\alpha_3 + \dots}}}$$

where the α_i are integers, $\alpha_i \neq 0$. We determine the integers α_i as follows:

$$x = \alpha_o + \sigma_o \quad \frac{1}{\sigma_o} = \alpha_1 + \sigma_1 \quad \frac{1}{\sigma_1} = \alpha_2 + \sigma_2$$

and so forth. Thus, we start by separating the ratio of the rates into two parts: the integer and the decimal fraction. The integer part is the first integer, α_o , in the continued-fraction expansion. The inverse of the decimal fraction is then separated into an integer and a decimal fraction. This integer becomes the second integer, α_1 , in our expansion. We repeat this process until the expansion is as detailed as we want it to be.

Now, we define a sequence of partial convergents of the continued fraction which turn out to be rational expressions. They give us the desired integer pairs, P_i, Q_i :

$$\frac{P_o}{Q_o} = \alpha_o = \frac{\alpha_o}{1}$$

$$\frac{P_1}{Q_1} = \alpha_o + \frac{1}{\alpha_1} = \frac{1 + \alpha_o \alpha_1}{\alpha_1}$$

$$\frac{P_2}{Q_2} = \alpha_o + \frac{1}{\alpha_1 + \frac{1}{\alpha_2}} = \frac{\alpha_o (1 + \alpha_1 \alpha_2) + \alpha_2}{1 + \alpha_1 \alpha_2}$$

$$\frac{P_3}{Q_3} = \frac{\alpha_o [\alpha_1 (1 + \alpha_2 \alpha_3) + \alpha_3] + 1 + \alpha_2 + \alpha_3}{\alpha_1 (1 + \alpha_2 \alpha_3) + \alpha_3}$$

We can compute $P_i Q_i$ recursively by

$$P_n = \alpha_n P_{n-1} + P_{n-2}$$

$$Q_n = \alpha_n Q_{n-1} + Q_{n-2}$$

which can be deduced from the examples above and proven by mathematical induction.

We can also understand the resonance phenomenon through physical reasoning by examining the simplest ideal case: the 24-hour geosynchronous orbit. Consider a satellite in a perfectly geostationary orbit in which the period is exactly one sidereal day. Because its motion synchronizes with the Earth's rotation, it will always be above the same point on the Earth's surface and will appear to stand still.

Now, let's recognize that the Earth's gravity field depends on longitude and consider only the dominant C_{22} -term corresponding to equatorial ellipticity. Recall the zonal harmonics in Fig. 7-5, but in a rotating reference frame with the Earth (Fig. 8-8). Symmetry clearly shows that when the satellite is on the extension of either axis of the equatorial ellipse (at positions S or U), the gravitational force is purely radial. These must then be equilibrium positions or stationary points in the rotating frame. On the other hand, when the satellite is off-axis, there will be a net tangential force, F , toward the nearest major axis (as indicated in Fig. 8-8).

On first examination, we might expect the satellite to accelerate in the direction of F . However, the drag paradox (see Sec. 8.6.2) dictates just the opposite, and the satellite will accelerate toward the nearest equilibrium position on the minor axis. Because it acquires momentum, the satellite will actually drift past S , and the direction of F will then be reversed. The drift will gradually be reversed. Hence, points S on the minor axis are positions of stable equilibrium ($\lambda \approx 75.3^\circ, 255.3^\circ$), whereas U are unstable positions ($\lambda \approx -14.7^\circ, 165.3^\circ$).

This analysis is important because many companies want satellites in geosynchronous orbit to achieve mission operations, but only two positions ($\lambda \approx 75.3^\circ, 255.3^\circ$) are naturally stable—and that's true only for the perturbation under investigation here! Thus, many geosynchronous satellites must exist away from the stable locations, and because they wander in longitude, they require continual stationkeeping (see Sec. 10.2.1) and much more onboard fuel to correct the longitudinal drift. The complexity and difficulty in launching satellites to geosynchronous orbit don't permit numerous launches and ease of service.

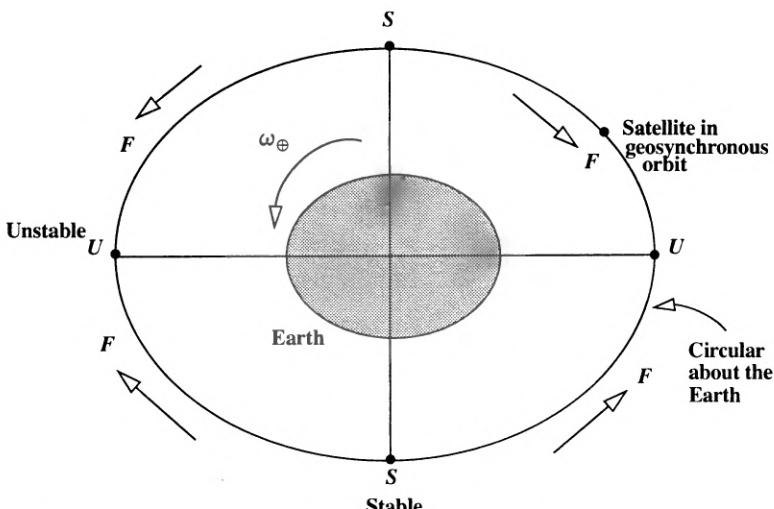


Figure 8-8. Polar View of an Equatorial Section of the Earth. (C_{22} only) F is the net tangential force on the satellite at the positions shown. C_{22} models a longitudinal asymmetry of the Earth. Both stable (S) and unstable (U) positions are identified.

We can also examine the paths of *nearly* circular, nearly equatorial orbits in the rotating coordinate frame. Their motion is essentially as shown in Fig. 8-9. Note the natural separation into three regions. In Region I, the orbital periods are less than 24 hours, and the satellite's apparent motion is to circle eastward around the Earth, that is, *advancing* monotonically eastward in longitude. Farther from the Earth, the period increases, and in the elongated-libration Region II, we have 24-hour orbits. Here the motion consists of a long-periodic oscillation, or circulation, about a stable equilibrium point. Beyond, in Region III, the periods become longer than 24 hours, and the apparent motion is a westward circulation about the Earth or a monotonic *regression* in east longitude. At extreme distances the motion will resemble that of the Moon.

Summary of Central-Body Effects

1. Secular perturbations in Ω , ω , and M are induced only by *even* zonal harmonics. The semimajor axis a , e , and i have no secular effects.
2. The main effects on the orbit are the secular motions of the node and perigee. Eq. (8-39) gives insight to the motion of perigee, which varies as $(4 - 5\sin^2(i))$. Hence, the line of apsides will advance or regress according to whether the inclination is less than or greater than i_ω , where $i_\omega = \sin^{-1}(2/\sqrt{5}) = 63.43^\circ$ or 116.6° . At these critical inclinations, perigee won't vary secularly due to J_2 perturbations.

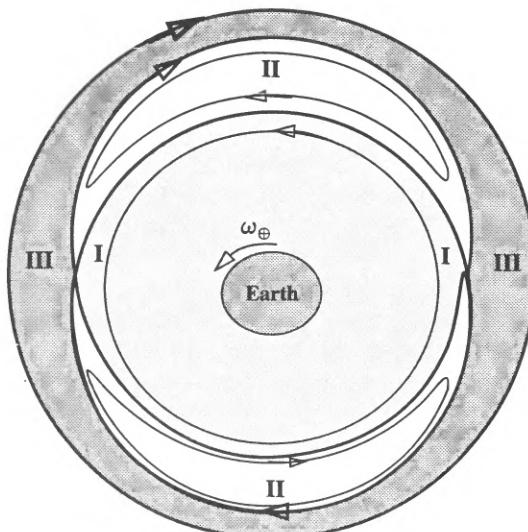


Figure 8-9. Motion of a Geosynchronous Satellite. Shows paths of nearly circular, small-inclination orbits in a coordinate frame rotating with Earth, as seen from the North Pole. Arrows show the satellite's direction of motion. Region I: orbital periods are less than 24 hr; Region II: periods are 24 hours; Region III: periods are greater than 24 hours.

3. For the polar orbit, $i = 90^\circ$, all secular and periodic terms in Ω and i vanish, indicating that the orbital plane remains fixed.
4. Periodic variations occur in all elements ($a, e, i, \Omega, \omega, M$) and are induced by *all* harmonics.
5. The presence of the divisor $[4 - 5\sin^2(i)]$ in the periodic terms [see Eq. (8-39)] means that the above solutions for e, i, Ω , and ω don't apply near the critical inclination, i_ω . Thus, we have to use numerical or special analytical techniques to accurately solve for the motion in this region.
6. In each case, the magnitude of the central-body perturbation increases with the orbit's increasing eccentricity and decreases with the orbit's increasing size, a .

8.6.2 Drag Analysis

Drag presents a challenge to accurate modeling, because the dynamics of the upper atmosphere aren't completely understood, in part due to the limited knowledge of the interaction of the solar wind and the Earth's magnetic field. In addition, drag models contain many parameters that are difficult to estimate with reasonable accuracy. Included in this list are the drag coefficient, atmospheric density, relative-velocity vector, and area of the satellite exposed to the instantaneous-velocity vector.

Equations for the Effects of Drag

To determine expressions for the effects of drag in terms of the orbital elements, you can approximate relations for the rate of change of the mean motion. If you assume perigee height remains relatively constant during satellite motion, you can begin with the equation for the perigee height ($r_p = a(1 - e)$). Taking the derivative produces

$$\frac{dr_p}{dt} = \frac{da}{dt}(1 - e) - a \frac{de}{dt} \approx 0$$

Solving for de/dt results in

$$\frac{de}{dt} = \frac{(1 - e)}{a} \frac{da}{dt}$$

To determine (da/dt) , consider the relation between the semimajor axis and the mean motion and take the derivative. Thus,

$$n^2 a^3 = \mu \quad 2n \frac{dn}{dt} a^3 + 3n^2 a^2 \frac{da}{dt} = 0$$

Solve for the semimajor axis to get

$$\dot{a} = -\frac{2a_o}{3n_o} \dot{n}$$

As a first approximation, assume $a = a_o$, $n = n_o$ and integrate the expressions using those initial values. As long as the changes in a and n aren't too large, the new semimajor axis is

$$a = a_o - \frac{2a_o}{3n_o} \dot{n} \Delta t \quad (8-49)$$

You can combine this expression with the rate of change for eccentricity to yield an approximate relation:

$$e = e_o - \frac{2(1 - e_o) \dot{n}_o}{3n_o} \Delta t \quad (8-50)$$

As mentioned, for sophisticated analyses, we need more accurate atmospheric density. Because the complete derivation for all classical orbital elements is very lengthy, I'll show only effects in the semimajor axis to demonstrate techniques and some of the required assumptions. Because drag is a nonconservative force, let's start with the Gaussian form of the Lagrange VOP equations [Eq. (8-24)]:

$$\frac{da}{dt} = \frac{2e \sin(\nu)}{n\sqrt{1-e^2}} F_R + \frac{2p}{nr\sqrt{1-e^2}} F_S$$

Drag opposes the velocity vector. Thus, we need to find the drag components in the *NTW* coordinate system because the *T*-axis is always aligned with the velocity vector. Unfortunately, the VOP equations are in the *RSW* system [Eq. (8-24)], requiring a transformation through the flight-path angle. Roy (1988, 325) shows the transformation using the relations for the flight-path angle from Eq. (2-58) and Eq. (2-59):

$$\begin{aligned}\frac{da}{dt} &= \frac{2e \sin(\nu)}{n\sqrt{1-e^2}} \left(-\frac{1}{2}\rho \frac{c_D A}{m} v_{rel}^2 \right) \sin(\phi_{fpa}) \\ &\quad + \frac{2p}{nr\sqrt{1-e^2}} \left(-\frac{1}{2}\rho \frac{c_D A}{m} v_{rel}^2 \right) \cos(\phi_{fpa})\end{aligned}$$

Substituting Eq. (2-58) and Eq. (2-59) to remove the flight-path angle leads to

$$\begin{aligned}\frac{da}{dt} &= -\rho \frac{c_D A}{m} v_{rel}^2 \left(\frac{1}{n\sqrt{1-e^2}} \right) \left[\frac{(e \sin(\nu))^2}{\sqrt{1+e^2+2e \cos(\nu)}} + \frac{p(1+e \cos(\nu))}{r\sqrt{1+e^2+2e \cos(\nu)}} \right] \\ &= -\rho \frac{c_D A}{m} v_{rel}^2 \left(\frac{\sqrt{1+e^2+2e \cos(\nu)}}{n\sqrt{1-e^2}} \right) \left[\frac{(e \sin(\nu))^2 + (1+e \cos(\nu))^2}{1+e^2+2e \cos(\nu)} \right]\end{aligned}$$

Simplifying yields

$$\frac{da}{dt} = -\rho \frac{c_D A}{m} v_{rel}^2 \left(\frac{\sqrt{1+e^2+2e \cos(\nu)}}{n\sqrt{1-e^2}} \right)$$

Notice the solution now depends on the true anomaly, but so does v_{rel} . To integrate, we must change the independent variable from time to true anomaly. Several intermediate relations are necessary, including Eq. (8-16) and the trajectory equation, Eq. (2-14):

$$\frac{dr}{dt} = \dot{r} = \frac{dr d\nu}{d\nu dt} = \frac{pe \sin(\nu)}{(1+e \cos(\nu))^2 r^2} \frac{h}{p} = \frac{h}{p} e \sin(\nu)$$

Now, let's find the relative velocity using the Pythagorean theorem:

$$\begin{aligned}v_{rel}^2 &= \dot{r}^2 + r^2 \dot{\nu}^2 = \frac{h^2}{p^2} e^2 \sin^2(\nu) + \frac{h^2}{p^2} (1+e \cos(\nu))^2 \\ &= \frac{n^2 a^2}{(1-e^2)} \left\{ e^2 \sin^2(\nu) + 1 + e^2 \cos^2(\nu) + 2e \cos(\nu) \right\} \\ &= \frac{n^2 a^2}{(1-e^2)} \{ 1 + e^2 + 2e \cos(\nu) \}\end{aligned}$$

Next, we'll need an expression for the time rate of change of the true anomaly, using Eq. (8-17):

$$\frac{dt}{d\nu} = \frac{r^2}{na^2\sqrt{1-e^2}} = \frac{a^2(1-e^2)^2}{na^2(1+e\cos(\nu))^2\sqrt{1-e^2}}$$

$$\frac{dt}{d\nu} = \frac{(1-e^2)^{3/2}}{n(1+e\cos(\nu))^2}$$

We can substitute to find a new expression for the rate of change of a :

$$\frac{da}{d\nu} = \frac{da}{dt} \frac{dt}{d\nu} = -\rho \frac{c_D A}{m} (1+e^2 + 2e\cos(\nu)) \frac{\sqrt{1+e^2 + 2e\cos(\nu)}}{(1+e\cos(\nu))^2} a^2$$

$$\frac{da}{d\nu} = -\rho \frac{c_D A a^2 (1+e^2 + 2e\cos(\nu))^{3/2}}{m (1+e\cos(\nu))^2}$$

Roy (1988, 325–327) shows equations for e , and i . He also shows that $da/d\nu$ and $de/d\nu$ result in secular effects. The presence of $\cos(\nu)$ means these are periodic variations. Again, let's change the variable from true anomaly to eccentric anomaly, E . Using Eq. (4-10) gives us

$$\frac{da}{dE} = \frac{da}{d\nu} \frac{d\nu}{dE} = -\rho \frac{c_D A a^2 (1+e\cos(E))^{3/2}}{m \sqrt{1-e\cos(E)}}$$

To find the secular rate of change, we must integrate over one period (360° of E):

$$\Delta a = -\frac{c_D A}{m} a^2 \int_0^{2\pi} \frac{\rho (1+e\cos(E))^{3/2}}{\sqrt{1-e\cos(E)}} dE$$

Many other formulations are possible. King-Hele (1987, 47) suggests expanding the result in a power series of $e\cos(E)$. He recognizes the validity of this expansion for $e < 0.2$. For larger values, see King-Hele (1987, 63). This process results in modified Bessel functions of the first kind of order j , J_j :

$$J_j(c) = \frac{1}{2\pi} \int_0^{2\pi} \cos(j\theta) \exp^{c\cos(\theta)} d\theta$$

Blitzer gives the change over one revolution in the elements due only to drag (don't confuse the Bessel functions with the zonal harmonics). Remember expressions are based on an exponential model of atmospheric density. Thus,

$$\Delta a_{rev} = -2\pi\delta a^2 \rho_p \left\{ J_o + 2eJ_1 + \frac{3e^2}{4}(J_0 + J_2) + \frac{e^3}{4}(3J_1 + J_3) + O(e^4) \right\} \exp^{-c}$$

$$\Delta e_{rev} = -2\pi \delta a \rho_p \left\{ J_1 + \frac{e}{2} (J_0 + J_2) - \frac{e^2}{8} (5J_1 + J_3) - \frac{e^3}{16} (5J_0 + 4J_2 - J_4) + O(e^4) \right\} \text{EXP}^{-c}$$

$$\Delta i_{rev} = -\frac{\pi a \omega_{\oplus} \delta \rho_p}{2n \sqrt{Q}} \sin(i) \left\{ (J_0 - 2eJ_1 + (J_2 - 2eJ_1) \cos(2\omega) + O(e^2)) \right\} \text{EXP}^{-c}$$

$$\Delta \Omega_{rev} = -\frac{\pi a \omega_{\oplus} \delta \rho_p}{2n \sqrt{Q}} \left\{ J_2 - 2eJ_1 + O(e^2) \right\} \sin(2\omega) \text{EXP}^{-c}$$

$$\Delta \omega_{rev} = -\Delta \Omega_{rev} \cos(i)$$

where we use some new variables

ρ_p = density at perigee

r_a = radius at apogee, $a(1+e)$

r_p = radius at perigee, $a(1-e)$

$$c = \frac{r_a - r_p}{H} = \frac{ae}{H}$$

$$\delta = \frac{Q A c_D}{m}$$

$$Q = \left(1 - \frac{r_p \omega_{\oplus}}{v_p} \cos(i) \right)^2 = 1 - \frac{2\omega_{\oplus} (1-e)^{3/2}}{n \sqrt{1+e}} \cos(i)$$

= factor which includes the rotation of the atmosphere ($0.9 \leq Q \leq 1.1$)

m = mass of satellite

J_s = modified Bessel function of the first kind of order s , in all cases having argument c

Effects of Drag

The predominant effect of drag is to shrink the orbit and in many cases, cause the satellite to reenter the atmosphere and hit the Earth. We can examine the effects of atmospheric drag through analogies and mathematical approximations. For satellites that are close to the Earth, the possibility of reentry is important in determining a satellite's lifetime.

Drag is a nonconservative force that acts to lower a satellite's orbit, much as the air acts to slow your progress as you ride a bicycle. On a perfectly calm day, there's still some resistance, and a satellite will always experience some force as it travels through the atmosphere, even if it only encounters the atmosphere near perigee as a Molniya satellite does. If you're squarely facing the wind, you feel much more resistance than if you're bent down. A satellite experiences varying effects from drag depending on its attitude (orientation). A strong wind is analogous to the density variations in the atmosphere (diurnal, solar, and geomagnetic for example). Mountainous terrain equates to the altitude of the satellite—the closer it is to the Earth, the greater the force.

A very common term used with drag effects is the *drag paradox*, in which a satellite speeds up as it gets closer to the Earth. To explain it from a mathematical perspective, let's consider the simple two-body forces acting on a satellite in a circular orbit. It has a period, P ; average kinetic energy, KE_{avg} ; potential energy, PE_{avg} ; mean motion, n ; and velocity, v . Remember, the radius and semimajor axis are the same for a satellite in a circular orbit; the potential energy is defined to be negative; and total energy, ξ , is negative (for a circle or ellipse). Thus,

$$\begin{aligned}\xi &= -\mu/2a \\ P &= 2\pi\sqrt{a^3/\mu} \\ KE_{avg} &= \mu/2a = -\xi \\ PE_{avg} &= -\mu/a = 2\xi \\ n &= \sqrt{\mu/a^3} \\ v &= \sqrt{\mu/a}\end{aligned}$$

The work done by the disturbing force \vec{F} during any small time interval (Δt) will alter the orbital energy by the distance through which the force acts, or

$$\Delta\xi = \vec{F} \cdot \vec{dist} = \vec{F} \cdot \vec{v}\Delta t = -F_T v \Delta t \quad (8-51)$$

where F_T is the tangential component of \vec{F} that usually opposes the velocity vector. We get the resulting effects on the various parameters using differentials:

$$\begin{aligned}\Delta\xi &= \frac{\mu}{2a^2}\Delta a \Rightarrow \Delta a = \frac{2a^2}{\mu}\Delta\xi \\ \Delta P &= 3\pi\sqrt{\frac{a}{\mu}}\Delta a = (3aP/\mu)\Delta\xi \\ \Delta KE_{avg} &= -\Delta\xi \\ \Delta PE_{avg} &= 2\Delta\xi \\ \Delta n &= -(3an/\mu)\Delta\xi \\ \Delta v &= -\sqrt{\mu/a}\Delta\xi\end{aligned} \quad (8-52)$$

These differentials determine the relative effects as the energy changes. In other words, any perturbing force that increases the total energy will enlarge the orbit, increase the period and potential energy, and decrease the average kinetic energy and mean-linear and angular velocities. But any force that drains energy from the satellite will shrink the orbit and increase the average velocity. Table 8-3 summarizes the results.

In the case of drag, the perturbing force directly opposes the motion, so $\Delta\xi < 0$. Accordingly, from Table 8-3, the orbit's size and period will decrease while the mean velocity increases. Because drag is strongest at perigee, where the velocity and atmo-

TABLE 8-3. Effect of Energy Changes Caused by Drag. This table shows the drag paradox: increasing energy enlarges the orbit but reduces the velocity, and vice versa.

Quantity	Positive $\Delta\xi > 0$	Negative $\Delta\xi < 0$
a, ρ, PE_{avg}	increases	decreases
KE_{avg}, n, v	decreases	increases

spheric density are greatest, the energy drain is also greatest at this point. The large energy drain at perigee causes the orbit to become increasingly circular with each revolution, as indicated in Fig. 8-10. Hence, the *drag paradox* is shown as the effect of atmospheric friction, which *increases* the satellite's average motion as it spirals inward.

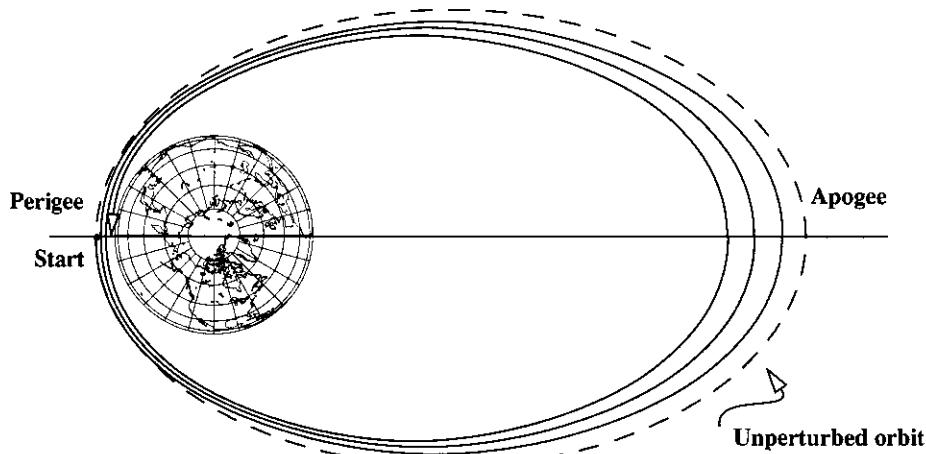


Figure 8-10. Contraction of Orbit Under Drag. This figure is greatly exaggerated but shows the general effect of drag on the satellite's orbit. The radius of perigee tends to remain constant, while apogee shrinks and the eccentricity approaches 0.0.

Obviously, the period can change rapidly for a highly drag-perturbed satellite. The change in orbital period, ρ , may be calculated to first order from the change in energy or semimajor axis in Eq. (8-52) (remembering that $\rho = 2\pi/n$):

$$\Delta\rho = \frac{3\rho}{2a}\Delta a = \frac{3\pi}{na}\Delta a \quad (8-53)$$

Secular changes occur in a, e, i , and apogee height, whereas long-periodic variations occur in i, Ω , and ω . In the case of highly eccentric orbits, a more pronounced effect arises

because of the coupling between third-body and drag perturbations. Long-periodic (third-body-induced) variations in perigee height can shift perigee to new positions, where the density, scale height, and hence lifetime, are very different. In fact, the initial orientation of the Sun, the Moon, and the satellite's orbit can lower perigee height enough to cause the satellite to reenter early, or they can increase the perigee height and vastly extend its lifetime (Cook and Scott, 1967). A program that can compare different model atmospheres and precise planetary-ephemerides is necessary to analyze these effects.

The effect of drag for a low-Earth satellite has important implications for a satellite's lifetime. Figure 8-11 graphs the orbital parameters most sensitive to drag (apogee and perigee heights, period, and eccentricity) during the lifetime of a typical satellite. King-Hele (1987) discusses the theory and applications of satellite lifetime at length for various orbit types and model atmospheres. The subject is complex enough to require an entire book. Our 1980 experience with Skylab reentry shows the last few orbits couldn't be accurately predicted—the remote Australian impact was fortunate, but unpredictable.

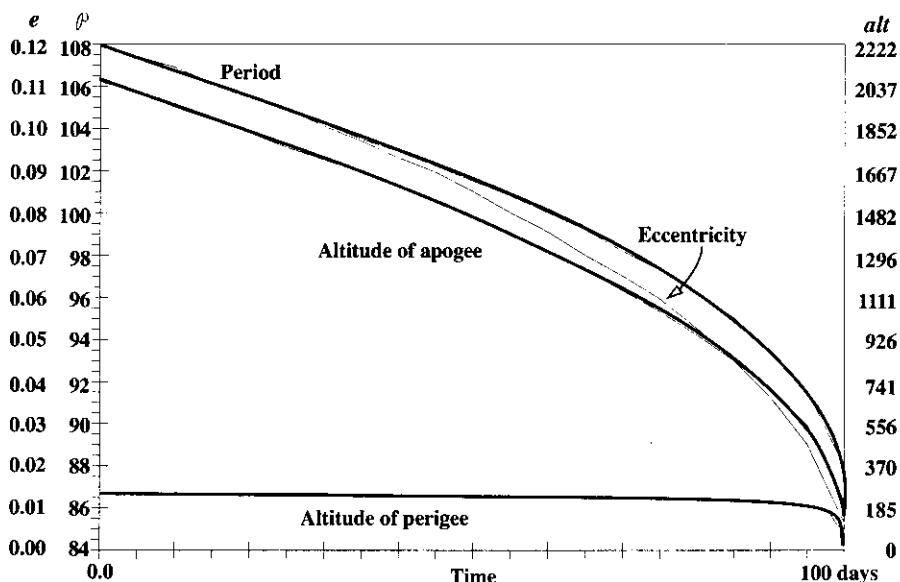


Figure 8-11. Notional Variations of Orbital Properties During a Satellite's Lifetime. In general, *certain* orbital parameters tend toward "zero," and eventually the satellite will reenter and crash. The left-hand axes are eccentricity and period (min) respectively, whereas the right-hand scale is altitude for perigee and apogee (km).

The difficult part of accurately determining drag involves finding the correct atmospheric density and satellite attitude. Unfortunately, determining atmospheric density is a very difficult part of the process. Our discussion of numerous atmospheric models should leave you with this impression. But the problems with determining a satellite's attitude

may not be so apparent. These calculations involve the cross-sectional area, drag coefficient, and satellite mass. The cross-sectional area is sometimes known for operating satellites, but as they near reentry, they may begin to tumble and have rapidly changing areas. The flight-path angle is only a very crude estimate of the orientation. Precise calculations should refer to texts devoted to attitude control, such as Wertz (1978), Kaplan (1976), and Kane, Likens, and Levinson (1983). As mentioned earlier, slight changes in the drag coefficient will greatly change the results of simulations. For this reason, the effective drag coefficient is almost always solved for in highly precise orbit determination (Chap. 9). Solving for c_D permits a refinement to correct errors in the atmospheric model. However, a difficulty arises because ρ , c_D , and A all appear in Eq. (7-24). Therefore, we can estimate a fixed value for the quantity $(\rho c_D)/A$, but this quantity also helps to absorb the errors in these parameters appearing in the propagation model used in the solution. Finally, the satellite's mass is usually constant, but as reentry becomes imminent, it can change. The causes range from ablation to solar panels tearing off.

Summary of Drag Effects

1. Secular changes occur in a , e , i .
2. Periodic changes occur in all orbital elements, especially i , Ω , ω .
3. Perigee height remains relatively constant but can be influenced by coupling with central-body or third-body effects.
4. Calculating atmospheric density is often the most difficult part of assessing and modeling the drag effect. Determining the correct ballistic coefficient (or drag coefficient) and the time-varying cross-sectional area (satellite attitude) are also very difficult.

8.6.3 Third-Body Analysis

Analyzing the third-body perturbations on a satellite's orbit is difficult because the periods are often very long. It's also important to use highly accurate ephemerides to locate the third body.

Equations for the Effects of Third-Body Perturbations

Detailed analyses of the luni-solar perturbations are complex because the third body's geometry changes continually. Here, I summarize results based on Cook's (1962) method, which assumes a circular orbit for the disturbing body and includes only contributions due to the P_2 term in the Legendre polynomial expansion of the potential. A subscripted "3" distinguishes the third body. Smith (1962) extends the theory to allow for $e_3 \neq 0$ and also includes higher-order terms in (a/a_3) . Taff (1985, 347–352) reproduces these results and includes corrected formulas from Geyling and Westerman (1971, 317–321) for the change in the elements due to a known change of true anomaly or eccentric anomaly. I won't

reproduce all the equations, but these expressions show the complexity of analytically modeling third-body perturbations.

We'll first need the direction cosines for the third body, which we find by examining the geometry. Using Fig. 8-12, find the direction cosines, A , B , and C :

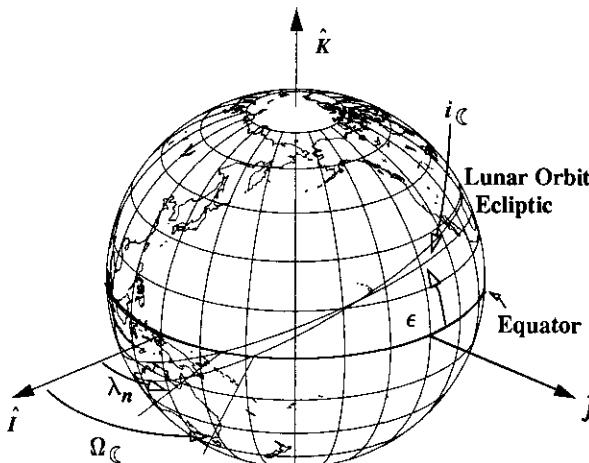


Figure 8-12. Parameters of the Lunar Orbit. This diagram shows the groundtrack of the Moon's orbit about the Earth, along with many of the angles needed to specify its location. Remember that the obliquity, ϵ , is about 23.5° and the Moon is inclined about 5.1° to the ecliptic.

$$A = \cos(\Omega - \Omega_3) \cos(u_3) + \cos(i_3) \sin(u_3) \sin(\Omega - \Omega_3)$$

$$B = \cos(i) [-\sin(\Omega - \Omega_3) \cos(u_3) + \cos(i_3) \sin(u_3) \cos(\Omega - \Omega_3)] \\ + \sin(i) \sin(i_3) \sin(u_3)$$

$$C = \sin(i) [\cos(u_3) \sin(\Omega - \Omega_3) - \cos(i_3) \sin(u_3) \cos(\Omega - \Omega_3)] \\ + \cos(i) \sin(i_3) \sin(u_3)$$

For the Sun, $i_3 = \epsilon \approx 23.5^\circ$, $\Omega_3 = 0.0^\circ$, $u_3 = \lambda_{M\odot}$, and for the Moon, $i_C \approx 5.1^\circ$, $\cos(i_3) = \cos(\epsilon)\cos(i_C) - \sin(\epsilon)\sin(i_C)\cos(\Omega_C)$, and

$$\sin(\Omega_3) = \frac{\sin(i_C) \sin(\Omega_C)}{\sin(i_C)}$$

$$u_3 = \lambda_{M\odot} - \Omega_C + \csc\left(\frac{\sin(\epsilon) \sin(\Omega_C)}{\sin(i_3)}\right)$$

The secular and long-periodic changes in the elements over *one revolution* of the satellite are sometimes given; however, it's simpler to provide the time rate of change for each element. We can always recover the per-revolution values by

$$\Delta c_{rev} = \frac{2\pi \dot{c}}{n} = \rho \dot{c}$$

Remember that Cook's development assumes the third body is fixed in the model, and he didn't provide an expression for the fast variable. This can be a limiting assumption in many scenarios. The secular and long-periodic *rates* of change of the elements are

$$\begin{aligned} \dot{a} &= 0 & \dot{r}_p &= -a\dot{e} \\ \dot{e} &= -\frac{15\mu_3 e \sqrt{1-e^2}}{4r_3^3 n^2} \{ 2AB \cos(2\omega) - (A^2 - B^2) \sin(2\omega) \} \\ \dot{i} &= \frac{3\mu_3 C}{4r_3^3 n^2 \sqrt{1-e^2}} \{ A [2 + 3e^2 + 5e^2 \cos(2\omega)] + 5Be^2 \sin(2\omega) \} \\ \dot{\Omega} &= \frac{3\mu_3 C}{4r_3^3 n^2 \sqrt{1-e^2} \sin(i)} \{ 5Ae^2 \sin(2\omega) + B [2 + 3e^2 - 5e^2 \cos(2\omega)] \} \\ \dot{\omega} &= -\dot{\Omega} \cos(i) + \frac{3\mu_3 \sqrt{1-e^2}}{2r_3^3 n^2} \left\{ 5ABS \sin(2\omega) + \frac{5}{2}(A^2 - B^2) \cos(2\omega) \right. \\ &\quad \left. - 1 + \frac{3(A^2 + B^2)}{2} \right\} + \frac{15\mu_3 a (A \cos(\omega) + B \sin(\omega))}{4r_3^3 n^2 e r_3} \left[1 - \frac{5}{4}(A^2 + B^2) \right] \end{aligned} \quad (8-54)$$

For moderately small eccentricities, second-order terms become important for the argument of perigee; that's when the last term is added.

The only secular changes will be in the node, the perigee, and the mean anomaly at epoch. Extracting these terms from the preceding equations, we obtain (for a third body in a circular orbit)

$$\begin{aligned} \dot{\Omega}_{sec} &= -\frac{3\mu_3 (2 + 3e^2) [2 - 3\sin^2(i_3)]}{16r_3^3 n \sqrt{1-e^2}} \cos(i) \\ \dot{\omega}_{sec} &= \frac{3\mu_3 [2 - 3\sin^2(i_3)]}{16r_3^3 n \sqrt{1-e^2}} \left\{ e^2 + 4 - 5\sin^2(i) \right\} \end{aligned} \quad (8-55)$$

Smith's (1962) equations, which include terms in e_3 , are

$$\begin{aligned}\dot{\Omega}_{sec} &= -\frac{3\mu_3(1-e^2)^3(1+5e^2)[2-3\sin^2(i_3)]}{8r_3^3n(1-e_3^2)^{3/2}} \cos(i) \\ \dot{\omega}_{sec} &= \frac{3\mu_3(1-e^2)^3[2-3\sin^2(i_3)]}{16r_3^3n(1-e_3^2)^{3/2}} \left\{ 4-5\sin^2(i)+5e^2\left(3-\frac{7}{2}\sin^2(i)\right) \right\}\end{aligned}\quad (8-56)$$

Effects of Third-Body Perturbations

To understand the long-term behavior of a satellite's orbit under the influence of the Sun or any other third body, first consider a qualitative example. Imagine both the satellite and the Sun smeared out into elliptical rings along their respective orbits (Fig. 8-13). The mutual gravitational attraction of the rings will create a torque about the mutual line of nodes (i.e., line of intersection), tending to turn the satellite ring into the ecliptic. The resulting gyroscopic effect of the torque on the spinning satellite ring will make the orbit precess about the pole of the ecliptic—specifically, the nodes will regress along the ecliptic. The amplitude of this precession varies depending on inclination, and often, it's so small that the effect is obscured by larger perturbations. Similarly, the Moon will cause the orbit to regress about an axis normal to the Moon's orbital plane. Because the Moon is inclined about 5° to the ecliptic, the precession will be slower. Both these effects are similar to the central-body effects discussed earlier. Moreover, because the force is conservative, we find that the third-body perturbations follow the characteristics of the central body, having similar secular and periodic effects.

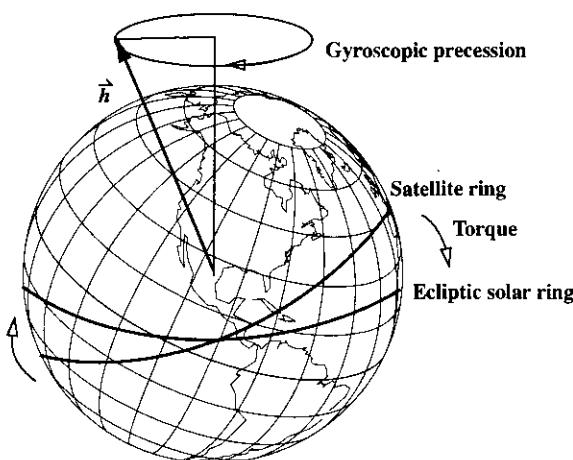


Figure 8-13. Third-Body Interactions. Imagine that the entire mass of a third body (the Sun, for instance) occupies a band about the planet. The resulting torque causes the satellite's orbit to precess like a gyroscope.

We've already seen that the Earth's oblateness causes the satellite's orbital plane to regress about the Earth's polar axis. For satellites *near* Earth, the oblateness effect's dominance means that the orbit will precess essentially about the polar axis. In particular, whereas oblateness makes the orbital plane regress about the Earth's polar axis, solar attraction causes a regression of the orbital plane about the ecliptic pole, and the Moon makes it regress about an axis normal to the Moon's orbital plane. The net effect is regression about a mean pole somewhere between the Earth's pole and the ecliptic pole. Oscillations occur in the nodal regression and in the inclination with very long periods because of the 18.6-year rotation of the Moon's orbital plane (i.e., node of the Moon along the ecliptic) about the ecliptic.

An interesting comparison results from examining the relative effects of the Moon and the Sun. Dividing the secular rates (nodal or periapsis) in Eq. (8-56) for the Moon by those for the Sun gives us

$$\frac{\dot{\Omega}_{\mathbb{C}}}{\dot{\Omega}_{\odot}} \approx \frac{\dot{\omega}_{\mathbb{C}}}{\dot{\omega}_{\odot}} \approx \frac{r_{\odot}^3 \mu_{\mathbb{C}} (2 - 3 \sin^2(i_{\mathbb{C}})) (1 - e_{\odot}^2)^{3/2}}{r_{\mathbb{C}}^3 \mu_{\odot} (2 - 3 \sin^2(i_{\odot})) (1 - e_{\mathbb{C}}^2)^{3/2}}$$

Using approximate values for the Sun and the Moon ($i_{\mathbb{C}} = 5.1^\circ$, $i_{\odot} = 0.0$, $e_{\mathbb{C}} = 0.05$, $e_{\odot} = 0.0$, $r_{\mathbb{C}} = 384,400$ km, $r_{\odot} = 149,598,023$ km, $\mu_{\mathbb{C}} = 4.9 \times 10^3$ km³/s², and $\mu_{\odot} = 1.327 \times 10^{11}$ km³/s²), we see that

$$\frac{\dot{\Omega}_{\mathbb{C}}}{\dot{\Omega}_{\odot}} \approx \frac{\dot{\omega}_{\mathbb{C}}}{\dot{\omega}_{\odot}} \approx 2.2$$

The ratio of the secular rates of the Moon's effects to those of the Sun is about 2.2.

This ratio is the same as the tidal force. The tidal force is caused by the Earth's rotation and the combined perturbative effects of the Sun and Moon. As the Earth rotates each day, it must "pull" the tidal bulge caused by the Moon pulling on the oceans. The net effect is energy dissipation through mechanical processes in the Earth's crust and the oceans. Thus, the Earth-Moon system loses some energy in the process. Over several thousand years, the Moon slowly recedes from the Earth, and its period increases to maintain a constant momentum. Ultimately, the tidal pull decreases. Perhaps lesser known is that the Earth's pulling on the Moon also influences the Moon's rotation rate. Therefore, the Moon's rotation is nearly the same as its orbital period, and thus we see only about 59% of one side of the Moon. This effect is known as *libration*.

Periodic variations occur in all elements caused by perigee precession and the motion of the disturbing body. To calculate these terms requires a reasonably accurate third-body ephemeris, i.e., knowledge of the third body's position as a function of time.

An important operational result of third-body perturbations involves the stationkeeping activities for a geosynchronous satellite. Allan and Cook (1964) have shown that inclinations for geosynchronous satellites will oscillate as $i \pm 15^\circ$ over many years. You can also consult Chobotov (1991, 252–253) for additional information.

For satellite orbits that dip into the Earth's atmosphere, a very important coupling occurs between the luni-solar and drag perturbations. The variation in eccentricity, Eq. (8-54), will shift the perigee height sinusoidally, causing the density to change. This, in turn, can strongly influence the satellite's lifetime. In fact, some launches are designed to cause a quicker decay of the launch vehicle's upper-stage orbit. In the case of the lunar perturbations, assuming the orbital elements don't change much during one lunar month, Moe (1960) calculates the mean rate of change of perigee height during this period to be

$$\Delta r_p = \frac{15\mu_{\mathbb{C}}ae\sqrt{1-e^2}}{8nr_{\mathbb{C}}^3} \sin(2\omega_{\mathbb{C}}) \sin(i_{\mathbb{C}}) \quad (8-57)$$

where $i_{\mathbb{C}}$ is the orbit's inclination to the Moon's orbital plane (\pm about 5.1° of the ecliptic), and $\omega_{\mathbb{C}}$ is the argument of the perigee from the ascending node on the ecliptic.

To calculate the solar perturbation, we can't invoke a similar averaging procedure because the elements are no longer assumed to be constant. Similar to the lunar effect, the Sun also causes sinusoidal variations in perigee height. When the combined lunar-solar changes of perigee are in the same direction, they have an important effect on a satellite's lifetime. For accurate results, evaluate Eq. (8-54) for each revolution, allowing for the secular changes in Ω and ω caused by the Earth's oblateness. Gedeon (1968) provides a series of curves with examples to help calculate variations in perigee height.

As the orbit's size increases, the oblateness effects decrease, whereas the third-body perturbations become more important. And at distances greater than geosynchronous altitude, the third-body effects become more dominant.

The full third-body disturbing function has already been presented and takes the same general form for the Sun, Moon, or other planetary body (assuming we use the correct coefficients and units). Using km, AU, or ER can be especially troubling for this process if we include more than one third-body perturbation. Thus, I recommend some kind of canonical units for standardization. Adding all of the contributing third-body perturbations gives us the total result.

Summary of Third-Body Effects

1. There are no secular, long-periodic, or m -monthly variations in the semi-major axis.
2. The only secular perturbations are in the node and perigee. For many orbits, the lunar effect is larger because it's closer to the Earth. The ratio of the lunar and solar effects is the same as the ratio of the lunar and solar tide-raising forces (~2.2). This is because the ratio is the same as the tidal-generating function, which derives from the third-body potential.
3. The long-periodic variations in e , i , Ω , and ω are completely associated with the motions of the satellite's perigee and the disturbing body.

4. For near-Earth orbits, the dominance of oblateness dictates that the orbital plane regresses very nearly about the polar axis. For higher orbits, however, the regression will be about some mean pole lying between the Earth's pole and the ecliptic pole.
5. Cook (1962) also points to the possible existence of 15 cases of resonance when the rate of change of any of the arguments in the periodic terms in Eq. (8-54) vanishes. Because resonance contributes large perturbations over a long period, such orbits can have reduced mission lifetimes.
6. For drag-perturbed orbits, the variations in perigee height caused by the Sun and Moon can greatly increase or decrease the satellite's lifetime.

8.6.4 Solar-Radiation Analysis

As mentioned earlier, solar-radiation effects depend on various parameters. It's very important to correctly determine reflectivity, c_R , the attitude with respect to the Sun, and the solar flux arriving at the satellite's position (including eclipse periods). The overall solar activity plays a large part in determining the last effect.

Equations for the Effects of Solar-Radiation Pressure

A lot of literature describes the effects of solar-radiation pressure on orbits, including theoretical studies and analyses of observational data. Blitzer extracts the following equations from the work by Cook (1962). He assumes that the disturbing acceleration, F_{SR} , is constant and given by Eq. (7-38) while the satellite is in sunlight and that it is zero in the Earth's shadow.

Cook expresses the perturbations in terms of the radial, transverse, and normal (R, S, W) components of the disturbing force, evaluated at perigee as if perigee is in sunlight. Because the normal component, W , is constant over the orbit, there is no subscript.

$$R_p = F_{SR} \left\{ \left[\cos^2\left(\frac{\epsilon}{2}\right) \cos(\omega + \Omega - \lambda_\odot) + \sin^2\left(\frac{\epsilon}{2}\right) \cos(\omega + \Omega + \lambda_\odot) \right] \cos^2\left(\frac{i}{2}\right) \right. \\ \left. + \left[\cos^2\left(\frac{\epsilon}{2}\right) \cos(\omega - \Omega + \lambda_\odot) + \sin^2\left(\frac{\epsilon}{2}\right) \cos(\omega - \Omega - \lambda_\odot) \right] \sin^2\left(\frac{i}{2}\right) \right. \\ \left. + \frac{1}{2} [\cos(\omega - \lambda_\odot) - \cos(\omega + \lambda_\odot)] \sin(i) \sin(\epsilon) \right\}$$

$$S_p = -F_{SR} \left\{ \left[\cos^2\left(\frac{\epsilon}{2}\right) \sin(\omega + \Omega - \lambda_\odot) + \sin^2\left(\frac{\epsilon}{2}\right) \sin(\omega + \Omega + \lambda_\odot) \right] \cos^2\left(\frac{i}{2}\right) \right. \\ \left. + \left[\cos^2\left(\frac{\epsilon}{2}\right) \sin(\omega - \Omega + \lambda_\odot) + \sin^2\left(\frac{\epsilon}{2}\right) \sin(\omega - \Omega - \lambda_\odot) \right] \sin^2\left(\frac{i}{2}\right) \right. \\ \left. - \frac{1}{2} [\sin(\omega + \lambda_\odot) - \sin(\omega - \lambda_\odot)] \sin(i) \sin(\epsilon) \right\}$$

$$W \sin(\omega) = -\frac{F_{SR}}{2} \left\{ [\cos(\omega + \Omega - \lambda_{\odot}) - \cos(\omega - \Omega - \lambda_{\odot})] \sin(i) \cos^2\left(\frac{\epsilon}{2}\right) \right. \\ \left. + [\cos(\omega + \Omega + \lambda_{\odot}) - \cos(\omega - \Omega - \lambda_{\odot})] \sin(i) \sin^2\left(\frac{\epsilon}{2}\right) \right. \\ \left. + [\cos(\omega + \lambda_{\odot}) - \cos(\omega - \lambda_{\odot})] \cos(i) \sin(\epsilon) \right\}$$

$$W \cos(\omega) = \frac{F_{SR}}{2} \left\{ [\sin(\omega + \Omega - \lambda_{\odot}) - \sin(\omega - \Omega + \lambda_{\odot})] \sin(i) \cos^2\left(\frac{\epsilon}{2}\right) \right. \\ \left. + [\sin(\omega + \Omega + \lambda_{\odot}) - \sin(\omega - \Omega - \lambda_{\odot})] \sin(i) \sin^2\left(\frac{\epsilon}{2}\right) \right. \\ \left. + [\sin(\omega + \lambda_{\odot}) - \sin(\omega - \lambda_{\odot})] \cos(i) \sin(\epsilon) \right\}$$

Notice these expressions use the obliquity of the ecliptic, $\epsilon = 23.5^\circ$, and the ecliptic longitude of the Sun, λ_{\odot} .

Let's use the true anomaly when the satellite exits the Earth's shadow, ν_{exit} , and the true anomaly on entering the shadow, ν_{ent} . If we separate common terms for clarity, the rates of change in the orbital elements and perigee height are

$$\tau_{ee} = \tan^{-1} \left(\frac{\sqrt{1-e^2} \tan\left(\frac{\nu_{ent}}{2}\right)}{1+e} \right) - \tan^{-1} \left(\frac{\sqrt{1-e^2} \tan\left(\frac{\nu_{exit}}{2}\right)}{1+e} \right)$$

$$r_{ee} = r_{ent} - r_{exit} \quad r_{ee2} = r_{ent}^2 - r_{exit}^2$$

$$r_{ee\nu} = r_{ent} \sin(\nu_{ent}) - r_{exit} \sin(\nu_{exit})$$

$$r_{ee\nu2} = r_{ent}^2 \sin(\nu_{ent}) - r_{exit}^2 \sin(\nu_{exit})$$

$$\dot{a} = \frac{1}{\pi n a} \left[r_{ee\nu} S_p + a \{ \cos(E_{ent}) - \cos(E_{exit}) \} R_p \right]$$

$$\dot{e} = \frac{n S_p}{2\pi\mu} \left[3a^2 \sqrt{1-e^2} \tau_{ee} - \frac{1}{2e} r_{ee\nu2} + \frac{a}{2e} (1-4e^2) r_{ee\nu} \right]$$

$$-\frac{n R_p}{4\pi\mu} \left[r_{ee2} + \frac{a(1-e^2)}{e^2} r_{ee} + \frac{1}{e} \{ r_{ent}^2 \cos(\nu_{ent}) - r_{exit}^2 \cos(\nu_{exit}) \} \right]$$

$$\dot{r}_p = \frac{n}{2\pi} [(1-e) \dot{a} - a \dot{e}]$$

$$\begin{aligned}
 i &= \frac{nW}{2\pi\mu} \left\{ \left[\frac{r_{ee\nu_2}}{2(1-e^2)} + \frac{1+2e^2}{2(1-e^2)} ar_{ee\nu} - \frac{3a^2e}{\sqrt{1-e^2}} \tau_{ee} \right] \cos(\omega) - \frac{r_{ee2}}{2e} \sin(\omega) \right\} \\
 \dot{\Omega} &= \frac{nW}{2\pi\mu \sin(i)} \left\{ \frac{r_{ee2}}{2e} \cos(\omega) \right. \\
 &\quad \left. + \left[\frac{r_{ee\nu_2}}{2(1-e^2)} + \frac{1+2e^2}{2(1-e^2)} ar_{ee\nu} - \frac{3a^2e}{\sqrt{1-e^2}} \tau_{ee} \right] \sin(\omega) \right\} \\
 \dot{\omega} &= -\frac{nR_p}{2\pi\mu e} \left[3a^2 \sqrt{1-e^2} \tau_{ee} + \frac{1}{2e} r_{ee\nu_2} - \frac{a}{2e} (1+2e^2) r_{ee\nu} \right] \\
 &\quad + \frac{nS_p}{4\pi\mu e^3} \left[e \{ r_{ent}^2 \cos(\nu_{ent}) - r_{exit}^2 \cos(\nu_{exit}) \} - a(1-e^2) r_{ee} \right] - \frac{\dot{\Omega}n}{2\pi} \cos(i)
 \end{aligned}$$

For a satellite in sunlight during the entire orbit, set $\nu_{ent} = \nu_{exit} + 360^\circ$ and $r_{ent} = r_{exit}$. Thus,

$$\begin{aligned}
 \dot{a} &= 0 \\
 \dot{e} &= \frac{3\sqrt{1-e^2}}{2na} S_p \\
 \dot{r}_p &= -a\dot{e} \\
 \dot{i} &= -\frac{3We \cos(\omega)}{2na\sqrt{1-e^2}} \\
 \dot{\Omega} &= -\frac{3We \sin(\omega)}{2na\sqrt{1-e^2} \sin(i)} \\
 \dot{\omega} + \dot{\Omega} \cos(i) &= -\frac{3\sqrt{1-e^2}}{2nae} R_p
 \end{aligned} \tag{8-58}$$

We must still determine the true anomaly on entering and exiting the shadow, as well as geocentric distances and times—a geometrical problem long familiar to astronomers. Appropriate methods for determining these quantities in terms of the satellite's position and the Sun's geocentric coordinates are in Sec. 3.5.

Effects of Solar-Radiation Pressure

We can gain an understanding of the general effects of solar-radiation pressure by examining the work done by the disturbing force. It will alter the orbit's energy and semimajor axis. From Eq. (8-52),

$$\Delta a = \frac{2a^2}{\mu} \Delta \xi = \frac{2a^2}{\mu} \vec{F} \cdot \vec{\Delta r} = -\frac{2a^2}{\mu} F \Delta r_\odot$$

where F is given by Eq. (7-38), and Δr_{\odot} is the radial displacement of the satellite relative to the Sun. Hence, over that part of the orbit during which the satellite moves toward the Sun, the energy and semimajor axis decrease; over that part of the orbit during which the satellite moves away from the Sun, both values increase. If the net displacement while the satellite is in the sunlight is Δr_{\odot} ,

$$\Delta a = -\frac{2a^2}{\mu} F \Delta r_{\odot} \quad (8-59)$$

Thus, for any orbit entirely in sunlight, the change in semimajor axis in one orbit revolution is zero. As with atmospheric drag, the disturbing force, F , is proportional to the satellite's area-to-mass ratio, so the perturbations will be greatest on a satellite with a large area-to-mass, such as the Echo balloon in the early 1960s.

Solar-radiation pressure effects give rise to "m-yearly" variations because of the apparent annual motion of the Sun. The eccentricity and argument of perigee are the main elements that vary. Neglecting shadow effects, it's obvious that the Sun's annual motion causes variations. In fact, Kaula showed that, for a satellite always in the sunlight, we can derive the disturbing acceleration from the gradient of a potential function similar to the zonal-harmonic analog of the third-body disturbing function. Unfortunately, most satellites experience some eclipse periods, which limits application of this technique.

Now let's compare the relative magnitudes of solar-radiation pressure and drag. Consider the acceleration of drag and create a ratio with the two specific forces (eliminating the minus signs):

$$\frac{a_{drag}}{a_{radiation}} = \frac{\frac{1}{2}\rho \frac{c_D A}{m} v_{rel}^2}{\frac{p_{SR} c_R A_{\odot}}{m}}$$

Now, let the coefficient of drag equal 2.0 and the reflectivity equal 1.0. Also assume the frontal area, A , and the area exposed to the Sun, A_{\odot} , are the same. Thus,

$$\frac{a_{drag}}{a_{radiation}} \approx \frac{\rho v_{rel}^2}{p_{SR}}$$

From Table 7-4, the atmospheric density at 800 km altitude is 1.170×10^{-14} kg/m³, and the velocity of a satellite in a circular orbit at 800 km is

$$v_{cs} = \sqrt{\frac{\mu}{r}} = 7451.8 \frac{\text{m}}{\text{s}}$$

If we use these values, the ρv^2 term becomes about 6.50×10^{-7} kg·m / s², which is nearly the same as the solar-radiation pressure at 800 km (see Sec. 7.6.4). Traditionally, 800 km is regarded as the approximate transition altitude between drag and solar-radiation pres-

sure (4.5×10^{-6} kg·m / s²). These numbers don't match exactly due to the variability in atmospheric density.

Summary of Solar-Radiation Effects

1. Radiation pressure induces periodic variations in all orbital elements, exceeding the effects of atmospheric drag at heights above about 800 km.
2. As with third-body attraction, induced changes in perigee height can seriously affect the satellite's lifetime.
3. The period for the variations can be as long as a year due to the annual motion of the Earth about the Sun (*m-yearly* effects).
4. The effects are usually small for most satellites, except for those with very low mass and large surface area.
5. The periodic effects become more complex if the satellite passes through the Earth's shadow.

8.7 Forming Analytical Solutions

The large variety of equations in the preceding sections should indicate that there are numerous analytical formulations for solving perturbed motion. Similar to the numerical solutions, the answer depends on the forces we model. Unfortunately, besides Newton's second law, there isn't an underlying equation like Eq. (7-46) that we use for *every* analytical solution. That's because we've performed a variation of parameters, some averaging, or another transformation to produce a simpler form of the equations of motion. We saw that this process usually relies on series expansions to express the motion—a source of practical difficulty. Orbital elements help us solve this problem because they give us insight into the behavior of the perturbative effect over time through the secular, short-periodic, and long-periodic variations discussed earlier. Although these distinctions help us decide which effects to model, the practical difficulty of an infinite series still remains.

When Lagrange investigated the secular and long-periodic motion of the planets, he retained first-order terms in the series by inspection. This was adequate for small perturbations over short time intervals. Unfortunately, as we include higher-order terms into the solution for larger perturbations, we need a systematic method to isolate the terms to retain. One technique is the method of averaging discussed earlier. In this method, we average the equations of motion to eliminate the short-periodic variations and form *mean* elements. This process of finding mean elements is an important step in forming semianalytical solutions. We'll explore this in greater detail in Sec. 8.8.

It's important to realize that each averaging process will produce "mean" elements that are specific to that process. This is important because most analytical techniques use mean elements, but all mean elements aren't the same. In Chap. 7, we saw that the numerical techniques assume osculating vectors that capture the true time-varying behavior (Fig. 8-1) of the state vector. In contrast, the initial elements for an analytical

theory should be mean elements that are consistent with the assumptions of the model—the averaging process and the fidelity of the disturbing forces. The elements need only to capture the “mean” behavior—the essential character of the time-varying motion. As before, selecting the terms that produce these mean elements takes on numerous forms.

I'll only present one complete simplified analytical technique (**PKEPLER**) which captures the secular perturbative effects resulting from J_2 . Other solution methods become very involved and a full derivation usually takes many pages. The techniques of Kozai and Brouwer are of interest because of their current use although semi-analytical routines are more accurate and efficient on modern computers. Unfortunately, complete derivations are quite lengthy, but they're readily available in the literature. I'll introduce the basic fundamentals of each of these routines.

8.7.1 Application: Perturbed Two-body Propagation

For general-purpose studies, a first-order approximation of Kepler's problem (**PKEPLER**) introduces a way to estimate a satellite's position under the influence of secular gravitational and drag perturbations. Although this is presented as an academic application, the concepts obviously derive from more complex theories, so they can give us insight into those techniques. Other techniques exist that adjust the f and g functions, but I won't describe them here. Consult Baker (1967, 198–208) and Burton and Melton (1992) for alternate approaches.

Basically, the technique I'll describe provides methods to account for a somewhat crude estimate of drag and central-body gravitational forces. Developing the routine is simple because it's adapted directly from the approach for classical orbital elements in Algorithm 30. The major variation depends on the available data. We can approximate the secular effects of drag on the semimajor axis, eccentricity, and mean anomaly if we have an estimate of the mean motion's rate of change. The remaining secular effects on node and argument of perigee are determined using the central-body gravitational model. The overall idea of **PKEPLER** is to determine the classical orbital elements, as in Algorithm 30, update the elements through the desired time, and reform the future state vector. Its major difference from Algorithm 30 is that, while updating the orbital elements, it adds the effects of perturbations (whenever possible).

The semimajor axis, mean anomaly, and eccentricity may be updated if an estimate of the mean-motion rate is known. The mean anomaly is found using a Taylor-series expansion, which is usually truncated after four terms:

$$M = M_o + n_o \Delta t + \frac{\dot{n}}{2} \Delta t^2 + \frac{\ddot{n}}{6} \Delta t^3$$

It isn't necessary to use an average value for the mean motion because the rest of the neglected perturbations and simplifications will mask any difference to the delta from the mean motion. This results in a crude approximation with a short time of validity. Accuracy will degrade quickly with time. The semimajor axis and eccentricity are updated using Eq. (8-49) and Eq. (8-50). Secular changes in longitude of ascending node and argument of

perigee are determined through the simplified equations developed earlier in this chapter [Eq. (8-37) and Eq. (8-39)]. With these newly created elements, we can determine the future (or past) state vector.

Implementing the PKEPLER Solution

The algorithm is very similar to Algorithm 30 but adds formulas required to update the classical orbital elements. We must apply the algorithm so special orbits (circular-inclined and equatorial) are possible. The solution is the same as in Algorithm 30, with the eccentric anomaly being set to the argument of latitude or the true longitude.

We can make this algorithm reasonably fast by storing some temporary variables. For example, we can store $n_o J_2 / p^2$ (μ is equal to $1.0 \text{ ER}^3/\text{TU}^2$). By eliminating R_{\oplus} (because it's 1.0 ER) as an independent quantity, no matter the gravitational model, we can delete several computational statements. Finding estimates of the mean motion rate of change can be a source of practical difficulty. We prefer numerical techniques to find \dot{n} and \ddot{n} , but we often use 2-line element sets because of their current availability. Be very careful when using these relations. For decades 2-line element sets from Space Command have transmitted the mean-motion rate *already divided by two*. Don't divide twice! The algorithm incorporates traditional values, *not* those contained in 2-line element sets. If we don't have an estimate of \dot{n} and \ddot{n} , we can simply omit the equations using those quantities. But remember, we'll no longer have any approximation for drag in the solution. If the orbit is circular, we must find a way to assign a value to the eccentric anomaly. We can use either the argument of latitude for inclined orbits, or the true longitude for equatorial orbits. Finally, notice that we omit parabolic and hyperbolic orbits because we seldom encounter them and because perturbations aren't usually significant for satellites speeding away from Earth on escape trajectories.

ALGORITHM 57: PKEPLER ($\overset{\curvearrowleft}{r}_o, \overset{\curvearrowleft}{v}_o, \Delta t, \overset{\curvearrowleft}{n}, \overset{\curvearrowleft}{\ddot{n}} \Rightarrow \overset{\curvearrowleft}{r}, \overset{\curvearrowleft}{v}$)

ELORB ($\overset{\curvearrowleft}{r}_o, \overset{\curvearrowleft}{v}_o \Rightarrow a_o, e_o, i_o, \Omega_o, \omega_o, v_o u, \lambda_{true}, \tilde{\omega}_{true}$)

```

[ IF  $e \neq 0$ 
    vtoAnomaly ( $v, e \Rightarrow E_o$ )
 ELSE
     $E_o = u$  or  $E_o = \lambda_{true}$ 
     $M_o = E_o - e_o \text{SIN}(E_o)$ 

```

$$p_o = a_o (1 - e_o^2), n_o = \sqrt{\frac{u}{a_o^3}}$$

Update for perturbations

$$\begin{aligned}
 a &= a_o - \frac{2a_o}{3n_o} \dot{n} \Delta t & e &= e_o - \frac{2(1-e_o)}{3n_o} \dot{n} \Delta t \\
 \Omega &= \Omega_o - \frac{3n_o R_{\oplus}^2 J_2}{2p_o^2} \cos(i) \Delta t \\
 \omega &= \omega_o + \frac{3n_o R_{\oplus}^2 J_2}{4p_o^2} \left\{ 4 - 5 \sin^2(i) \right\} \Delta t \\
 M &= M_o + n_o \Delta t + \frac{\dot{n}}{2} \Delta t^2 + \frac{\ddot{n}}{6} \Delta t^3 \\
 p &= a(1-e^2) \\
 \text{KepEqtnE } (M, e \Rightarrow E) \\
 \text{IF } e \neq 0 \\
 &\quad \text{Anomalytor } (E, e \Rightarrow \nu) \\
 \text{ELSE} \\
 &\quad u = E \text{ or } \lambda_{true} = E \\
 \text{RANDV } (p, e, i_o, \Omega, \omega, \nu(u, \lambda_{true}, \tilde{\omega}_{true}) \Rightarrow \vec{r}, \vec{v})
 \end{aligned}$$

8.7.2 Kozai's Method

Kozai's basic theory in 1959 was to use Lagrange's VOP equations and solve them using an ad hoc averaging technique. His method was one of the first to account for the Earth's asphericity ($J_2 \dots J_5$ only), but he neglected drag effects. Indeed, little observational data even existed for the upper atmosphere, and the importance (and difficulty) of modeling effects of drag wasn't widely recognized at that time. Even so, the U.S. Space Command often uses a variation of his method, mainly for its inherent speed. One must remember that the accuracy of this technique is limited because of the restricted models of disturbing forces. Knowles (1995) shows data suggesting the accuracy of these types of theories is at best, about 1 km in total position error at epoch, with significant growth over short time intervals. Finally, using these equations is no small task! Today's fast personal computers all but eliminate the need for such analytical theories because numerical-integration techniques work so well.

I'll present only the equations we need because the complete derivations are involved and aren't necessary for this book. The overall idea is the same as for Algorithm 30: the orbital elements are updated individually, and the position and velocity vectors are reconstructed at the desired future (or past) time. Differences arise in treating secular and short- and long-periodic variations. The overall solution combines the secular (dot terms), short-

periodic (subscripted SP), and long-periodic terms (subscripted LP). Because only zonal harmonics are considered, there are no m -dailies or linear-combination terms. Notice the similarity to expressions for the semimajor axis and mean motion, as well as to the first-order approximation for the mean anomaly (Eq. (8-41)).

$$a = \bar{a} + \Delta a_{SP} \quad \bar{a} = a_o \left\{ 1 - \frac{3R_{\oplus}^2 J_2 \sqrt{1-e_o^2}}{4p^2} (2 - 3 \sin^2(i_o)) \right\}$$

$$e = e_o + \Delta e_{LP} + \Delta e_{SP}$$

$$i = i_o + \Delta i_{LP} + \Delta i_{SP}$$

$$\omega = \omega_o + \dot{\omega} \Delta t + \Delta \omega_{LP} + \Delta \omega_{SP}$$

$$\Omega = \Omega_o + \dot{\Omega} \Delta t + \Delta \Omega_{LP} + \Delta \Omega_{SP}$$

$$M = M_o + \bar{n} \Delta t + \Delta M_{SP} \quad \bar{n} = n_o \left\{ 1 + \frac{3R_{\oplus}^2 J_2 \sqrt{1-e_o^2}}{4p^2} (2 - 3 \sin^2(i_o)) \right\}$$

$$n_o^2 a_o^3 = \mu$$

Kozai (1959, 370) begins by showing the short-periodic variations caused by the second zonal coefficient. Taff (1985, 332–340) also presents the theory, including comments about its particular type: first-order, second-order, and so on. Note that I've substituted as follows to convert his notation, A_2 , A_3 , and A_4 . Note the correction, 35 instead of 55, to the conversion listed in Taff. We must include the Earth's radius only to keep the correct units in each expression. Also note that the unity values of canonical units (μ , R_{\oplus} , ...) permit us to drop many mathematical operations.

$$A_2 = \frac{3J_2 R_{\oplus}^2}{2} \quad A_3 = -J_3 R_{\oplus}^3 \quad A_4 = -\frac{35J_4 R_{\oplus}^4}{8}$$

The *secular* variations are of interest, and they are virtually identical to those of Merson [Eq. (8-38) and Eq. (8-40)]; however, due to the method of averaging used, the J_2^2 term is different, and Kozai didn't include effects of J_6 . The J_2^2 terms are

$$\begin{aligned} \dot{\Omega}_{J_2^2} \Rightarrow & \frac{3J_2^2 R_{\oplus}^4 n \cos(i)}{32p^4} \left\{ -216 - 24e^2 + 288\sqrt{1-e^2} \right. \\ & \left. + (240 - 30e^2 - 432\sqrt{1-e^2}) \sin^2(i) \right\} \end{aligned}$$

$$\dot{\omega}_{J_2^2} \Rightarrow \frac{9nJ_2^2R_{\oplus}^4}{384p^4} \left\{ 192 + 56e^2 - 192\sqrt{1-e^2} \right. \\ \left. + (-172 + 288\sqrt{1-e^2}) \sin^2(i) + e^2 \sin^4(i) \right\}$$

Kozai's expressions for short-periodic variations are in Sec. 8.6.

8.7.3 Brouwer's Method

In the same journal in which Kozai's theory appeared, Brouwer (1959) proposed a similar method. It also didn't include drag. Here, I'll present some equations we need, plus several changes developed over the years. You'll find many variations of Brouwer's theory in use today, but remember, the Brouwer theory has limited force models and a general accuracy of about 1 km at epoch (Knowles, 1995).

Brouwer's original work has been refined over time. The first major refinement came in 1961 when Brouwer and Hori published a simplified treatment of drag as well as effects of the nonspherical Earth ($J_2 \dots J_5$ only). Next, because Brouwer's theory uses Delaunay variables, singularities exist for zero eccentricity, zero inclination, and critical inclination. Lyddane solved the zero eccentricity and inclination problems in 1963 by adding a term to some of the formulas. Deprit (1969) used a similar approach to treat the critical inclination. Finally, because some centers use this theory often in daily operations, we need a way to account for the effects of third bodies. Hujasak (1979) addressed this disturbing force, along with geosynchronous resonance. All these developments give Brouwer's method the appearance of being a superb analytical theory, but it has several significant shortcomings. It's fine for general use in applications needing limited accuracy. But as computational power increases and satellite systems demand more accuracy, its effectiveness diminishes—it's still an analytical series approximation accounting for a large but incomplete subset of dominant perturbations, so it's only moderately accurate. The U.S. Air Force and Navy operate similar versions of this theory. Recently, Cefola (1996) has shown a way to include m -daily effects in the Navy's theory, with remarkable success and improved accuracy.

As suggested earlier, Brouwer's use of Delaunay variables and the perturbation method was a major departure from Kozai's theory. Kozai used ad hoc averaging, whereas Brouwer used the method of canonical transformations. The underlying approach used techniques introduced in Sec. 8.7.2 on the Hamiltonian VOP. Brouwer did two canonical transformations to simplify the equations of motion. His notation consisted of a series of primes and double primes to represent the transformed Delaunay variables from each canonical transformation. I'll use subscripts.

Brouwer refers to his elements as *mean elements*. However, they differ from Kozai's elements and other so-called mean elements because they depend on the assumptions of the perturbation method. All theories strive to compute the same osculating elements but may differ subtly in the way they separate secular, short-periodic, and long-periodic varia-

tions. This is of little consequence *within* a theory because we're interested only in the final results. However, if we mix elements or formulas from different theories, their varying assumptions lead to unintended results.

The particular definition for these mean quantities determines the form of the resulting equations. Thus, if we examine the secular rates of change for the orbital elements from the general variation of parameters (Sec. 8.3), from Kozai theory (Sec. 8.7.2), and from Brouwer's theory, we'll find three different expressions! What's worse, they're all correct, based on the assumptions for each theory and the way we applied the averaging. The bottom line? Be very cautious when using the term *mean variables*, document how you get them, and use the proper theory with those elements.

Brouwer's basic premise was to eliminate M , Ω , and ω from the Hamiltonian. Because the equations of motion [Eq. (8-27)] are very simple, if we could drop all three of these values, the equations of motion would reduce to L , h , and H as constant values and M , Ω , and ω as secular rates in the Hamiltonian formulation. Ω is not in the original Hamiltonian, so two successive [canonical] transformations remove M and ω , respectively. Brouwer shows the first transformation with a single prime and the second with a double prime.

Brouwer and Clemence (1961, 398–414) give the algorithm for propagating a satellite's orbit. I use a slightly different notation, keeping the familiar symbols for the classical orbital elements instead of potentially confusing alternatives. The overall approach is similar to that of Kozai, with updates in orbital elements for secular, long-periodic, and short-periodic variations. However, Brouwer uses different “mean” or initial values:

$$\begin{aligned} a &= \bar{a} + \Delta a_{SP} + \Delta a_{LP} \\ e &= e_o + \Delta e_{SP} + \Delta e_{LP} \\ i &= i_o + \Delta i_{SP} + \Delta i_{LP} \\ \omega &= \omega_o + \dot{\omega}\Delta t + \Delta\omega_{SP} + \Delta\omega_{LP} \\ \Omega &= \Omega_o + \dot{\Omega}\Delta t + \Delta\Omega_{SP} + \Delta\Omega_{LP} \\ M &= M_o + n\Delta t + \Delta M_{SP} + \Delta M_{LP} \quad n_o^2 a_o^3 = \mu \end{aligned}$$

Brouwer makes several substitutions to simplify the resulting equations, but these substitutions hamper comparisons with other methods. As mentioned earlier, I'll present a few equations in a common notation and format, so we can more easily compare this method with others. The substitutions are listed below, in case you want to compare from scratch (bars denote “mean” elements):

$$\eta = \sqrt{1 - \bar{e}^2} \quad \theta = \cos(\bar{i})$$

$$\gamma_2 = \frac{J_2 R_\oplus^2}{2\bar{a}} \quad \gamma'_2 = \frac{\gamma_2}{\eta^4} = \frac{J_2 R_\oplus^2}{2\bar{p}^2}$$

$$\begin{aligned}\gamma_3 &= \frac{-J_3 R_{\oplus}^3}{\bar{a}^3} & \gamma'_3 &= \frac{\gamma_3}{\eta^6} = \frac{-J_3 R_{\oplus}^3}{\bar{p}^3} \\ \gamma_4 &= \frac{-3J_4 R_{\oplus}^4}{8\bar{a}^4} & \gamma'_4 &= \frac{\gamma_4}{\eta^8} = \frac{-3J_4 R_{\oplus}^4}{8\bar{p}^4} \\ \gamma_5 &= \frac{-J_5 R_{\oplus}^5}{\bar{a}^5} & \gamma'_5 &= \frac{\gamma_5}{\eta^{10}} = \frac{-J_5 R_{\oplus}^5}{\bar{p}^5}\end{aligned}$$

Now let's find the secular variations. Remember, Brouwer defines three elements to have no secular variations caused by the zonal harmonics, $\dot{a} = \dot{e} = \dot{i} = 0$. The remaining three variables are very similar to those presented earlier [Eq. (8-38) to Eq. (8-42)] and to Kozai, except for the order (J_6 was not included), and the significantly different J_2^2 term, based on the specific averaging techniques. The J_2^2 term of the *secular* variations for longitude of ascending node and argument of perigee (respectively) are

$$\begin{aligned}\dot{\Omega}_{J_2^2} &\Rightarrow \frac{3J_2^2 R_{\oplus}^4 n \cos(i)}{32\bar{p}^4} \left\{ -9 - e^2 - 6\sqrt{1-e^2} + \left(10 - \frac{5}{4}e^2 + 9\sqrt{1-e^2} \right) \sin^2(i) \right\} \\ \dot{\omega}_{J_2^2} &\Rightarrow \frac{9n J_2^2 R_{\oplus}^4}{384\bar{p}^4} \left\{ 1062 + 168e^2 + 576\sqrt{1-e^2} + (-2382 - 108e^2 - 1584\sqrt{1-e^2}) \sin^2(i) \right. \\ &\quad \left. + (1290 - 135e^2 + 1080\sqrt{1-e^2}) \sin^4(i) \right\}\end{aligned}$$

The long-periodic variations are identical to Kozai's formulas.

8.8 Semianalytical Solutions

Speed has been the main concern for propagation theories. So analytical theories, which were relatively easy to use and gave reasonable results, dominated numerical methods for centuries. In fact, until modern times, numerical methods were not realistic for most problems. The analytical techniques were accurate enough for scientific uses of the day, but we never could have gotten to the Moon without numerical techniques. Today's computers have made numerical integration possible and very popular for operational systems. These systems depend on numerical-integration programs, with sophisticated mathematical models of perturbing accelerations, which yield very precise results. Unfortunately, even with very fast computers, we often can't afford the lengthy processing times of certain long-range analyses and detailed studies. As an example, consider the computational burden of simulating a five-year mission on a workstation carried out by Sabol (1994) in which he analyzed the ELLIPSO constellation (six satellites with $e = 0.346\ 08$, $a = 10,559$ km). Wallace (1995) did a similar five-year investigation of the Teledesic constellation of

840 satellites in nearly circular orbits at 700 km altitude. Numerical techniques *could* have provided the answers, but only with much more number crunching. Thus, semianalytical theories are a significant step forward for large *and* accurate long-term studies.

Today, we trade-off speed and accuracy by combining the best features of numerical and analytical techniques. A semianalytical approach gets the best speed and accuracy by taking advantage of the characteristics of the effects present on a satellite's orbit. The underlying approach is to separate the short-periodic contributions from the long-periodic and secular effects so we can numerically propagate the mean element rates. This is desirable because the short-periodic contributions constrain the integrator step sizes, which allows us to numerically integrate the equations of motion governing the long-periodic and secular effects with large step sizes (typically on the order of a day). Because the short-periodic contributions are 2π periodic in the fast variable, we model them analytically using Fourier series. Thus, we recover the short-periodic contributions at the integrator's step-size times and then combine them with the secular and long-periodic contributions. If we want short-periodic variations at times other than the integrator times (which is often the case), we use an efficient interpolation scheme.

Many semianalytical techniques exist. Of note are the Hoots Analytic Dynamic Ephemeris Theory (HANDE, Hoots, 1982 and Hoots and France, 1987), and the Semianalytical Liu Theory (SALT, Liu, 1973 and 1974). Each of these routines is designed for satellites in very low orbits around the Earth and is well-suited for that purpose. We'll focus on the Draper Semianalytical Satellite Theory, which is more complete, applicable to all orbit classes, and highly flexible.

8.8.1 The Draper Semianalytical Satellite Theory (DSST)

Paul Cefola and his colleagues at the Charles Stark Draper Laboratory developed this theory in the mid-1970s and early 1980s. Many conference proceedings and technical reports detail the approach. Cefola (1972), McClain (1992, 1978) and Danielson et al. (1993, 1994, and 1995) present excellent summaries of the mathematical technique used in this theory. Early (1986) presents a broad overview of the rationale behind the theory. Finally, Barker et al. (1995) and Fonte et al. (1995) have thoroughly compared analytical and numerical techniques with DSST, showing the accuracy and effectiveness of DSST.

DSST has many distinctions over other semianalytical techniques, including an *extensive* treatment of perturbing forces (central body including tesseral harmonics, drag, third-body, solar-radiation, and others); great flexibility, so we can tailor the algorithm to the application; recent improvements to documentation; and wide use. Although some applications don't require the ability to incorporate many force models, others do. It's useful to have a theory which accounts for all significant perturbations. History has shown that, once a theory gains widespread use, the inherent assumptions of the theory are soon forgotten, and it's often used inappropriately. A theory that handles *many* different scenarios lessens the possibility for failure due to unintended application.

We can state DSST's basic philosophy rather simply. We first establish equations of motion for the osculating elements. We then model conservative perturbations through

Lagrange's VOP equations [Eq. (8-12)] and nonconservative forces through Gauss's VOP equations [Eq. (8-25)]. To simplify the discussion, we'll restrict ourselves to a single disturbing perturbation depending on a small parameter, ϵ . Including a delta function, the VOP equations take the general form

$$\begin{aligned}\frac{d\hat{a}_i}{dt} &= n(\hat{a}) \delta_{i6} + \epsilon F_i(\hat{a}_1, \hat{a}_2, \dots, \hat{a}_5, \lambda_M) \\ \delta_{i6} &= 0 \quad i \neq 6 \quad i = 1, 2, \dots, 6 \\ \delta_{i6} &= 1 \quad i = 6\end{aligned}\tag{8-60}$$

We can write the osculating elements as

$$\hat{a}_i = a_i + \eta_i(a_1, a_2, \dots, a_5, \lambda_M) \quad i = 1, 2, \dots, 6\tag{8-61}$$

where a_i are mean elements containing the secular and long-periodic variations and η_i are 2π periodic in λ_M , so we can apply averaging. The η_i are the short-periodic variations. The essence of averaging consists of integrating the equations of motion over a particular variable to eliminate its explicit appearance in the equations of motion. Applying the generalized method of averaging to the equations of motion will remove the short-periodic terms, as Nayfeh (1973) points out. The short-periodic variations are assumed to contain all the high-frequency components in the osculating elements, \hat{a}_i , so the mean elements vary slowly with time.

More specifically, we define the relationship between the osculating elements, the mean elements, and the short-periodic variations through a *near-identity transformation*, Eq. (8-62). This transformation expresses the osculating elements in terms of mean elements

$$\hat{a}_i = a_i + \sum_{j=1}^m \epsilon^j \eta_{ij}(a_1, a_2, \dots, a_5, \lambda_M)\tag{8-62}$$

where $\epsilon^j \eta_{ij}$ represents a small short-periodic variation of order j in element a_i . Note that ϵ is a naturally occurring small parameter associated with the perturbing force— J_2 for example.

The *mean* elements are assumed to be governed by VOP equations of the form

$$\frac{da_i}{dt} = n(a_1) \delta_{i6} + \sum_{j=1}^m \epsilon^j A_{ij}(a_1, a_2, \dots, a_5, \lambda_M)\tag{8-63}$$

where $n(a_1)$ is the *mean* mean motion. The first term applies only to the fast variable, which is usually the *mean* mean longitude, λ_M .

We must now relate the (as yet) undefined functions A_{ij} and η_{ij} to the osculating VOP equations. The terms ϵF_i give the osculating element's rate of change due to the perturbing

forces as functions of the osculating elements. We now expand the osculating rate functions, F_i , about the mean elements, and the osculating mean motion about the *mean* mean motion. In other words, we expand the disturbing acceleration about the mean elements. Expanding with Taylor's theorem as a power series in $\Delta a = \hat{a} - a$ and rearranging the result as a power series in ϵ yields

$$\begin{aligned} F_i(\hat{a}, \hat{\lambda}_M) &= F_i(a, \lambda_M) + \sum_{j=1}^m \epsilon^j f_{ij}(a_1, a_2, \dots, a_5, \lambda_M) \\ n(\hat{a}_1) &= n(a_1) + \sum_{j=1}^m \epsilon^j N_j(a_1) \\ f_{i1} &= \sum_{j=1}^6 \frac{\partial F_i}{\partial a_j} \eta_{j1} \quad f_{i2} = \sum_{j=1}^6 \frac{\partial F_i}{\partial a_j} \eta_{j2} + \frac{1}{2} \sum_{j=1}^6 \sum_{k=1}^6 \frac{\partial^2 F_i}{\partial a_j \partial a_k} \eta_{j1} \eta_{k1} \\ &\dots \\ N_1 &= -\frac{3\eta}{2a_1} n(a_1) \quad N_2 = \left\{ -\frac{3\eta_{12}}{2a_1} + \frac{15}{8} \left(\frac{\eta_{11}}{a_1} \right)^2 \right\} n(a_1) \\ &\dots \end{aligned}$$

Thus the osculating VOP equations are expressed in the mean elements as

$$\frac{d\hat{a}_i}{dt} = \delta_{i6} \left(n + \sum_{j=1}^m \epsilon^j N_j \right) + F_i(a, \lambda_M) + \sum_{j=1}^m \epsilon^j f_{ij}(a_1, a_2, \dots, a_5, \lambda_M) \quad (8-64)$$

Next, differentiating with respect to time, the near-identity transformation yields another expression for the right-hand-side of the VOP equations expressed in mean elements.

$$\frac{d\hat{a}_i}{dt} = \frac{da_i}{dt} + \sum_{j=1}^m \epsilon^j \sum_{k=1}^6 \frac{\partial \eta_{ij}}{\partial a_k} \frac{da_k}{dt}$$

Substituting the mean element VOP equations and rearranging terms yields

$$\frac{d\hat{a}_i}{dt} = n(a_1) \delta_{i6} + \sum_{j=1}^m \epsilon^j \left\{ A_{ij} + n(a_1) \frac{\partial \eta_{ij}}{\partial \lambda_M} + \sum_{l=1}^{m-j} \epsilon^l \sum_{k=1}^6 A_{ki} \frac{\partial \eta_{ij}}{\partial a_k} \right\}$$

We now set the right-hand-side of the equation equal to the right-hand-side of Eq. (8-64). Because both expressions are power series in ϵ , terms of the same power in ϵ must be equal. The results for $j = 1, 2$ are:

$$\begin{aligned} A_{i1} + \frac{\partial \eta_{i1}}{\partial \lambda_M} n(a_1) &= F_i(a_1, a_2, \dots, a_5, \lambda_M) + N_1 \delta_{i6} \\ A_{i2} + \frac{\partial \eta_{i1}}{\partial \lambda_M} n(a_1) &= f_{i1}(a_1, a_2, \dots, a_5, \lambda_M) + N_2 \delta_{i6} - \sum_{k=1}^6 \frac{\partial \eta_{i1}}{\partial a_k} A_{k1} \end{aligned} \quad (8-65)$$

These equations relate (at each order of ϵ) the unknown functions A_{ij} and the partial derivatives of the (as yet) unknown short-periodic variations to the known terms of the power series expansion of the osculating perturbing function, f_{ij} . Both sides of these equations still retain the fast variable dependance. We eliminate this dependance by averaging over the fast variable (λ_M) on the interval $-\pi$ to π . Let $\langle \cdot \rangle$ denote the averaging operation with respect to λ_M . Because the functions η_{ij} are 2π periodic in λ_M , so are their partial derivatives. Furthermore, we may specify that η_{ij} have zero mean— $\langle \eta_{ij} \rangle = 0$. This constraint also devolves to the partial derivatives. Consequently, averaging Eq. (8-65) gives

$$\begin{aligned} A_{i1} &= \langle N_1 \rangle \delta_{i6} + \langle F_i(a_1, a_2, \dots, a_5, \lambda_M) \rangle \quad j = 1 \\ A_{i2} &= \langle N_2 \rangle \delta_{i6} + \langle f_{i1}(a_1, a_2, \dots, a_5, \lambda_M) \rangle \end{aligned} \quad (8-66)$$

thus the first order mean element rates, A_{i1} , are obtained by averaging the VOP perturbing function, $F_i(a, \lambda_M)$. The higher order mean element rates are obtained by averaging the functions f_i constituting the power series expansion of F about the mean elements.

The final step determines the short-periodic variations η_{ij} . Subtracting Eq. (8-66) from Eq. (8-65) yields a partial differential equation for the η_{ij} . For $j = 1$, we obtain

$$n(a_1) \frac{\partial \eta_{i1}}{\partial \lambda_M} = N_1^S \delta_{i6} + F_i^S(a, \lambda_M)$$

$$\begin{aligned} F_i^S(a, \lambda_M) &= F_i(a, \lambda_M) - \langle F_i(a, \lambda_M) \rangle \\ N_1^S &= N_1 - \langle N_1 \rangle \end{aligned}$$

Note that F_i^S and N_1^S have zero mean by definition (i.e. $\langle F_i^S \rangle = 0$). Integration yields the function η_{i1} and an arbitrary integration “constant” that’s a function of the slow variables $C(a_1, a_2, \dots, a_5)$.

$$\eta_{i1} = \frac{1}{n} \int \{ N_1^S \delta_{i6} + F_i^S(a, \lambda_M) \} d\lambda_M + C(a_1, a_2, \dots, a_5)$$

Because the function C is arbitrary, we are free to choose it as desired. In fact, we have already done so due to the constraint that η_{ij} have a zero mean. This requires that $C \equiv 0$.

$$\langle \eta_{i1} \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} \eta_{i1}(a_1, a_2, \dots, a_5, \lambda_M) d\lambda_M = 0$$

Notice that we could have chosen C differently. If we did, our choice might not permit the average of the partial derivatives to vanish when we averaged Eq. (8-65). This would change the definition of the mean element rates, A_{ij} (for $j \geq 2$), and consequently the definition of the mean elements themselves. This is the cause of many different definitions of mean elements in the literature. It also explains why the mean elements are specific to a particular analytical theory and why the mean elements and short-periodic variations must be consistent so that their sum correctly approximates the osculating elements.

Several perturbations exist for real-world applications. Consequently, we must replace the single VOP perturbing function, F_i , in the preceding analysis by multiple functions.

$$\epsilon F_i \leftarrow \sum_{\alpha} v_{\alpha} F_{i\alpha} \quad (8-67)$$

The sum is taken over all the perturbations of interest. Each v_{α} is a small parameter associated with $F_{i\alpha}$. The mean element rates and short-periodic variations must also be generalized. Consult McClain (1992) or Danielson et al. (1994, 1995) for more details.

Most of the perturbations that affect satellites can be averaged, including effects of the central body, third-body point masses, atmospheric drag, and solar-radiation pressure. But they must be 2π periodic in λ_M . Averaging can be performed either analytically or numerically using quadratures. Numerical averaging is particularly appropriate for the non-conservative perturbations such as drag and solar-radiation pressure with shadow effects.

Some perturbations don't lend themselves to averaging. Remember that the disturbing potential (or force) must be 2π periodic in the variable we wish to eliminate. Perturbations we can't average include atmospheric drag with a tumbling asymmetrical spacecraft; solar-radiation pressure for satellites similar to those in the drag section; continuous thrust with rapid, nonperiodic changes in direction; and impulsive thrust. These perturbations are typical of powered flight, which, in general, isn't required to be either slowly varying or 2π -periodic in λ_M . Averaging methods can't predict the effects of these perturbations.

The averaging is significant because the high frequencies of the short-periodic terms drive step-size requirements for numerical techniques. To generate accurate ephemerides, we use small step sizes because of the high-frequency component of the motion. However, if we remove the short-periodic variations, we can numerically propagate the remaining equations of motion for *mean* elements [Eq. (8-63)] using much larger step sizes than with the equations of motion for *osculating* elements. These larger step sizes depend on the frequency of the long-periodic terms. For example, a typical low-altitude satellite using a conventional numerical routine (70×70 gravity field) might require an integration step size on the order of 30 seconds. With the semianalytical theory, the integration step size would be on the order of one day (86,400 seconds)! Then, we can independently construct the short-periodic contribution to the motion using analytical or numerical methods.

Writing the osculating equations of motion in terms of the *mean* elements, Eq. (8-61), doesn't imply that we've removed the short-periodic variations. For example, osculating equations of motion written in terms of mean elements still contain short-periodic variations; short-periodic variations are caused by the fast variable—even if it's a *mean* fast

variable. But the averaged equations of motion, which give the *mean* element rates, do imply we've removed the short-periodic variations.

Expansions of gravitational potentials for the central body and third body were originally developed by Slutsky (1982) and by Proulx et al. (1981). Danielson, Neta, and Early (1994, 1995) derive various first-order and second-order rates for the mean elements. They also discuss short-periodic variations (including atmospheric drag and solar-radiation pressure). Finally, they discuss truncation algorithms and numerical methods needed in semianalytical satellite theories.

To use this satellite theory in highly precise orbit determinations, we must have a way to compute the partial derivatives of the observations with respect to the solve-for parameters (see Sec. 9.5.1). This also applies to special and general perturbations. The solve-for vector includes the epoch values of the mean elements and other physical constants (for example, the coefficients of drag). By the chain rule, computing the partial derivatives reduces to computing the partial derivatives of the osculating elements at an arbitrary time with respect to the epoch values of the mean elements and the physical constants included in the solve-for vector. Mathematically, this matrix of partial derivatives is

$$\frac{\partial \hat{a}}{\partial \hat{a}_o} = [\mathbf{I} + \mathbf{B}_1] [\mathbf{B}_2: \mathbf{B}_3] + [0: \mathbf{B}_4]$$

\mathbf{B}_1 = matrix of partials of the short-periodic variations, $\epsilon\eta_i$, at time, t , with respect to the mean elements at time, t

\mathbf{B}_2 = matrix of partials of the mean elements at time, t , with respect to the mean elements at the epoch time

\mathbf{B}_3 = matrix of partials of the mean elements at time, t , with respect to the physical constants in the solve-for vector

\mathbf{B}_4 = matrix of partials of the short-periodic variations at time, t , with respect to the physical constants in the solve-for vector

Analogous to the semianalytical theory for the satellite's motion, the \mathbf{B}_2 and \mathbf{B}_3 matrices solve the variational differential equations for the averaged dynamics, which we can integrate numerically with a very large step size. The short-periodic partial derivatives can be computed using the same technique for computing the short-periodic variations, or by finite-difference methods.

8.9 Practical Considerations

We've examined many techniques to solve the perturbation problem. It's useful to look at some concepts that transcend the techniques of Chap. 7 and this chapter. In particular, we must address the issue of expected accuracy from any technique. Specific concerns about numerical techniques are in Chap. 7, but I'll extend the discussion in this section to examine how errors in propagation methods change over time. Finally, the availability of initial data is important to selecting a method for determining orbits.

8.9.1 Expected Accuracy of Propagation Techniques

As we design a satellite mission, success depends on certain accuracy from the orbit determination. This accuracy helps us develop an *error budget* which summarizes acceptable errors that won't sacrifice the mission. Determining an error budget for a particular satellite is a challenging task. Among our concerns are the availability of data and the *true* mission requirements. Often, requirements are written based on the current system because no other data exists, and rather than flag deficiencies, it's sometimes less confrontational to show compliance despite shortcomings.

A note about units is necessary for this discussion because several conventions exist. The data in Table 8-6 shows variations expressed in meters. The result could also have presented angular separations as seen from a site on the ground. Moreover, any angular discrepancies may be given in arcseconds, radians, or degrees. The angular approach may not be quite as intuitive, but it is independent of the satellite's altitude. Figure 8-14 shows the interrelationship among errors in linear units, angular units, and the satellite's altitude.

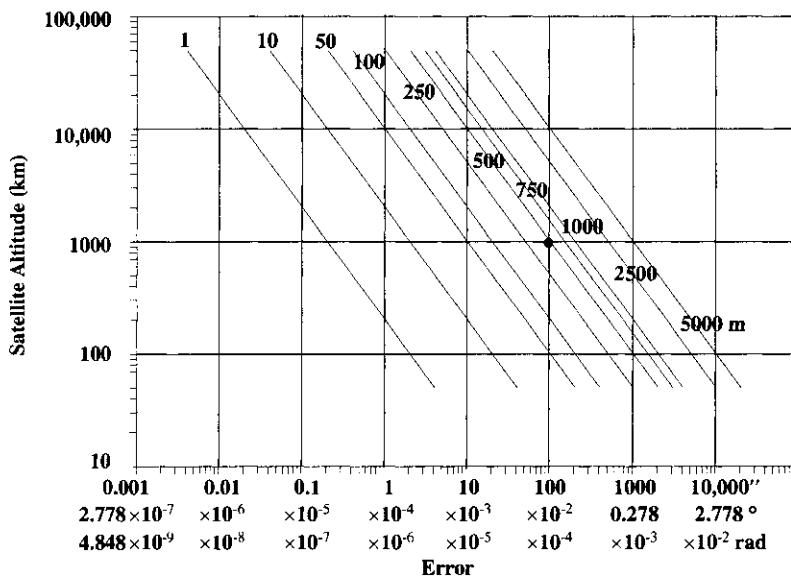


Figure 8-14. Error Values for Combinations of Altitude and Angular Separation. This chart helps us determine the error at the satellite's altitude for a given angular separation in arcseconds, degrees, or radians. For example, a 100" (0.0278° or 4.848×10^{-4} rad) angular error for a satellite at 1000 km is about 500 m.

In determining an orbit from observational data, we should always strive to match the propagator to the accuracy of the available data. Knowles (1995) shows a chart (see Fig. 8-15), which relates the potential accuracy of various computational techniques to sources of observational data in order to arrive at an expected error for the resulting system.

Remember, this chart compares the accuracy of existing routines, not the theoretical differences between numerical and analytical techniques. We'll discuss the relationship of the propagator and the processing of observations in Sec. 9.8.2.

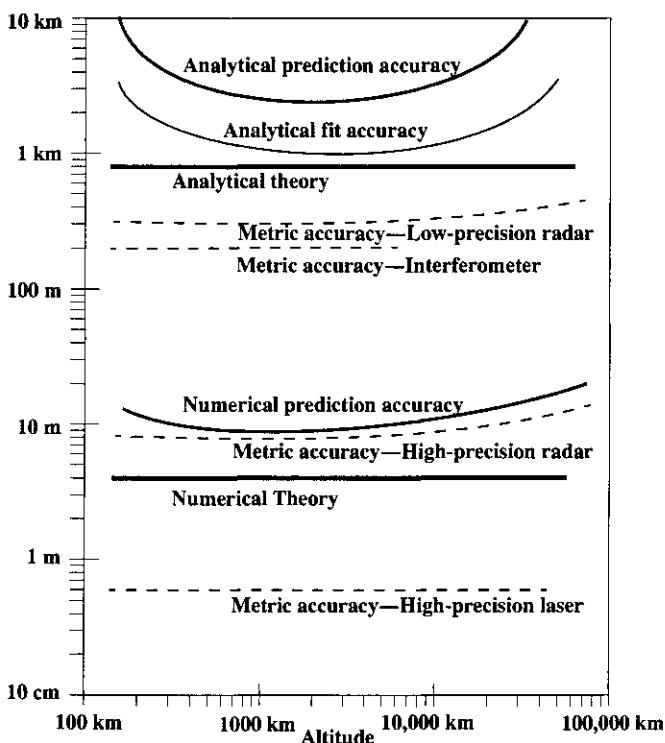


Figure 8-15. Errors in Determining a Satellite's Position. This figure (Knowles, 1995) shows general trends for typical accuracies of numerical and analytical theories (solid lines). Semianalytical theories range from nearly numerical to nearly analytical, depending on their force models. Observations (dashed lines) tend to degrade at higher altitudes. Notice how predictions and propagations degrade the theoretical accuracies, especially for analytical theories with significant drag and third-body effects.

From the discussions of differential-correction and propagation techniques in Chap. 9, we should recognize that processing actual data will result in an estimated state that's slightly less accurate than the potential accuracy of a theory. Subsequent propagations will further degrade the accuracy, as we'll see in Sec. 9.8. [See also Junkins et al., (1996).] To illustrate the use of Fig. 8-15, consider the design of a new space-based surveillance satellite. This type of satellite would require very accurate orbit determination because errors in the satellite's position will translate into errors in the sensor data we obtain. Thus, we may not be able to use existing analytical theories or less precise observations at all.

Finally, we must know what to expect from a certain propagation method. The errors from propagation methods are separate from the error growth discussed in Sec. 9.8. Propagation errors *add* to the covariance error growth. Thus, for this section, we'll assume the highly accurate propagation method is exact and contains no covariance errors. This topic is sensitive because each orbit has different characteristics and thus, like Table 8-6, is intended to provide only an intuitive feel for some propagation accuracies. We'll examine the radial, cross-track, and along-track errors from a highly accurate numerical method for five satellites: LACE, TOPEX, LAGEOS, Molniya, and a geosynchronous satellite. Satellite acronyms are in Appendix A, and they are introduced briefly in Sec. 10.2.

First, we must understand how the propagation is set up. It's very difficult to find consensus about absolute truth in this problem. Thus, I've used a "full-up" numerical-propagation routine implementing a 50×50 JGM-2 gravity field, a Jacchia-Roberts atmosphere with actual solar characteristics, the complete JPL DE-245 solar and lunar ephemerides, and solar-radiation pressure to calculate the reference orbits. The initial epoch for all simulations is January 18, 1991, at 09:36 UTC. This time corresponds to the occurrence of a solar maximum to enhance the effects of drag. For comparison, I've used the simplified numerical integrator introduced in Sec. 7.7.1 (J_2, J_3, J_4 , Sun, Moon, and drag). We usually examine the radial, along-track, and cross-track variations for each component of the position and velocity vectors and therefore use the RSW coordinate system. The root-sum-square (RSS) gives us an overall assessment of the positional differences. Figure 8-16 shows these variations for LACE, using the propagation scheme mentioned above.

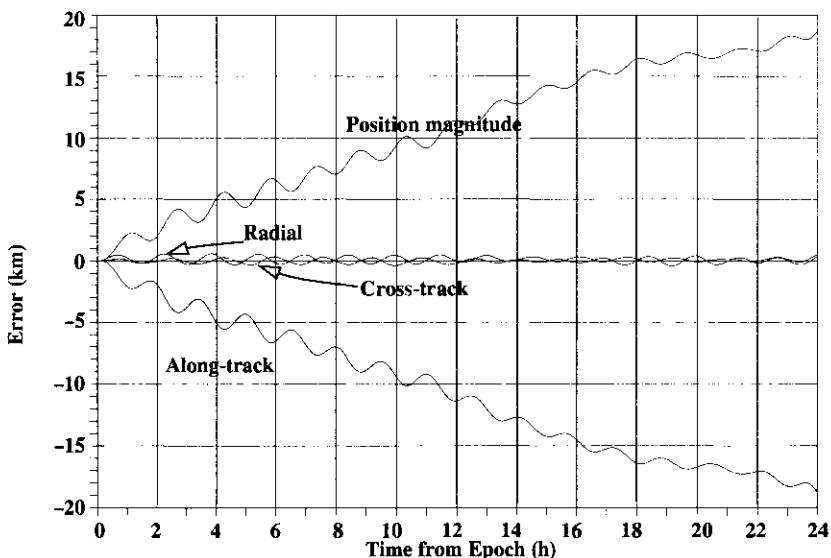


Figure 8-16. Comparison of Propagation Errors for LACE. This figure shows RSW components and RSS position differences for a simplified numerical propagator and force model truncation. The reference orbit uses a complete numerical-integration program. Atmospheric drag causes along-track deviations to be the greatest.

Because I chose 1991 and the satellite is relatively low in the atmosphere, the variations from drag are expected to be large due to the solar maximum. It's important to notice the large oscillatory variations as time advances. We can consider the "full-up" numerical propagation to be truth for this analysis. In actuality, it varies about 10 m in total RMS from the true orbit. A common misconception is that the error always grows at the same rate for limited theories. From Fig. 8-16 to Fig. 8-20, however, we'll see that performance of a simplified propagator varies considerably for different satellite orbits.

Next, we can examine the performance of a simplified propagator for a satellite that's a little higher, such as TOPEX in Fig. 8-17. TOPEX is at an altitude of about 1300 km. Notice that the errors are much smaller than those from the previous example. This is reasonable because drag is no longer affecting the orbit. Another interesting observation is that the largest errors now occur in the along-track and cross-track directions.

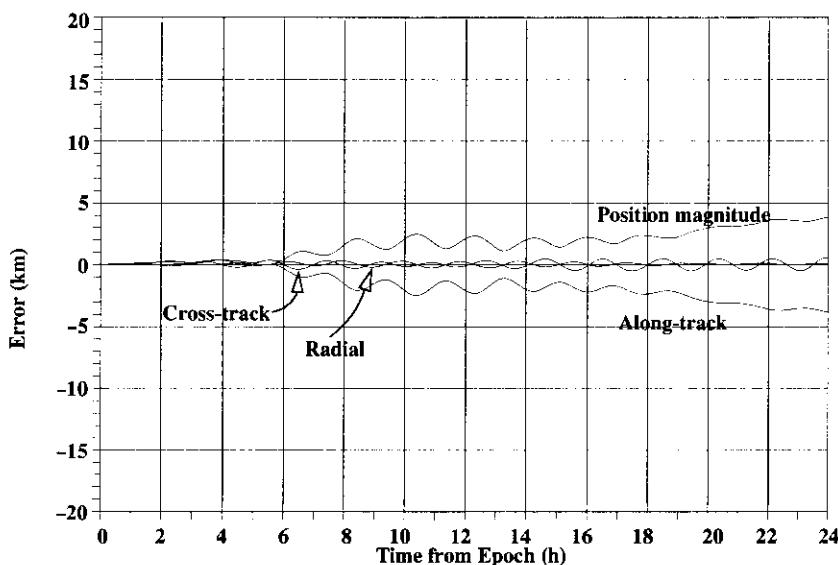


Figure 8-17. Comparison of Propagation Errors for TOPEX. This figure shows RSW components and RSS position differences for a simplified numerical propagator and force model truncation. The reference orbit uses a complete numerical-integration program. Without significant drag perturbations, the overall error is greatly reduced.

Figure 8-18 shows data for LAGEOS. This satellite orbits at an altitude of about 6000 km, and is used mainly for calibration and scientific research. Notice how well-behaved the orbit is—even a simplified numerical integrator can capture the behavior nicely. This is because the satellite is well above any significant drag regime, and the Earth's gravitational disturbances diminish with altitude due to the $(R_\oplus/a)^3$ term in the potential. Finally, the orbit is low enough so the third-body disturbances are small and approximately equal to the central-body non-spherical disturbances. The largest error is still in the radial compo-

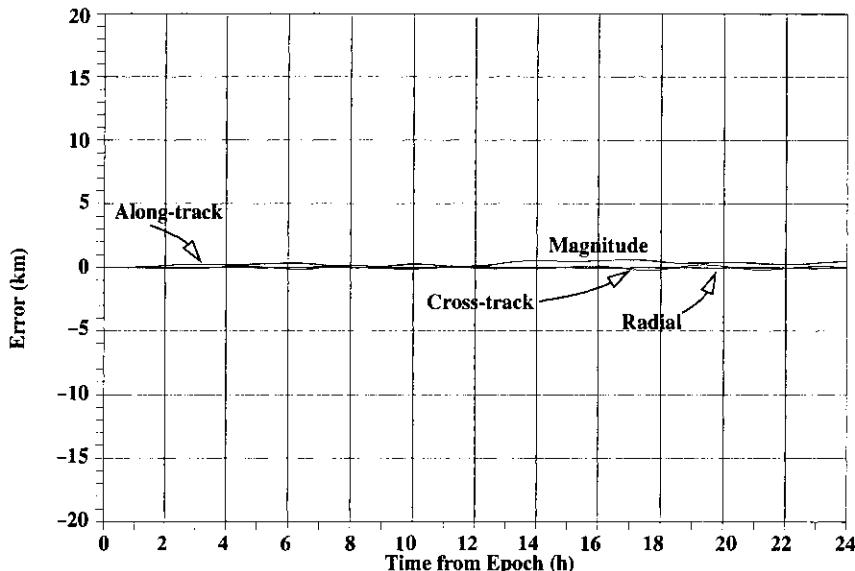


Figure 8-18. Comparison of Propagation Errors for LAGEOS. This figure shows RSW components and RSS position differences for a simplified numerical propagator and force model truncation. The reference orbit is generated using a complete numerical-integration program. The lack of significant perturbations makes LAGEOS a perfect candidate for calibration studies.

nent. Coincidentally, LAGEOS's corner-cube retro-reflectors make it desirable as a calibration satellite, and it is often used in satellite laser-ranging tests.

As shown in Fig. 8-19, the Molniya's high eccentricity and variable distance make it particularly difficult to model. The structure of the error graph is very different for this orbit. We now have a combination of drag acting at perigee and third-body forces acting at apogee. The satellite's position changes very rapidly at perigee, and we can see a pronounced signature of the 12-hour orbit. Each time the satellite passes through perigee, all three components differ significantly from the "truth" model. The effect is enhanced because we're using a simple integrator with a fixed step size, and the large variations result partly from this feature.

Lastly, we examine a geosynchronous orbit in Fig. 8-20. The behavior is more "regular" than in the Molniya case because the orbit is at high altitude and is nearly circular, but there are still large variations. Notice the slower rate of change for the variations. The effects of third-body perturbations are usually much slower (longer period) than those for drag and central-body variations on low-Earth orbits. Over periods of weeks and months, you would observe slowly varying motion in the orbital elements due to the m -monthly and m -yearly periodic variations and other related terms.

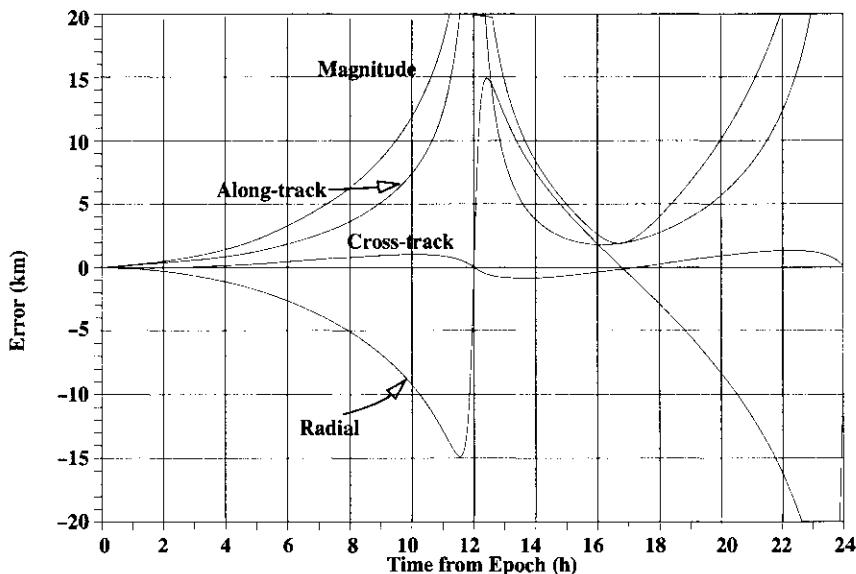


Figure 8-19. Comparison of Propagation Errors for Molniya. This figure shows RSW components and RSS position differences for a simplified numerical propagator and force model truncation. The reference orbit uses a complete numerical-integration program. The fast velocity at perigee shows up on each 12-hour orbit as a large variation.

8.9.2 Initial Conditions, Sources, and Conversions

Initial conditions are just as important as the physical data; however, they usually receive much less attention in the literature. First and foremost, any simulation must have an initial state vector to use as input data. Of course, the exact form of this state vector varies widely, but generally, we may view it as being either position and velocity vectors in some coordinate system or one of the sets of orbital elements. Often, the data also includes satellite and dynamic parameters (mass, area, c_D , c_R).

The main sources for initial conditions of orbiting satellites are the state vectors that satellite owner-operators provide and the U.S. Space Command's 2-line elements sets. Appendix D.5 lists some addresses. Many highly-accurate observations and ephemerides are available on the Internet. Be very careful when using any element set to ensure you don't misuse time (*UTC*, *TAI*, *TDB*, ...); coordinate systems (*B1950*, *J2000*, ...); theories (Kozai, Brouwer, numerical, ...); and assumptions (mean or osculating, two-body, ...). Everything should be consistent!

Confusion often arises concerning the theories. Unfortunately, much operational code was developed with little documentation and few configuration controls. As a result, some theories carry the same name but have different formulations. Many analysts try to use elements consistent with one theory in another, or they attempt a conversion—and that's the problem: *transforming the given elements in one theory to those consistent with*

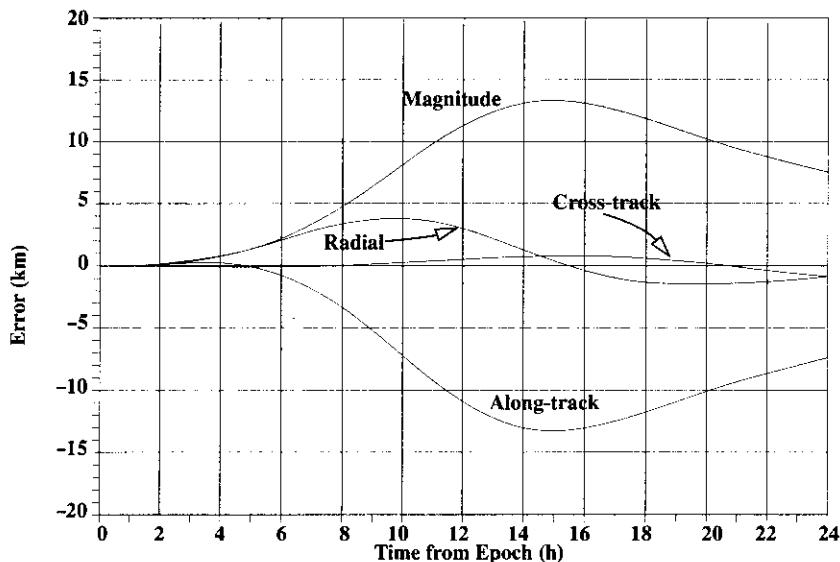


Figure 8-20. Comparison of Propagation Errors for a Geosynchronous Satellite. I've shown RSW components and RSS position differences for a simplified numerical propagator with force model truncation. The reference orbit is generated using a complete numerical-integration program. Third-body effects dominate the motion here, and the effects tend to have longer periods.

another theory, and vice versa. To convert correctly, we must understand clearly the definitions and assumptions for each element set and develop a consistent transformation method. Unfortunately, you won't always find this level of detail.

A related area of confusion arises between two-body elements, mean elements, and osculating elements. We often consider these three to be equal, but there are some subtle differences. **Two-body elements** usually are those derived from and used with two-body equations of motion. As such, no perturbations influence or change the orbital elements. **Osculating elements**, on the other hand, are the *instantaneous* elements for a satellite under the influence of perturbations. **Mean elements** result from averaging the effect of the perturbations over a specified interval of time. The confusion arises because the osculating elements describe the perturbed problem and are defined for a particular instant of time, but they're often used with simplified (two-body, for instance) propagation schemes. Such use nullifies assumptions inherent to the osculating elements. Of course, we can use osculating elements in this manner, but we have to be aware that comparisons with future (or past) perturbed elements are of limited accuracy and eventually aren't meaningful. Mean elements suffer a similar confusion. Thus, we should always try to identify the correct form of an orbital-element set.

There are numerous conversions: osculating vectors, 2-line, Kozai, and Brouwer mean elements, and so on. A commonly available data set is the 2-line element set discussed

previously (Sec. 2.4.2). For the casual user, these element sets work well with methods described in Hoots and Roehrich (1980) and are an easy way to simulate many satellites with some precision. As an aside, an alternative appears in the work of Cefola, Fonte, and Shah (1996), who have enhanced the U.S. Navy's analytical theory. This theory simulates some of the effects of drag by using the mean motion rates or the ballistic coefficient, as well as some third-body and periodic and resonant central-body effects.

For highly accurate solutions, two problems arise. First, it's usually difficult to obtain timely element sets—datasets are often days to weeks old. Remember, mean elements capture only the smoothed time-varying effects (free of high frequencies contributions), while osculating elements capture all time-varying effects. When the epoch of the element set and the simulation time are separated by a finite time (the length depends on the orbit), and when we use a limited analytical theory, large errors (hundreds of kilometers in position) are likely for many orbits. The second problem involves the theory used. The 2-line element sets are *usually* determined by the U.S. Space Command using a simplified version of Brouwer's theory, but they're published with parameters consistent with Kozai's theory. Besides the obvious disconnect in using two theories, the differences in these theories require certain conversions. An undesirable side effect of many conversions is loss of accuracy because we often have to iterate for each conversion.

Converting Elements—General

As mentioned earlier, the main reason for converting between osculating and mean elements is to be consistent between theory and initial conditions of each algorithm. Analytical theories typically use *mean* elements because they primarily model the dominant secular variations and long-periodic terms. Most short-periodic terms are small and are usually ignored. Depending on the orbit, this simplification may result in elements which differ substantially from the satellite's instantaneous state vector. On the other hand, numerical methods use osculating values because they are consistent with the actual equations of motion. Consequently, they capture all contributions to the motion (secular, long-, and short- periodic). I'm not trying to confuse you, but a subtlety does exist. Numerical techniques require precise inputs to generate precise outputs. On the other hand, analytical theories require data consistent with that theory. Because the short-periodic variations aren't usually included in analytical theories, using osculating elements can actually produce worse results than using mean elements consistent with the theory!

Let's consider a simple example. J_2 causes an 8–10 km short-periodic variation in the semi-major axis for low-Earth satellites. Thus, if the osculating a is used in place of a mean a , the assumed a could be off by almost 10 km. This causes an error in the satellite's mean motion, which produces a secular drift in the calculated in-plane position specified by the mean anomaly. The drift, to first order, can grow rapidly. It's given by

$$\Delta M = -\frac{3n}{2a} \Delta a \Delta t$$

The conversion from *mean* to *osculating* elements is well defined [Eq. (8-61)] for complete analytical theories with short periodic variations. Unfortunately, it's more difficult to find mean elements when osculating values are known. In addition, converting between mean element theories can introduce additional problems. See Der (1996) for additional information. I'll briefly discuss both cases.

Osculating to Mean Element Conversions

In Chap. 7 we saw that numerical techniques are becoming more popular. Thus, it's important that we understand how to convert an osculating state vector to a mean element set. This situation usually arises because we need the speed of a semianalytical technique to perform accurate and long term studies. I'll introduce two options.

A *single-point conversion* solves Eq. (8-61) for the mean elements.

$$a_i = \hat{a}_i - \epsilon \eta_{ij} (a_1, a_2, \dots, a_5, \lambda_M)$$

Although the short-periodic variation has mean elements as arguments, we can use the osculating elements instead. This introduces an error of $O(\epsilon)$ in evaluating η_{ij} . However, the resulting error in the mean elements is only $O(\epsilon^2)$ because of the product $\epsilon \eta_{ij}$, and because we have already ignored errors of second order in Eq. (8-61). Usually, we use this first estimate to re-evaluate the short-periodic variation and again subtract from the osculating elements to refine our answer. Additional iterations beyond the first few rarely provide additional accuracy. We can implement this process with higher order contributions if the formula for η_{ij} are available. This process is often adequate for general use.

The *precise conversion of elements* uses a least-squares differential-correction to estimate the mean elements (See Chap. 9 for details on estimation). A high precision (numerical) propagator generates an ephemeris for a day or two. The vectors are treated as observations in the least squares process. For semianalytical techniques, this requires a mean element propagator and short-periodic models. The differences (residuals) between the numerical and semianalytical or analytical ephemerides are analyzed to obtain corrections to the initial mean elements. We iterate this process until the changes become small. The initial estimate may come from the single point conversion described above, by averaging, brute force, or by numerical quadrature, for the osculating elements over a single revolution. The method adjusts the initial mean elements to partially compensate for the force model truncation in the analytical theory. This *can* be a benefit, but it may also yield undesired effects.

Mean Element to Mean Element Conversions

Ideally, mean elements in one theory should be converted to approximate osculating elements which are then converted to mean elements in the new theory using the *single point conversion* mentioned previously. This requires short-periodic variations for each analytical theory, preferably consistent in terms of the order of ϵ . In some cases, these variations don't exist to the same order, or the theories are formulated in different elements sets (e.g. classical vs. Delaunay). Special conversion formula can be developed, however, there is

significant potential for introducing errors in these “ad hoc” processes. For instance, conversion between element sets should never be applied to mean elements alone. Rather, they should be applied to the osculating equivalent of the mean and short-periodic variations. For example, if we try to convert *mean* classical orbital elements to *mean* equinoctial elements using Eq. (2-62), we introduce a small non-linear bias in the answer by neglecting the short-periodic variations in the transformation. You can see this by expanding an element (e.g. $e^2 = h_e^2 + k_e^2$) and adding a small delta to each mean element. The resulting “mean elements” aren’t “mean” in the same *sense* as the original elements.

The *precise conversion of elements* is similar to the osculating to mean conversion, but we replace the numerically propagated ephemeris with one from an analytical theory: mean elements and the short-periodic variations. These values are processed with the second analytical model using a batch least-squares estimation process to find the best estimate of the initial mean elements for the second theory. We must still iterate.

8.10 Summary of Perturbation Effects

Let’s combine all the results so we can determine major categories and effects for certain classes of orbits. Figure 8-21 shows several types of orbits.

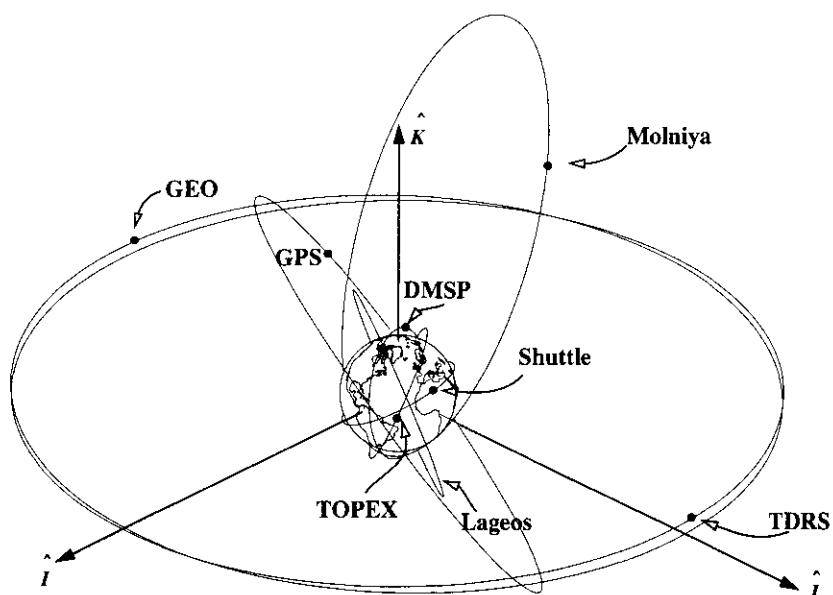


Figure 8-21. Representative Satellite Orbits. This figure shows various satellite orbits to the correct scale. The GEO orbit is exactly equatorial, whereas the TDRS orbit is inclined about 2° . Note I've shown only one GPS orbit plane. In reality, 24 satellites are deployed in six planes. Notice the extreme height of apogee for the Molniya orbit and its location over the Soviet Union.

First, we must categorize the types of orbits so we can accurately evaluate the perturbing forces. Next, it's useful to specify which elements (of any orbit) are affected by certain types of perturbations. Then, in a final table, we'll combine both previous tables and show the dominant forces on each class of satellite.

I want to be sure we cover most practical orbit types as listed in Table 8-4. I've included the space station Mir with other LEO satellites because the two satellites treat perturbations differently. For Mir, satellite reboost maneuvers occur almost continually to correct the station's orbit. LEO satellites usually drift until the error becomes unacceptable for a particular mission. Notice I've included some categories that aren't in common orbits today because future technology may make these less popular classes acceptable.

TABLE 8-4. Typical Satellite Classes. This table shows some orbits I've used throughout this book. I selected TDRS as a geosynchronous example. The Molniya, Prognoz, and Ellipso orbits are the only eccentric orbits I've shown because there's too much variety when looking at all possible combinations. Ellipso and ICO are future systems.

Type	Altitude Apogee (km)	Altitude Perigee (km)	e	i (°)	n (rev/day)	Catalog Number
LEO	286	205	0.0062	varies	16.104 638	varies
Shuttle	300	250	0.0010	28.500	16.000 000	varies
GFZ	385	376	0.0006	51.652	15.623 923	23558
Mir	390	376	0.0010	51.651	15.615 284	16609
LACE	467	450	0.0012	43.103	15.357 473	20496
Ofeq	670	364	0.0221	143.367	15.162 636	23549
Hubble Space Telescope	599	590	0.0006	28.469	14.910 760	20580
Landsat	705	700	0.0004	98.329	14.570 540	14780
DMSP	811	800	0.0008	98.622	14.257 934	13736
TOPEX	1343	1331	0.0008	66.039	12.809 310	22076
Vanguard 2	3042	555	0.1520	32.880	11.740 821	11
Ellipso	7835	526	0.3461	116.565	8.001 448	
LAGEOS I	5946	5840	0.0043	109.850	6.386 641	8820
ICO	10,360	10,349	0.0003	45.000	4.010 902	
GPS	20,230	20,129	0.0019	55.341	2.005 802	19802
Molniya	39,869	353	0.7459	63.069	2.013 683	13964
Geostationary, TDRS-6	35,805	35,766	0.0005	0.004	1.002 746	22314
GEO - GSTAR-3	35,813	35,760	0.0006	7.386	1.002 697	19483
LES-8	35,845	35,725	0.0014	15.143	1.002 754	8746
Prognoz	194,208	7273	0.8726	28.730	0.247 629	20413

Next, we examine the orbital elements and which perturbing forces affect their values, shown in Table 8-5. Notice I've divided the perturbations only into secular, short-periodic, and long-periodic effects.

TABLE 8-5. Summary of Perturbation Effects. This table shows the types of effects for the perturbing forces on the orbital elements. General effects are also listed.

Element	All zonal	Even zonals	Sectorial Tesserai	Drag	Third body	Solar Radiation
a	Periodic	Periodic	Periodic	Secular & periodic	Periodic	Periodic
e	Periodic	Periodic	Periodic	Secular & periodic	Periodic	Periodic
i	Periodic		Periodic	Secular & periodic	Periodic	Periodic
Ω	Periodic	Secular	Periodic	Periodic	Secular & periodic	Secular & periodic
ω	Periodic	Secular	Periodic	Periodic	Secular & periodic	Secular & periodic
M	Secular & periodic	Secular & periodic	Periodic	Periodic	Secular & periodic	Secular & periodic
GENERAL:				<ul style="list-style-type: none"> - Periodic changes result from atmospheric density differences and a rotating atmosphere - r_p almost constant except for coupling with third-body - Changes in value of c_D can cause large changes in the satellite orbit - Satellite attitude is important in finding cross-sectional area, A - Predictions become very difficult near reentry due to changes in A, c_D, and m - Low mass, large area yields greater effect 		
ZONAL:				<ul style="list-style-type: none"> - Ratio of secular effects Moon/Sun ~2.2 - Orientation of perigee and node important for periodic effects - No secular, LP, or m-monthly variations in a - Orbital planes precesses about Earth's polar axis (for LEO), mean pole between Earth and ecliptic pole for deep space - For drag perturbed orbits, satellite lifetime can be effected - i drifts to about 15° in 27 years for geosynchronous satellites 		
SECTORIAL/TESSERAL:				<ul style="list-style-type: none"> - Entry/exit from Earth shadow causes complexity in defining motion - Perigee height changes can affect satellite lifetime - Some variations can be up to a year - Evaluating satellite attitude is important to finding exposed area to the Sun, A_{SR} - Low mass, large area yield a greater effect 		
Resonance - Commensurability between satellite period and Earth rotation. Shallow (days), Deep (years). Two types,						
- Repeat groundtrack, $q = 0$ and $l - 2p \neq 0$						
- Anomalistic, $q \neq 0$ and $l - 2p = 0$						
-- Even tesserai only						
-- Stable GEO, $\lambda = 75.3^\circ, 255.3^\circ$						
-- Unstable GEO, $\lambda = -14.7^\circ, 165.3^\circ$						

Finally, let's combine the previous two tables to list the dominant perturbing forces on a satellite, based on its orbit. Table 8-6 shows the result.

TABLE 8-6. Dominant Perturbation Effects on Satellite Orbits. This table shows some of the classes of satellite orbits used throughout this book and identifies *representative* values due to the main sources of perturbations. Periodic effects show maximum amplitudes, whereas secular variations depict changes per day. Some periodic calculations are from Cefola (1997) using DSST. These numbers are intended only as a rough guide—actual numbers can vary significantly depending on the satellite's orientation, and the specific orbital elements.

Orbit Class	Central Body		Drag		Third Body		Solar Radiation
Secular Effects	Ω °/day	ω °/day	a m/day	e /day	Ω °/day	ω °/day	m/day
LEO	-5.7	6.5	5000				
Shuttle	-5	5					
Mir	-5.0	3.8					
Landsat	1	-3.1		100			
DMSP	1	-2.9					
TOPEX	-2.1	-0.5					
LAGEOS I	0.3	-0.2					
ICO	-0.1	0.1					
GPS	~	~					
Molniya	-0.2		100				
GEO							
Periodic Effects	SP m	m-daily m	lin com m	Res m			
LEO							
Shuttle	7000	470	50	Deep	33	0.5	~
Mir	6000	380	46		3	0.3	~
Landsat	9100	610	36		~	~	~
DMSP	9000	590	134		~	~	~
TOPEX	7200	445	334		~	1	~
LAGEOS I	4800	115	18			0.1	~
ICO	2000	40	23			10	10
GPS	1700	17	18	Deep		100	12
Molniya	32,000	105	900	Deep	50	250	10
GEO	1600	~	16	Deep		716	50

Problems

1. Herrick (1972, Chapters 16 and 17) introduces a grave symbol (e.g., \ddot{x}) to determine the Lagrange planetary equations of motion. Form equations of motion with this technique for a simple harmonic oscillator ($\ddot{x} + w^2x = 0$) and for the orbital problem using equinoctial elements.
2. In Chaps. 7 and 8, we discussed two forms of the potential function; classical and Kaula's. List advantages and disadvantages in using each for (a) a complete, general-purpose propagation program, (b) analysis of specific gravitational effects, and (c) a simplified classroom analysis tool.
3. You've been assigned as a new Branch Chief for Satellite Modeling & Simulation. Your first task is to show your management the locations of all GPS, Molniya, TDRS, and DMSP satellites for one month. What propagation scheme should you use and why?
4. What are the J_2 secular rates of change for Ω and ω for a satellite in a 500 km, circular, equatorial orbit? Does the inclination make a difference? Why? Is your answer different for an eccentric orbit ($e = 0.2$)? Provide a mathematical basis for your answers.
5. You're doing a first-order analysis on a new satellite in an elliptical ($e = 0.2$) orbit at 700 km altitude. Can you design the orbit so no maneuvers are necessary to maintain it? Hint: consider secular J_2 perturbations only—can the effects of $\dot{\Omega}$ counteract $\dot{\omega}$? If it can't be done at 700 km, is another altitude feasible?
6. Including perturbations, what's the uncertainty in problem 5 in Chap. 4?
7. You're at work with a telecommunications company. You have almost 900 satellites in circular orbits at an altitude of about 700 km. Your boss comes in and says he wants to know where all the satellites are to within 50 m. You have dense observations (200 per pass) from several highly accurate radars, can access SLR data, and have new well-coded, documented, and validated algorithms for numerical differential correction and for the propagator. Can you proceed? (Hint: review estimation techniques in Chap. 9)
8. Find analytical expressions for the rates of change in the orbital elements due to the influence of changes in velocity.
9. Find expressions for the long-periodic rates of change for orbital elements using Kozai or Brouwer. Can any terms dominate the rate of change? If so, what are they? If not, why not?
10. Do a conversion of a 2-line element set from mean to osculating values. Are these elements useful? Explain. [Hint: research Hoots and Roehrich (1980), Fieger (1987), Herriges (1988), Dyar (1993), Cefola, Fonte, and Shah (1996), Fonte (1993), and Der (1996)].

CHAPTER 9 STATISTICAL ORBIT DETERMINATION

- 9.1 Historical Background
- 9.2 Introduction to Sensor Systems
- 9.3 Linear Least Squares
- 9.4 Nonlinear Least Squares
- 9.5 Application: Orbit Determination With Differential Correction
- 9.6 Sequential-Batch Least Squares
- 9.7 Kalman Filtering
- 9.8 Practical Considerations

9.1 Historical Background

To this point, most of our discussion has centered on simple, two-body determination of a satellite's orbit. But we need a way to account for the real-world anomalies during subsequent processing. Actually determining the orbit for a satellite (or, historically, a planet), presents great challenges.

The evolution of statistical orbit determination began hundreds of years ago with Kepler (c. 1610) and Legendre (c. 1750). Gauss (c. 1810) gave it a firm analytical and computational basis. Many improvements and innovations took place between these original foundations and current theories, but the fundamental principles are the same.

Much is made of the rediscovery of the asteroid Ceres in 1801. This event was important because it was the first time astronomers had used observations to formulate an orbit and predict a future position. More importantly, Ceres appeared where it was supposed to—within certain limits, of course.

The events leading to this discovery are somewhat clouded but still insightful. Sorenson (1970, 63) and Bell (1965, 259) suggest that Gauss originally discovered, or invented, the process of least squares in 1795. Unfortunately, Gauss didn't publish his results. Adrian Marie Legendre (1752–1833), working independently, formulated his own results in 1806 in a work titled *Nouvelles méthodes pour la détermination des orbites des comètes* (*New Methods for the Determination of the Orbits of Comets*). Although this work didn't completely describe the process, when Karl Friedrich Gauss (1777–1855) published his work *Theoria Motus Corporum Coelestium* (*Theory of Motion of Heavenly Bodies about the Sun in Conic Sections*) in 1809, Legendre was upset. He wrote a letter to Gauss and essentially scolded him for not doing a better literature search! Although we often think of

Gauss as being the root of many important mathematical discoveries, he was genuinely concerned about Legendre's letter. Still, over the centuries, Gauss has been credited with discovering the least-squares method—with some help from Legendre.

Gauss's method closely resembles modern ones. Sorenson (1970, 64) discusses the individual assumptions and observations Gauss used in forming his theory. He also quotes from *Theoria Motus* ([1809] 1963, 260): "the most probable value of the unknown quantities will be that in which the sum of the squares of the differences between the actually observed and computed values multiplied by numbers that measure the degree of precision is a minimum."

It's a true credit to Gauss's genius that his method enjoys such wide popularity and use. The ensuing three centuries produced many useful results, but they were all adjustments to, or extensions of, his basic premises.

An important discovery in statistical estimation methods came about during the 20th century. In 1912, Sir Ronald A. Fisher (1890–1962) introduced the concept of the **maximum-likelihood method**, extending Gauss's principle to cover non-Gaussian error statistics. Although not a complete theory, it provided the stimulus for work by Andrei N. Kolmogorov (1903–1987) and Norbert Weiner (1894–1964) in 1941 and 1942, respectively. According to Sorenson, their independent methods began the next advancement in estimation theories.

The next major event occurred in 1958 when Peter Swerling (1929–) published a RAND corporation report discussing a new recursive algorithm for statistical orbit determination. Sorenson (1970, 66) notes that the same controversy which surrounded Gauss and Legendre erupted in 1960 when Rudolf E. Kalman (1930–) published his landmark work, *A New Approach to Linear Filtering and Prediction Problems*. Although Swerling's paper contained the same equations with minor changes, at first, no one made anything of their similarities. But later, Kalman filtering became more popular, and many studies tried to improve and extend the original concepts. In 1963, Swerling wrote to the *AIAA Journal* and complained as Legendre had done. To date, the method retains Kalman's name. Although the recursive formulations of Swerling and Kalman are very popular today, they are really only a variation on the fundamental contributions of Gauss and Legendre.

This historical sketch isn't complete without discussing some of the technology required to use Gauss's theory, Kalman filters, and so forth. Without improvements in observational instruments, many of the techniques and methods in this chapter would be of little use. If, around 1800, Gauss had been trying to estimate the orbit for a satellite in low-Earth orbit instead of a distant minor planet, he may never have published his results. The observations wouldn't have been accurate enough to support his calculations. Fortunately, Ceres proved to be an appropriate subject for the observational techniques of the time.

The United States' first modern event in space surveillance occurred with Sputnik I when the Millstone Hill radar detected its signals. The Doppler method to determine the range rate of Sputnik [credited to Richard Anderle (1926–) of the Naval Surface Weapons Center] also permitted a refined estimate for J_2 just two weeks after launch. Thus, Anderle began modern orbit determination and established an important technique to define the

Earth's gravitational field. The U.S.'s first surveillance *tracking* of a satellite also occurred with the Millstone Hill radar when it successfully tracked Sputnik II. Over the next few decades, many types of radars and sensor devices were designed, mainly to detect ballistic missiles. During the Cold War, apprehension over a possible undetected first strike made early warning a primary mission. Even today, the ranges and sensor limitations of some tracking systems make them better suited for missile detection than for space surveillance. Still, the designers of the original systems had enough foresight to over-design the sensors, so they continue to work remarkably well for both missions.

Highly accurate orbit determination from observations remains important today. The use of laser technology for tracking satellites, resupplying power, and communicating demands more accuracy. Each satellite acquisition requires highly accurate ephemerides to locate the satellite because the divergence of the laser is often very small. Improved ephemerides can reduce and often eliminate costly search time.

This chapter introduces statistical orbit determination within real-world constraints. An important issue is our ability to predict the long-term evolution of an object's position and velocity, including realistic measures of uncertainty. Because objects as small as flecks of paint have damaged the Space Shuttle's windows, accurately determining their orbits is increasingly important. In part, this process requires estimation. Estimation is intimately tied to *initial orbit determination* (Chap. 6), *prediction* (Chap. 8), and *uncertainty estimates*, which we'll discuss here. We'll start by introducing terms and operating characteristics for sensor systems.

9.2 Introduction to Sensor Systems

Operational planners depend on sensor systems to support mission analysis and design, and ultimately, mission success. In this chapter, I'll describe sensors used to observe objects *in space*, including space-based sensors observing other satellites. Chapter 10 discusses the opposite problem of observing the Earth (space-based *surveillance* and *reconnaissance*). We'll define ground-based *space surveillance* as the general practice of determining the locations of objects in space.

A sensor may not be able to measure directly the quantity we want to observe; thus, we must know the technical routines described in previous chapters to analyze real-world systems. Figure 9-1 shows the basic geometry for a sensor site. Of course, if we use the topocentric (SEZ) coordinate system, a "site" can be either ground-based or space-based.

It's important to differentiate types of sensors, so let's extend Muolo's (1993, 97–103) definitions of sensors for the U.S. Space Command. In terms of usage, three distinct categories exist: dedicated, collateral, and contributing. Large organizations use *dedicated sensors* for primary missions, such as space surveillance, or geodetic and scientific research. *Collateral sensors* have other primary missions, but we may still use them to support space surveillance. For example, the U.S. Space Command's ballistic-missile-warning sensors routinely track low-Earth satellites. Lastly, *contributing sensors* mainly are those not under a primary organization's control. Most university, commercial, and

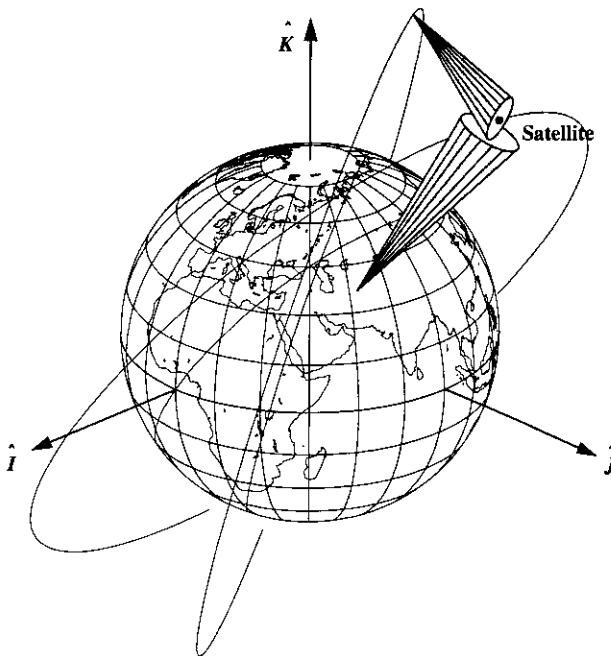


Figure 9-1. Sensor Geometry. This figure shows the geometry for acquiring a satellite. The ground-based sensor can't observe the satellite when it's below the local horizon. The “site” in orbit doesn't experience the same limitations, but it does allow the existence of negative (below the horizon) elevations.

other sensors are in this category. It's possible for dedicated sensors from one organization to function as collateral or contributing sensors for another organization.

We also need a **tracking system** or **network**, which uses globally distributed sensors to acquire data on satellites of interest for processing into ephemerides. This system is necessary because the techniques in Chap. 6 aren't adequate to calculate sufficiently accurate orbits needed by most applications; we require additional observations to improve existing orbits. This chapter discusses several mathematical techniques for combining new observations with earlier orbit solutions to arrive at improved orbit estimates.

Several networks support various missions. I'll describe a few of them. One network to consider is the **Space Surveillance Network** (SSN), which developed from ballistic-missile warning assets. It has become increasingly important in determining the location of active and passive objects in space. In 1996, Space Command was tracking more than 9000 objects. Many people believe this number could grow to well over 50,000 as they locate and catalog smaller particles of debris. Safety is the main reason to increase the number of objects being tracked—debris may hit a spacecraft and destroy a mission.

The former Soviet Union has a large network, the **Russian Space Surveillance System** (RSSS). It's designed to operate efficiently with geographically distributed observations. Khutorovsky (1996) describes the operation of the sensors. The RSSS contains radar, phased array, and optical sensors.

The **Air Force Satellite Control Network** (AFSCN) is made up of sensors which contribute to space support. **Space support** consists of command and control functions for the satellite and on-orbit support to satellite operations. Most of the AFSCN's sensors are dedicated because each system has specific requirements.

The **Deep Space Network** (DSN), managed by the Jet Propulsion Laboratory, locates objects *far away* from the Earth. NASA is the primary user of this system for interplanetary probes. The term *far* is obviously relative. Recall the definition of deep space from Chap. 2, which places these objects near geosynchronous orbit. The DSN usually handles satellites outside the Earth's sphere of influence, so the definition varies. Three ground stations handle operational activities: the Goldstone complex in Southern California, the Spanish complex in Madrid, Spain, and the Australian complex near Canberra.

NASA uses a network of almost 50 dedicated and contributing stations known as the SLR Network. Each of the **Satellite Laser Ranging** (SLR) sites very accurately measures a satellite's range. These sites typically track satellites which have corner-cube retro-reflectors that reflect the laser back in the same direction from which it was received. The Laser Geodynamics Satellites (LAGEOS) help calibrate the system.

The French space agency, Centre National d'Études Spatiales (CNES) has developed the DORIS system to precisely track and determine orbits for the TOPEX/Poseidon, SPOT-2, and other satellites requiring high accuracy. **DORIS** is a one-way Doppler system in which the satellite receives signals from several of the 50 CNES ground stations. These stations are in addition to the SLR sites. Master stations are in Toulouse, France, and Kourou, French Guiana.

Table 9-1 lists some of the categories of sensors and the sites that handle them. NAVSPASUR is a network of nine stations which use radar interferometry to set up a "fence." Any satellite passing through the fence triggers detection. The system's output is direction cosines that are usually transformed to azimuth and elevation. Additional processing will sometimes yield a crude estimate of range, but it's usually two to three times less accurate than the angular data. Originally, Baker-Nunn cameras used very precise optics to photograph satellites. Comparing the resulting time-tagged photographs to star catalogs yielded highly accurate angular measurements. Ground-based Electro-Optical Deep Space Surveillance (GEODSS) sites have replaced the older system. They work on the same principle, except the system uses charge-coupled device cameras to permit digital image processing. The output angles are right ascension and declination.

The actual visibility of a sensor site may be limited in effective range, azimuth, or elevation because of mechanical (design), geographical, and political constraints. The U.S. Air Force Academy is an example of a location which is limited in azimuth and elevation due to the high mountain range to the west. Sensors' range limitations are extremely important because satellites can easily be too far away from the site. Tracking systems can search only a limited amount of space. Many of the resources serve specific missions

TABLE 9-1. Representative Space Sensors Grouped by Usage. Data is combined from Taff (1985, 132–135), Muolo (1993, 97–103), Vallado (1989, App. A) and *Almanac* (annual, App. J). Abbreviations are as follows: Ground-based Electro-Optical Deep Space Surveillance (GEODSS), Navy Space Surveillance (NAVSPASUR), Maui Optical Tracking and Identification Facility (MOTIF), Advanced Research Projects Agency (ARPA) - Lincoln C-Band Observables Radar (ALCOR), ARPA Long-Range Tracking and Instrumentation Radar (ALTAIR), and Air Force Maui Optical Station (AMOS).

Site	Type	Latitude (°)	Longitude (°)	Altitude (m)	Sensor #
SSN Dedicated Sensors					
Socorro, NM	GEODSS	33.82	-106.66	1510.2	210
Maui, HI	GEODSS	20.71	-156.26	3058.6	230
Diego-Garcia, Indian	GEODSS	-7.41	72.45	-61.2	240
Eglin, FL	Phased Array	30.57	-86.21	34.7	399
NAVSPASUR, 9 sites	Interferometer	~33.0	-87 to -117	10 to 1400	741–749
MOTIF, HI	Optical	20.71	-156.26	3059.0	951
SSN Collateral Sensors					
Fylingdales, UK	Phased Array	54.37	-0.67	338.9	344–346
Pirincli, Turkey	Radar	37.91	39.99	890.0	337
Ascension, Atlantic	Radar	-7.91	-14.40	56.1	354
Clear, AK	Radar	64.29	-149.19	213.3	359
Antigua, West Indies	Radar	17.14	-61.79	0.5	363
Cape Cod, MA	Phased Array	41.75	-70.54	80.3	386, 387
Beale, CA	Phased Array	39.14	-121.35	115.7	388, 389
Thule, Greenland	Phased Array	76.57	-68.30	424.7	394, 395
Cavalier, ND	Phased Array	48.72	-97.90	347.3	396
Kaena Point, HI	Radar	21.57	-158.27	300.2	932
SSN Contributing Sensors					
Kwajalein Atoll	Radar - ALCOR,	9.39	167.48	62.7	333
	Radar - ALTAIR	9.39	167.48	62.7	334
Millstone, MA	Radar	42.62	-71.49	123.1	369
Haystack, MA	Radar	42.62	-71.49	115.7	372
AMOS, HI	Optical	20.71	-156.26	3058.2	952
DSN Sensors					
Goldstone, CA		35.39	-116.848	1036.0	
		40.41	-3.685	670.0	
		-35.32	149.008	767.0	

(launch corridors, missile warning, and so on) and are greatly constrained; thus, all the sensors combined can't search the entire volume in which satellites operate. Hence, the *Starship Enterprise* could achieve a parking orbit around the Earth, and we might never know it's there! Table 9-2 lists representative limits for some of the sensor sites we've identified.

TABLE 9-2. Physical Limitations of Representative Sensors. This table includes representative range, azimuth, and elevation values, but I've left out minimum-range values because they're not very useful. Minimum elevations often include an allowance for tropospheric distortions. Some azimuth ranges go through 360°. (Vallado, 1989)

Sensor	Range (km)		Azimuth (°)		Elevation (°)	
	Maximum	Minimum	Maximum	Minimum	Maximum	
Socorro, NM*	—	0	360	20	90	
Maui, HI*	—	0	360	20	90	
Diego-Garcia, Ind Ocean*	—	0	360	20	90	
Eglin, FL	13,210	145	215	1	90	
NAVSPASUR, 9 sites†	—	90 only	270 only	0	90	
Fylingdales, UK	4820	285	189	4	70	
Pirincli, Turkey	5100	0	360	2	86	
Ascension, Atlantic Ocean	1900	0	360	1	90	
Clear, AK	4910	170	110	1	90	
Antigua, West Indies	2550	0	360	0	90	
Cape Cod, MA	5555	347	227	3	80	
Beale, CA	5555	126	6	3	80	
Thule, Greenland	5555	297	177	3	80	
Cavalier, ND	3300	313	62	2	45	
Kaena Point, HI	6380	0	360	0	90	
Kwajalein Atoll	4500	0	360	1	90	
Millstone, MA	40,744	0	360	0	90	

*Output data is right ascension and declination and acquisition depends on optical cross-section size.

†Output data is direction cosines in two directions and range depends on radar cross-section size.

Calibration is very important when using real-world observations. Tracking data is imperfect, often containing biases and noise. Calibration identifies the biases and deduces the noise statistics. This information is necessary to accurately weight data so we can improve the accuracy of the solution. Table 9-3 lists representative noise and biases for some sensors. Later, we'll discuss how we use this data.

TABLE 9-3. Representative Calibration Values for Sensors. Noise statistics and biases are listed for some common sensors. Realize these values are only estimates and change dramatically over time mainly because components age. (Vallado, 1989)

Sensor	Bias			Noise		
	Range (m)	Azimuth (°)	Elevation (°)	Range (m)	Azimuth (°)	Elevation (°)
Socorro, NM*	—	0.0017	0.0010	—	0.0033	0.0027
Maui, Hawaii*	—	0.0023	0.0017	—	0.0037	0.0030
Diego-Garcia, In*	—	0.0017	0.0023	—	0.0043	0.0043
Eglin, FL	4.3	0.0100	0.0094	32.1	0.0154	0.0147
NAVSPASUR	—	0.0011	0.0011	—	0.0110	0.0110
Fylingdales, UK	2253.0	0.0690	0.0255	2199.0	0.1034	0.0364
Pirincliuk, Turkey	409.0	0.0264	0.0193	38.80	0.0337	0.0290
Ascension, Atl	92.2	0.0133	0.0085	101.7	0.0283	0.0248
Clear, Alaska	153.2	0.0355	0.0171	62.5	0.0791	0.0240
Antigua, West Ind	80.0	0.0081	0.0045	92.5	0.0224	0.0139
Cape Cod, MA	70.8	0.0130	0.0075	26.0	0.0260	0.0220
Beale, CA	92.0	0.0160	0.0130	35.0	0.0320	0.0330
Thule, Greenland	70.8	0.0130	0.0075	26.0	0.0260	0.0220
Cavalier, ND	49.9	0.0036	0.0038	28.0	0.0125	0.0086
Kaena Point, Ha	80.0	0.0081	0.0045	92.5	0.0224	0.0139
Kwajalein Atoll	108.0	0.0140	0.0161	162.9	0.0318	0.0129
Millstone, MA	150.0	0.0001	0.0001	150.0	0.0100	0.0100

*Output data is right ascension and declination.

Given the limitations of the sensors and their effective ranges, we'll see that sensors must be spread throughout the world to cover a particular satellite. The primary mission usually drives where sensors are. For example, NASA uses sensors on the Eastern and Western Test Ranges to support launch operations. The geodetic community uses satellite laser-ranging and other sites around the world to support scientific studies including geoid, tidal, polar motion, and plate tectonic modeling. Figure 9-2 shows some of the sensors used to track satellites.

Regardless of the data's source, limitations, and form, we should make our estimation algorithms general enough to process it into a useful result. In other words, estimation must be versatile in predicting, filtering, and smoothing data. For estimation, *predicting* is simply using existing observations to find future states, which sounds like the propagation techniques of Chap. 4. We'll see later that estimation techniques are intimately tied to propagation methods. *Filtering* is determining the current state using current (and past) observations. In this book, I'll use filtering to refer to all Kalman-filter applications and

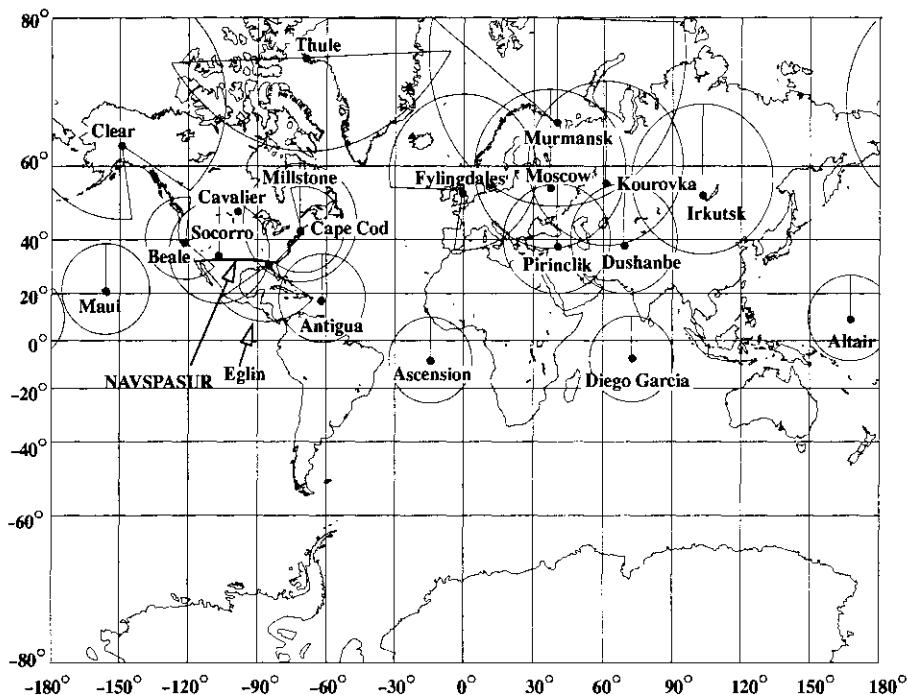


Figure 9-2. Representative Worldwide Sensors. Sensors are around the world, but coverage isn't particularly good in the Southern Hemisphere and over the oceans. I've given the sites notional 2500 km range limits, not the values in Table 9-2.

sequential-batch methods that update the state and covariance matrices. Finally, *smoothing* techniques improve previous states based on existing data. We sometimes apply this process to the observations themselves, and we call it *pre-processing*. Often, least squares and traditional sequential-batch least squares are used as smoothing techniques because they determine the state at a particular epoch time. As an aside, for Kalman-filter smoothing techniques, additional considerations arise. We won't consider these applications in detail. Consult Gelb (1989, 200–203) and Bierman (1977, 211–232) for more development.

We sometimes make two distinctions in estimation. One describes the system's evolution over time using deterministic equations of motion, i.e., assuming the system's dynamics are exact and considering only the measurements as error sources. This method is *deterministic* and assumes you can accurately model everything about the system. Unfortunately, many forces (such as atmospheric conditions or gravity anomalies) acting on a satellite in the real world are incompletely known, so this method is unacceptable for highly accurate missions. On the other hand, the *stochastic* approach uses information in addition to measurements to account for the fact that neither the mathematical models nor

the measurements are perfectly known but are corrupted by random processes. We can then combine information from the dynamics, uncertainty with the force model, and the measurement errors to obtain the best estimate possible. This combination is the basis of estimation.

The following sections describe several ways to statistically calculate an orbit. The underlying approach is stochastic because we use additional observations which include noise to estimate the best answer. We'll begin by examining smoothing techniques. *Linear least squares* introduces the notation and fundamental concepts. *Nonlinear least squares* is the more general problem we encounter in orbit determination. This problem is very difficult to solve completely, so we linearize and combine least squares with an iterative processing solution called differential correction. The concept of *sequential least squares* is based on least-squares evaluations of sequential batches of observations. It uses statistical information from previous least-squares solutions to help process succeeding batches of data. Finally, we'll explore the filtering techniques for the *Kalman* and *Extended Kalman filters*. Because the methods of solution require iteration or incremental updates to the state, *all* the techniques in this chapter are often referred to as *differential correction*. Be aware that differential correction is used interchangeably with linear and nonlinear least squares, and sometimes even with filtering processes. Finally, both least squares and sequential least-squares techniques process groups of data, or *batches*, so they're sometimes called *batch least squares* and *sequential-batch* methods. I'll keep the batch descriptor with the sequential least-squares case to distinguish it from Kalman-filtering applications.

9.3 Linear Least Squares

Linear unweighted least squares is the simplest estimation technique. It assumes that all data is given equal weighting, or importance. Remember that it's really a smoothing technique. We'll use an example problem to discuss the method. The problem consists of a highly polished steel marble rolling across a horizontal table marked off by an x - y coordinate frame. The objective is to find a trajectory equation, $y = f(x)$, to model the rolling marble, using data taken from sensors along the axes. Assume the marble leaves the origin as depicted in Fig. 9-3.

The eight data points shown are our information, but the data contains noise. How do we find $y = f(x)$? Two possible solutions are

1. With eight points, fit a seventh-order polynomial through the points.
2. Using our understanding of the problem's physical nature, choose a more appropriate mathematical model and fit the measurements "as well as possible."

The stochastic solution is the second, defining "as well as possible" to mean "minimize the sum of the squares of the residuals" (least squares). The residuals will be the difference

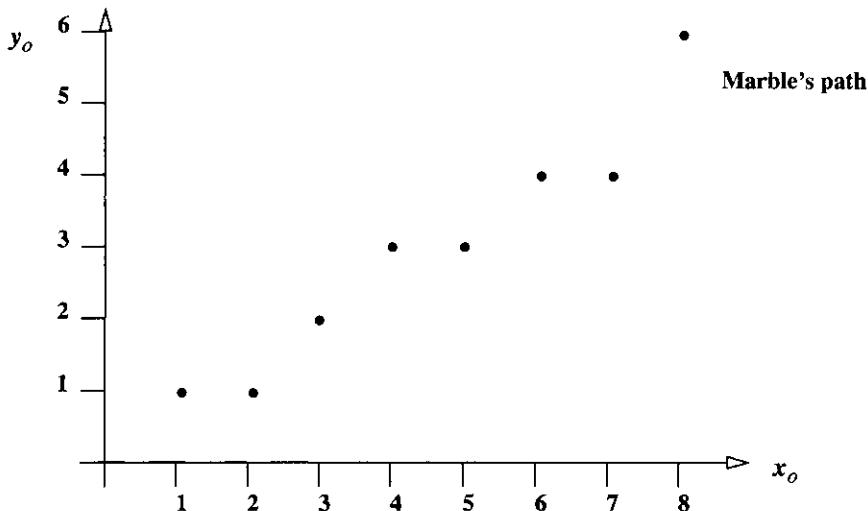


Figure 9-3. Path of a Marble Across a Table. To illustrate the process of least squares, consider a randomly moving marble on a table. The x and y locations are the actual measurements over time.

in the actual observations and those obtained from the state vector solution. The first step in the solution is to define all variables:

$x_o \Rightarrow$ Observed values of the *independent* variable

$y_o \Rightarrow$ Observed values of the *dependent* variable

$N \Rightarrow$ Number of data points or observations

$i \Rightarrow i^{\text{th}}$ data point subscript

$y_c \Rightarrow$ Computed value of dependent variable, based on $y_c = f(x)$ for any value of x

$r = y_o - y_c \Rightarrow$ **residual**, the difference between observed and computed values of the dependent variable (the difference between the model and reality)

Don't confuse the residual, \bar{r} , with the position magnitude often used in astrodynamics. We'll use the same symbol for both, but the context of a particular equation will establish the correct meaning.

We can now determine the solution. With the second approach, model the motion as a straight line:

$$y_{c_i} = \alpha + \beta x_{o_i}$$

Define each residual, \bar{r}_i , by

$$\bar{r}_i = y_{o_i} - y_{c_i} = y_{o_i} - (\alpha + \beta x_{o_i})$$

so the least-squares criterion (for N observations) or the **cost function**, J , is

$$J = \sum_{i=1}^N \bar{r}_i^2 = \text{a minimum} \quad (9-1)$$

We use a squared function because a parabola (\bar{r}_i^2) has a minimum whereas a line (\bar{r}_i) doesn't.

The task is to determine the parameters α and β in the equation below such that the sum of the squares of the residuals, \bar{r}_i^2 , is a minimum. Notice these parameters are known as the *state* for this problem, and they define the location of the object of interest. Don't confuse β used in these examples with azimuth used later on in this chapter.

To find this minimum, set the first derivative, with respect to the state, of the least-squares criterion to zero and solve for the unknowns α and β . This result indicates the extreme values of the function. Mathematically, the function [Eq. (9-1)] is a parabola with no maximum and one minimum; thus, the first derivative will provide a global minimum. The computed y_i equation contains α and β in a *linear* relationship, so

$$J = \sum_{i=1}^N \bar{r}_i^2 = f(\alpha, \beta) = \sum_{i=1}^N (y_{o_i} - (\alpha + \beta x_{o_i}))^2$$

At this point, we could simply differentiate the cost function and determine the answer directly. Let's take a different approach for this case to show the individual steps. Because this function has two variables, both partial derivatives are equal to 0 at the minimum, so you can split the equation in two:

$$\frac{\partial}{\partial \alpha} \sum_{i=1}^N \bar{r}_i^2 = \sum_{i=1}^N \frac{\partial \bar{r}_i^2}{\partial \alpha} = \sum_{i=1}^N 2 \bar{r}_i \frac{\partial \bar{r}_i}{\partial \alpha} = 0 \quad \frac{\partial}{\partial \beta} \sum_{i=1}^N \bar{r}_i^2 = \sum_{i=1}^N \frac{\partial \bar{r}_i^2}{\partial \beta} = \sum_{i=1}^N 2 \bar{r}_i \frac{\partial \bar{r}_i}{\partial \beta} = 0$$

$$\sum_{i=1}^N \bar{r}_i \frac{\partial \bar{r}_i}{\partial \alpha} = 0 \quad \sum_{i=1}^N \bar{r}_i \frac{\partial \bar{r}_i}{\partial \beta} = 0$$

Now, solve these two equations for α and β . If you use the residual definition,

$$\bar{r}_i = y_{o_i} - (\alpha + \beta x_{o_i})$$

the first equation becomes

$$\sum_{i=1}^N \bar{r}_i \frac{\partial \bar{r}_i}{\partial \alpha} = \sum_{i=1}^N \bar{r}_i \frac{\partial (y_{o_i} - \alpha - \beta x_{o_i})}{\partial \alpha} = \sum_{i=1}^N \bar{r}_i (-1) = 0$$

and the second equation is

$$\sum_{i=1}^N \bar{r}_i \frac{\partial \bar{r}_i}{\partial \beta} = \sum_{i=1}^N \bar{r}_i \frac{\partial (y_{o_i} - \alpha - \beta x_{o_i})}{\partial \beta} = \sum_{i=1}^N \bar{r}_i (-x_{o_i}) = 0$$

Find the two resulting simultaneous equations by evaluating the summation from $i = 1$ to N (in this case $N = 8$):

$$\begin{aligned}\bar{r}_1(-1) + \bar{r}_2(-1) + \dots + \bar{r}_8(-1) &= 0 \\ \bar{r}_1(-x_{o_1}) + \bar{r}_2(-x_{o_2}) + \dots + \bar{r}_8(-x_{o_8}) &= 0\end{aligned}$$

These equations include two unknowns, α and β . To solve for α and β , we must separate them from everything else. First, put the equations in matrix notation and divide by -1 :

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ x_{o_1} & x_{o_2} & \dots & x_{o_8} \end{bmatrix} \begin{bmatrix} \bar{r}_1 \\ \bar{r}_2 \\ \vdots \\ \bar{r}_8 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Substitute the definition of the residual, \bar{r}_i , to get

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ x_{o_1} & x_{o_2} & \dots & x_{o_8} \end{bmatrix} \begin{bmatrix} y_{o_1} - (\alpha + \beta x_{o_1}) \\ y_{o_2} - (\alpha + \beta x_{o_2}) \\ \vdots \\ y_{o_8} - (\alpha + \beta x_{o_8}) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Notice α and β have appeared, but they're in the middle of a matrix. Apply the distributive law to the second matrix:

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ x_{o_1} & x_{o_2} & \dots & x_{o_8} \end{bmatrix} \begin{bmatrix} y_{o_1} \\ y_{o_2} \\ \vdots \\ y_{o_8} \end{bmatrix} - \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_{o_1} & x_{o_2} & \dots & x_{o_8} \end{bmatrix} \begin{bmatrix} \alpha + \beta x_{o_1} \\ \alpha + \beta x_{o_2} \\ \vdots \\ \alpha + \beta x_{o_8} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Separating α and β results in

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ x_{o_1} & x_{o_2} & \dots & x_{o_8} \end{bmatrix} \begin{bmatrix} y_{o_1} \\ y_{o_2} \\ \vdots \\ y_{o_8} \end{bmatrix} - \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_{o_1} & x_{o_2} & \dots & x_{o_8} \end{bmatrix} \begin{bmatrix} 1 & x_{o_1} \\ 1 & x_{o_2} \\ \vdots & \vdots \\ 1 & x_{o_8} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Notice the new matrix is the transpose of the left-hand matrix. Rearrange the matrix addition to place the state vector on one side:

$$\underbrace{\begin{bmatrix} 1 & 1 & \dots & 1 \\ x_{o_1} & x_{o_2} & \dots & x_{o_8} \end{bmatrix}}_{A^T} \underbrace{\begin{bmatrix} 1 & x_{o_1} \\ 1 & x_{o_2} \\ \vdots & \vdots \\ 1 & x_{o_8} \end{bmatrix}}_{\hat{X}} = \underbrace{\begin{bmatrix} 1 & 1 & \dots & 1 \\ x_{o_1} & x_{o_2} & \dots & x_{o_8} \end{bmatrix}}_{A^T} \underbrace{\begin{bmatrix} y_{o_1} \\ y_{o_2} \\ \vdots \\ y_{o_8} \end{bmatrix}}_b$$

Define these matrices in a symbolic form. Keep in mind what specific numbers they contain (let N be the number of observations. In this case $N = 8$).

A \Rightarrow partial-derivative matrix ($N \times 2$)

$A^T \Rightarrow A$ transpose ($2 \times N$)

\hat{X} \Rightarrow solution, state vector, or state space (2×1)

b \Rightarrow observation matrix ($N \times 1$)

In symbolic form then, the equation is

$$A^T \hat{X} = A^T b$$

These equations are the basis of least squares. They're often called the **normal equations**, and they trace their history back to Gauss. Notice that although A is an $N \times 2$ matrix, $A^T A$ is a 2×2 matrix. Thus if we form $A^T A$ directly, we don't have to handle large matrices for cases where we have many observations. Although this is of small benefit to the simple examples in this section, for practical orbit determination, we'll see that it's very important. I'll demonstrate the summation technique in the example.

We still want to solve for \hat{X} (α and β), the parameters that define the trajectory's orientation (state). Notice these don't define the *shape* of the trajectory because we already did that using our insight into the physics of the problem (straight-line motion). Similarly, in orbit determination, we have already defined the orbit's shape as a conic section. We need to determine only its six orbital elements. Although A and A^T aren't usually square matrices, the matrix product $A^T A$ is always square. Thus, we may invert it, provided it's positive definite (not singular). Solving for \hat{X} gives us

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \hat{X} = (A^T A)^{-1} A^T b$$

and the general solution for the linear case of \hat{X} is

$$\hat{X} = (A^T A)^{-1} A^T b \quad (9-2)$$

The matrix inverse, $(A^T A)^{-1}$, is called the *covariance matrix*. We'll discuss this later. Notice that when we put in the test values for each variable, we find the state, \hat{X} . The overall process is sometimes called **parameter estimation** because we are ultimately trying to determine the parameters, α and β .

▼ Example 9-1: Using Linear Least Squares (1).

GIVEN: Data from Fig. 9-3.

FIND: Best estimate of the state

By visually examining Fig. 9-3, we can list the observation data:

x_{o_i}	1	2	3	4	5	6	7	8
y_{o_i}	1	1	2	3	3	4	4	6

We must assume a mathematical model for the motion. Assume linear motion as

$$y_{c_i} = \alpha + \beta x_{o_i}$$

The $A^T A$ and $A^T b$ matrices are found by summing the following relations:

$$A^T A = \begin{bmatrix} \sum_{i=1}^N 1 & \sum_{i=1}^N x_{o_i} \\ \sum_{i=1}^N x_{o_i} & \sum_{i=1}^N x_{o_i}^2 \end{bmatrix} = \begin{bmatrix} 8 & 36 \\ 36 & 204 \end{bmatrix}$$

$$A^T b = \begin{bmatrix} \sum_{i=1}^N y_{o_i} \\ \sum_{i=1}^N x_{o_i} y_{o_i} \end{bmatrix} = \begin{bmatrix} 24 \\ 136 \end{bmatrix}$$

Finding the inverse is usually the most difficult part of the problem and can cause many numerical difficulties as the dynamics become more complex. For this example, we can do so by solving simultaneous equations through $(A^T A)^{-1} (A^T A) = I$ or by an even simpler approximation. Use the matrix rules in Sec. C.2 to find the inverse. The determinant is $(8)(204) - (36)(36) = 336$. With this value, find

$$(A^T A)^{-1} = \frac{\begin{bmatrix} 204 & -36 \\ -36 & 8 \end{bmatrix}}{336} = \begin{bmatrix} 0.6071 & -0.1071 \\ -0.1071 & 0.0238 \end{bmatrix}$$

The linear least-squares solution from Eq. (9-2) is

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = (A^T A)^{-1} A^T b$$

Solving results in

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 0.6071 & -0.1071 \\ -0.1071 & 0.0238 \end{bmatrix} \begin{bmatrix} 24 \\ 136 \end{bmatrix} = \begin{bmatrix} 0.0 \\ 0.667 \end{bmatrix}$$

and the final answer is

▲ $y_{c_i} = 0.0 + 0.667x_{o_i}$

It's important to distinguish between a *curve-fit* and an *estimate* using physical models. An ad hoc curve-fit may fit the data successfully but have little or no value for prediction. Correctly modeling the physics is very important to estimation.

9.3.1 Error Analysis

If our measurements and models were perfect, we wouldn't need estimation at all! Unfortunately, the real world is imperfect with errors arising from many sources. By examining estimation from observation through processing, we can find the major sources of error.

Instrument / Measurement Error

All observations are subject to error. Some errors, such as those due to sensor maintenance, equipment failure, or operator error, aren't particularly well suited for mathematical modeling and usually result in the observation being ignored. Other factors may corrupt data. Consider a satellite's pass over a ground station and recall the general limitations discussed in Fig. 9-1. If the maximum elevation of the pass is only 5°, the sensor may not be able to observe the satellite. Even if it can, the extremely low elevation may lead to less accurate results because the atmospheric distortion from looking through a larger slice of the atmosphere results in imprecise modeling.

Measurement errors are easy to quantify. ***Measurement error, R***, is loosely defined to be the variation in the observations about their true value. It consists of three main subcategories: noise, biases, and nonrandom time-varying errors (drift)—as shown in Fig. 9-4.

Biases are a *constant* offset from the true value, whereas ***noise*** is a statistical indication (actually a standard deviation) of the random variation about the measured “mean.” In other words, noise is scattering about the observed mean value. ***Drift*** represents a slow variation over time of the observed mean value over some appropriate interval. The nice feature of biases is that, if we know them, we can simply subtract (or add) them to each observation. We can estimate measurement noise in estimation, but we'd like to know its value *a-priori* to streamline the process. We can estimate the time-varying errors of drift with properly formed Kalman filters, but we can't predict them.

To illustrate the effect of noise and biases, let's continue with Example 9-1. Suppose that the *y* bank of sensors consistently outputs answers with an error of +1 unit and that this bias is unknown to the observer. The intercept of the straight line will have a +1 unit error. The result for Example 9-1 would then be

$$y_{c_i} = 1 + 0.667x_{o_i}$$

Although this example seems to show estimation will detect the bias, this is true *only* because the system is linear *and* the *y* observation isn't coupled to the *x* observation. In general, least-squares techniques *won't* filter out instrument bias, though we can *solve for* biases as part of the state space. On the other hand, we can design estimation techniques to determine biases as part of the solution formula. Because we can simply add or subtract

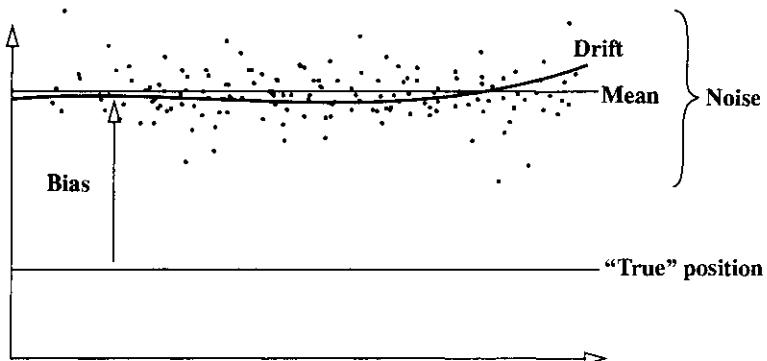


Figure 9-4. Representation of Noise, Biases, and Drift. The mean of the measured values is the average of all observations but may not be the same as the true position. The difference is the bias. The noise is a random variation about the mean.

correct biases from observational data, thereby improving accuracy, it's extremely important to use an accurate method.

Time-varying errors are discussed less often, but they're still important. The inaccuracy of clocks over time is one cause of these errors. Tracking systems such as GPS (and the older Navy Transit system) rely on a transmitting and receiving clock. Doppler measurement errors are a function of the behavior of both clocks. Unfortunately, all clocks suffer from white noise in their frequency, which integrates to a random walk in phase (time). Crystal clocks (quartz, for instance) also suffer from random walk in frequency. Atomic clocks suffer from an error called *flicker noise*. Each of these errors affects range and Doppler measurements. Other time-varying errors result from local tropospheric effects, such as humidity levels as a weather front passes over a sensor. Ionospheric effects depend on the time of day. Temperature affects some sensors—for example, temperature gradients can cause unequal contraction and expansion on optics, or on the faces of large phased-array radars.

Mathematical Modeling

The next main source of error is mathematical modeling during data processing. Although we have some very precise models for propagation (complex atmospheres and gravitational models), we can't model everything perfectly. For instance, *post-fit* solutions of the TOPEX satellite produce less than a 2 cm error in slant-range residual if they use the new JGM-3 Gravity Model coefficients in a differential-correction program with complete numerical propagation. *Pre-processed* solutions are typically 100 times worse! ***Process noise***, v , is the error in the mathematical modeling of the system dynamics. It limits our ability to model accurately the system's true dynamics before and after a given epoch. Closely related is the ***power-density matrix*** or the ***second moment of the process noise***,

\mathbf{Q} , which is a covariance of acceleration errors induced by the mathematical modeling of the system dynamics. We'll discuss these parameters in Sec. 9.7.

Even our best theories can't exactly model the atmosphere, the Earth's shape, or the satellite's attitude. Process noise tries to account for each of these errors. Much of the literature implies that it's difficult to develop the proper models for process noise. This is not entirely true. In fact, if you write orbit-acceleration errors as a function of the gravity field coefficients, thrust uncertainty as a function of time, and so on, you can characterize the errors in those models and *derive* a model for process noise. But these adjustments can be complex.

The end result of process-noise error is that the dynamics may introduce more error into the solution than the measurements. One difficulty is that we often assume the process-noise statistics are random. Unfortunately, they're often systematic errors such as we might expect from an incomplete gravity model. These errors can be highly correlated with time and not well approximated by white noise.

Statistical Information

Several concepts from probability theory will help us understand the estimation process. You may need to review fundamental concepts because I'll explain only briefly those relevant to astrodynamics. Consult texts such as Walpole (1974) for more information and formal derivations of statistics. Figure 9-5 illustrates some concepts.

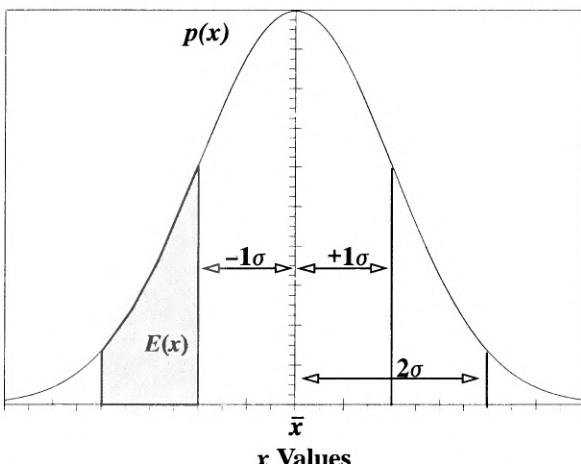


Figure 9-5. Gaussian Distribution. I've shown the mean, \bar{x} , and two bands of the standard deviation, σ . Each standard deviation consists of positive and negative components. The area under the curve is the expectation, $E(x)$. The area between any two points gives us the probability, or expectation; thus, a $\pm 1\sigma$ region encompasses 0.683, or 68.3% (see also Table 9-4). The total area under the curve is unity. Notice each band of standard deviation occurs at equal displacements from the mean.

The basis for estimation is the likelihood of an event occurring described by the *probability density function*, $p(x)$. The probability of an event, $\xi < x < \xi + d\xi$, occurring for a random variable, ξ , is $p(\xi)d\xi$. The probability distribution is subject to the condition that the integral over all possible values is equal to one. Thus,

$$\int_{-\infty}^{\infty} p(\xi) d\xi = 1$$

The *mean*, \bar{x} , is defined as the expected value of an event occurring over all possibilities. Zarchan (1994, 68) says “we can also think of the mean value of x as the sum (integral) of all values of x , each being weighted by its probability of occurrence [$p(x)$].” The usual averages for a finite sample can be shown, for large N , to approximate the corresponding integral expectations. For example, we define the *mean* as

$$\bar{x} \equiv E(x) = \int_{-\infty}^{\infty} \xi p(\xi) d\xi \equiv \frac{1}{N} \sum_{i=1}^N x_i \quad (9-3)$$

The *Law of large numbers* permits us to approximate \bar{x} using the sample mean. This is the basis for *Monte-Carlo methods*.

Recalling the original least-squares criteria, it makes sense also to examine the squared expected value, $E(x^2)$. Notice that the approximation of the finite sample is now divided by $(N - 1)$ because we are trying to get a sense of the sample’s variation. A variation is measured between two quantities, so we must reduce N by one. Moreover, if there is only one measurement, the expected value can’t vary. If there are two or more measurements, variation may occur from one less degree of freedom than the number of observations. So

$$E(x^2) \equiv \int_{-\infty}^{\infty} \xi^2 p(\xi) d\xi \equiv \frac{1}{N-1} \sum_{i=1}^N x_i^2$$

From this expression we can determine the *root mean square* (RMS), which is simply the square root of the mean squared expected value. Note that the RMS does *not* involve the mean and considers only the expected value of the observations. Analysis of RMS values is widely used—often instead of standard deviations, which we’ll introduce shortly.

$$\text{RMS} \equiv \sqrt{E(x^2)} \equiv \sqrt{\frac{1}{N-1} \sum_{i=1}^N x_i^2} \quad (9-4)$$

The literature is somewhat confusing about the actual formula for the RMS. Selby (1975, 573) clearly states the divisor is N , whereas Neter (1982, 61) says notations vary. Many

operational programs assume the sample size (N) is very large, so that using N in place of $N - 1$ matters little. We'll adhere to the $N - 1$ convention to ensure commonality with the standard deviation when the mean is zero (see the next paragraph). For Example 9-1, the RMS of the residuals is equal to 0.436 435.

Although the RMS doesn't involve the mean value, the *variance*, σ^2 , represents the variability of the expected value of each variable from its mean:

$$\begin{aligned}\sigma^2 &\equiv E \{ [x - E(x)]^2 \} = E \{ [x - \bar{x}]^2 \} \\ &= \int_{-\infty}^{\infty} (\xi - \bar{x}) p(\xi) d\xi \equiv \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2\end{aligned}\quad (9-5)$$

The *standard deviation*, σ , is the positive square root of the variance; it measures the dispersion of the data. Be aware that the RMS and the standard deviation usually *aren't* the same. They are equal *only* when the mean is zero. Many astrodynamical applications use these quantities interchangeably, but it's very important to keep the two distinct. With a zero mean, the RMS and the standard deviation will be the same (remember to use $N - 1$). Assuming samples have a zero mean is widely accepted in astrodynamics.

The most important probability density function is the *Gaussian*, or *normal* distribution (Fig. 9-5). It's the most common measurement model and has many elegant properties. Zarchan (1994, 70) notes:

It can be shown that the resultant probability density function of a sum of Gaussian distributed random variables is also Gaussian. In addition, under certain circumstances, it can also be shown that the sum of independent random variables, regardless of individual density function, tends toward Gaussian as the number of random variables gets larger. That is in fact why so many random variables are Gaussian distributed.

The final statement suggests the basis of the *central limit theorem*, which says large sample sizes will converge to Gaussian distributions. If the errors are truly random and the model accurately captures the physics, more data will produce better estimates. Of course, for cases with large errors in the system model, such as the least-squares example mentioned previously, using a linear model for parabolic data won't improve with more data.

We've introduced the mean and standard deviations for a single scalar random variable. The results generalize to vectors. Given a vector with random variables as components, its mean is the vector with components given by the mean of the random variables. The concept of standard deviations generalizes to a square covariance matrix, P , which we'll discuss shortly. For a vector of two or more scalar functions, we can use a mass distribution as an excellent analogy to the probability distribution. The mean is analogous to the location of the center of mass, and the covariance matrix is analogous to the matrix for the moment of inertia. We saw applications of these analogies in Chap. 7.

When we consider random vectors, we form the mean using concepts from Eq. (9-3), only now we have a matrix solution. Suppose we have a state model in which the observations are y , the state is x , and we have arbitrary constants, m and b . In this case,

$$\mathbf{y} = \mathbf{mx} + \mathbf{b} \quad (9-6)$$

Using Eq. (9-3) we can determine the mean of the state,

$$\bar{\mathbf{y}} \equiv E(\mathbf{y}) = E[\mathbf{mx} + \mathbf{b}] = mE(\mathbf{x}) + \mathbf{b} = m\bar{\mathbf{x}} + \mathbf{b}$$

From Eq. (9-5), we find the variance as follows. Notice that it's a square matrix due to the squared term in the expectation operator:

$$\mathbf{P}_y \equiv E[(\mathbf{y} - \bar{\mathbf{y}})(\mathbf{y} - \bar{\mathbf{y}})^T] = E[m(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^Tm^T]$$

Using properties from linear-error theory and simplifying gives us

$$\begin{aligned}\mathbf{P}_y &= mE[(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T]m^T \\ \mathbf{P}_y &= m\mathbf{P}_x m^T\end{aligned} \quad (9-7)$$

This result is usually known as the *similarity transformation*; it allows us mathematically to describe the covariance matrix, \mathbf{P}_y .

The standard deviations resulting from the square root of the diagonal terms in the covariance matrix indicate the error of the estimate. We can use these standard deviations to eliminate spurious data points. Table 9-4 shows probabilities of a given sample or data point being within a certain standard deviation of the mean, for one- to three-dimensional Gaussian probability distributions. The two-dimensional case represents the volume formed by rotating the bell curve (in Fig. 9-5) about the vertical axis associated with the mean value, $\bar{\mathbf{x}}$.

TABLE 9-4. Probability Values. Probability values are given for different standard deviations as a function of the dimension of the state space. For example, 95.5% of the observations are within 2σ of the mean for a one-dimensional system (Junkins, 1977, Appendix B). Remember that regardless of dimension, the probability distribution is normalized to 1.0.

Dimension	1σ	2σ	3σ
1	0.683	0.955	0.997
2	0.394	0.865	0.989
3	0.200	0.739	0.971

We can select certain data points for elimination by checking for those with residuals outside some number of standard deviations from the mean. Of course, we must either process the data once or already know the standard deviation before doing this editing. Eliminating data can be very tricky, especially if data is sparse. As an example of spurious data points, consider Example 9-2.

▼ **Example 9-2. Using Linear Least Squares (2).**

Suppose observation 7 differs from Example 9-1:

x_{o_i}	1	2	3	4	5	6	7	8
y_{o_i}	1	1	2	3	3	4	7	6

Solve for α and β again. The A matrix is the same as before, as is A^T and $(A^T A)^{-1}$. But the b matrix has changed, $b^T = \begin{bmatrix} 1 & 1 & 2 & 3 & 3 & 4 & 7 & 6 \end{bmatrix}$.

$$A^T b = \begin{bmatrix} \sum_{i=1}^N y_{o_i} \\ \sum_{i=1}^N x_{o_i} y_{o_i} \end{bmatrix} = \begin{bmatrix} 27 \\ 157 \end{bmatrix}$$

The inverse matrix is still the same, so

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 0.6071 & -0.1071 \\ -0.1071 & 0.0238 \end{bmatrix} \begin{bmatrix} 27 \\ 157 \end{bmatrix} = \begin{bmatrix} -0.429 \\ 0.845 \end{bmatrix}$$

giving an answer of

$$y_{c_i} = -0.429 + 0.845x_{o_i}$$

This fitted line represents the *expected motion*, whereas the residuals are errors off the line. Next, calculate the residuals:

$$\bar{r}_i = y_{o_i} - y_{c_i} = y_{o_i} - (\alpha + \beta x_{o_i}) = y_{o_i} - (-0.429 + 0.845x_{o_i})$$

Continue this process for each observation:

y_{o_i}	1	1	2	3	3	4	7	6
x_{o_i}	1	2	3	4	5	6	7	8
y_{c_i}	0.417	1.262	2.107	2.952	3.797	4.643	5.488	6.333
\bar{r}_i	0.583	-0.262	0.107	0.048	-0.797	-0.643	1.512	-0.333

Because least-squares techniques lead to average, or unbiased estimates, we assume a zero mean and calculate the RMS using the residuals. Thus,

$$\text{RMS} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N \bar{r}_i^2} = \sqrt{\frac{3.869}{7}} = 0.743\ 45$$

Because this is an unbiased estimate, the mean is zero, and the RMS is the standard deviation. Once we've completed one pass through the data, we can check for points with a residual outside, say, a 2σ range. The 2σ (0.743) limit is for illustration only and is never used in the real world. It's much more common to use a 4σ to 5σ criterion depending on your application. A specific problem may impose different restrictions. Observation 5 is close, so we'll keep it. Observation 7 has a residual of 1.512 , so we'll remove it as nonrepresentative. Using all data points except 7 and recalculating the least-squares

▲ solution for α and β gives us the improved answer: $y_{c_i} = -0.131 + 0.721x_{o_i}$.

Summarizing, Example 9-2 illustrates detecting and eliminating a bad data point with the following results. Before eliminating point 7,

$$y_{c_i} = -0.429 + 0.845x_{o_i} \quad \text{RMS} = 0.743\,452$$

$$\alpha = -0.428 \pm 0.7792 \text{ and } \beta = 0.845 \pm 0.1543$$

After eliminating point 7, the values become

$$y_{c_i} = -0.131 + 0.721x_{o_i} \quad \text{RMS} = 0.346\,725$$

$$\alpha = -0.131 \pm 0.7970 \text{ and } \beta = 0.721 \pm 0.1694$$

Notice that, although the RMS dropped, our confidence in the answer got worse because there was less data! Usually, we want greater confidence in the answer. In the example, this is the best we can do with only sparse data, and it points out a basic limitation. It's important to distinguish between a "curve fit" and an estimate using physical models. With an arbitrary curve, we may fit the data well but obtain a model with no value for prediction. Perhaps the most difficult part of estimation is getting the correct physical model which not only fits the data but also gives us reasonable confidence in the accuracy of our predictions. Incorrect physical models can cause the unestimated parameters to "lie." Another problem is using the RMS to edit the data with too tight a tolerance. If we used the new RMS (0.346) we might lose observations that are crucial to the results, hence the larger ranges for standard deviations used to edit the data.

Consider Parameters

For least squares, we can include the effect of the uncertainty of unestimated parameters on the covariance (but not the converged state estimate) through ***consider parameters***. These parameters appear as extra variables in the state space but they aren't estimated through processing. Consider parameters provide a way to incorporate *a priori* estimates of certain parameters through their associated confidence measures. We won't discuss this process in detail because the Kalman filter is a much more efficient way to do this analysis. Also see Bierman (1977, 162–182) and Long et al. (1989, Chap. 8) for more information.

9.3.2 Linear Weighted Least Squares

The computations for *weighted least squares* don't differ very much from the unweighted case. However, introducing weights permits us to account for differences in the accuracy of measurements. We'll use the same example from the linear case to illustrate the principle; however, instead of being given the observations, suppose two *different* types of sensor equipment take the data. Set A is an expensive, sophisticated system providing highly accurate measurements, and Set B is a cheap, simple system with less desirable noise characteristics. If we consider the data as is, the residuals for Set B will be larger than those from Set A. We want to weight the fit toward Set A, because its data is better. Suppose the

manufacturers provide sensor accuracies (standard deviations) with $\sigma_A < \sigma_B$. This assumption could be a source of practical difficulties because the manufacturer of a cheap sensor usually won't accurately characterize the sensor errors. Because the fit is based on residuals, we will *weight the residuals*, using the inverses of the standard deviations. We don't really weight individual observations; rather, we weight observation classes, usually by a sensor type or location. Thus, all the observations (of the same type) from a particular sensor are assumed to have similar characteristics. Gauss placed this heuristic choice on rigorous theoretical grounds, and under reasonable assumptions, we can prove this choice leads to the most likely trajectory estimate. Recognize this weight choice is simply the inverse of the measurement noise squared (Table 9-3). Later, we'll use a measurement noise matrix, \mathbf{R} , that is equal to the square of the noise values. Thus, a small standard deviation will weight the sensor more.

$$w_A = \frac{1}{\sigma_A^2} \text{ and } w_B = \frac{1}{\sigma_B^2} \text{ so that } w_A > w_B$$

The weighted residuals are $w_A \bar{r}_i$ for Set A and $w_B \bar{r}_i$ for Set B. We still assume a linear trajectory and calculate the residuals as in the linear case:

$$y_{c_i} = \alpha + \beta x_{o_i}$$

$$b_c = f(x) = \alpha + \beta x$$

$$w_i \bar{r}_i = w_i (b_{o_i} - b_{c_i}) = w_i (b_{o_i} - (\alpha + \beta x_{o_i}))$$

Now, let's write the weighted residuals using matrix notation, remembering that each observation will have weights associated with the state-vector parameters.

$$w_i = \begin{bmatrix} \frac{1}{\sigma_A^2} & 0 \\ 0 & \frac{1}{\sigma_B^2} \end{bmatrix} = \begin{bmatrix} w_A^2 & 0 \\ 0 & w_B^2 \end{bmatrix}$$

$$\bar{\mathbf{r}} = \mathbf{b} - \mathbf{AX}$$

$$\mathbf{A} = \begin{bmatrix} 1 & x_{o_1} \\ 1 & x_{o_2} \\ \vdots & \vdots \\ 1 & x_{o_8} \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \quad \mathbf{W} = w_i^T w_i = \begin{bmatrix} w_1^2 & 0 & \dots & 0 \\ 0 & w_2^2 & & \\ \vdots & & \ddots & \\ & & & 0 \\ 0 & \dots & 0 & w_8^2 \end{bmatrix}$$

We denote the *weighting matrix*, W , to simplify the notation. Applying the least-squares criterion [Eq. (9-1)] to the weighted residuals produces the cost function:

$$J = \sum_{i=1}^N w_i^2 \bar{r}_i^2 = \bar{r}^T W \bar{r} = (\mathbf{b} - \mathbf{A}\mathbf{X})^T W (\mathbf{b} - \mathbf{A}\mathbf{X})$$

To find the minimum, we first expand the cost function:

$$J = (\mathbf{b} - \mathbf{A}\mathbf{X})^T W (\mathbf{b} - \mathbf{A}\mathbf{X}) = \mathbf{b}^T W \mathbf{b} - 2\mathbf{b}^T W \mathbf{A} \mathbf{X} + \mathbf{X}^T \mathbf{A}^T W \mathbf{A} \mathbf{X}$$

Now, let's use the matrix rules for differentiation (Appendix C) and set the derivative of the cost function equal to zero. Because we'll be getting the best estimate of the state, we'll still use $\hat{\mathbf{X}}$ for the result.

$$\frac{\partial J}{\partial \mathbf{X}} = -2\mathbf{b}^T W \mathbf{A} + 2\hat{\mathbf{X}}^T \mathbf{A}^T W \mathbf{A} = 0$$

Let's simplify by taking the transpose of both sides and distributing the transpose operator. Noting that $W^T = W$, we get

$$(\hat{\mathbf{X}}^T \mathbf{A}^T W \mathbf{A})^T = (\mathbf{b}^T W \mathbf{A})^T$$

$$\mathbf{A}^T W \mathbf{A} \hat{\mathbf{X}} = \mathbf{A}^T W \mathbf{b}$$

The solution state is then

$$\hat{\mathbf{X}} = (\mathbf{A}^T W \mathbf{A})^{-1} \mathbf{A}^T W \mathbf{b} = P \mathbf{A}^T W \mathbf{b} \quad (9-8)$$

The definitions are the same as the unweighted case, with one addition:

$$W \Rightarrow \text{weighting matrix } (N \times N)$$

Besides providing the best estimate of the state (α and β in the example), the least-squares method also gives us statistical confidence in the answers. The **covariance matrix**, $P = (\mathbf{A}^T W \mathbf{A})^{-1}$, contains the estimates for the closeness of the fit with the actual observations and is the same size as in the unweighted case, Eq. (9-2). We call this the *variance / covariance matrix* of the estimate, or more simply, the covariance matrix because it contains both variances and covariances. The variances, the squares of the standard deviations, describe the closeness of the “fit” with the data. Note that these standard deviations are *not* the same as the pseudo-variances in the power-density matrix associated with process noise we discuss later. For the simple or unweighted least-squares problem, we calculate the covariance matrix as the matrix $(\mathbf{A}^T \mathbf{A})^{-1}$. For the weighted case, we include the weighting matrix, W , and use $(\mathbf{A}^T W \mathbf{A})^{-1}$; however, the principles here remain valid. On page 713, we'll discuss developing a more general form with the Kalman filter. The covariance matrix is square and symmetrical, and it contains variances and covariances of the estimated parameters. Using an example state having three estimated parameters— α , β , γ —gives us the covariance matrix of the form

$$\mathbf{P} \equiv (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} = E(\left[\mathbf{X} - \bar{\mathbf{X}}\right]^2) = \begin{bmatrix} \sigma_\alpha^2 & \mu_{\alpha\beta}\sigma_\alpha\sigma_\beta & \mu_{\alpha\gamma}\sigma_\alpha\sigma_\gamma \\ \mu_{\beta\alpha}\sigma_\alpha\sigma_\beta & \sigma_\beta^2 & \mu_{\beta\gamma}\sigma_\beta\sigma_\gamma \\ \mu_{\gamma\alpha}\sigma_\alpha\sigma_\gamma & \mu_{\gamma\beta}\sigma_\beta\sigma_\gamma & \sigma_\gamma^2 \end{bmatrix}$$

where σ_α is the standard deviation defined in Eq. (9-5), σ_α^2 is the variance, $\mu_{\alpha\beta}$ is the correlation coefficient of α with β ($\mu_{\beta\alpha} = \mu_{\alpha\beta}$), and so forth.

The diagonal terms are the variances of the estimate of the state parameters. The square roots of the variances are the standard deviations for each element of the state space. You should *always* include these values when discussing the results of an estimation calculation because they establish statistical confidence in each element. Some systems don't use this data because it is deemed inaccurate when obtained from multiple passes—an opinion that is only partly correct. The covariance may indeed be inaccurate due to poorly known statistics (noise, biases, and drift) and improper propagation techniques used during differential correction, but these can be fixed. Whatever the fidelity, we can use the covariance matrix to indicate trends, such as which direction to search if a sensor doesn't immediately acquire a satellite.

The off-diagonal elements are called covariance terms; they contain the **correlation coefficients** ($\mu_{xy}^2 = (1/\sigma_x\sigma_y) E\{(x-\bar{x})(y-\bar{y})\}$). They are sometimes just as important as the variances. The **correlation coefficients** represent the degree of independence among the elements, with zero being complete independence and positive signs indicating a direct correlation. Negative signs imply an inverse relationship. We want the correlations to be zero or, at least, very small. Consider solving for a satellite's classical orbital elements and the drag coefficient, c_D . We've explored c_D in greater detail in Sec. 7.6.2, but it adjusts the effect of drag-model density to match the observed drag effects on the satellite motion. Because drag depends primarily on atmospheric density which in turn depends on altitude, the satellite's altitude is directly related to the density calculations. Thus, the semimajor axis and c_D are highly correlated in the estimation process.

Another example occurs in the problem of determining a GPS orbit—the correlations between clock-phase (time) error and the GPS's radial position are large and negative, indicating they are almost indistinguishable from the observations. Remember that these correlations depend on the available observation types. It's possible that other observation types might yield smaller correlation coefficients between parameters—for example, an onboard accelerometer to measure an estimate of c_D . For satellites flying in formation, the relative position accuracy can be orders of magnitude better than the absolute position error because errors in the gravity model are correlated 0.99+ for closely spaced objects. Correlations can be extremely important.

Using the covariance matrix in Example 9-1 gives us

$$(\mathbf{A}^T \mathbf{A})^{-1} = \begin{bmatrix} 0.6071 & -0.1071 \\ -0.1071 & 0.0238 \end{bmatrix} \quad y_c = 0.000 + 0.667x_o$$

We can examine several quantities in these equations. First, the answer represents a line (we guessed a straight line for the dynamics) with minimized residuals. The covariance matrix tells us the bounds of the answer. Remember to find the square roots of each term (0.6071 and 0.0238).

$$\alpha = 0.000 \pm 0.7792$$

$$\beta = 0.667 \pm 0.1543.$$

These are both one-sigma confidence intervals. Notice how much uncertainty there is in the answer! Without knowing the standard deviations, the user would never know how much variation exists in the solution parameters, hence the importance of providing statistical information with the answer. The off-diagonal terms show the two coefficients correlate only slightly (-0.1071), but they're not completely independent.

9.4 Nonlinear Least Squares

We have developed a tool for the linear case that can give us the best answer based on model selection and measurement data corrupted with Gaussian noise. We can apply least-squares techniques to nonlinear problems by linearizing the problem, obtaining an approximate solution, and then iterating to refine the answer. We're now estimating *corrections* to our initial estimate of the state. As an example, assume an ant is rolling a marble across a table in a sine-wave pattern. Figure 9-6 shows the expected path of the marble. Our job is to determine the amplitude, α , and the phase shift, β , of the sine wave.

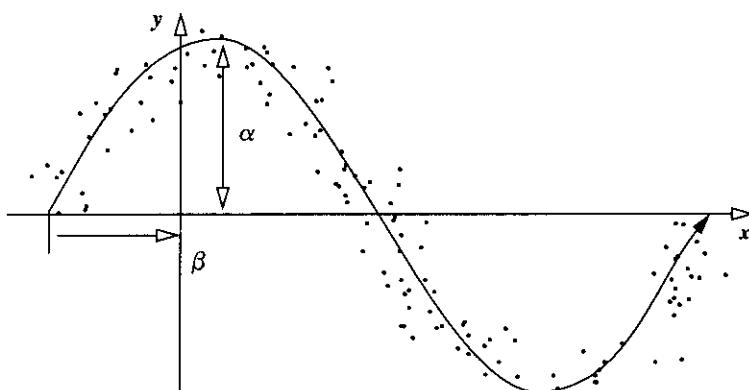


Figure 9-6. Sinusoidal Path of a Marble. If we now view a sinusoidal path $y = f(x) = \alpha \sin(x + \beta)$, the estimation problem becomes nonlinear because we can't separate α and β .

Again, assume our sensors measure y as a function of x but give us noisy data because they're not perfect. The measurement and residual equations for this problem are

$$y = f(x) = \alpha \sin(x + \beta)$$

$$\bar{r}_i = y_{o_i} - \alpha \sin(x_{o_i} + \beta)$$

In this case, the measurement-state relationship, $y = \alpha \sin(x + \beta)$, is now a nonlinear function of the state, α , β . Let's proceed through the least-squares development and see if this nonlinearity makes any difference. Apply the least-squares criterion of Eq. (9-1) to get

$$\sum_{i=1}^N \bar{r}_i \frac{\partial \bar{r}_i}{\partial \alpha} = 0$$

$$\sum_{i=1}^N \bar{r}_i \frac{\partial \bar{r}_i}{\partial \beta} = 0$$

$$\frac{\partial \bar{r}_i}{\partial \alpha} = -\sin(x_{o_i} + \beta)$$

$$\frac{\partial \bar{r}_i}{\partial \beta} = -\alpha \cos(x_{o_i} + \beta)$$

Put the summations in matrix form:

$$-1 \begin{bmatrix} \sin(x_{o_1} + \beta) & \sin(x_{o_2} + \beta) & \dots & \sin(x_{o_N} + \beta) \\ \alpha \cos(x_{o_1} + \beta) & \alpha \cos(x_{o_2} + \beta) & \dots & \alpha \cos(x_{o_N} + \beta) \end{bmatrix} \begin{bmatrix} \bar{r}_1 \\ \bar{r}_2 \\ \vdots \\ \bar{r}_N \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Using the equation for \bar{r}_i , divide by -1 , and use the definition for the A^T matrix. But notice that now the partial-derivative matrix results in expressions, not simply the observations:

$$A^T \begin{bmatrix} y_{o_1} - \alpha \sin(x_{o_1} + \beta) \\ y_{o_2} - \alpha \sin(x_{o_2} + \beta) \\ \vdots \\ y_{o_N} - \alpha \sin(x_{o_N} + \beta) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Now split the matrix to find

$$A^T \begin{bmatrix} y_{o_1} \\ y_{o_2} \\ \vdots \\ y_{o_N} \end{bmatrix} - A^T \begin{bmatrix} \alpha \sin(x_{o_1} + \beta) \\ \alpha \sin(x_{o_2} + \beta) \\ \vdots \\ \alpha \sin(x_{o_N} + \beta) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

The next step is to separate the state, α and β , from the right-hand matrix and solve for it, but there is one major difficulty: you can't separate α and β from the nonlinear relation-

ship, $\alpha \sin(x + \beta)$. In fact, you could separate α but not β . Therefore, the equation is linear in α and nonlinear in β .

Thus, you're stuck if you can't come up with some technique of handling a nonlinear relationship for the measurement state. Fortunately, a Taylor series allows you to approximate the nonlinear equations with linear equations, provided we can neglect the higher-order terms in the Taylor series. So, if you compute $y = f(x)$ as an expansion about a nominal α_n and β_n ,

$$y_c = f(\alpha, \beta, x_o) = g(\alpha, \beta) \text{ for any given } x_o$$

The Taylor series is

$$\begin{aligned} y_c &= y|_{\alpha_n \beta_n} + (\alpha - \alpha_n) \frac{\partial y}{\partial \alpha}|_{\alpha_n \beta_n} + (\beta - \beta_n) \frac{\partial y}{\partial \beta}|_{\alpha_n \beta_n} + \frac{(\alpha - \alpha_n)^2}{2!} \frac{\partial^2 y}{\partial \alpha^2}|_{\alpha_n \beta_n} \\ &\quad + \frac{(\beta - \beta_n)^2}{2!} \frac{\partial^2 y}{\partial \beta^2}|_{\alpha_n \beta_n} + \dots \end{aligned}$$

Because we neglect higher powers (second order and above) of $(\alpha - \alpha_n)$ and $(\beta - \beta_n)$ in the linearization, the formulation provides corrections to a known state as $\Delta\alpha = \alpha - \alpha_n$ and $\Delta\beta = \beta - \beta_n$. Consequently, the nonlinear least squares problem requires an *a priori* estimate of the state for solution. Now,

$$y_{c_i} = y_{n_i} + \Delta\alpha \frac{\partial y_{n_i}}{\partial \alpha} + \Delta\beta \frac{\partial y_{n_i}}{\partial \beta}$$

where

$$y_{n_i} = y_i|_{\alpha_n \beta_n} \quad \frac{\partial y_{n_i}}{\partial (\)} = \frac{\partial y_i}{\partial (\)}|_{\alpha_n \beta_n}$$

Now, calculate the values of the observations and the partial derivatives using the initial estimates of the state (α_n, β_n) from above. Thus, you can write the matrix equation as

$$\mathbf{A}^T \begin{bmatrix} y_{o_1} \\ y_{o_2} \\ \vdots \\ y_{o_N} \end{bmatrix} - \mathbf{A}^T \begin{bmatrix} y_{n_1} + \Delta\alpha \frac{\partial y_{n_1}}{\partial \alpha} + \Delta\beta \frac{\partial y_{n_1}}{\partial \beta} \\ y_{n_2} + \Delta\alpha \frac{\partial y_{n_2}}{\partial \alpha} + \Delta\beta \frac{\partial y_{n_2}}{\partial \beta} \\ \vdots \\ y_{n_N} + \Delta\alpha \frac{\partial y_{n_N}}{\partial \alpha} + \Delta\beta \frac{\partial y_{n_N}}{\partial \beta} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

where

$$\frac{\partial y_{n_i}}{\partial \alpha} = \sin(x_{o_i} + \beta_n) \quad \frac{\partial y_{n_i}}{\partial \beta} = \alpha_n \cos(x_{o_i} + \beta_n)$$

Notice, in the linear case, the partial-derivative matrix contained partial derivatives of the residuals, $\partial r_i / \partial \alpha$, but in this case, it contains partials of the measurements, $\partial y_{n_i} / \partial \alpha$. Apply the distributive law and separate the state parameters:

$$A^T \begin{bmatrix} y_{o_1} - y_{n_1} \\ y_{o_2} - y_{n_2} \\ \vdots \\ y_{o_N} - y_{n_N} \end{bmatrix} - A^T \begin{bmatrix} \frac{\partial y_{n_1}}{\partial \alpha} & \frac{\partial y_{n_1}}{\partial \beta} \\ \frac{\partial y_{n_2}}{\partial \alpha} & \frac{\partial y_{n_2}}{\partial \beta} \\ \vdots & \vdots \\ \frac{\partial y_{n_N}}{\partial \alpha} & \frac{\partial y_{n_N}}{\partial \beta} \end{bmatrix} \begin{bmatrix} \Delta \alpha \\ \Delta \beta \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Consider some interesting features of this equation. The matrix containing $y_{o_i} - y_{n_i}$ looks like the b matrix, except it contains differences between the measured and nominal y values. We'll call it a *residual* matrix but give it a slightly different symbol \tilde{b} . The matrix containing the observation partials is the transpose of A^T , so it's the A matrix. We evaluate the A matrix using the nominal state vector. The matrix containing $\Delta\alpha$ and $\Delta\beta$ corresponds to the \hat{X} matrix, except it's now the *corrections* to α and β . We'll call it $\delta\hat{x}$. Substituting the newly defined matrices gives us

$$A^T \tilde{b} - A^T A \delta\hat{x} = 0$$

and the corrections to the state are

$$\delta\hat{x} = (A^T A)^{-1} A^T \tilde{b} \quad (9-9)$$

We can keep the matrices small by applying the summation principles as in the linear case. Eq. (9-9) is identical to the linear least-squares equation, except that

1. A , \tilde{b} , and $\delta\hat{x}$ are defined in the derivation
2. A , A^T , and \tilde{b} contain α_n and β_n (the previous estimates)
3. It's an approximate solution because we've used a truncated Taylor series.

Points (2) and (3) imply the potential to iterate to improve the estimate whereas in the linear case, we obtained solutions to α and β directly. The steps for the most general form of differential correction using Gaussian least squares are:

1. Pick an initial nominal state (α_n and β_n in this case). This is important because, if we're to find a minimum, the initial nominal value must be near

the desired minimum value. Otherwise, the iteration may diverge or, in some cases, converge on an incorrect value. Consider a simple quartic function having four real roots near two relative maxima. Locating a specific root is difficult unless the initial guess is very close to the desired root. In some cases, you may converge on an incorrect solution. The *radius of convergence* is the range, or *interval*, of estimated parameter values which will yield the desired answer. Initial guesses outside the interval won't converge on the desired answer.

2. Compute y_{n_i} corresponding to each x_{o_i}
3. Compute each residual $\bar{r}_i = y_{o_i} - y_{n_i}$
4. Compute each partial derivative, $\partial y_{n_i} / \partial \alpha$ and $\partial y_{n_i} / \partial \beta$, using α_n, β_n
5. Form A, A^T and \tilde{b}
6. Solve for $\Delta\alpha$ and $\Delta\beta$
7. Find $\alpha_{n_{new}} = \alpha_{n_{old}} + \Delta\alpha$ and $\beta_{n_{new}} = \beta_{n_{old}} + \Delta\beta$
8. If you meet a stopping criterion, quit. Otherwise, go to step (2).

Because we're correcting the state a bit on each iteration, this process is termed *differential correction*.

The last step is knowing when to quit. Remember, our basic goal is to minimize the sum of the residuals squared. Thus, it makes sense to use this sum as our criterion for stopping. We quit when the RMS of the residuals stops changing, within a certain tolerance:

$$\text{RMS} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N r_i^2} = \sqrt{\frac{\tilde{b}^T \tilde{b}}{N-1}} \quad (9-10)$$

Notice we don't use the covariance matrix as an exit criterion. One may think that, as confidence in the solution grows, the process should stop. However, the covariance isn't a good measure because we set up the process by minimizing the residuals, not the covariance. The standard deviations should decrease but don't always tend toward zero. Setting up a percentage for the RMS, we quit when

$$\left| \frac{\text{RMS}_{old} - \text{RMS}_{new}}{\text{RMS}_{old}} \right| \leq \epsilon \quad (9-11)$$

where the convergence tolerance, ϵ , is a function of the problem's nonlinearity and the measurement error. Program the test condition with care— RMS_{old} could be identically zero. We seldom require more than six or seven iterations to achieve convergence if a satisfactory initial estimate is available and if we determine the derivatives correctly.

We must also address the weighting problem. Weighting appears in the solution of the nonlinear problem exactly as it does in the linear case. The differential-correction equation for nonlinear, weighted least-squares becomes

$$\delta\hat{x} = (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W} \tilde{\mathbf{b}} = \mathbf{P} \mathbf{A}^T \mathbf{W} \tilde{\mathbf{b}} \quad (9-12)$$

where again, for our state-space dimension of 2, with N observations,

$\mathbf{A} \Rightarrow$ partial-derivative matrix ($N \times 2$)

$\mathbf{A}^T \Rightarrow$ \mathbf{A} transpose ($2 \times N$)

$\delta\hat{x} \Rightarrow$ corrections to the state vector (2×1)

$\tilde{\mathbf{b}} \Rightarrow$ residual matrix ($N \times 1$)

$\mathbf{W} \Rightarrow$ weighting matrix ($N \times N$)

$\mathbf{P} = (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \Rightarrow$ covariance matrix (2×2)

Error analysis (Sec. 9.3.1) is the same as for the linear problem. The covariance matrix $(\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1}$ contains the variances and covariances of the solution parameters (state space) and indicates a measure of observability of the parameters from the available data types, as well as the effect of noisy data on the estimated parameters. It is independent from the actual data. Again, these statistical values will be accurate *only* if our choice of models matches reality and the calibration values are correct. Our choice of models depends on our understanding of the *physics of the problem*. It's also significant that the region within which the implicit linearity assumption holds must include the region of uncertainty (i.e., the $\pm 3\sigma$ region). Otherwise, the covariance matrix may not represent the actual errors. Although least squares is a powerful mathematical tool, it's helpless in the face of a poor physical understanding of the problem.

In astrodynamics, covariance analysis, sometimes called error analysis, is used to evaluate how the type, amount, quality, and distribution of observation data affect the performance of the estimation process. This is often done as part of mission planning to determine required sensor assets and models to use in data reduction and planning, for mission support. Actual data isn't needed—just observation models and partial derivatives, noise characteristics, and assumed tracking schedules and sensor locations.

Example 9-3 solves a problem with a nonlinear measurement-state relationship, so

$$y_{c_i} = \alpha x_{o_i}^\beta$$

Note how to obtain the initial nominal values in this problem.

▼ Example 9-3. Using Nonlinear Least Squares.

GIVEN:

x_{o_i}	1	2	3	4
y_{o_i}	2.5	8.0	19.0	50.0

If we assume the data above represents a function [a quick plot suggests rapid divergence of $y(x)$ which a parabola may fit],

$$y_{c_i} = \alpha x^{\beta}_{o_i}$$

As a first guess, assume the curve exactly matches the largest two pairs of data. Use the last two points because they're farther apart and may better represent the coefficients. Thus,

$$19.0 = \alpha_n (3)^{\beta_n} \quad 50.0 = \alpha_n (4)^{\beta_n}$$

By dividing these two equations, you get $2.63 = 1.333^{\beta}$, so β is $\ln 2.63 / \ln 1.333 = 3.363$. Either equation will then yield a solution for α . The simultaneous solution yields

$$\alpha_n = 0.4721 \quad \beta_n = 3.3634$$

The assumed nonlinear form of solution implies it will require differential correction, and you won't be able conveniently to break apart the state vector in the matrix relations. The nominal values are already determined from the assumed dynamics:

$$y_{n_i} = \alpha_n (x_{o_i})^{\beta_n}$$

Fortunately, analysis easily gives you the partial derivatives:

$$\frac{\partial y_{n_i}}{\partial \alpha_n} = (x_{o_i})^{\beta_n} \quad \frac{\partial y_{n_i}}{\partial \beta_n} = \alpha_n (x_{o_i})^{\beta_n} (\ln x_{o_i})$$

Solution rests in solving the equation

$$\delta \hat{x} = (A^T W A)^{-1} A^T W \bar{b} \text{ or } (A^T A)^{-1} A^T \bar{b}$$

Notice we've assumed the weights are unity values. It's important to note that for equally weighted measurements, this solution simplifies to the linear form of Eq. (9-2), although we still have corrections to the state, rather than the complete state. The $A^T A$ matrix is found by summing the values as shown. For example, the 1,1 component is $(1)^2 + (10.2915)^2 + 40.2476^2 + (105.9147)^2 = 12,944.71$.

$$A^T A = \begin{bmatrix} \sum_{i=1}^4 \left(\frac{\partial y_{n_i}}{\partial \alpha_n} \right)^2 & \sum_{i=1}^4 \frac{\partial y_{n_i}}{\partial \alpha_n} \frac{\partial y_{n_i}}{\partial \beta_n} \\ \sum_{i=1}^4 \frac{\partial y_{n_i}}{\partial \alpha_n} \frac{\partial y_{n_i}}{\partial \beta_n} & \sum_{i=1}^4 \left(\frac{\partial y_{n_i}}{\partial \beta_n} \right)^2 \end{bmatrix} = \begin{bmatrix} 12,944.71 & 8216.22 \\ 8216.22 & 5251.58 \end{bmatrix}$$

$$(A^T A)^{-1} = \begin{bmatrix} 0.011\ 08 & -0.017\ 34 \\ -0.017\ 34 & 0.027\ 31 \end{bmatrix}$$

Now, let's find the $A^T \bar{b}$ matrix using the summation process. The 1,1 component is $1(2.5 - 0.4721) + 10.2915(8 - 4.8587) + 0 + 0 = 34.3598$. The last two terms are zero because we used them to find the nominal value.

$$A^T \bar{b} = \begin{bmatrix} \sum_{i=1}^4 \frac{\partial y_{n_i}}{\partial \alpha_n} (y_{o_i} - y_{n_i}) \\ \sum_{i=1}^4 \frac{\partial y_{n_i}}{\partial \beta_n} (y_{o_i} - y_{n_i}) \end{bmatrix} = \begin{bmatrix} 34.3598 \\ 10.5796 \end{bmatrix}$$

Putting everything together gives us

$$\delta \hat{x} = \begin{bmatrix} \Delta\alpha \\ \Delta\beta \end{bmatrix} = \begin{bmatrix} 0.01108 & -0.01734 \\ -0.01734 & 0.02731 \end{bmatrix} \begin{bmatrix} 33.9071 \\ 10.2919 \end{bmatrix} = \begin{bmatrix} 0.1973 \\ -0.3067 \end{bmatrix}$$

Therefore, the updated values for the state are

$$\alpha_{new} = 0.4721 + 0.1973 = 0.6692$$

$$\beta_{new} = 3.3634 - 0.3067 = 3.0567$$

$$RMS_{new} = \sqrt{\frac{(2.0279)^2 + (3.1416)^2 + (0)^2 + (0)^2}{3}} = 2.1589$$

After the second iteration,

$$\delta \hat{x} = \begin{bmatrix} \Delta\alpha \\ \Delta\beta \end{bmatrix} = \begin{bmatrix} 0.0633 \\ -0.0180 \end{bmatrix}, \hat{X} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 0.7327 \\ 3.0387 \end{bmatrix}$$

The new $RMS_{new} = 2.7498$.

Checking the percentage of change of the RMS, we find

$$\Delta_{RMS} = \left| \frac{2.1589 - 2.7498}{2.1589} \right| = 0.2737$$

We haven't converged, so we need another iteration. After the third iteration,

$$\delta \hat{x} = \begin{bmatrix} \Delta\alpha \\ \Delta\beta \end{bmatrix} = \begin{bmatrix} 0.0021 \\ -0.0008 \end{bmatrix}, \hat{X} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 0.7348 \\ 3.0379 \end{bmatrix}$$

$$RMS_{new} = 1.8267 \text{ and } \Delta_{RMS} = \frac{2.7498 - 1.8267}{2.7498} = 0.3357$$

Notice that, although the RMS_{new} has gone down, the percentage change is still large, so we still haven't converged on the solution. After the fourth iteration,

$$\delta \hat{x} = \begin{bmatrix} \Delta\alpha \\ \Delta\beta \end{bmatrix} = \begin{bmatrix} 0.00001 \\ -0.00001 \end{bmatrix}$$

$$\hat{X} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 0.7349 \\ 3.0378 \end{bmatrix}$$

$$RMS_{new} = 1.8258 \text{ and } \Delta_{RMS} = \frac{1.8267 - 1.8258}{1.8267} = 0.00049$$

We've converged on the solution. The best estimate of the trajectory is

▲ $y_i = 0.7349 (x_i)^{3.0378}$

9.5 Application: Orbit Determination With Differential Correction

The differential-correction technique described previously (nonlinear least squares) is a powerful tool which can accurately estimate an orbit's state from measurements of the motion made by radar, optical, or other data sources. A minimum of six elements completely define the orbit. The astrodynamics community typically uses the six elements that define the orbit and augments them with solve-for parameters. The *solve-for* parameters allow us to solve for dynamical and measurement model parameters like the drag coeffi-

cient and measurement biases. In estimation, *state space* refers to the set six elements and all the solve-for parameters. When applying the technique to orbit determination, however, we must handle several new situations:

1. We can choose from several state spaces: $\{a, e, i, \Omega, \omega, \nu\}$, $\{k_e, h_e, n, \lambda_{mean}, \rho_e, q_e\}$, $\{r_I, r_J, r_K, v_I, v_J, v_K\}$, and so on. (Refer to Sec. 2.4.) Each state space completely describes an orbit and is therefore valid. Remember that certain sets of orbital elements have conditions for which at least one element is undefined—or they have *singularities*. There are several nonsingular element sets, including position and velocity vectors. The equinoctial elements also have nonsingular partial derivatives which make them a very attractive choice when designing a complete system. We often include *solve-for* parameters such as the drag coefficient, sensor biases, and station locations, in the state space. We'll use the position and velocity vectors in this development for several reasons, including ease in finding the osculating orbital elements and improved operation for special orbits such as circular and elliptical equatorial orbits.
2. We may have more than one measurement at a time (ρ, β, el , etc.). Remember, β is the azimuth.
3. The measurements are nonlinear, complex functions of the state. Recall Sec. 3.4.3 (**RAZEL**).

Assume a sensor can measure range, azimuth, and elevation at each time, t_i , so that

$$\bar{y}_{o_i} = \begin{bmatrix} \rho_o \\ \beta_o \\ el_o \end{bmatrix} \text{ at times } = t_i$$

To form residuals, we need to compute some predicted measurements from the position and velocity vectors:

$$\bar{y}_{c_i} = \begin{bmatrix} \rho_c \\ \beta_c \\ el_c \end{bmatrix} \text{ at time } = t_i$$

Because \bar{y}_{c_i} is a nonlinear function of the position and velocity vectors, we use a first-order Taylor series (as in the least-squares method). In this approach, we express the computed measurement as a Taylor-series expansion about some nominal point. For this problem, the vector expression at $t = t_i$ is

$$\bar{y}_{c_i} = \bar{y}_{n_i} + \Delta r_I \frac{\partial \bar{y}_{n_i}}{\partial r_I} + \Delta r_J \frac{\partial \bar{y}_{n_i}}{\partial r_J} + \Delta r_K \frac{\partial \bar{y}_{n_i}}{\partial r_K} + \Delta v_I \frac{\partial \bar{y}_{n_i}}{\partial v_I} + \Delta v_J \frac{\partial \bar{y}_{n_i}}{\partial v_J} + \Delta v_K \frac{\partial \bar{y}_{n_i}}{\partial v_K}$$

Notice the $\bar{y}_n = f(r_I, r_J, r_K, v_I, v_J, v_K, t_i)$ are functions of the nominal vector at each observation time. In addition, each position and velocity vector component is at time t_i . The requirement for nominal vectors means that we must determine the initial orbit (Chap. 6, **SITE-TRACK**, **GIBBS**, **LAMBERT**, etc.). Define the residuals, \bar{r}_i , and don't confuse them with position vectors. I'll always use the overbar to indicate residuals.

$$\bar{r}_i = \bar{y}_{o_i} - \bar{y}_{c_i} = \bar{y}_{o_i} - \left(\bar{y}_{n_i} + \Delta r_I \frac{\partial \bar{y}_{n_i}}{\partial r_I} + \dots + \Delta v_K \frac{\partial \bar{y}_{n_i}}{\partial v_K} \right)$$

Assume each measurement is weighted using its appropriate standard deviation. Use Table 9-3 for the noise (standard deviation) estimates.

$$w_\rho = \frac{1}{\sigma_\rho} \quad w_\beta = \frac{1}{\sigma_\beta} \quad w_{el} = \frac{1}{\sigma_{el}}$$

As an aside, we can also subtract out the biases (again from Table 9-3) from each observation if they're reasonably accurate. Then, minimize the cost function

$$J = \sum_{i=1}^N (w_i \bar{r}_i)^T (w_i \bar{r}_i)$$

where you now have three weights for each observation triplet:

$$w_i = \begin{bmatrix} w_\rho & 0 & 0 \\ 0 & w_\beta & 0 \\ 0 & 0 & w_{el} \end{bmatrix}, i = 1 \dots N$$

Set the first derivatives equal to zero. Remember, the two vector equations produce six scalar equations:

$$\sum_{i=1}^N w_i^T w_i \bar{r}_i \frac{\partial \bar{r}_i}{\partial r} = 0 \quad \sum_{i=1}^N w_i^T w_i \bar{r}_i \frac{\partial \bar{r}_i}{\partial v} = 0$$

Notice, because $\frac{\partial \bar{r}_i}{\partial r_I} = -\frac{\partial \bar{y}_{n_i}}{\partial r_I}$, etc., you can put these equations in matrix form and, depending on the number of observations, N :

$$-1 \begin{bmatrix} \frac{\partial \bar{y}_{n_1}}{\partial r_I} & \frac{\partial \bar{y}_{n_2}}{\partial r_I} & \cdots & \frac{\partial \bar{y}_{n_N}}{\partial r_I} \\ \frac{\partial \bar{y}_{n_1}}{\partial r_J} & & \frac{\partial \bar{y}_{n_N}}{\partial r_J} & \\ \vdots & & \vdots & \\ \frac{\partial \bar{y}_{n_1}}{\partial v_K} & \cdots & \cdots & \frac{\partial \bar{y}_{n_N}}{\partial v_K} \end{bmatrix} \begin{bmatrix} w_1^2 \bar{r}_1 \\ w_2^2 \bar{r}_2 \\ \vdots \\ w_N^2 \bar{r}_N \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

If you define the *partial-derivative matrix* as A^T and the weighting matrix, W , as the product of $w_i^T w_i$, the matrix algebra is the same as for the nonlinear least-squares case. Thus, you end up with the same solution as for the nonlinear case [Eq. (9-12)], with the new definitions. I've shown a summation for the \tilde{b} matrix to simplify computational processing. I'll discuss this further in Sec. 9.5.4.

$$\delta \bar{x} = \begin{bmatrix} \Delta r_I \\ \Delta r_J \\ \Delta r_K \\ \Delta v_I \\ \Delta v_J \\ \Delta v_K \end{bmatrix} \quad \tilde{b} \approx \begin{bmatrix} \bar{r}_1 \\ \bar{r}_2 \\ \vdots \\ \bar{r}_N \end{bmatrix} = \sum_{i=1}^N \begin{bmatrix} \rho_{o_i} - \rho_{n_i} \\ \beta_{o_i} - \beta_{n_i} \\ e l_{o_i} - e l_{n_i} \end{bmatrix}$$

9.5.1 Calculating the Partial-Derivative Matrix (A)

In the previous section, we chose to represent each observation as a vector triplet (ρ_i, β_i, el_i) . Therefore, for our state space of position and velocity vectors, the partial-derivative matrix, A , looks like this (I've shown a summation notation for computational efficiency):

$$A = \frac{\partial \text{observations}}{\partial X_o} = \sum_{i=1}^N \begin{bmatrix} \frac{\partial \rho_i}{\partial r_{I_o}} & \frac{\partial \rho_i}{\partial r_{J_o}} & \frac{\partial \rho_i}{\partial r_{K_o}} & \frac{\partial \rho_i}{\partial v_{I_o}} & \frac{\partial \rho_i}{\partial v_{J_o}} & \frac{\partial \rho_i}{\partial v_{K_o}} \\ \frac{\partial \beta_i}{\partial r_{I_o}} & \frac{\partial \beta_i}{\partial r_{J_o}} & \frac{\partial \beta_i}{\partial r_{K_o}} & \frac{\partial \beta_i}{\partial v_{I_o}} & \frac{\partial \beta_i}{\partial v_{J_o}} & \frac{\partial \beta_i}{\partial v_{K_o}} \\ \frac{\partial el_i}{\partial r_{I_o}} & \frac{\partial el_i}{\partial r_{J_o}} & \frac{\partial el_i}{\partial r_{K_o}} & \frac{\partial el_i}{\partial v_{I_o}} & \frac{\partial el_i}{\partial v_{J_o}} & \frac{\partial el_i}{\partial v_{K_o}} \end{bmatrix} \quad (9-13)$$

The position and velocity vectors are the nominal state at epoch t_o ; thus, these partials relate quantities at many different times to a state vector at one *epoch* time. In essence, then, the A matrix describes how changes in the initial state (position and velocity vectors) affect the observations (measurements). These are sometimes called sensitivity partial derivatives. Often the largest contribution to these changes is the satellite's periodic motion as it moves through its orbit. In high-precision studies, we may need to include how perturbations affect this overall motion.

How do we compute these nonlinear, time-dependent derivatives? The first step is to break up the partial derivative using the chain rule:

$$A = \frac{\partial \text{observations}}{\partial \hat{X}_o} = \frac{\partial \text{observations}}{\partial \hat{X}} \frac{\partial \hat{X}}{\partial \hat{X}_o} = H\Phi \quad (9-14)$$

Eq. (9-14) is an important relation because it distinguishes the observation partial derivatives from the partial derivative of the state over time. This latter matrix is called the matrix of *variational equations*, or the *state-transition matrix*, Φ . There are subtle differences in the formulation of the variational equations and the Φ matrix so we'll use the variational notation for least-squares applications, and reserve the Φ notation for filter applications. They both relate the variation in the state at time, t , \bar{X} , to variations in the state at time t_o , X_o . I'll use a subscript "o" for the epoch time only to simplify the notation. I also don't specify the time dependance (t) with the state vector for clarity. We calculate the two matrices in Eq. (9-14) using analytical or numerical techniques, or by finite differencing. We'll discuss each option shortly.

We can also numerically approximate the complete A matrix [Eq. (9-14)]; it's fairly easy but only approximate. The numerical process is called *finite differencing* because we take small differences of the state to determine their effect on the system. The nominal state at epoch is used to generate position and velocity vectors at the times of the observations. Six additional trajectories are determined as we vary each component of the position and velocity vectors. When the observations are found from both trajectories, we have an approximation to the partial derivative. That is, we take the partial derivative of the observations (at the observation times) with respect to the state at the epoch time. It's best to modify each state element with a percentage of that element rather than a fixed delta when using position and velocity vectors in the state vector. This technique eliminates the sensitivity to large and small values within the state vector. But it's really not appropriate for the semimajor axis because a very small change produces a large effect on the motion due to the change in the mean motion. In this case, it's best to modify a by a few or several tens of meters to control the overall effect on the orbit. Remember to be consistent in calculating the A matrix; i.e., the order of differences must be correct. If you determine each change to the state vector as the modified state minus the nominal state,

$$\delta_i = \hat{X}_{mod_i} - \hat{X}_{nom_i}$$

you must also calculate each observation as the modified observation minus the nominal observation. The approximation for each observation and element of the state becomes

$$\frac{\partial \text{observations}}{\partial \hat{X}_o} \approx \frac{\text{obs}_{mod} - \text{obs}_{nom}}{\delta_i}$$

Another approach, ***central differencing***, takes both positive and negative deviations from the nominal state vector to determine the approximation:

$$\frac{\partial \text{observations}}{\partial \hat{X}_o} \approx \frac{f(\hat{X} + \delta_i) - f(\hat{X} - \delta_i)}{2\delta_i} \quad (9-15)$$

This method ensures a more accurate estimate of the derivative at the desired time but requires more computational time because we need twelve trajectories (for a six-parameter state space). I've shown an example of the propagations of the state at epoch and an observation time in Fig. 9-7.

In algorithm form, finite differencing looks like this:

FOR $j = 1$ to the number of states

$$\textbf{PKEPLER}(\vec{r}_{nom}, \vec{v}_{nom}, \Delta t \Rightarrow \vec{r}_i, \vec{v}_i)$$

$$\textbf{RAZEL}(\vec{r}_i, \vec{v}_i, \vec{r}_{site_i}, \phi_{gd}, \theta_{LST_i} \Rightarrow \rho_n, \beta_n, el_n, \dot{\rho}_n, \dot{\beta}_n, \dot{el}_n)$$

Modify each component of the nominal state vector

$$\delta_j = \hat{X}_{nom_j} (1.01) \text{ modify by } 1\% \text{ of the original value}$$

$$\hat{X}_{mod_j} = \hat{X}_{nom_j} + \delta_j$$

$$\textbf{PKEPLER}(\vec{r}_{mod}, \vec{v}_{mod}, \Delta t \Rightarrow \vec{r}_i, \vec{v}_i)$$

$$\textbf{RAZEL}(\vec{r}_i, \vec{v}_i, \vec{r}_{site_i}, \phi_{gd}, \theta_{LST_i} \Rightarrow \rho_m, \beta_m, el_m, \dot{\rho}_m, \dot{\beta}_m, \dot{el}_m)$$

Find elements of the A matrix for each observed value (i.e., ρ , β , el):

$$\frac{\partial \text{obs}}{\partial \text{component}} \approx \frac{\delta \text{obs}_j}{\delta_j} = \frac{\text{obs}_{mod} - \text{obs}_{nom}}{\delta_j}$$

Reset the modified component \hat{X}_{mod_j} to its original value \hat{X}_{nom_j} .

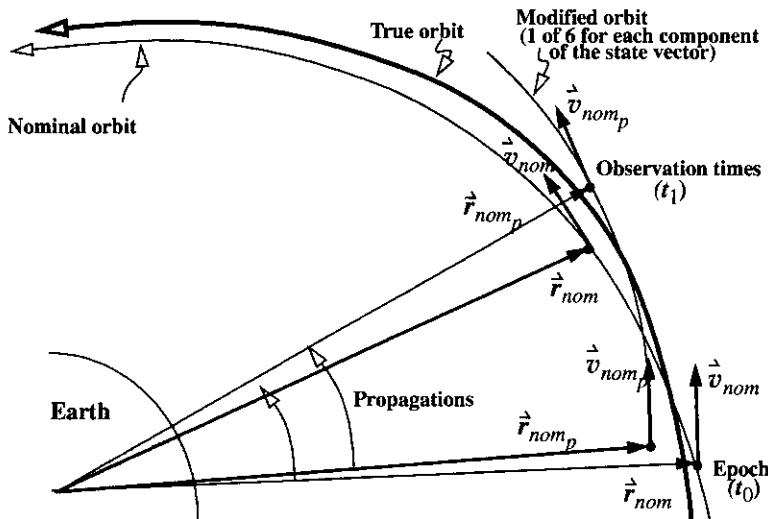


Figure 9-7. Propagations Required to Calculate the A Matrix Numerically. We find the partial derivative of the observations with respect to the state by modifying the initial state, propagating the nominal and modified states to each observation time and finding the difference of the observations. This figure shows one movement of the nominal and modified state vectors to an observation time. Although any epoch time may be chosen, we usually choose either the beginning or the end of a batch of data.

9.5.2 Calculating the Observation Partial Derivatives

For higher precision, the most common way to evaluate the partial derivatives of the observations with respect to the state is to do it analytically. This calculation is important because we'll use the same operation in the Kalman filter. This matrix, H , is called the *observation partials matrix*, or sometimes the *sensitivity matrix* because it indicates how the observations are affected by changes in the state. As an example, consider expressions for range, azimuth, and elevation. Recalling the calculations of Algorithm 15 (**RAZEL**), we'll have trouble writing a single expression relating these changes. Let's begin by writing the slant-range vector in terms of the *IJK* position vector:

$$\begin{bmatrix} \rho_S \\ \rho_E \\ \rho_Z \end{bmatrix} = \begin{bmatrix} r_I \sin(\phi_{gd}) \cos(\theta_{LST}) + r_J \sin(\phi_{gd}) \sin(\theta_{LST}) - r_K \cos(\phi_{gd}) \\ -r_I \sin(\theta_{LST}) + r_J \cos(\theta_{LST}) \\ r_I \cos(\phi_{gd}) \cos(\theta_{LST}) + r_J \sin(\phi_{gd}) \cos(\phi_{gd}) + r_K \sin(\phi_{gd}) \end{bmatrix}$$

We find the observation quantities as in Algorithm 15:

$$\rho = |\vec{r}_{IJK} - \vec{r}_{siteIJK}| \quad \cos(\beta) = \frac{-\rho_S}{\sqrt{\rho_S^2 + \rho_E^2}} \quad \sin(el) = \frac{\rho_Z}{\rho}$$

The partial derivatives for one observation ρ are relatively easy:

$$\frac{\partial \rho}{\partial r_I} = \frac{r_I}{r}$$

Partial derivatives for azimuth and elevation are more difficult due to trigonometric terms and additional components in the slant-range vector. I won't list individual formulas because they depend on specific observation models. See Long et al. (1989, Chaps 4 and 7) and Shaver (1980) for detailed equations.

9.5.3 Calculating the Variational Equations

When we find the observation partial derivatives separately, we need a more rigorous approach to the variational equations. We then combine the two matrices to find the partial-derivative matrix [Eq. (9-14)]. The *variational equations* are the partial derivatives of the state at time t with respect to the epoch state at t_o , $\partial \hat{X} / \partial \hat{X}_o$, and we can calculate them analytically, by numerical integration, or by using finite differences. (I won't detail finite differences because the concept is the same as discussed in Sec. 9.5.1.) Analytical methods assume analytical expressions for the state as functions of the time and epoch elements. The partial derivatives are obtained from these expressions. For solution, analytical methods assume limited force models, usually restricted to two-body motion or two-body plus the J_2 oblateness perturbation.

Consider the force model limited to two-body motion. If we use cartesian variables and consider variations only in the six epoch state parameters \dot{r}_o, \dot{v}_o , the variational equations become

$$\frac{\partial \hat{X}}{\partial \hat{X}_o} = \begin{bmatrix} \frac{\partial \hat{r}}{\partial r_o} & \frac{\partial \hat{r}}{\partial v_o} \\ \frac{\partial \hat{r}_o}{\partial r_o} & \frac{\partial \hat{v}_o}{\partial v_o} \\ \frac{\partial \hat{v}}{\partial r_o} & \frac{\partial \hat{v}}{\partial v_o} \\ \frac{\partial \hat{v}_o}{\partial r_o} & \frac{\partial \hat{v}_o}{\partial v_o} \end{bmatrix}$$

where each of the elements of the above matrix are themselves 3×3 matrices, e.g.,

$$\frac{\partial \hat{r}}{\partial r_o} = \begin{bmatrix} \frac{\partial r_I}{\partial r_{I_o}} & \frac{\partial r_I}{\partial r_{J_o}} & \frac{\partial r_I}{\partial r_{K_o}} \\ \frac{\partial r_J}{\partial r_{I_o}} & \frac{\partial r_J}{\partial r_{J_o}} & \frac{\partial r_J}{\partial r_{K_o}} \\ \frac{\partial r_K}{\partial r_{I_o}} & \frac{\partial r_K}{\partial r_{J_o}} & \frac{\partial r_K}{\partial r_{K_o}} \end{bmatrix}$$

The analytical solution is provided by the Taylor series used in the classical f and g series (see Sec. 4.3.1 on page 254). Thus,

$$\begin{aligned}\hat{\vec{r}}(t) &= \hat{\vec{r}}_o + \hat{\vec{v}}_o(t) \Delta t + \frac{1}{2} \frac{d\hat{v}}{dt} \Big|_{t=t_o} \Delta t^3 + \dots \\ \hat{\vec{v}}(t) &= \hat{\vec{v}}_o + \frac{d\hat{v}}{dt} \Big|_{t=t_o} \Delta t + \dots\end{aligned}$$

Ignoring higher order terms, the variational equations are given by

$$\left. \frac{\partial \hat{X}}{\partial \hat{X}_o} \right|_{\text{2-body}} = \begin{bmatrix} 1 + \frac{3\mu\Delta t^2 r_I}{2r_o^5} & \frac{3\mu\Delta t^2 r_I}{2r_o^5} & \frac{3\mu\Delta t^2 r_I}{2r_o^5} & \Delta t & 0 & 0 \\ \frac{3\mu\Delta t^2 r_I}{2r_o^5} & 1 + \frac{3\mu\Delta t^2 r_J}{2r_o^5} & \frac{3\mu\Delta t^2 r_I}{2r_o^5} & 0 & \Delta t & 0 \\ \frac{3\mu\Delta t^2 r_I}{2r_o^5} & \frac{3\mu\Delta t^2 r_I}{2r_o^5} & 1 + \frac{3\mu\Delta t^2 r_K}{2r_o^5} & 0 & 0 & \Delta t \\ \frac{3\mu\Delta t^2 r_I}{r_o^5} & 0 & 0 & 1 & 0 & 0 \\ 0 & \frac{3\mu\Delta t^2 r_J}{r_o^5} & 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{3\mu\Delta t^2 r_K}{r_o^5} & 0 & 0 & 1 \end{bmatrix}$$

The analysis is much simpler using the classical orbital elements, a , e , i , Ω , ω , and $M = nt + M_o$. If \hat{a} represents orbital elements instead of vectors, the variational equations are

$$\frac{\partial \hat{a}}{\partial \hat{a}_o} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ -\frac{3n}{2a} \Delta t & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Analytical representations for the variational equations using two-body and J_2 dynamics are also in the literature; see Shaver (1980, 303–325) for an equinoctial formulation.

These analytical representations are often used for estimation, depending on the particular problem. They provide surprisingly good results—particularly the two-body and J_2 formulations—considering the limited dynamical modeling. This model is sufficient for many applications.

But we sometimes wish to incorporate more dynamics into the variational equations, whether it be a 4×4 or 12×12 gravity field, or some other perturbations such as drag. Also, we may wish to estimate some force-model parameters along with the state such as the drag coefficient (c_D) or possibly gravity-model coefficients. The most accurate technique uses numerical integration to obtain the variational equations. In essence, the partial derivatives of the equations of motion are integrated to yield the result.

Let's assume we're given the state at time \hat{X} and the epoch state \hat{X}_o , possibly augmented by dynamical parameters (c_D , C_{lm} , S_{lm} , for example), which define

$$\hat{X}_o = (\dot{\hat{r}}_o, \dot{\hat{v}}_o, c_D, C_{lm}, S_{lm})$$

We want to obtain $\partial \dot{\hat{X}} / \partial \hat{X}_o$. The equations of motion, $\dot{\hat{X}} = \vec{f}(\dot{\hat{r}}_o, \dot{\hat{v}}_o, c_D, C_{lm}, S_{lm})$, depend on the initial state and the dynamical parameters in the force models. Taking the partial derivatives with respect to the solve-for state vector yields

$$\frac{\partial \dot{\hat{X}}}{\partial \hat{X}_o} = \frac{\partial \vec{f}}{\partial \hat{X}_o}(\dot{\hat{r}}_o, \dot{\hat{v}}_o, c_D, C_{lm}, S_{lm}) \quad (9-16)$$

Rewriting the left-hand side using operator notation and interchanging the order of the operators yields

$$\frac{\partial \dot{\hat{X}}}{\partial \hat{X}_o} = \frac{\partial}{\partial \hat{X}_o} \left[\frac{d}{dt} [\hat{X}] \right] = \frac{d}{dt} \left[\frac{\partial [\hat{X}]}{\partial \hat{X}_o} \right] \quad (9-17)$$

Thus, Eq. (9-17) shows that the partial derivatives with respect to \hat{X}_o of the equations of motion are the same as the time rate of change of the variational equations. Combining Eq. (9-16) and Eq. (9-17) yields a first-order system of differential equations for the variational equations:

$$\frac{d}{dt} \left[\frac{\partial \hat{X}}{\partial \hat{X}_o} \right] = \frac{\partial}{\partial \hat{X}_o} \left(\frac{d \hat{X}}{dt} \right) = \frac{\partial \vec{f}(\dot{\hat{r}}_o, \dot{\hat{v}}_o, c_D, C_{lm}, S_{lm})}{\partial \hat{X}_o} \quad (9-18)$$

This is an $n \times n$ matrix of first-order differential equations where n is the number of solve-for parameters. The equations are numerically integrated, (usually) with the same algorithms used for the equations of motion, and chained with the observation partial derivatives to find the desired partial derivatives for Eq. (9-14). Eq. (9-18) implies six times the number of equations, which typically means six times the computational cost. To reduce the computational burden, the force model used to integrate the variational

equations is often reduced from that used to integrate the equations of motion. As an example, precision orbit determination using a 21×21 gravity model might use a 2×2 or 4×4 field in the variational equations. Using a full gravity model in the variational equations seldom yields significantly better results than those obtained with prudently reduced models, but it does require the most computations.

Let's look at an example to demonstrate the process. We find the equations of motion from Eq. (7-46), but for now, let's use the state notation.

$$\frac{d\hat{\mathbf{X}}}{dt} = \dot{\hat{\mathbf{X}}}_{2-body} + \dot{\hat{\mathbf{X}}}_{nonspherical} + \dot{\hat{\mathbf{X}}}_{drag} + \dot{\hat{\mathbf{X}}}_{3-body} + \dot{\hat{\mathbf{X}}}_{SR} + \dot{\hat{\mathbf{X}}}_{other}$$

Then, the partial derivatives with respect to the epoch state are

$$\frac{\partial}{\partial \hat{\mathbf{X}}_o} \left(\frac{\partial \hat{\mathbf{X}}}{\partial t} \right) = \frac{\partial \dot{\hat{\mathbf{X}}}_{2-body}}{\partial \hat{\mathbf{X}}_o} + \frac{\partial \dot{\hat{\mathbf{X}}}_{nonspherical}}{\partial \hat{\mathbf{X}}_o} + \frac{\partial \dot{\hat{\mathbf{X}}}_{drag}}{\partial \hat{\mathbf{X}}_o} + \dots$$

Remember that our state is composed of position and velocity vectors as well as solve-for parameters, but I don't show any solve-for parameters in this example.

$$\hat{\mathbf{X}}_{2-body} = \begin{bmatrix} \mathbf{r} \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} \mathbf{r}_I \\ \mathbf{r}_J \\ \mathbf{r}_K \\ \mathbf{v}_I \\ \mathbf{v}_J \\ \mathbf{v}_K \end{bmatrix} \quad \text{and} \quad \dot{\hat{\mathbf{X}}}_{2-body} = \begin{bmatrix} \mathbf{v} \\ a \end{bmatrix} = \begin{bmatrix} \mathbf{v}_I \\ \mathbf{v}_J \\ \mathbf{v}_K \\ -\frac{\mu \mathbf{r}_I}{r^3} \\ -\frac{\mu \mathbf{r}_J}{r^3} \\ -\frac{\mu \mathbf{r}_K}{r^3} \end{bmatrix}$$

By Eq. (9-18) we have

$$\frac{\partial}{\partial \hat{\mathbf{X}}_o} \left(\frac{d\hat{\mathbf{X}}}{dt} \right) = \begin{bmatrix} \frac{\partial \mathbf{v}_I}{\partial r_{I_o}} \frac{\partial \mathbf{v}_I}{\partial r_{J_o}} \frac{\partial \mathbf{v}_I}{\partial r_{K_o}} \frac{\partial \mathbf{v}_I}{\partial v_{I_o}} \frac{\partial \mathbf{v}_I}{\partial v_{J_o}} \frac{\partial \mathbf{v}_I}{\partial v_{K_o}} \\ \frac{\partial \mathbf{v}_J}{\partial r_{I_o}} \dots \\ \vdots \\ \frac{\partial \mathbf{a}_K}{\partial r_{I_o}} \frac{\partial \mathbf{a}_K}{\partial r_{J_o}} \frac{\partial \mathbf{a}_K}{\partial r_{K_o}} \frac{\partial \mathbf{a}_K}{\partial v_{I_o}} \frac{\partial \mathbf{a}_K}{\partial v_{J_o}} \frac{\partial \mathbf{a}_K}{\partial v_{K_o}} \end{bmatrix}$$

The two-body contribution is obtained from the partial derivatives of the two-body acceleration, which depends only on the position, r .

$$\frac{\dot{\hat{X}}_{2\text{-body}}}{\partial \hat{X}_o} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ -\frac{\mu}{r^3} + \frac{3\mu r_I^2}{r^5} & \frac{3\mu r_I r_J}{r^5} & \frac{3\mu r_I r_K}{r^5} & 0 & 0 & 0 \\ \frac{3\mu r_I r_J}{r^5} & -\frac{\mu}{r^3} + \frac{3\mu r_J^2}{r^5} & \frac{3\mu r_J r_K}{r^5} & 0 & 0 & 0 \\ \frac{3\mu r_I r_K}{r^5} & \frac{3\mu r_J r_K}{r^5} & -\frac{\mu}{r^3} + \frac{3\mu r_K^2}{r^5} & 0 & 0 & 0 \end{bmatrix}_{t=t_k}$$

The equations of motion due to the nonspherical Earth depend only on the satellite's position through the partial derivatives of the potential function. Thus, the partial derivatives of the acceleration with respect to the velocity vanish. The partial derivatives of the velocity with respect to the position also vanish. Now,

$$\frac{\dot{\hat{X}}_{\text{nonspherical}}}{\partial \hat{X}_o} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ \frac{\partial a_I}{\partial r_{I_o}} \frac{\partial a_I}{\partial r_{J_o}} \frac{\partial a_I}{\partial r_{K_o}} & 0 & 0 & 0 \\ \frac{\partial a_J}{\partial r_{I_o}} \frac{\partial a_J}{\partial r_{J_o}} \frac{\partial a_J}{\partial r_{K_o}} & 0 & 0 & 0 \\ \frac{\partial a_K}{\partial r_{I_o}} \frac{\partial a_K}{\partial r_{J_o}} \frac{\partial a_K}{\partial r_{K_o}} & 0 & 0 & 0 \end{bmatrix}$$

Remember that the acceleration is the gradient of the potential function, so in spherical coordinates of the satellite (r, ϕ_{gc}, λ) ,

$$\dot{\vec{a}} = \frac{\partial U}{\partial \vec{r}} = \frac{\partial U}{\partial r} \frac{\partial \vec{r}}{\partial r} + \frac{\partial U}{\partial \lambda} \frac{\partial \vec{r}}{\partial \lambda} + \frac{\partial U}{\partial \phi_{gc}} \frac{\partial \vec{r}}{\partial \phi_{gc}}$$

Notice we use the chain rule to obtain partial derivatives with respect to cartesian state components. Long et al. (1989, 4-16 to 4-20) show results for partial derivatives of a highly accurate numerical propagator. The gravitational acceleration is determined using Eq. (7-17). Remember that we evaluate the acceleration in a true-of-date system. (U is the aspherical potential *excluding* the two-body term):

$$\begin{aligned}\frac{\partial \vec{a}}{\partial \vec{r}} &= \frac{\partial}{\partial \vec{r}} \left(\frac{\partial U}{\partial r} \right) \frac{\partial r}{\partial \vec{r}} + \frac{\partial}{\partial \vec{r}} \left(\frac{\partial U}{\partial \phi_{gc}} \right) \frac{\partial \phi_{gc}}{\partial \vec{r}} + \frac{\partial}{\partial \vec{r}} \left(\frac{\partial U}{\partial \lambda} \right) \frac{\partial \lambda}{\partial \vec{r}} \\ &\quad + \frac{\partial U}{\partial r} \frac{\partial^2 r}{\partial \vec{r}^2} + \frac{\partial U}{\partial r} \frac{\partial^2 \phi_{gc}}{\partial \vec{r}^2} + \frac{\partial U}{\partial r} \frac{\partial^2 \lambda}{\partial \vec{r}^2}\end{aligned}$$

In matrix notation, the first set of partial derivatives are

$$\frac{\partial}{\partial \vec{r}} \begin{bmatrix} \frac{\partial U}{\partial r} \\ \frac{\partial U}{\partial \phi_{gc}} \\ \frac{\partial U}{\partial \lambda} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 U}{\partial r^2} & \frac{\partial^2 U}{\partial r \partial \phi_{gc}} & \frac{\partial^2 U}{\partial r \partial \lambda} \\ \frac{\partial^2 U}{\partial \phi_{gc} \partial r} & \frac{\partial^2 U}{\partial^2 \phi_{gc}} & \frac{\partial^2 U}{\partial \phi_{gc} \partial \lambda} \\ \frac{\partial^2 U}{\partial \lambda \partial r} & \frac{\partial^2 U}{\partial \lambda \partial \phi_{gc}} & \frac{\partial^2 U}{\partial \lambda^2} \end{bmatrix} \begin{bmatrix} \frac{\partial r}{\partial \vec{r}} \\ \frac{\partial \phi_{gc}}{\partial \vec{r}} \\ \frac{\partial \lambda}{\partial \vec{r}} \end{bmatrix} \quad (9-19)$$

Differentiate Eq. (7-21) (again excluding the two-body term) to find the partial derivatives of the matrix:

$$\frac{\partial^2 U}{\partial r^2} = \frac{\mu}{r^3} \sum_{l=2}^{\infty} \left(\frac{R_{\oplus}}{r} \right)^l (l+2)(l+1) \sum_{m=0}^l P_{lm} [\sin(\phi_{gc})] \{ C_{lm} \cos(m\lambda) + S_{lm} \sin(m\lambda) \}$$

$$\begin{aligned}\frac{\partial^2 U}{\partial r \partial \phi_{gc}} &= \frac{\partial^2 U}{\partial \phi_{gc} \partial r} \\ &= -\frac{\mu}{r^2} \sum_{l=2}^{\infty} \left(\frac{R_{\oplus}}{r} \right)^l (l+1) \sum_{m=0}^l \{ P_{l,m+1} [\sin(\phi_{gc})] - m \tan(\phi_{gc}) P_{l,m} [\sin(\phi_{gc})] \} \\ &\quad \times \{ C_{lm} \cos(m\lambda) + S_{lm} \sin(m\lambda) \}\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 U}{\partial r \partial \lambda} &= \frac{\partial^2 U}{\partial \lambda \partial r} \\ &= -\frac{\mu}{r^3} \sum_{l=2}^{\infty} \left(\frac{R_{\oplus}}{r} \right)^l (l+1) \sum_{m=0}^l m P_{lm} [\sin(\phi_{gc})] \{ S_{lm} \cos(m\lambda) - C_{lm} \sin(m\lambda) \}\end{aligned}$$

$$\frac{\partial^2 U}{\partial \phi_{gc}^2} = \frac{\mu}{r} \sum_{l=2}^{\infty} \left(\frac{R_{\oplus}}{r} \right)^l \sum_{m=0}^l \{ C_{lm} \cos(m\lambda) + S_{lm} \sin(m\lambda) \} [\tan(\phi_{gc}) P_{lm+1} [\sin(\phi_{gc})]] \\ + \{ m^2 \sec^2(\phi_{gc}) - m \tan^2(\phi_{gc}) - l(l+1) P_{lm} [\sin(\phi_{gc})] \}$$

$$\frac{\partial^2 U}{\partial \phi_{gc} \partial \lambda} = \frac{\partial^2 U}{\partial \lambda \partial \phi_{gc}} = \frac{\mu}{r} \sum_{l=2}^{\infty} \left(\frac{R_{\oplus}}{r} \right)^l \sum_{m=0}^l m \{ S_{lm} \cos(m\lambda) - C_{lm} \sin(m\lambda) \} \\ \times \{ P_{lm} [\sin \phi_{gc}] - m \tan(\phi_{gc}) P_{lm} [\sin \phi_{gc}] \}$$

$$\frac{\partial^2 U}{\partial \lambda^2} = -\frac{\mu}{r} \sum_{l=2}^{\infty} \left(\frac{R_{\oplus}}{r} \right)^l \sum_{m=0}^l m^2 P_{lm} [\sin \phi_{gc}] \{ C_{lm} \cos(m\lambda) + S_{lm} \sin(m\lambda) \}$$

The partial derivatives on the right-hand side of Eq. (9-19) are given in Eq. (7-22). Determine the second derivatives using Eq. (7-22):

$$\frac{\partial^2 r}{\partial r^2} = \frac{1}{r} \left[I - \frac{\dot{r} \dot{r}^T}{r^2} \right]$$

$$\frac{\partial^2 \phi_{gc}}{\partial r^2} = -\frac{1}{(r_I^2 + r_J^2)^{3/2}} \left[\left(\frac{\partial r_K}{\partial r} \right)^T - \frac{r_K \dot{r}}{r^2} \right] \left[r_I \left(\frac{\partial r_I}{\partial r} \right) + r_J \left(\frac{\partial r_J}{\partial r} \right) \right] \\ - \frac{1}{r^2 \sqrt{r_I^2 + r_J^2}} \left[r \left(\frac{\partial r_K}{\partial r} \right) + r_K I - \frac{2r_K \dot{r} \dot{r}^T}{r^2} \right]$$

$$\frac{\partial^2 \lambda}{\partial r^2} = -\frac{2}{(r_I^2 + r_J^2)} \begin{bmatrix} -r_J \\ r_I \\ 0 \end{bmatrix} \left[r_I \left(\frac{\partial r_I}{\partial r} \right) + r_J \left(\frac{\partial r_J}{\partial r} \right) \right] + \frac{1}{(r_I^2 + r_J^2)} \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Identify the partial derivatives of the ECEF vector components with respect to the total vector as direction vectors [for example, $\partial r_I / \partial \dot{r} = (1, 0, 0)$].

We find the partial derivatives for drag, third-body, and solar-radiation pressure similarly, but I won't detail the formula here due to their length. We've spent a lot of time discussing how to calculate these matrices because they're a very important part of precise differential correction, and we also need them for Kalman-filtering applications.

9.5.4 Implementing Least-Squares Techniques

The least-squares process consists of trying to minimize a cost function that depends on actual and computed observations that, in turn, depend on dynamics and the initial state. The residuals are the difference between the actual and computed observations. We weight the observations according to their perceived accuracy. The cost function is a scalar function that depends on the state vector through the computed observations. We treat the actual observations as constants. We can think of the cost function as a manifold in an $n+1$ dimensional space with wells corresponding to the minima of the cost function. In this analogy, n is the number of solve-for parameters in the state vector. The initial condition gives us a starting location. The differential correction process uses the partial-derivative matrix as a directional derivative that specifies the direction from the current position on the manifold, to the nearest local minimum. The iterative nature allows us to find the local minimum, although convergence isn't guaranteed.

Developing a generic routine for differential correction isn't trivial because we must take into account many variations:

1. Form of the state vector: position and velocity vectors or orbital elements, and solve-for parameters—usually determined by the specific problem.
2. Interval of convergence, determination of the nominal state vector, and initial estimates
3. Quantity, quality, and availability of observational data
4. Calculation of the partial-derivative matrix: analytical or numerical
5. Choice of propagation routines
6. Ability to consider dynamic modeling errors—*consider parameters*
7. Tolerances and criteria for ending the computation

We've discussed the importance of the interval of convergence to a solution—the nominal orbit is very important. In the case of initial orbit determination, we know nothing about the orbit and we have only observational data. In these cases, the nominal orbit comes from one of the techniques in Chap. 6. Although the **SITE-TRACK** algorithm tends to indicate that we can form position and velocity vectors from each observation, it's much more common to have data that allows us to form only the position vector.

Algorithm 58 shows a technique to form a nominal state vector from observational data. It's simply an averaging technique, in which all the vectors are formed using **HGIBBS**, **GIBBS**, or **LAMBERT**, and then are moved to a common time and divided by the number of observations. Although we show a simplified propagator (**PKEPLER**), we could use any propagator including numerical techniques. Except for cases in which the data spans a large time interval, **PKEPLER** is probably sufficient to form an initial guess. Remember that although Algorithm 58 gives an initial estimate of the state vector, it's more convenient to already have an initial estimate of the satellite's state vector so we can proceed directly. Except for new launches, this is usually the case. The initial value of the

nominal orbit can strongly affect the success of differential correction. Many orbits won't converge without a fairly close initial estimate, especially with limited data.

ALGORITHM 58: NOMINAL STATE ($\phi_{gd}, \lambda, h_{ellp}$ obs, $t_i \Rightarrow \dot{\vec{r}}_{nom}, \dot{\vec{v}}_{nom}$)

LSTIME ($JD_i, UT_i, \lambda \Rightarrow \theta_{LST_i}, \theta_{GST_i}$)

SITE-TRACK ($\phi_{gd}, \theta_{LST_i}, h_{ellp}, \rho_i, \beta_i, el_i \Rightarrow \dot{\vec{r}}_i$) at each t_i

GIBBS ($\dot{\vec{r}}_1, \dot{\vec{r}}_2, \dot{\vec{r}}_3 \Rightarrow \dot{\vec{v}}_{2i}$) at each middle time t_i (or **HGIBBS** or **LAMBERT**)

PKEPLER ($\dot{\vec{r}}_{2i}, \dot{\vec{v}}_{2i}, \Delta t_i \Rightarrow \dot{\vec{r}}_{2io}, \dot{\vec{v}}_{2io}$) at epoch times

Average $\dot{\vec{r}}_{2io}, \dot{\vec{v}}_{2io}$ to find the nominal state vector $\bar{\vec{X}}_{nom} = \begin{bmatrix} \dot{\vec{r}}_{nom} \\ \dot{\vec{v}}_{nom} \end{bmatrix}$

Estimating a state vector depends on the quantity, availability, and type of observational data. In general, we should have many more observations than the number of parameters in the state space. Although techniques exist for calculating the initial orbit using small amounts of data, we need more observations to improve the confidence of the result. Unfortunately, these minimally determined solutions offer no means to reduce the influence of measurement noise on the result or establish any confidence in the answer. Seven basic types or combinations of data are usually available (see Sec. 6.2). The type of data is important because we must find observation partial derivatives for each type. We should also use statistical parameters (noise and bias) from Table 9-3 during the differential correction process for improved accuracy.

Choosing between analytical and numerical partial derivatives evokes a lot of discussion and ultimately rests with the treatment of Eq. (9-14). For highly accurate simulations, we usually determine the observation partial derivatives analytically and evaluate the variational equations either numerically or analytically. Analytical formulations for the variational equations are often more popular than numerical techniques due to the increased speed. That's because analytical partials *rarely* include higher-order gravitational effects beyond J_2 , or any form of solar-radiation pressure, complex drag models, and so on.

Reduced-force models for the variational equations usually suffice because we're only trying to evaluate the sensitivity of the state to establish the direction of our next correction in the iteration. Less accurate partial derivatives usually mean a few more iterations to arrive at the minimum. In general, these comments also apply to the differences between numerical and analytical partial derivative formulations. Thus, complex force models aren't usually needed for the variational equations.

Despite the flexibility for evaluating the partial-derivative matrix, we need our estimation techniques to be consistent with the propagation algorithms that use the final result.

This may seem to contradict the previous discussion for the variational equations, but a subtlety exists. The consistency of propagation routines is important for all *propagations of the state*, or predictions. Remember from Fig. 8-15 that there is a large difference between the prediction performance of propagation routines. For highly accurate simulations, we should use fully configured numerical techniques in the differential-correction program to achieve the best accuracy and then use the same propagator for subsequent predictions.

The choice of propagator is also tied to the accuracy of the observations. A good rule is to use a propagator that's an order of magnitude more accurate than your best observations. We should extract as much information as possible from the observations. Cefola, Fonte, and Shah (1996) show that using approximate analytical propagation methods with moderately accurate observations results in hundreds of meters of error. In cases like this, the differential correction produces a biased state estimate that absorbs the error of the theory and compromises the estimate of the noise present in the observations!

For computer implementations, we can use an accumulation method for calculating $A^T W A$ and $A^T W \tilde{b}$ as in the linear and nonlinear examples. Suppose we have 1000 observations of ρ , β , and el —a common occurrence in orbit estimation. For a six-parameter state space, our A matrix will be 6×3000 , W will be 3000×3000 , and \tilde{b} will be 3000×1 . This imposes significant memory requirements on a program. However, we can evaluate each observation independently and accumulate the result. This process reduces the matrix sizes to 6×3 , 3×3 , and 3×1 respectively—a significant savings!

Finally, the most difficult question may be “when to quit”. Because orbit determination isn't linear, we slightly change the orbital elements for each step and iterate until the residuals stop changing. However, the magnitude of each residual, $\Delta\rho$, $\Delta\beta$, and Δel , may differ vastly from the others; therefore, we must do some scaling to keep the largest ones from controlling the solution. We can do this with the weights (W_i matrices); thus, if

$$\text{RMS}_{new} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N w_i^2 r_i^2}$$

we define the weighted RMS as

$$\text{RMS}_{new} = \sqrt{\frac{1}{N-1} \frac{\sum_{i=1}^N (\Delta\rho_i)^2}{\sigma_\rho^2} + \frac{\sum_{i=1}^N (\Delta\beta_i)^2}{3} + \frac{\sum_{i=1}^N (\Delta el_i)^2}{\sigma_{el}^2}} = \sqrt{\frac{\tilde{b}^T W \tilde{b}}{n_{obs}(N-1)}}$$

We've added the n_{obs} to this expression to account for each type of measurement associated with an observation (for example $n_{obs} = 3$ as shown). We stop iterating the differential correction whenever the change in the weighted RMS becomes small, as in Eq. (9-11). Remember our earlier discussion about using standard deviations. Although we use the

RMS to end the iterations, the standard deviation is crucial to interpreting the results. We can adjust the RMS tolerance for convergence to meet the needs of a particular problem. Because differential correction is inherently noisy, unrealistic tolerances (say 1×10^{-8}) may actually cause the iterations to become unstable. A percentage change of the RMS values will usually eliminate this difficulty when coupled with a limit of, say, 1×10^{-3} .

The description in Algorithm 59 is valid for a number of different applications. I've shown a general propagation technique in the algorithm so you'll remember to use whatever propagation technique you've planned to use after the differential correction is complete. Finally, as with the **HGIBBS** method, you should be careful to preserve accuracy when using the *JD* as the time argument. The Modified Julian Date may be more accurate.

ALGORITHM 59: Differential Correction (Obs at t_i , $\bar{X}_{nom} \Rightarrow \hat{X}_o$)

FOR $i = 1$ to the number of observations

Propagate the nominal state to each t_i and find nominal observations

$$\bar{X}_{nom_i} = \int_{t_k}^{t_{k+1}} \dot{\bar{X}}_{nom} dt + \bar{X}_{nom}$$

$$RAZEL (\dot{\bar{r}}_{nom_i}, \dot{\bar{v}}_{nom_i}, \dot{\bar{r}}_{site_i}, \phi_{gd}, \theta_{LST_i} \Rightarrow \rho_n, \beta_n, el_n, \dot{\rho}_n, \dot{\beta}_n, \dot{el}_n)$$

Find the \tilde{b} matrix as observed minus nominal data at each t_i

$$\frac{d\hat{X}}{dt} = \dot{\bar{X}}_{2-body} + \dot{\bar{X}}_{nonspherical} + \dot{\bar{X}}_{drag} + \dot{\bar{X}}_{3-body} + \dot{\bar{X}}_{SR} + \dot{\bar{X}}_{other}$$

$$A = \frac{\partial \text{observations}}{\partial \hat{X}} \frac{\partial \hat{X}}{\partial \hat{X}_o} = H \frac{\partial}{\partial \hat{X}_o} \left[\frac{d\hat{X}}{dt} \right]$$

$$\delta\hat{x} = (A^TWA)^{-1}A^TW\tilde{b} = PA^TW\tilde{b}$$

Check RMS for convergence.

Update the state vector and repeat if not converged:

$$\hat{X}_{nom} = \hat{X}_{nom} + \delta\hat{x}$$

Here's an example to compare other techniques.

▼ **Example 9-4. Using Differential Correction for Least-Squares Orbit Determination.**

GIVEN: Observations from Kaena Point on January 29, 1995 for GEOS-III (see Table 9-5).

FIND $\hat{\mathbf{X}}, \hat{\mathbf{P}}$ at an epoch at the first observation, using the first ten observations.

TABLE 9-5. Example Observations. Data for GEOS-III (#7734) is from Kaena Point, Hawaii.

Sat	Yr Mo D	Time (UTC)	Range (km)	Az (°)	EI (°)
7734	1995 1 29	02:38:37.000	2047.502 00	60.4991	16.1932
7734	1995 1 29	02:38:49.000	1984.677 00	62.1435	17.2761
7734	1995 1 29	02:39:02.000	1918.489 00	64.0566	18.5515
7734	1995 1 29	02:39:14.000	1859.320 00	65.8882	19.7261
7734	1995 1 29	02:39:26.000	1802.186 00	67.9320	20.9351
7734	1995 1 29	02:39:38.000	1747.290 00	70.1187	22.1319
7734	1995 1 29	02:39:50.000	1694.891 00	72.5159	23.3891
7734	1995 1 29	02:40:03.000	1641.201 00	75.3066	24.7484
7734	1995 1 29	02:40:15.000	1594.770 00	78.1000	25.9799
7734	1995 1 29	02:40:27.000	1551.640 00	81.1197	27.1896
7734	1995 1 29	02:40:39.000	1512.085 00	84.3708	28.3560
7734	1995 1 29	02:40:51.000	1476.415 00	87.8618	29.4884
7734	1995 1 29	02:41:03.000	1444.915 00	91.5955	30.5167
7734	1995 1 29	02:41:15.000	1417.880 00	95.5524	31.4474
7734	1995 1 29	02:41:27.000	1395.563 00	99.7329	32.2425
7734	1995 1 29	02:41:39.000	1378.202 00	104.0882	32.8791
7734	1995 1 29	02:41:51.000	1366.010 00	108.6635	33.3788
7734	1995 1 29	02:42:03.000	1359.100 00	113.2254	33.5998

The raw observations will usually arrive in a condensed format which you can expand. Be sure to get the proper information to do so. We're using only the first ten points for this problem, but we'll use the remaining data with future examples. Because the process is very long when each step is shown, I'll show matrices only for the first iteration, include subsequent iterations with minimal information, and discuss the final answer to point out important results. If we use Algorithm 58, the first ten observations (at the epoch time) give us

$$\bar{\mathbf{X}}_{nom}^T = [0.9020 \ 0.4192 \ 0.5394 \ 0.5469 \ -0.2433 \ -0.7248] \text{ (ER, ER/TU)}$$

We'll use finite differencing to find the partial-derivative matrix because this is just an example. The first iteration using $N = 10$ observations produces accumulated matrices as

$$\begin{aligned}
 A^T W A &= \sum_{i=1}^{10} \begin{bmatrix} A^T \\ 6 \times 3 \end{bmatrix} \begin{bmatrix} W \\ 3 \times 3 \end{bmatrix} \begin{bmatrix} A \\ 6 \times 6 \end{bmatrix} \\
 &= \begin{bmatrix} 2.918 \times 10^9 & 3.914 \times 10^9 & 1.937 \times 10^9 & 2.412 \times 10^8 & 4.300 \times 10^8 & 1.927 \times 10^8 \\ 3.914 \times 10^9 & 3.797 \times 10^{10} & 1.823 \times 10^{10} & 3.748 \times 10^8 & 2.698 \times 10^9 & 1.096 \times 10^9 \\ 1.937 \times 10^9 & 1.823 \times 10^{10} & 1.047 \times 10^{10} & 1.748 \times 10^8 & 1.127 \times 10^9 & 5.773 \times 10^8 \\ 2.412 \times 10^8 & 3.748 \times 10^8 & 1.748 \times 10^8 & 2.475 \times 10^7 & 4.271 \times 10^7 & 1.854 \times 10^7 \\ 4.300 \times 10^8 & 2.698 \times 10^9 & 1.127 \times 10^9 & 4.271 \times 10^7 & 2.630 \times 10^8 & 9.930 \times 10^7 \\ 1.927 \times 10^8 & 1.096 \times 10^9 & 5.773 \times 10^8 & 1.854 \times 10^7 & 9.930 \times 10^7 & 4.898 \times 10^7 \end{bmatrix} \\
 A^T W \bar{b} &= \sum_{i=1}^{10} \begin{bmatrix} A^T \\ 6 \times 3 \end{bmatrix} \begin{bmatrix} W \\ 3 \times 3 \end{bmatrix} \begin{bmatrix} \bar{b} \\ 1 \end{bmatrix} = \begin{bmatrix} 56,947.184 \\ -202,969.757 \\ -38,036.831 \\ 4320.123 \\ -15,666.380 \\ -1318.018 \end{bmatrix}
 \end{aligned}$$

where we've used the weights (noise) from Table 9-3. Be sure to convert the units appropriately.

$$W = \begin{bmatrix} \frac{1}{\sigma_\beta^2} & 0 & 0 \\ 0 & \frac{1}{\sigma_\rho^2} & 0 \\ 0 & 0 & \frac{1}{\sigma_{el}^2} \end{bmatrix} = \begin{bmatrix} \frac{1}{80^2} & 0 & 0 \\ 0 & \frac{1}{0.0081^2} & 0 \\ 0 & 0 & \frac{1}{0.0045^2} \end{bmatrix}$$

The update to the state is

$$\hat{\delta x} = (A^T W A)^{-1} A^T W \bar{b} = P A^T W \bar{b}$$

$$\hat{\delta x}^T = [2.418 \quad -1.686 \quad 2.038 \quad 7.994 \quad -8.841 \quad 1.642] \times 10^{-5}$$

and the nominal state becomes

$$\hat{x}_{nom} = \hat{x}_{nom} + \hat{\delta x} \quad \hat{x}_{nom}^T = [0.9020 \quad 0.4191 \quad 0.5394 \quad 0.5470 \quad -0.2434 \quad -0.7246] \text{ (ER, ER/TU)}$$

Subsequent iterations produce

$$\hat{\delta x} = \begin{bmatrix} 0.008403 \\ 0.522820 \\ -0.024703 \\ -1.379670 \\ 1.331090 \\ -0.649389 \end{bmatrix} \times 10^{-6} = \begin{bmatrix} -0.055683 \\ -0.092550 \\ 0.328402 \\ 1.352650 \\ 0.053096 \\ -0.192100 \end{bmatrix} \times 10^{-8}$$

The final answer and covariance matrix are

$$\hat{r}_1 = 0.902\ 015 \hat{i} + 0.419\ 145 \hat{j} + 0.539\ 390 \hat{k} \text{ ER}$$

$$= 5753.173 \hat{i} + 2673.361 \hat{j} + 3440.304 \hat{k} \text{ km}$$

$$\hat{v}_1 = 0.546\ 996 \hat{i} - 0.243\ 417 \hat{j} - 0.724\ 599 \hat{k} \text{ ER/TU}$$

$$= 4.324\ 207 \hat{i} - 1.924\ 299 \hat{j} - 5.728\ 216 \text{ km/s}$$

$$\hat{P} = \begin{bmatrix} 1.797 \times 10^{-9} & 1.091 \times 10^{-10} & -2.069 \times 10^{-10} & -1.710 \times 10^{-8} & -7.126 \times 10^{-10} & 8.480 \times 10^{-10} \\ 1.091 \times 10^{-10} & 9.936 \times 10^{-10} & -1.530 \times 10^{-9} & -2.518 \times 10^{-11} & -8.805 \times 10^{-9} & 1.322 \times 10^{-8} \\ -2.069 \times 10^{-10} & -1.530 \times 10^{-9} & 2.638 \times 10^{-9} & 1.067 \times 10^{-9} & 1.369 \times 10^{-8} & -2.421 \times 10^{-8} \\ -1.710 \times 10^{-8} & -2.518 \times 10^{-11} & 1.067 \times 10^{-9} & 2.232 \times 10^{-7} & -6.680 \times 10^{-9} & -1.568 \times 10^{-8} \\ -7.126 \times 10^{-10} & -8.804 \times 10^{-9} & 1.369 \times 10^{-8} & -6.680 \times 10^{-9} & 9.475 \times 10^{-8} & -1.511 \times 10^{-7} \\ 8.480 \times 10^{-10} & 1.322 \times 10^{-8} & -2.421 \times 10^{-8} & -1.567 \times 10^{-8} & -1.511 \times 10^{-7} & 3.186 \times 10^{-7} \end{bmatrix}$$

If we look at the standard-deviation values of r_1 , for example, we see that our answer has a standard deviation of about 270 m [$(1.797 \times 10^{-9})^{1/2} 6,378,136.3$]. Although this answer seems inaccurate, we used only ten observations.

At this point, we haven't considered the coordinate system. Therefore, we've "assumed" a generic inertial frame. If we use these vectors to point a sensor, we must be sure what the system requires—true-of-date, mean-of-date, and so on (see Sec. 1.7). We'll probably have to transform the coordinates.

Now, let's compare the results to answers verified through actual experiments. Incidentally, this test was the first successful dark-pass illumination by the HI-CLASS program and it increased accuracy by nearly a full order of magnitude over existing operational methods (Phillips Laboratory, 1995).

$$\hat{r}_1 = 5753.0076 \hat{i} + 2674.0502 \hat{j} + 3440.0874 \hat{k} \text{ km}$$

$$\hat{v}_1 = 4.326\ 418 \hat{i} - 1.925\ 781 \hat{j} - 5.726\ 373 \hat{k} \text{ km/s}$$

Notice the similarity between the answers from our simple example and the "complex" answer above. The standard-deviation RMS from the actual experiment was about 4 m, which coincided with the station bias solved for during the differential correction. If we rerun the problem using a range bias of 8 m, our answers change slightly, but they better represent the true answer. The standard deviation is the same, but our answer is about 45 m closer. The main differences arise from using true-of-date coordinates and precise propagation formulas in the differential correction.



9.6 Sequential-Batch Least Squares

Suppose we have 1,000 observations and we use Eq. (9-12) in an iterative loop to find the best answer to the state space of the system. But just as we finish converging on the answer, we receive thirty new observations. We could redo the process, using 1,030 observations, but that is a waste of time because we've already extracted the information from the first 1,000. The following solution involves looking at methods of sequentially processing successive batches of data.

The *batch least-squares*, *sequential-batch estimator*, the *Kalman filter*, and the *Extended Kalman filter* will all process the new data; but first, some definitions. We must address two main issues: how to incorporate the new data and how to propagate the state estimate and covariance to the current time, if desired. Each technique mentioned above will incorporate new data; however, it's possible to form a sequential-batch technique that doesn't move the state and covariance forward in time. I'll define ***sequential-batch*** routines as those which incorporate only new data. A *variation* of this technique can move the state and covariance through time, but this isn't the usual approach. The Kalman-filter approaches always combine the new data and move the state and covariance through time.

Using the previous example, we have iteratively determined the final state update at time t_o using 1,000 observations through time, t_k , so let's identify each variable as such:

$\hat{X}(0|k) \Rightarrow$ the *best state estimate* (or answer) at t_o , based on observations through t_k

$\delta\hat{x}(0|k) \Rightarrow$ the converged state-update estimate at t_o , based on observations through t_k .

This is simply the last correction vector from the differential correction (batch least squares). It normally approaches zero after iteration.

$\tilde{b}_k \Rightarrow$ the *residual matrix* containing residuals through t_k

$A_k \Rightarrow$ the *partial-derivative matrix* relating all observations through t_k to the state at t_o

$\hat{P} \equiv (A^T W A)^{-1} \Rightarrow$ the *covariance matrix*. It's expressed as $\hat{P}(0|k)$, implying it relates observations through t_k to an epoch state at t_o .

We now have thirty observations at a new time, t_n , with the sensor weights associated with each piece of data, W_{new} . Summarizing, we have a state-update estimate, $[\delta\hat{x}(0|k)]$, based on observations through time t_k ; some statistical values indicating the accuracy of the estimate, $P(0|k)$; some observations, y_n , at time, t_n ; and some statistical values estimating the accuracy of the new observations, W_{new} . The logical approach is to combine everything in an expanded least-squares process, treating the old state-error estimate as yet another set of observations. Looking at the residual equation applied to this approach, we see the residual matrix and the partial-derivative matrix are now

$$\tilde{b} = \begin{bmatrix} [\tilde{b}]_k \\ \vdots \\ [\tilde{b}]_{new} \end{bmatrix} \quad A = \begin{bmatrix} [A]_k \\ \vdots \\ [A]_{new} \end{bmatrix}$$

Notice that the $[A]_{new}$ matrix accounts for the new observations, and we find it just as we do the original A matrix. The partial derivatives are for the new observation with regard to the state at the original epoch, t_o , to account for the new pieces of data. The weighting matrix is as before:

$$W = \begin{bmatrix} W_k & 0 \\ 0 & W_{new} \end{bmatrix}$$

with the components of the W_{new} matrix being the reciprocal of the variances of the observations. The nonlinear, update equation [Eq. (9-12)] now becomes

$$\delta\hat{x}(0|k+n) = \left([A_k^T A_{new}^T] \begin{bmatrix} W_k & 0 \\ 0 & W_{new} \end{bmatrix} \begin{bmatrix} A_k \\ A_{new} \end{bmatrix} \right)^{-1} [A_k^T A_{new}^T] \begin{bmatrix} W_k & 0 \\ 0 & W_{new} \end{bmatrix} \begin{bmatrix} \tilde{b}_k \\ \tilde{b}_{new} \end{bmatrix}$$

Expanding terms gives us

$$\delta\hat{x}(0|k+n) = (A_{new}^T W_{new} A_{new} + A_k^T W_k A_k)^{-1} (A_{new}^T W_{new} \tilde{b}_{new} + A_k^T W_k \tilde{b}_k)$$

Recognize the presence of the covariance matrix, $\hat{P}_k^{-1} = A_k^T W_k A_k$, in Eq. (9-12). Also,

$$\hat{P}_k^{-1} \delta\hat{x}(0|k) = A_k^T W_k \tilde{b}_k$$

Substituting these relations gives us a new state correction:

$$\delta\hat{x}(0|k+n) = (A_{new}^T W_{new} A_{new} + \hat{P}_k^{-1})^{-1} (A_{new}^T W_{new} \tilde{b}_{new} + \hat{P}_k^{-1} \delta\hat{x}(0|k)) \quad (9-20)$$

and we define the updated covariance matrix as

$$\hat{P}_{k+n} = \hat{P}(0|k+n) = (A_{new}^T W_{new} A_{new} + \hat{P}_k^{-1})^{-1} \quad (9-21)$$

so $\delta\hat{x}(0|k+n)$ is the improved correction to $\hat{X}(0|k)$ due to additional data through time t_n . But we must not forget to iterate! We still have a nonlinear problem and a linearized solution, however, so we have to be careful when iterating. In the previous example, the nominal estimate resulting from processing 1,000 observations is probably very close to the actual answer. Adding thirty points may improve the accuracy, but the iterations can actually diverge after two or three passes. That's because the noise in the new observations is potentially larger than the minimized error resulting from the previous least-squares estimation. If this occurs, the sequential-batch estimation may ignore the new observations.

Observe the size of the matrices. From the augmented matrices used in developing this section, we might think a computer implementation would require larger A and b matrices. But careful examination of $\delta\hat{x}$ in Eq. (9-20) and Eq. (9-21) shows that although A and b are larger, the matrices we use are the same size! The k subscripted values are the results of initial differential-correction processing. The new subscripted values depend on the number of added observations, but when multiplied, the resulting product is the same size

as the k subscripted values. The discussion for directly accumulating the $\mathbf{A}^T \mathbf{W} \mathbf{A}$ and $\mathbf{A}^T \mathbf{W} \mathbf{b}$ matrices from Algorithm 59 also applies to sequential batch formulations.

We can rearrange the inverse in Eq. (9-21) to reduce the complexity of the matrix inversion so it has a dimension of the size of the next subset of measurements. This form is similar to Eq. (9-32), which we'll discuss in the Kalman-filter section. Remember that \mathbf{W} usually doesn't require calculation because it's the matrix for known measurement errors. The solution [Eq. (9-22)] is known as *Schur's identity* (Baker, 1967, 495). It reduces the complexity of the matrix inversion when there are fewer observations than the state-space dimension.

$$\hat{\mathbf{P}}_{k+n} = \hat{\mathbf{P}}_k - \hat{\mathbf{P}}_k \mathbf{A}_{new}^T (\mathbf{W}_{new}^{-1} + \mathbf{A}_{new} \hat{\mathbf{P}}_k \mathbf{A}_{new}^T)^{-1} \mathbf{A}_{new} \hat{\mathbf{P}}_k \quad (9-22)$$

Now suppose we get another batch of observations at a new time, t_{n+1} . We can use the sequential batch [Eq. (9-20) and Eq. (9-21)] just by saving the matrix on the left-hand brackets (before inverting) as $\hat{\mathbf{P}}^{-1}(0|k)$ and replacing n with $n + 1$ in all the subscripts. We would then have $\delta\hat{\mathbf{x}}(0|k+n+1)$, the best estimate for the state update after using observations through t_{n+1} (after iteration, of course).

The problem still remains of how to propagate $\delta\hat{\mathbf{x}}$ and $\hat{\mathbf{P}}(0|K)$ to the current time because the state-update estimate is still calculated at the epoch time. The *state-transition matrix* (from the next section) is sometimes used to propagate our state update and covariance.

$$\hat{\mathbf{P}} \approx \Phi \hat{\mathbf{P}}_o \Phi^T$$

The sequential estimation technique [Eq. (9-32)] is similar, but it also invokes an additional component (process noise) in the formulation. We'll discuss the Kalman filter shortly. The result is a predicted estimate using the new data. Remember that Φ is a linearized approximation for use in propagating the covariance matrix, and we may introduce unnecessary errors.

Continuing with Example 9-4, let's look at Example 9-5, which presents the sequential-batch solution:

▼ Example 9-5. Using Sequential-Batch Estimation for Differential Correction.

GIVEN: Observations from Kaena Point on January 29, 1995 for GEOS-III

FIND: $\hat{\mathbf{x}}, \hat{\mathbf{P}}$

Now, suppose we examine the last eight observations in Table 9-5 on page 700. We know the predicted value of the state from Example 9-4 but must find the accumulated matrices. For the first iteration, we could choose any epoch, but let's keep the original epoch from Example 9-4. The weighting matrix remains the same as in Example 9-4 because the observations are from the same site. As an aside, if we had observations from a different station, we could easily insert the noise parameters during the accumulation process.

$$\begin{aligned} A^T W A &= \sum_{i=1}^8 \begin{bmatrix} A^T \\ 6 \times 3 \end{bmatrix} \begin{bmatrix} W \\ 3 \times 3 \\ 3 \times 6 \end{bmatrix} \begin{bmatrix} A \\ 6 \times 3 \end{bmatrix} \\ &= \begin{bmatrix} 8.855 \times 10^9 & 1.335 \times 10^{10} & 1.745 \times 10^9 & 1.796 \times 10^9 & 2.784 \times 10^9 & 2.299 \times 10^8 \\ 1.335 \times 10^{10} & 3.193 \times 10^{10} & 3.560 \times 10^9 & 2.597 \times 10^9 & 6.447 \times 10^9 & 4.246 \times 10^8 \\ 1.745 \times 10^9 & 3.560 \times 10^9 & 2.536 \times 10^9 & 2.937 \times 10^8 & 5.901 \times 10^8 & 4.833 \times 10^8 \\ 1.796 \times 10^9 & 2.597 \times 10^9 & 2.937 \times 10^8 & 3.733 \times 10^8 & 5.558 \times 10^8 & 3.621 \times 10^7 \\ 2.784 \times 10^9 & 6.447 \times 10^9 & 5.901 \times 10^8 & 5.558 \times 10^8 & 1.338 \times 10^9 & 6.120 \times 10^7 \\ 2.299 \times 10^8 & 4.246 \times 10^8 & 4.833 \times 10^8 & 3.621 \times 10^7 & 6.120 \times 10^7 & 9.696 \times 10^7 \end{bmatrix} \end{aligned}$$

and

$$A^T W \tilde{b} = \sum_{i=1}^8 \begin{bmatrix} A^T \\ 6 \times 3 \end{bmatrix} \begin{bmatrix} W \\ 3 \times 3 \\ 3 \times 1 \end{bmatrix} \begin{bmatrix} \tilde{b} \\ 3 \end{bmatrix} = \begin{bmatrix} 181,338.222 \\ -9756.766 \\ 9478.221 \\ 41,653.052 \\ -3484.222 \\ 2091.122 \end{bmatrix}$$

Using Eq. (9-20) and Eq. (9-22), we find the updated state values. Remember that, although these parameters contain additional observations, they are still referenced to the original epoch.

$$\delta \hat{x}_{k+n} = (A_{new}^T W_{new} A_{new} + A_k^T W_k A_k)^{-1} (A_{new}^T W_{new} \tilde{b}_{new} + A_k^T W_k \tilde{b}_k)$$

$$\hat{P}_{k+n} = \hat{P}_k - \hat{P}_k A_{new}^T (W_{new}^{-1} + A_{new} \hat{P}_k A_{new}^T)^{-1} A_{new} \hat{P}_k$$

The first iteration produces

$$\delta \hat{x} = \begin{bmatrix} 3.1195 \times 10^{-25} \\ -1.8672 \times 10^{-25} \\ 3.4015 \times 10^{-6} \\ -5.4548 \times 10^{-6} \\ 6.4144 \times 10^{-5} \\ -6.2091 \times 10^{-5} \end{bmatrix}$$

As expected, the additional data adds little to the answer from Example 9-4, so we don't need more iterations. The final answer and covariance matrix are

$$r_1 = 0.902\,015 \hat{i} + 0.419\,145 \hat{j} + 0.539\,394 \hat{k} \text{ ER}$$

$$= 5753.173 \hat{i} + 2673.361 \hat{j} + 3440.328 \hat{k} \text{ km}$$

$$\dot{v}_1 = 0.546\,991 \hat{i} - 0.243\,353 \hat{j} - 0.724\,661 \hat{k} \text{ ER/TU}$$

$$= 4.324\,164 \hat{i} - 1.923\,794 \hat{j} - 5.728\,711 \text{ km/s}$$

$$\hat{P} = \begin{bmatrix} 0 & 0 & 0 & -1.686 \times 10^{-29} & 3.583 \times 10^{-10} & -3.583 \times 10^{-10} \\ 0 & 0 & 0 & 3.583 \times 10^{-10} & -2.072 \times 10^{-10} & -1.510 \times 10^{-10} \\ 0 & 0 & 3.583 \times 10^{-10} & -2.072 \times 10^{-10} & -2.989 \times 10^{-10} & 1.479 \times 10^{-10} \\ 0 & 3.583 \times 10^{-10} & -2.072 \times 10^{-10} & -2.989 \times 10^{-10} & -1.747 \times 10^{-9} & 1.894 \times 10^{-9} \\ 3.583 \times 10^{-10} & -2.072 \times 10^{-10} & -2.989 \times 10^{-10} & -1.746 \times 10^{-9} & 1.153 \times 10^{-9} & 7.413 \times 10^{-10} \\ -3.583 \times 10^{-10} & -1.510 \times 10^{-10} & 1.478 \times 10^{-10} & -1.894 \times 10^{-9} & 7.413 \times 10^{-10} & 9.038 \times 10^{-11} \end{bmatrix}$$

▲ Notice the similarity to the answer from Example 9-4.

9.7 Kalman Filtering

A major characteristic of differential correction and sequential-batch differential correction is that the converged state estimate and covariance matrix are based on processing a batch of data, spread over some time interval which may be minutes, hours, days, or even weeks. These techniques don't lend themselves to problems in which the forces are incompletely modeled. For example, variable atmospheric drag due to changing solar flux and geomagnetic activity causes slow and fast variations in the orbit, respectively. A second characteristic is that the estimate is always associated with a particular epoch. Thus, you may have to predict from the epoch state to a new time. A more significant issue is how to propagate the state and covariance matrix over the time interval to provide the statistical information at the new epoch. Covariance propagation is one application in which accurate information about the process noise is useful, but this must be weighed against the quality of that process-noise model. Kalman and Extended Kalman filters solve some of these problems but also introduce some new ones.

Kalman's contribution combined statistical mathematics with linear system theory to produce a recursive algorithm for computing "maximum-likelihood" estimates of a system state. Simply speaking, the Kalman filter is a technique for computing the best estimate of the state of a time-varying process. It uses a predictor-corrector technique ideally suited for computer applications, given imperfect observations and uncertain dynamics. It differs from the least-squares technique in two very important and useful ways. First, it continuously updates the epoch time, thus estimating the state at each successive observation time. Of course this assumes we have data available over a period of time. Second, it carries all the information concerning past measurements in its current state and covariance estimates and therefore doesn't need to reprocess all of the past measurement information at each time step. This is the sequential nature we saw in the sequential-batch method.

The Kalman filter and Extended Kalman filter have proven to be extremely useful for problems in which data streams continuously, such as in an attitude-estimation system or with continuous observations from a space-based sensor, but these techniques have been difficult to use for orbit determination. Many investigators conclude that filters are difficult to "tune" (selecting an appropriate model for process noise); thus, they often ignore

new data and diverge—a process known as *smugness*. Recent investigators, following an approach by Jim Wright, define models for process noise which derive from the statistical properties of the errors in the physical force models. I introduce this work separately in Sec. 9.7.4.

From Fig. 9-8, suppose we want to use new observations at time t_{k+1} and we have already calculated the best estimate of the converged state and the covariance (\hat{X}_k and \hat{P}_k) from data at the original epoch, t_k .

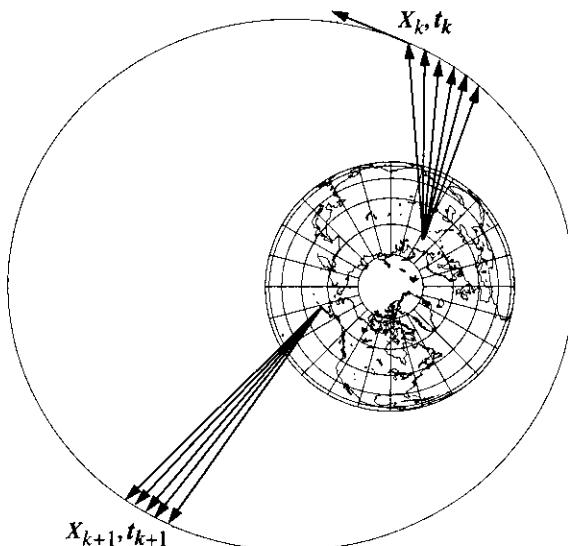


Figure 9-8. Observations from Multiple Passes. Differential correction minimizes the sum of the squares of the residuals by referring all corrections to a single state at a *fixed epoch*, t_k . The Kalman filter finds a state update at each observation time, t_{k+1} .

The propagation of the converged state and covariance to the current time assumes several forms. We'll discuss three—the Kalman filter, linearized Kalman filter, and the extended Kalman filter. The Kalman filter is used with linear systems. Linear dynamics simplify the propagation considerably. Unfortunately, satellite orbit determination is a nonlinear problem. Thus, we can choose a Linearized or an extended Kalman filter. These methods approach the propagation differently. The linearized Kalman filter uses a reference trajectory and calculates the differences to that trajectory. Thus the term *linearized* refers to the fact that we propagate the differences to the state, δx . This concept is very similar to Encke's method. The extended Kalman filter uses the current state estimate for each propagation instead of a single reference trajectory. In both cases, we use a two-step process because we have a nonlinear system. We *predict* the new state based on previous data, and then *update* that result with the new observations. I'll develop the equations for linear and nonlinear propagation and explore each of these concepts next.

Linearizing the Dynamics—The State-Transition Matrix, Φ

To linearize the equations of motion, let's begin by assuming we have two neighboring trajectories. Suppose also that we know one and are interested in finding the other. This reduces to finding the difference between them. By linearizing the equations, we can develop a simplified approximation to compute the differences over a limited time interval, much as we did in Encke's method in Chap. 7. The only constraint is that the two trajectories remain close enough that neglected higher order terms (Eq. (9-25)) are negligible. If the two trajectories are X and Y , we can express the initial conditions and derivative of a six-parameter state as

$$\begin{aligned} X(t_o) &= X_o & \dot{X} &= \hat{f}(X) \\ Y(t_o) &= Y_o & \dot{Y} &= \hat{f}(Y) \end{aligned} \quad (9-23)$$

where f simply denotes a function of the state. Letting the difference in the two trajectories be δx gives us

$$Y = X + \delta x \quad (9-24)$$

Taking the derivative and using Eq. (9-23) result in $\dot{Y} = \hat{f}(X + \delta x)$. Now, because $\hat{f}(X + \delta x)$ is nonlinear, let's expand the function in a Taylor series about X , ($\delta x = Y - X$):

$$\dot{Y} = \hat{f}(X) + \frac{\partial \hat{f}(X)}{\partial X} \delta x + \frac{\partial \hat{f}^2(X)}{2! \partial X^2} \delta x^2 + \dots \quad (9-25)$$

Here, the partial derivative of $\hat{f}(X)$ is a 6×6 matrix that depends on time. Next, take the time derivative of Eq. (9-24) and substitute into the left-hand side of Eq. (9-25):

$$\dot{X} + \delta \dot{x} = \hat{f}(X) + \frac{\partial \hat{f}(X)}{\partial X} \delta x + \frac{\partial \hat{f}^2(X)}{2! \partial X^2} \delta x^2 + \dots$$

Using Eq. (9-23), we can simplify this expression as

$$\delta \dot{x} = \frac{\partial \hat{f}(X)}{\partial X} \delta x + u = F(t) \delta x + u \quad (9-26)$$

I've introduced two new quantities here. The neglected terms of the second order and higher are represented by a vector, u , and $F(t)$ is the matrix of partial derivatives of the state rates as indicated. Eq. (9-26) represents the linearized dynamics and its solution is the time-varying difference between the two neighboring trajectories. Let's examine the formation of the F matrix and the u matrix.

We can find a relationship for F by assuming a solution to Eq. (9-26) having the form

$$\delta x = \Phi(t, t_o) \delta x_o \quad (9-27)$$

Notice we've introduced the ***state-transition matrix***, Φ , which propagates the state differences through time and represents the partial derivatives of the state at time t with respect

to the state at t_o . We've also let $\mathbf{u} = 0$ for now. The form looks similar to the variational equations. We can differentiate Eq. (9-27) and substitute into Eq. (9-26), allowing us to find a first-order differential equation and an initial condition:

$$\begin{aligned}\dot{\Phi}(t, t_o) &= \mathbf{F}(t)\Phi(t, t_o) \\ \Phi(t_o, t_o) &= \mathbf{I}\end{aligned}\tag{9-28}$$

The initial condition results from Eq. (9-27) if the time interval is zero. The following matrix operations apply:

$$\begin{aligned}\Phi(t, t_o) &= \Phi^{-1}(t_o, t) \\ \Phi(t_1, t_o) &= \Phi(t_1, t)\Phi(t, t_o)\end{aligned}\tag{9-29}$$

An important restriction is that Φ be invertible; thus, the determinant of Φ can't equal zero.

For the orbit problem, the state-transition matrix is often obtained by numerically integrating Eq. (9-28) along with the equations of motion. The fidelity of the integration for Eq. (9-28) is usually not as important as for the equations of motion because we're only trying to evaluate the small state displacement, $\delta\mathbf{x}$, from our nominal trajectory using the state-transition matrix. Errors in $\delta\mathbf{x}$ tend to be small because $\delta\mathbf{x}$ is small. Thus, the propagation of errors in $\delta\mathbf{x}$ are also small. However, the nominal state components are not small so we want to propagate them accurately so our best estimate of the state (the sum of the nominal state with $\delta\mathbf{x}$) is accurate.

Under certain circumstances, we can obtain an analytical formulation for $\Phi(t_o, t)$. Junkins (1975, 45–48) shows the general case for analytical development, as well as how to derive the Peano-Baker method for determining the state-transition matrix. One special case occurs when the \mathbf{F} matrix is constant—how constant remains open for discussion because it depends on the specific application. Such systems are called *stationary systems* because \mathbf{F} is time independent. Another useful tool is the *Cayley-Hamilton Theorem* which states that all square matrices satisfy their own characteristic equations (Gelb, 1989, 19–20). This theorem lets us represent an exponential of a matrix as a power series whose argument is a matrix. To demonstrate this, let's expand $\delta\mathbf{x}$ in a Taylor series:

$$\delta\mathbf{x} = \delta\mathbf{x}_o + \dot{\delta\mathbf{x}}_o(t - t_o) + \frac{1}{2!}\ddot{\delta\mathbf{x}}_o(t - t_o)^2 + \frac{1}{3!}\dddot{\delta\mathbf{x}}_o(t - t_o)^3 + \dots$$

But from Eq. (9-26) and its derivatives, we know that

$$\begin{aligned}\dot{\delta\mathbf{x}}_o &= \mathbf{F}\delta\mathbf{x}_o \\ \ddot{\delta\mathbf{x}}_o &= \mathbf{F}\dot{\delta\mathbf{x}}_o = \mathbf{F}^2\delta\mathbf{x}_o \\ \dddot{\delta\mathbf{x}}_o &= \mathbf{F}\ddot{\delta\mathbf{x}}_o = \mathbf{F}^3\delta\mathbf{x}_o\end{aligned}$$

Substituting into the Taylor series and simplifying give us

$$\delta \mathbf{x} = \delta \mathbf{x}_o \left\{ \mathbf{I} + \mathbf{F}(t - t_o) + \frac{1}{2!} \mathbf{F}^2(t - t_o)^2 + \frac{1}{3!} \mathbf{F}^3(t - t_o)^3 + \dots \right\}$$

The expression in the brackets is the definition of a matrix exponential; thus, we can rewrite the Taylor series expansion and the state-transition matrix:

$$\left. \begin{aligned} \delta \mathbf{x} &= \text{EXP}^{F(t, t_o)} \delta \mathbf{x}_o \\ \Phi(t, t_o) &= \text{EXP}^{F(t, t_o)} \end{aligned} \right\} F \text{ constant only} \quad (9-30)$$

This follows by comparison with Eq. (9-27).

Eq. (9-30) is now a linearized equation for the dynamical system (we didn't include the truncation matrix, \mathbf{u}), and it represents the time-varying difference (to first order) between the two trajectories. Now recognize the definition of \mathbf{F} in Eq. (9-26) as the time rate of change of the variational equations we developed in Eq. (9-17). Thus,

$$\mathbf{F}(t) = \frac{\partial \dot{\bar{X}}(X)}{\partial \bar{X}_o} = \frac{d}{dt} \left(\frac{\partial X}{\partial \bar{X}_o} \right)$$

We can calculate the state-transition matrix from the exponential relation in Eq. (9-30), or by numerically integrating Eq. (9-28). The discussion from Sec. 9.5.3 also applies. \mathbf{F} is very complex for detailed force models, so we obtain it from the partial derivative of the acceleration vector. Also, we can find the \mathbf{F} matrix using finite differences. I'm presenting this method last because it's usually not regarded as robust enough for highly accurate simulations. Its primary use is academic—for illustration of the basic concepts. For small time differences, we approximate \mathbf{F} by applying the secant method (a first-order slope):

$$\mathbf{F} = \frac{\frac{\dot{\bar{X}}}{\bar{X}}}{\frac{\Delta \bar{X}}{\Delta \bar{X}_o}} \approx \frac{\dot{\bar{X}}_o}{\Delta \bar{X}_o}$$

meaning that a small change in \bar{X}_o will result in a small change for $\dot{\bar{X}}_o$. Notice this approximation for F implies that a larger change in \bar{X}_o will result in a change in $\dot{\bar{X}}_o$ that is *directly proportional* to the effect produced by the smaller change. For nonlinear systems, this change *won't* be proportional, and the finite difference will yield erroneous results.

For \mathbf{F} , we change each component of the state at time t_k by some small amount [1% usually works well, Eq. (9-15)] and calculate the resulting derivative of the state. Recall this is simply velocity and acceleration, so you can find the values using analytical or numerical techniques.

$$\mathbf{F} \equiv \begin{bmatrix} \frac{\dot{\bar{X}}_{1_c}(t) - \dot{\bar{X}}_{1_n}(t)}{\bar{X}_{1_c}(t) - \bar{X}_{1_n}(t)} & \cdots & \frac{\dot{\bar{X}}_{1_c}(t) - \dot{\bar{X}}_{1_n}(t)}{\Delta \bar{X}_6(t)} \\ \frac{\dot{\bar{X}}_{2_c}(t) - \dot{\bar{X}}_{2_n}(t)}{\Delta \bar{X}_1(t)} & \cdots & \frac{\dot{\bar{X}}_{2_c}(t) - \dot{\bar{X}}_{2_n}(t)}{\Delta \bar{X}_6(t)} \\ \vdots & & \vdots \\ \frac{\dot{\bar{X}}_{6_c}(t) - \dot{\bar{X}}_{6_n}(t)}{\Delta \bar{X}_1(t)} & \cdots & \frac{\dot{\bar{X}}_{6_c}(t) - \dot{\bar{X}}_{6_n}(t)}{\Delta \bar{X}_6(t)} \end{bmatrix}_{t=t_k}$$

$\bar{X}_{1_c}(t)$ is the first changed state at t_k , and $\bar{X}_{1_n}(t)$ is the nominal state at the same time.

We must also evaluate the truncation term from Eq. (9-26). Over small time intervals, we will assume u is essentially constant. Thus, using Euler's method (first-order approximation) results in

$$\delta x = u \delta t$$

The state-transition matrix allows us to map this error to any subsequent time, and the total effect on the state is found by summing over all time intervals. Notice that we have introduced a subtle change in the definition of u . It's no longer the neglected higher order dynamics, but rather a statistical measure of the uncertainty in our dynamics due to force model truncation and other errors. We can find the most general expression by taking the limit as the time intervals approach zero, so

$$v = \int_{t_o}^t \Phi(t, t_o) u dt$$

This equation defines process noise. The **process noise**, v , is the uncertainty in the dynamics used in the propagation process.

Estimating the State, Predicted Covariance, and State Update

Using the linearized dynamics, the resulting updates can be much faster, but a subtlety exists. Don't confuse the nonlinear propagation of the state with the linearized approximation for the update to the state. When you propagate the state by numerically integration, this is a nonlinear process. When you use Φ to propagate the state update, you've assumed linearized dynamics for the solution. Remember that v is the noise due to the linearization and it has a zero mean. The resulting linear approximation for the propagated state update becomes

$$\delta\tilde{x}_{k+1} = \Phi\delta\hat{x}_k + v \quad (9-31)$$

We use the state-transition matrix, Φ , to propagate the covariance matrix so we can see how good our state is at the new time. Referencing Gelb (1989, 75–76), the covariance matrix is really the expected value of the square of the errors in the state, so we can estimate it by

$$\hat{P}_k = E(\delta\hat{x}_k \delta\hat{x}_k^T)$$

and the predicted value of the covariance is based on the predicted state update at the new time, t_{k+1} . This includes data only through t_k . Thus,

$$\bar{P}_{k+1} = E(\delta\hat{x}_{k+1} \delta\hat{x}_{k+1}^T)$$

Because we've already found the predicted value for the state update at the new time, we can substitute for it:

$$\bar{P}_{k+1} = E\{(\Phi\delta\hat{x}_k + v)(\Phi\delta\hat{x}_k + v)^T\}$$

$$\bar{P}_{k+1} = E(\Phi\delta\hat{x}_k \delta\hat{x}_k^T \Phi^T + \Phi\delta\hat{x}_k v^T + v \delta\hat{x}_k^T \Phi^T + vv^T)$$

Let's assume the state update and the error in the dynamics are uncorrelated. This means the expected value of their product is zero, and

$$\bar{P}_{k+1} = \Phi E(\delta\hat{x}_k \delta\hat{x}_k^T) \Phi^T + E(vv^T)$$

We know the expected value of Φ is $E(\Phi) = \Phi$ (the modeled dynamics aren't random), and the factor of the first expectation operator is really the definition for the covariance at time t_k . The last expectation-operator term is the *second moment of the process noise*, Q , which accounts for the process-noise error in propagating the covariance matrix through time:

$$\bar{P}_{k+1} = \Phi \hat{P}_k \Phi^T + Q \quad (9-32)$$

We've reached the first goal: propagating the state, state-update estimate, and covariance matrix to the new time. Eq. (9-31) and Eq. (9-32) perform this propagation, which results in *predictions* of the state, state update, and covariance at the current time. Unlike differential correction and sequential-batch differential correction, which require iteration to process the data over time, these equations usually aren't iterated. They also don't yet contain any of the new information from time t_{k+1} ; therefore, we must refine or correct the values by refining the solution.

Generalizing the Update Equations

We can refine the predicted results and find generalized expressions by rewriting the result for sequential differential correction—[Eq. (9-20) and Eq. (9-21)]. Doing so gives us equations to correct our estimate for the data through t_{k+1} :

$$\begin{aligned}\hat{\delta\dot{x}}_{k+1} &= (\mathbf{H}^T \mathbf{W} \mathbf{H} + \bar{\mathbf{P}}_{k+1}^{-1})^{-1} (\mathbf{H}^T \mathbf{W} \tilde{\mathbf{b}} + \bar{\mathbf{P}}_{k+1}^{-1} \delta\dot{x}_{k+1}) \\ \hat{\mathbf{P}}_{k+1} &= (\mathbf{H}^T \mathbf{W} \mathbf{H} + \bar{\mathbf{P}}_{k+1}^{-1})^{-1}\end{aligned}\quad (9-33)$$

There are a few differences. Notice each quantity has a $k+1$ subscript because now we've updated them to the current time. The new data is added (appended to the residual matrix, $\tilde{\mathbf{b}}$), but we've used \mathbf{H} instead of \mathbf{A} for the *partial-derivative matrix*. They both represent partial derivatives of the observations with respect to the state, but \mathbf{H} uses the current state whereas \mathbf{A} uses the epoch state. We determine \mathbf{H} just as we do in the case for nonlinear least squares. We usually calculate this matrix analytically, based on the observation models. Thus,

$$\mathbf{H} = \frac{\partial \text{observations}_{k+1}}{\partial \bar{\mathbf{X}}_{k+1}}$$

If each observation consists of a triplet $\{\rho, \beta, el\}$, each matrix we accumulate will be (3×6) for a state vector of six parameters. Remember that we can approximate \mathbf{H} with finite differencing. The same caution from the least-squares case applies in that we must be consistent when applying the delta, perturbed minus nominal, or vice versa.

In a simple sense, we could use the correction equations [Eq. (9-33)] directly, but examining the covariance update shows the need to invert a 6×6 matrix (for a 6×1 state vector). It would be nice to reduce the order of the matrix inversion to improve computational efficiency. In addition, we may encounter observability constraints if we have fewer observations than the state space. The solution relies on the *Matrix Inversion Lemma*, originally derived by Bryson and Ho (1975, 349–359). The process begins with the updated covariance [Eq. (9-21)]:

$$\hat{\mathbf{P}} = (\mathbf{H}^T \mathbf{W} \mathbf{H} + \bar{\mathbf{P}}^{-1})^{-1}$$

Now, write the inverse of the original equation:

$$\hat{\mathbf{P}}^{-1} = \mathbf{H}^T \mathbf{W} \mathbf{H} + \bar{\mathbf{P}}^{-1}$$

Next, pre-multiply by $\hat{\mathbf{P}}$, and post-multiply by $\bar{\mathbf{P}}$ to get

$$\hat{\mathbf{P}} \hat{\mathbf{P}}^{-1} \bar{\mathbf{P}} = \hat{\mathbf{P}} \mathbf{H}^T \mathbf{W} \mathbf{H} \bar{\mathbf{P}} + \hat{\mathbf{P}} \bar{\mathbf{P}}^{-1} \bar{\mathbf{P}}$$

and simplify to produce

$$\bar{\mathbf{P}} = \hat{\mathbf{P}} \mathbf{H}^T \mathbf{W} \mathbf{H} \bar{\mathbf{P}} + \hat{\mathbf{P}} \quad (9-34)$$

Although we won't use this expression now, notice you could rearrange Eq. (9-34) as $\hat{P} = \bar{P} - \hat{P}H^TWH\bar{P}$. Post-multiplying Eq. (9-34) by H^TW to get

$$\bar{P}H^TW = \hat{P}H^TWH\bar{P}H^TW + \hat{P}H^TW$$

and factoring yields

$$\bar{P}H^TW = \hat{P}H^TW(H\bar{P}H^TW + I)$$

The key now is to recognize the need to factor out a W matrix. Therefore, replace the identity matrix with $W^{-1}W$:

$$\bar{P}H^TW = \hat{P}H^TW(H\bar{P}H^TW + W^{-1}W)$$

Factor a W out of $H\bar{P}H^TW + W^{-1}W$, remembering the *order* of multiplication is important:

$$\bar{P}H^TW = \hat{P}H^TW(H\bar{P}H^T + W^{-1})W$$

Post-multiply both sides by W^{-1} :

$$\bar{P}H^T = \hat{P}H^TW(H\bar{P}H^T + W^{-1})$$

Post-multiply both sides by $(H\bar{P}H^T + W^{-1})^{-1}$:

$$\bar{P}H^T(H\bar{P}H^T + W^{-1})^{-1} = \hat{P}H^TW(H\bar{P}H^T + W^{-1})(H\bar{P}H^T + W^{-1})^{-1} = \hat{P}H^TW$$

Solve for the right-hand side:

$$\hat{P}H^TW = \bar{P}H^T(H\bar{P}H^T + W^{-1})^{-1}$$

Now, define $K \equiv \hat{P}H^TW$ as the *Kalman gain matrix* (K), so that

$$K = \bar{P}H^T(H\bar{P}H^T + W^{-1})^{-1} \quad (9-35)$$

By taking the inverse weighting matrix $W^{-1} = R$ as the measurement-noise matrix, we find that R leads to maximum-likelihood estimation. There is a subtlety in that they are equal only if the observations are uncorrelated, which is usually true. W is a function of the statistics only, while R admits correlations. In this book, we will adopt the convention of including R in Kalman filtering algorithms. Replacing $\hat{P}H^TW$ in Eq. (9-34) identified several steps back ($\hat{P} = \bar{P} - \hat{P}H^TWH\bar{P}$) results in

$$\hat{P} = \bar{P} - KHP = [I - KH]\bar{P} \quad (9-36)$$

This last formula is the basic covariance relation we want. It can exhibit numerical problems (a loss of positive definiteness). Be aware that certain cases may require special

processing to avoid these problems. One way of guaranteeing the positive definiteness of the covariance is to use the *Joseph algorithm* (Gelb, 1989, 305):

$$\hat{\mathbf{P}} = (\mathbf{I} - \mathbf{K}\mathbf{H}) \bar{\mathbf{P}} (\mathbf{I} - \mathbf{K}\mathbf{H})^T + \mathbf{K}\mathbf{R}\mathbf{K}^T \quad (9-37)$$

Thus, to guarantee positive definiteness requires the cost of additional computations. Another elegant algorithm was developed by Potter in 1962 (Battin, 1987, 659) and extended by Bierman (1977, 68–112). Their work on formulating square-root filters (see Sec. 9.8.3) yields very stable algorithms, especially for onboard processing with limited computational power.

Although no inversions appear explicitly in the final answer [Eq. (9-36) or Eq. (9-37)], the definition of \mathbf{K} does contain an inverse. Inspection shows the success of our work above. Because \mathbf{H} is the partial-derivative matrix of the observations, a state vector of 6×1 will produce an \mathbf{H} of 1×6 for each observation. However, remember our observation matrix may consist of three values— ρ , β , el ; in this case, \mathbf{H} would be a 3×6 . Continuing with the matrices in the expression shows the matrix we'll invert is a 3×3 instead of a 6×6 . We can further increase this speed if our observations are uncorrelated. For instance, if errors in azimuth don't affect the accuracy of range, we can process each element of the observation vector individually, and the resulting matrix to invert is a 1×1 ! A division compared to even a 3×3 inversion is a tremendous savings and one of the real benefits of a Kalman filter. These observations mask an important issue—minimal stability or the performance of the filter when only a few observations are used. The Joseph algorithm is more numerically stable than the classical technique of Eq. (9-36).

Our final task is to update the state using the corrected value of $\hat{\mathbf{P}}$ and the predicted value of $\delta\hat{\mathbf{x}}$. If we substitute the expression for $\hat{\mathbf{P}}$ into the state update in Eq. (9-33), we may write the estimate of the state as

$$\delta\hat{\mathbf{x}} = \hat{\mathbf{P}}(\mathbf{H}^T \mathbf{W} \tilde{\mathbf{b}} + \bar{\mathbf{P}}^{-1} \delta\bar{\mathbf{x}})$$

Expanding gives us

$$\delta\hat{\mathbf{x}} = \hat{\mathbf{P}}\mathbf{H}^T \mathbf{W} \tilde{\mathbf{b}} + \hat{\mathbf{P}}\bar{\mathbf{P}}^{-1} \delta\bar{\mathbf{x}}$$

Now, let's use the definition of the Kalman gain and Eq. (9-36) to produce

$$\delta\hat{\mathbf{x}} = \mathbf{K}\tilde{\mathbf{b}} + (\mathbf{I} - \mathbf{K}\mathbf{H})\delta\bar{\mathbf{x}}$$

Rearranging terms results in

$$\delta\hat{\mathbf{x}} = \delta\bar{\mathbf{x}} + \mathbf{K}(\tilde{\mathbf{b}} - \mathbf{H}\delta\bar{\mathbf{x}})$$

Notice the term $\mathbf{H}\delta\bar{\mathbf{x}}$ represents the matrix of predicted values of the residuals, and $\tilde{\mathbf{b}}$ is the familiar residual matrix (observed minus nominal). Therefore, we're correcting the predicted estimate with the error in the predicted residuals. The Kalman gain serves as the

sensitivity of the estimate to the residuals. Now, let's summarize the various forms of the Kalman-filter equations.

Types of Filters

The filter process begins with an initial state and covariance matrix at a given epoch. The filter can't generate this initial data from the observations, so it relies on batch least-squares techniques to find them. Our goal is to update the state and covariance to the current time step-by-step, using each observation and either a linear or an “extended” approach. Filter algorithms provide analytical formulas for covariance propagation and avoid the need to propagate by numerical integration. I'll introduce three implementations of Kalman filters. The linear approach is designed only for linear systems, but it illustrates the process. The Linearized and Extended Kalman filters are the approaches for orbit determination.

9.7.1 Kalman Filter (Linear System)

We've derived the basic equations for the Kalman filter, following Gelb (1989, 107–119). When we apply a Kalman filter to a linear system, certain simplifications occur. Most notable is that we can use the state-transition matrix, Φ , to propagate the state exactly because the equations of motion are linear. Although we don't use white process noise for the predicted state update because it vanishes on average, we keep the second moment of the process noise, Q , to ensure the covariance remains open and doesn't become smug during processing. Remember that linear equations of motion rarely approximate real-world spacecraft systems. Algorithm 60 summarizes the Kalman filter for a linear system.

ALGORITHM 60: Kalman Filter—Linear System

$$(\hat{X}_k, \hat{P}_k, z_k, Q, R_k \Rightarrow \hat{X}_{k+1}, \hat{P}_{k+1})$$

$$\Phi(t_{k+1}, t_k) = \frac{\partial \hat{X}_{k+1}}{\partial \hat{X}_k}$$

$$H = \frac{\partial \text{observations}_{k+1}}{\partial \bar{X}_{k+1}}$$

Prediction

$$\bar{X}_{k+1} = \Phi \hat{X}_k$$

Predicted State

$$\bar{P}_{k+1} = \Phi \hat{P}_k \Phi^T + Q$$

Predicted Error Covariance

Update

$$\mathbf{K}_{k+1} = \bar{\mathbf{P}}_{k+1} \mathbf{H}_{k+1}^T \left[\mathbf{H}_{k+1} \bar{\mathbf{P}}_{k+1} \mathbf{H}_{k+1}^T + \mathbf{R}_k \right]^{-1} \quad \text{Kalman Gain}$$

$$\hat{\mathbf{X}}_{k+1} = \bar{\mathbf{X}}_{k+1} + \mathbf{K}_{k+1} \left[\mathbf{z}_{k+1} - \mathbf{H}_{k+1} \bar{\mathbf{X}}_{k+1} \right] \quad \text{State Estimate}$$

$$\hat{\mathbf{P}}_{k+1} = [\bar{\mathbf{P}}_{k+1} - \mathbf{K}_{k+1} \mathbf{H}_{k+1}] \bar{\mathbf{P}}_{k+1} \quad \text{Error Covariance Estimate}$$

Because the equations of motion are linear, we don't require a Taylor series expansion about a known trajectory. We use the state, \mathbf{X} , instead of the corrections to the state, $\delta\mathbf{x}$. The $\tilde{\mathbf{b}}$ matrix is replaced with the \mathbf{z} matrix of observations just as we did in the linear least squares example. In the terms of nonlinear *differential correction*, the residual matrix $\tilde{\mathbf{b}}$ is really $(\mathbf{z} - \mathbf{H}\mathbf{X})$. The state is updated each time we receive an observation, and the Φ and \mathbf{H} matrices are almost always calculated analytically for linear systems. Φ may or may not depend on time, based on the dynamics model we choose. We use no noise in the state update when we assume white Gaussian noise with zero mean ($E(\mathbf{v}) = 0$). Example 9-6 illustrates this process.

The presence of process noise and measurement noise gives us the chance to change our statistical hypothesis for a problem—a process known as *tuning* the filter. Tuning lets us satisfy certain tests, such as the *McReynolds filter-smoother* test. Consult Bierman (1977, 214, 225-231) for additional information on this test.

▼ Example 9-6. Applying a Linear Kalman Filter.

A high-speed rail system uses a radar to measure range to a platform where it is designed to stop. The train is decelerating at a constant unknown rate, \ddot{a} (see Fig. 9-9). The train engineer's job is to estimate position and velocity information from the radar's range measurements ($\sigma_{range} = 5$ m).

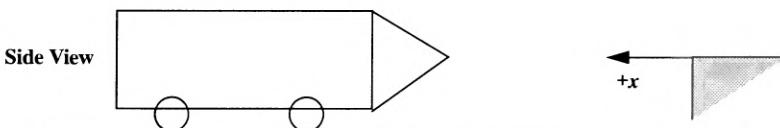


Figure 9-9. Schematic View of a High-speed Rail System for a Problem Applying a Linear Kalman Filter. This figure corresponds to Example 9-6, which looks at a high-speed rail system and tries to determine a train's position and velocity. We'll consider only one dimension.

Set up the equations of motion. The first step is to recognize this is a linear problem. Because we must accurately assess the acceleration to describe position and velocity accurately, we'll also estimate the train's deceleration. The state vector will be

$$\bar{x} = \begin{bmatrix} r \\ v \\ a \end{bmatrix} = \begin{bmatrix} r \\ \dot{r} \\ \ddot{r} \end{bmatrix}$$

We need initial conditions. Assume the initial conditions with 1σ uncertainties as given below:

$$r_o = 600 \text{ m} \pm 15 \text{ m}$$

$$v_o = -45 \text{ m/s} \pm 5 \text{ m/s}$$

$$a_o = 1.5 \text{ m/s}^2 \pm 1 \text{ m/s}^2$$

This expression defines the best estimate of the state:

$$\hat{X}_o = \begin{bmatrix} 600 \\ -45 \\ 1.5 \end{bmatrix}$$

and the variances (standard deviation squared) allow us to form the best estimate of the covariance matrix:

$$\hat{P}_o = \begin{bmatrix} 225 & 0 & 0 \\ 0 & 25 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Because the observations are range, the H matrix is simply the range component, which seems to imply the H matrix is easy to determine. It's easy only because the range measurement is the same as the range position in the state vector. If the observations had been angular measurements, the H matrix would have been more complex. Now,

$$H = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$$

The process-noise matrix, Q , is a noise vector with some statistics. We can use engineering judgment to guess its value. Because this is a well-defined problem with only minor unknowns, such as wind and bearing friction, our model will be fairly accurate, so we'll guess our noise statistics are small. Therefore, let's define

$$\text{range} = 1 \text{ m}$$

$$\text{velocity} = 0.1 \text{ m/s}$$

$$\text{acceleration} = 0.1 \text{ m/s}^2$$

$Q(t)$ now becomes a constant matrix whose diagonal consists of the variances of our guessed dynamics and power-density-matrix statistics:

$$Q(t) = \begin{bmatrix} \text{range}^2 & 0 & 0 \\ 0 & \text{velocity}^2 & 0 \\ 0 & 0 & \text{acceleration}^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.01 & 0 \\ 0 & 0 & 0.01 \end{bmatrix}$$

Because we're measuring only range, the measurement-noise matrix, $R(t)$, reduces to one term: $R(t) = \sigma^2 = 25 \text{ m}^2$.

Don't confuse this standard deviation with the *initial* standard deviations for the initial conditions. Finally, because we've chosen the dynamics model to be a simple acceleration, we can find the position, velocity, and acceleration by knowing acceleration and integrating. This may not always be true, but if it is:

$$a(t) = a_o$$

$$v(t) = a_o \Delta t + v_o$$

$$r(t) = \frac{1}{2} a_o \Delta t^2 + v_o \Delta t + r_o$$

or in matrix form,

$$\Phi = \begin{bmatrix} 1 & \Delta t & \frac{\Delta t^2}{2} \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix}$$

As the train approaches the platform, the engineer gets the following readings from his radar:

	$t_{\text{observation}}$ (s)	r_{observed} (m)
t_0	0	600
t_1	5	385
t_2	10	234
t_3	15	85
t_4	20	2

Now, we're ready to set up the Kalman-filter equations. We'll employ numerical subscripts to identify which time step value we're using.

Predicting and correcting the state at time $t = 5$ seconds ($i = 1, \Delta t = 5$ s)

$$\bar{X}_1 = \Phi \hat{X}_0 = \begin{bmatrix} 1 & \Delta t & \frac{\Delta t^2}{2} \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 600 \\ -45 \\ 1.5 \end{bmatrix} = \begin{bmatrix} 1 & 5 & 12.5 \\ 0 & 1 & 5 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 600 \\ -45 \\ 1.5 \end{bmatrix} = \begin{bmatrix} 393.75 \\ -37.50 \\ 1.50 \end{bmatrix}$$

$$\bar{P}_1 = \Phi \hat{P}_0 \Phi^T + Q(t) = \begin{bmatrix} 1 & 5 & 12.5 \\ 0 & 1 & 5 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 225 & 0 & 0 \\ 0 & 25 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 5 & 1 & 0 \\ 12.5 & 5 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.01 & 0 \\ 0 & 0 & 0.01 \end{bmatrix}$$

$$\bar{P}_1 = \begin{bmatrix} 1007.25 & 187.5 & 12.5 \\ 187.5 & 50.01 & 5 \\ 12.5 & 5 & 1.01 \end{bmatrix}$$

The simplicity of the H matrix allows us more simply to determine the Kalman gain. The quantity $\bar{P}H^T$ will always be the first column of the covariance matrix. In addition, $\bar{P}H\bar{P}^T$ will always be a 1×1 matrix, consisting of the first row and column value from the covariance matrix. The inverse is simply the reciprocal of this quantity. Keeping these shortcuts in mind, we find that correction of the state at time $t = 5$ seconds is

$$K_1 = \bar{P}_1 H^T [\bar{P}H\bar{P}^T + R]^{-1} = \begin{bmatrix} 1007.25 \\ 187.50 \\ 12.50 \end{bmatrix} \frac{1}{1032.25} = \begin{bmatrix} 0.975 & 78 \\ 0.181 & 64 \\ 0.012 & 11 \end{bmatrix}$$

Again, the simplicity of the H matrix permits some simplifications in determining the best estimates:

$$\hat{X}_1 = \bar{X}_1 + K_1 [\bar{z} - H\bar{X}_1] = \begin{bmatrix} 393.75 \\ -37.50 \\ 1.50 \end{bmatrix} + \begin{bmatrix} 0.975 & 78 \\ 0.181 & 64 \\ 0.012 & 11 \end{bmatrix} [385 - 393.75] = \begin{bmatrix} 385.211 & 92 \\ -39.089 & 37 \\ 1.394 & 04 \end{bmatrix}$$

$$\hat{P}_1 = [I - K_1 H] \bar{P}_1 = \begin{bmatrix} 0.024 & 22 & 0 & 0 \\ -0.181 & 64 & 1 & 0 \\ -0.012 & 11 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1007.25 & 187.5 & 12.5 \\ 187.5 & 50.01 & 5 \\ 12.5 & 5 & 1.01 \end{bmatrix} = \begin{bmatrix} 24.394 & 53 & 4.541 & 05 & 0.302 & 74 \\ 4.541 & 05 & 15.952 & 12 & 2.729 & 47 \\ 0.302 & 74 & 2.729 & 47 & 0.858 & 63 \end{bmatrix}$$

Notice the range uncertainty ($\sigma = 4.939$ 08 vs. 5.0) has decreased from the uncertainty on the initial conditions.

Predicting and correcting the state at time $t = 10$ seconds ($i = 2$, $\Delta t = 5$ s)

$$\bar{X}_2 = \Phi \hat{X}_1 = \begin{bmatrix} 1 & 5 & 12.5 \\ 0 & 1 & 5 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 385.211 & 92 \\ -39.089 & 37 \\ 1.394 & 04 \end{bmatrix} = \begin{bmatrix} 207.190 & 60 \\ -32.119 & 16 \\ 1.394 & 04 \end{bmatrix}$$

$$\bar{P}_2 = \Phi \hat{P}_1 \Phi^T + Q(t) = \begin{bmatrix} 1 & 5 & 12.5 \\ 0 & 1 & 5 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 24.394 & 53 & 4.541 & 05 & 0.302 & 74 \\ 4.541 & 05 & 15.952 & 12 & 2.729 & 47 \\ 0.302 & 74 & 2.729 & 47 & 0.858 & 63 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 5 & 1 & 0 \\ 12.5 & 5 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.01 & 0 \\ 0 & 0 & 0.01 \end{bmatrix}$$

$$\bar{P}_2 = \begin{bmatrix} 952.522 & 241.835 & 24.683 \\ 241.835 & 64.723 & 7.023 \\ 24.683 & 7.023 & 0.867 \end{bmatrix}$$

$$K_2 = \bar{P}_2 H^T [H \bar{P}_2 H^T + R]^{-1} = \begin{bmatrix} 952.522 \\ 241.835 \\ 24.638 \end{bmatrix} \frac{1}{977.522} = \begin{bmatrix} 0.974 & 42 \\ 0.247 & 39 \\ 0.025 & 25 \end{bmatrix}$$

$$\hat{X}_2 = \bar{X}_2 + K_2 [\bar{z} - H\bar{X}_2] = \begin{bmatrix} 207.190 & 60 \\ -32.119 & 16 \\ 1.394 & 04 \end{bmatrix} + \begin{bmatrix} 0.974 & 42 \\ 0.247 & 39 \\ 0.025 & 25 \end{bmatrix} [234 - 207.190 & 60] = \begin{bmatrix} 233.314 & 50 \\ -25.486 & 61 \\ 2.070 & 99 \end{bmatrix}$$

$$\hat{P}_2 = [I - K_2 H] \bar{P}_2 = \begin{bmatrix} 0.025 & 58 & 0 & 0 \\ -0.247 & 39 & 1 & 0 \\ -0.025 & 25 & 0 & 1 \end{bmatrix} \begin{bmatrix} 952.522 & 241.835 & 24.683 \\ 241.835 & 64.723 & 7.023 \\ 24.683 & 7.023 & 0.867 \end{bmatrix} = \begin{bmatrix} 24.360 & 6.185 & 0.631 \\ 6.185 & 4.894 & 0.916 \\ 0.631 & 0.916 & 0.245 \end{bmatrix}$$

We can continue the example at 15 and 20 seconds. The final state and covariance are

$$\hat{X}_4 = \bar{X}_4 + K_4 [\bar{z} - H\bar{X}_4] = \begin{bmatrix} -22.304 & 37 \\ -19.264 & 26 \\ 1.110 & 63 \end{bmatrix} + \begin{bmatrix} 0.901 & 42 \\ 0.172 & 22 \\ 0.014 & 67 \end{bmatrix} [2 - (-22.304 & 37)] = \begin{bmatrix} -0.395 & 88 \\ -15.078 & 54 \\ 1.467 & 37 \end{bmatrix}$$

Notice that the train has passed the barrier ($r = -0.395$ m). Now,

$$\hat{P}_4 = [I - K_4 H] \bar{P}_4 = \begin{bmatrix} 0.098 & 58 & 0 & 0 \\ -0.172 & 22 & 1 & 0 \\ -0.014 & 67 & 0 & 1 \end{bmatrix} \begin{bmatrix} 228.606 & 43.676 & 3.722 \\ 43.676 & 9.334 & 0.855 \\ 3.722 & 0.855 & 0.095 \end{bmatrix} = \begin{bmatrix} 22.535 & 4.305 & 0.367 \\ 4.305 & 1.812 & 0.214 \\ 0.367 & 0.214 & 0.041 \end{bmatrix}$$

Table 9-6 summarizes the data for our problem.

TABLE 9-6. Residuals and Standard-Deviation Values for Example 9-6: Kalman Filter (Linear System). This table shows the values for each state update of Example 9-6.

Observation	$z - H\bar{X}$ (m)	Standard Deviation		
		position (m)	velocity (m/s)	acceleration (m/s ²)
0		15	5	1
1	-8.75	4.939 08	3.994 01	0.926 62
2	26.8093	4.935 65	2.212 23	0.494 97
3	-46.7687	4.842 51	1.704 69	0.292 23
4	24.3043	4.747 16	1.346 27	0.201 81

If we try different initial conditions and statistics, we'll find that the updated estimates at each time depend on

- 1. The number of measurements and the time between them.
- 2. The assumed statistics: measurement noise, R , and process noise, Q .
- 3. The initial conditions.



9.7.2 Linearized Kalman Filter (*LKF*)

When the Kalman filter is adapted for nonlinear systems, many of the same implementation problems arise as in nonlinear differential correction. Consult (Gelb, 1989, 182–189) for additional background.

The linear Kalman filter approach assumes that a reference trajectory generated from the *a priori* initial state vector is sufficiently close to the actual trajectory. This is similar to the idea behind Encke's method for orbit determination. The original state vector and covariance matrix define the reference trajectory over the time interval of the simulation. The displacements or corrections that we apply to the reference trajectory are estimated at each observation time. The *LKF* continues to provide the *current* state and covariance information, but doesn't recalculate the reference trajectory at each point. Rectification points can be inserted if the estimated trajectory drifts too far, but remember that we'll have to recalculate a new reference trajectory and reinitialize the covariance matrix. This ephemeris generation requires additional computing resources, so we don't want to re-initialize too often—in fact, a main benefit of the linearized Kalman filter is the reduced computational burden. This reduced burden makes it especially desirable for autonomous navigation systems and similar tasks.

With the linearized approach, we use the *initial* state vector (input) for all initial calculations, which keeps the state and the update separated. Although this method helps us determine where errors occur during a run, it can't handle large changes. Basically, then, the technique is useful for systems which require current estimates of the state and speed. Consider Algorithm 61. For the first iteration, we assume the predicted state update, $\delta\hat{x}_k$, is zero. Note that I've omitted the process noise, v , as an input. Typical practice assumes it

has a normal distribution with a zero mean. I've also shown the more accurate technique to determine Φ in the algorithm, but you can use an approximation.

$$\Phi(t_{k+1}, t_k) = \frac{\partial \bar{X}_{k+1}}{\partial \bar{X}_k} \approx \mathbf{I} + \mathbf{F}\Delta t + \mathbf{F}^2 \frac{\Delta t^2}{2!} + \dots$$

Alternatively, we can find Φ by numerical integration [Eq. (9-28)] that we can perform with the prediction of the state. Finally, I've shown **PKEPLER** as the propagator, but you can use any propagator that's suitable for your problem.

ALGORITHM 61: Linearized Kalman Filter

$$(\hat{X}_o, \hat{P}_k, Q, R_k, \text{obs} \Rightarrow \hat{X}_{k+1}, \hat{P}_{k+1})$$

$$H = \frac{\partial \text{observations}_{k+1}}{\partial \bar{X}_{k+1}}$$

$$F = \frac{\partial \dot{\bar{X}}_{t_k}}{\partial \bar{X}_{t_k}}$$

$$\Phi(t_{k+1}, t_k) = \mathbf{I} + \int_{t_k}^{t_{k+1}} F(t_{k+1}) \Phi(t_{k+1}, t_k) dt$$

Prediction

$$\mathbf{PKEPLER}(\hat{X}_o, t_{k+1} - t_o \Rightarrow \hat{X}_{k+1}) \equiv \Phi \hat{X}_o \quad \text{Predicted State}$$

$$\delta \bar{x}_{k+1} = \Phi \delta \hat{x}_k \quad \text{Predicted State Update}$$

$$\bar{P}_{k+1} = \Phi \hat{P}_k \Phi^T + Q \quad \text{Predicted Error Covariance}$$

Update

$$K_{k+1} = \bar{P}_{k+1} H_{k+1}^T \left[H_{k+1} \bar{P}_{k+1} H_{k+1}^T + R_k \right]^{-1} \quad \text{Kalman Gain}$$

$$\delta \hat{x}_{k+1} = \delta \bar{x}_{k+1} + K_{k+1} \left[\tilde{b}_{k+1} - H_{k+1} \delta \bar{x}_{k+1} \right] \quad \text{State Update Estimate}$$

$$\hat{P}_{k+1} = \bar{P}_{k+1} - K_{k+1} H_{k+1} \bar{P}_{k+1} \quad \text{Error Covariance Estimate}$$

$$\hat{X}_{k+1} = \bar{X}_{k+1} + \delta \hat{x}_{k+1} \quad \text{State Estimate}$$

This is the *classic* use of a Kalman filter for nonlinear systems. Notice the addition of the state-error estimate, $\delta\bar{x}$, from the linear system. Just as the nonlinear least-squares case includes a change in the state, the *LKF* produces incremental changes to the state. In other words, the *LKF* is linearizing about a nominal state just as the nonlinear least squares did. The Q matrix accounts for the uncertainty in the Φ matrix in attempting to model the nonlinear dynamics of the problem. The \tilde{b} matrix contains the residuals as in the problem using nonlinear differential correction (actual observations minus the observations from the state \bar{X}). Notice the use of $H\delta\bar{x}$ because we're now comparing it to the residual matrix, \tilde{b} . We find the \tilde{b} matrix as the measured observations minus the predicted value of the observations, $H(\bar{X})$, which for the orbit problem means finding ρ , β , and el from the state vector.

Although the state is updated in the update section, you *don't* use the update in the subsequent prediction! This simply means you're updating the state for whatever mission application is driving the *LKF*, but all computations are based on a nominal trajectory, without incorporating updates. This method is extremely fast because the reference trajectory is computed only once and stored ahead of time, allowing you to calculate H and Φ as required. The drawback is that the errors will grow with time (the reference trajectory will diverge from the truth), and at some point, you must reinitialize the state and generate a new reference trajectory. The Extended Kalman filter, discussed next, avoids some of these difficulties. Earlier, we saw that we often propagate the state using a numerical integrator. For the orbit problem, because there are simplified methods to propagate the state to the new time and older computers were slow, a past solution chose *not* integrate the state but instead used the Φ matrix for two-body motion and let $\delta\bar{x}$ absorb the dynamics errors, as well as the state corrections based on the observations. This idea should sound very familiar—it's the concept behind Encke's method for analyzing orbital perturbations. The rectification point is similar to the restart mentioned above. However, the consequences of these assumptions must be evaluated for each application and for current technology. The common approach may no longer be valid.

We still label this technique as differential correction even though iteration isn't present. We could reprocess the observations, but a performance improvement isn't guaranteed. As later observations arrive, the state vector continually changes to reflect that data.

9.7.3 Extended Kalman Filter (*EKF*)

Although the *LKF* is adequate for many applications, it may not be adequate for very nonlinear systems. Some orbit-determination problems have specific needs which we can meet using an Extended Kalman filter (often known as the Kalman-Bucy method). The *EKF* reduces divergence encountered by the *LKF* when large differences occur in the observational data. It does this by updating the reference state trajectory at each observation time. Consequently, the predicted state update is zero, $\delta\bar{x}_{k+1} = 0$, because there is no state update to be propagated. The prediction would then be $\delta\bar{x}_{k+1} = v$. In practice, v is assumed to have a zero mean. Thus, it's treated as zero for predictions but still influ-

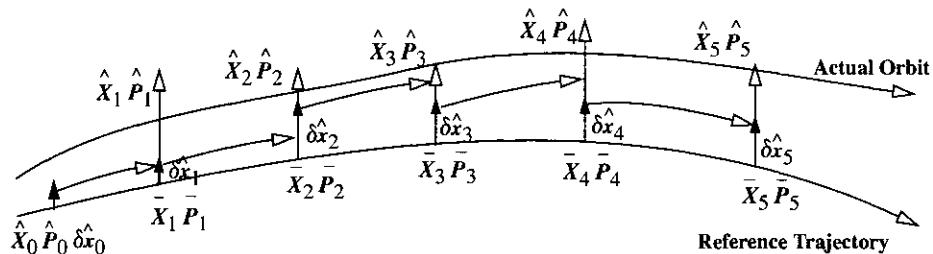


Figure 9-10. Representation of a Linearized Kalman Filter. The predicted, estimated, and state updates vary over time, so I've intentionally shown the values varying, not always converging, over time. Notice how we use only the state update estimates as inputs to determine subsequent state vectors.

ences covariance predictions. However, we must still include the second moment of the process noise, \mathbf{Q} . I show these features in Algorithm 62. Notice the continuing requirement to know the state and covariance before running the algorithm.

Unfortunately, the *EKF* requires that we restart the integrator at each observation time, and this can be computationally expensive. The choice of integrator has important implications for the type of propagator because multi-step methods have complex start-up techniques. I've shown *PKEPLER* to indicate one choice of propagator. A highly-accurate numerical technique may be needed for precise applications. Also notice that we do not use the *a priori* state estimate for the state propagations. The state transition matrix is used only to update the state update and the covariance matrix. I've also shown the more rigorous method to find Φ using numerical integration. The approximation will often work, but it does introduce additional error into the solution. Finally, because \mathbf{F} is no longer constant, we must recalculate it with each new observation. This adds additional processing at each step.

ALGORITHM 62: Extended Kalman Filter ($\hat{\mathbf{X}}_k, \hat{\mathbf{P}}_k, \mathbf{Q}, \mathbf{R}_k \Rightarrow \hat{\mathbf{X}}_{k+1}, \hat{\mathbf{P}}_{k+1}$)

$$\mathbf{H} = \frac{\partial \text{observations}_{k+1}}{\partial \hat{\mathbf{X}}_{k+1}}$$

$$\mathbf{F} = \frac{\partial \dot{\hat{\mathbf{X}}}_{t_k}}{\partial \hat{\mathbf{X}}_{t_k}}$$

$$\Phi(t_{k+1}, t_k) = \mathbf{I} + \int_{t_k}^{t_{k+1}} \mathbf{F}(t_{k+1}) \Phi(t_{k+1}, t_k) dt$$

Prediction

$\text{PKEPLER } (\hat{X}_k, t_{k+1} - t_k \Rightarrow \hat{X}_{k+1})$	Predicted State
$\delta\bar{x}_{k+1} = 0$	Predicted State Update
$\bar{P}_{k+1} = \Phi\hat{P}_k\Phi^T + Q$	Predicted Error Covariance

Update

$K_{k+1} = \bar{P}_{k+1}H_{k+1}^T \left[H_{k+1}\bar{P}_{k+1}H_{k+1}^T + R_k \right]^{-1}$	Kalman Gain
$\hat{x}_{k+1} = K_{k+1}\bar{x}_{k+1}$	State Update Estimate
$\hat{P}_{k+1} = \bar{P}_{k+1} - K_{k+1}H_{k+1}\bar{P}_{k+1}$	Error Covariance Estimate
$\hat{X}_{k+1} = \bar{X}_{k+1} + \delta\hat{x}_{k+1}$	State Estimate

Note how the prediction of the state uses the previous estimate of the state. Unlike the *LKF*, this filter uses the updated state for subsequent calculations, and the predicted state update, $\delta\bar{x}$, is not used. This technique simplifies the equations but doesn't make the filter run any faster. In fact, an *EKF* will usually run *slower* than an *LKF* because the Φ matrix must be recalculated at each step. In addition, the predicted value of the state must be updated at each step. As mentioned in the *LKF*, however, we can avoid a complex propagation by using a perturbed two-body approach. Except in some time-critical mission situations, increased computer speed has virtually eliminated concern about this additional processing, and the *EKF* enjoys wide popularity for its accuracy and speed. The δ matrix is the residual matrix defined for the *LKF* and the nonlinear differential correction. Although no iteration is required, it's still considered differential correction because we're estimating corrections to the previous estimate of the state. Figure 9-11 graphically depicts the state and state updates over time.

If a real-world filter diverges, it's usually because the filter mismodels the application (either deterministically or statistically), or the initial filter estimates were outside the radius of convergence. For example, if the filter doesn't "know" a maneuver has occurred, it will diverge in the first track following the maneuver, which becomes a mechanism for detecting the unknown maneuver. Another example of potential divergence could be an *EKF* using a white-noise hypothesis for process noise, but this is an incorrect assumption because the process noise is time-correlated. In this case, the filter might diverge or, at least, it won't converge. A filter may not converge if the initial state update is very large, outside the interval of convergence (in which case iterated least squares will also diverge). A filter may *not* converge if the measurement representation is incorrect. An example is if we assume measurement errors are due only to white noise and a constant bias, when the bias actually varies greatly over time. Least squares also behaves erratically in this case. It

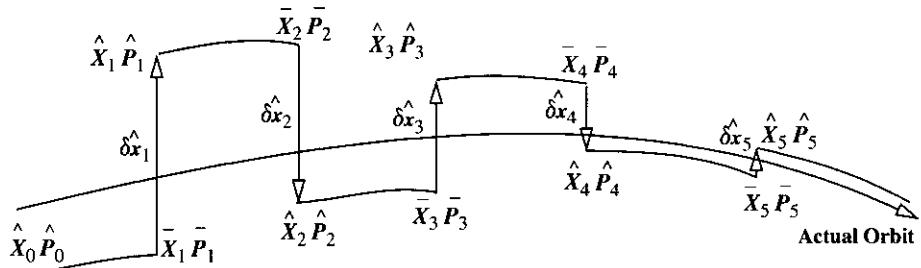


Figure 9-11. Representation of an Extended Kalman Filter. In this ideal representation, notice how all the predicted, estimated, and state updates converge over time. Normally, filter convergence is not guaranteed.

can perpetually reject half the data on one iteration, then accept it and reject the other half on the following iteration.

9.7.4 Successful Filtering Applications

Swerling and Kalman were interested in generalized filtering problems, whereas Gauss was interested in orbit determination. Applying filters to orbit determination has not always succeeded. Failures generate a negative “folklore” about filters, which is evident in today’s debates about orbit-determination methods. For example, during the Mariner project, an onboard *EKF* was used for navigation. When the ground controllers sent a mid-course correction, the filter completely ignored it—because it had become *smug*. Smug means the P matrix (and therefore the Kalman-Gain matrix) was very small, so the *EKF* believed that all states were very well known and no updates were necessary! Several retries didn’t get the data accepted. The ground controllers had to turn the filter off and restart it to initialize P . Incidents of this type have given filters a reputation of being fragile, subject to instability, diverging, and restricted to problems having dense data. More recently filters have been very successful. The orbit-determination algorithm that is part of the GPS’s ground-control system is a filter, and the Jet Propulsion Laboratory uses a filter to obtain very precise ephemerides for GPS satellites and for the TOPEX satellite.

Filters have a persistent reputation of being hard to tune, especially in setting the Q parameters for process noise. But a new and challenging approach strives to define the process noise as a function of the system’s physical processes. One example is the work of Wright (1981, 1994). Wright recognized that process noise in the orbit problem is a real physical phenomenon that arises from uncertainty in the force models. Wright shows how to derive a generalized process-noise function for the errors in the gravity model, whether due to truncating the model (which we do to save computer time) or to our inability to perfectly model the Earth’s gravity. For the gravity problem, Wright relies on work done by Kaula (1959), a geodesist, who derived a statistical model of acceleration error at the

Earth's surface. A simple example illustrates the concept. If we model a 4×4 gravity field, the process noise error is the contribution of the gravity field *above* 4×4 . If we limit the field to 50×50 , we can calculate the process noise "exactly." These physically defined filters don't enjoy wide support, partly because people must know astrodynamics *and* estimation to exploit the concept. The authors of physically defined filters claim that most issues of stability, robustness, and fragility have been solved. They advocate deriving process-noise estimates from the problem definition, as opposed to adopting a filter structure and hoping that it fits the problem. They would say that the Mariner filter failed because it wasn't properly designed for the problem.

Effectively, Wright notes that the state updates and dynamics are correlated:

$$E[\Phi \delta \hat{x}_k v^T] \neq 0$$

and the process noise isn't white, as Kalman postulated. He uses equinoctial orbital elements and gives the expression for the predicted error covariance adding a symmetric matrix of the orbit error covariance function (1994b).

$$\bar{P}_{k+1} = \Phi \hat{P}_k \Phi^T + P_{k+1, k}^{ff}$$

This appears to be of the same form as the Kalman time update [Eq. (9-32)], but the final matrix is really a complex function of the spherical harmonic potential, variances in gravity, and autocorrelations (first moments of the distribution, Gelb, 1989, 36) that derive from them. Consult Wright (1981, 1994) for the remaining formulation.

9.7.5 Summary of Kalman Filters

The Kalman filter and Extended Kalman filters have good and bad points when compared to least squares and sequential least squares. Remember that, for most applications, we're really comparing filtering and smoothing techniques. The following are advantages of the Kalman filter over least squares and sequential least squares:

1. It's a very fast, recursive process that is ideally suited for on-line control of a system because it doesn't require iteration (Viking Lander, Apollo Lunar Landers). In view of the interest and success of the Global Positioning System (that we'll discuss in Chap. 10), these systems are being integrated into military tactics. This integration often needs orbit determination in real time, which means a filter.
2. It estimates the current state and covariance matrix, which has many benefits. Hujšak (1995) has shown that a Kalman filter can be used effectively to process *through* a maneuver, and the post-maneuver orbit accuracy improves with each observation.
3. It may require less memory storage and processor time, but that depends on the computer software.

4. It allows solving for measurement biases simultaneously in the state space, which accommodates the true time-varying bias. This is true only if the bias is observable enough--there must be enough data to determine the bias. Least-squares techniques can fail if we try to solve for biases simultaneously with the orbit, usually because we don't have enough data. Insufficient data can also affect filters.
5. It can model time-varying coefficients of drag and solar-radiation pressure. Although a least-squares technique can solve for a constant bias or average parameter during the fit span, the approach is at best only piecewise continuous. This advantage also assumes we have enough data to arrive at a solution.

But Kalman filters do have a few disadvantages:

1. The covariance matrix, and hence the Kalman gain, can become too small, causing it to ignore new data. As mentioned earlier, this is called filter *smugness* or *divergence*, which we can overcome in various ways. The quick solution is to increase Q . The classical covariance recursion is numerically unstable in many applications, which only adds to this problem. Investigators like Jim Wright would reexamine how to form Q from the dynamics of the problem. Others would invoke the Joseph algorithm, fading-memory filters, or square-root formulas, which I'll introduce shortly.
2. The filter isn't *self-starting*; to begin, it requires an initial estimate of the state, X , covariance, P , plus the process noise, v , power-density matrix, Q , and measurement error, R . Notice we don't need an estimate for the state update because we assume it's zero on the first pass. We can use least-squares differential correction as the starting point, but this may be time-consuming.
3. We should determine the Q matrix from the physics of the problem. This can be very difficult depending on the fidelity of our dynamics model.
4. If our data is very noisy, we may not be able to converge on a solution because noisy data has a large matrix for measurement error, R , which causes us to ignore new data if the filter is smug (see #1 above). If the data is very poor, both a well-designed filter *and* a least-squares technique will fail.

We've discussed some of the issues in a recurring debate which centers on whether to use a least-squares or a filter approach. Either algorithm may apply to a given problem, so carefully review all mission requirements to help determine which one best meets your needs.

9.8 Practical Considerations

Precise statistical orbit determination is extremely complicated. The real world presents many difficulties which compound existing problems. I'll examine four aspects of estimation: data, implementation, results, and applications.

9.8.1 Data

We've previously discussed that sensor observations will not always be available when needed. Maintenance, downtime, and various personnel actions can influence quality and, in some cases, prevent the observations themselves from reaching the user. Unfortunately, we can't predict these effects inside a very large system. One practical solution is trying to gather as much raw information as possible, so we'll have a backup whenever normal procedures fail. Recall the results from Example 9-2. When we removed the obviously incorrect point, the RMS improved but the standard deviation worsened. If we had had more data, RMS and standard deviation *might* have both decreased.

Closely related is the amount of data observed for a satellite. You might be tempted to believe continuous data exists for all satellites. This is just not true; data is often limited for satellite orbits. A difficulty arises with geosynchronous and repeating-groundtrack satellites. As we saw in Chap. 6, if we can observe only a small arc of a satellite's orbit, it's much more difficult to determine an accurate answer. This difficulty can lead to mismodeling of the orbit. Too little data can also result in our inability to estimate additional state parameters. For instance, we need more data if we want to solve for biases in addition to the state vector.

In general, more tracking data should give better accuracy. However, this assumes that the biases are identified and removed, and that the relative accuracy of each observation is known and used in the weighting process. If the orbit degrades with additional data, the model and the calibration are probably the cause. Additionally, suppose a sensor reports only two or three data points. Obviously the sensor receives much more data—often hundreds or thousands of points per pass. One approach to obtain more data could be to task more sensors to observe the satellite and report the sparse observations. Although that would give us a slight improvement, we could incorrectly conclude that sparse data from many sites permits *highly accurate* orbit determination. This is simply false because there is a trade-off between data density, number of observing stations, and the length of the desired predictions. Quite often, *denser data* (even from single sites) can actually improve the quality of the orbits and reduce the overall sensor tasking when combined with accurate biases and proper numerical processing. Fonte (1995) showed that dense, real-world observations from a single station could produce orbits accurate to less than 10 m for a 12-hour prediction. This indicates we need dense observations (perhaps on the order of hundreds per pass) to produce precise orbits (10 m to 100 m for all satellites). Sparse data (less than five observations per pass) even from multiple sensors can actually increase the error to over 400 m for low-Earth satellites.

We should also examine the general accuracy we can get on a satellite's pass. Figure 9-12 shows a greatly exaggerated satellite pass in which three points of the orbit are high-

lighted. Although Fig. 9-12 shows only one pass, many passes at low, medium, and high elevations can eliminate some of the errors resulting from using just one pass. The types of errors may not be obvious. When the satellite is very low to the horizon, a small vertical error can result in a large uncertainty about where the satellite is in its orbit—the along-track component. As the satellite travels over the site, a small horizontal error will become a large plane change or a cross-track error. This effect is magnified because the satellite is usually closest to the site at its maximum elevation, or *culmination*. If we combine these results, we get an *error ellipsoid* about the satellite. In general, along-track errors are greatest because we can't get precise timing information. Cross-track errors are usually better, typically resulting from a sensor's misalignment. Radial errors are usually the smallest.

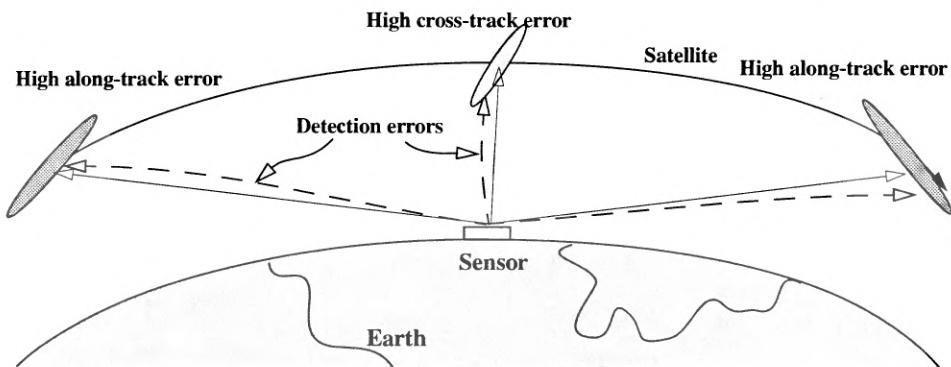


Figure 9-12. Sensor Viewing Accuracies on a Pass. Several viewing constraints affect the accuracy of the data a sensor receives. Over many passes, the advantages and disadvantages tend to cancel out. Whenever the satellite is low on the horizon, the along-track values are most difficult to resolve. Whenever the satellite is directly above the site, the cross-track errors tend to dominate.

The question often arises, “Which type of data is more accurate?” Admittedly, range data is very important. As we saw in Chap. 6 (initial orbit determination), the availability of range is crucial to elegantly transforming data. Does this mean we should ignore angular data? Certainly not! By including all available data, we’re much more likely to find the global minima and the best solution. We can estimate a trade-off between range and angular measurements using Fig. 8-14. Consider a satellite at 1000 km altitude. If our range measurements are accurate to about 50 m, we would need angular measurements of at least $10''$ to provide the same precision. Of course, there are subtle differences in the processing and the answers vary with the satellite’s altitude, but this technique allows us to form an initial evaluation.

The ideal availability of observations is to have simultaneous range (and range rate) data from multiple locations. The resulting triangulation would be very accurate, but

unfortunately, the process isn't practical except for some applications of GPS. An alternate approach uses multiple sensors to obtain observations. This process allows the satellite's motion to enhance the *observability*, and gives geometric diversity to the observations. **Observability** indicates the sensitivity of the state parameters to the observations. We can evaluate it by examining the H matrix. The partial derivatives give us the observability for each measurement. Partials which vanish, or have small magnitudes, provide no information to the estimation process. We should eliminate them because they can affect the final answer. Remember, this sensitivity analysis requires no actual observation measurements. We must evaluate only the observation partial derivatives, which depend on the observation models and the *a priori* state vector. To illustrate observability, consider a geosynchronous satellite observed from the Earth. The relative geometry between the satellite and the site remains essentially fixed. If only range measurements are available (as is often the case), they will be very insensitive to any out-of-plane motion.

Sensor Calibration

Determining the actual sensor biases and noise statistics is difficult, but it strongly affects the accuracy of differential correction. Although we tend to view data in Table 9-3 as *constant values*, they're actually only estimates of the true time-varying values calibrated through many mathematical operations. Some systems (such as GPS and SLR) do this technically demanding process well; others do it poorly. A high-quality approach appears in the Lincoln Laboratory's studies of calibration biases, in which they use LAGEOS and SLR data to produce a definitive reference. You should always try to apply external references, if they exist! The fact that the noise and biases change over time—with different satellites and in certain geometries—doesn't mean the values should be ignored; instead, we should examine them closely to detect large variations which may indicate more severe (and correctable) problems. Remember, least-squares techniques will estimate the average values, whereas filters can better track the time-varying values. In each case, we must have enough data to permit a solution to be found. Too little data doesn't permit us to determine the state and bias values.

The value of correct calibration has been known for decades. Recently, Fonte (1995) has suggested that using incorrect biases can produce almost a 70 m error in a low-Earth satellite's location. Miscalibration may result in a very good site appearing bad and vice versa. It would be unwise to ignore this large an error if we could solve it simply by determining the correct calibration values.

9.8.2 Implementing a Differential-Correction Technique

Tying any estimation scheme to a propagator presents yet another practical concern. Most real-world estimation systems are very complex, especially if we use highly accurate and analytical partials. Linking this large system with a highly accurate propagation scheme may also present some unique challenges. First, we must consider which parameters the computer program requires as formal parameters. A good system should always use a

highly accurate propagator. Any technical changes or updates shouldn't need lengthy (and costly) periods of reprogramming.

The next consideration is operator knowledge. The nuances of a mathematical technique *may* be lost by operators who never study the technical details. This could lead to incorrect system-level decisions by people based on their knowledge of a particular software program, and not the technical details. It's nice to view an operational system as being able to reproduce theoretical accuracies, but limitations in any large system make this view unrealistic. Often, operators don't have the luxury of time to adjust, recalculate, and iterate many times until they get a precise solution.

For differential correction, you should never use a propagator whose accuracy is less than the best observations you receive. In fact, your propagator should be at least an order of magnitude *more accurate* than your best observations. This will ensure that the observation residuals are minimized, not the theory! On the contrary, for partial derivative calculations, using high-fidelity force models beyond those necessary to capture your best observations offers little advantage in accuracy but greatly increases the computer's burden. Don't arbitrarily reduce the force models without evaluating (and documenting!) the rationale.

Post-processing vs. Prediction

The results we obtain from observational data differ greatly between post-processing and prediction. ***Post-processing*** refers to differential-correction runs using observations to determine satellite positions in the past. All atmospheric and solar parameters have been measured. Because we know the physical parameters (atmosphere and solar radiation), the estimation processing can "concentrate" on the satellite parameters and is very accurate. Many papers within the literature discuss orbit determination for the TOPEX orbit achieving radial position accuracies around 3 cm. We expect this accuracy for a many low-Earth orbits without *significant* drag perturbations in a *post-processing* mode.

Prediction involves the opposite problem—that of using observations to determine a satellite's future positions without knowing all physical parameters. Future locations for the TOPEX satellite can't reliably produce orbit accuracies better than about 3 m for one-day predictions. Although this seems small, consider that it's 100 times worse than the post-processed answers! Here, the estimation process ends up incorporating poorly known physical parameters with the state—a process that usually degrades the accuracy. The large discrepancy exists because we don't have accurate predictions for atmospheric and solar parameters. It's extremely important to recognize which type of processing is taking place when trying to assess a particular scheme's accuracy.

Update Intervals and Fit Spans

Tracking systems often use special methods to determine when to update a satellite's orbital elements. I'll define an ***update interval*** as the amount of time which may pass before we must take more observations and *update* the element set. We can look at this

interval in two ways: (1) one site can observe, process, and re-acquire satellites, and (2) many sites can take lots of observations, and each can update a centralized set of elements.

For the case in which we have only one site, our options are limited. We must take data and process it with enough regularity to continue successful acquisition. Multiple sites offer the opportunity to skip observation passes among the sites.

The length of the update interval depends on the quality and quantity of observations, the estimation and propagation schemes, and the type of sensors. Ideally this interval should occur when the propagation theories would introduce enough error to keep the sensor from re-acquiring the satellite, or when the most stringent accuracy requirement is no longer met for the satellite. Unfortunately, this is often not the case; the update-interval times are often several times shorter than the actual requirement. If the forces are modeled correctly, much longer update intervals are possible, sometimes weeks in length. Propagation techniques with poorly modeled forces require shorter update intervals because the estimation process can't accommodate the large errors resulting from inadequate force models that grow over time. Considering the real-world availability of data, this difficulty can be important.

Closely related is the concept of a fit span. A *fit span* is defined as the period of time over which we must collect all the data selected for a differential correction. Be careful not to confuse this with the update interval. The fit span often includes many update intervals, as shown in Fig. 9-13. Schumacher and Glover (1995) describe a method to determine the fit span for analytical theories based on the satellite's period and rate of decay. Other methods use the eccentricity and the period. For nonmaneuvering satellites,

<u>Period</u>	<u>Fit Span</u>
$P > 800 \text{ min}$	30 days
$600 < P < 800 \text{ min}$	15 days
$\dot{P} > -0.0005 \text{ min/day}$	10 days
$-0.001 < \dot{P} < -0.0005 \text{ min/day}$	7 days
$-0.01 < \dot{P} < -0.001 \text{ min/day}$	5 days
$\dot{P} < -0.01 \text{ min/day}$	3 days

where we've used

$$\dot{P} = \frac{-4\pi}{n^2} \left(\frac{\dot{n}}{2} \right) (1440 \text{ min/day})$$

The bottom line is that we'd like to get as much data as possible. Once the orbit is established, we continue to use a high-fidelity propagator to preserve the accuracy and allow larger update intervals. We can adjust the fit span to meet accuracy requirements.

An important example where we want to adjust the fit span occurs when we try to separate the effects of atmospheric drag, c_D , from the estimate of the semimajor axis. There is a strong correlation between the semimajor axis and drag. We'll see later that this drag effect is quadratic in the along-track direction, while a (through M) causes linear variations in the along-track motion. If the fit span is too short, we can't separate the contribu-

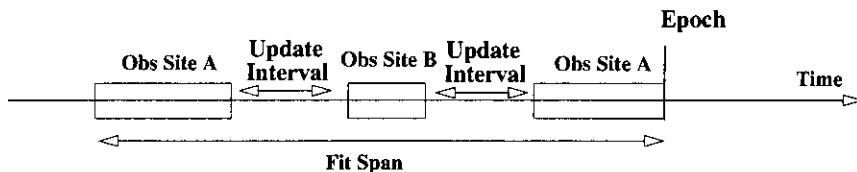


Figure 9-13. Depiction of Fit Span and Some Included Update Intervals. Remember that the fit span may be several days long. Each pass may be only 10 to 15 minutes for low-Earth and 8 to 10 hours for deep-space orbits; update intervals can range from hours to days but are usually shorter than the fit span. The epoch is usually placed at the end of data so it's current.

tion from drag. As the time increases, we begin to be able to separate the quadratic and linear changes in the along-track motion.

In general, for incomplete theories (those which have ignored or truncated many perturbation terms of the original derivations), longer fit spans are often used to sample enough points to average the mismodeled forces. Unfortunately, the longer the fit span, the greater the potential error introduced by the propagation routines used to simulate the satellite's motion through time. If we use a shorter fit span, data may be insufficient to accurately separate the periodic variations from the secular effects. This can potentially cause a bias in our estimate of the initial state. In addition, each estimate causes a discontinuity in the orbit solution. A “correct” fit span is difficult to find for all orbits and ultimately involves a trade-off between the accumulated errors and the discontinuity of solutions (Gabor, 1993, 62). The overall solution is to sample for a time period that covers all relevant perturbations.

The type of measurements is also important for the estimation algorithm. “Traditional” measurements (e.g., ρ , β , el) have distinct initial orbit determination techniques. We can also process “nontraditional” observations using a nongeometric approach. For example, consider processing range-only data—a very common operation. If we have an initial estimate of the state, we can propagate the nominal state vector and generate computed range “observations.” By comparing these values to the actual observations, we can complete a least-squares solution. This process enables us to process a variety of data, including range-only, range-rate only, and even position and velocity vectors. We can also use the latter method to convert between mean and osculating orbital elements (Sec. 8.9.2).

9.8.3 Analyzing the Results of Differential Corrections

The principal method to evaluate an estimation solution is through the covariance matrix. If you've used correct numerical techniques, calibration, and so on, the biases of your observations should be slightly better than your covariance matrix (noise and drift are the difference).

Although the covariance matrix gives us a first look at differential correction's accuracy *at the epoch*, over time the error will grow, perhaps significantly. However, some variations, like cross-track errors, are periodic and do not grow over extended time periods. Error growth depends on force model errors, and on initial errors in the covariance. Figure 9-14 shows the growth of an example satellite's covariance matrix. Der and Danchick (1996) show numerical examples for this growth on several satellites. Junkins et al. (1996) also show the shape of the covariance matrix over time.

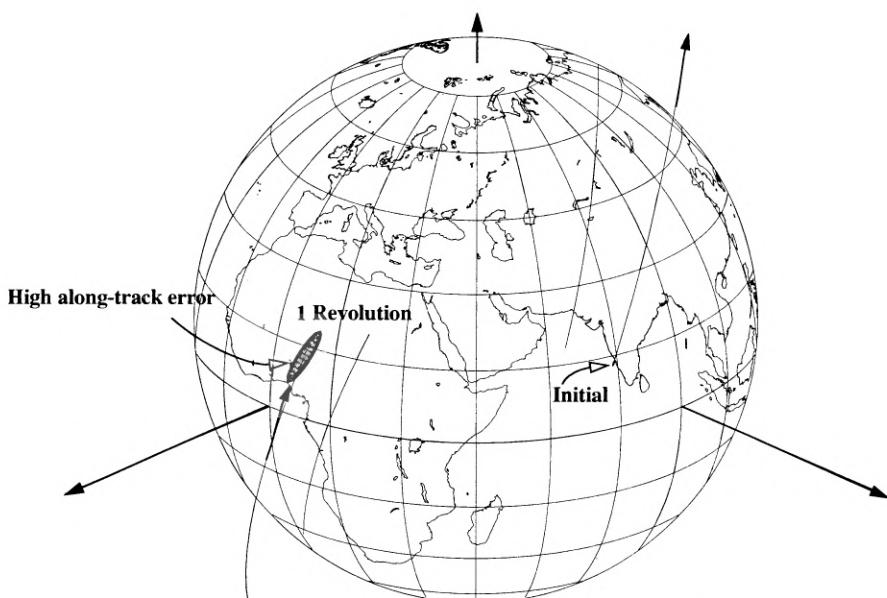


Figure 9-14. Growth in Error Covariance. I've shown an exaggerated view of how an initial error at epoch grows over time (one revolution). The three-dimensional object around the satellite depicts the uncertainty with which the satellite's position is known. Accurately determining this time-varying quantity is extremely important to finding probabilities associated with orbital events. The initial error (RMS) is about 74 km, and the final error is about 740 km.

The growth in the error covariance is best understood by examining both velocity and position errors. If the initial state contains velocity errors, they integrate into the position over time. Errors in the position affect the gravity accelerations and introduce additional velocity errors. Recall that the covariance matrix contains standard deviations (squared) for each element of the state space. This means we are given an uncertainty for each element at the epoch time. As we propagate the state space through time, the growth of these errors gradually erode our confidence in the answer. Even if our original state-space answer was perfect, it would still degrade over time due to force model errors. The presence of an initial error band only accelerates that degradation. If we use a Kalman filter,

we can try to estimate the amount of the error. The problem then shifts to the original determination of the covariance data. If it comes from least squares, it may not be consistent with the filter approach because least squares usually accounts for un-modeled dynamic errors by adjusting the initial state vector. Thus, the uncertainty in the final estimate of the state vector reflects (in part) the average error in the dynamics. Finally, it's important to recognize that filter applications using a fading-memory filter don't preserve the correct covariance matrix [Gelb, 1989, 287].

Convergence Issues

The iterative process of estimation doesn't always converge. Sometimes we can perform additional processing to overcome this problem. Editing the noisiest observations sometimes corrects the problem, but we must be careful, especially if there are only limited observations. We can also improve the accuracy of the partial derivatives and the fidelity of the mathematical algorithms (e.g., matrix inversion). If the estimation still doesn't converge, our initial estimate may be too far off, or the observation quantity or quality may be the cause. Although the observation quantity is usually the problem, insufficient calibration and poor observability can affect the process.

For systems using data from observations over long time periods, the covariance matrix can become zero, and even negative. Negative matrices are fatal for computations, and they usually occur if the filter is started incorrectly, or the filter tuning is wrong. These problems lead to severe difficulties for matrix inversion. There are several solutions. For sequential batch and least-squares techniques, square root matrix inversion algorithms, singular value decomposition, and fading-memory filters are useful. For Kalman filters, fading-memory, Joseph-algorithm [Eq. (9-37)], and square-root formulas work well.

Fading-memory filters weight older information less than the current information. We want to ensure the covariance matrix is never negative, or zero. For the Kalman filter, Gelb (1989, 286) shows how to evaluate a fading-memory variable, m_{fade} , which is

$$m_{fade} = \exp^{\frac{\Delta t_{obs}}{\tau_{age}}}$$

Here, we use the time between observations, Δt_{obs} , and the constant-age weighting factor, τ_{age} . In essence, we're increasing the measurement error for the older observations. The effect produces an exponential decay on the use of older data. The only difference is in the predicted covariance matrix, which becomes

$$\bar{P}_{k+1} = M_{fade} \Phi \hat{P}_k \Phi^T + Q_k \quad (9-38)$$

To use a fading-memory filter for least-squares techniques, recall the sequential-batch formulation [Eq. (9-20) and Eq. (9-21)]. The previous data could be counted less by simply multiplying by a square matrix, M_{fade} , with values slightly less than 1.0 on the diagonals and zero elsewhere. The difference of each coefficient from unity would represent the amount of significance for each term. This simplistic example is just one of many existing

methods. The qualitative reason for weighting older data is to tell the Kalman filter that the dynamical model is imperfect, and therefore the extrapolated dynamics are suspicious. This is reflected, to some degree, in M_{fade} .

Square-root filters are a more robust numerical approach to propagate poorly conditioned (near singular) covariance matrices. They create another matrix, S , which is related to the covariance by the *Cholesky decomposition*

$$\mathbf{P} = \mathbf{S}\mathbf{S}^T$$

Gelb (1989, 306–308) derives an update recursion for \mathbf{S} to replace Eq. (9-36) and shows an example of a poorly conditioned covariance matrix for which the conventional approach produces a \mathbf{P} matrix with negative diagonal values (a singular matrix). The square-root formulation processes the information without difficulty. Consult Gelb (1989), Bierman (1977, 68–112), or Maybeck (1979, 368–405) for a more detailed discussion of square-root filters.

A related issue is **uncorrelated tracks**, which occur when we can't associate new observations with existing orbits. Khutorovsky (1996) describes “the cause of mistakes in allocation of radar measurements to orbits or tracked satellites can be only the errors of predicted orbit.” Once again, we see the importance of propagation methods in the estimation process.

9.8.4 Applications: Detecting Maneuvers

We've examined maneuvers in detail in Chaps. 5 and 6. However, maneuvers can present a difficulty for differential-correction techniques. It's unlikely that we'll observe a satellite *during* a maneuver, but if we obtain data on both sides of a maneuver, the resulting orbit could contain significant errors. For the case where a maneuver is planned, we can easily obtain data after the event is completed and update the orbit. Hujšak (1995) has shown that a Kalman filter can actually process *through* a maneuver. This capability is a significant advancement to space surveillance. Unknown maneuvers do present some difficulty. One approach is to take a few observations each day to ensure the orbit never degrades too long. Unfortunately, this is a very inefficient use of resources and larger maneuvers (like geosynchronous transfer orbits) may still place the satellite outside a sensor's field of view. I've suggested that numerical methods are preferred in all aspects of differential correction, and this allows us to extend the update interval to ease tasking requirements on sensors. You could argue that a maneuver occurring right after the last observations could affect your ability to re-find the satellite. Remember, however, that the Navy's interferometer system is a “fence” which detects all objects passing by. Even if a satellite maneuvered after the last observations, the next revolution would expose a difference from the predicted entry/exit time at the fence. In fact, the fence performs particularly well in detecting tough maneuvers like geosynchronous transfer orbits. Of course, I'm assuming that we're using highly accurate propagation techniques to maintain the orbit. The difference in time of arrival at the fence could be used to perform either a least-squares or filter solution, re-estimate the orbit, and dynamically assign assets to track the satellite. This

technique allows us to rapidly re-establish a satellite's orbit after an unknown maneuver, while minimizing unnecessary sensor taskings.

Problems

1. Your company has launched a new satellite from Cape Kennedy into a circular, 45° , inclined orbit at about 500 km altitude. Which sensor sites would (a) have the best chance at initial acquisition one revolution after launch and (b) be most likely to acquire the satellite during routine orbit maintenance? Why?
2. What are the terms of the F matrix (analytical technique) for a state space of position and velocity vectors including solar-radiation pressure? (Hint: see Chap. 7 for a discussion of the equations of motion)
3. We know that drag is a difficult problem to model. How can you get the highest accuracy (~ 30 m), assuming your observations are reasonably good (~ 10 m)? Should you process more data, process more often, solve for c_D , develop new atmospheric models, determine new coefficients to "absorb" the unknown effects to drag, use a filter with realistic process noise, other? Hint: review Sec. 8.6.2 and list the trade-offs for each factor that affect atmospheric drag.
4. Design a least-squares technique for the angles-only method of orbit determination (Sec. 6.4.2). Using the following data from a site at 40° N, 110° W, 2000 m high, what are the orbital elements?

JD	h:min:s	α_t°	δ_t°
1996 2 26	07:40:00.0	48.7539698	-4.5195195
1996 2 26	07:00:00.0	53.7702524	-4.2555627
1996 2 26	08:20:00.0	58.7873608	-3.9997397
1996 2 26	08:40:00.0	63.8054128	-3.7540275
1996 2 26	09:00:00.0	68.8244967	-3.5203280
1996 2 26	09:20:00.0	73.8446671	-3.3004530
1996 2 26	09:40:00.0	78.8659412	-3.0961098
1996 2 26	10:00:00.0	83.8882968	-2.9088868
1996 2 26	10:20:00.0	88.9116713	-2.7402413

5. Sec. 9.5.1 introduced how to calculate the analytical expressions for the partials of the observations with respect to the state. What are the remaining partials for range, β , and el for a state vector consisting of position and velocity vectors? How does your answer change for right ascension and declination?

6. Your corporate manager has placed you in charge of a large (but old) tracking system. You control many assets (radar/optical, etc.) and relate well to the SLR community. Your system uses a sequential-batch (fading-memory) differential correction with a moderate analytical propagation theory (J_2 , J_3 , J_4 , exponential drag, some patches for certain orbits). Most of your computer software is poorly written and hosted on an obsolete computer, but you have access to newly created numerical routines and a brand new distributed processing work center. Your company is trying to win a contract to supply highly accurate element sets to a new communication-satellite company (about 800 satellites in ~800 km circular orbits), and you invoke a quality-improvement program. Of the following options, which have a good chance of increasing your chances of winning the contract? Explain each.
 - a) Do nothing—it's too hard and you'll rely on old customers
 - b) Switch to a better propagator (numerical/semianalytical)
 - c) Write another patch for your existing analytical theory
 - d) Switch to full numerical processing on the new distributed processing system
 - e) Obtain better calibration data
 - f) Lengthen or shorten the update intervals
 - g) Lengthen or shorten the fit spans
 - h) Improve the star catalog
 - i) Build new sensors or close older ones
 - j) Take more or fewer observations per pass
 - k) Use a Kalman-filter technique with a mathematically derived process noise
 - l) Create a Web location for all observations to reside so each sensor can numerically process data in near-real time and place the updated osculating vectors and covariance matrix on the Web
 - m) Do something else
7. For Example 9-4, what are the look angles for a site at Kaena Point on January 31, 1995 at 21:00 local time (*UTC*)? Would you expect your answers to be different for a site at Ascension? Explain your answer. (Hint: use propagation techniques from Chaps. 7 and 8 and see Sec. 10.7.)
8. Use Eq. (9-7) and Eq. (9-8) to derive an expression for the covariance matrix.
9. Use the observations from Kaena Point on January 29, 1995 for GEOS-III in Table 9-5 to find how the answer changes as eight more observations arrive, after those calculated in Example 9-4. Program both the linearized and extended Kalman filters and compare their answers at each observation time. Remember that the update changes with each new observation time.

CHAPTER 10 REAL-WORLD MISSION ANALYSIS

- 10.1 Introduction
- 10.2 Mission Orbits
- 10.3 Geometries for Surveillance and Reconnaissance
- 10.4 Designing and Maintaining Mission Orbits
- 10.5 Navigation—the Global Positioning System
- 10.6 Predicting Satellite Look Angles
- 10.7 Determining Close Approaches
- 10.8 Rise / Set

10.1 Introduction

Military applications in space began when the Soviet Union launched the first Sputnik satellite in October, 1957. The utility of high-altitude platforms was well known and largely responsible for the 17th century's development in balloons and the 20th century's development of aircraft. But applying such platforms to space poses interesting and new challenges to military planners.

This chapter provides insight into some real-world applications which may be useful when putting together these “back-of-the-envelope” programs. We’ll examine *specialized orbit design* (Secs. 10.4 and 10.6) and *applications* (Secs. 10.7–10.8). The section on specialized orbits focuses on the analysis required to craft an orbit for a particular mission. There are actually two phases. First, we examine the effect of perturbations on the nominal orbit over a certain time interval. Then, we determine appropriate maneuvers to keep the orbit within desired tolerances. I’ll explore the maneuver portion only with the repeat-groundtrack orbits because it’s very lengthy for most precise applications. Although there are many classes of orbits, this chapter treats just three. The mission-applications algorithms are useful for analyzing orbital motion and characteristics related to other satellites and sensors. Although most orbit analyses consist of propagation and orbit transfers, which we’ve already discussed, we need to know about other analytical tools that help engineers plan certain studies.

First, we’ll look at the overall concepts of reconnaissance and surveillance from space. I’ll then discuss some specialized classes of orbit design. Larson and Wertz (1992, 177–181) define several types of specialized orbits: Sun-synchronous, navigation, frozen, and repeat-groundtrack satellites. Keep in mind that most mission-design concepts are based on a simplified analysis of the effect of perturbations on the orbit. Actual mission planning

would consider these results only a starting point for developing detailed mission studies. A very important mission that uses these techniques is surveillance of the Earth, using satellites such as Landsat, RADARSAT, and SPOT.

Satellite mission design and support tie together many of the concepts presented in this book. Figure 10-1 shows the interactions.

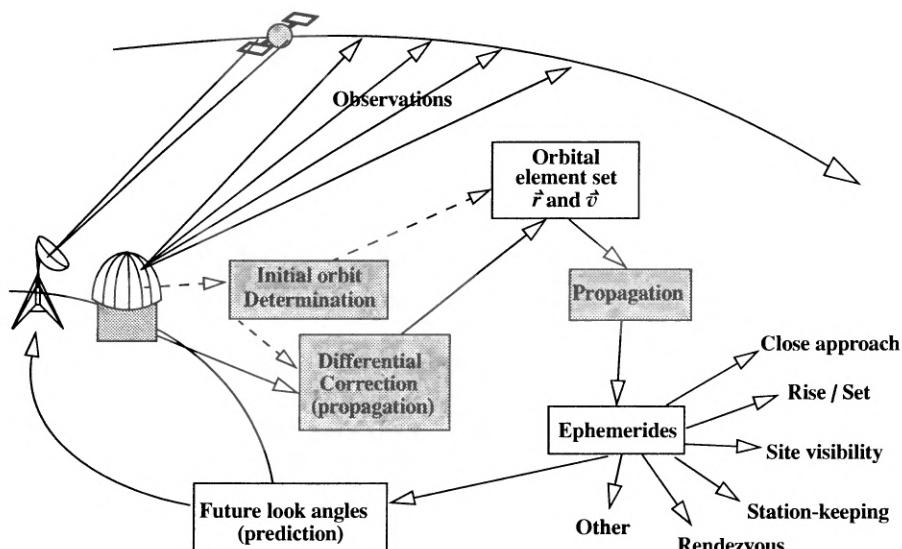


Figure 10-1. Developing Mission Support for Satellites. This figure shows the interrelations of initial orbit determination (Chap. 6), propagation (Chap. 8), and differential correction (Chap. 9). Analysis of mission applications, shown at the lower right, requires observations, orbital element sets, and ephemerides. Future observations complete the cycle.

Figure 10-1 also shows the flow of operations for an orbit-determination system. I've shown initial orbit determination as an option before differential correction. Because we consider differential correction to be the routine processing of observations, we need an initial estimate of the orbit. If we lose track of a satellite (from either an unknown maneuver or faulty predictions), subsequent tracks are called ***Uncorrelated Tracks*** (UCT) because the data doesn't match any of the existing information. When this occurs, we must use techniques for initial orbit determination to locate the new orbit. Hujšak (1995) has shown that it's possible to perform orbit determination over time intervals that encompass maneuvers, provided we have post-maneuver observations. But it's important to detect the maneuver quickly because the satellite can become "lost" in a short time. Once we determine the orbital elements, we can use propagation techniques to create an ephemeris for a satellite. To complete the cycle, we must create future look angles for a sensor to reacquire the satellite. We'll examine this technique first in the mission applications.

Next, we'll continue with a very important tool involving the prediction of close approaches between two satellites. As the number of satellites increases and their associated hardware (debris) accumulates in space, very precise methods of determining close approaches are necessary to safeguard payloads and astronauts on every mission. The problem is generally known as ***close-approach determination***. Although some basic theories have existed for about 20 years, increased computational power and new techniques are changing the way we address this problem.

Finally, we want to know the precise times when a satellite will rise and set from a particular sensor's field of view, so we can take quick and accurate observations. This problem is the ***rise/set*** problem.

10.1.1 Satellite Populations

It's useful to understand the current satellite populations with regard to active payloads, debris, and relative distributions depending on certain orbit parameters: inclination, eccentricity, and mean motion. We need to know these satellite distributions when developing certain programs. For instance, if most satellites are in low-Earth orbits, developing a collision-avoidance scheme for a geosynchronous satellite won't make much sense.

Figure 10-2 shows the distribution of existing satellite orbits by inclination. Notice the large concentration of orbital inclinations in the 60° to 130° range because this region allows satellites to observe and communicate with large sections of the Earth. The band occurring near 100° is mainly for Sun-synchronous orbits used mostly by Earth resource satellites. A smaller population shows up with geosynchronous satellites near $i = 0^\circ$.

The eccentricity distribution is less pronounced because most satellites are in nearly circular orbits. Although some eccentric orbits are more difficult to achieve, maintain, and track, the distribution is starting to change. Many satellites use circular orbits, but Figure 10-3 shows that more than 13% of satellites are in moderately to highly eccentric orbits (e greater than 0.1).

We can obtain information about the period and altitude by examining the values and distribution of satellite mean motions. Figure 10-4 shows this distribution. Notice that the plot shows equivalent values for mean motion, semimajor axis, and period.

Two comments are necessary. First, recognize that the distributions change (sometimes dramatically) over time. Unfortunately, some operational computer programs were developed assuming a single snapshot of the satellite population, with the distribution remaining the same. Secondly, we must develop a system architecture to be as general as possible. Some programs may be streamlined by eliminating certain constraints, so developers have to flag any assumptions in an attempt to keep the program or system from being misused in the future. This dilemma isn't easy to avoid, and often hasn't been!

10.2 Mission Orbits

Besides understanding the distribution of satellites by orbital elements, it's also important to understand the types of satellite missions we design. I've listed several missions in

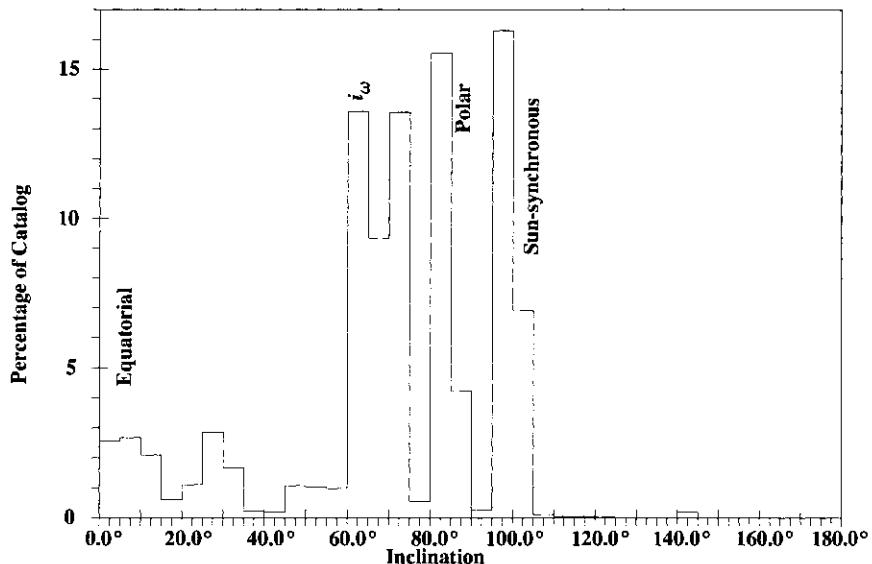


Figure 10-2. Inclination of Earth Satellite Orbits. This figure shows how satellite orbits are distributed based on inclination. Notice the large concentration of satellites near the critical inclination, i_{ω} , as well as the polar and Sun-synchronous regions.

Table 10-1 and representative orbits that satisfy each one. I'll discuss many of the specialized orbits required for these satellites in later sections.

The particular missions may seem simple, but I'll briefly review them. **Communication** satellites are an integral part of our society. Few imagined the explosion of television and telephone communications after launch of ECHO 1, the first communications satellite, on August 12, 1960. We also depend heavily on **weather** satellites for information on approaching storms and temperatures over the next few days. TIROS 1 was the first meteorological satellite—launching on April 1, 1960. **Navigation** satellites are helping redefine how we determine any location on the Earth. Industries like airlines and automobiles now use navigational information from satellites to display precise positions. Caprara (1986, 5) lists the functions of **military** satellites as communication, early warning, nuclear detection, ocean surveillance, reconnaissance, weather, electronic intelligence, research, and others. **Scientific** satellites perform a broad range of activities to study the Earth and our environment.

Each of these missions uses different orbits, as shown in Table 10-1. The specialized orbits contain **low-Earth** (LEO) to **geosynchronous** (GEO) orbits. I introduced these in Chap. 2. The types of low-Earth satellites are varied, whereas geosynchronous orbits are nearly circular and nearly equatorial, with periods of one solar day. **Semi-synchronous** orbits have orbital periods of 12 hours and are especially useful in navigation satellites.

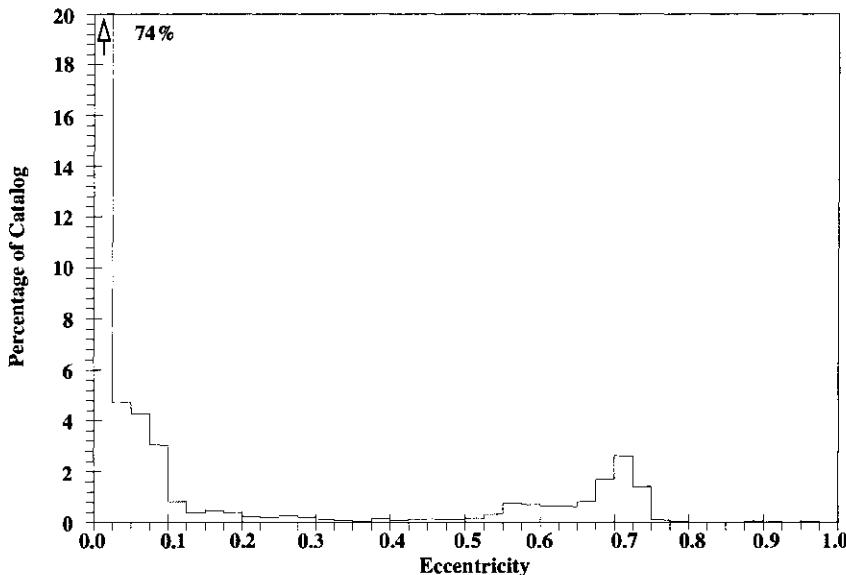


Figure 10-3. Eccentricity of Earth Satellite Orbits. This figure shows the satellite population distributed by eccentricity. Although many satellites are still in roughly circular orbits, about 13% of the catalog has $e > 0.1$. The first band contains 74% of the satellites, but I've only shown 20% for clarity in the remaining bands.

The broad category of *specialized orbits* represents several types of orbits designed to meet certain mission constraints. *Sun-synchronous* orbits keep the line of nodes of the satellite's orbit fixed relative to the Sun. This is useful for satellites needing constant illumination or certain lighting conditions on the solar panels. *Repeat-groundtrack* orbits provide periodic coverage over a specified ground location. Another example is an orbit that minimizes altitude variations by fixing the eccentricity and the argument of perigee. This type of orbit is useful for sensitive instruments with limited operational ranges.

Frozen orbits are specialized orbits that try to fix one or more orbital elements in the presence of perturbations. This concept is very useful in designing particular mission orbits. The Sun-synchronous and minimum altitude variation orbits are examples of frozen orbits. We design frozen orbits by using the dominant disturbing forces to achieve the desired effect. This isn't an exact process because other disturbing forces tend to disrupt the frozen orbit's geometry. However, because these other forces aren't the major disturbing force, the disruptions aren't rapid or large enough to prevent small periodic corrections from maintaining the desired frozen orbit. Reducing the size of stationkeeping maneuvers has important implications for minimizing satellite fuel requirements.

Finally, I'll identify an "other" category to capture various orbital designs not already mentioned. An example of these orbits is the former Soviet Union's development of a spe-

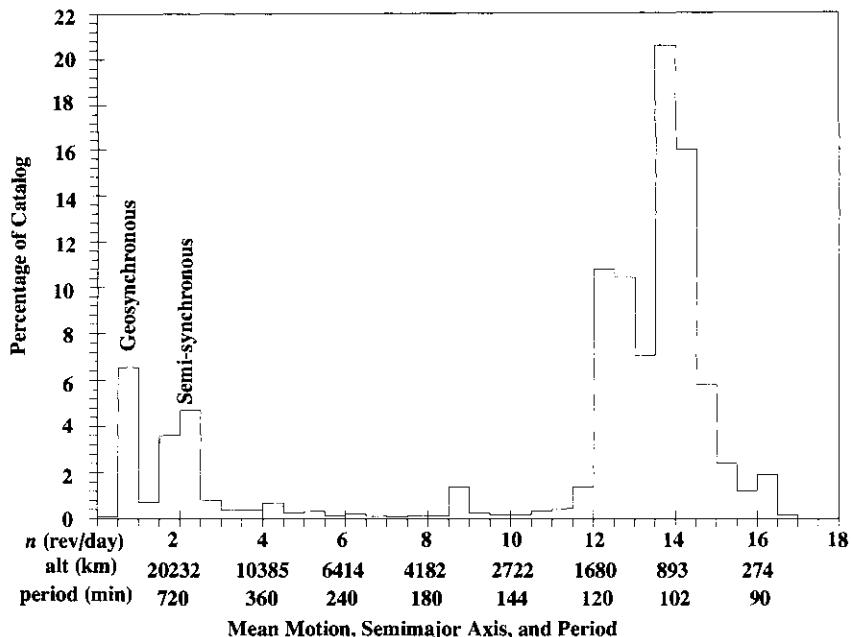


Figure 10-4. Mean Motion of Earth-Orbiting Satellites. This figure shows the space population distributed by mean motion. We determine two important quantities from mean motion: the orbital period and semimajor axis (altitude).

cialized orbit to suit their communication needs in the early 1960s. The *Molniya* orbit has a 12-hour period and an eccentricity of about 0.7. With the critical inclination (63.4°), the argument of perigee remains nearly constant. This allows apogee to be “fixed” over the high-latitude location of the former Soviet Union. This long duration is ideal for communications.

Finally, it’s important to recognize the present and future distribution of satellite orbits when planning operational software. As shown in Table 10-2, objects in space are either payloads (active and inactive) or debris. **Debris** includes rocket bodies, platforms, and the cables and devices used to secure and release satellites for deployment. The disparity between the orbiting and reentered objects occurs mainly because of satellite-deployment strategies and the desired target orbit. Typically, a deep-space launch leaves two to three times more objects in space than does its near-Earth counterpart. Communication satellites make up a large part of the active catalog.

10.2.1 Stationkeeping

Any discussion including perturbations must also consider stationkeeping maneuvers in a few main categories: space stations, low-Earth and mid-Earth-orbit satellites, and GEO

TABLE 10-1. Satellite Mission Orbits. This table shows several satellite missions and some specialized orbits they use. I've listed some common orbits and representative satellites from Caprara (1986). I've separated the orbits into four broad classes, which I call *specialized* orbits because certain orbital elements are designed to change very little. (SS = Sun synchronous, RG = repeat-groundtrack, MA = minimum-altitude variation, O = other.)

Mission/ “Frozen” Parameter	Semi- Synch			SS <i>i</i>	RG* β_Ω	MA $e + \omega$	O <i>varies</i>	Example Satellites
	LEO	Synch	GEO					
Communication	X	X	X				X	Intelsat, Satcom, Comstar, Molniya
Weather	X		X	X				GOES, NOAA, TIROS
Navigation		X						GPS, GLONASS
Military	X	X	X	X	X	X		DSCS, DMSP, TDRS, Milstar, DSP
Earth and ocean resources	X			X	X	X		Landsat, SPOT, RADARSAT, TOPEX
Scientific	X		X	X		X		Explorer, Hubble, GEOS, Lageos

*Periodic maneuvers are required to maintain the groundtrack.

TABLE 10-2. Orbiting and Decayed Satellites. The table shows the five most active countries in space and their contribution to its population (Janes, 1994, 149). Although accounting for drag is a challenge for the astrodynamist, drag reduces the satellite population through decay and eventual atmospheric reentry.

Country	Orbiting			Reentered		
	Payloads	Debris	Total	Payloads	Debris	Total
Former Soviet Union	1275	2395	3670	1606	9557	11163
United States	622	2691	3313	644	2860	3504
European Space Agency	24	137	161	4	450	454
Japan	49	52	101	9	72	81
People's Republic of China	10	79	89	23	78	101

satellites. Each of the mission orbits discussed will evolve over time due to perturbations. We'll see later that we can use the dominant perturbations to help plan stable mission orbits. Other orbits require periodic maintenance maneuvers. I won't discuss each category at length, but I'll introduce the relevant concepts and cite other references for more information.

Stationkeeping requirements for a space station are important because the technology to place a large station in a deep-space Earth orbit isn't, and won't soon be, mature. Space stations such as Skylab (c. 1978) and Mir (1986–date) are affected by many perturbations, especially atmospheric drag. The solution is to do nearly continual stationkeeping maneuvers to maintain the proper orbit. Doing so keeps each adjustment relatively small. These operations also require supply ships that periodically arrive with more fuel. Without resupply, the stations wouldn't last very long.

Stationkeeping for low- and mid-Earth-orbit satellites has multiple options. First, resupply often isn't practical, so considerable effort goes into designing orbits which require little adjustment due to the precise balancing of perturbations. The process is loosely termed "frozen orbits," which we'll examine later. The other main option is to do periodic maneuvers to maintain the correct orbit. This requires proper mission analysis *before* launch and must include accurate long-term analysis of perturbations. This is a prime example of the need for accurate semianalytical routines to analyze long-term motion. Based on the results of such studies, the mission may be designed with the proper fuel budget to do its task and keep the proper orbit for the designed lifetime.

Stable equilibrium locations for geosynchronous satellites are important because many satellites are in these orbits. The overall process is often called *east-west stationkeeping* because we try to maintain the longitude for the satellite. This is very important for communication satellites. We want to design the orbit (if possible) to keep stationkeeping velocity as low as possible. The qualitative analysis on page 597 (resonance) has shown that under the influence of the C_{22} harmonic, four positions of equilibrium exist in an Earth-fixed frame that is symmetrically situated on the extensions of the principal axes of the equatorial ellipse in Fig. 8-8 or Fig. 8-9. The unstable equilibrium longitudes are -14.7° and 165.3° , and the stable values are about 75.3° , and 255° .

Finally, as we saw in Chap. 8, the effects of third-body perturbations cause a long-periodic motion in the inclination for geosynchronous satellites. The process is usually called *north-south stationkeeping* because the perturbation results in changes that are normal to the equator. For north-south stationkeeping, Chobotov (1991, 326) indicates that the period is about 27 years to complete a half period in the inclination of about $\pm 15^\circ$. The satellite loses about 1° of inclination per year for ten years, and then it slows to reach 15° after about 17 years. He also suggests that we can maintain the inclination to about 0.1° with about four or five maneuvers each year. Each maneuver is about 50 m/s. Curtis (1994, 21) indicates that COMSAT was the first satellite to use this north-south motion for maneuver planning. By changing the satellite's attitude to compensate for the inclination change, the yearly fuel requirements went from 16.8 kg to 1.4 kg!

An important question arises: what happens when satellites become inoperable at geosynchronous altitude? Communication satellites are populating the geosynchronous belt with increasing frequency, causing tighter restrictions on location. Obsolete satellites are often left to drift, subject only to natural forces. Unfortunately, some of the satellites that drifted away several years ago are now coming back towards $i = 0^\circ$, increasing the possibility of colliding with replacement satellites. This process points to the importance of accurately analyzing perturbation effects.

10.3 Geometries for Surveillance and Reconnaissance

I introduced surveillance and reconnaissance in Chap. 9 when we examined ground-based measurements of satellites. We can also view these topics when the observing takes place *from space*. The distinction between surveillance and reconnaissance is subtle but important. **Surveillance** is observing events from space; **reconnaissance** is observing a particular area based on a specific tasking. Traditionally, military operations have covered both, but with the increased use of space, civilian applications are rapidly emerging. Applications include mapping, Earth-resources technology, communications, and environmental monitoring.

Reducing and analyzing sensor data from satellites requires us to use spherical geometry and understand basic sensor geometries. Some general formulas apply to various astrodynamics problems. Consider the Earth shown in Fig. 10-5 with two locations identified. The relationship between the two is best explained using launch and target concepts because many of the relevant equations were originally derived for ballistic missiles. As described earlier, the azimuth, β , denotes the direction from one location to another. There are two azimuths, one for the *short way* ($0^\circ < \beta < 180^\circ$), and one for the *long way* ($180^\circ < \beta < 360^\circ$). We usually want to find the ground range, Λ , between the two locations.

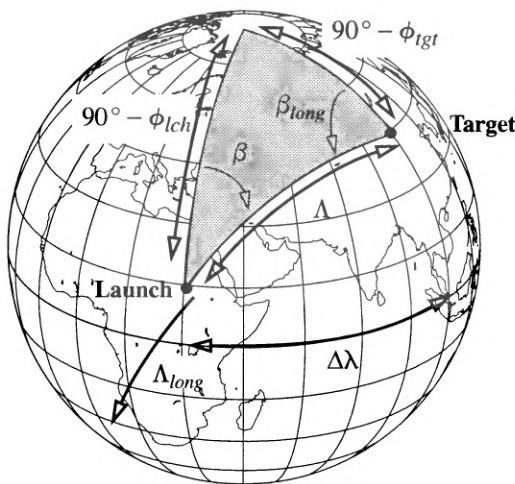


Figure 10-5. Earth Geometry. We can develop several important formulas from the spherical triangles formed by two locations on the Earth. Of primary interest are the azimuth, β , and the ground range to the target, Λ . These values are unique for both the short and long way to the target.

Before proceeding, note that ground range may be in radians, degrees, or kilometers. We often use approximations for this process which introduce some errors in the result. The formula developed in Eq. (C-22) and written using the ground range angle,

$$\Lambda_{\text{km}} = R_{\oplus} \Lambda_{\text{radians}} \quad (10-1)$$

applies to this problem. Remember, this relation is for a spherical Earth. Although I've shown kilometers in Eq. (10-1), the units for the Earth's radius determine the final answer as long as we use radians on the right-hand side of the equation. We developed a ground-range formula in Eq. (7-12):

$$\cos(\Lambda) = \sin(\phi_{tgt}) \sin(\phi_{lch}) + \cos(\phi_{tgt}) \cos(\phi_{lch}) \cos(\Delta\lambda) \quad (10-2)$$

This formula always gives the answer for the *short way* to the target or an angle that is less than 180° . From the law of sines for spherical triangles,

$$\frac{\sin(\beta)}{\sin(90 - \phi_{tgt})} = \frac{\sin(\Delta\lambda)}{\sin(\Lambda)}$$

Solving results in an equation we can use with ATAN2 to resolve the quadrant.

$$\sin(\Lambda) = \frac{\sin(\Delta\lambda) \cos(\phi_{tgt})}{\sin(\beta)} \quad (10-3)$$

The direction (azimuth) of the ground trace uses the cosine law,

$$\cos(90^\circ - \phi_{tgt}) = \cos(90^\circ - \phi_{lch}) \cos(\Lambda) + \sin(90^\circ - \phi_{lch}) \sin(\Lambda) \cos(\beta)$$

which we can rearrange to get

$$\cos(\beta) = \frac{\sin(\phi_{tgt}) - \sin(\phi_{lch}) \cos(\Lambda)}{\cos(\phi_{lch}) \sin(\Lambda)} \quad (10-4)$$

Then, let's find a sine expression using the sine law, as in our development of Eq. (10-3):

$$\sin(\beta) = \frac{\sin(\Delta\lambda) \cos(\phi_{tgt})}{\sin(\Lambda)} \quad (10-5)$$

If we know the azimuth, β , and ground range, Λ , our problem is to determine the latitude and longitude of the target. Let's begin by solving the cosine equation for the target latitude:

$$\sin(\phi_{tgt}) = \cos(\beta) \cos(\phi_{lch}) \sin(\Lambda) + \sin(\phi_{lch}) \cos(\Lambda) \quad (10-6)$$

The longitude relies on solving Eq. (10-3) for the change in longitude. We can resolve angles without quadrant checks if we find the cosine expression by solving Eq. (10-2):

$$\begin{aligned} \sin(\Delta\lambda) &= \frac{\sin(\beta) \sin(\Lambda)}{\cos(\phi_{tgt})} \\ \cos(\Delta\lambda) &= \frac{\cos(\Lambda) - \sin(\phi_{lch}) \sin(\phi_{tgt})}{\cos(\phi_{lch}) \cos(\phi_{tgt})} \end{aligned} \quad (10-7)$$

Otherwise, we must specify the direction (long or short) to determine longitude. Then, we add the change in longitude to the launcher's initial longitude to find the target longitude:

$$\lambda_{tgt} = \lambda_{lch} + \Delta\lambda$$

The next area to explore is the sensor's geometry and field of view (Fig. 10-6). I define **field of view** ($\tilde{\Lambda}_{FOV}$) as the angle of the total area that a sensor can observe. The overbar distinguishes the fact that we're using the diameter of the total area. Some sensors are called **staring sensors** because they constantly observe one location beneath the satellite. In this case, $\tilde{\Lambda}_{FOV}$ is the total area that the satellite sees at any time. **Scanning sensors** observe many different locations depending on the sensor's orientation to the satellite. These sensors are often used with drive mechanisms that oscillate with respect to the satellite's velocity vector. The total area for this type of system can be much larger than any instantaneous observable region.

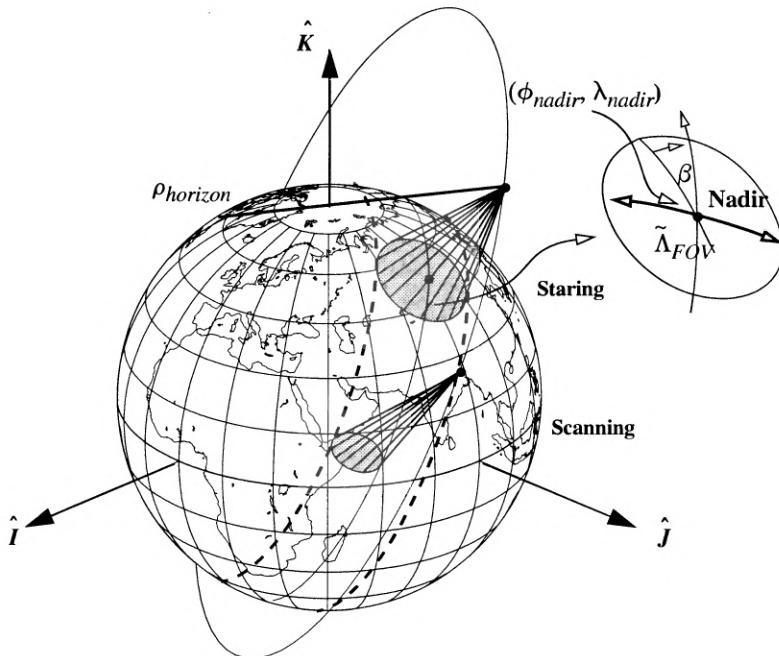


Figure 10-6. Geometry for Field of View (FOV). The FOV problem consists of achieving mission objectives within a satellite sensor's limit. An arbitrary 1500 km radius circle about the staring-satellite's subpoint depicts this concept. Over an interval of time, the scanning sensor can view the same area as the staring sensor, but it can't continually observe the entire region of the staring sensor.

The general problem associated with calculating field of view is determining how much of the Earth we can see, given certain parameters about the viewing satellite. Several specific cases arise, and we'll explore how to describe the geometries. Consider the basic geometry shown in Fig. 10-7 and remember we're still using a spherical Earth.

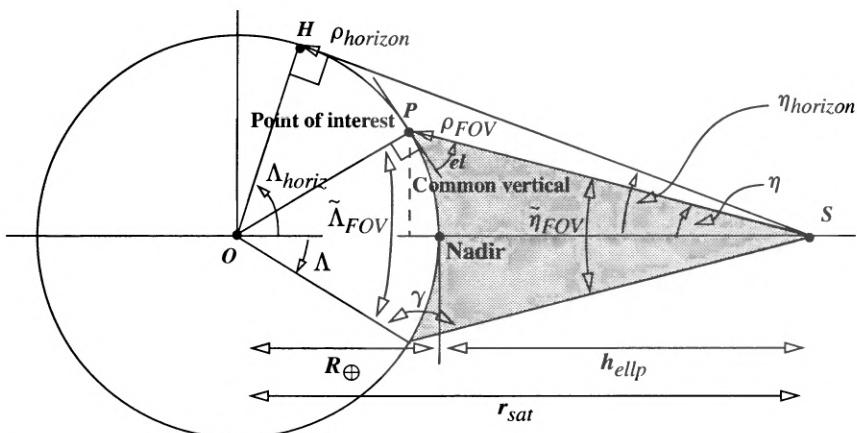


Figure 10-7. Basic Geometry for Field of View—Nadir Pointing. We require several parameters to determine the location of the Earth's horizon and the ground-range angle for a point of interest, P . From the nadir point defined directly beneath a satellite, S , at an altitude of h_{ellp} , measure the boresight angle, η , outward from the radial direction. For general calculations, an intermediate angle, γ , and the common vertical line are required to determine the ground range, Λ . I use the tilde overbar to signify the total field-of-view angles visible—for both staring and scanning sensors.

The point directly below the satellite is called the **nadir point**, and many subsequent calculations for angles reference this point. A **boresight angle**, η , locates the angular displacement from the nadir direction to an object of interest. A **phase angle**, β_{ph} , is sometimes used to specify where the sensor is looking. It has the same symbol as azimuth because it's measured from the instantaneous position of due north from the satellite.

When forming equations for calculations, it's extremely important to consider all quadrants. A spaceborne sensor doesn't have some of the limitations of sensors for Earth-based systems. In particular, elevation values are often negative.

Typically, the satellite's position vector is known, permitting us to determine the latitude and longitude of the nadir point ($\phi_{nadir}, \lambda_{nadir}$) using Algorithm 23 in Chap. 3. Next, it's often important to determine the various boresight angles: specifically, the horizon from the satellite, the maximum displacement the sensor can view, or the location of another satellite. We find the horizon-boresight angle, $\eta_{horizon}$, using Fig. 10-7. From the right triangle shown, we can directly determine

$$\sin(\eta_{horizon}) = \frac{R_{\oplus}}{r_{sat}}$$
 (10-8)

The sine expression is sufficient because the angle will never exceed $\pm 90^\circ$. As an aside, measurements made near the limit of a sensor's field of view ($\eta \approx \tilde{\eta}_{FOV}/2$) aren't usually as accurate as those near the center. This is usually because of mechanical and electrical design limitations.

We can also develop the ground-range angle on the Earth that's visible to the sensor using planar trigonometry. As a check, this quantity should never be more than one-fourth of the Earth's circumference.

$$\cos(\Lambda_{horiz}) = \frac{R_{\oplus}}{R_{\oplus} + h_{ellp}}$$

For the slant range to the horizon, we also use plane trigonometry:

$$\rho_{horizon} = r_{sat} \cos(\eta_{horizon})$$
 (10-9)

To find another expression, let's use the Pythagorean theorem on the right triangle (*SHO*) in Fig. 10-7 and simplify:

$$\rho_{horizon} = \sqrt{r_{sat}^2 - R_{\oplus}^2} = \sqrt{2R_{\oplus}h_{ellp} + h_{ellp}^2}$$

Get a general expression for the slant range by examining the oblique triangle (*SPO*) in Fig. 10-7. Now, find the intermediate angle, γ , using the sine law for oblique triangles:

$$\sin(\gamma) = \frac{r_{sat} \sin(\eta)}{R_{\oplus}}$$
 (10-10)

Remember to consider the quadrants because the angle (γ) will always be larger than 90° .

Find the slant range to any point on the Earth by using an oblique triangle formula for sides from Eq. (C-21).

$$\rho = R_{\oplus} \cos(\gamma) + r_{sat} \cos(\eta)$$
 (10-11)

As a check, the slant-range value is always between h_{ellp} and $\rho_{horizon}$, unless we're observing another satellite.

The opposite problem of a site observing a satellite with a minimum or maximum elevation constraint is found using the law of sines for oblique triangles.

$$\frac{\sin(90^\circ + el)}{R_{\oplus} + h_{ellp}} = \frac{\sin(90^\circ - (\Lambda + el))}{R_{\oplus}}$$

Thus,

$$\cos(\Lambda) = \frac{R_{\oplus} \cos(el)}{R_{\oplus} + h_{ellp}} - \cos(el)$$

Solve for the *general boresight angle*, η , by rearranging Eq. (10-11) and substituting the satellite's radius and the range from a target location. See Fig. 10-8. These operations yield

$$\cos(\eta) = \frac{\rho^2 + r_{sat}^2 - R_\oplus^2}{2\rho r_{sat}} \quad (10-12)$$

Notice that you must know the slant range, ρ , to the point of interest. If the point of interest is *not* constrained to the Earth's surface, you must change R_\oplus to be the distance from the center of the Earth to the point of interest, $R_\oplus + h_{ellip}$.

Now let's examine a method to determine the latitude and longitude values of the specific ground locations that are visible to the satellite. The only point you know immediately is the nadir point directly beneath the satellite. To obtain the latitude and longitude of other points, you must determine a ground distance and direction from the nadir. Use a sensor-based frame (SEZ at the satellite) to express these quantities and convert the cartesian coordinates to an angular separation, η , and azimuth, β . Next, convert η to Λ using Eq. (10-13). Then use β to find the particular latitude and longitude from Eq. (10-6) and Eq. (10-7), using the spherical triangle in Fig. 10-5. If you examine all possible azimuths (0° to 360°) from the nadir point, you can find the locus of points the satellite can see.

Two cases exist—staring directly at the nadir point and scanning off to a side. The solution for both cases is to find the ground range, Λ_{FOV} , that corresponds to a given boresight angle, $\tilde{\eta}_{FOV}$. Let's begin by finding the intermediate angle from Eq. (10-10) for the given boresight angle. Next, find the slant range to the satellite using Eq. (10-11). The common vertical point of the field-of-view angles in Fig. 10-7 permits you to use the sine law and find the ground range from the nadir point:

$$\sin(\Lambda) = \frac{\rho \sin(\eta)}{R_\oplus} \quad (10-13)$$

The more general case occurs for a scanning sensor, as shown in Fig. 10-8. The same concepts apply, but finding the sensor's ground limits is more complicated because the geometry is no longer symmetrical about the satellite's radius vector or nadir, requiring more processing to find the ground-range angle. By carefully examining Fig. 10-8, we can do two calculations using the center boresight angle, η_{center} , to determine maximum and minimum boresight angles, and ultimately, maximum and minimum ground-range angles. The difference in ground-range angles ($\Lambda_{max} - \Lambda_{min}$) is the value we seek.

Let's look at an example to show how we can use these relations.

▼ Example 10-1: Calculating Field of View.

GIVEN: Satellite in a circular orbit ($i = 50^\circ$) at an altitude of 800 km. $\tilde{\eta}_{FOV} = 25.0^\circ$. The current nadir location is 60°N and 40°E .

FIND: Minimum and maximum values of latitude and longitude for both $\eta_{center} = 0^\circ$ and 40° .

At 800 km altitude, the position vector is

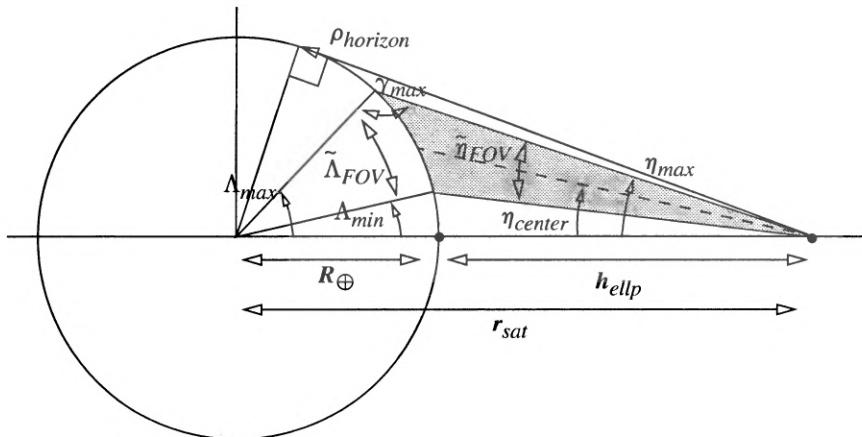


Figure 10-8. Geometry for Field of View—Oblique Pointing. This figure shows a scanning sensor, as well as another angle, η_{center} , we need to calculate ground ranges for oblique pointing. When η_{center} is zero, the diagram reverts to the geometry of Fig. 10-7.

$$r_{\text{sat}} = 6378.1363 + 800 = 7178.1363 \text{ km} = 1.125\,428 \text{ ER}$$

First, determine the horizon-boresight angle:

$$\sin(\eta_{\text{horizon}}) = \frac{R_{\oplus}}{r_{\text{sat}}} = \frac{1}{1.125\,428}, \eta_{\text{horizon}} = 62.6917^\circ$$

We can also determine the ground range that is visible to the sensor by

$$\cos(\Lambda_{\text{horiz}}) = \frac{R_{\oplus}}{R_{\oplus} + h_{\text{ellp}}} = \frac{1}{1.125\,428} \Rightarrow \Lambda_{\text{horiz}} = 27.3082^\circ$$

Therefore, the sensor can see $(27.3082)(\pi/180)(R_{\oplus}) = 3039 \text{ km}$.

It's useful to find the maximum slant range so you can check later calculations.

$$\rho_{\text{horizon}} = r_{\text{sat}} \cos(\eta_{\text{horizon}}) = 1.125\,428 \cos(62.6917) = 0.516\,323 \text{ ER} = 3293.178 \text{ km}$$

Find the ground-range angle by first finding the intermediate angle, γ . Because you're looking directly down, use half the complete boresight angle:

$$\sin(\gamma) = \frac{r_{\text{sat}} \sin(\eta)}{R_{\oplus}} = \frac{1.125\,428 \sin(12.5)}{1}, \gamma = 14.098\,36^\circ, 165.901\,64^\circ$$

You must use the larger angle because the triangle is oblique. To find the slant range, use half the FOV angle because you're looking straight down. Be sure it's just slightly larger than the orbital altitude of 800 km.

$$\rho = R_{\oplus} \cos(\gamma) + r_{\text{sat}} \cos(\eta) = (1) \cos(165.901\,64) + 1.125\,428 \cos(12.5)$$

$$\rho = 0.128\,87 \text{ ER} = 821.97 \text{ km}$$

Now find the ground range from the satellite subpoint:

$$\sin(\Lambda) = \frac{\rho \sin(\eta)}{R_{\oplus}} = \frac{0.12887 \sin(12.5)}{1}, \Lambda = 1.5984^\circ = 177.92193 \text{ km}$$

You'll always use the smaller value because a satellite can't observe more than a 90° arc of the Earth's surface. Simply add the ground-range angle to the satellite nadir's location to find the locations of minimum and maximum latitudes:

$$\phi_{tgtmax} = 61.5984^\circ, \phi_{tgtmin} = 58.4016^\circ, \lambda = 40^\circ\text{W}$$

If you needed the east-west variations, you could add the ground-range angle to the longitude values. Now, if $\eta_{center} = 40^\circ$, the analysis changes. First, find the maximum and minimum ground-range angles from the nadir point. Several intermediate parameters are needed. It's important to determine the maximum and minimum boresight angles:

$$\eta_{max} = \eta_{center} + \frac{\tilde{\eta}_{FOV}}{2} = 52.5^\circ, \text{ and } \eta_{min} = \eta_{center} - \frac{\tilde{\eta}_{FOV}}{2} = 27.5^\circ$$

Use the maximum and minimum values ($\eta = \eta_{max}, \eta_{min}$) and calculate the following parameters. Be careful to check quadrants.

$$\sin(\gamma) = \frac{r_{sat} \sin(\eta)}{R_{\oplus}}, \gamma_{max} = 116.76482^\circ, \text{ and } \gamma_{min} = 148.69021^\circ$$

$$\rho = R_{\oplus} \cos(\gamma) + r_{sat} \cos(\eta), \rho_{max} = 0.234788 \text{ ER, and } \rho_{min} = 0.143897 \text{ ER}$$

$$\sin(\Lambda) = \frac{\rho \sin(\eta)}{R_{\oplus}}, \Lambda_{max} = 10.73517^\circ, \text{ and } \Lambda_{min} = 3.80977^\circ$$

The total ground range is the difference between the maximum and minimum ranges:

$$\Lambda_{max} - \Lambda_{min} = 6.92538^\circ = 770.93067 \text{ km}$$

Notice that the ground range observed is larger than the 340 km (twice 177 km) for the observation beneath the satellite. Determine the ground locations the same as before, but because the center of the field-of-view circle is no longer at the nadir point, you must first find the center of the sensor's field-of-view. This involves knowing the direction the sensor is looking. If you assume the satellite is looking perpendicular to the velocity vector, you can locate the center as follows. Because the inclination is 50° , the azimuth for a side-looking sensor is 140° , or 320° . Let's use 140° . Now use Eq. (10-6) to find the center that the sensor is observing.

$$\sin(\phi_{tgt}) = \cos(\beta) \cos(\phi_{lch}) \sin(\Lambda) + \sin(\phi_{lch}) \cos(\Lambda)$$

$$\sin(\phi_{tgt}) = \cos(140^\circ) \cos(60^\circ) \sin(6.92538^\circ) + \sin(60^\circ) \cos(6.92538^\circ)$$

$$\phi_{tgt} = 54.1519^\circ$$

From Eq. (10-7),

$$\sin(\Delta\lambda) = \frac{\sin(\beta) \sin(\Lambda)}{\cos(\phi_{lch})} = \frac{\sin(140^\circ) \sin(6.92538^\circ)}{\cos(60^\circ)} \quad \Delta\lambda = 7.9865^\circ$$

$$\lambda_{tgt} = 47.9865^\circ$$

You can now find the new minimum and maximum latitudes by simply subtracting and adding $\Delta\lambda$.

$$\phi_{tgtmax} = 57.6146^\circ, \text{ and } \phi_{tgtmin} = 50.6892^\circ$$

You can find the locus of points at the edge of the sensor's field of view by repeated use of Eqs. (10-6) and (10-7), letting the azimuth vary from 0° to 360° . For example, at $\beta = 20^\circ$,



$$\phi_{tgt} = 57.3871^\circ, \text{ and } \lambda_{tgt} = 50.1831^\circ$$

10.4 Designing and Maintaining Mission Orbits

Operational mission analysis requires that in addition to two-body motion, we account for the effect of perturbations on the evolution of the orbit over time. To do such mission analysis, we must first be able to determine the size and source of perturbations on the orbit. Obviously, the magnitudes of the perturbations will vary with the orbit, but certain ones usually require attention. From Chap. 8 we know some significant secular perturbations result for Ω and ω , due to the even zonal harmonics (ignoring the periodic perturbations), and for a and e , due to drag. We'll consider these secular perturbations as a first cut at planning. Recognize that actual mission planning incorporates resonances, third-body effects, complex drag models, and so on. Mission analysis that includes these factors is ideally suited for semianalytical techniques (Sec. 8.8.1) because it often requires accurate predictions for the distant future.

As discussed in Table 10-1, I've grouped a broad set of satellite orbits into a category called *specialized orbits*. I discuss three major orbital types within this section—Sun-synchronous, repeat groundtrack, and minimum altitude variations. The term frozen orbits is sometimes applied to these classes; however, I feel there is adequate rationale to separate these three categories. I discuss the general concept of a frozen eccentricity and argument of perigee as a separate option for orbital design. An example at the end of this section demonstrates how we can combine aspects of orbital design to create a unique satellite orbit. For further information, Boelitz (1991) discusses aspects of maneuver planning for specialized orbits.

10.4.1 Sun-Synchronous Orbit

The first specialized orbit we'll consider is the ***Sun-synchronous*** orbit, which maintains a constant orientation towards the Sun during a year. We use this orbit for various reasons—from wanting a constant source of solar radiation for the solar cells to maintaining a certain viewing condition when observing the Earth. Consider the geometry in Fig. 10-9.

For low-Earth satellites, the dominant secular motion of Ω is caused by J_2 . Thus if we can design the orbit so the nodal rate matches the average rate of the Sun's motion projected on the equator, the satellite's line of nodes will maintain a constant angular separation with respect to the Sun. Moreover, we can design this separation to be within any desired value between 0° to 360° by selecting the initial value of Ω . Recall Eq. (8-37) which gave the secular perturbations for the longitude of ascending node as a result of J_2 perturbation. In one year (365.242 189 7 days, see page 5), the Earth completes 360° around the Sun.

$$\begin{aligned} \frac{d\Omega}{dt}_{desired} &= \frac{360^\circ}{365.242\ 189\ \bar{7}} = 0.985\ 647\ 36 \frac{\text{deg}}{\text{day}} = 0.017\ 202\ 79 \frac{\text{rad}}{\text{day}} \\ &= 1.606\ 378\ 006 \times 10^{-4} \frac{\text{rad}}{\text{TU}} \end{aligned}$$

Setting these two relations equal results in

$$\frac{1.606\ 378\ 006 \times 10^{-4}}{n} = 1.606\ 378\ 006 \times 10^{-4} \sqrt{\frac{a^3}{\mu}} = -\frac{3R_{\oplus}^2 J_2}{2a^2(1-e^2)^2} \cos(i)$$

$$a^{7/2} = -\frac{3R_{\oplus}^2 J_2 \sqrt{\mu}}{2(1.606\ 378\ 006 \times 10^{-4})(1-e^2)^2} \cos(i) \quad (10-14)$$

Be careful of the units in this relation! As shown, the units of a will be the same as the Earth's radius if μ is in canonical units. Notice the relation between the semimajor axis and the inclination. Solution of Eq. (10-14) shows the semimajor axis (altitude), eccentricity, and inclination for a Sun-synchronous orbit. See Fig. 10-10.

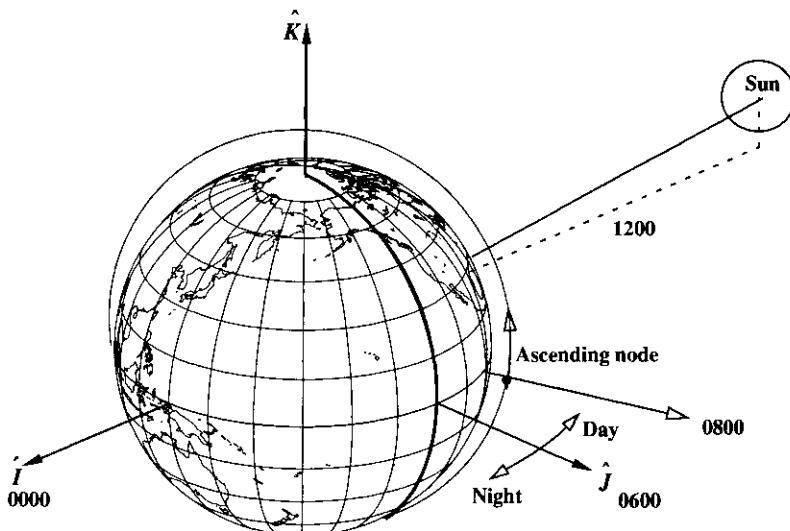


Figure 10-9. Geometry for Sun-Synchronous Orbits. We use a nomenclature based on the time of the ascending node to distinguish Sun-synchronous orbits. The orbit shown is an 0800 orbit because the satellite reaches the ascending node at 8:00 A.M. local time. I've shown the Sun offset from the equatorial plane for generality.

In practice, we solve Eq. (10-14) for the inclination because the semimajor axis and eccentricity are usually determined by other mission constraints. For example, repeat-groundtrack orbits define the semimajor axis, and limits on the sensor's range may impose a maximum limit on a . Low-Earth orbits usually imply small eccentricities because larger eccentricity values (that don't result in Earth impact) extend the orbit type to mid-Earth orbits. In addition, frozen orbits can fix the eccentricity to a certain value, thus minimizing altitude variations from perigee to apogee without the need for special stationkeeping.

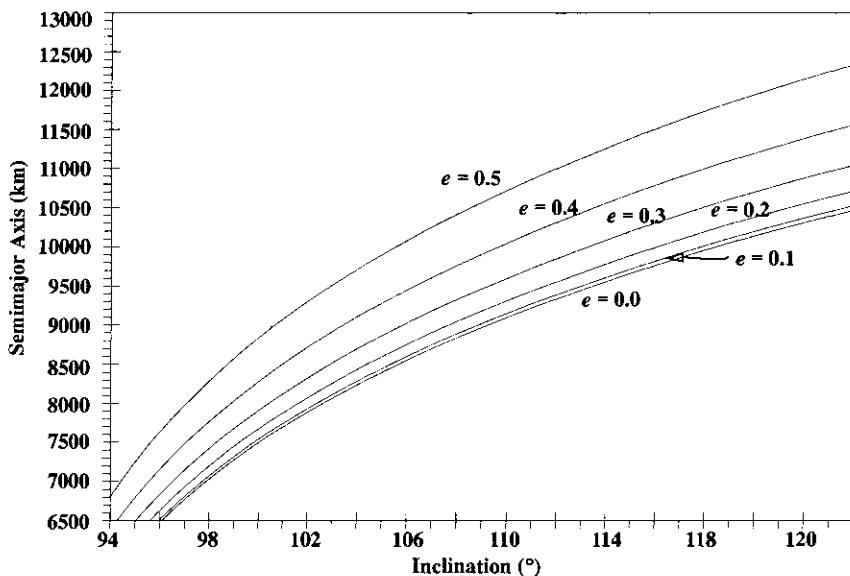


Figure 10-10. Elements for Sun-Synchronous Orbits. This figure shows that choosing two orbital elements from i , a , and e determines the third element needed for a Sun-synchronous orbit. The semimajor axis and radius of perigee must be greater than the Earth's radius to be realistic.

burns to maintain a near-zero eccentricity. Therefore, the problem becomes one of determining the inclination given other parameters.

Now consider an example calculation based on RADARSAT information from *Jane's* (1994, 385).

▼ Example 10-2. Calculating a Sun-synchronous Orbit.

GIVEN: The Canadian government launched a satellite on November 4, 1995 (RADARSAT) to monitor the polar icecap's motion. A Sun-synchronous orbit was selected to permit appropriate viewing of the polar regions.

FIND: If the altitude is 800 km, what inclination must the satellite occupy for a circular orbit? What is the altitude if the eccentricity is 0.2 at that same inclination?

First, convert the altitude to a semimajor axis, because you're looking at a circular orbit.

$$a = 800.0 + 6378.1363 \text{ km} = 1.125\ 32 \text{ ER}$$

Remember, the rate must be in rad/TU.

$$\cos(i) = -\frac{a^{7/2} 2 (1.606\ 378\ 006 \times 10^{-4}) (1-e^2)^2}{3R_{\oplus}^2 J_2 \sqrt{\mu}} = -\frac{(1.125\ 32)^{7/2} 2 (1.606\ 378\ 006 \times 10^{-4})}{3(1^2)(0.001\ 082\ 63)\ \sqrt{1}}$$

$$i = 98.6^\circ$$

If you keep $i = 98.6^\circ$ and change the eccentricity to 0.2,

$$a^{7/2} = \frac{-3(1^2)(0.001\ 082\ 63)\sqrt{1}}{2(1.606\ 378\ 006 \times 10^{-4})(1 - 0.2^2)^2} \cos(98.6^\circ)$$

$a = 1.151\ 879$ ER = 7346.846 km, or apogee and perigee altitudes of 674.8 km \times 1262.6 km

▲ Notice that the perigee altitude is still within the sensor limits, but apogee may place the sensor past its range of performance.

Effect of Nodal Drift

The secular rate of Ω is caused mainly by the J_2 perturbation and depends on the elements a , e , and i . It's particularly sensitive to changes in the inclination—more so than to changes in a and e . The effect is given by

$$\Delta\dot{\Omega}_{sec} = \frac{\partial\dot{\Omega}}{\partial t}\Delta i + O(\Delta i)^2$$

where using Eq. (8-37) yields

$$\begin{aligned}\Delta\dot{\Omega}_{sec} &= \frac{3}{2}n\left(\frac{R_\oplus}{p}\right)^2 J_2 \sin(i)\Delta i + O(\Delta i)^2 \\ &= -\dot{\Omega}_{sec} \tan(i)\Delta i + O(\Delta i)^2\end{aligned}\quad (10-15)$$

The primary drift in the inclination is caused not by J_2 but by the Sun because the Sun-synchronous constraint causes a resonance effect due to the critical argument $\Omega - \dot{\alpha}_\odot \approx 0$ (Sec. 8.5.2), or equivalently, $\dot{\Omega} - (\dot{\omega}_\odot + \dot{M}_\odot) \approx 0$. This causes a slowly growing oscillation with a large amplitude that particularly affects the satellite orbit's inclination. Because of this oscillation and the tight constraints most Sun-synchronous missions place on the Ω -Sun angle, periodic orbit-adjustment burns are required to correct and maintain this angle within a specified tolerance.

We'll consider two Ω maintenance strategies to maintain the Sun-synchronous orientation. The first is to do a noncoplanar burn to kick the node back to the desired value. Thus, if Ω_o is the desired value and Ω is the current value, the desired change is $\Delta\Omega = \Omega - \Omega_o$. We can show that a velocity impulse in the direction normal to the orbital plane, along the W -axis in the RSW system Δv_W , causes a change in the node of

$$\Delta\Omega = \frac{1}{na} \frac{\sin(u)}{\sin(i)} \Delta v_W + O(e^2)$$

This is a different form of Algorithm 37. Unfortunately, it has disadvantages. Even though we've reestablished Ω , the inclination is still perturbed by Δi from the desired Sun-synchronous value. Thus we also have to correct the inclination back to the nominal value to maintain the original orbital design. This requires separate burns for i and Ω to achieve maximum efficiency for each or a combined burn at reduced efficiency. Either way, we require more Δv and consequently more fuel.

A more desirable approach is to take advantage of the fact that Δi causes nodal drift. Specifically, we can correct the drift caused by the solar perturbation ($\Delta i_{correct} = -\Delta i_{drift}$)

and correct $\dot{\Omega}$ back to the nominal Sun-synchronous value. However, this leaves the nodal drift, $\Delta\Omega$. We can correct the nodal drift from the current value back to the nominal value through $(-\Delta\Omega)$ by noting that the inclination change, Δi , caused the node to drift through $\Delta\Omega$. Thus, if we correct the inclination by an additional $(-\Delta i)$, we should drift back to the nominal Ω . Therefore, our total correction to the inclination is $-2\Delta i$ where Δi is the current drift. This strategy allows us to adjust only one element (i) and thus is a more efficient burn strategy. McClain (1990) shows that a velocity impulse normal to the orbital plane (W -axis in the *RSW* system) causes an instantaneous change in the inclination given by

$$\Delta i = \frac{1}{na} \cos(u) \Delta v_W + O(e^2) \quad (10-16)$$

Inspection of Eq. (10-16) shows that the efficiency of the Δv is maximized by burning at the equator ($u = 0$). This is a general form of the inclination-only maneuvers in Algorithm 36. Solving for Δv_W to restore the nominal values for node and node rate gives us

$$\Delta v_W = -2\Delta i \frac{na}{\cos(u)}$$

where Δi is the observed drift in the orbital inclination.

In practice we can extend the time between these periodic adjustments to inclination by recognizing that we have an acceptable window ($\Omega_o \pm \Delta\Omega$) about the nominal Ω value. When the node drifts to one end of this window, we perform a burn to correct the node to the other end of the window. This effectively doubles the time between maneuvers to adjust the inclination.

Recall that Eq. (10-15) describes the delta in the nodal drift rate away from the nominal value required for the Sun-synchronous orbit. This difference is due to the displacement in the inclination from the nominal value. Both of these deltas ($\Delta\Omega$, Δi) depend on time. In particular, the inclination drift at time t is given by

$$\Delta i(t) = \Delta i_o + \frac{di}{dt}(t - t_o) \quad (10-17)$$

where Δi_o is the amount of drift at time t_o . The inclination rate di/dt is evaluated using the third-body disturbing function R_{\odot} for the Sun [Eq. (8-34)] restricted to the dominant contribution ($l = 2$) and to the term producing the resonance effect. We'll briefly discuss this technique later on.

We can consider di/dt to be constant over the time intervals associated with maintaining the Sun-synchronous orbit (several months to a few years). The nodal drift is obtained by integrating Eq. (10-15):

$$\Delta\Omega_{sec} = - \int_{t_o}^t \dot{\Omega}_{sec} \tan(i) \Delta i dt \quad (10-18)$$

Consistent with other approximations already made, we assume nominal (constant) values for $\dot{\Omega}_{sec}$ and $\text{TAN}(i)$. Substituting Eq. (10-17) into Eq. (10-18) and integrating yield

$$\Delta\Omega = A_1(t - t_o) + A_2 \frac{(t - t_o)^2}{2} \quad (10-19)$$

$$A_1 = -\dot{\Omega}_{sec} \text{TAN}(i) \Delta i_o$$

$$A_2 = -\dot{\Omega}_{sec} \text{TAN}(i) \frac{di}{dt}$$

Note that, for burn-planning, Δi_o is not the accumulated inclination drift at the time of excursion from the nodal window. Rather, it's the unknown change in inclination we must obtain from the inclination-adjustment maneuver to achieve our planned reverse drift to the opposite side of the Ω window. Eq. (10-19) and Fig. 10-11 also show that the drift is quadratic in the time.

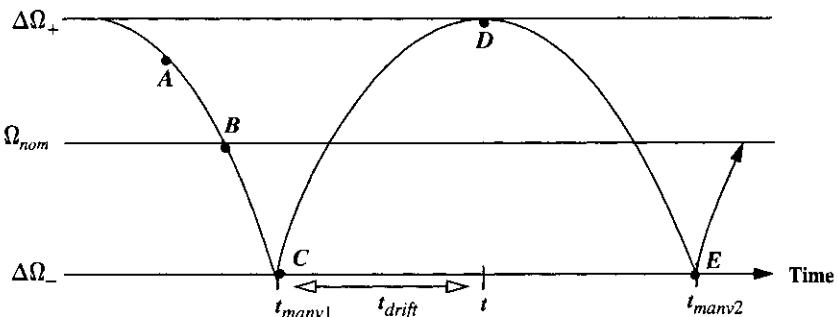


Figure 10-11. Profile of Nodal Drift. This figure shows the change in the nodal drift, $\Delta\Omega$, over time. There are usually limits placed about the nominal value (\pm) within which the mission is successful. The satellite mission begins at point A with orbit insertion. The first burn occurs at point C and we drift until the second maneuver at point E.

We want to determine the time of our maximum reverse nodal drift which we want to occur at the boundary $\Omega_o - \Delta\Omega$. This maximum reverse drift satisfies the condition

$$\frac{d\Delta\Omega}{dt} = 0$$

Taking the derivative of Eq. (10-19), setting it to zero, and solving it yields $t_{drift} = -A_1/A_2$. Note that A_1 and A_2 must differ in sign because $t_{drift} > 0$. We want to drift through the interval $-2\Delta\Omega$ over the interval $[t_o, t_{drift}]$, so

$$-2\Delta\Omega = A_1 t_{mav} + \frac{A_2}{2} t_{mav}^2 = A_1 \left(\frac{-A_1}{A_2} \right) + \frac{A_2}{2} \left(\frac{-A_1}{A_2} \right)^2 = -\frac{1}{2} \frac{A_1^2}{A_2}$$

Solving for A_1 yields

$$A_1 = \pm 2\sqrt{A_2 \Delta \Omega}$$

We determine the proper sign to guarantee that A_1 and A_2 differ in sign, which means A_1 must have a sign opposite to that of di/dt . The desired inclination adjustment, Δi_a , comes from the definition of A_1 :

$$\Delta i_a = -\frac{\dot{\Omega}_{sec} \tan(i)}{A_1}$$

The approximate time interval between inclination adjustments is $2t_{drift}$.

The inclination rate in Eq. (10-17) is given by the Lagrange VOP equations [Eq. (8-12)]:

$$\frac{di}{dt} = \frac{1}{na^2\sqrt{1-e^2}\sin(i)} \left\{ \cos(i) \frac{\partial R}{\partial \omega} - \frac{\partial R}{\partial \Omega} \right\}$$

Taking the dominant term in Kaula's third-body disturbing function [Eq. (8-34)] for the Sun and further restricting the indices $(l,m,p,h,q,q') = (2,2,1,0,0,0)$ to the resonant term of interest yield

$$R_\odot = \frac{1}{12} \left(\frac{\mu_\odot}{a_\odot} \right) \left(\frac{a}{a_\odot} \right)^2 F_{2,2,1}(i) F_{2,2,0}(i_\odot) H_{2,1,0}(e) G_{3,0,0}(e_\odot) \\ \times \cos[-2(\omega_\odot + M_\odot) + 2(\Omega - \Omega_\odot)]$$

Now $\dot{\omega}_\odot \approx \dot{\Omega}_\odot \approx 0$, and for the Sun-synchronous case, $\dot{\Omega} \approx \dot{M}_\odot$. So the rate of the argument becomes very small in magnitude, which indicates the resonance. Taking the partial derivatives of R_\odot with respect to ω and Ω and substituting the result into the VOP equation yield the desired rate. Clearly $\partial R / \partial \omega = 0$ because ω doesn't appear in the equation.

10.4.2 Repeat-Groundtrack Orbit

Repeat-groundtrack orbits are those which retrace their groundtrack over a certain time interval. They're important to missions that periodically revisit a particular point on the Earth, such as those studying gravity, the atmosphere, or the movement of the polar ice cap. This section examines how to design repeat-groundtrack orbits and presents ways to determine fuel requirements for preserving certain orbital characteristics in the face of real-world perturbations. Because we're trying to establish a commensurability with the satellite and the Earth's rotation, we'll have to develop specific definitions for the period of the satellite. In addition, we'll see that the concepts of orbital resonance applies as a result of this commensurability.

The anomalistic period is a fundamental quantity used to specify the orbital period. The ***anomalistic period*** (perigee-to-perigee ρ_K) is the time required for the satellite to complete one revolution with respect to perigee including the drift rate for argument of perigee (the time to advance 360° in mean anomaly). In the absence of perturbations, the anomalistic period is equal to the ***Keplerian period*** [Eq. (2-16)]. Thus we can view the anomalistic period as a perturbed or osculating Keplerian period.

$$\rho_K = \frac{2\pi}{n} = 2\pi\sqrt{\frac{a^3}{\mu}} \quad (10-20)$$

The Nodal Period

A repeat-groundtrack orbit depends on a commensurability between the time interval it takes the satellite to make two successive equator crossings called the satellite ***nodal period***, and the period of the Earth's rotation with respect to the ascending node—the ***nodal period of Greenwich***. Carter (1991) gives the nodal period of Greenwich as

$$\rho_{\theta_G} = \frac{2\pi}{\omega_{\oplus} - \dot{\Omega}} \quad (10-21)$$

McClain (1990) shows a way to determine the nodal period (node-to-node ρ_Ω). Because the nodal crossings are tied to the satellite's mean motion and the argument of perigee, it's reasonable to assume those quantities affect the period. McClain finds

$$\rho_\Omega = \frac{2\pi}{\dot{M} + \dot{\omega}} = \frac{2\pi}{n + \dot{M}_o + \dot{\omega}} = \frac{2\pi}{n} \left(\frac{1}{1 + \frac{\dot{M}_o + \dot{\omega}}{n}} \right) = \rho_K \left(\frac{1}{1 + \frac{\dot{M}_o + \dot{\omega}}{n}} \right)$$

Expressions for \dot{M}_o [Eq. (8-41)] and $\dot{\omega}$ [Eq. (8-39)] are already known. However, consider near-Earth satellites routinely used to study Earth resources, such as Landsat and RADARSAT, with nearly circular orbits, $e \approx 0.0011$. Because the eccentricity is of the same order as J_2 , $e^2 \approx J_2^2 \ll 1$. As a result, we can simplify the secular rate expressions (deleting terms of e^2) and add them together:

$$\dot{M}_o + \dot{\omega} \approx \frac{3nJ_2}{4} \left(\frac{R_{\oplus}}{a} \right)^2 \{ (2 - 3\sin^2(i)) + (4 - 5\sin^2(i)) \}$$

The nodal period is now

$$\rho_\Omega \approx \rho_K \frac{1}{1 + \frac{3J_2}{2} \left(\frac{R_{\oplus}}{a} \right)^2 (3 - 4\sin^2(i))} + O(e^2)$$

Truncating at a first-order series expansion (because x is of the order J_2) gives us

$$\frac{1}{1+x} = 1 - x + O(x^2)$$

which yields the final result:

$$\rho_{\Omega} \equiv \rho_K \left[1 - \frac{3J_2}{2} \left(\frac{R_{\oplus}}{a} \right)^2 (3 - 4 \sin^2(i)) \right] \quad (10-22)$$

Effect of Nodal Motion on Equator Crossings

The purpose of a repeat groundtrack is to periodically revisit and observe locations on the Earth. These locations may be few in number or many, but all repeat-groundtrack orbits require a set of fixed equator-crossing points and a more or less fixed orbital inclination. The fixed equator crossing points “tie” the groundtrack to the Earth at the equator. As long as the orbital inclination remains unchanged, the groundtrack remains fixed at the higher latitudes. For most of this discussion, we’ll assume the inclination remains fixed, although it does drift slowly due to perturbations. We’ll consider the effect of this drift later.

The two types of equator crossings are ascending, which cross from the Southern Hemisphere to the Northern Hemisphere, and descending, which cross in the opposite direction. Unless otherwise stated in this discussion, I’ll always refer to equator crossing points of the same type. The usual convention is to use the ascending crossings, so I’ll concentrate on them.

Two key properties of the groundtrack are the number of *equator crossing points*, k_{cp} , and the *frequency*, k_{rep} , with which they occur (see Chap. 4 for example groundtracks). Usually, equator crossing points aren’t overflowed in contiguous order (the distinction between passes and revolutions in Chap. 4) because low-Earth satellites don’t have enough time to return to the ascending node before the neighboring crossing point has rotated underneath the node. Rather, the distance between *consecutive* crossing points is greater than the distance between *contiguous* crossing points.

Consider the Landsat 6 spacecraft, which failed to achieve mission orbit from launch. The mission orbit was designed to provide global coverage of the Earth every 16 days ($k_{rep} = 16$ d) at a density of roughly 172 km along the equator. This required 233 crossing points because $k_{cp} = 233 \approx 2\pi R_{\oplus} / 172$ km. Assuming the spacecraft starts over the equator at a time t_o , it takes 232 orbit revolutions (relative to the equator) to cover the 233 equator crossing points, or $232/16 = 14.5$ revolutions per day. In practice, it’s an iterative process to find a because we can choose the crossing point distance as well as the time needed to repeat the orbit. Sometimes a satellite must operate at a certain altitude. Example 10-3 demonstrates this concept. The time interval between two successive arrivals at the ascending node is called the *nodal period*, ρ_{Ω} , in Eq. (10-22).

The spacing, or the *groundtrack shift*, λ_S , between consecutive equator crossing points is defined by how far the Earth rotates during the nodal period, relative to the ascending node located in inertial space. If the Earth’s rotation, ω_{\oplus} , and nodal regression, Ω , are measured in radians per unit time, the angular spacing (in radians) is given by

$$\lambda_S = (\omega_{\oplus} - \dot{\Omega}) \rho_{\Omega} \quad (10-23)$$

We convert to kilometers by multiplying by the Earth's radius. For Landsat 6, $\lambda_S = 24.721\ 03^\circ$, or 2751.93 km [$2\pi R_{\oplus}(16) \div 233$]. Because the Earth is rotating to the East, the location of the next consecutive equator crossing will be west of the previous one. Thus, although λ_S gives the magnitude, the direction of the shift is westward.

Note that we defined the groundtrack-shift parameter, λ_S , between two successive node arrivals as depending on the Earth's rotation relative to the ascending node. If the node is fixed in inertial space (unperturbed), the shift depends only on the Earth's rotation. However, perturbations (most notably J_2) cause the node to regress secularly to the west along the equator. Remember that we select the satellite's nodal period, ρ_{Ω} , so it arrives back at the ascending node as the Earth's eastward rotation brings the next crossing point under the node. If the node moves to the west due to nodal regression, the desired equator crossing point arrives under the node before the satellite does. Over time, the target crossing point moves eastward, and when the satellite arrives at the node, it flies over a location west of the target location (see Fig. 10-12). This unwanted westward groundtrack shift is due to the nodal regression. That's why Eq. (10-23) measures the Earth's rotation relative to the *moving* node (Ω_{sec}) and why we must also account for changes in the nodal drift ($\Delta\dot{\Omega}_{sec}$) caused mainly by third-body perturbations. Although the change in the nodal drift is much smaller, it does grow over time and eventually causes unacceptable error in the repeat-groundtrack geometry. Later, we'll discuss how to correct this error through stationkeeping maneuvers.

Effect of Drag on Repeat-Groundtrack Orbits

Drag perturbations also affect the nodal period, ρ_{Ω} . Although ρ_{Ω} depends on the elements a , e , and i , the primary dependence is due to the semimajor axis through the mean motion. For low-Earth orbits, drag causes decay of the semimajor axis, which makes the nodal and Keplerian periods change. Basically, a change in the nodal period, $\Delta\rho_{\Omega}$, will cause a change in the groundtrack-shift parameter, $\Delta\lambda_S$ per revolution, similar to Eq. (10-23)

$$\Delta\lambda_S = (\omega_{\oplus} - \dot{\Omega}) \Delta\rho_{\Omega}$$

The change in ρ_{Ω} due to a decaying semimajor axis

$$\Delta\rho_{\Omega} = \frac{\partial\rho_{\omega}}{\partial a} \Delta a + O(\Delta a)^2 \cong \frac{3\pi}{na} \Delta a + O(J_2)$$

caused mainly by the change in the Keplerian period, ρ_K , and higher-order effects on the order of J_2 . Consequently, a decay in the semimajor axis ($\Delta a < 0$), causes the nodal period to decrease ($\Delta\rho_{\Omega} < 0$), and the satellite arrives back over the equator sooner than desired. It crosses the equator east of the desired crossing point. This eastward shift continues to grow as the semimajor axis decays.

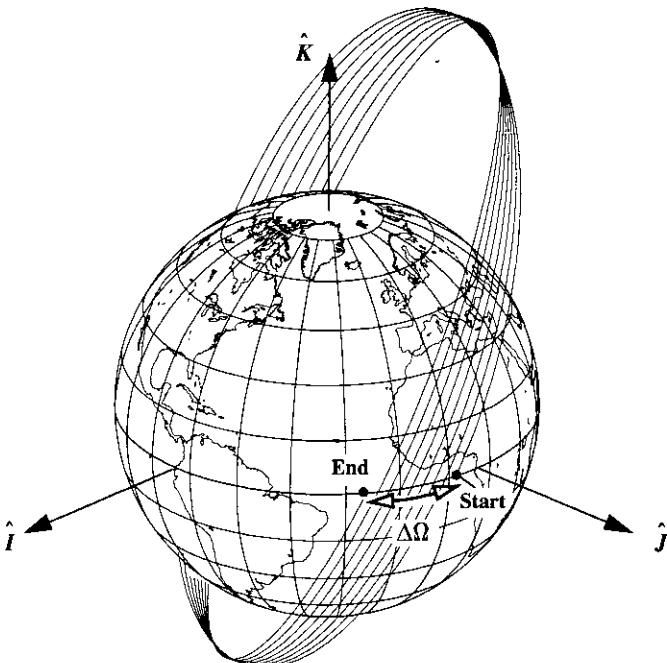


Figure 10-12. Drift in Ascending Node. This is an exaggerated view of how J_2 secular perturbations affect the orbital plane. The figure shows J_2 magnified 20 times on a nonrotating Earth.

For convenience, let's assume that the semimajor axis's rate of decay is constant. Also for convenience, I'll drop the subscript on ρ_Ω and use ρ_i to denote the nodal period at the end of revolution i . Consider the decay of the nodal period starting from the nominal value ρ_o . After one revolution the value is $\rho_1 = \rho_o - \Delta\rho$, where $\Delta\rho$ is the change in the period due to the change in the semimajor axis. However, the time of arrival is changed only by the average change in ρ over one revolution because $\Delta\rho = 0$ at the beginning. Thus, the accumulated change in the time of arrival (Δt) relative to the desired time is $\Delta t = \Delta\rho/2$. With no further decay, because $\rho_1 = \rho_o - \Delta\rho$, the change in the time of arrival after the next revolution would be $\Delta\rho + \Delta\rho/2$: the average period decay over the first revolution summed with the $\Delta\rho$ change of the new constant nodal period over the second revolution. However, if decay continues over the second revolution, the time of arrival shortens further by the average period decay over this interval; thus, $\Delta t = \Delta\rho + \Delta\rho/2 + \Delta\rho/2$. The resulting nodal period is now $\rho_2 = \rho_o - 2\Delta\rho$. Without further decay, after the third revolution the accumulated change in the time of arrival will be due to the earlier accumulated change over revolutions 1 and 2, and the difference between the changed nodal period at the beginning of revolution 3, with the nominal value ρ_o is $2\Delta\rho + 3\Delta\rho = 5\Delta\rho$. If decay takes place over this third revolution, the accumulated Δt over this revolution is $\Delta\rho_n/2$, which we add to the

$5\Delta P$ to obtain $5\Delta P + \Delta P/2$. The relationship becomes apparent in Table 10-3. Keep in mind that the change in the satellite's time of arrival over the equator, $\Delta t_{rev} = \Delta P/2$, is due to the change in nodal period caused by the semimajor axis's decay over one revolution—relative to that already expected at the beginning of the revolution. Also, the change in the satellite's time of arrival due to two distinct nodal periods P_o and P is their difference, $\Delta t_{diff} = P - P_o$.

TABLE 10-3. Effect of Drag on a Satellite's Period. The effect of drag on a satellite's orbital period is shown as a periodic effect (the change in nodal period on each revolution), a difference from the previous revolution, and the accumulated effect from the original period.

Revolution #	Nodal Period at End of Revolution P_i	$\Delta t = \Delta t_{diff} + \Delta t_{rev}$ over 1 Revolution	Accumulated Δt_{acc} from Revolution 0
0	P_o	0	0
1	$P_1 = P_o - \Delta P$	$\Delta P + \Delta P/2$	$\Delta P + \frac{\Delta P}{2} = \frac{3}{2}\Delta P$
2	$P_2 = P_1 - \Delta P$ $= P_o - 2\Delta P$	$2\Delta P + \Delta P/2$	$\frac{5\Delta P}{2} + \frac{3\Delta P}{2} = \frac{8}{2}\Delta P$
3	$P_3 = P_2 - \Delta P$ $= P_o - 3\Delta P$	$3\Delta P + \Delta P/2$	$\frac{15}{2}\Delta P$
j	$P_j = P_o - j\Delta P$	$j\Delta P + \Delta P/2$	$\sum_{k=1}^j \left(k\Delta P + \frac{\Delta P}{2} \right) = \left(\frac{1}{2}j^2 + j \right)\Delta P$

Thus we deduce that, if ΔP_Ω is the change in nodal period per revolution, the accumulated effect on the groundtrack-shift parameter after j revolutions is

$$\sum_{k=1}^j \Delta \lambda_{S_k} = \frac{1}{2} (j^2 + j) (\omega_\oplus - \dot{\Omega}) \Delta P_\Omega$$

The change in the groundtrack shift grows quadratically with the number of revolutions and results in severe decay of the repeat groundtrack under strong drag conditions. Consequently, periodic orbit-adjustment burns must be made to restore the semimajor axis if we are to maintain the repeat groundtrack.

Planning Orbit-Maintenance Maneuvers

Let's consider a strategy for developing an orbit-adjustment burn plan to maintain our repeat groundtrack. We'll consider perturbations in the elements a and i that affect the

nodal period (drag) and the nodal regression drift rate (central-body and third-body). We won't consider the effects of perturbations in the eccentricity because they're usually small for repeat-groundtrack orbits of near-Earth satellites in nearly circular orbits.

As we have seen, perturbations acting on the satellite's orbital motion cause the equator crossing points to drift in longitude. The main cause of this groundtrack drift at the equator is decay in the semimajor axis caused by atmospheric drag, as well as drift in the orbital inclination due to the effect of the Sun's gravitational attraction. Assuming small perturbations in the spacecraft's motion so we can neglect higher-order terms, the groundtrack drift after one revolution is given by a truncated Taylor series expansion [using Eq. (10-23)]

$$\begin{aligned}\Delta\lambda_S &= \frac{\partial\lambda_S}{\partial a}\Delta a + \frac{\partial\lambda_S}{\partial i}\Delta i \\ \frac{\partial\lambda_S}{\partial a} &= R_{\oplus}(\omega_{\oplus} - \dot{\Omega})\frac{\partial\beta_{\Omega}}{\partial a} - \frac{\partial\dot{\Omega}}{\partial a}R_{\oplus}\beta_{\Omega} \quad (10-24) \\ \frac{\partial\lambda_S}{\partial i} &= R_{\oplus}(\omega_{\oplus} - \dot{\Omega})\frac{\partial\beta_{\Omega}}{\partial i} - \frac{\partial\dot{\Omega}}{\partial i}R_{\oplus}\beta_{\Omega} \\ \Delta a &= \Delta a_o + \frac{da}{dt}(t - t_o) \quad \Delta i = \Delta i_o + \frac{di}{dt}(t - t_o)\end{aligned}$$

The average rate of the drift over one revolution is

$$\frac{\Delta\lambda_S}{\beta_{\Omega}} = \frac{1}{\beta_{\Omega}} \left[\frac{\partial\lambda_S}{\partial a}\Delta a_o + \frac{\partial\lambda_S}{\partial i}\Delta i_o + \left(\frac{\partial\lambda_S}{\partial a}\frac{da}{dt} + \frac{\partial\lambda_S}{\partial i}\frac{di}{dt} \right)(t - t_o) \right]$$

Substituting a continuous representation for the discrete representation above yields

$$\begin{aligned}\frac{d\lambda_S}{dt} &= k_1 + k_2(t - t_o) \\ \frac{d\lambda_S}{dt} &= \frac{d}{dt}(\Delta\lambda_S) = \frac{d}{dt}(\lambda_S - \lambda_{Snom}) = \frac{d\lambda_S}{dt} - \frac{d\lambda_{Snom}}{dt} = \frac{d\lambda_S}{dt} - 0 \quad (10-25) \\ k_1 &= \frac{1}{\beta_{\Omega}} \left(\frac{\partial\lambda_S}{\partial a}\Delta a_o + \frac{\partial\lambda_S}{\partial i}\Delta i_o \right) \quad k_2 = \frac{1}{\beta_{\Omega}} \left(\frac{\partial\lambda_S}{\partial a}\frac{da}{dt} + \frac{\partial\lambda_S}{\partial i}\frac{di}{dt} \right)\end{aligned}$$

We get the cumulative groundtrack drift at any time t by integrating the above expression for the groundtrack-drift rate. It yields a quadratic expression,

$$\Delta\lambda_S = \lambda_S - \lambda_{S0} = k_1(t - t_o) + \frac{1}{2}k_2(t - t_o)^2 \quad (10-26)$$

where $\Delta\lambda_S$ is the groundtrack drift from the initial value λ_{S0} over the time interval $t - t_o$. Usually, these values are offsets from the nominal groundtrack value, λ_{Snom} in Eq. (10-23), at their respective times. To maximize the interval between groundtrack maintenance

burns we take advantage of the entire window about the nominal groundtrack within which the mission is successful. The quadratic's drift profile is shown in Fig. 10-13. We burn at C to reverse the groundtrack drift by changing the semimajor axis. The Δa is chosen so that by the time the groundtrack drifts to the maximum eastward displacement at point D , the effect of drag has reduced a to the nominal value for the repeat groundtrack. As the decay effects continue, the groundtrack drift reverses direction until the next burn at point E .

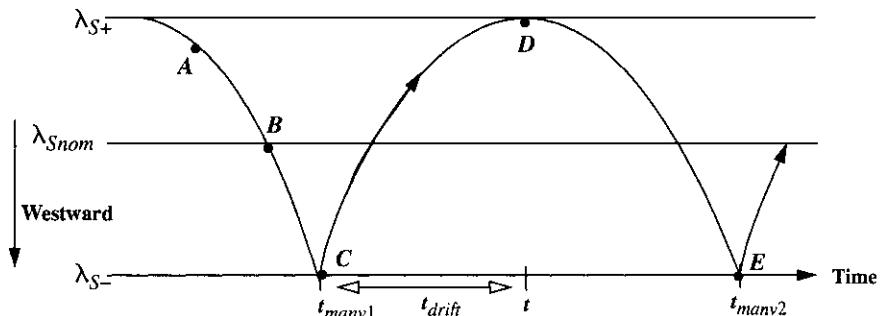


Figure 10-13. Profile of Groundtrack Drift. This figure shows the change in the groundtrack-shift parameter, λ_s , over time as corrected by periodic burns. There are usually limits placed about the nominal value (\pm) within which the satellite mission is successful. Suppose A is the initial groundtrack displacement resulting from orbit insertion. We can maximize the efficiency of our maneuvers by allowing the groundtrack to drift to point C . The burn reverses the groundtrack drift until point D where drag again begins to change the direction to the west. Example 10-3 begins at point B .

Note that this quadratic profile was derived from a first-order Taylor series expansion given in Eq. (10-24); thus, the coefficients are evaluated at time t_o , usually the burn time t_{manv1} . The coefficient k_1 depends on offsets in the semimajor axis and inclination, and k_2 depends on the drift rates in these elements. We assume the element rates are constant in this analysis: they depend only on the natural perturbing forces acting on the spacecraft's motion and hence can't be altered. However, a burn can yield offsets in the elements a and i and thus affect the coefficient k_1 . The most efficient technique is to alter the value of the semimajor axis; thus, the following discussion assumes no correction to the inclination, $\Delta i_o \equiv 0$. On the other hand, it also assumes that an impulsive change in the semimajor axis, Δa_o , is obtained by an impulsive change to the velocity, Δv , at time t_{manv1} .

The quadratic expression [Eq. (10-26)] provides the basis for finding the Δv necessary to cause the groundtrack to drift from the current offset value λ_{S-} at the burn time t_{manv1} , to the maximum allowed westward drift λ_{S+} after an elapsed time interval t_{drift} . First, we select the maximum westward displacement of the groundtrack. Next, we determine the free parameters in the coefficients (k_1, k_2) to achieve the desired effect because these parameters must be related to the Δv . We must first determine the time of maximum west-

ward drift. The maximum groundtrack drift occurs at time $t - t_{manv1} = t_{drift}$ and satisfies the condition $d\lambda_S / dt = 0$ which, in view of Eq. (10-25), yields

$$t_{drift} = \frac{k_1}{k_2}$$

Clearly, the values of k_1 and k_2 must differ in sign. Evaluating the quadratic's drift profile [Eq. (10-26)] at this time obtains the maximum westward drift value:

$$\lambda_{S+} - \lambda_{S-} = -k_1 \frac{k_1}{k_2} + \frac{1}{2} k_2 \left(-\frac{k_1}{k_2} \right)^2 = -\frac{1}{2} \frac{1}{k_2} k_1^2$$

If we take the left-hand side of this equation to be positive, k_2 must be of opposite sign, or $k_2 \leq 0$. Solving this expression for k_1 gives

$$k_1 = \sqrt{2k_2(\lambda_{S-} - \lambda_{S+})}$$

The desired offset in semimajor axis is obtained by inverting the expression for k_1 in Eq. (10-25). Adding this offset to the original value gives us the corrected semimajor axis to maintain the repeat-groundtrack criteria (remember that $\Delta t_o \equiv 0$)

$$\Delta a = k_1 \rho_\Omega \left(\frac{\partial \lambda_S}{\partial a} \right)^{-1}$$

Assuming the velocity change is along the velocity vector, the conversion of the semimajor axis increment to the magnitude of the velocity increment is given by

$$\Delta v = \frac{n^2 a}{2v} \Delta a$$

If we also neglect the orbital eccentricity (remember we assumed it was small), the velocity direction coincides with the transverse direction (normal to the position vector in the orbital plane and positive in the direction of motion, $v = n/a$), and the expression becomes

$$\Delta v = \frac{n}{a} \Delta a + O(e)$$

It's important to note that we've corrected the nodal period caused by decay in both the semimajor axis and inclination by adjusting only the semimajor axis through the orbit-adjustment burn. We didn't correct the inclination because the inclination drift is quite small and very slow compared to the semimajor axis's decay due to drag. Also, inclination-adjustment burns require out-of-plane burns and are expensive in terms of Δv and fuel. They should be avoided as long as possible.

Actually, we were able to compensate for the effect of inclination drift by adjusting the semimajor axis—not back to the nominal value, but to a biased value near the nominal value. This biased value modified the Keplerian-period part of the nodal period to compensate for the effect of the drift in the inclination. This is a clever way to avoid an unde-

sirable out-of-plane burn, but it has one drawback in certain circumstances. Many spacecraft that fly repeat-groundtrack orbits also fly Sun-synchronous orbits. In these cases, the solar perturbation eventually causes enough inclination drift so we must adjust the inclination to maintain the Sun-synchronous constraint. When we do this burn, the nodal period abruptly changes because of our specially selected semimajor axis used to compensate for inclination drift. Thus, soon after we adjust the inclination (12 hours or so), we have to adjust the semimajor axis to reestablish the correct nodal period by adjusting a back to its nominal value.

I've shown approximations for the partial derivatives of the nodal period and nodal rate with respect to a and i in Algorithm 63. The rates of change for a and i are usually found from numerical analyses using the force models selected for the mission. These usually include central-body, drag, third-body, and solar-radiation effects. However, drag is the dominant force affecting a and the third body has the main effect on i (for Sun-synchronous cases).

ALGORITHM 63: Repeat Groundtrack ($a, e, i, \lambda_{S+}, \lambda_{S-}, da/dt, di/dt \Rightarrow a_{new}$)

$$p = a(1 - e^2)$$

$$\dot{\Omega} = -\frac{3nR_\oplus^2 J_2}{2p^2} \cos(i) \quad n = \sqrt{\frac{\mu}{a^3}}$$

$$\frac{\partial \rho_\Omega}{\partial a} = \frac{3\pi}{na} \left[1 + \frac{1}{2} J_2 \left(\frac{R_\oplus}{a} \right)^2 (4 \cos^2(i) - 1) \right] + O(e)$$

$$\frac{\partial \rho_\Omega}{\partial i} = \frac{12\pi}{n} J_2 \left(\frac{R_\oplus}{a} \right)^2 \sin(2i) + O(e)$$

$$\frac{\partial \dot{\Omega}}{\partial a} = -\frac{7}{2} \frac{\dot{\Omega}}{a} \quad \frac{\partial \dot{\Omega}}{\partial i} = -\dot{\Omega} \tan(i)$$

$$\rho_\Omega \equiv \frac{2\pi}{n} \left[1 - \frac{3J_2}{2} \left(\frac{R_\oplus}{a} \right)^2 (3 - 4 \sin^2(i)) \right]$$

$$\lambda_{Snom} = (\omega_\oplus - \dot{\Omega}) \rho_\Omega$$

$$\frac{\partial \lambda_S}{\partial a} = R_\oplus (\omega_\oplus - \dot{\Omega}) \frac{\partial \rho_\Omega}{\partial a} - \frac{\partial \dot{\Omega}}{\partial a} R_\oplus \rho_\Omega$$

$$\frac{\partial \lambda_S}{\partial i} = R_\oplus (\omega_\oplus - \dot{\Omega}) \frac{\partial \rho_\Omega}{\partial i} - \frac{\partial \dot{\Omega}}{\partial i} R_\oplus \rho_\Omega$$

$$k_2 = \frac{1}{\rho_\Omega} \left(\frac{\partial \lambda_S da}{\partial a dt} + \frac{\partial \lambda_S di}{\partial i dt} \right) \quad k_1 = \sqrt{2k_2(\lambda_{S-} - \lambda_{S+})}$$

$$t_{drift} = -\frac{k_1}{k_2}$$

$$\Delta a = k_1 \rho_\Omega \left(\frac{\partial \lambda_S}{\partial a} \right)^{-1}$$

$$\Delta v = \frac{n}{a} \Delta a \text{ for small } e$$

10.4.3 Minimum Altitude Variation Orbits

Sometimes, satellites carry sensors that require nearly constant altitudes, either at one location or, more often, globally. We usually have a constraint that dictates the variation in altitude, $r_p + \Delta r < \text{tolerance}$. A satellite's altitude, h_{ellp} , depends on two components: the satellite's geocentric distance, r , and the geocentric distance of the subsatellite point on the Earth's surface, r_{site} , so that $h_{ellp} = r - r_{site}$. From Eq. (2-14) we can find r as a function of the orbital elements, and r_{site} depends on the model of the Earth's surface (recall the discussion in Chap. 3 with Algorithm 23). The difference in r_{site} between a spherical and an ellipsoidal Earth model amounts to a little more than 21 km at the poles and vanishes at the equator. Remember that we specify height above a reference ellipsoid. Because reference ellipsoids are symmetrical in longitude, r_{site} depends only on the geodetic latitude, ϕ_{gd} . Thus, variations in the altitude for a fixed site are caused by variations in the satellite's geocentric distance Δr . Global variations must also consider variations in the reference ellipsoid, but the reference ellipsoid is very well known, so

$$\Delta h_{ellp} = \Delta r - \Delta r_{ellipsoid}$$

The primary global variation in r results from the eccentricity of the ellipse where two-body motion occurs. This maximum difference is determined by examining the apogee and perigee radii. It is $\Delta r_{max} = 2ae$.

It would seem, at first glance, that specifying a mission orbit with vanishing eccentricity would eliminate this variation; however, perturbations don't permit a vanishing eccentricity to persist. Most notably, the zonal harmonics cause long-periodic and short-periodic variations in the eccentricity. Because the osculating eccentricity can't be negative, the mean eccentricity must be greater than zero and the osculating eccentricity must be greater than the mean eccentricity half the time (on average). Under certain circumstances, the mission orbit can be designed to nullify the long-periodic oscillations (which contribute most of the variation in the eccentricity and in the argument of perigee). We call this a **frozen-eccentricity orbit**—it minimizes global variations in altitude. We'll discuss this in more detail later. We define the satellite's geocentric distance relative to the reference frame's origin (center of the Earth)—not the Earth's surface. Thus, if we want to analyze the altitude above a given point on the Earth's surface, we must introduce a dependence on the location without corrupting the expression for r . Remember that the reference ellipsoid

depends only on ϕ_{gd} . We can introduce a latitude dependence using the argument of latitude, u . Eliminating ν from the trajectory equation [Eq. (2-14)] in favor of u yields

$$r = \frac{a(1-e^2)}{1+e\cos(u-\omega)} \quad (10-27)$$

This expression yields r as a function of the argument of latitude, u and the argument of perigee, in addition to a , e . We treat u and ω as mutually independent parameters in the following analysis.

For two-body motion, $r = r(u)$ will remain constant at a fixed value ($u = u_o$). The variation around the orbit is due to the orbit's eccentricity ($\Delta r_{max} = 2ae$). For perturbed motion, variations in a , e , and ω will cause variations in r . The perturbation in u (independent of the perturbation in ω) reflects a change in the satellite's arrival time at the location specified by $u = u_o$, not a change in $r(u)$. Because this applies to any value of u along the orbit, it also applies to the global problem, so we don't consider the perturbation in u for effects on the altitude variation.

The effect on r of perturbations in a , e , and ω to first order is

$$\Delta r = \frac{\partial r}{\partial a} \Delta a + \frac{\partial r}{\partial e} \Delta e + \frac{\partial r}{\partial \omega} \Delta \omega + \text{higher-order terms} \quad (10-28)$$

Let's consider each of these contributions separately.

The semimajor axis's decay due to drag maps directly into altitude decay. Using Eq. (10-27),

$$\Delta r = \frac{\partial r}{\partial a} \Delta a + O(\Delta a)^2 = \frac{r}{a} \Delta a + O(\Delta a)^2 = \frac{(1-e^2) \Delta a}{1+e\cos(u-\omega)} + O(\Delta a)^2 \quad (10-29)$$

Because $u - \omega = \nu$, the maximum effect occurs at apogee ($\nu = \pi$), where the denominator takes on its minimum value. The minimum effect occurs at perigee. We can also use Eq. (10-29) to find the change in altitude over a fixed latitude location associated with $u = u_o$.

Eccentricity variations are mainly due to the zonal harmonics; they affect the satellite's geocentric distance by

$$\Delta r = \frac{\partial r}{\partial e} \Delta e + O(\Delta e)^2$$

where the partial derivative has two components—one depends on the explicit appearance of e in the expression for r ; the other is from the implicit dependence of u (through ν) on e . Remember that $\nu = \nu(M, e)$.

$$\frac{\partial r}{\partial e} = \left. \frac{\partial r}{\partial e} \right|_{\text{explicit}} + \frac{\partial r}{\partial u} \frac{\partial u}{\partial e}$$

The implicit term accounts for the difference in r due to a displacement of Δu along the orbit, which is caused by Δe . Any displacement in Δu can be viewed as a time-of-arrival change and thus does not contribute to the altitude variation upon actual arrival at a fixed

Earth location. Because this applies to any location, it also applies to the global-variation problem. Consequently, to first order,

$$\Delta r = -\frac{r}{a(1-e^2)} [r \cos((u-\omega) + 2ae)] \Delta e \quad (10-30)$$

We can determine the maximum effect on Δr due to Δe by analyzing this expression. This yields the location in the orbit where the maximum effect occurs. Using this information and the maximum Δr permitted, we determine the maximum permissible Δe to satisfy the constraint on global variations in altitude. When we reach this value, we do an eccentricity-adjustment burn to restore the nominal value. As we saw on page 293, the most efficient eccentricity-adjustment burn uses thrust in the transverse (normal to the radius vector) direction while at perigee or apogee:

$$\Delta e = \frac{\sqrt{1-e^2}}{na} (\cos(E) + \cos(\nu)) \Delta v_T$$

We can solve this expression to obtain the Δv required for the eccentricity restoration. For cases of small eccentricity, the transverse direction differs from the direction of the velocity vector by order (e), and the Δe correction simplifies to

$$\Delta e = \frac{2\Delta v}{na} \cos(M) + O(e)$$

where M is the mean anomaly, and the Δv is applied along direction of the velocity vector. We determine the sign of the eccentricity correction using the direction of the velocity vector and the location of the burn (perigee or apogee). For positive Δv (increasing the velocity vector), the change in the eccentricity is positive for perigee burns and negative for apogee burns. The opposite holds for a Δv applied against the velocity vector, which decreases the satellite's velocity.

Perigee drift, due mainly to the zonal harmonics, can also cause variations in the altitude by changing the argument of perigee. Figure 10-14 shows the general geometry depicting variations in perigee location.

The perigee drift causes a variation in the distance given by

$$\Delta r = \frac{\partial r}{\partial \omega} \Delta \omega + O(\Delta \omega)^2 = -\frac{r^2}{a(1-e^2)} e \sin(u-\omega) \Delta \omega + O(\Delta \omega)^2$$

We can analyze this equation, to first order in $\Delta \omega$, to find the maximum change in r . The variation vanishes at $\nu = u - \omega = 0^\circ, 180^\circ$ and it's a maximum at $\nu = u - \omega = 90^\circ, 270^\circ$. Thus,

$$|\Delta r_{max}| = ae(1-e^2)|\Delta \omega| \quad (10-31)$$

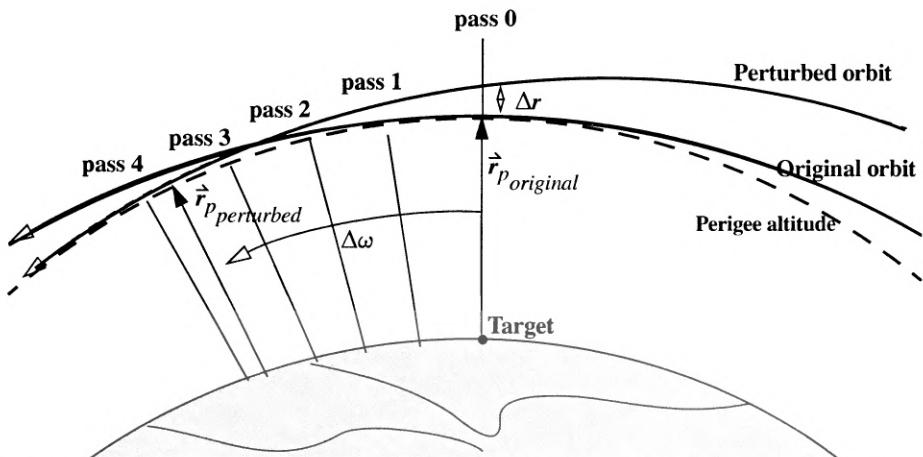


Figure 10-14. Perigee Drift. This figure is an exaggerated view of how J_2 secular perturbations cause perigee to drift. On each pass, perigee drift causes the satellite's altitude over the target to increase. We must correct the orbit between passes three and four to keep the satellite within mission limits. The perturbed orbit is shown at the exact point where the target will exit the satellite's field of view.

Note that, for small e , $|\Delta r_{max}| \approx ae\Delta\omega$ and thus is small but not necessarily negligible for nearly circular orbits.

Given a maximum permissible Δr , we can invert Eq. (10-31) to determine the maximum perigee drift before we must do an orbit-adjustment burn to restore the perigee to its nominal value. We estimate the time until this burn using the allowable drift $|\Delta\omega|$ and the drift rate $|\dot{\omega}|$.

$$\Delta t_{max} = \left| \frac{\Delta\omega}{\dot{\omega}} \right| \quad (10-32)$$

The analysis for a fixed location on the Earth requires us to evaluate the partial derivative $\partial r / \partial \omega$ at the particular argument of latitude $u = u_o$. Then,

$$\Delta\omega = \Delta r_{max} \div \left| \frac{\partial r}{\partial \omega} \right|_{u=u_o}$$

and the time of allowed drift is the same as above.

However, note that this analysis breaks down for Earth locations near the orbit's perigee or apogee because $\sin(u - \omega) = 0$ for $u - \omega = 0, \pi$. Consequently, $\partial r / \partial \omega = 0$, so to first order in $\Delta\omega$ the perigee drift doesn't affect the altitude. That means we have no information on the allowed drift interval because $|\Delta\omega / \dot{\omega}| = 0$ to first order. This is an important case because many reconnaissance missions rely on placing the argument of perigee

over a location of interest to minimize the altitude. For these special cases, we must extend the analysis to second order in $\Delta\omega$. Let

$$\Delta r = \frac{\partial r}{\partial \omega} \Delta\omega + \frac{1}{2!} \frac{\partial^2 r}{\partial \omega^2} (\Delta\omega)^2 + O(\Delta\omega)^3$$

The first-order partial derivative obtained above vanishes at perigee and apogee $\partial r / \partial \omega = 0$. The second-order partial derivative is

$$\frac{\partial^2 r}{\partial \omega^2} = -\frac{2re^2 \sin^2(u-\omega)}{a(1-e^2)} \frac{\partial r}{\partial \omega} + \frac{r^2 e \cos(u-\omega)}{a(1-e^2)}$$

Evaluating this expression at perigee ($u - \omega = 0$) yields

$$\left. \frac{\partial^2 r}{\partial \omega^2} \right|_{u-\omega=0} = ae \left(\frac{1-e}{1+e} \right)$$

Thus, at perigee and apogee,

$$\Delta r = -\frac{ae}{2} \left(\frac{1-e}{1+e} \right) (\Delta\omega)^2 + O(\Delta\omega)^4$$

Note that the effect of the third-order term $O(\Delta\omega)^3$ also vanishes if we evaluate $\partial^3 r / \partial \omega^3 \Big|_{u-\omega=0}$ because Δr is an even function of $\Delta\omega$ displacements about perigee (apogee), $\Delta r(\Delta\omega) = \Delta r(-\Delta\omega)$.

Again, assuming a maximum variation in altitude, Δr_{max} , we invert the second-order result to obtain

$$\Delta\omega = \pm \sqrt{\frac{2}{ae} \left(\frac{1+e}{1-e} \right) \Delta r_{max}} \quad (10-33)$$

This is the maximum drift from perigee. To get the maximum drift time, we use Eq. (10-32) with the new expression for $\Delta\omega$.

Once we obtain $\Delta\omega$ from either the first-order or second-order analysis, as appropriate, we must find the Δv value to achieve the desired correction. The amount of Δv required depends on the magnitude of the correction $|\Delta\omega|$, the direction of the thrust vector in the orbital plane, and the location in the orbit where we do the burn. We can show that the most efficient burn is in the transverse direction at $\nu = \pm 90^\circ$. This is the *S*-axis in the *RSW* system. Thus,

$$\Delta\omega = \frac{\sqrt{1-e^2}}{nae} \left(1 + \frac{r}{p} \right) \sin(\nu) \Delta v_s$$

For small e cases, the difference between the transverse direction and the velocity direction is of order $O(e)$, so we can neglect it. Also, the above expression simplifies to

$$\Delta\omega = \frac{2}{nae} \sin(M) \Delta v_S + O(e)$$

where Δv_S is applied along the velocity vector. Thus, at $\nu = M = \pm 90^\circ$,

$$\Delta\omega = \pm \frac{2}{nae} \Delta v_S + O(e)$$

The particular value of ν or M depends on the required sign of the perigee correction, and the Δv_S magnitude is

$$\Delta v_S = \frac{nae}{2} |\Delta\omega| + O(e)$$

Typically, $e \approx 0.001$, so the error incurred by assuming a small eccentricity is on the order of 0.1%. This is smaller than typical variations in thruster performance between burns.

Algorithm 64 shows the required equations to determine the time before burning and the magnitude of the burn.

ALGORITHM 64: Minimum Altitude Variation ($\Delta r, a, e, i \Rightarrow k_{max}$)

$$n = \sqrt{\frac{\mu}{a^3}} \quad \dot{\omega} = \frac{3nR_\oplus^2 J_2}{4a^2 (1-e^2)^2} (4 - 5 \sin^2(i))$$

$$\Delta r = \frac{(1-e^2) \Delta a}{1 + e \cos(u - \omega)}$$

$$\Delta r = \frac{r}{a(1-e^2)} [r \cos(u - \omega) - 2ae] \Delta e$$

$$\Delta r_{max} = \frac{ae^2(1-e^2)}{1+5e^2} [2e + \sqrt{1-e^2}] \Delta \omega$$

$$\Delta\omega_{max} = \sqrt{2 \frac{\Delta r}{ae} \left(\frac{1+e}{1-e} \right)}$$

$$k_{max} = \left| \frac{\Delta\omega_{max}}{\dot{\omega}} \right|$$

$$\Delta v_S = \frac{nae}{2} |\Delta\omega_{max}|$$

10.4.4 Frozen-Orbit Eccentricity and Argument of Perigee

As mentioned earlier, this technique is used to design orbits that minimize global variations in altitude by nulling the long-periodic variations in the eccentricity and argument of perigee. Consider the simplified example of an Earth satellite perturbed only by the J_2 and J_3 zonal harmonics (see Cutting et al., 1978). The long-periodic argument of perigee rate obtained after averaging the short-periodic variations is

$$\frac{d\omega}{dt} = \frac{3n}{(1-e^2)^2} J_2 \left(\frac{R_\oplus}{a} \right)^2 \left(1 - \frac{5}{4} \sin^2(i) \right) \theta$$

$$\theta = 1 + \frac{J_3}{J_2} \left(\frac{R_\oplus}{a} \right) \left(\frac{1}{(1-e^2)} \right) \left(\frac{\sin^2(i) - e \cos^2(i)}{\sin(i)} \right) \frac{\sin(\omega)}{e}$$

The long-periodic eccentricity rate is

$$\frac{de}{dt} = -\frac{3}{2} \frac{n}{(1-e^2)^2} J_3 \left(\frac{R_\oplus}{a} \right)^3 \sin(i) \left(1 - \frac{5}{4} \sin^2(i) \right) \cos(\omega)$$

Note that $d\omega/dt$ vanishes at critical i_ω or when $\theta = 0$. For given values of a and i , values of $e = e_o$ and $\omega = \omega_o$ exist which satisfy the condition $\theta = 0$. However, a nonvanishing eccentricity rate will cause the eccentricity value to change so the condition $\theta = 0$ is no longer satisfied. Hence, de/dt becomes nonzero and ω drifts from the value selected to help satisfy the condition $\theta = 0$.

The eccentricity rate vanishes if $i = 0$, $i = i_\omega$, or $\omega_o = 90^\circ, 270^\circ$. Any of these conditions results in a *frozen eccentricity*. The *argument of perigee is frozen* for $i = i_o, i_\omega$, and for specific values of e and ω that satisfy $\theta = 0$. Most Earth-observation missions don't fly in the equatorial plane or at critical inclination. For these cases, we select $\omega_o = 90^\circ$ or 270° , which nulls the eccentricity rate, and we determine the value of e that satisfies $\theta = 0$. For $\omega_o = 90^\circ$, McClain (1987) shows that, if we neglect terms of order $O(J_2 e)$,

$$e_o \approx -\frac{1}{2} \frac{J_3}{J_2} \left(\frac{R_\oplus}{a} \right) \sin(i) \sin(\omega)$$

For real-world applications, we extend the analysis to include other zonal harmonics. Continuing the example of the Landsat 6 mission (page 766), $J_2 = -0.108\ 265 \times 10^{-2}$ and $J_3 = -0.254\ 503 \times 10^{-5}$ results in $e_o \approx 0.001\ 05$. Including zonal harmonics through J_{21} changes this value only to $e_o \approx 0.001\ 15$.

These frozen-orbit elements are essentially single-averaged, mean elements resulting from eliminating the short-periodic variations. If these values are used to establish the initial values of the osculating elements,

$$e_{osc} = e_{mean} + \Delta e_{SP}$$

$$\omega_{osc} = \omega_{mean} + \Delta \omega_{SP}$$

The long-periodic motion in these two elements is reduced, resulting in passive control for the eccentricity, the argument of perigee, and the altitude (globally). Of course the passive altitude control assumes no decay in the semimajor axis.

If the initial values of the mean eccentricity and argument of perigee are displaced from the frozen-orbit values (e_o, ω_o), they will exhibit long-periodic oscillations about these values. The amplitude of these oscillations will depend on the amount of the displacement. If we generate time histories for the mean eccentricity and argument of perigee through numerical integration and use them to plot the quantities $e\cos(\omega)$ and $e\sin(\omega)$, we obtain a closed-circle trajectory centered on the frozen-orbit values. Different initial values for e and ω will produce other trajectories in concentric circles, provided e and ω are close enough to the frozen values. For more distant starting values, the trajectories appear as sinusoidal waves. Thus, if we start close to the frozen-orbit values, we'll remain close.

You can use this behavior to refine the initial estimate of the frozen-orbit eccentricity obtained from the simplified J_2, J_3 analysis. Using the first estimate for e_o , we generate time histories for the mean elements using numerical integration of the mean element rates obtained from the single-averaged VOP equations of motion, or an equivalent method. We do this with a comprehensive zonal-harmonic model. We don't include other perturbations because the phenomena of a frozen-orbit eccentricity/perigee depends only on the perturbing force from the zonal harmonics. Using the element histories for mean e and ω , we generate the circular trajectories by plotting $x = e\cos(\omega)$, $y = e\sin(\omega)$; then, determine the coordinates of the center x_o, y_o and solve for e_o :

$$e_o = \sqrt{x_o^2 + y_o^2}$$

It's not necessary to solve for ω_o because only two values are permissible. The correct value is easily determined from the J_2, J_3 analysis, and it doesn't change because of additional zonal harmonics.

The frozen-orbit geometry is surprisingly stable, suffering little decay due to other natural perturbations—even drag (see McClain, 1987). However, one disturbance—orbital adjustment maneuvers—will disrupt the frozen-orbit geometry. Most frozen-orbit conditions for eccentricity/perigee are used by Earth-observing missions that also employ repeat-groundtrack orbits and Sun-synchronous orbits. We aren't concerned with inclination-adjustment burns because these are out-of-plane and don't affect the eccentricity or argument of perigee. However, the altitude- or semimajor-axis-adjustment burn to maintain a repeat-groundtrack will disrupt the frozen values for eccentricity and argument of perigee. The burn plane must be developed to avoid disrupting the frozen-orbit geometry. Several solutions exist—see Cutting et al. (1978) for one approach.

10.4.5 Application: Designing a Specialized Orbit

Let's examine a photo-reconnaissance satellite over a hypothetical target at the Colosseum in Rome, Italy. The site coordinates are 41.52°N and 12.37°E at an altitude, H_{MSL} , of 152 m. The satellite could launch from anywhere, but for this example, we'll say it starts at

epoch exactly at perigee over the Colosseum. The satellite has a scanning sensor with a total field of view of $\tilde{\eta}_{FOV} = 20^\circ$, it can scan at $\pm 15^\circ$ off the nadir direction, and it operates at an altitude of 160 ± 30 km. The satellite has a ballistic coefficient equal to 100 kg/m^2 .

To design our specialized orbit, we must 1) analyze the mission to determine the initial orbital elements, 2) correct the orbital elements to achieve a repeat-groundtrack orbit, and 3) plan stationkeeping maneuvers to maintain the desired orbit within altitude restrictions over the site. The first step develops the initial orbital elements to place the satellite in a two-body orbit with perigee over the target. Then, we'll adjust the elements to achieve a repeat-groundtrack orbit accounting for the dominant secular perturbations. Finally, the effects of the perturbations are analyzed and a maneuver schedule is developed that allows us to maintain the orbital altitude over the target. The sum of each maneuver determines how much fuel the mission requires.

Developing the Initial Elements

We need an orbit that meets the constraints using only two-body dynamics. The critical limits on the camera, $160 \text{ km} \pm 30 \text{ km}$ (closer is better but we don't want to reenter), require us to use an oblate Earth. A spherical model would cause altitude errors of about 15 to 20 km for extreme northern or southern latitudes and errors of 5 to 15 km for mid-latitude targets. This amount of error exceeds the operating limits of our hypothetical camera. With the target latitude, ϕ_{tgt} , and longitude, λ_{tgt} , we must find the Earth's radius at the target latitude. Figure 10-15 shows the general geometry.

We find the site radius's magnitude using Eq. (1-16). The radius of perigee is found knowing the radius of the site (on an oblate Earth) and the best camera altitude, $alt_{perigee}$. We'll use an "o" subscript for the initial values because they'll change later.

$$r_p = r_{site} + alt_{perigee} = a_o(1.0 - e_o) \quad (10-34)$$

Finding the semimajor axis relies on two factors: the desired Earth crossing points, and the repeat period for the ground track. We start by examining the sensor characteristics and the designed operating altitude for the sensor. For example, if a satellite can see 50 km on either side of the satellite on each pass, we may use 100 km for the equator crossing value. The circumference of the Earth will yield the number of crossings required for the satellite to complete one "pass". There are multiple answers in these calculations. If a satellite can operate at a higher altitude, it can cover more ground area, and we can usually reduce the repeat time interval. On the other hand, small sensor coverage regions usually mean longer repeat intervals. Some satellites repeat over much larger periods like TOPEX, whose repeat cycle is 127 days.

The mean motion is found by dividing the number of crossings by the repeat interval, $n = k_{cp}/k_{rep}$. In practice, this gets us close to our desired operating altitude, but we can iterate with these equations until the precise values are obtained. Now, let's find the semimajor axis of the orbit by using Eq. (2-19).

In some cases (Example 10-3), we want to place the satellite directly over the target at orbit insertion. This requires us to adjust the eccentricity once a and r_p are found.

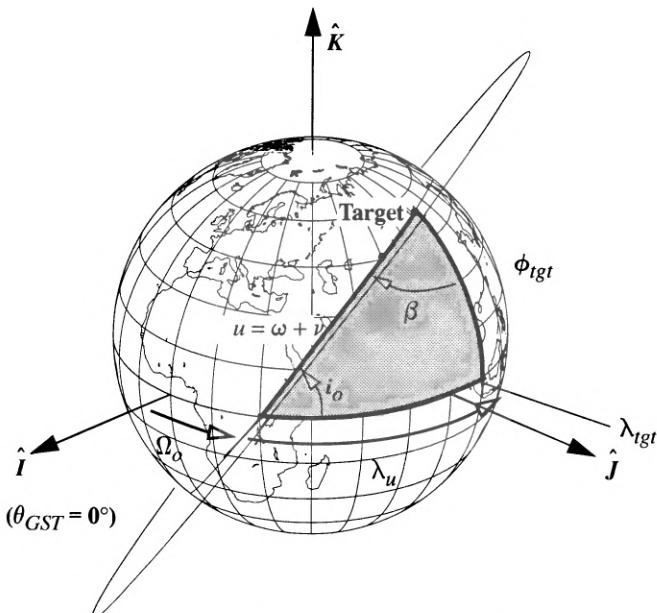


Figure 10-15. Geometry for Injecting a Satellite into Orbit. The inclination is a critical parameter in determining a strategy for satellite injection. Ω depends on the time (GST) and longitude (λ_u) from the initial placement (target), whereas the remaining orbital elements arise from mission constraints. The argument of latitude ($u = \omega + \nu$) permits generality for orbit types. We assume the true anomaly is zero at epoch. The azimuth, β , of the satellite from the target is measured clockwise from north. It's equal to 180° plus the angle shown.

$$e_o = \frac{a_o - r_p}{a_o} \quad (10-35)$$

The inclination is selected considering target latitude, sensitivity of the secular rates for ω and Ω , fuel efficiency, and the available inclinations for direct launches from the launch sites. Given the target latitude, ϕ_{tgt} , inclination, i , must be

$$\phi_{tgt} < i_o < 180.0^\circ - \phi_{tgt} \quad (10-36)$$

so the satellite will pass over the target. Anticipating the effects of major perturbations, we might choose 63.4° or 90.0° for inclination, where $\dot{\omega}_{sec}$ or $\dot{\Omega}_{sec}$, respectively, is zero. Fuel efficiency dictates a launch in as easterly a direction as possible to take advantage of the Earth's rotation. For this reason, retrograde orbits are often avoided unless required by some other operational considerations (such as needing the satellite to be Sun-synchronous). The final consideration is to select from a range of inclinations (Sec. 5.4.1 in Chap.

5) that permit a direct launch from the launch site. If we choose an inclination outside these ranges, we'll have to do a plane change after injection, which requires a lot of fuel.

The longitude of the ascending node, Ω , is defined by λ_u , λ_{tgt} , and Greenwich Sidereal Time, θ_{GST} , as shown in Fig. 10-15. The first step in solving for Ω is to use equations from the launch-window problem (Sec. 5.4.1) to solve for the satellite's azimuth [β , Eq. (5-11)] as it passes over the target and to find the angle between the node and the target's longitude meridian [λ_u , Eq. (5-15)].

The next step is to solve for θ_{GST} , given the start time of the mission. Algorithm 1 determines the GST and LST , but be sure to use Algorithm 2 to determine the correct JD input for the routine. From Fig. 10-15, $LST = \Omega + \lambda_u$. Finally, solve for Ω :

$$\Omega_o = \theta_{GST} + \lambda_{tgt} - \lambda_u \quad (10-37)$$

The argument of perigee is defined by the target latitude and orbit inclination, if we assume perigee is above the target. Figure 10-15 depicts the geometry necessary to calculate the argument of perigee. Using the sine law from spherical trigonometry [Eq. (C-25)] gives us $\text{SIN}(\omega_o) / \text{SIN}(90^\circ) = \text{SIN}(\phi_{tgt}) / \text{SIN}(i_o)$. Rearranging,

$$\text{SIN}(\omega_o) = \frac{\text{SIN}(\phi_{tgt})}{\text{SIN}(i_o)} \quad (10-38)$$

The last parameter, mean anomaly, is chosen so the satellite starts at perigee over the target at the epoch time, so $M_o = 0.0^\circ$.

▼ Example 10-3. Finding Preliminary Orbital Elements.

Given: Roman Colosseum, Rome, Italy, $\phi_{tgt} = 41.52^\circ\text{N}$, $\lambda_{tgt} = 12.37^\circ\text{E}$, $Site_{alt} = 0.152\text{ km}$, $alt_{perigee} = 160\text{ km} \pm 30\text{ km}$, 12:00 UT , on April 2, 2000
Scanning sensor with $\tilde{\eta}_{FOV} = 20^\circ$ and $\eta_{center} = \pm 15^\circ$

Find: Preliminary orbital elements

Begin by finding the ellipsoidal Earth's coordinates. From the beginning of Algorithm 45,

$$r_\delta = 0.749\,765\text{ ER}, r_K = 0.659\,621\text{ ER} \text{ and } |r_{site}| = \sqrt{r_\delta^2 + r_K^2} = 0.998\,557\,5\text{ ER} = 6368.936\text{ km}$$

This allows you to find the radius of perigee:

$$r_p = 6368.936\text{ km} + 160\text{ km} = 6528.936\text{ km} = 1.023\,643\text{ ER}$$

You can now solve either the global Earth observation problem, or the problem of observing a particular site (this problem). I'll introduce both processes because they unite several methods from this chapter. For the global observation problem, begin by calculating the limiting field of view. With a perigee height of 160 km and an $\tilde{\eta}_{FOV} = 20^\circ$, a scanning sensor can see a total angle of $10^\circ \pm 15^\circ$ (remember that the sensor sees half $\tilde{\eta}_{FOV}$ about the center angle). We can then use Eq. (10-10), Eq. (10-11), and Eq. (10-13) to find the maximum amount of drift before the Colosseum will be out of view. Notice we don't use the radius of perigee above a specific site for the global problem; $r_{sat} = 6538.14\text{ km} = 1.025\,086\text{ ER}$.

$$\text{SIN}(\gamma) = \frac{r_{sat}\text{SIN}(\eta)}{R_{\oplus}} = \frac{(160 + 6378.1363)\text{ SIN}(25^\circ)}{6378.1363}, \gamma = 25.672\,079^\circ, 154.327\,920^\circ$$

You use the larger angle because the triangle is oblique and γ must be larger than 90° . To find the slant range, use the same ($10^\circ \pm 15^\circ$) FOV angle. As a check, be sure it's just slightly larger than the orbital altitude of 160 km, and that it's within the limits of our sensor (160 ± 30 km).

$$\rho = R_{\oplus} \cos(\gamma) + r_{sat} \cos(\eta) = \cos(154.327920) + 1.0250857 \cos(25^\circ)$$

$$\rho = 0.027754 \text{ ER} = 177.024616 \text{ km}$$

Now find the maximum ground range from the satellite subpoint to the edge of the area that's visible to the sensor. Remember you can convert the angle to a distance with Eq. (10-1) or Eq. (C-22).

$$\sin(\Lambda_{max}) = \frac{\rho \sin(\eta)}{R_{\oplus}} = \frac{177.024616 \sin(25^\circ)}{6378.1363}, \Lambda_{max} = 0.672066^\circ = 74.814 \text{ km}$$

You can now determine an acceptable equator crossing value. A first estimate could double the maximum ground range (149 km), but we are free to adjust this value based on mission constraints. Let's use about 243 km. Using the circumference of the Earth, find the number of crossing points— $k_{cp} = 2\pi R_{\oplus} / 243 = 164.918$ crossings. Rounding to the nearest integer, you'll repeat in one less, 164. Now choose the length of time to repeat the groundtrack. This process is iterative depending on your sensor characteristics. For instance, let's choose a 10 day repeat period (k_{rep}) that gives us $n_o = 164 / 10 = 16.4 \text{ rev/day}$. You could also choose a 7 day repeat, but the sensor would have to cover 346 km on each pass to maintain approximately the same perigee radius. This may not be within the sensor capabilities. Another value is 480 km in 5 days. Using 10 days, find the semimajor axis from the mean motion. Don't forget to convert the units of the mean motion to canonical units, $n_o = 0.9622364 \text{ rad/TU}$.

$$a_o = \left(\mu \left(\frac{1}{n_o} \right)^2 \right)^{1/3} = \left(1 \left(\frac{1}{(16.4)(2\pi)0.009338092} \right)^2 \right)^{1/3} = 1.025996 \text{ ER} = 6543.94 \text{ km}$$

Now find the eccentricity to make our perigee height exactly 160 km—it should be small.

$$e_o = (a_o - r_p) / a_o = (1.025996 - 1.023643) / 1.025996 = 0.0022933$$

When you design the orbit to observe a single location (this problem), you use the radius of perigee above the site and a pre-determined eccentricity. We'll choose $e_o = 0.001$, which lets us find a_o and n_o .

$$a_o = r_p / (1 - e_o) = 1.023643 / (1 - 0.001) = 1.024668 \text{ ER} = 6535.4713 \text{ km}$$

$$n_o = \sqrt{\frac{\mu}{a_o^3}} = \sqrt{\frac{1}{1.024668^3}} = 0.964107197 \text{ rad/TU}$$

To continue, you need to select an inclination. Eq. (10-38) allows you to choose either the inclination or the argument of perigee. For this example, let's arbitrarily choose $i_o = 45.0^\circ$, which means you can get a direct launch from most space centers. Remember, for a direct launch, the target latitude must be less than or equal to the orbit's inclination to avoid costly plane-change maneuvers.

At this point, you must iterate using concepts from analyzing resonance in Chap. 8. You can vary the eccentricity somewhat to yield a small range of a values, however, e cannot vanish because J_2 perturbations impose a minimum value for the mean e of about 0.001. Notice that the eccentricity we found for the global case was slightly larger. For now, let's assume $e_o = 0.001$. You can use this value, with the perigee radius to find the semimajor axis, and a new value of the mean motion. Accurate specification of a repeat groundtrack requires us to evaluate the resonance condition from page 598. Using $M = n + M_o$,

$$\frac{n + \dot{\omega} + \dot{M}_o}{\omega_{\oplus} - \dot{\Omega}} \approx \frac{P}{Q}$$

The initial guess is chosen using two-body motion. Candidate values of P and Q are found using this ratio (n/ω_{\oplus}) in a continued fraction. Using $n/\omega_{\oplus} = 0.964107 / 0.058833 = 16.387017 \text{ rev/sidereal day}$, the continued fraction is found by successive subtraction and reciprocal operations. It is

$$\frac{n}{\omega_{\oplus}} = 16.387017 = 16 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{2 + \dots}}}}$$

You can examine successive convergents to obtain the potential pairs of P and Q . You select a set of P and Q and refine the mean motion and semimajor axis using an alternate form of the resonance condition above. You can also use the relation of the anomalistic and nodal periods to find the mean motion. I've shown a simplified formula using Eq. (10-20), Eq. (10-21), and Eq. (10-22).

$$\frac{Q_i}{P_i} = \frac{\theta_{\theta_G}}{\theta_{\Omega}} \text{ and } n = \frac{P_i}{Q_i} \left\{ \frac{2\pi}{\theta_{\theta_G}} \right\} \frac{\theta_{\Omega}}{\theta_K} = \frac{P_i}{Q_i} (\omega_{\oplus} - \dot{\Omega}) \left\{ 1 - \frac{3J_2}{2} \left(\frac{R_{\oplus}}{a_o} \right)^2 \left(3 - 4 \sin^2(i_o) \right) \right\}$$

Now evaluate this expression to find the semimajor axis and eccentricity. Using the equations on page 600-601 to find P_i and Q_i , and the initial values $a_o = 6535.4713$ km and $e_o = 0.001$, the sequence of convergents give us

P_i	Q_i	Iteration	θ_{θ_G} TU	θ_{Ω} TU	θ_K TU	n_o rad/TU	a_o km	e_o
16	1	0	104.915 552	6.507 022	6.517 102	0.956 726	6569.0409	0.006 105
		1	104.948 259	6.557 325	6.567 380	0.956 443	6570.3375	0.006 301
33	2	0	104.915 552	6.507 022	6.517 102	0.986 624	6435.6536	invalid
49	3	0	104.915 552	6.507 022	6.517 102	0.976 658	6479.3595	invalid
82	5	0	104.915 552	6.507 022	6.517 102	0.980 645	6461.7883	invalid
213	13	0	104.915 552	6.507 022	6.517 102	0.979 725	6465.8353	invalid

Because a_o , e_o , and n_o are related, we can iterate these equations, but you seldom need more than one iteration. I've shown one iteration on the first case because it's the only valid combination for this problem. The other cases require semimajor axis values less than the perigee radius. The only way to satisfy the other repeat cycles is to move perigee from over the target, but this affects our original problem.

To get Ω_o , first find the intermediate values:

$$\sin(\beta) = \frac{\cos(i_o)}{\cos(\phi_{tgt})} = \frac{\cos(45^\circ)}{\cos(41.52^\circ)}, \beta = 70.8068^\circ$$

$$\cos(\lambda_u) = \frac{\cos(\beta)}{\sin(i_o)} = \frac{\cos(70.8068)}{\sin(45.0^\circ)}, \lambda_u = 62.2943^\circ$$

Next, find the Greenwich Sidereal Time from Algorithm 1. The $JD = 2,451,637.0$.

$$LSTIME(JD, UT, \lambda \Rightarrow \theta_{LST}, \theta_{GST}), \theta_{GST} = 11.140 176^\circ$$

$$\Omega_o = \theta_{GST} - \lambda_u + \lambda_{tgt} = -38.784 170^\circ$$

$$\sin(\omega_o) = \frac{\sin(\phi_{tgt})}{\sin(i_o)} = \frac{\sin(41.52^\circ)}{\sin(45.0^\circ)}, \omega_o = 69.628 610^\circ$$

Finally, set $M_o = 0^\circ$ so the satellite starts over the target. The preliminary values are

$$a_o = 1.030 134 \text{ ER} = 6570.3375 \text{ km}, e_o = 0.006 301, i_o = 45.0^\circ$$

$$\Omega_o = -38.784 170^\circ, \omega_o = 69.628 610^\circ, M_o = 0.000^\circ$$

Remember that this orbit is designed for a particular Earth equator crossing distance. It will repeat in 10 days but perturbations will change the orientation, so you'll have to make some adjustments.



Maintaining the Repeat Groundtrack

Our next task is to adjust the orbital elements to set up a repeat-groundtrack orbit. Remember that this process will yield orbital elements that account for certain perturbations, but we'll have to perform periodic maneuvers to adjust the orbit.

▼ **Example 10-4. Maintaining the Repeat-Groundtrack Orbit.**

Continuing with Example 10-3, find the semiparameter and the rate of nodal regression from J_2 .

$$p = a(1 - e^2) = 1.030\ 134(1 - 0.006\ 301^2) = 1.030\ 093 \text{ ER}$$

$$\dot{\Omega} = -\frac{3nR_{\oplus}^2 J_2}{2p^2} \cos(i) = -0.001\ 035 \text{ rad/TU} = -6.350\ 776 \text{ %/day}$$

You will require several partial derivatives to find the solution.

$$\frac{\partial \rho_{\Omega}}{\partial a} = \frac{3\pi}{na} \left[1 + \frac{1}{2} J_2 \left(\frac{R_{\oplus}}{a} \right) (4 \cos^2(i) - 1) \right] = 9.570\ 755 \text{ TU/ER}$$

$$\frac{\partial \dot{\Omega}}{\partial a} = -\frac{7\dot{\Omega}}{2a} = 0.003\ 517 \text{ rad/(TU·ER)}$$

$$\frac{\partial \rho_{\Omega}}{\partial i} = \frac{12\pi}{n} J_2 \left(\frac{R_{\oplus}}{a} \right) \sin(2i) = 0.041\ 426 \text{ TU}$$

$$\frac{\partial \dot{\Omega}}{\partial i} = -\dot{\Omega} \tan(i) = 0.001\ 035 \text{ rad/TU}$$

Find the anomalistic period.

$$\rho_{\Omega} \equiv \frac{2\pi}{n} \left[1 - \frac{3J_2}{2} \left(\frac{R_{\oplus}}{a} \right)^2 (3 - 4\sin^2(i)) \right] = 6.559\ 270 \text{ TU} = 88.201\ 531 \text{ min}$$

The groundtrack shift should be approximately equal to $2\pi R_{\oplus} k_{rep} / k_{cp}$.

$$\lambda_S = (\omega_{\oplus} - \dot{\Omega}) \rho_{\Omega} = (0.058\ 833\ 599 - 0.001\ 035) 6.559\ 270 = 0.392\ 695 \Rightarrow 2504.6604 \text{ km}$$

The partial derivatives give you the sensitivity of λ_S to changes in a and i .

$$\frac{\partial \lambda_S}{\partial a} = R_{\oplus} (\omega_{\oplus} - \dot{\Omega}) \frac{\partial \rho_{\Omega}}{\partial a} - \frac{\partial \dot{\Omega}}{\partial a} R_{\oplus} \rho_{\Omega} = 0.549\ 921$$

$$\frac{\partial \lambda_S}{\partial i} = R_{\oplus} (\omega_{\oplus} - \dot{\Omega}) \frac{\partial \rho_{\Omega}}{\partial i} - \frac{\partial \dot{\Omega}}{\partial i} R_{\oplus} \rho_{\Omega} = -0.004\ 309$$

Next you need an estimate of the rate of change for the semimajor axis and the inclination. You can get da/dt and di/dt from numerical or analytical techniques (see Chap. 7 and Chap. 8). Let's use the simplified numerical technique in Chap. 7.

$$da/dt = -0.174\ 562 \text{ km/rev} = -0.000\ 027\ 4 \text{ rad/TU}$$

$$di/dt = 0.000\ 132 \text{ %rev} = 0.000\ 002\ 3 \text{ rad/TU}$$

The two k -parameters are found using tolerances that are chosen for the maximum allowable groundtrack drift. Let λ_{Snom} vary by about ± 50 km. This puts a 100 km tolerance band about the nominal groundtrack shift parameter, λ_{Snom} . Even if the satellite drifts this far, the sensor will still be able to see the Colosseum ($\Lambda_{max} = \pm 74$ km).

$$k_2 = \frac{1}{\rho_{\Omega}} \left(\frac{\partial \lambda_S}{\partial a} \frac{da}{dt} + \frac{\partial \lambda_S}{\partial i} \frac{di}{dt} \right) = -0.000\ 002\ 296 \text{ ER/TU}^2$$

$$k_1 = \sqrt{2k_2(\lambda_{S+} - \lambda_{S-})} = \sqrt{2(-0.000\ 002\ 296)} \frac{(-50 - 50)}{6378.1363} = 0.000\ 268\ 3 \text{ ER/TU}$$

This allows you to find the correction to the semimajor axis.

$$\Delta a = k_1 \rho_{\Omega} \left(\frac{\partial \lambda_S}{\partial a} \right)^{-1} = 0.000\ 268\ 3 (6.559\ 270)(1/0.549\ 921) = 0.003\ 200 \text{ ER} = 20.413 \text{ km}$$

Finally, you must find out how often to maneuver, and the magnitude of each burn.

$$t_{drift} = -\frac{k_1}{k_2} = -\frac{0.000\ 268\ 3}{-0.000\ 002\ 296} = 116.862\ 6 \text{ TU} = 1571.433\ 9 \text{ min}$$

$$\Delta v = \frac{n}{a} \Delta a = \frac{0.956\ 443}{1.030\ 134} 0.003\ 200 = 0.002\ 972 \text{ ER/TU} = 23.4911 \text{ m/s}$$

Remember that you let the satellite drift for twice the drift time, so you'll have to perform a maneuver every 3142 min, and it will be about 23 m/s. This is a large maneuver, and you must correct the orbit  about every other day—a consequence of choosing a low perigee height.

Maintaining Minimum Altitude Variations

The corrections to maintain a repeat groundtrack require us to periodically adjust the semimajor axis. Unfortunately, every time we change the semimajor axis, the location of perigee changes. In addition, perturbations also cause the perigee location to change. The requirement for the camera to be 160 km ± 30 km over the target is soon violated by perigee drift unless the orbit is very nearly circular. Even here, semimajor axis and eccentricity variations nullify the initial analysis over time.

Algorithm 64 shows the procedure to maintain a small change in the location of perigee. There were three main factors: global variations in altitude based on satellite height, changes in e from drag and central-body perturbations, and the change in the perigee location. We won't consider the global variations about the orbit because we're only interested in the location over the target at perigee. We also won't consider the changes in e , but realize that the zonal harmonics can cause fluctuations in e . In addition, long-periodic effects will change e and very small changes in e can result in very large variations in altitude. If more detailed analyses suggest these variations become too large, we may need to explore a frozen-eccentricity orbit design.

Example 10-5. Maintaining Minimum Altitude Variations.

GIVEN: The results of Example 10-4.

FIND: Corrections to minimize altitude variations.

The first step is to find the secular perturbations resulting from J_2 , $\dot{\omega} = 6.736\ 015^\circ/\text{day} = 0.001\ 098 \text{ rad/TU}$. The maximum change in distance is ± 30 km. Thus the maximum allowable apsidal drift is

$$\Delta\omega_{max} = \sqrt{2 \frac{\Delta r}{ae} \left(\frac{1+e}{1-e} \right)} = \sqrt{2 \frac{30}{(6570)(0.006\ 301)} \left(\frac{1+0.006\ 301}{1-0.006\ 301} \right)} = 1.211\ 444 = 69.410\ 675^\circ$$

The amount of time until this condition occurs is

$$k_{max} = \left| \frac{\Delta\omega_{max}}{\dot{\omega}} \right| = \left| \frac{1.211\ 444}{0.001\ 098} \right| = 1103.482 \text{ TU} = 14,838.353 \text{ min} = 10.304\ 412 \text{ days}$$

Now find the change in velocity.

$$\Delta v = \frac{nae}{2} \Delta\omega_{max} = \frac{0.956\ 443 (1.030\ 134) 0.006\ 301}{2} 1.211\ 444$$

$$\Delta v = 0.003\ 761 \text{ ER/TU} = 29.729 \text{ m/s}$$

 Notice the long time delay between maneuvers. In addition, each burn is rather small. We could further reduce these maneuvers by incorporating frozen-eccentricity designs.

The preceding three examples introduce some of the considerations necessary for designing a specialized orbit. In practice, these calculations are a first step to determining actual orbital elements. Notice how we can combine the different types of orbits depending on our mission requirements. The examples didn't explore Sun-synchronous and frozen orbits. A Sun-synchronous constraint would give us approximately equal solar lighting conditions on a desired pass interval, while the frozen eccentricity could be difficult to achieve with the repeat groundtrack and minimum altitude variations constraints.

10.5 Navigation—the Global Positioning System

Humans have needed accurate navigation systems for thousands of years. Much of the ancient work in astronomy was aimed at improving navigation at sea. An interesting series of events occurred when the British Parliament offered £20,000 for anyone who could determine longitude within about $\frac{1}{4}^{\circ}$ after a six-month voyage at sea. John Harrison (1693–1776) eventually won the prize with his invention, the marine chronometer. He proved the accuracy of his system in 1761, and although he wasn't paid in full until 1773, his clock worked superbly (determine longitude to about $1\frac{1}{4}^{\circ}$). This instrument permitted vastly improved astronomic observations, and modern versions are still used today. This section provides a broad overview and description of the *Navigation System with Timing And Ranging* (NAVSTAR), which is better known as the *Global Positioning System* (GPS). Although GPS is a constellation of 24 satellites, we won't dwell on issues of constellation design. Modern navigation satellites mark a tremendous technological advance that promises to revolutionize the way we think about space. For instance, accurate positions from GPS signals are helping farmers determine the best places to put fertilizer on million-acre ranches.

10.5.1 Historical Background

Although the history of navigation reaches back before the Christian era, the first practical radio-navigation systems relied on radio transmitters located on the ground to broadcast modulated carrier waves to distant receivers. Those early ground-based systems, most of which were installed on the European continent during World War II, provided reasonably accurate position fixes for Allied and Axis military forces, but they pale compared to today's capabilities in space.

The underlying concept for satellite aids to navigation dates back to the 1950s and 1960s with the Navy's TIMATION Program and the Air Force's 621B Project. TIMATION developed a practical way to track satellites, and 621B provided ground-based radio transmitters used by pilots for triangulation. The first successful constellation of space-based radio transmitters for navigation, the Transit Satnav System, came from Johns Hopkins University in the early 1960s and used satellites produced by RCA. The U.S. Department of Defense financed this low-altitude constellation (1070 km polar orbits) but encouraged civilian use in 1973. Soon military users were outnumbered by more than 30 to 1. The system used single, two-dimensional fixes to achieve accuracies of about 500 m.

This accuracy was soon inadequate, so a new system was planned to provide accuracies under 10 m. These relatively demanding requirements drove the engineering design of the GPS satellites, which required a new way of looking at navigation, communication, and time transfer. In particular, we know from Einstein's theories that any two clocks which move with respect to each other, or are at different points in a gravitational field, will appear to tick at different rates. Although this effect is small, the highly accurate tolerances for GPS require us to correct for these clock dilations.

Implications of General and Special Relativity

Before we explore GPS, it's instructive to see how it works and why it has certain constraints. To do so, we must introduce the concepts of *special relativity* and *general relativity*. Let's briefly explore why.

In 1818, Karl Friedrich Gauss drafted a proposal to his government to do a geodetic survey of the Kingdom of Hanover (Buhler, 1987, 81). For over a decade, Gauss and his colleague Schumacher constructed a *triangulation* of this two-dimensional part of the Earth's surface. Although the military and economic benefits of such maps were obvious, Gauss must have been intrigued by measuring the Earth's true shape because he devoted a lot of time to the project. The technique of triangulation is simple in principle but cumbersome in practice. Starting from a known baseline, one constructs a network of triangles. In Gauss's case, trees were cut to open lines of sight. Trigonometry is then used to construct the map and to calculate the nonuniform curvature of the two-dimensional geometry.

Today, the triangles or two-dimensional simplexes of ordinary geodesy that Gauss used are replaced by their four-dimensional counterpart, or *4-simplex*. Figure 10-16 shows the situation. A 4-simplex (four-dimensional triangle) has five vertices. Four of these vertices are the GPS satellites; the remaining vertex is the observer's receiving antenna. Although satellites form a spatial tetrahedron (six solid lines interconnecting the four satellites obtained by cross-link ranging or modeling), the observer is connected to the satellites by just four radio signals. These signals provide information about the observer's position and velocity relative to the satellite in space-time. In other words, the observer can triangulate using the base tetrahedron formed by the four satellites.

You're probably wondering why we're introducing curved space-time geometry and four-dimensional triangles. Why was our familiar three-dimensional paradigm for navigation and communication replaced by these new model concepts? The answer lies in precise tracking of time—a problem we've seen throughout this book. Remember that time is a major factor in determining the position error for a satellite and is responsible for the along-track variations, which usually exceed cross-track and radial displacements.

Precise global synchronizing of clocks is integral to GPS operations. The GPS satellites' atomic clocks are moving fast enough with respect to the clocks on Earth and to users' clocks that they require corrections for the special relativistic effects on their synchronization. At the same time, the orbital radii of the GPS satellites are large enough to cause a difference between the gravitational potentials of clocks on the satellites and on the ground. These differences produce gravitational effects on the clock synchronization

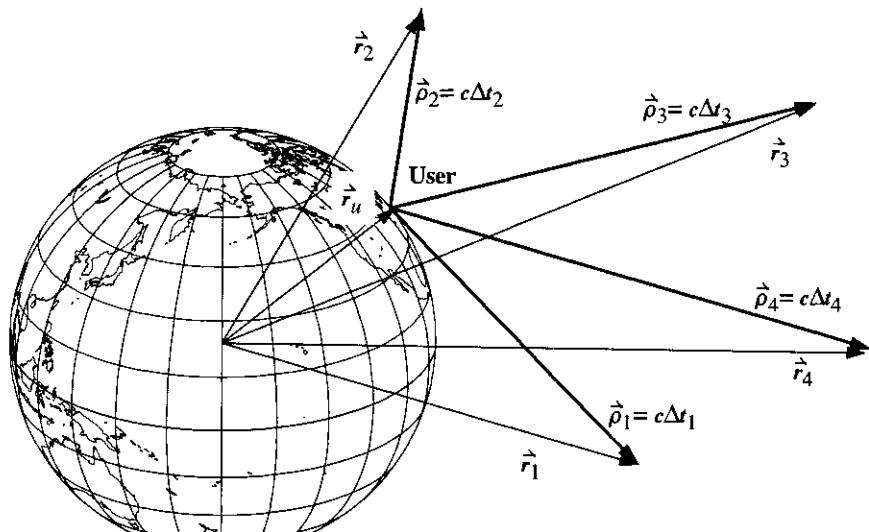


Figure 10-16. Navigating with the GPS—the 4-Dimensional Simplex. A GPS receiver starts its navigation solution by measuring four pseudoranges to four widely dispersed GPS satellites (ρ_i , $i = 1 \dots 4$). Then it substitutes these four measured pseudoranges into a system of four equations in four unknowns. The first three unknowns are the receiver's three mutually orthogonal position coordinates. The fourth unknown is the clock offset error.

roughly equal to the special relativistic effects. We can consistently treat both effects by using Einstein's 1915 theory of general relativity.

Shortly after Einstein introduced his theory of special relativity, Herman Minkowski (1864–1909) represented it geometrically. Minkowski's flat space-time model of special relativity emphasized the unity of space and time. In 1908, he spoke at the 80th assembly of the German Natural Scientists and Physicians in Cologne. His words encapsulated what Einstein's theory means to us: "Henceforth space by itself, and time by itself are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality."

The *theory of special relativity* unifies time with space. Here we use the speed of light ($c = 3 \times 10^{10}$ cm/s) to equate time with distance. A clock accuracy of one nanosecond translates to approximately 33 cm because that's the distance light travels in one nanosecond.

Einstein described a curved space-time theory of gravity (*general relativity*) that geometrically linked mass-energy and curved space-time geometry. Imagine a steel ball placed on a rubber membrane. The steel ball causes the rubber membrane to curve and distort under its weight. Just as the heavy ball causes the membrane to curve under its weight, the Earth causes a dimple in space-time. If we place a small marble on this curved surface and give it a tangential kick, it "orbits" around the steel ball, spiraling as if in orbit. Just as

the curvature of the membrane causes the marble to orbit the ball, the curvature of space causes satellites (e.g., the GPS satellites) to orbit the Earth.

We can approximate the curved space-time geometry enveloping the Earth by a **Schwarzschild geometry**. This technique is named after Karl Schwarzschild (1873–1916), who developed the concept in 1916, just months after Einstein formed his general theory of relativity. It's an exact solution of Einstein's equations for an approximately spherical mass such as the Earth. In this geometry the rate of a clock such as a GPS clock ($\Delta\tau_{GPS}$) is slower than that of an identical clock ($\Delta\tau$) placed far away. Using the position and velocity magnitudes for the GPS satellites,

$$\Delta\tau_{GPS} = \left\{ 1 - \left(\frac{M_\oplus}{r_{GPS}} - \frac{v_{GPS}^2}{2} \right) \right\} \Delta\tau \quad (10-39)$$

The first fraction is the *redshift* correction, and the second is the *second-order Doppler* correction. The magnitudes of both corrections are about the same. Combining these correction terms offsets the frequency approximately one part in 4×10^9 . Therefore, relativistic corrections to GPS are on the order of nanoseconds and accumulate over time. Every nanosecond error in time introduces a 33 cm error. For this reason, the GPS clock frequencies are physically offset from 10.23 MHz to 10.229 999 995 45 MHz to match the differential redshift effect between the Earth-bound receiver and the satellite. More sophisticated analysis shows that other relativistic effects don't accumulate enough over a few hundred years to warrant incorporation into the GPS system. Consult Wheeler (1990), Logsdon (1992), or Hoffmann-Wellenhof et al. (1994) for further information.

10.5.2 System Introduction

The Global Positioning System has three segments: the *space segment*, the *control segment*, and the *user segment*. The space segment consists of 24 satellites in six orbital planes, with four satellites in each plane. Each GPS satellite is a space-based satellite for radio navigation which moves in a semisynchronous (12-hour) orbit. The nearly circular orbits provide global coverage, which allows real-time position, velocity, and timing services for low-altitude satellites and terrestrial users. It's sometimes called the *Navstar* system, but we'll simply use GPS.

The orbital slots are indicated by a letter for the orbital plane (A–F) and a number for the position within the plane (1–4). Therefore, the GPS satellites are in orbital slots ranging from A1 to F4. GPS satellites fall into five classes: Block I, Block II, Block IIA, Block IIR, and Block IIF. Figure 10-17 represents a snapshot in time to show how the GPS satellites are positioned in the constellation.

Block I satellites were first launched in 1978. The original constellation was at 63° inclination, but all Block I satellites have now been replaced. The Block II satellites were the first to occupy orbits at 55° inclination. These satellites were launched beginning in 1982, about the same time as the former Soviet Union's Glonass navigation system. Some Block II satellites had mechanisms to keep unauthorized users from accessing some sig-

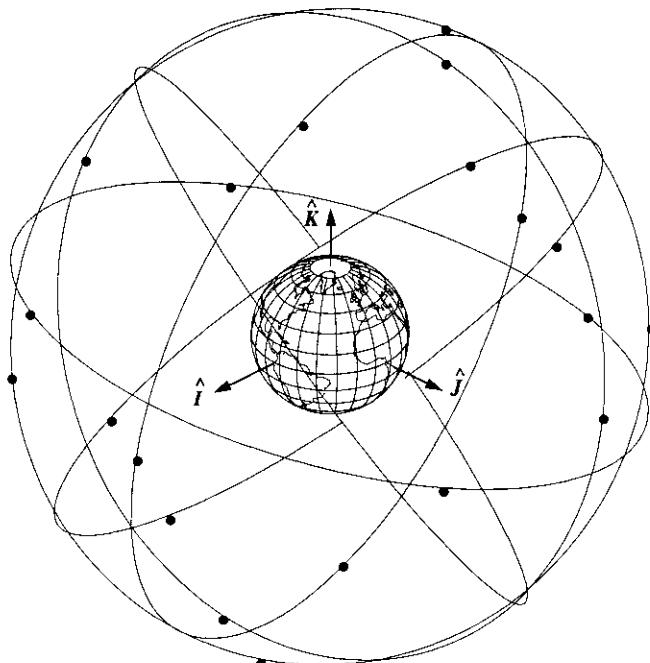


Figure 10-17. The GPS Constellation of 24 Satellites. The 24 satellites in the GPS constellation travel around the Earth in 12-hour circular orbits with $i = 55^\circ$.

nals. The Block IIA (advanced) satellites talk to each other, so the constellation is more autonomous. Some Block IIA satellites have laser-ranging retro-reflectors in order to obtain very precise orbits. The U.S. Department of Defense officially declared Initial Operational Capability on December 8, 1993, when 24 GPS satellites—Block I, Block II, and Block IIA—were operational and available for navigation.

Block IIR (replenishment/replacement) satellites feature hydrogen-maser atomic clocks instead of the cesium and rubidium clocks. These new clocks are ten times more precise. The improved satellite-to-satellite communications, onboard computing, and cross-link ranging to other GPS satellites will improve the constellation's autonomy from ground control. The Block IIF (follow-on) satellites will be launched after the year 2000 and will further improve autonomy over the Block IIR satellites.

The *Operational Control System* that constitutes the control segment of GPS consists of the Master Control facility at the Consolidated Space Operations Center (CSOC) at Falcon Air Force Base in Colorado, five world-wide monitoring stations, and three ground-control stations. The CSOC collects tracking data from the monitor stations and calculates the satellite ephemerides and clock corrections using a Kalman filter. The results are then passed to one of the three ground-control stations for uploading to the satellites. If a ground station can't transmit, the satellites have stored navigation messages, so position-

ing remains accurate. For precise ephemerides, five more sites from the Defense Mapping Agency augment the tracking data from the monitor stations. Nonmilitary tracking networks exist to provide post-processed, precise ephemerides. One example is the International GPS Service (IGPS), with tracking stations co-located with Very Long Baseline Interferometry (VLBI) sites.

The user segment consists of GPS receivers and the user community. The GPS receivers convert the satellites' signals into position, velocity, and time estimates for navigating, positioning, distributing time, or geodesy. Four satellites are required to solve for the position and the time at any instant. I'll discuss other applications in the applications section.

10.5.3 Signals from GPS Satellites

Each of the 24 satellites in the GPS constellation transmits its navigation signals on the same two L-band frequencies, called L_1 (1575.42 MHz) and L_2 (1227.60 MHz). Two *pseudorandom noise* PRN-codes, the C/A-Code and the P-Code, are superimposed on these carrier bands. These codes provide—along with satellite ephemerides (broadcast ephemerides)—ionospheric modeling coefficients, status information, system time, and satellite-clock corrections. The measured travel times of the signals from the satellites to the receivers are used to compute the pseudorange. Pseudorange is one of three *observables*, or the actual parameters we can detect with a receiver. We'll discuss pseudorange, the carrier phase, and linearly combined observables later. The atomic clocks on each satellite produce the fundamental L-band frequency, 10.23 MHz. The L_1 and L_2 carrier frequencies are generated by multiplying the fundamental frequency by 154 and 120, respectively.

The *Coarse-Acquisition code*, C/A, sometimes called the *Standard Positioning Service*, is a pseudorandom noise code that is modulated onto the L_1 carrier. The wavelength of the C/A signal is approximately 300 m. Because initial tests using the C/A code resulted in better-than-expected positions, the U.S. Department of Defense introduced *Selective Availability* (SA), to deny full system accuracy to unauthorized users. This technique intentionally corrupts the GPS satellites' clocks and the broadcast ephemerides. Errors are introduced into the fundamental frequency of the GPS clocks. This clock "dithering" affects corrections to the satellites' clocks, as well as the pseudorange observables. Errors are introduced into the broadcast ephemerides by truncating the orbital information in the navigation message. Clock dithering contributes an error signal to the pseudoranges with 50 m amplitude and periods of several minutes. Truncating the broadcast ephemerides contributes an error signal with 50–150 m amplitude and periods of some hours (Hoffman-Wellenhoff, 1994).

The *Precision*, P-code, sometimes called the *Precise Positioning Service*, is modulated onto the L_1 and L_2 carriers. The wavelength of the P-code is about 30 m. The P-code is referred to as the Y-code if it's encrypted. Y-code actually combines the P-code with a W-encryption code and requires a Department of Defense-authorized receiver to decode it. Originally, the encryption was intended as a means to safeguard the signal from being corrupted by interference, jamming, or signals with a falsified GPS signature. Because of the

protection against spoofing, the encryption is referred to as *anti-spoofing (A/S)*, which is either on or off—it's not a variable effect like selective availability.

Pseudorange Observables

A main observable available is the *pseudorange*—the distance between the GPS satellite at some transmit time, t_{trans} , and the receiver at some receive time, t_{rec} . Because the transmit and receive times are different, it's impossible to measure the true range between the satellite and the receiver. Because the GPS signals travel at the speed of light, we measure the pseudorange as the speed of light times the time interval.

$$\rho = c(t_{rec} - t_{trans})$$

The time is indicated with a lower case t to show that it's clock time. We model the clock time as differing from the true time, T , by a bias, α , a drift, β , and a drift rate, γ . Thus, with T_{ref} equal to an external reference at the time α , β , and γ were determined,

$$t = T + \alpha + \beta(T - T_{ref}) + \gamma(T - T_{ref})^2$$

After substituting for the clock time and grouping the clock-correction terms (δt_{rec} and δt_{trans}), the pseudorange becomes

$$\rho = c(T_{rec} - T_{trans}) + c(\delta t_{rec} - \delta t_{trans})$$

We now see the pseudoranges as a function of the “true” pseudorange (the difference between the position vectors of the satellite and the receiver at the true transmit and receive times) and the correction terms.

Let's substitute the magnitude of the true pseudorange using an orthogonal coordinate system with the origin at the center of the Earth (typically Earth-Centered, Earth-Fixed for surveying and point positioning). Now, the model becomes

$$\rho = |\vec{r}_{GPS}(T_{trans}) - \vec{r}_{rec}(T_{rec})| + c(\delta t_{rec} - \delta t_{trans}) \quad (10-40)$$

The satellite coordinates (first term) and clock corrections (last term) are available from the broadcast ephemeris. Thus, four unknowns remain. We can find the position of the receiver and the receiver's clock correction by using four simultaneous pseudoranges at one instant of time. In a more general sense, taking observations over multiple times, the relation between unknowns and observables is $3 + k \leq i_{GPS}$. Here, i_{GPS} is the number of satellites used and k is the number of times at which simultaneous observations are taken. Using more satellites or more observation times leads to an over-determined system. This suggests using a batch least squares routine for differential correction. If we assume we know the receiver's location, we can find the single unknown receiver clock correction from a single observation. We call this *time determination*.

Carrier-Phase Observables

Another observable is based on the *carrier phase* of the signal. It observes the signal frequency and doesn't require any information contained on the signal. The relation between phase, ϕ , and frequency, f , is

$$\phi(T) = fT - \phi_o$$

where ϕ_o is the initial phase resulting from the clock errors ($f\delta t$). To detect the phase, a receiver clock generates a signal with a reference frequency similar to the one being broadcast by the GPS satellite. This signal is then “beat” against the received frequency from the GPS satellite. The phase for the reference frequency, f_r , is

$$\phi_r(T) = f_r T - \phi_{ro}$$

and the phase for the received signal originating at the satellite is

$$\phi_s(T) = f_s T - f_s \frac{\rho_{true}}{c} - \phi_{so}$$

where the phase of the satellite's received signal is actually associated with $T - \rho_{true}/c$, the time of transmission, not with time T , the true time of receipt. By beating the phases together (taking the difference), we obtain an integer number of cycles, N , plus an instantaneous fractional beat phase. Thus,

$$\phi_s - \phi_r = N + \Delta\phi$$

where $\Delta\phi$ is what's actually observed. If we rearrange the previous equations, the phase observable equation becomes

$$N + \Delta\phi = (f_s - f_r) T - \frac{f_s}{c} \rho_{true} + c(\delta t_{rec} - \delta t_{trans})$$

If we assume that $f_s = f_r$ (by design) and recall that $c = \lambda f$, we find that

$$-\lambda\Delta\phi = \rho_{true} + c(\delta t_{rec} - \delta t_{trans}) + \lambda N$$

As a shorthand notation, we'll define $\rho_\phi = -\lambda\Delta\phi$. Note that by comparing ρ_ϕ with Eq. (10-40), we can form $\rho = \rho_\phi + \lambda N$ where we have a distance equal to N cycles of wavelength λ , and a fraction of $\Delta\phi$.

$$\begin{aligned}\rho_\phi &= \rho_{true} + c(\delta t_{rec} - \delta t_{trans}) + \lambda N \\ \rho_\phi &= -\lambda\Delta\phi\end{aligned}\tag{10-41}$$

The integer number of cycles, N , is typically not known and varies for every receiver-satellite combination. Because of the ambiguous nature of N , it's referred to as the **integer ambiguity**. As long as the connection between the receiver and the satellite isn't broken, N remains constant while the fractional beat phase changes over time. We can solve for it by

using the code pseudoranges or estimating it. The loss of signal lock between a GPS satellite and the receiver is referred to as “cycle slip.” If the signal lock is re-established, a new ambiguity will exist. We must solve for it separately from the original ambiguity.

ρ_{true} is the same as it was for the code-based pseudorange. If the satellite’s location and clock correction are known from the broadcast ephemeris or some other source, four unknowns represent the position and clock correction of the receiver, and other unknown ambiguity terms exist for every satellite-receiver combination. When using the phase observables, figuring out how many observations are required becomes tricky. The number of unknowns needs to be less than or equal to the number of observations. Therefore, we must have observables taken at more than one time to find the position and the clock offset and all ambiguity terms. We can express the relation as

$$3 + k + i_{GPS} \leq k i_{GPS}$$

where i_{GPS} is the number of GPS satellites and k is the number of observation times. We need at least two observation times, requiring five satellites. If observations are taken over three times, we need three satellites. When using only three satellites, it’s important to allow ample time between observations to ensure good geometric observability or low *geometric dilution of precision*. We’ll discuss this idea later. Once again, if more observations are available, we could use a differential-correction routine.

Other Effects on Pseudoranges

Other corrections besides the clock offsets will appear as pseudorange biases, including the atmospheric effects of the ionosphere and troposphere. A first-order ionospheric correction may be computed from the linear combination of signals from both L₁ and L₂. I’ll say more about it later. Some models try to account for all effects of the ionosphere but require much effort in modeling the atmosphere’s total electron count, which depends strongly on time. The troposphere has a “dry” part and a “wet” part, depending on the amount of water vapor. The dry part constitutes 90% of the tropospheric refraction, whereas the wet part represents 10%. However, the models for the dry troposphere are more accurate than the models for the wet troposphere. Therefore, the errors in the wet troposphere have a larger effect on the pseudorange bias than the errors in the dry troposphere. The largest effects of the troposphere can be avoided by prescribing an elevation mask for the receiver, thereby avoiding signals from low-elevation satellites. With a 15-degree elevation mask, four to eight satellites will be simultaneously observable at any instant of time.

Receiver-specific biases due to the offset of the antenna’s phase center should be considered as well. The phase center of an antenna is where the signal is essentially received and the point to which the measurement refers. T_{rec} indicates the true time that the signal arrives at the phase center. Thus, we need to know the distance between the antenna’s phase center and a reference marker in order to correctly interpret the results. The phase center varies from antenna to antenna and frequency to frequency. Therefore, different phase centers exist for the L₁ and L₂ signals, as well as for signals constructed from a lin-

ear combination of the L_1 and L_2 frequencies. Manufacturers of receivers typically supply the location of the offset of the antenna phase of various receiver types for each frequency; however, this value will vary between receivers of the same type due to production inconsistencies. Also, the phase center will vary for a single antenna receiving a single frequency because of the azimuth, elevation, and intensity of the signal. This variation of the phase center for an antenna is more important than an incorrect offset which could simply be solved for. Therefore, for highly accurate work, we may need to re-evaluate the location of the antenna-phase offset.

When satellite signals reflect off another surface and then arrive at the receiver simultaneously with nonreflected signals, this is known as **multipath** (the same phenomena that distorts radio and television reception in mountainous regions). Multipath affects code-based pseudoranges more than phase-based pseudoranges. Mathematical models to account for the multipath effect are impractical because the multipath effect depends so much upon the geometry of the situation. Instead of accounting for multipath, we should avoid multipath by placing the antenna as far from reflective objects as possible. If we can't avoid multipath, we can estimate the bias from the pseudoranges free of ionospheric effects or we can remove the effects through digital filtering, wideband antennas, or radio-antenna ground planes or by choosing an antenna that filters based on signal polarization.

We consider special- and general-relativity effects because the reference frames fixed with the GPS satellite and the receiver are accelerated compared to the reference frame fixed in the Earth. The Earth's gravitational field causes perturbations in the satellite's orbits and space-time curvature of the satellite's signal. The acceleration of the reference frames causes perturbations in the fundamental frequency of the satellite and receiver clocks.

Because the ionosphere's first-order effect contributes the most to the pseudorange bias, we'll investigate the correction using two approaches. We can linearly combine the L_1 and L_2 observables to form a new signal that is free of ionospheric effects. Or we can correct one of the signals to remove these effects. Although the derivation for the linearly combined observable is easier to do, in practice the correction method is probably easier to use.

Linearly Combined Observables

We can combine the first two observables to eliminate an undesirable effect, or aid resolution of others, such as ambiguity resolution. One of the more common combinations of signals provides a first-order correction to the effects of the ionosphere. The first-order contribution of the ionosphere to the pseudorange bias is related to the inverse of the frequency squared. Substituting values into this relation yields a wavelength of about 48.5 cm for the ***ionospheric-free signal*** or a frequency of 618.8 MHz ($60.5 f_0$).

There are other linear combinations of the observables besides the one that creates the ionospheric-free observable. Adding the two phases results in the narrow lane, whereas subtracting the two phases results in the wide lane. The signal ambiguity for the wide lane is more easily resolved than the individual ambiguities. Linear combinations of code

pseudoranges and phase pseudoranges can be used to compensate for cycle slips or to smooth pseudoranges over time. Remember that linear combinations of observables will affect the amount of noise in your results. To resolve the ambiguities, you may want to solve for the ambiguity using a combination, and then switch to phase-only processing once the ambiguity is resolved.

Determining Position

If the timing pulses generated by a GPS receiver were always perfectly synchronized with the clocks carried onboard the satellites, ranging measurements to three widely dispersed satellites would be enough to determine the user's position vector, \vec{r}_u . But in almost all cases, they're not, so *four* ranging measurements are required instead of only *three*.

Each of the four separate ranging measurements is corrupted by *clock-bias errors*, $(\delta t_{rec} - \delta t_{trans})$, in each receiver's clock, so we call them "pseudoranges" (ρ_i) to emphasize that they're corrupted values.

Figure 10-16 showed how a GPS receiver measures the four signal travel times and then multiplies each of them by the speed of light to obtain the four necessary pseudoranges, ρ_i . These four pseudorange values are then substituted into this system of four equations in four unknowns. We can regard each of these four equations as the algebraic characterization of the diagonal of a parallelepiped. From Fig. 10-16, we can form the vector relation $\vec{\rho}_i = \vec{r}_i - \vec{r}_u$ where \vec{r}_u is the unknown user location. Thus, with an unknown clock bias, $(\delta t_{rec} - \delta t_{trans})$,

$$(\rho_i - c \cdot (\delta t_{rec} - \delta t_{trans}))^2 = (r_{iI} - r_{uI})^2 + (r_{iJ} - r_{uJ})^2 + (r_{iK} - r_{uK})^2 \quad i = 1 \dots 4 \quad (10-42)$$

The GPS receiver determines the position vectors, \vec{r}_i , of each of the four satellites. It starts by extracting the 16 ephemeris constants from each satellite's 50 bps data stream. Then it uses these ephemeris constants in evaluating a simple set of closed-form equations. The four unknowns, \vec{r}_u and the clock bias, are permanently embedded in the four simultaneous range equations [Eq. (10-42)]. Ordinary mathematical manipulations won't provide an explicit solution with the four unknowns isolated on the left-hand side of each equation. However, a first-order Taylor series expansion, followed by a computer-based mathematical iteration, yields a workable solution that can provide acceptable levels of accuracy for most practical applications.

The range equation defining the distance between the GPS receiver and satellite number one can be written as follows:

$$\rho_1 = \sqrt{(r_{1I} - r_{uI})^2 + (r_{1J} - r_{uJ})^2 + (r_{1K} - r_{uK})^2} + c \cdot (\delta t_{rec} - \delta t_{trans}) \quad (10-43)$$

If we can provide an acceptably accurate estimate of \vec{r}_{uest} and $(\delta t_{rec} - \delta t_{trans})_{est}$ for the four unknown variables, we can expand this equation in a first-order Taylor series. For GPS applications, initial estimates for \vec{r}_u within 100 kilometers will usually do. The first guess for the clock bias is usually 0.0.

$$\begin{aligned}\Delta\rho_1 &= \rho_1 - (r_{ulesp}, r_{ujesp}, r_{ukesp}, \delta t_{recest}) \\ &= \frac{\partial\rho_1}{\partial r_{ul}}\Delta r_I + \frac{\partial\rho_1}{\partial r_{uj}}\Delta r_J + \frac{\partial\rho_1}{\partial r_{uK}}\Delta r_K + \frac{\partial\rho_1}{\partial \delta t_{recest}}\Delta \delta t_{recest} + u_1\end{aligned}\quad (10-44)$$

At this point, u_1 represents the truncation error in the first-order Taylor series expansion. Later, after iteration, u_1 will converge toward the statistical error along the line-of-sight vector between the receiver and satellite number one. Similar series expansions for the other three pseudoranging equations provide additional equations that we solve using the estimation techniques in Chap. 9.

Dilution of Precision

A common term in the GPS community is the ***Dilution of Precision*** (DOP), which expresses the geometry of the GPS satellites relative to the receiver. The farther apart the angles from the receiver to the satellites, the better the positioning accuracy. The angles mark out a volume in space which is related to the inverse of the dilution of precision. Therefore, the lower the DOP, the larger the volume in space. Ideally, each of the position vectors of the four satellites relative to the user should be orthogonal to the others. This presents a problem given the limitations of a three-dimensional world, so the best approach is to make three position vectors orthogonal. The satellites define a triangle in space. We choose the fourth position vector to be at the vertex of a right tetrahedron with triangles as the base. A more mathematical approach to the DOP uses the covariance matrix. The ***geometric dilution of precision*** (GDOP), is equal to the root sum square of the standard deviations of the state, which is equivalent to the square root of the trace of the covariance.

Other subsets of the DOP include the ***position dilution of precision***, which is the root sum square of the standard deviations of only the position. The ***time dilution of precision*** is the standard deviation of the time. The ***horizontal dilution of precision*** becomes the root sum square of the standard deviations for the elements of the state existing in the tangential plane, and the ***vertical dilution of precision*** is the standard deviation of the component perpendicular to the tangential plane. The horizontal and vertical dilutions result from a rotation of the position covariance into the topocentric (SEZ) coordinate system.

Differential Techniques

The U.S. Department of Defense envisioned the civilian community using the GPS network and thus constrained them to the Coarse-Acquisition code. Combining the less accurate code and selective availability meant less accuracy for the civilian community. Civilian engineers had to find a way to improve the obtainable accuracy from GPS. The solution came in the form of ***differential GPS***, DGPS, which calculates corrections to the signals. The basic concept of DGPS uses two receivers, one at a known location and one at an arbitrary position, that see at least four satellites in common. By fixing the location of one of the receivers, we can find the other location either by computing corrections to the satellite's position or by computing corrections to the pseudoranges.

By using DGPS, we can remove the effects of selective availability (SA). For short-baseline distances between receivers, we can remove some of the biases from the atmosphere as well. This cancellation effect is the result of both receivers seeing the same portion of the atmosphere. If we know one receiver's location, we can calculate the bias in the pseudorange to the known receiver and use it to correct the solution of the other receiver's location. The DGPS system designed by the U.S. Coast Guard calculates the biases at a known receiver's location and then broadcasts them on a radio frequency.

Double-Difference Processing

Double differences are formed by differencing the signals from two satellites for each receiver. These differences are then "differenced" to arrive at the solution. This technique is similar to DGPS. It uses double-differenced observables to eliminate the effects of selective availability and other biases in solving for baseline distances and relative positions. Double differences are used mainly for surveying and geodetic research that observes phases; however, they aren't limited to those applications. The real advantage of double differencing is that clock corrections (and therefore clock dithering) are eliminated from the observations at similar times. Be careful when investigating other sources about double differencing. Some approaches form the double differences from the phase-derived pseudoranges instead of the phases themselves. As an aside, *single differences* can be used to remove satellite clock corrections.

At this point, the GPS system may seem to have an incredible number of variables. It does! The system is very complex because it must produce accurate timing and positional data in the presence of terms for special and general relativity. I've only sketched the system. I encourage you to study it further and recommend Wheeler (1990), Logsdon (1992), or Hoffmann-Wellenhof et al. (1994) for more information.

10.6 Predicting Satellite Look Angles

A very useful application of the perturbation techniques discussed so far is the prediction problem. For this book, **PREDICT** is defined as the calculation of future times and angles when a satellite will be visible from a ground station. It applies widely, including the requirement to provide a sensor with predicted look angles (azimuth and elevation, or right ascension and declination) with which it may observe the satellite. It also allows us to view a satellite ourselves if the lighting conditions permit.

In this section I'll develop very few new techniques and rely heavily on algorithms from the rest of the book. The given data consists of the satellite's state vector (position and velocity vectors) at some given epoch and the desired interval for which the calculations should proceed for a given sensor site. Depending on accuracy requirements, we must always distinguish geocentric and topocentric values, proper coordinate systems, and so on. In general, propagation routines move the satellite's state vector to the start time, if different from the epoch, and then throughout the interval. At each step through the interval, we calculate the angular measurements and visibility parameters. The usual output is azimuth and elevation but may include right ascension and declination values.

Visibility checks, which are the only new techniques, divide into several cases. Figure 10-18 shows the geometry. The proximity of the site to the terminator is important when determining how the satellite will appear (visually) from the site. The **terminator** is the line on the Earth that separates night and day. As Fig. 10-18 shows, there's a period of time after the site passes the terminator when the site is in the dark, but the satellite is illuminated by the Sun. You can see this usually isn't a very long interval, and it depends on the satellite's altitude, but a sunlit satellite is very useful to help acquire satellites if the orbit isn't well-known. Of course, as the satellite's altitude increases, the satellite becomes dimmer (its *magnitude* increases) and we'll eventually not be able to see it.

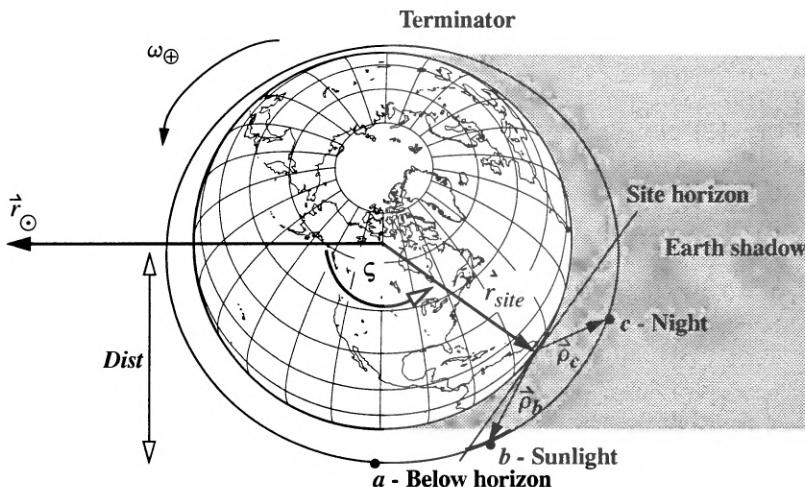


Figure 10-18. Determining Visibility. The visibility of a low-Earth satellite from a ground station, r_{site} , depends on several things: the Earth obstructs the view, *a*; the satellite is in the sunlight, *b*; and the satellite is in the Earth's shadow, *c*. We assume the Earth's shadow is cylindrical. Some observations are unaffected by the Earth's shadow.

Three conditions are necessary to determine visibility. To establish these conditions, we first find the slant-range vector from the site to the satellite ($\vec{r}_{sat_{SEZ}} = \vec{r}_{sat_{SEZ}} - \vec{r}_{site}$). This allows us to find the first condition in which the satellite is above or below the horizon. We determine this simply by looking at the third component of the range vector in the SEZ coordinate system. It's positive when the satellite is above the horizon. The next condition requires determining if the site is in sunlight or in the Earth's shadow, which we can do by taking the dot product of the Sun and site vectors ($\vec{r}_{\odot} \cdot \vec{r}_{site}$). If the result is positive, the site is in the Sun; if not, it's in the shadow. The final condition consists of determining whether the satellite is in the Sun and the site is in the shadow. We must determine the angle, ζ , from the Sun to the satellite and then the perpendicular distance (*Dist*) from the centerline of the Sun to the satellite [$Dist = r \cos(\zeta - 90^\circ)$]. I've described this

angle in Sec. 3.8, page 194. If the perpendicular distance is larger than the Earth's radius, the satellite is in the Sun and is visible to the sensor site. By arranging each of the tests in the proper order, we can account for all visibility outcomes. The algorithm shows the proper order. I use three labels for visibility: *Visible* means the satellite is sunlit and the site is in the dark; *radar night* means the satellite isn't sunlit and the site is in the dark; and *radar Sun* means the satellite and site are in the sunlight.

Before presenting the algorithm, let's examine some other constraints. Perhaps most important are the limitations imposed by the viewing site. This process embodies the discussion in Sec. 9.2. Sites that are very close to large mountains, such as the U.S. Air Force Academy, may limit elevations on certain azimuths. Although the algorithm may predict a viewing opportunity, physical limitations may prevent a successful sighting.

Related to site-viewing limits are the orbital geometries, often referred to as *satellite passes*. We define a *satellite pass* as the period of time in which a satellite flies over the observing site. The mathematical quantities determining a satellite pass are time, range, azimuth, and elevation. Other angles, such as right ascension and declination, commonly replace azimuth and elevation angles. It's desirable to select a pass which has a reasonable length of time—say five to ten minutes for a low-Earth satellite. A pass of only two minutes may not be long enough to get adequate data. It may also be at only a few degrees of elevation, which also limits the accuracy because of atmospheric refraction. High elevations diminish the limitations on the observations and are usually coupled with the closest approaches.

Taking these constraints into account, we can write Algorithm 65 to predict look angles for any satellite visible from a ground station. I've shown calculations for azimuth and elevation, but you can also find topocentric right-ascension and declination. The choice of propagator can use any of the techniques discussed in Chap. 7 or Chap. 8. Be sure the units of the Sun's position vector are compatible with the rest of the algorithm.

ALGORITHM 65: *PREDICT* ($\vec{r}_o, \vec{v}_o, t_o, \vec{r}_{site}, \Delta t \Rightarrow t_i, \rho_i, \beta_i, el_i, Vis_i$)

PKEPLER ($\vec{r}_o, \vec{v}_o, \Delta t = t_{start} - t_o \Rightarrow \vec{r}, \vec{v}$)

LOOP

PKEPLER ($\vec{r}, \vec{v}, \Delta t = t_{curr} - t_{start} \Rightarrow \vec{r}, \vec{v}$)

LSTIME ($\lambda, JD \Rightarrow \theta_{LST}, \theta_{GST}$)

SITE ($\phi_{gd}, h_{ellp}, \theta_{LST} \Rightarrow \vec{r}_{site}, \vec{v}_{site}$)

$$\hat{\rho} = \vec{r} - \vec{r}_{site}$$

Rotate $\hat{\rho}$ to SEZ and determine visibility

```

IF  $\hat{r}_Z > 0$ 
  SUN ( $JD \Rightarrow \hat{r}_{\odot}$ )
  IF  $\hat{r}_{\odot} \cdot \hat{r}_{site} > 0$ 
    Vis = 'Radar Sun'
  ELSE
    SIN ( $\zeta$ ) =  $\frac{|\hat{r}_{\odot} \times \hat{r}|}{|\hat{r}_{\odot}| |\hat{r}|}$ 
    Dist =  $|\hat{r}| \cos(\zeta - 90^\circ)$ 
    IF Dist > 1.0
      Vis = 'Visible'
      Find  $\alpha, \delta, \alpha_t, \delta_t, \beta, el$ , as appropriate
    ELSE
      Vis = 'Radar night'
  ELSE
    Vis = 'not visible'
  JD =  $JD + \Delta t$ 
  t =  $t + \Delta t$ 
  UNTIL  $t > t_{end}$ 
-----
```

▼ **Example 10-6. Predicting Look Angles (PREDICT).**

GIVEN: April 1, 1997, 4:36 pm local time, a site at Massachusetts Institute of Technology $\phi_{gd} = 42.38^\circ$, $\lambda = -71.13^\circ$, $h_{ellp} = 24$ m, and the Mir space station. Use the orbital elements from Chap. 2, Sec. 2.4.2, but at the time given here.

FIND: Look angles from the site for the two days.

We must first find the position and velocity vectors. Using Algorithm 6, find

$$\hat{r} = 6585.0396\hat{i} + 1568.1846\hat{j} + 9.1163\hat{K} \text{ km}$$

$$\hat{v} = -1.115776\hat{i} + 4.631681\hat{j} + 6.014957\hat{K} \text{ km/s}$$

As an aside, notice that the 2-line elements sets are usually propagated so the satellite is near the equator, thus the small K -component. Now determine an ephemeris for the satellite during the time of interest using the following formulas:

$$PKEPLER (\hat{r}_o, \hat{v}_o, \Delta t = t_{start} - t_o \Rightarrow \hat{r}, \hat{v})$$

$$LSTIME (\lambda, JD \Rightarrow \theta_{LST}, \theta_{GST})$$

$$SITE(\phi_{gd}, h_{ellp}, \theta_{LST} \Rightarrow \vec{r}_{site}, \vec{v}_{site}) \text{ and } \vec{\rho} = \vec{r} - \vec{r}_{site}$$

Because we'll get many vectors throughout the time, let's look at the process for one set of vectors on April 2, 1997 at 0:40:0.00 UTC (JD = 2,450,540.5278). Determine visibility as follows. The position and velocity vectors are

$$\vec{r} = -2928.748\hat{i} + 3612.284\hat{j} + 4912.409\hat{k} \text{ km}$$

$$\vec{v} = -6.8946\hat{i} - 2.5076\hat{j} - 2.2661\hat{k} \text{ km/s}$$

Rotating the position vector to the *SEZ* coordinate system and using

$$\vec{r}_{site} = -2987.135\hat{i} + 3652.788\hat{j} + 4276.893\hat{k} \text{ km}$$

$$\vec{\rho} = -515.4778\hat{i} - 19.5573\hat{j} + 377.9011\hat{k} \text{ km}$$

Because $\vec{\rho}_Z > 0$, we must find coordinates for the Sun to determine the visibility.

$$SUN(JD \Rightarrow \vec{r}_{\odot}), \vec{r}_{\odot} = 0.9765\hat{i} + 0.1957\hat{j} + 0.0848\hat{k} \text{ AU}$$

Be sure to convert this vector to ER before finding the dot product, $\vec{r}_{\odot} \cdot \vec{r}_{site}$. We can find the Sun-satellite angle because the dot product (-6764.377 ER) is less than zero.

$$\sin(\zeta) = \frac{|\vec{r}_{\odot} \times \vec{r}|}{|\vec{r}_{\odot}| |\vec{r}|}, \zeta = 75.1188^\circ$$

The distance quantity is

$$Dist = |\vec{r}| \cos(\zeta - 90^\circ) = 1.0250 \text{ ER}$$

Because *Dist* is greater than 1.0, the satellite is visible. We can find $\alpha, \delta, \alpha_t, \delta_t, \beta, el$, as appropriate. We continue this process for the remaining time interval to arrive at the answers in Table 10-4. .

TABLE 10-4. Prediction Values for Example 10-6. This table shows sample visibility parameters for three passes in the example problem.

Visibility	Range km	β °	el °	Date	UTC (h:min:s)
Radar Sun	1939.0553	308.7953	3.4753	Apr 1, 97	23:00:00
Radar Sun	1336.2945	330.6628	11.7486	Apr 1, 97	23:02:00
Radar Sun	1109.2059	12.8207	16.5096	Apr 1, 97	23:04:00
Visible	1450.4773	50.7766	9.7693	Apr 1, 97	23:06:00
Visible	2095.7887	69.1973	1.8027	Apr 1, 97	23:08:00
Visible	2033.5055	308.4148	2.4780	Apr 2, 97	00:36:00
Visible	1246.2822	318.0732	13.4250	Apr 2, 97	00:38:00
Visible	639.4761	357.8273	36.2248	Apr 2, 97	00:40:00
Radar Night	850.2544	81.8094	24.4537	Apr 2, 97	00:42:00
Radar Night	1578.3136	101.2887	7.7160	Apr 2, 97	00:44:00
Visible	2143.8520	293.8868	1.3452	Apr 2, 97	02:12:00



10.7 Determining Close Approaches

On July 24, 1996, a French satellite (Cerise) collided with a piece of debris from an Ariane rocket-body that was launched in 1986. As we deploy more satellites, the ability to determine periods when they are close to one another is a much more significant challenge for engineers concerned with orbital safety. For discussion, we'll call the satellite of interest the *primary* and the other satellite the *secondary*. Our basic problem is to determine two quantities: the time of closest approach and the relative distance between the two satellites.

My discussions of differential correction and propagation in this book highlight that our knowledge of satellite states is inexact. Even with very good observations, the resulting element sets have errors at epoch. Furthermore, as element sets are propagated through time, those initial errors can and do grow rapidly. Engineers must have a robust routine to determine periods of close approach and a method to determine the statistical probability.

Orbit analysts run the Space Shuttle against several hundred tracked satellites to identify those which may come dangerously close to the Shuttle. After tabulating future close encounters, the Air Force's orbit analysts can alert NASA's mission controllers of the conjunction times as well as their relative separation. On Shuttle Missions 44 and 48, debris was predicted to enter a safety zone—sometimes called an *error ellipsoid* (triaxial spheroid) or a warning football—resulting in NASA's commanding the Space Shuttle away from the approaching objects. Analysis of a planned maneuver by Mission 57 predicted a close encounter with a spent Cosmos booster (Kennedy, 1993), so NASA again commanded the Shuttle to avoid it.

An obvious solution to the close-approach problem is to step sequentially along the orbits of two satellites—to form an ephemeris for each—and then difference the position vectors to determine their relative distances. Indeed, we use this approach to construct a *truth* table of encounters. Although straightforward, this approach takes many computations, especially for a lot of satellites.

I'll present a method developed by Alfano and Negron (1993) and Alfano (1994), called the *Alfano-Negron Close Approach Software (ANCAS)*. This method determines close approaches between two arbitrary trajectories. It has several advantages over traditional geometrical and brute-force methods. *ANCAS* uses a relative-distance function, for which each point and slope matches in the blending functions. Our main goal here is determining the distance between the two satellites. The method uses any propagation scheme, permitting us to select a desired speed and accuracy. Finally, the answers result from closed-form solutions and eliminate the need for iteration. I'll introduce a method developed by Carlton-Wipperf (1990) to determine the probability that a satellite is at the predicted location. This particular *combination* of routines marks the most accurate way to analyze close-approach periods.

Besides predicting times and distances, the close-approach technique has other uses. As we launch satellites and move them between orbits, we must do close-approach calculations. A less obvious application is reconstructing a maneuver. If a satellite does an unknown maneuver, some sensors may lose the satellite and actually track it as a new

satellite. But, by running the close-approach technique backwards, we can usually find the maneuver time and separate the two satellites.

10.7.1 Finding the Close-Approach Functions

The first calculation compares the apogee and perigee values of the orbits. This provides a rough cut to determine if close approaches are possible. Even with a fast algorithm, it doesn't make sense to compare circular low-Earth and GEO satellites for close approaches. If the absolute value between the largest perigee and the smallest apogee is greater than the desired distance tolerance, the two satellites will never approach one another, unless a thrusting maneuver is applied. Thus,

$$|r_{p_{max}} - r_{a_{min}}| > Dist \quad (10-45)$$

This test is good for all closed orbits, including those in different planes. Orbit that satisfy this initial filter need no further processing. Figure 10-19 illustrates the geometry. We could use many other geometrical filters, but they require a lot of computing and still include evaluating each satellite's ephemerides.

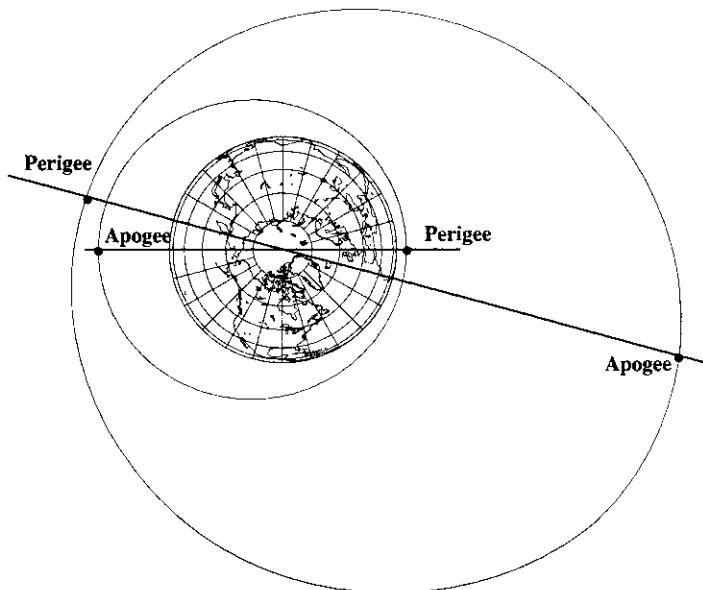


Figure 10-19. Apogee-Perigee Filter. The example shows two orbits, one with $e = 0.2$ and $a = 8637.018$ km and the other with $e = 0.35$ and $a = 17,444.459$ km. Even with perturbations, the apogee of the smaller orbit ($r_a = 10,364$ km) will never approach the perigee of the larger ($r_p = 11,339$ km). These orbits lie in the same plane, but this concept works for orbits in different planes.

Distance and Range-rate Functions

Let's begin with the geocentric equatorial (*IJK*) position vectors of the *primary*, \vec{r}_p , and *secondary*, \vec{r}_s , satellites at time t . The *relative-distance vector*, \vec{r}_d , and its time derivatives become

$$\begin{aligned}\vec{r}_d &= \vec{r}_s - \vec{r}_p \\ \dot{\vec{r}}_d &= \dot{\vec{r}}_s - \dot{\vec{r}}_p \\ \ddot{\vec{r}}_d &= \ddot{\vec{r}}_s - \ddot{\vec{r}}_p\end{aligned}\tag{10-46}$$

where *any* orbit propagator can produce the ephemerides for the primary and secondary satellites. We can take large time steps with the propagator because of the splining techniques we use later. If the satellite's acceleration vectors aren't available, we can approximate the *IJK* components using Eq. (7-46), but this simplification includes only J_2 perturbative forces, making it only an approximation. We define the *distance function*, $f_d(t)$, and its time derivatives using dot products, and we square the distance function so we don't need to evaluate a square root:

$$\begin{aligned}f_d(t) &= \vec{r}_d \cdot \vec{r}_d \\ \dot{f}_d(t) &= 2(\dot{\vec{r}}_d \cdot \vec{r}_d) \\ \ddot{f}_d(t) &= 2(\ddot{\vec{r}}_d \cdot \vec{r}_d + \vec{r}_d \cdot \ddot{\vec{r}}_d)\end{aligned}$$

We evaluate these equations for a sequence of times until two steps containing a minimum are found. Satellite close approaches occur whenever $f_d(t)$ is at a local minimum—that is, when $\dot{f}_d(t) = 0$ and $\ddot{f}_d(t) > 0$. The final condition implies the two satellites are in the same half-plane of the orbit. To determine these times of closest approach, compute the coefficients α_{cd} of the *derivative* function for the range-rate polynomial equation $P_{cd}(\tau)$ [Eq. (C-39), see Algorithm 66]. If you use the derivative of the function, the cubic splining in Eq. (C-39) still applies—you simply use the first and second derivatives instead of the function and first derivative. Remember that τ varies from 0.0 to 1.0. Now, extract the real, distinct root(s), τ_{droot} , of $P_{cd}(\tau)$ on the interval 0.0 to 1.0. Use the “*d*” subscript to indicate the distance function. This is necessary to distinguish the results from the ellipsoidal function we'll examine shortly. If

$$\left. \frac{dP_{cd}(\tau)}{d\tau} \right|_{\tau = \tau_{droot}} > 0$$

a local minimum exists. You must still determine the time and range with τ_{droot} known. Use quintic splines to capture the contribution of acceleration and to keep the solution accurate. Form the quintic spline, P_{qi_f} , using Eq. (C-45) (see Algorithm 66) and the *I*-

components of Eq. (10-46). Form P_{qi_J} and P_{qi_K} similarly. The quintic polynomial uses τ_{droot} and doesn't require you to find the root. The minimum distance is

$$d_{close} = \sqrt{P_{qi_I}^2(\tau_{droot}) + P_{qi_J}^2(\tau_{droot}) + P_{qi_K}^2(\tau_{droot})}$$

and the associated time of the close approach is

$$t_{close}(\tau_{droot}) = t_n + \tau_{droot}\Delta t$$

where t_n is an endpoint of the time interval containing the minima, and $\Delta t = t_{n+1} - t_n$.

The Ellipsoidal Function

Once we have a candidate close approach, we must identify an error ellipsoid to model a "safety zone" around the primary satellite. We need this zone because propagation and differential correction techniques (Chaps. 7, 8, and 9) introduce some errors into the solution. Although we get a number indicating the satellite's location at a point in time, a volume is actually defined by the propagation of the covariance matrix. The propagation of the covariance depicts the satellite's probability of occupying a certain location within the overall region. This zone introduces a length of time during which the two satellites are nearby. Typically, the propagation error is greatest in the along-track direction because time errors greatly displace the satellite along the orbital path. This tends to create an elongated zone (football shaped, see Fig. 10-20) about the primary satellite's orbit. The zone is characterized by three axes. We often set these values arbitrarily because many systems don't have covariance information. For precise work, you should try to get accurate covariance estimates and use them when characterizing the error region. If you've done your differential-correction processing correctly, your error region should be fairly small. You want to create this region so the major axis is colinear with the primary satellite's velocity vector, \vec{v}_p . This is the *NTW* coordinate system.

It's important to obtain a function that determines the entry and exit conditions for the secondary satellite from the ellipsoidal region about the primary satellite. We're only interested in a two-dimensional result, so if we project \vec{r}_d along the primary axis (T) using the Pythagorean theorem, we can use the general form of an ellipse to achieve the desired result.

$$\frac{N^2}{a^2} + \frac{T^2}{b^2} - 1 = 0 \quad N = \frac{\vec{r}_d \cdot \vec{v}_p}{\vec{v}_p} \quad T^2 = \vec{r}_d \cdot \vec{r}_d - N^2$$

Projecting \vec{r}_d onto the \vec{r}_p unit vector gives us the *ellipsoidal function*,

$$f_e(\tau) = \frac{\frac{(\vec{r}_d \cdot \vec{v}_p)^2}{(\vec{v}_p \cdot \vec{v}_p)}}{a^2} + \frac{(\vec{r}_d \cdot \vec{r}_d) - \frac{(\vec{r}_d \cdot \vec{v}_p)^2}{(\vec{v}_p \cdot \vec{v}_p)}}{b^2} - 1$$

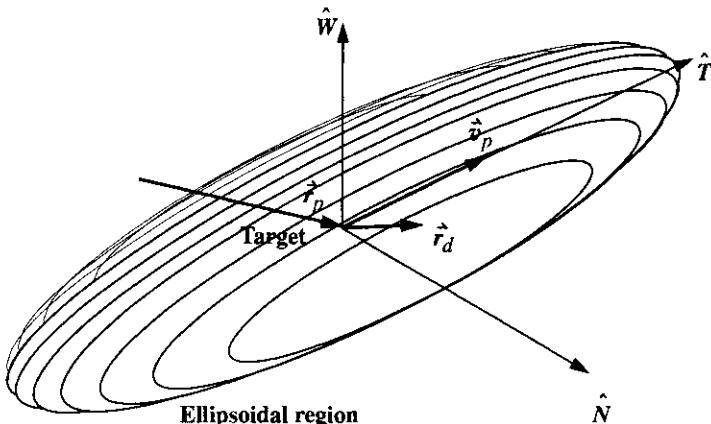


Figure 10-20. Geometry for the Ellipsoidal Function. The ellipsoidal region depicts our imprecise knowledge with respect to where the satellite is over time. As we predict farther into the future, the region expands. Because time is often a main source of error, the along-track component (in the T direction) is usually the largest.

where several conditions are possible. If $f_e(\tau) > 0$, the secondary is out of range. $f_e(\tau) = 0$ occurs as the secondary satellite enters or exits; the satellite is at the surface of the ellipsoidal region. $f_e(\tau) < 0$ means the secondary is within the specified range of the primary. We need the primary satellite's velocity only to orient the NTW coordinate system. It's not part of the overall logic. For a sphere, the minor and major axes are equal, so we don't need the projections, and the ellipsoidal function is

$$f_e(\tau) = \frac{|\vec{r}_d \cdot \vec{r}_d|}{a^2} - 1$$

Let's examine the entry and exit points more closely. To determine the times of entry and exit, you must compute the coefficients α_{ce} from Eq. (C-40) (see Algorithm 66) for the cubic polynomial equation $P_{ce}(\tau)$. To do so, you need two endpoints, p_1, p_4 , and two intermediate points, p_2, p_3 , for the interval. The endpoints are simply

$$\begin{aligned} p_1 &= f_e(0), \quad (t = t_n) \\ p_4 &= f_e(1), \quad (t = t_{n+1}) \end{aligned}$$

But you'll need to reconstruct the intermediate points from the distance functions of Eq. (10-46), assuming you can't just get the function values by invoking the propagator. This reconstruction involves determining the polynomial coefficients, P_{ce} , as before. Also, you must reconstruct the primary velocity and acceleration vectors at times t_n and t_{n+1} to produce the cubic coefficients, α_{ce} . The intermediate points become

$$p_2 = f_e(\tau_1), \tau_1 = 1/3$$

$$p_3 = f_e(\tau_2), \tau_2 = 2/3$$

Note that τ_1 and τ_2 are evenly spaced on the interval. If a minimum distance occurs in the interval, the values change to distribute the other root by halving the largest remaining interval.

$$\begin{aligned} \tau_2 &= \tau_{droot} \\ \text{IF } \tau_{droot} < 0.5 \\ \tau_1 &= (1 + \tau_{droot})/2 \\ \text{ELSE} \\ \tau_1 &= \tau_{droot}/2 \end{aligned}$$

By including the minimum distance, you'll force the $P_{ce}(\tau)$ curve to pass through that point; this ensures an entry or exit won't be missed due to ill-conditioned endpoints, as long as the quintic functions accurately represent relative distance components. Extract the real root(s), τ_{eroot} , of $P_{ce}(\tau)$ on the interval 0.0 to 1.0.

Use a refining process to find the entry or exit time if a valid root exists. This step improves the accuracy of the $P_{ce}(\tau)$ curve in the region of τ_{eroot} by reconstructing the intermediate point p_2 at the root. If you find a second root, recompute p_3 ; otherwise, it remains associated with the minimum distance. Then, recompute new α_{ce} coefficients for the $P_{ce}(\tau)$ curve from Eq. (C-39). Also recompute the real root(s), τ_{eroot} , of $P_{ce}(\tau)$ on the interval from 0.0 to 1.0, so the time of entry or exit becomes ($\Delta t = t_{n+1} - t_n$)

$$t(\tau_{eroot}) = t_n + \tau_{eroot}\Delta t$$

An exit exists if

$$\left. \frac{dP_{ce}(\tau)}{d\tau} \right|_{\tau=\tau_{eroot}} > 0$$

Otherwise, it's an entry.

Implementing the Close-Approach Algorithm

The solution relies on combining cubic and quintic splining. Because $P_{cd}(\tau)$ is a cubic equation defining the range rate between t_n and t_{n+1} , you may find more than one real root in the time interval that satisfies the inequality constraint. This condition indicates that two minima occurred in the interval and poses no computational problems. To compute faster, tailor the polynomial-solution method of Sec. C.4.1 to provide the best performance by ignoring all complex roots. Higher-order polynomials could fit the data more accurately, but you'd lose the ability to find roots in closed form. Depending on the specific values of a problem, this can dramatically increase computer runtime, so you should try to reduce extraneous computations of this polynomial.

Algorithm 66 shows the complete process including determination of the entrance and exit from an error ellipsoid around the primary satellite. If you don't need this calculation, you can save many computations. Don't take too large a time step. Trial cases suggest that roots were missed when true anomaly changed more than 40° in a single time step. For two minima to occur in a single time interval, either the primary or secondary would most likely exceed this angle. Finally, although I show **PKEPLER** as the propagation algorithm, you can use any propagator.

ALGORITHM 66: ANCAS ($\hat{\mathbf{r}}_s, \hat{\mathbf{v}}_s, \hat{\mathbf{r}}_p, \hat{\mathbf{v}}_p, JD_{start}, JD_{stop}, \Delta t \Rightarrow t_{close}, d_{close}, \tau_{entry}, \tau_{exit}$)

Do an initial filter to eliminate obvious candidates:

$$|r_{p_{max}} - r_{a_{min}}| > Dist$$

Generate ephemerides for the primary and secondary:

$$\text{PKEPLER}(\hat{\mathbf{r}}_s, \hat{\mathbf{v}}_s, \Delta t \Rightarrow \hat{\mathbf{r}}_s, \hat{\mathbf{v}}_s)$$

$$\text{PKEPLER}(\hat{\mathbf{r}}_p, \hat{\mathbf{v}}_p, \Delta t \Rightarrow \hat{\mathbf{r}}_p, \hat{\mathbf{v}}_p)$$

$$\hat{\mathbf{r}}_d = \hat{\mathbf{r}}_s - \hat{\mathbf{r}}_p$$

$$\dot{\hat{\mathbf{r}}}_d = \hat{\mathbf{v}}_s - \hat{\mathbf{v}}_p$$

$$\ddot{\hat{\mathbf{r}}}_d = \ddot{\hat{\mathbf{r}}}_s - \ddot{\hat{\mathbf{r}}}_p$$

$$f_d(t) = \hat{\mathbf{r}}_d \cdot \dot{\hat{\mathbf{r}}}_d$$

$$\dot{f}_d(t) = 2(\dot{\hat{\mathbf{r}}}_d \cdot \dot{\hat{\mathbf{r}}}_d)$$

$$\ddot{f}_d(t) = 2(\ddot{\hat{\mathbf{r}}}_d \cdot \dot{\hat{\mathbf{r}}}_d + \dot{\hat{\mathbf{r}}}_d \cdot \ddot{\hat{\mathbf{r}}}_d)$$

Search for the root for the time of minimum range, τ_{droot} :

Roots occur where $\dot{f}_d = 0$, and $\ddot{f}_d > 0$

$$P_{cd}(\tau) = \alpha_{c3}\tau^3 + \alpha_{c2}\tau^2 + \alpha_{c1}\tau + \alpha_{c0}$$

Using Eq. (C-39) with $f_d = \dot{f}_d$ and $\dot{f}_d = \ddot{f}_d$,

Solve $P_{cd}(\tau) = 0$ to find τ_{droot}

Find the minimum distance at τ_{droot} . Solve the three quintics, one for each vector component. (I'll show scalar α terms and vector $\hat{\mathbf{r}}_d$ to indicate this process.)

$$P_{qi}(\tau) = \alpha_{qi5}\tau_{droot}^5 + \alpha_{qi4}\tau_{droot}^4 + \alpha_{qi3}\tau_{droot}^3 + \alpha_{qi2}\tau_{droot}^2 + \alpha_{qi1}\tau_{droot} + \alpha_{qi0}$$

Using Eq. (C-45) with $f_d = \dot{\vec{r}}_d$, $\dot{f}_d = \ddot{\vec{r}}_d$, $\ddot{f}_d = \dddot{\vec{r}}_d$

Solve $P_{qi}(\tau) = 0$ to find τ_{droot}

$$d_c = \sqrt{P_{qi_I}^2(\tau_{droot}) + P_{qi_J}^2(\tau_{droot}) + P_{qi_K}^2(\tau_{droot})}$$

$$t_c(\tau_{droot}) = t_n + \tau_{droot}\Delta t$$

Now check for the error-ellipsoid region:

$$f_e(\tau) = \frac{\frac{(\dot{\vec{r}}_d \cdot \dot{\vec{v}}_p)^2}{(\vec{v}_p \cdot \vec{v}_p)} - \frac{(\dot{\vec{r}}_d \cdot \dot{\vec{v}}_p)^2}{(\vec{v}_p \cdot \vec{v}_p)}}{a^2} + \frac{b^2 - 1}{b^2}$$

$$p_1 = f_e(0)$$

$$p_2 = f_e(\tau_1), \tau_1 = 1/3$$

$$p_3 = f_e(\tau_2), \tau_2 = 2/3$$

$$p_4 = f_e(1)$$

$$P_{ce}(\tau) = \alpha_{c3}\tau^3 + \alpha_{c2}\tau^2 + \alpha_{c1}\tau + \alpha_{c0}$$

Using Eq. (C-40),

Solve $P_{ce}(\tau) = 0$ to find τ_{droot}

IF τ_{droot} exists, THEN find exactly the close approach time

$$\tau_2 = \tau_{droot}$$

$$\left[\begin{array}{l} \text{IF } \tau_{droot} < 0.5 \\ \quad \tau_1 = (1 + \tau_{droot})/2 \\ \text{ELSE} \\ \quad \tau_1 = \tau_{droot}/2 \end{array} \right]$$

$$\tau_1 = (1 + \tau_{droot})/2$$

ELSE

$$\tau_1 = \tau_{droot}/2$$

Solve $P_{qi}(\tau) = 0$ using Eq. (C-45) to find τ_{eroot}

$$t_{\text{entry/exit}}(\tau_{eroot}) = t_n + \tau_{eroot}\Delta t$$

▼ Example 10-7. Determining a Close Approach.

GIVEN: Classical orbital elements of the U.S. Space Shuttle Endeavor and a Soviet Cosmos rocket booster shown in Table 10-5. This example is from Alfano (1994) as part of a prediction analysis for Mission 57 in 1993.

FIND: Times and distances of closest approach of Endeavor and the Cosmos booster.

TABLE 10-5. Orbital Elements for Example 10-7. This table shows elements for the Space Shuttle Endeavor and Cosmos rocket booster.

	Endeavor	Cosmos Booster
a	6821.120 045 km (1.069 453 36 ER)	6850.377 771 km (1.074 040 55 ER)
e	0.003 849 877	0.001 603 018
i	28.444 697 8°	64.993 716 31°
Ω	299.761 799 28°	207.226 889 39°
ω	65.952 836 21°	264.831 399 57°
M	189.452 940 53°	64.460 082 15°
UTC (yr/day/h/min/s)	1993 174 12:32:07.091	1993 174 12:32:07.941

Develop a truth ephemeris by advancing the satellites at intervals of 0.1 s, solving for three parameters at each step: relative minimum and sphere and ellipsoid entrance/exit times. Model orbital motion using **PKEPLER**. After running the close-approach algorithm, find the times and ranges of close approach.

Using results from Table 10-6, a planned Space Shuttle maneuver was delayed 45 min to prevent a close approach with the Cosmos booster; Table 10-7 shows the predicted conjunction. It's interesting to note the timing accuracy of this method given the brevity of the closest encounter: 23 s within a 100 km sphere, with a close-approach distance of 4.5 km, and only 2.6 s within a 25 × 100 km ellipsoid! This brevity is due to the object's approach angle (110° off the primary velocity vector) with a closure rate of 8.7 km/s, where a timing error of 0.1 s can result in a range error of 870 m. Also, the 300 s time difference is chosen to capture all potential close approaches.

TABLE 10-6. Results for Example 10-7. This table shows the minimum approach distances and times for the Shuttle and the Cosmos booster. Note the nearest approach is more than 4 km. The truth model was generated using **PKEPLER** with a 0.1 s step size. The date is day 174 of 1993 (June 23).

Truth		Cubic Spline ($\Delta t = 300$ s)		Absolute Difference	
Time (h:min:s)	Range (m)	Time (h:min:s)	Range (m)	Time (s)	Range (m)
10:11:42.2	446,064.3	10:11:42.3	446,662.1	0.1	597.8
10:58:29.2	327,301.7	10:58:29.2	326,573.1	0.0	728.6
11:45:20.3	133,258.1	11:45:20.3	133,557.0	0.0	298.9
12:32:07.9	4445.1	12:32:08.0	4672.8	0.1	227.7
13:18:58.4	213,053.0	13:18:58.4	213,788.6	0.0	735.6

TABLE 10-7. Solutions for Entry and Exit within a 100 km Sphere (Example 10-7).
The simulation date is for day 174 of 1993 (June 23). I've shown answers for an ellipsoid and a sphere.

	100 km Sphere		25 × 100 km Ellipsoid	
	Truth (h:min:s)	Cubic Spline ($\Delta t = 600$ s)	Truth (h:min:s)	Cubic Spline ($\Delta t = 300$ s)
Entry	12:31:56.4	12:31:56.5	12:32:06.4	12:32:06.4
Exit	12:32:19.4	12:32:19.5	12:32:09.0	12:32:09.0
Duration (s)	23.0	23.0	2.6	2.6



10.7.2 Statistical Analysis

The next major concern is the statistical probability of the close approach. Moreover, *ANCAS* tells you when the orbits will approach each other, but the satellite positions used in the *ANCAS* algorithm are uncertain. The *ANCAS* algorithm doesn't tell you how those uncertainties affect the time and distance of closest approach. Carleton-Wipperf (1990) discusses how to determine probabilities for the close-approach problem. Consult this work for additional information.

Incorporating probability into the answers of a close-approach prediction is important to the operational planner. Suppose analysis indicates a piece of debris will pass within 3 km of the Space Shuttle. Should the Shuttle be commanded to move? This question is the basis for a lot of analysis and discussion needed to determine the correct option—only one step of which is the actual probability calculations.

First, consider the available satellite data and the accuracy of the ephemerides. Typically, if we use an analytical theory during differential correction, the resulting satellite ephemeris can have errors on the order of tens of kilometers or more. Using an analytical theory often negates any covariance analysis which we could use to approximate the accuracy of the ephemerides. It's desirable to have osculating vectors with an accurate covariance matrix derived from mathematical estimates of the real world. This process is mathematically intensive but readily available through applications of least squares or the Kalman filter introduced in Chap. 9.

If we assume we can obtain accurate covariance information about the time of the close approach, we can calculate the probability of collision. This leads to the final area of interest—the policy and legal considerations in maneuvering the satellite. If a satellite is predicted to collide with another, what probability is acceptable, who should maneuver, and who must determine the potential collision in the first place? These questions are not easily resolved, but in all cases, we want an accurate technical solution.

10.8 Rise / Set

When only a few satellites were in orbit, the early days of satellite operations consisted of establishing communications whenever a vehicle entered the ground station's visibility window. With a tabulation of viewing times and antenna-pointing parameters, satellite operators were prepared to transmit and receive data from the vehicle. These operations have changed a lot. Today, mission effectiveness relies greatly on spaceborne tracking and cross-link communications. Manufacturers design satellites to interact with operators, even when outside the visibility window of a ground station. Mission designers are concerned with the ways in which a satellite deployment strategy *loads* a ground station, as well as the frequency and duration whenever several satellites vie for station time. This is a key orbit-design issue because the satellite with the highest priority usually wins competitions for ground support. Also, as the satellite's altitude decreases, the opportunities to track or communicate with the vehicle from a ground station become more restricted, further complicating satellite scheduling. These issues call for a computationally efficient routine to generate a timetable of visibility periods—one that can process all orbit types against any ground location.

We define the *rise/set* problem as the process of determining the times at which a satellite rises and sets with respect to a ground location. The easiest solution uses a numerical method to determine visibility periods for the site and satellite by evaluating *IJK* position vectors of each. It advances vectors by a small time increment, Δt , and checks visibility at each step. A drawback to this method is computation time, especially when modeling many perturbations and processing several satellites.

Escobal ([1965] 1985, 162–174) proposed a faster method to solve the rise/set problem by developing a closed-form solution for unrestricted visibility periods about an oblate Earth. He assumes infinite range, azimuth, and elevation visibility for the site. Escobal transforms the geometry for the satellite and tracking station into a single transcendental equation for time as a function of eccentric anomaly. He then uses numerical methods to find the rise and set anomalies, if they exist.

More recently, Lawton (1987) developed a method to find visibility periods for vehicles in circular or nearly circular orbits by approximating an unrestricted visibility function with a Fourier series. Exploiting the sinusoidal nature of the visibility curve generated by satellites with orbital eccentricities less than 0.1, he determined the local shape of this curve and then used a numerical search to locate viewing times. This method works well for less eccentric orbits but fails for more elliptical orbits because the visibility function isn't periodic.

This section presents an analytical method from Alfano (1993) and Alfano, Negron, and Moore (1992). It rapidly determines rise/set times of a satellite for a ground station and includes restrictions for range, azimuth, and elevation. The algorithm uses pairs of fourth-order polynomials (quartic blending, Sec. C.4.3) to construct functions that represent the restricted parameters (range, azimuth, and elevation) versus time for an oblate Earth. It can produce these functions from either uniform or arbitrarily spaced data points. The viewing times are obtained by extracting the real roots of localized quintic

polynomials. This algorithm works for all orbital eccentricities and perturbed satellite motion, provided the functions remain continuous. We can substitute parabolic blending for quartic blending to reduce complexity, but this decreases accuracy for the same time step. If we take smaller steps to regain accuracy, processing time increases.

Functions for Azimuth, Elevation, and Range

The first few steps are identical to an approach which simply produces ephemerides for the site and satellite. The main difference is the ability to use much larger step sizes (on the order of five to ten minutes) by using quartic blending. Although you can choose the time intervals, don't choose one that's too large. If the interval is too large, you'll miss crossing times.

Let $\hat{r}_{site}(t)$ be the *IJK* position vector of the ground station and $\hat{r}_{sat}(t)$ be the satellite's corresponding position at time t . Define the vector from the site to the satellite, $\hat{\rho}_{IJK}(t)$, as

$$\hat{\rho}_{IJK}(t) = \hat{r}_{sat}(t) - \hat{r}_{site}(t) \quad (10-47)$$

This vector is easily transformed to the topocentric horizon coordinate system [Eq. (1-33)], yielding $\hat{\rho}_{SEZ}(t)$.

At first glance, one might think the range, azimuth, and elevation solutions from Algorithm 15 would be sufficient. But they're *not* for this problem because quartic blending works best with "smooth" functions. We'll define a "smooth" function as one which allows us to recover the rise/set parameters with quartic blending. Moreover, the presence of elevation and azimuth present unique difficulties when developing a universal solution. Consider a satellite rising at an azimuth of 350° and setting at an azimuth of 30° . You could use many checks and conditions to account for all possible occurrences passing through 360° and ensure a smooth transition to 0° , but that's a large task. A better solution is to develop functions that don't depend on the discontinuities of trigonometric arguments. If you know the limiting values of range, azimuth, and elevation (ρ_{lim} , high and low values of β_{lim} and el_{lim}), you can empirically determine the range, azimuth, and elevation functions.

Because range varies evenly throughout the orbit, you can use it directly. You can remove azimuth discontinuities by examining the south and east components of the slant-range vector. The elevation function uses the differences between actual and limiting values to create a slowly varying equation. Selecting these functions is very important to the final results—unfortunately, it's not easy to find them from scratch. If you chose the right functions, you may be able to use a lower-order blending or splining relation, which can greatly decrease the computation. As mentioned earlier, you can examine the problem geometry and use judgment to find the functions, but actual testing is the final means of validating the process. Here, let's choose the range as the standard value because it varies so little. The azimuth equation [Eq. (3-9)] has a step discontinuity when the function changes from 359° to 0° , so use the tangent expression to capture the time-varying compo-

nents that define β . Finally, for el , use a trigonometric equation that compares the limiting value with the time-varying values.

$$\begin{aligned} f_{range}(t) &= \rho(t) - \rho_{\text{LIM}} \\ f_{\beta_{lim}}(t) &= \rho_E(t) + \rho_S(t) \tan(\beta_{lim}) \Big|_{\text{High,Low}} \\ f_{el_{lim}}(t) &= \cos^{-1} \left[\frac{\cos(el(t))}{\lvert \mathbf{r}_{sat}(t) \rvert} \right] - el(t) - \cos^{-1} \left[\frac{\cos(el_{lim})}{\lvert \mathbf{r}_{sat}(t) \rvert} \right] + el_{lim} \Big|_{\text{High,Low}} \end{aligned} \quad (10-48)$$

The azimuth and elevation equations represent two functions for the high and low restrictions. For sensors which have complete visibility, use 0° and 360° , and -90° and 90° respectively. Any time one of these functions vanishes, determine the time of crossing from the polynomial root. Also, check signs for azimuth limits to test the roots validity. A sign check is like a quadrant check, but it avoids expensive trigonometric checks. These limit functions were designed to reduce the variability of the standard azimuth and elevation values, which increases the accuracy of the curve fit. Fig. 10-21 shows representative examples.

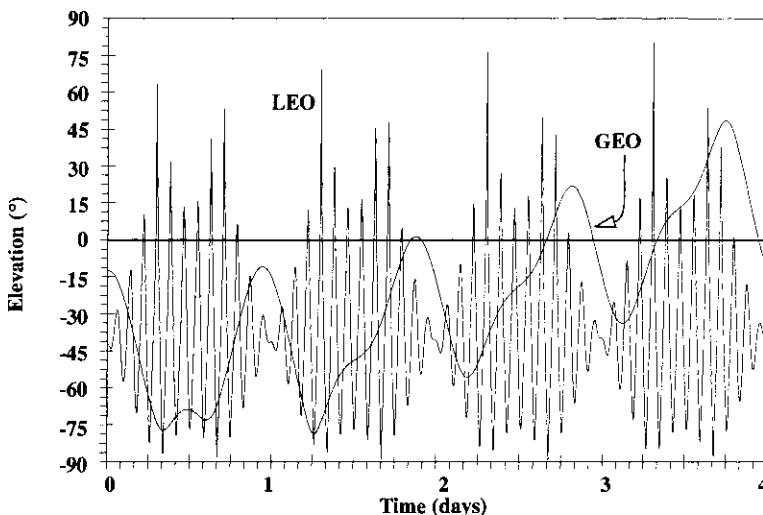


Figure 10-21. Elevation Versus Time. The standard functions of azimuth and elevation vary too much to permit an accurate solution with only a fourth-order polynomial fit. The curves show elevation values for low-Earth and GEO orbits. Notice the slight long-periodic variation evident in the GEO orbit. That's because it's a 55° inclined orbit.

Implementing the Rise/Set Algorithms

Because generating the ephemeris and determining rise/set are distinct problems, you can generate satellite ephemeris files to any desired accuracy from existing orbit-propagation methods—numerical, analytical, and semianalytical. You can also integrate this logic onboard the satellite computer to support autonomous operations; the computer generates timetables for a target set and then activates various instruments and operating modes once a site comes into view.

Algorithm 67 shows the method. Generating site and satellite ephemerides is identical to the simple numerical approach but permits larger time steps. Eq. (C-44) (see Algorithm 67) shows the relations for setting up a quartic blend. Unfortunately, the solution involves a quintic polynomial which has no analytical solution, so you have to process intelligently to avoid unnecessary calculations.

By choosing six consecutive values, ρ_S , ρ_E , and ρ_Z from a reference file, you can easily compute the α_{qr} coefficients for the limit functions for $F_{qr}(\tau)$ and then find the real unique root(s), τ_{root} . Notice that the points don't have to be equally spaced in time because you can take uneven time spacing and blend them like the rise/set functions. If the root(s) is within bounds, compute a different set of coefficients, α_{qr} , using the times that correspond to the original chosen values. This resulting $F_{qr}(\tau)$ equation yields the in-view or out-of-view time associated with τ_{root} . When a crossing time has occurred, again use quartic blending to determine $\rho_S(\tau_{root})$, $\rho_E(\tau_{root})$ and $\rho_Z(\tau_{root})$. Then, you can use those values to produce information on range, azimuth, and elevation from Eq. (10-48). Remember to form polynomials and recover the time of each occurrence.

We have several constraints to examine, so let's look at a graphic representation of what we're trying to do. Figure 10-22 shows the overall situation, as it applies to the rise/set problem.

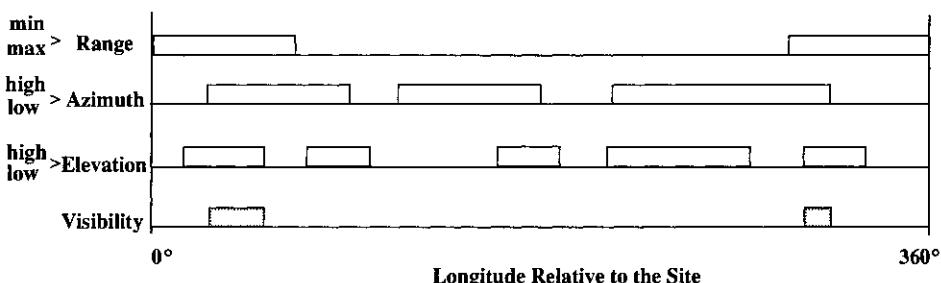


Figure 10-22. How Restrictions Affect Site Visibility. This figure shows how multiple restrictions add to yield the final answer. Rectangles show when a respective constraint is met. The satellite is visible only when all three restrictions are met. We formulate the solution for longitude values from 0° to 360°. The range function will always be symmetrical because the longitude is relative to the site. Azimuth and elevation aren't symmetrical.

Once we evaluate each of the constraints, we must “add” the results to determine actual rise and set times. In Fig. 10-22, there are only small visibility opportunities due to the many constraints.

Algorithm 67 allows us to calculate many quantities only once, thereby speeding the overall operation. Although I show propagation with the **PKEPLER** algorithm, you can use *any* propagator.

ALGORITHM 67: RISE/SET ($\vec{r}_{sat}, \vec{v}_{sat}, \phi_{gd}, \lambda, h_{ellip}, JD, \rho_{lim}, \beta_{lim}, el_{lim} \Rightarrow t_{rise/set}, \beta_{rise/set}, el_{rise/set}$)

$$C_{\oplus} = \frac{R_{\oplus}}{\sqrt{1 - e_{\oplus}^2 \sin^2(\phi_{gd})}} \quad S_{\oplus} = C_{\oplus}(1 - e_{\oplus}^2)$$

$$r_{\delta} = (C_{\oplus} + h_{ellip}) \cos(\phi_{gd})$$

$$r_K = (S_{\oplus} + h_{ellip}) \sin(\phi_{gd})$$

Generate a file of position vectors for the satellite and site, plus the transformation matrix and function values:

PKEPLER ($\vec{r}_{sat}, \vec{v}_{sat}, \Delta t \Rightarrow \vec{r}_{sat}(t), \vec{v}_{sat}(t)$)

LSTIME ($JD, UT1, \lambda_{East} \Rightarrow \theta_{LST}, \theta_{GST}$)

$$\vec{r}_{site}(t) = \begin{bmatrix} r_{\delta} \cos(\theta_{LST}) \\ r_{\delta} \sin(\theta_{LST}) \\ r_K \end{bmatrix}$$

$$\begin{bmatrix} SEZ \\ IJK \end{bmatrix} = \begin{bmatrix} \sin(\phi_{gd}) \cos(\theta_{LST}) & \sin(\phi_{gd}) \sin(\theta_{LST}) & -\cos(\phi_{gd}) \\ -\sin(\theta_{LST}) & \cos(\theta_{LST}) & 0 \\ \cos(\phi_{gd}) \cos(\theta_{LST}) & \cos(\phi_{gd}) \sin(\theta_{LST}) & \sin(\phi_{gd}) \end{bmatrix}$$

$$\vec{\rho}_{IJK}(t) = \vec{r}_{sat}(t) - \vec{r}_{site}(t)$$

$$\vec{\rho}_{SEZ}(t) = \begin{bmatrix} SEZ \\ IJK \end{bmatrix} \vec{\rho}_{IJK}(t)$$

$$\rho(t) = |\vec{\rho}_{SEZ}|$$

$$\sin(el(t)) = \frac{\rho_Z(t)}{\rho(t)}$$

Evaluate the functions at each time:

$$f_{range}(t) = \rho(t) - \rho_{lim}$$

$$f_{\beta_{lim}}(t) = \rho_E(t) + \rho_S(t) \tan(\beta_{lim}) \Big|_{High, Low}$$

$$f_{el_{lim}}(t) = \cos^{-1} \left[\frac{\cos(el(t))}{|\vec{r}_{sat}(t)|} \right] - el(t) - \cos^{-1} \left[\frac{\cos(el_{lim})}{|\vec{r}_{sat}(t)|} \right] + el_{lim} \Big|_{High, Low}$$

Search for rise/set times.

For each of the five functions, select six points to plug into Eq. (C-44):

$$f_{qr}(\tau) = \alpha_{qr5}\tau^5 + \alpha_{qr4}\tau^4 + \alpha_{qr3}\tau^3 + \alpha_{qr2}\tau^2 + \alpha_{qr1}\tau + \alpha_{qr0} \quad 0.0 \leq \tau \leq 1.0$$

Solve $P_{qr}(\tau) = 0$ to find each τ_{root} of the high/low constraints on ρ , β , and el

Combine results and determine τ_{root} for each common time.

Recover original values for range, azimuth, and elevation:

$$\rho(t) = |\vec{\rho}_{SEZ}|$$

$$\sin(el) = \frac{\rho_Z}{\rho}$$

IF Elevation $\neq 90^\circ$

$$\sin(\beta) = \frac{\rho_E}{\sqrt{\rho_S^2 + \rho_E^2}} \quad \cos(\beta) = \frac{-\rho_S}{\sqrt{\rho_S^2 + \rho_E^2}}$$

ELSE

Undefined

$$t_{rise/set}(\tau_{root}) = t_n + \tau_{root}\Delta t$$

▼ Example 10-8. Predicting Satellite Rise and Set Times.

GIVEN: Satellites with data shown in Table 10-8.

FIND: Rise and set times for one day for a site at the U.S. Air Force Academy initialized at the vernal equinox (meaning it's on the I axis). The station coordinates are 104.883°W , 39.007°N , and $h_{ellp} = 2.918 \text{ km}$.

For quartic blending, advance the site and satellite in three-minute intervals and check the limit functions for crossing times.

Table 10-9 shows the viewing times for a one-day simulation using the parameters of Table 10-8. I didn't include timetables for objects 1, 2, 4, and 7 because those objects weren't visible to the restricted site. Quartic blending (with three-minute time steps) reduces computer processing time by 89% over a five-second stepsize using **PKEPLER**. It also allows us to match point, slope, and acceleration at each point. Notice this method's results differ by less than one second from that of a truth model using a 0.1-second step.

TABLE 10-8. Orbital Elements for Example 10-8. For each satellite, $\omega = \Omega = M = 0.0^\circ$.

Object	n (rev/solar day)	e	i ($^\circ$)
1	1.002 721 41	0.000 003 2	0.0956
2	8.365 892 35	0.008 015 8	90.0175
3	0.248 919 61	0.936 306 0	64.9874
4	0.214 672 09	0.066 812 8	57.3500
5	13.376 596 79	0.014 507 2	90.2619
6	16.097 692 32	0.007 874 2	82.8709
7	1.002 719 20	0.000 310 9	0.0099
8	12.415 524 16	0.003 649 8	74.0186
9	13.841 508 48	0.004 896 4	144.6414

TABLE 10-9. One-day Timetable Solutions for Rise/Set of Satellites 6, 8, and 9, Respectively (for Example 10-8). The quartic blending uses *PKEPLER* with a 3 m time step. The truth simulation uses a 0.1 s stepsize. We could also present the corresponding range, azimuth, and elevation values for each rise and set time.

Quartic Blending		Truth Simulation		Absolute Difference	
In View (s)	Out of View (s)	In View (s)	Out of View (s)	In View (s)	Out of View (s)
354.7	648.1	354.7	648.1	0.0	0.0
706.6	714.1	706.7	714.2	0.1	0.1
41,430.9	41,689.1	41,431.0	41,689.2	0.1	0.1
47,617.3	48,003.7	47,617.4	48,003.8	0.1	0.1
84,436.3	84,650.8	84,436.3	84,650.8	0.0	0.0
356.2	644.7	356.2	644.8	0.0	0.1
7,539.8	7,911.0	7,539.9	7,911.0	0.1	0.0
37,783.3	37,855.8	37,783.3	37,855.9	0.0	0.1
44,244.3	44,820.3	44,244.5	44,820.3	0.2	0.0
83,997.9	84,052.9	83,988.2	84,052.9	0.3	0.0
52,035.4	52,284.1	52,035.5	52,284.2	0.1	0.1
58,962.6	58,299.9	57,962.7	58,300.0	0.1	0.1
63,878.0	64,219.2	63,878.0	64,219.3	0.0	0.1
69,725.3	70,114.0	69,725.3	70,114.1	0.0	0.1
75,511.5	75,981.5	75,511.5	75,981.5	0.0	0.0

Problems

1. Can η_{FOV} equal Λ_{FOV} ? If so, how?
2. For Example 10-3, suppose $i = 130^\circ$. How do your answers change? Consider nodal, perigee, and drag corrections.
3. You're tracking a satellite with orbital elements equal to

$$\hat{r}_{IJK} = 191.955\,268 \hat{i} + 1069.525\,308 \hat{j} + 7098.497\,371 \hat{k} \text{ km}$$

$$\hat{v}_{IJK} = 7.206\,807\,1 \hat{i} - 1.780\,491\,4 \hat{j} + 0.060\,407\,9 \hat{k} \text{ km/s}$$

February 15, 1997, 01:30:55.0 UTC

After you've lost the satellite, your boss hands you its elements at a future time: 1 day, 1.539 936 9 hours later,

$$\hat{r}_{IJK} = 6078.803\,133 \hat{i} - 754.062\,501 \hat{j} + 3748.213\,782 \hat{k} \text{ km}$$

$$\hat{v}_{IJK} = 3.641\,897\,3 \hat{i} - 1.770\,935\,1 \hat{j} - 6.256\,019\,3 \hat{k} \text{ km/s}$$

You suspect a maneuver has occurred. What is the Δv , when did it occur, and what's the change in period? Explain your answer, especially if the trajectories don't intersect.

4. On your first assignment, you've been tasked to tell the boss when your building at Omaha, Nebraska (41.16°N , 95.57°W) will be visible to a certain satellite. It's Friday, and you have the current osculating state vector. The general needs the answer by 09:00 Monday morning. How do you proceed?
5. You're asked to design a new satellite-reconnaissance orbit for the satellite in Example 10-3 through Example 10-5. You have limited fuel to make the necessary periodic corrections and decide to try the following options to reduce your total change in velocity for a 60-day mission:

a. An elliptical orbit ($e = 0.3$).

b. Use improved optics to change the required perigee altitude to 250 ± 40 km

As an additional exercise, determine the fuel requirements using the initial problem parameters (Example 10-3), but calculate the answers for a Sun-synchronous orbit (0650 h), and a frozen eccentricity orbit ($e = 0.01$). Be sure to support your answers with numerical values.

6. You're in charge of a company trying to design a new sensor for use in space surveillance. You have a good design, but want to be sure you select the correct location for the sensor. The blending and splining techniques used in this chapter provide an idea to develop a *site-visibility* program that takes into account the average time certain satellites will be visible to a sensor. Alfano, Negron, and Wright presented a solution in 1992, and Alfano and Negron extended the solution in 1993 to include multiple restrictions from the sensor site. Program the technique and evaluate its accuracy against long simulations of the Rise/Set algorithm presented in this chapter.

7. A large company is considering its requirements for future systems to do surveillance tracking in space. After an extensive review, they conclude they'll be able to complete all their surveillance tasks, as long as they build three new highly accurate radars and one new interferometer "fence" system. Assume the data from Fig. 10-2 to Fig. 10-4 represents the distribution of satellites. Find the best locations for the new sensors that will minimize tasking. Assume the fence will occupy the same position at about 33°N. Hint: assume differential-correction processing of the data is accurate, including numerical propagations, dense observations, treatment of calibration coefficients, and current solar parameters.
8. Continuing with the concept in Problem 7, explain how you could minimize the number of uncorrelated tracks for an orbit-determination system in the real world if the system must cover all orbit types. Consider whether it's: 1) better to make the standard accuracy of each satellite very accurate or inaccurate; 2) easier to detect an unknown maneuver with accurate or inaccurate ephemerides; 3) better to use all existing data for post-maneuver processing, 4) better to use a Kalman filter or a least-squares technique. Hint: use the general information on tracking systems in Chap. 9, and the information from Chap. 10 concerning the distribution of types of satellite orbits. Remember that an interferometer fence will provide moderately accurate angles-only measurements on any satellite passing through its web.
9. Following the discussion from Problem 7, assume your study indicates the sensors need even higher accuracy. You've worked with NASA's satellite-laser-ranging (SLR) sites and think they will do the job. Unfortunately, you must be careful about using lasers to obtain ranging data from a satellite. In particular, you don't want to hit another satellite if you miss the one you're trying to track. This problem is known as the *laser clearinghouse* problem, and it can be solved by brute force—propagating each satellite and determining line-of-sight information at each time step. Unfortunately, this is very inefficient and requires significant computer resources. Your task is to develop a new laser-clearinghouse algorithm using the blending and splining functions presented and used in this chapter. This will permit much larger time steps for the propagation techniques, free the technique from any particular propagator, reduce computer resources, and speed execution and accuracy. In particular, pay close attention to the function you derive for the blending and splining techniques.
10. You're a defense attorney representing a company whose satellite is accused of colliding with GALAXY 4 and knocking out television coverage just as Denver was about to win the Super Bowl on January 26, 1997 at 16:00 Mountain Standard Time. Information consists of element sets for each satellite, about eight hours before the alleged event. The osculating state vectors for each satellite are as follows. Each satellite has a spherical uncertainty region with minor axes equal to ± 25 m, RMS.

Defense Satellite: January 25, 1997, 12:47 UTC

$$\dot{\vec{r}}_{IJK} = 31,497.390\,167 \hat{i} + 28,026,945\,431 \hat{j} + 276,112\,685 \hat{k} \text{ km}$$

$$\dot{\vec{v}}_{IJK} = -2.044\,055 \hat{i} + 2.296\,802 \hat{j} - 0.017\,964 \hat{k} \text{ km/s}$$

GALAXY 4: January 18, 1997, 13:43:45.402 UTC

$$\dot{\vec{r}}_{IJK} = 28,478.560\,600 \hat{i} + 31,090.200\,400 \hat{j} + 4.008\,900 \hat{k} \text{ km}$$

$$\dot{\vec{v}}_{IJK} = -2.267\,700 \hat{i} + 2.076\,500 \hat{j} + 0.000\,300 \hat{k} \text{ km/s}$$

Assume the jury will comprehend and act on factual information—not court-room drama. If you can win the case, how do you proceed? If not, how will the prosecution present the true facts of the case?

11. You're launching a new satellite and want to evaluate the potential for close approaches with other satellites. Using the following data for the initial placement in orbit, should your launch proceed?

$$a = 7047.9490 \text{ km}$$

$$a = 7048.0192 \text{ km}$$

$$e = 0.000\,899$$

$$e = 0.001\,470\,2$$

$$i = 98.1025^\circ$$

$$i = 98.4535^\circ$$

$$\Omega = 141.7519^\circ$$

$$\Omega = 334.7433^\circ$$

$$\omega = 67.4104^\circ$$

$$\omega = 119.3840^\circ$$

$$M = 292.8048^\circ$$

$$M = 240.8797^\circ$$

$$\text{Epoch} = \text{July 23, 1997}$$

$$\text{Epoch} = \text{July 23, 1997}$$

$$09:25:32.358 \text{ UTC}$$

$$08:15:32.965 \text{ UTC}$$

APPENDIX A

DICTIONARY OF SYMBOLS AND ACRONYMS

- A.1 Dictionary of Symbols
- A.2 Dictionary of Acronyms

A.1 Dictionary of Symbols

The many “standard” symbols in astrodynamics have been in use for centuries. I attempt to keep many of these common symbols, give alternatives wherever appropriate, and stay consistent throughout the book. Some symbols have duplicate meanings, but the appropriate meaning should be apparent from the context. I use a few major notations throughout the book.

Mean Values	over bar	\bar{r}
First Derivative	dot	\dot{r}
Second Derivative	double dot	\ddot{r}
Third Derivative	triple dot	\dddot{r}
Vector	angular symbol	\hat{r}
Unit Vector	hat	$\hat{\mathbf{r}}$
Matrix	bold	\mathbf{r}
Matrix (predicted)	over bar	$\bar{\mathbf{r}}$
Matrix (best estimate)	hat	$\hat{\mathbf{r}}$

The following list of symbols is alphabetical—lowercase, then uppercase; Arabic, then Greek letters.

a	Semimajor axis
\dot{a}	Acceleration vector (often with subscripts)
a_i	$i = 1 \dots 4$ intermediate range-only
\hat{a}_i	Osculating orbital elements
a_i	Satellite's initial acceleration magnitude
a_i, a_{iu}	Coefficients of Gaussian initial orbit determination ($i = 1, 3$)
a_p	3-hourly value of the geomagnetic planetary amplitude
a_v	Vehicle acceleration (low-thrust)
b	Semi-minor axis
b	Observation matrix (observations)
\tilde{b}	Residual matrix (observed-nominal)

c	Chord length between two vectors
c	Constant in energy equation, assumed to be zero for astrodynamics
c	Half the distance between foci in a conic section
c	Speed of light
c_i	Coefficients of techniques for determining initial orbits ($i = 1, 2, 3$)
c_i	Arbitrary orbital element
c_2, c_3	Functions of universal variable
c_D	Coefficient of drag
c_R	Coefficient of reflectivity (used with solar-radiation calculations)
c_v	Control variable (low-thrust)
d	General distance
d_i	Intermediate values in Gaussian angles-only orbit determination
e	Eccentricity
e_e	Eccentricity of Earth's orbit around the Sun
e_F	Eccentric for minimum eccentricity orbit (fundamental ellipse)
e_{\oplus}	Eccentricity of Earth's ellipsoidal approximation (see also flattening f)
el	Elevation, measured positively from the local horizontal
f	f function—Kepler problem
f	Flattening coefficient for the Earth (see also e_{\oplus})
f	Parameter in Encke's formulation
f	Frequency of electromagnetic radiation
f	General mathematical function notation, usually with an argument e.g. $f(x)$
f_d	Distance function
f_d	Doppler shifted frequency
f_i	Power Series of F_i
f_r	Reference frequency
f_r	Retrograde factor for equinoctial elements (+1 for direct orbits, -1 for retrograde orbits)
g	Gravity
g	g function—Kepler problem
$g_{equator}$	Equatorial value of gravity
g_{im}	Gaussian geomagnetic coefficient
g_{normal}	Normal component of gravity
g_p	Poincaré orbital element
g_{pole}	Polar value of gravity

g_{th}	Theoretical gravity, sometimes with a descriptor as in (th -84) for WGS-84 values
h	Specific angular momentum
h	Step size
h	Intermediate argument in Gaussian initial orbit determination
h	Intermediate parameter in third-body accelerations
h_e	Equinoctial element
h_{ellp}	Height above the reference ellipsoid
h_{lm}	Gaussian geomagnetic coefficient
h_o	Base altitude
h_{sys}	Specific angular momentum for n -body problem, perpendicular to the invariable plane
i	Inclination
i	Index (general)
j	Index (general)
k_e	Equinoctial element
k_g	Intermediate value used in calculating theoretical gravity
k_i	Intermediate values of numerical integrators
k_l	Love numbers
k_p	Geomagnetic planetary index
k_{int}	Interceptor satellite's additional revolutions
k_{tgt}	Target satellite's additional revolutions
l	Index (general)
t	Gravitational index
t	Intermediate value of Gaussian initial orbit determination
l_p	Poincaré orbital element
m	General mass; usually with a subscript added
m	Intermediate value of Gaussian initial orbit determination
m_p	Propellant mass fraction
m_{sat}	Mass of satellite
m_\oplus	Mass of the Earth
\dot{m}	Specific mass flow rate (actual mass flow rate \div initial mass)
n	Mean motion
\hat{n}	Node vector, points to ascending node
\bar{n}	Mean mean motion
n_p	Parabolic mean motion

p	Semiparameter
p_e	Equinoctial orbital element
p_o	Pressure
$p_{parabola}$	Parabolic semiparameter
p_{SR}	Solar-radiation pressure
Δp	Change in pressure
q_e	Equinoctial element
r	Position vector's magnitude
r_a	Radius of apogee
r_d	Relative distance function
r_{op}	Parabolic mean radius in Battin's initial orbit determination
r_p	Radius of perigee
r_p	Radius of primary
r_s	Radius of secondary
r_K	Position vector's vertical component
r_δ	Position vector's equatorial component
\dot{r}	Radial component of velocity (<i>not</i> the velocity magnitude)
\bar{r}_i	Residual in differential correction; observed minus computed value
\mathbf{r}_{site}	Matrix of three site position vectors
s	Ground range
s	Intermediate variable (minimum inclination and the parabolic time of flight)
s	Semiperimeter
s	Step-size adjustment parameter
s	Time regularization variable
t	General time
t_{acc}	Accumulated time
t_{drift}	Time to drift between maneuvers for specialized orbit design
t_m	Transfer direction of motion between two position vectors (+1, $\Delta v < 180^\circ$) short way, (-1, $\Delta v > 180^\circ$) long way
t_p	Parabolic flight time between two positions
t_s	Switching time (continuous thrust)
u	Argument of latitude (circular inclined orbits); positive in direction of satellite's motion
u	Argument of Taylor series of f and g series development
u	Truncation error in development of process noise

u_x, u_y	Instantaneous position of the North Pole (x / y)
u_M	Argument of latitude (circular inclined orbits) using mean anomaly
v	Magnitude of satellite's velocity
ν	Uncertainty in dynamics model (process noise)
v_{acc}	Accumulated velocity
v_{esc}	Escape velocity
v_L	Velocity of the launch site
v_R	Radial velocity in the <i>RSW</i> system
v_S	Tangential velocity in the <i>RSW</i> system
v_∞	v -infinity; velocity at the sphere of influence
w	Intermediate angle for calculating the parabolic time of flight
w	Weighting value for least squares, reciprocal of the value for standard deviation
x	Radial position (relative motion)
x	Independent value in differential correction
x	Component of position vector in <i>P</i> -axis direction
x_p	Polar motion component, measured positively from the North pole along $\lambda = 0^\circ$
x_1, x_2	Intermediate values of Gaussian initial orbit determination
y	Sector-to-triangle ratio of Gaussian initial orbit determination
y	Component of position in direction of velocity vector (relative motion)
y	General value of the function in differential correction and numerical integration
y	Component of position vector in <i>Q</i> -axis direction
y_c	Vertical component of a circle
y_e	Vertical component of an ellipse
y_h	Vertical component of a hyperbola
y_{he}	Vertical component of a equatorial hyperbola
y_p	Polar motion component, measured positively from the North pole along $\lambda = 90^\circ \text{W}$
z	Rotation angle for precession
z	Cross-plane position (relative motion)
A	Avogadro number
A	General area, usually with a subscript
A	Cross-sectional area-drag

A	Partial derivative matrix
A_{ij}	Mean-elements rate of change
A_{bm}	Intermediate value for third-body potential function
A_p	Daily value of the geomagnetic planetary amplitude
A_{SR}	Cross-sectional area—solar radiation
A_\odot	Cross-sectional area—solar-radiation pressure
B	Parabolic anomaly
B	Intermediate vector in Gibbsian orbit determination
B	Intermediate parameter in third-body potential derivation
B	Summation of Legendre polynomials in third-body acceleration
\vec{B}	Magnetic field
\vec{B}	Vector constant of integration ($= \mu e$) used in deriving the trajectory equation
B^*	Ballistic coefficient and perigee density (See also BC)
B_{bm}	Intermediate value for third-body potential function
C	Dot product of middle line-of-sight unit vector and middle site vector
C	Jacobian constant used in the three-body problem
C	Constant in the Hills relative motion problem
C	$\hat{L}_2 \cdot \vec{r}_{site2}$ in Gaussian angles-only orbit determination
C	Intermediate value in double- r iteration
C_\oplus	Auxiliary variable for site determination
$C_{2,0}$	Unnormalized zonal harmonic of degree 2, gravitational coefficient
C_{lm}	Unnormalized gravitational coefficient, degree l , order m
\bar{C}^*_{lm}	Normalized gravitational coefficient used for geoid undulation calculations, degree l , order m . Even zonal harmonics minus geometric value.
D	Determinant value
D	Constant in the Hill's relative motion problem
D	Intermediate value in Gibbsian orbit determination
D	Diameter of sensor aperture
D_\odot	Mean elongation of the Sun, positive east
D_\odot	Mean elongation of the Moon, positive east
$D_{1,2}$	Middle position determinant, Gaussian initial orbit determination
$D_{3,4}$	Middle velocity determinant, Gaussian initial orbit determination
E	Eccentric anomaly, positive in direction of satellite's motion
E	Complete elliptic integral of the second kind
$E()$	Expected value

E_i	Intermediate vector in range-only orbit determination
F	Forces
F	Mean elongation of the Moon
F, F'	Foci of a conic section
F	Partial derivative matrix (Kalman filter)
F	Inclination functions of Kaula gravitational potential
F_g	Force of gravity
F_i	Intermediate parameter of double- r iteration ($i = 1 \dots 4$)
F_i	VOP Perturbing function
F_{tmp}	Inclination function
$F_{10.7}$	Solar flux value from 10.7 cm radiation
F_{EUV}	Solar flux value from extreme ultraviolet radiation
G	Gravitational constant
G	Eccentricity functions of Kaula gravitational potential (also called Hansen coefficients)
G_{tmp}	Eccentricity function
G_p	Poincaré element
H	Hyperbolic anomaly
H	Scale height
H	Observation matrix at current time
\mathcal{H}	Hamiltonian
H_d	Delaunay element
$H_{pq'}$	Third-body formulation of Hansen coefficients
H_p	Poincaré element
H_{MSL}	Height above mean sea level
I	Identity matrix
I	Polar moment of inertia
I_{sp}	Specific impulse (usually in seconds)
J_i	Zonal harmonic coefficient ($i = 2$, etc.)
J_i	Modified Bessel function of the first kind
J	Cost Function, estimation
K	Kalman gain matrix
K	Complete elliptic integral of the first kind
K_p	Planetary index (magnetic field)
\hat{L}	Line-of-sight vector, usually with a subscript; L is a matrix of three \hat{L}
L_d	Delaunay element

L_e	Equinoctial element
L_g	Intermediate variable in Gibbs formulation
L_i	Lagrange point for the three-body problem ($i = 1 \dots 5$)
L_p	Poincaré element
M	Mean anomaly, positive in the direction of satellite's motion
M	Molecular mass, general mass
M	Matrix of $(L^{-1}r_{site})$ in Gaussian angles-only orbit determination
M	Matrix of slant-range vectors in range-only orbit determination
M_{fade}	Matrix of fading-memory values for differential correction
M_\oplus	Mean anomaly of the Earth, positive in the direction of satellite's motion
M_\odot	Mean anomaly of the Sun, positive in the direction of satellite's motion
N	Number of observations in statistical orbit determination
N	Ambiguity
N	Intermediate vector in Gibbsian orbit determination
N_\oplus	Geoid undulation height
N_ϕ	Radius of curvature in meridian—see also C_\oplus
P	Parabolic anomaly
P	Integer parameter in resonance calculations
\wp	Orbital period (Keplerian)
$P_{l,0}$	Legendre polynomials, degree l
$P_{l,m}$	Associated Legendre functions, degree l , order m
ρ_{res}	Resonance period
ρ_K	Anomalistic period (also Keplerian period)
ρ_Ω	Nodal period (also nodal period of Greenwich with subscript G)
\hat{P}	Predicted covariance matrix
$\hat{\hat{P}}$	Best estimate of covariance
Q	Accuracy estimate of double-r iteration
Q	Integer parameter in resonance calculations
Q	Power density matrix resulting from process noise
Q	Matrix used in Lambert's time formulation
R	Disturbing potential ($U - U_{2\text{-body}}$)
R	Measurement noise matrix, also the inverse of W matrix
R	Universal gas constant

R	Ratio of orbit's final radius over its initial radius
R_{\oplus}	Radius of Earth
R^*	Ratio of orbit's intermediate radius over its final radius (bi-elliptic transfers)
$[R]$	Rotation matrix
S	Intermediate value in double- r iteration
S	Alternate matrix for square root formulations
S_{lm}	Unnormalized gravitational coefficient
S_{bnpq}	Kaula gravitational function
$\tilde{S}_n, S_n^{\text{II}}$	First and second summed back differences
S_{\oplus}	Auxiliary function for site determination
T	Time of perigee passage
T	Lambert's time function
T	Absolute temperature (K)
T_c	Night-time temperature of the exospheric atmosphere
T_{TDB}	Julian centuries of barycentric dynamical time, TDB
T_{TDT}	Julian centuries of terrestrial dynamical time, TDT
T_{TAI}	Julian centuries of atomic time, TAI
T_{UT1}	Julian centuries of universal time, $UT1$
$U_{2\text{-body}}$	Gravitational potential (two-body)
U	Gravitational potential (aspherical), includes two-body portion
V	Magnetic field potential
W	Weighting matrix
\bar{X}	Predicted state vector; sometimes just position and velocity vector, sometimes with solve-for parameters
\hat{X}	Best estimate of the state vector
Z_{ij}	Cross-product of vectors i and j
α	General angle
α	Reciprocal of semimajor axis (Kepler's problem in universal variables)
α	Right ascension, measured positively to the east
α_{xx}	Coefficients for <i>non-summed</i> numerical integration methods (Adams-Basforth AB , Adams-Moulton AM , Störmer S , Cowell C)
α_{cop}	Coplanar angle measuring amount of separation, given three vectors
α_{DMS}	Degree-minute-second angular value in radians or degrees
α_L	Lead angle, positive in the direction of the target satellite's motion

α_e, α_h	Intermediate angles (elliptical and hyperbolic) used in Lambert problem
α_t	Topocentric right ascension, measured positively to the east
α_{12}, α_{23}	Separation angle between two vectors
β	Azimuth, measured positively east from north
β	General angle
β_e, β_h	Intermediate angles (elliptical and hyperbolic) used in Lambert's problem
β_k	Intermediate summation for approximation of third-body acceleration
β_{xx}	Coefficients for <i>summed ordinate</i> numerical integration methods (Adams-Bashforth AB, Adams-Moulton AM, Störmer S, Cowell C)
χ	Universal variable
δ	Declination, measured positively north from the equator
δ	General parameter representing a small change to a variable
δ_\odot	Declination of the Sun, measured positively north
δ_b	Back difference values of numerical integrators
δ_t	Topocentric declination, measured positively north from the local equator
$\bar{\delta x}$	Predicted state correction
$\hat{\delta x}$	Best estimate for state correction
ϵ	Obliquity of the ecliptic (mean)
ϵ	General small parameter
ϵ	Small parameter in Encke's formulation
$\bar{\epsilon}$	True obliquity of the ecliptic
ϵ_{impq}	Phase lag
ϵ_Θ	Resonance parameter
$\Delta\epsilon$	Change in obliquity of the ecliptic, used in nutation calculations
ϕ	General latitude, specified with subscripts
ϕ_{as}	Astronomical latitude, measured positively north from the equator
ϕ_c	In-plane control angle, positive in the radial direction
$\phi_{ecliptic}$	Ecliptic latitude, measured positively north from the ecliptic
ϕ_{fpa}	Flight-path-angle, measured positively from the local horizontal to the velocity vector
ϕ_{gc}	Geocentric latitude, measured positively north from the equator
ϕ_{gd}	Geodetic latitude, measured positively north from the equator
ϕ_{inc}	Incidence angle for radiation
ϕ_{rd}	Reduced latitude, measured positively north from the equator

ϕ_{SR}	Solar radiation angle for reflection
γ	Payload axis angle; positive if the new inclination is larger than the original
γ	Intermediate angle in topocentric right-ascension and declination conversions
η	Intermediate parameter in Battin's initial orbit determination
η_{center}	Sensor field of view to the center
$\tilde{\eta}_{FOV}$	Total sensor field of view angle
η_{FOV}	Sensor field of view angle
$\eta_{horizon}$	Sensor field of view to the horizon
η_i	Arbitrary vector component ($i = 1, 2, 3$)
η_{ij}	Short-periodic variations
φ	Turning angle for hyperbolic orbits
κ_{\odot}	Gaussian constant of gravitation for the Sun
λ	Longitude, measured positively east from the Greenwich meridian
λ	Wavelength of electromagnetic radiation
λ	Temporary variable for analyzing parabolic time of flight
$\Delta\lambda$	Change in terrestrial longitude for a groundtrack, relative to the starting location
$\lambda_{ecliptic}$	Ecliptic longitude
λ_{true}	True longitude at epoch (circular equatorial orbits) measured in one plane; positive
λ_u	Longitudinal component of satellite position, measured positively from the ascending node to the longitude of a satellite's subpoint
$\lambda_{vR}, \lambda_{vS}$	Lagrange multipliers in orbit raising (λ dynamics)
λ_M	Mean longitude at epoch (circular equatorial orbits) measured in two planes ($\Omega + \omega + M$)
$\lambda_{M\odot}$	Mean longitude for the Sun
$\lambda_{M\odot}$	Mean longitude for the Moon
λ_S	Groundtrack shift parameter
$\lambda_{S+}, \lambda_{S-}$	Maximum and minimum values of groundtrack shift parameter
λ_{\odot}	Celestial longitude for the Sun
μ	Gravitational parameter, usually with a subscript
μ^*	Mass ratio for the restricted three-body problem
μ_{xy}	Correlation coefficients between x and y
μ_{\oplus}	Gravitational parameter for the Earth
μ_{\odot}	Gravitational parameter for the Sun

ν	True anomaly
θ	General angle (complement of ϕ_{fpa} , angular separation angles, etc.)
θ_p	Polar angle of satellite in low-thrust transfer
θ_{AST}	Apparent sidereal time, measured positively to the east from the Greenwich meridian along true equator (also <i>AST</i>)
θ_{BW}	Beamwidth
θ_{GST}	Greenwich mean sidereal time, measured positively to the east from the Greenwich meridian (also <i>GST</i>)
θ_{LST}	Local mean sidereal time, measured positively to the east from the site (also <i>LST</i>)
ρ	Range magnitude (site to satellite)
ρ	Atmospheric density
$\hat{\rho}$	Range vector (site to satellite)
ρ_n	Density of the atmosphere at night
σ	Standard deviation, usually with a subscript for estimation theories
σ	Rate parameter fro mean anomaly
τ	Scaling parameter in the SIGHT problem
τ_i	Delta time between a given time (t) and a middle time ($i = 1, 3$)
τ_{phase}	Time in a phasing orbit
τ_{min}	Location of minimum distance in SIGHT algorithm
τ_{wait}	Wait time for a rendezvous
τ_{trans}	Transfer time of flight
τ_{HMS}	Time of hour-minute-seconds in radians or degrees
τ_{UT}	Universal time in a seconds or HHMM.SSSSSS format
ω	Argument of periapsis; positive in the direction of a satellite's motion
ω_{int}	Angular velocity of a satellite—usually specifies interceptor or target
ω_S	Rotation of synodic frame with respect to barycentric frame
ω_\oplus	Rotational velocity of the Earth
$\tilde{\omega}$	Longitude of periapsis, measured in two planes, in the direction of a satellite's motion ($\Omega + \omega$)
$\tilde{\omega}_{true}$	True longitude of periapsis, measured in one plane, in the direction of a satellite's motion
$\tilde{\omega}_M$	Mean longitude of periapsis, using the mean anomaly, measured in one plane in the direction of satellite motion
ξ	Specific mechanical energy (see also SME)
ξ_i	Arbitrary vector component ($i = 1, 2, 3$)
ξ_{sys}	Specific mechanical energy for n -body problem

$\xi(x)$	Continued fraction in Battin's initial orbit determination
ψ	Universal variable argument
$\Delta\psi$	Change in longitude (nutation)
ζ	Precession angle
Φ	State transition matrix
ϑ	Angle of Hyperbolic asymptotes
ϑ	Angle between normal component of velocity vectors in maneuver calculations (positive in direction of the target satellite's motion)
ϑ	Phase angle (rendezvous calculations) from target satellite to interceptor satellite. Positive sense is in direction of the target satellite's motion.
ϑ	Thrust control angle in low thrust. Also vehicle pitch angle
Π_{lm}	Conversion term for normalizing gravitational coefficients
$\Pi_{lmSchmidt}$	Conversion term for normalizing geomagnetic coefficients with Schmidt's method
Θ_{lmpq}	Kaula's gravitational argument
$\Theta_{3\text{-body}}$	Third-body gravitational argument
Λ	Total range angle
$\tilde{\Lambda}_{FOV}$	Total sensor ground range visible to a sensor
Λ_{FOV}	Sensor ground range visible to a sensor
Θ	Rotation angle for precession
Θ	Angle between velocity vectors in maneuver calculations
Σ_i	Summation of previous function values for numerical integrators
ς	Angular distance between the Sun and satellite vectors (solar radiation-pressure and PREDICT problem). Specified with a type—civil, nautical, or astronomical—when used as a sunrise sunset parameter
Ω	Longitude of ascending node; positive east from the I axis.
Ω_c	Longitude of ascending node of the Moon; positive from the I axis
Ψ	Precession angle
Υ	Vernal equinox
ρ	Parallax angle, sometimes with a subscript for geocentric, horizontal, or heliocentric
$\nabla_i()$	Individual back values based on a function—the satellite acceleration
\Rightarrow	Shows we're using an algorithm described elsewhere, or the result of an algorithm
\Leftarrow	Shows replacement of a value during an iteration

A.2 Dictionary of Acronyms

A few items occur as both symbols and acronyms. Again, I've tried to be consistent in identifying these acronyms.

AAS	American Astronautical Society
ABC	Theoretical perfectly inertial coordinate system
AFSCN	Air Force Satellite Control Network
AIAA	American Institute of Aeronautics and Astronautics
ALCOR	ARPA Lincoln C-Band Observables Radar
ALTAIR	ARPA Long Range Tracking and Instrumentation Radar
ARPA	Advanced Research Projects Agency
AMOS	Air Force Maui Optical Station
A/S	Anti-Spoofing
AST	Apparent sidereal time; positive to the east from the vernal equinox along the true equator (also θ_{AST})
ANCAS	Alfano-Negrone Close Approach Software
ΔAT	Difference of <i>TAI</i> minus <i>UTC</i>
ATAN2	Mathematical procedure which determines the correct quadrant for a trigonometric operation using the sine and cosine arguments.
AU	Astronomical unit. Mean distance of the Earth from the Sun.
BC	Ballistic coefficient ($m/c_D A$) (see also B^*)
BIH	Bureau International de l'Heure
B1950	Coordinate system using the <i>FK4</i> star catalog
C/A	Coarse Acquisition
CCD	Charge-coupled device
CEP	Celestial ephemeris pole
CIO	Conventional International Origin (locates the North pole)
CIS	Conventional inertial system (See also <i>ECI</i>)
DGPS	Differential GPS
DMS	Degree-minute-second
DMSP	Defense Meteorological Support Program
DOP	Dilution of Precision
DSCS	Defense Support Communication Satellite
DSN	Deep Space Network
DSP	Defense Support Program
DSST	Draper Semianalytical Satellite Theory

DU*	Distance Unit normalized to the ratio of orbits (R)
ECEF	Earth-centered, Earth-fixed coordinate system. Fixed to a meridian of longitude (see also <i>EFG</i>)
ECI	Earth-centered inertial coordinate system. Often used interchangeably with <i>IJK</i> . Identified with a subscript for precise applications—J2000, mod-mean of date, tod-true of date, wPM-with polar motion, w/oPM-without polar motion.
ED	Elapsed days
EFG	Earth-fixed Greenwich coordinate system. Fixed to the Greenwich meridian
EKF	Extended Kalman filter
EQW	Equinoctial coordinate system
$EQ_{equinox}$	Equation of equinoxes
EQ_{time}	Equation of time
ER	Earth radius
ET	Ephemeris time
EUV	Extreme ultraviolet (radiation)
FK4	Star catalog which the B1950 reference frame uses
FK5	Star catalog which the J2000 reference frame uses
FMS	Fictitious mean Sun
FRAC	Fraction (decimal) of a real number (i.e., FRAC (1.57) = 0.57)
GEM-#	Goddard Earth Model with a designator (e.g., 10)
GEO	Geosynchronous Earth Orbit
GEODSS	Groundbased Electro-Optical Deep Space Surveillance
GEOS	Geodetic Earth Orbiting Satellite
GFZ	<i>Geo Forschungs Zentrum</i> , Laser geodetic satellite for Germany
GHA	Greenwich hour angle, usually with a descriptor, measured positively to the <i>west</i> from the Greenwich meridian
GLONASS	Global Orbiting Navigation Satellite System
GMT	Greenwich mean time, <i>not UT</i>
GOES	Geostationary Operational Environmental Satellite
GPS	Global Positioning System
GST	Greenwich mean sidereal time, measured positively to the east from the vernal equinox (see also θ_{GST})
HMS	Hour, minute, second
IERS	International Earth Rotation Service
IJK	Geocentric equatorial coordinate system

IPMS	International Polar Motion Service
IJK _t	Topocentric equatorial coordinate system
INT	Truncation of a real number (i.e., INT (1.57) = 1.0)
ITRF-##	International Terrestrial Reference Frame associated with a year, e. g. 90
JD	Julian date, often with a subscript for the type of time (<i>UTC</i> , <i>UT1</i> , <i>TDB</i>)
J2000	Coordinate system at January 1, 2000, using the <i>FK5</i> star catalog
JGM-#	Joint Gravitational Model, with revisions identified, 2, 3, ...
KE	Kinetic energy
KF	Kalman filter
LACE	Low Power Atmospheric Compensation experiment
LAGEOS	Laser Geodynamics Satellite
LANDSAT	Earth Resources Technology Satellite
Laser	Light Amplification by Stimulated Emission of Radiation
LEO	Low-Earth Orbit
LHA	Local hour angle (usually with a descriptor—Sun, Moon, etc.) measured positively to the west from the site
LOS	Line of sight
LST	Local mean sidereal time, measured positively to the east from a particular site (see also θ_{LST})
MAX	Maximum value from a set (e.g., MAX(3.2, 1.5, 6.3) = 6.3)
MIN	Minimum value from a set (e.g., MIN(3.2, 1.5, 6.3) = 1.5)
MJD	Modified Julian date (<i>JD</i> − 2,400,000.5)
MOD	Modulus Operation. For example, 8 MOD 3 = 2 and 9 MOD 3 = 0
MSL	Mean sea level
MST	Mean solar time
NASA	National Aeronautics and Space Administration
NOAA	National Oceanic Atmospheric Administration
NTW	Normal tangential coordinate system
PAM	Payload assist module
PE	Potential energy (See also $U_{2\text{-body}}$, U)
POE	Precision Orbit Ephemerides
PRN	Pseudo-random noise
PQW	Perifocal coordinate system (periapsis, normal)
PV	Distance from primary to vertex in eclipse calculations
RADAR	Radio Detecting and Ranging
RADARSAT	Canadian observation satellite

RMS	Root-mean-square
ROT#	Rotation of a vector through an angle. The axis is usually identified as 1, 2, or 3.
ROT	Rotation through a series of separate rotations as in a coordinate transformation
ROUND	Approximation of a real number to nearest integer value (i.e., $1.57 \Rightarrow 2$)
RSSS	Russian Space Surveillance System
RSW	Satellite coordinate system (radial, direction of velocity, normal)
SEZ	Topocentric horizon coordinate system
SF	Solar Flux
SFU	Solar flux unit
SI	Système International
SIGN	Returns the sign of an input number, for example $\text{SIGN}(-157.93) = -1.0$
SLR	Satellite Laser Ranging
SOI	Sphere of influence
SSN	Space Surveillance Network
TAI	International atomic time
TCB	Barycentric coordinate time
TCG	Geocentric coordinate time
TDB	Terrestrial barycentric time
TDRS	Tracking and Data Relay Satellite. An extra S at the end refers to the complete system of TDRS satellites (Tracking and Data Relay Satellite System).
TDT	Terrestrial dynamical time
TOPEX	Ocean Topography Experiment
TRUNC	Truncation to nearest integer value towards zero. [e.g., $\text{TRUNC}(1.57 = 1)$]
TT	Terrestrial time
TU	Time unit, also called a Herg in some applications
TU*	Time Unit normalized to the ratio of orbits (R)
UI	Length of update interval
USNO	US Naval Observatory
UT	Universal time (general). Specified for precise applications.
UTC	Coordinated universal time, “clock” time
UT0	Observation of solar time at a particular site
UT1	UT0 corrected for longitudinal variations; most common
UT2	UT1 corrected for seasonal variations; not very common anymore

$\Delta UT1$	Difference of $UT1$ minus UTC
VLBI	Very long baseline interferometry
VOP	Variations of parameters
WGS-##	World Geodetic Survey with two digits indicating the year completed, e.g., -66, -72, -84, etc.
WWV	Worldwide broadcast station for time
WWVH	Worldwide broadcast station for time in Hawaii
XYZ	Heliocentric coordinate system (can have its origin at the barycenter in which case we use a subscripted B)
xyz	Coordinate system for Hills, barycentric (B), synodic (S). Individual components are labeled x , y , and z .
YMD	Year, month, day

B.1 Jacchia-Roberts Atmosphere

B.2 Soviet Cosmos Atmosphere

We present two representative atmospheric models to show how complex it is to model atmospheric density. The Jacchia-Roberts formulation is very common, but also very complex. The Soviet model produces remarkably accurate answers, with much less computational requirements.

B.1 Jacchia-Roberts Atmosphere

Several technical papers present the theory for the Jacchia atmosphere, but Long et al. (1989, 4-35 to 4-56) give us a clear integrated discussion. From Long et al. we determine the atmospheric density using a modified analytical expression of the Jacchia-Roberts theory. The overall approach is to model the atmospheric temperature. Corrections for molecular composition and solar activity are added to the basic temperature formula, and the density is obtained for various regions in the atmosphere.

B.1.1 Evaluating Temperature

First, approximate the exospheric temperature using the *nighttime global exospheric temperature*, T_c , excluding all effects of geomagnetic activity:

$$T_c \text{ (K)} = 379 + 3.24\bar{F}_{10.7} + 1.3 [F_{10.7} - \bar{F}_{10.7}]$$

where $F_{10.7}$ is the average daily solar flux at a 10.7 cm wavelength for the day of interest and $\bar{F}_{10.7}$ is an 81-day running average of $F_{10.7}$ values, centered on the day of interest. Because the effect of solar flux on atmospheric density lags one day behind the observed values, T_c calculations can (at best) use values which are one day old. In computational programs, we can calculate these values on each day of a simulation before running the program. We use the resulting value of T_c to determine the *uncorrected exospheric temperature*, T_{unc} :

$$T_{unc} = T_c \left\{ 1 + 0.3 \left[\sin^{2.2}(\theta) + (\cos^{2.2}(\eta) - \sin^{2.2}(\theta)) \cos^3\left(\frac{\tau}{2}\right) \right] \right\}$$

Now, let's find the Sun's declination, δ_\odot , local hour angle of the Sun from the upper culmination, LHA_\odot , geodetic latitude of the satellite, ϕ_{gd} , and τ ($-180^\circ < \tau < 180^\circ$):

$$\eta = \frac{|\phi_{gd} - \delta_\odot|}{2} \quad \theta = \frac{|\phi_{gd} + \delta_\odot|}{2}$$

$$\tau = LHA_\odot - 37.0^\circ + 6.0^\circ \sin(LHA_\odot + 43.0^\circ)$$

$$LHA_{\odot} = \frac{180.0^\circ}{\pi} \left\{ \frac{r_x r_J - r_y r_I}{|r_x r_J - r_y r_I|} \cos^{-1} \left(\frac{r_x r_I + r_y r_J}{\sqrt{r_x^2 + r_y^2} \sqrt{r_I^2 + r_J^2}} \right) \right\}$$

$$\phi_{gd} = \tan^{-1} \left\{ \frac{1}{(1-f)^2} \left[\frac{r_K}{\sqrt{r_I^2 + r_J^2}} \right] \right\} \quad (B-1)$$

Notice we don't iterate to determine latitude. We use Eq. (1-18). Both the satellite vector, $\hat{r}(r_I r_J r_K)$, and the Sun's unit vector, $\hat{r}_{\odot}(r_x r_I r_J)$, are in true-of-date (*ECI*_{tod}) coordinates. We actually determine LHA_{\odot} from the dot product and the two vectors for the Sun and the satellite. We're not using the standard cosine approach because we need only the *I* axis and *J* axis to determine the angle from the satellite to the Sun. The initial fraction corrects the sign of the angle. More complicated methods are available to determine the *LHA*, but the precision here doesn't need the extra accuracy.

We now correct for the geomagnetic activity and its effect on temperature. The *correction factor for exospheric temperature*, ΔT_{corr} , depends on the geomagnetic index, k_p , and is calculated for altitudes of at least 200 km. The actual value of k_p was originally set with a 6.7-hour lag, during which the molecular interactions build up and the change in density would be noticed. Newer research suggests a 3-hour delay may be more appropriate.

$$\Delta T_{corr} = 28.0^\circ k_p + 0.03 \exp^{k_p}$$

For altitudes below 200 km,

$$\Delta T_{corr} = 14.0^\circ k_p + 0.02 \exp^{k_p}$$

Thus the *corrected exospheric temperature*, T_{corr} is

$$T_{corr} = T_{unc} + \Delta T_{corr}$$

and the *inflection point temperature*, T_x , is

$$T_x = 371.6678^\circ + 0.051\,880\,6 T_{corr} - 294.3505^\circ \exp^{(-0.002\,162\,22 T_{corr})} \quad (B-2)$$

Using the base value temperature $T_o = 183^\circ$ K, Jacchia defines the empirical temperature function for altitudes below 125 km as

$$T(h_{ellp})_{0-125} = T_x + \frac{T_x - T_o}{35^4} \sum_{n=0}^4 C_n h_{ellp}^n \quad (B-3)$$

$$C_0 = -89,284,375.0 \quad C_1 = 3,542,400.0 \quad C_2 = -52,687.5$$

$$C_3 = 340.5 \quad C_4 = -0.8$$

where the coefficients c_i have units of $1 / \text{km}^i$, and i is the index (0 to 4). I've included a subscript with $T_{x(125)}$ to remind you that it's the inflection point temperature at an altitude of 125 km.

Jacchia defines the region above 125 km in altitude with an empirical, asymptotic function for temperature (remember that h_{ellp} is in km).

$$T(h_{ellp})_{125-} = T_x + \frac{2}{\pi} (T_{corr} - T_x) \times \tan^{-1} \left\{ 0.95\pi \left(\frac{T_x - T_o}{T_{corr} - T_x} \right) \left(\frac{h_{ellp} - 125}{35} \right) \left(1 + 4.5 \times 10^{-6} (h_{ellp} - 125)^{2.5} \right) \right\} \quad (\text{B-4})$$

B.1.2 Robert's Corrections to Temperature

Roberts replaced the temperature expression in Eq. (B-4) with

$$T(h_{ellp})_{125-} = T_{corr} + (T_{corr} - T_x) \exp \left\{ \left(\frac{T_x - T_o}{T_{corr} - T_x} \right) \left(\frac{h_{ellp} - 125}{35} \right) \left(\frac{l}{R_{pole} + h_{ellp}} \right) \right\} \quad (\text{B-5})$$

where the spacecraft's altitude, h_{ellp} , and the Earth's polar radius, R_{pole} , are in km, and the parameter $l = 12,315.3554$ km is by Jacchia. We can now integrate the result in closed form to analyze perturbations. While developing systems to determine orbits, scientists at the Draper Laboratory noticed that Roberts's asymptotic function made densities above 125 km disagree. Therefore, Draper Laboratory uses a least-squares curve fit for the best value of l , determined by

$$l = \sum_{j=0}^4 l_j T_{corr}^j$$

with coefficients:

$$\begin{aligned} l_0 &= 0.103\ 144\ 5 \times 10^5 \\ l_1 &= 0.234\ 123\ 0 \times 10^1 \\ l_2 &= 0.157\ 920\ 2 \times 10^{-2} \\ l_3 &= -0.125\ 248\ 7 \times 10^{-5} \\ l_4 &= 0.246\ 270\ 8 \times 10^{-9} \end{aligned}$$

Analysis by Draper Laboratory personnel show that the maximum deviation from Jacchia's values is about 6.7%. Although l is often discontinuous at $h_{ellp} = 125$ km, few (if any) satellites can remain in orbit at 125 km without significant thrusting to overcome the effects of drag!

B.1.3 Jacchia's Corrections for Density

For altitudes below 200 km, we include the *geomagnetic effect* on density:

$$(\Delta \log_{10} \rho)_G = 0.012k_p + 1.2 \times 10^{-5} \exp^{kp} \quad (\text{B-6})$$

where k_p is the geomagnetic planetary index.

We determine the number of years from 1958 using the Julian date of 1958 (2,436,204). JD_{1958} is the number of days from January 1, 1958:

$$T_{1958} = \frac{JD_{1958}}{365.2422}$$

so we can calculate the correction for *seasonal latitudinal variation* in the lower thermosphere.

$$\begin{aligned} (\Delta \log_{10} \rho)_{LT} &= 0.014 (h_{ellp} - 90) \sin(2\pi T_{1958} + 1.72) \sin(\phi_{gd}) \\ &\quad |\sin(\phi_{gd})| \exp^{-0.0013(h_{ellp} - 90)^2} \end{aligned} \quad (\text{B-7})$$

The correction for *semi-annual variations* in density use an intermediate value

$$\tau_{SA} = T_{1958} + 0.09544 \left\{ \left[\frac{1}{2} + \frac{1}{2} \sin(2\pi T_{1958} + 6.035) \right]^{1.65} - \frac{1}{2} \right\}$$

The correction is

$$\begin{aligned} (\Delta \log_{10} \rho)_{SA} &= (5.876 \times 10^{-7} h_{ellp}^{2.331} + 0.06328) \exp^{-0.002868h_{ellp}} \\ &\times \{0.02835 + [0.3817 + 0.17829 \sin(2\pi\tau_{SA} + 4.137)] \sin(4\pi\tau_{SA} + 4.259)\} \end{aligned} \quad (\text{B-8})$$

Be careful of the units because the trigonometric terms use radians.

The correction for density because of *seasonal variations of helium by latitude* is

$$(\Delta \log_{10} \rho)_{He} = 0.65 \frac{\delta_{\odot}}{\epsilon} \left[\sin^3 \left(\frac{\pi}{4} - \frac{\phi_{gd} \delta_{\odot}}{2|\delta_{\odot}|} \right) - 0.35355 \right]$$

with the obliquity of the ecliptic, ϵ , and the declination of the Sun from before.

B.1.4 Robert's Corrections to Density

Roberts (1971) noticed that atmospheric density varies significantly in the range around 90 km to 125 km. Thus, he sectioned the atmosphere into three altitude bands to account for major variations in density: 90–100 km, 100–125 km, and above 125 km. The values of density from 0 to 90 km are important mainly for calculating reentry and are usually handled by standard atmospheres (See Sec. 7.6.2).

90 to 100 km

For altitudes between 90 and 100 km, Roberts uses the Jacchia temperature profile in Eq. (B-3) and integrates with partial fractions to obtain a standard density.

$$\rho_{std} \Big|_{90-100} (h_{ellp}) = \frac{\rho_{90} T_{90}}{M_{90}} \frac{M(h_{ellp})}{T(h_{ellp})} F_1^k \text{EXP}^{kF_2} \quad (\text{B-9})$$

This equation assumes atmospheric mixing is predominant, and has several constants: the mean molecular mass at $h_{ellp} = 90$ km is $M_{90} = 28.826\ 78$ gm/mol, and the value of the density is assumed to be $\rho_{90} = 3.46 \times 10^{-6}$ kg/m³. The constant k is

$$k = \frac{-35^4 g_{SL} R_{pole}^2}{R(T_x - T_o) C_4}$$

using the gravity at sea level ($g_{SL} = 9.80\ 665$ m/s²), polar radius ($R_{pole} = 6356.766$ km), the universal gas constant ($R = 8.314\ 32$ Joules / K·mol), and C_4 from Eq. (B-3).

We determine the mean molecular mass using coefficients in Table B-1 and

$$M(h_{ellp}) = \sum_{n=0}^6 A_n h_{ellp}^n$$

TABLE B-1. Coefficients of Mean Molecular Mass. The units are the reciprocal of the index in km. For example, $n = 2$, units are /km².

Index _n	A _n	Index _n	A _n
0	-435,093.363 387	4	-0.089 587 909 95
1	28,275.564 639 1	5	0.000 387 375 86
2	-765.334 661 08	6	-0.000 000 697 444
3	11.043 387 545		

Then, we solve for F_1 and F_2 in Eq. (B-9) using quadratic expressions:

$$F_1 = \left(\frac{h_{ellp} + R_{pole}}{90 + R_{pole}} \right)^{P_1} \left(\frac{h_{ellp} - x_{r1}}{90 - x_{r1}} \right)^{P_2} \left(\frac{h_{ellp} - x_{r2}}{90 - x_{r2}} \right)^{P_3}$$

$$\times \left(\frac{h_{ellp}^2 - 2x_{r3}h_{ellp} + x_{r3}^2 + x_{i3}^2}{8100 - 180x_{r3} + x_{r3}^2 + x_{i3}^2} \right)^{P_4}$$

$$F_2 = (h_{ellp} - 90) \left[A_6 + \frac{p_5}{(h_{ellp} + R_{pole}) (90 + R_{pole})} \right] \\ + \frac{p_6}{x_{i3}} \text{TAN}^{-1} \left[\frac{x_{i3} (h_{ellp} - 90)}{x_{i3}^2 + (h_{ellp} - x_{r3}) (90 - x_{r3})} \right]$$

There are two real (x_{r1} , x_{r2}) roots, and two complex conjugate roots (x_{r3} , $\pm x_{i3}$). We take only the positive imaginary component, and calculate the roots from the polynomial function and coefficients

$$f(h_{ellp}) = \sum_{n=0}^4 C_n^* h_{ellp}^n \quad C_o^* = \frac{35^4 T_x}{C_4 (T_x - T_o)} + \frac{C_0}{C_4} \\ C_n^* = \frac{C_n}{C_4} \quad (1 \leq n \leq 4)$$

for values of C_n given by Eq. (B-3). We need other quantities to complete the next relations:

$$X^* = -2x_{r1}x_{r2}R_{pole}(R_{pole}^2 + 2x_{r3}R_{pole} + x_{r3}^2 + x_{i3}^2)$$

$$V = (R_{pole} + x_{r1})(R_{pole} + x_{r2})(R_{pole}^2 + 2x_{r3}R_{pole} + x_{r3}^2 + x_{i3}^2)$$

$$U(r_i) = (r_i + R_{pole})^2(x_{ri}^2 - 2x_{r3}r_i + x_{r3}^2 + x_{i3}^2)(x_{r1} - x_{r2})$$

$$W(r_i) = x_{r1}x_{r2}R_{pole}(R_{pole} + r_i) \left(R_{pole} + \frac{x_{r3}^2 + x_{i3}^2}{r_i} \right)$$

Long et al. (1989, 4-43) corrected $W(r_i)$ given above from Roberts's original presentation. The parameters p_i in the functions F_i above are

$$p_2 = \frac{S(x_{r1})}{U(x_{r1})} \quad p_3 = \frac{-S(x_{r2})}{U(x_{r2})} \quad p_5 = \frac{S(-R_{pole})}{V}$$

$$p_4 = \frac{1}{X^*} \{ B_0 - x_{r1}x_{r2}R_{pole}^2 [B_4 + B_5(2x_{r3} + x_{r1} + x_{r2} - R_{pole})] \\ + W(x_{r1})p_2 - x_{r1}x_{r2}B_5R_{pole}(x_{r3}^2 + x_{i3}^2) + W(x_{r2})p_3 \\ + x_{r1}x_{r2}(R_{pole}^2 - x_{r3}^2 - x_{i3}^2)p_5 \}$$

$$p_6 = B_4 + B_5 (2x_{r3} + x_{r1} + x_{r2} - R_{pole}) - p_5 - 2(x_{r3} + R_{pole})p_4 \\ - (x_{r2} + R_{pole})p_3 - (x_{r1} + R_{pole})p_2$$

$$p_1 = B_5 - 2p_4 - p_3 - p_2$$

Using coefficients in Table B-2, the coefficients B_n and the functions $S(h_{ellp})$ are given by

$$B_n = \alpha_n + \beta_n \frac{T_x}{T_x - T_o} \quad S(h_{ellp}) = \sum_{n=0}^5 B_n h_{ellp}^n$$

TABLE B-2. Coefficients. These coefficients help determine the functions that lead to the density between 100 and 125 km.

n	α_i	β_i
0	3,144,902,516.672 729	-52,864,482.179 109 69
1	-123,774,885.483 291 7	-16,632.508 473 368 28
2	1,816,141.096 520 398	-1.308 252 378 125
3	-11,403.310 794 892 67	0.0
4	24.364 986 121 055 95	0.0
5	0.008 957 502 869 707 995	0.0

100 to 125 km

Roberts assumes diffusive equilibrium above altitudes of 100 km and integrates Eq. (B-3) by partial fractions for each atmospheric constituent to obtain the standard density between 100 km and 125 km.

$$\rho_{std|100-125}(h_{ellp}) = \sum_{i=1}^5 \rho(100) \frac{M_i}{M_s} \mu_i \left[\frac{T(100)}{T(h_{ellp})} \right]^{1+\alpha_i} F_3^{M_i k} \text{EXP}^{M_i k F_4} \quad (\text{B-10})$$

The mean molecular mass, M_s , at sea level is 28.96 gm/mol. We also need the parameters from Table B-3 (μ_i , α_i , M_i). We could evaluate the density at 100 km [$\rho(100)$] by Eq. (B-9), but this approach is intensive as we saw in the last section! Long et al. (1989, 4-45) presents a polynomial fit as an alternate approach to find $\rho(100)$.

$$\rho_{std|100-125}(100) = M_s \sum_{n=0}^6 \xi_n T_{corr}^n$$

$$\begin{aligned}\xi_0 &= 0.198\ 554\ 9 \times 10^{-10} & \xi_1 &= -0.183\ 349 \times 10^{-14} \\ \xi_2 &= 0.171\ 173\ 5 \times 10^{-17} & \xi_3 &= -0.102\ 147\ 4 \times 10^{-20} \\ \xi_4 &= 0.372\ 789\ 4 \times 10^{-24} & \xi_5 &= -0.773\ 411\ 0 \times 10^{-28} \\ \xi_6 &= 0.702\ 694\ 2 \times 10^{-32}\end{aligned}$$

We don't consider hydrogen significant in determining density below 125 km. Now, find the temperature at 100 km, $T(100)$, using Eq. (B-3):

$$T(100) = T_x + 35^{-4} \sum_{n=0}^4 C_n (100)^n (T_x - T_o) = -0.945\ 855\ 89 (T_x - T_o)$$

The parameters, k and C_n , in Eq. (B-10) are the same as in the 90–100 km case. The functions F_3 , F_4 are slightly different.

$$\begin{aligned}F_3 &= \left(\frac{h_{ellp} + R_{pole}}{R_{pole} + 100} \right)^{q_1} \left(\frac{h_{ellp} - x_{r1}}{100 - x_{r1}} \right)^{q_2} \left(\frac{h_{ellp} - x_{r2}}{100 - x_{r2}} \right)^{q_3} \\ &\quad \times \left(\frac{h_{ellp}^2 - 2x_{r3}h_{ellp} + x_{r3}^2 + x_{i3}^2}{10000 - 200x_{r3} + x_{r3}^2 + x_{i3}^2} \right)^{q_4}\end{aligned}$$

$$\begin{aligned}F_4 &= \left[\frac{q_5 (h_{ellp} - 100)}{(h_{ellp} + R_{pole}) (R_{pole} + 100)} \right] \\ &\quad + \frac{q_6}{x_{i3}} \text{TAN}^{-1} \left[\frac{x_{i3} (h_{ellp} - 100)}{x_{i3}^2 + (h_{ellp} - x_{r3}) (100 - x_{r3})} \right]\end{aligned}$$

The parameters q_i are defined similar to the p_i parameters with the F_1 and F_2 functions.

$$q_2 = \frac{1}{U(x_{r1})} \quad q_3 = \frac{-1}{U(x_{r2})} \quad q_5 = \frac{1}{V}$$

$$q_4 = \frac{1 + x_{r1}x_{r2}(R_{pole}^2 - x_{r3}^2 - x_{i3}^2)q_5 + W(x_{r1})q_2 + W(x_{r2})q_3}{X^*}$$

$$q_6 = -q_5 - 2(x_{r3} + R_{pole})q_4 - (x_{r2} + R_{pole})q_3 - (x_{r1} + R_{pole})q_2$$

$$q_1 = -2q_4 - q_3 - q_2$$

where the roots (x_{r1} , x_{r2} , x_{r3} , x_{i3}), X^* , and the U and W functions were previously determined.

Above 125 km

To get the total standard atmospheric density at altitudes h_{ellp} above 125 km, we must integrate the diffusion differential equation using the temperature profile, [Eq. (B-5)], to achieve the individual effect of the *five* basic atmospheric constituents (Table B-3) on the standard density.

$$\rho_{std}^{125-}(h_{ellp}) = \sum_{i=1}^5 \rho_i(125) \left(\frac{T_x}{T(h_{ellp})} \right)^{1+a_i+\gamma_i} \left(\frac{T_{corr}-T_x}{T_{corr}-T_o} \right)^{\gamma_i}$$

$$\gamma_i = \frac{M_i g_o R_{pole}^2}{R^3 T_{corr}} \left(\frac{T_{corr}-T_x}{T_x-T_o} \right) \left(\frac{35}{6481.766} \right)$$
(B-11)

The parameter g_o is the mean surface gravity, R_{pole} is the polar radius, R is the *universal gas constant*, and ρ_i is the constituent *mass density*. M_i is the constituent *molecular mass* in grams/mole, and a_i is the constituent *thermal diffusion coefficient* (see Table B-3).

TABLE B-3. Atmospheric Constituents and Related Constants. It's important to know properties of the atmosphere's individual components. The constituent number density is multiplied by $M_s/\rho(100)$ and divided by Avogadro's number.

index <i>i</i>	Constituent	Molecular mass (M_i) gm/mol	Thermal diffusion coefficient (a_i)	Constituent density (μ_i)
1	N ₂	28.0134	0	0.781 10
2	Ar	39.948	0	0.009 343 2
3	He	4.0026	-0.38	0.614 71 × 10 ⁻⁵
4	O ₂	31.9988	0	0.161 778
5	O	15.9994	0	0.095 544
6	H	1.007 97	0	

Although we can explicitly determine the constituent mass density, d_i (125 km) from Eq. (B-10), Draper Laboratory uses a polynomial curve fit approximation in the Goddard Trajectory Determination System (Long et al., 1989, 4-49).

$$\log_{10}[d_i(125)] = \sum_{j=0}^6 \delta_{ij} T_{corr}^j \quad d_i = \frac{p_i}{M_i}$$

Here, δ_{ij} are curve-fit coefficients in Table B-4, and Avogadro's number are used. This final correction for helium is

TABLE B-4. Curve-Fitting Coefficients, δ_{ij} . The degree of the polynomial (j) and the index (i) helps us determine the constants used to find the constituent mass density.

	(i=1) N ₂	(i=2) Ar	(i=3) He	(i=4) O ₂	(i=5) O
$j=0$	$0.109\ 315\ 5 \times 10^2$	$0.804\ 940\ 5 \times 10^1$	$0.764\ 688\ 6 \times 10^1$	$0.992\ 423\ 7 \times 10^1$	$0.109\ 708\ 3 \times 10^2$
$j=1$	$0.118\ 678\ 3 \times 10^{-2}$	$0.238\ 282\ 2 \times 10^{-2}$	$-0.438\ 348\ 6 \times 10^{-3}$	$0.160\ 031\ 1 \times 10^{-2}$	$0.611\ 874\ 2 \times 10^{-4}$
$j=2$	$-0.167\ 734\ 1 \times 10^{-5}$	$-0.339\ 136\ 6 \times 10^{-5}$	$0.469\ 431\ 9 \times 10^{-6}$	$-0.227\ 476\ 1 \times 10^{-5}$	$-0.116\ 500\ 3 \times 10^{-6}$
$j=3$	$0.142\ 022\ 8 \times 10^{-8}$	$0.290\ 971\ 4 \times 10^{-8}$	$-0.289\ 488\ 6 \times 10^{-9}$	$0.193\ 845\ 4 \times 10^{-8}$	$0.923\ 935\ 4 \times 10^{-10}$
$j=4$	$-0.713\ 978\ 5 \times 10^{-12}$	$-0.148\ 170\ 2 \times 10^{-11}$	$0.945\ 198\ 9 \times 10^{-13}$	$-0.978\ 218\ 3 \times 10^{-12}$	$-0.349\ 073\ 9 \times 10^{-13}$
$j=5$	$0.196\ 971\ 5 \times 10^{-15}$	$0.412\ 760\ 0 \times 10^{-15}$	$-0.127\ 083\ 8 \times 10^{-16}$	$0.269\ 845\ 0 \times 10^{-15}$	$0.511\ 629\ 8 \times 10^{-17}$
$j=6$	$-0.229\ 618\ 2 \times 10^{-19}$	$-0.483\ 746\ 1 \times 10^{-19}$	0.0	$-0.313\ 180\ 8 \times 10^{-19}$	0.0

$$[\rho_{\text{He}}(h_{\text{ellp}})]_{\text{corrected}} = \rho_{\text{He}}(h_{\text{ellp}}) 10^{(\Delta \log_{10} \rho)_{\text{He}}}$$

We must also account for the concentration of hydrogen whenever the altitudes are above 500 km.

$$\rho_{\text{H}}(h_{\text{ellp}}) = \rho_{\text{H}}(500) \left[\frac{T_{500}}{T(h_{\text{ellp}})} \right]^{(1 + a_{\text{H}} + \gamma_{\text{H}})} \left[\frac{T_{\text{corr}} - T(h_{\text{ellp}})}{T_{\text{corr}} - T_{500}} \right]^{\gamma_{\text{H}}}$$

We calculate the hydrogen density at 500 km, $\rho_{\text{H}}(500)$, using

$$\rho_{\text{H}}(500) = \frac{M_{\text{H}}}{A} 10^{[73.13 - (39.4 - 5.5 \log_{10}(T_{500})) (\log_{10}(T_{500}))]}$$

where γ_{H} is determined from Eq. (B-11), A is Avogadro's number, $6.022\ 57 \times 10^{23}$, T_{500} is the temperature at 500 km from Eq. (B-5), and the molecular mass and thermal-diffusion coefficient for hydrogen are in Table B-3.

We then apply the correction terms of Eq. (B-6), Eq. (B-7), and Eq. (B-8) to the standard density:

$$(\Delta \log_{10} \rho)_{\text{corr}} = (\Delta \log_{10} \rho)_G + (\Delta \log_{10} \rho)_{SA} + (\Delta \log_{10} \rho)_{LT}$$

giving the final corrected density as

$$\rho(h_{\text{ellp}}) = \rho_{\text{std}}(h_{\text{ellp}}) 10^{(\Delta \log_{10} \rho)_{\text{corr}}} \quad (\text{B-12})$$

B.2 Soviet Cosmos Atmosphere

This model is constructed empirically from observations of the orbital motion of the Soviet Cosmos satellites. The model includes the dependence of the density on solar flux and geomagnetic activity as well as the diurnal and semiannual density variations. This model is valid for satellites in the altitude range of 160–600 km. It's sometimes called the

GOST model. Consult Amelina et al. (1996) and Voiskovskii et al. (1973) for more information.

The Soviet formula models the actual atmospheric density as a product of five factors, where each factor corresponds to a particular density variation.

$$\rho = \rho_n k_1 k_2 k_3 k_4 \quad (\text{B-13})$$

where ρ_n is the *night-time density profile* assumed to be exponentially decreasing with respect to altitude, according to

$$\rho_n = \exp^{a_1 - a_2 \sqrt{h_{ellp} - a_3}} \quad (\text{B-14})$$

The coefficients in this expression were determined empirically and are listed in Table B-5. The altitude of the satellite in km is listed as the variable h_{ellp} . The k_1 factor in Eq. (B-13) represents the dependence of the atmospheric density on the solar flux ($F_{10.7}$). The factor k_2 accounts for the diurnal effect, and k_3 takes into account the semiannual effect of the density variation. The factor k_4 represents the effect of geomagnetic activity index (a_p) (Voiskovskii, et al., 1973). In determining the confidence interval, only daily averages of the density are needed, and the dependence on the semiannual cycle is not needed if the error of the predicted density is normalized by the predicted nominal density. This normalization not only makes the density a stationary process but also removes the need for the inclusion of the night-time density (ρ_n) in the calculations (Carter, McClain and Cefola, 1987). The k_1 and k_4 factors given below are

$$k_1 = 1 + \frac{(b_1 + b_2 h_{ellp}) (F_{10.7} - \bar{F})}{\bar{F}}$$

$$k_4 = 1 + \left(c_1 + c_2 h_{ellp} + c_3 \exp^{\left\{ \frac{-(h_{ellp} + c_4)^2}{c_5^2} \right\}} \right) \left(\cos^{m_1} \left(\frac{\Psi_1}{2} \right) + c_6 \cos^{m_2} \left(\frac{\Psi_2}{2} \right) \right)$$

with

$$\cos(\Psi_1) = \frac{1}{r} \left(z \sin(\delta_\odot) + \cos(\delta_\odot) \left\{ x \cos(\gamma_1) + y \sin(\gamma_1) \right\} \right)$$

$$\cos(\Psi_2) = \frac{1}{r} \left(-z \sin(\delta_\odot) + \cos(\delta_\odot) \left\{ x \cos(\gamma_2) + y \sin(\gamma_2) \right\} \right)$$

$$\gamma_1 = \delta_\odot + \phi_1 - (\theta_{LSTM} - \omega_\oplus(t - 3^{hr}))$$

$$\gamma_2 = \delta_\odot + \phi_2 - (\theta_{LSTM} - \omega_\oplus(t - 3^{hr}))$$

$$k_3 = 1 + (A_1 + A_2 h_{ellp}) A(d)$$

$$k_4 = 1 + (e_1 + e_2 h_{ellp}) \ln\left(\frac{a_p}{\bar{a}_p}\right)$$

where the coefficients are listed in Table B-5. The one parameter which must be solved for through linear interpolation is the numerical correction ($A(d)$) which is specified for 10-day intervals from the beginning of the year. The variable \bar{F} represents the mean solar flux for the period under consideration. The Soviet density model only allows for a mean solar flux of 75, 100, 125, or 150. Thus, the actual average solar flux must be rounded to one of these values. The term \bar{a}_p represents the mean geomagnetic index dependent on the mean solar flux which is shown in Table B-5.

TABLE B-5. Soviet Density Model Parameters. Coefficients are listed for different values of the mean solar flux \bar{F} .

	$\bar{F} = 75$	$\bar{F} = 100$	$\bar{F} = 125$	$\bar{F} = 150$
a_1	-14.030	-15.095	-17.028	-16.072
a_2	0.9108	0.8299	0.7198	-16.072
a_3	59.77	68.92	93.36	70.33
b_1	-0.630	-0.750	-0.710	-0.765
b_2	0.00506	0.00560	0.00562	0.00571
c_1	0.130	0.00217	0.00257	0.00199
c_2	0.00014	0.00217	0.00257	0.00199
c_3	3.733	3.784	4.048	4.698
c_4	-507.95	-566.11	-632.63	-707.58
c_5	189.85	200.97	230.76	278.35
c_6	-0.041	-0.047	-0.038	-0.012
m_1	4.2	4.1	4.4	5.2
m_2	6.0	6.0	5.9	5.9
φ_1	37.4°	34.2°	34.5°	33.8°
φ_2	325.9°	318.0°	308.0°	322.2°
A_1	-0.602	-0.526	-0.513	-0.607
A_2	0.00669	0.00636	0.00631	0.00670
e_1	-0.132	-0.130	-0.128	-0.115
e_2	0.00108	0.00104	0.00095	0.00089
\bar{a}_p	2	2	3	4

APPENDIX C

MATHEMATICAL FUNDAMENTALS

- C.1 Vector Fundamentals
- C.2 Matrix Fundamentals
- C.3 Trigonometric Fundamentals
- C.4 Numerical Techniques

C.1 Vector Fundamentals

Because astrodynamics operates mainly in three dimensions, vector relations and operations are very important to understanding the algorithms in the book. Although a vector's most general form can have multiple dimensions, I'll present equations in this section as though all vectors have only three components.

First, we can simply determine the vector's magnitude by taking the square root of the sum of the squares of each component.

$$|\vec{a}| = \sqrt{a_i^2 + a_j^2 + a_k^2} \quad (\text{C-1})$$

Unit vectors are often used in angles-only orbit determination and as unit axes for coordinate systems. The magnitude of the unit vector must be 1.0, and the vector will always have all three components, although individual components may be zero.

$$\hat{a} = \frac{a_i}{|\vec{a}|}\hat{I} + \frac{a_j}{|\vec{a}|}\hat{J} + \frac{a_k}{|\vec{a}|}\hat{K} \quad (\text{C-2})$$

Dot products determine the projection of one vector on another. We can determine them from either the individual components or the vectors and the included angle. The presence of an included angle between the vectors helps a lot in determining orbital elements.

$$c = \vec{a} \cdot \vec{b} = |\vec{a}||\vec{b}|\cos(\alpha) = a_i b_i + a_j b_j + a_k b_k \quad (\text{C-3})$$

Cross products determine the vector that is normal to two vectors. There is also an included angle, but because the relation uses the sine, the relation doesn't help much in determining the included angle as quadrant ambiguity exists. As with the dot product, we can solve this relation using either the components of the vector or the vectors and the included angle.

$$\begin{aligned} \vec{c} &= \vec{a} \times \vec{b} = (a_j b_k - a_k b_j)\hat{I} - (a_i b_k - a_k b_i)\hat{J} + (a_i b_j - a_j b_i)\hat{K} \\ |\vec{c}| &= |\vec{a}||\vec{b}|\sin(\alpha) \end{aligned} \quad (\text{C-4})$$

A variant of the cross product involves multiple cross products. Be careful of the parenthesis because the operation is not commutative.

$$\hat{\vec{a}} \times (\hat{\vec{b}} \times \hat{\vec{c}}) = (\hat{\vec{a}} \cdot \hat{\vec{c}}) \hat{\vec{b}} - (\hat{\vec{a}} \cdot \hat{\vec{b}}) \hat{\vec{c}} \quad (\text{C-5})$$

$$(\hat{\vec{a}} \times \hat{\vec{b}}) \times \hat{\vec{c}} = (\hat{\vec{a}} \cdot \hat{\vec{c}}) \hat{\vec{b}} - (\hat{\vec{b}} \cdot \hat{\vec{c}}) \hat{\vec{a}} \quad (\text{C-6})$$

C.2 Matrix Fundamentals

This book uses a row (r)—col (c) notation for *all* indices of matrices. The *size* of a matrix is defined to be row \times col. Matrices are equal if *every* element of the matrices are equal, and of course, if the two matrices are the same size.

Matrix inversion usually is mathematically intensive, and clumsy in computer code. For small matrices, simpler options exist. We can invert 2-by-2 matrices by (1) taking the *transpose*—interchanging the rows and columns, (2) changing the signs of the off-diagonal terms, and (3) dividing by the determinant of the original matrix. For 3-by-3 matrices, we take the *adjoint* and divide by the determinant of the original matrix. The *adjoint* is the transpose of the submatrices of each element. These submatrices are called *cofactors*, and we find them by crossing out the row and column of the particular element and evaluating the remaining determinant. We must also change the sign if $r + c$ is odd. For larger matrices, consult standard texts on mathematics and computer programming.

Differentiating matrices is useful in differential correction applications. It permits significant savings in terms of the derivations, but requires several rules. In particular, Junkins (1975, A-3) shows partial derivatives of two forms we often encounter in estimation. The solution rests in using a gradient with the matrix of interest. Given a state vector, X , and arbitrary matrices C , A , and B , and vectors y and z , we have

$$y = C^T X \quad z = A^T X + X^T B X \quad (\text{C-7})$$

The gradients of y and z are

$$\nabla_x y = C^T \quad \nabla_x z = A^T + 2X^T B$$

I won't review specifics of basic matrix operations, but I will introduce some we can use to reduce determinants. From Reid (1983, 441),

1. If two rows (columns) in a matrix are interchanged, the determinant will change its sign.
2. The value of the determinant is unchanged if any scalar multiple of any row (or column) is added to any other row (or column).
3. If any row (or column) of a matrix is identically zero, the value of the determinant is zero.

4. Multiplying any row (or column) of a matrix by a nonzero scalar (k) multiplies the determinant by k .
5. $\det(A) = \det(A^T)$

Lastly, it's useful to review *Cramer's rule* and its applicability to solving systems of equations. Suppose we have two equations with unknown parameters x and y .

$$a_1x + b_1y = c_1$$

$$a_2x + b_2y = c_2$$

We can solve for the parameters by forming determinants by replacing each x and y column with the column of c values, respectively, and dividing by the determinant, D , of the a and b coefficients. Thus,

$$x = \frac{\begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix}}{D} \quad y = \frac{\begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}}{D} \quad D = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$$

C.3 Trigonometric Fundamentals

This section lists some trigonometric relations used to develop the algorithms in the book. All equations come from Selby (1975, 227–240).

Begin by considering the Pythagorean relationships:

$$\begin{aligned} \cos^2(\alpha) + \sin^2(\alpha) &= 1 \\ 1 + \tan^2(\alpha) &= \sec^2(\alpha) \\ 1 + \cot^2(\alpha) &= \csc^2(\alpha) \end{aligned} \tag{C-8}$$

These relations also exist for hyperbolic functions:

$$\cosh^2(\alpha) - \sinh^2(\alpha) = 1 \tag{C-9}$$

Standard reduction formulas are

$$\begin{aligned} \sin(\alpha) &= \cos(90^\circ - \alpha) = \sin(180^\circ - \alpha) \\ \cos(\alpha) &= \sin(90^\circ - \alpha) = -\cos(180^\circ - \alpha) \\ \tan(\alpha) &= \cot(90^\circ - \alpha) = -\tan(180^\circ - \alpha) \\ \cot(\alpha) &= \tan(90^\circ - \alpha) = -\cot(180^\circ - \alpha) \end{aligned} \tag{C-10}$$

Selby (1975, 228–229) shows a simple way to extend the results of the reduction formulas, establishing a rule so

$$f(\pm\alpha + n90^\circ) = \pm g(\alpha) \tag{C-11}$$

where

n may be any integer, positive, negative, or zero

f is any one of the six trigonometric functions: SIN, COS, TAN, COT, SEC, or CSC
 α may be any real angle measure

If n is even, then g is the same function as f . If n is odd, then g is the co-function of f .

Half-angle formulas are

$$\begin{aligned} \text{TAN}\left(\frac{\alpha}{2}\right) &= \frac{1 - \cos(\alpha)}{\sin(\alpha)} = \frac{\sin(\alpha)}{1 + \cos(\alpha)} \\ \sqrt{2}\cos\left(\frac{\alpha}{2}\right) &= \frac{\sin(\alpha)}{\sqrt{1 - \cos(\alpha)}} \Rightarrow \cos\left(\frac{\alpha}{2}\right) = \sqrt{\frac{1 + \cos(\alpha)}{2}} \\ \sqrt{2}\sin\left(\frac{\alpha}{2}\right) &= \frac{\sin(\alpha)}{\sqrt{1 + \cos(\alpha)}} \Rightarrow \sin\left(\frac{\alpha}{2}\right) = \sqrt{\frac{1 - \cos(\alpha)}{2}} \\ \text{TAN}^2(\alpha) &= \frac{1 - \cos(2\alpha)}{1 + \cos(2\alpha)} \\ \text{SIN}^2(\alpha + \beta) &= \frac{1}{2} - \frac{1}{2}\cos(2(\alpha + \beta)) \end{aligned} \quad (\text{C-12})$$

It's sometimes useful to expand sum and difference relations:

$$\begin{aligned} \text{SIN}(\alpha + \beta) &= \text{SIN}(\alpha)\text{COS}(\beta) + \text{COS}(\alpha)\text{SIN}(\beta) \\ \text{SIN}(\alpha - \beta) &= \text{SIN}(\alpha)\text{COS}(\beta) - \text{COS}(\alpha)\text{SIN}(\beta) \\ \text{COS}(\alpha + \beta) &= \text{COS}(\alpha)\text{COS}(\beta) - \text{SIN}(\alpha)\text{SIN}(\beta) \\ \text{COS}(\alpha - \beta) &= \text{COS}(\alpha)\text{COS}(\beta) + \text{SIN}(\alpha)\text{SIN}(\beta) \\ \text{TAN}(\alpha + \beta) &= \frac{\text{TAN}(\alpha) + \text{TAN}(\beta)}{1 - \text{TAN}(\alpha)\text{TAN}(\beta)} \\ \text{TAN}(\alpha - \beta) &= \frac{\text{TAN}(\alpha) - \text{TAN}(\beta)}{1 + \text{TAN}(\alpha)\text{TAN}(\beta)} \end{aligned} \quad (\text{C-13})$$

We can express trigonometric terms using other forms:

$$\begin{aligned} \text{SIN}(\alpha) &= \text{TAN}(\alpha)\text{COS}(\alpha), & \text{COS}(\alpha) &= \text{COT}(\alpha)\text{SIN}(\alpha) \\ \text{TAN}(\alpha) &= \text{SIN}(\alpha)\text{SEC}(\alpha), & \text{COT}(\alpha) &= \text{COS}(\alpha)\text{CSC}(\alpha) \\ \text{SEC}(\alpha) &= \text{CSC}(\alpha)\text{TAN}(\alpha), & \text{CSC}(\alpha) &= \text{SEC}(\alpha)\text{COT}(\alpha) \end{aligned} \quad (\text{C-14})$$

Double-angle formulas are useful in reduction of relations and appear often in the perturbation derivations:

$$\begin{aligned} \text{SIN}(2\alpha) &= 2\text{SIN}(\alpha)\text{COS}(\alpha) = \frac{2\text{TAN}(\alpha)}{1 + \text{TAN}^2(\alpha)} \\ \text{TAN}(2\alpha) &= \frac{2\text{TAN}(\alpha)}{1 - \text{TAN}^2(\alpha)} \end{aligned}$$

$$\begin{aligned}\cos(2\alpha) &= \cos^2(\alpha) - \sin^2(\alpha) = 2\cos^2\alpha - 1 \\ &= 1 - 2\sin^2(\alpha) = \frac{1 - \tan^2(\alpha)}{1 + \tan^2(\alpha)} \\ \cot(2\alpha) &= \frac{\cot^2(\alpha) - 1}{2\cot(\alpha)}\end{aligned}\tag{C-15}$$

Series expansions are especially helpful in universal-variable techniques:

$$\sin(\alpha) = \sum_{k=0}^{\infty} \frac{(-1)^k (\alpha)^{(2k+1)}}{(2k+1)!} = \alpha - \frac{\alpha^3}{3!} + \frac{\alpha^5}{5!} - \frac{\alpha^7}{7!} + \frac{\alpha^9}{9!} - \frac{\alpha^{11}}{11!} + \frac{\alpha^{13}}{13!} - \dots\tag{C-16}$$

$$\cos(\alpha) = \sum_{k=0}^{\infty} \frac{(-1)^k (\alpha)^{2k}}{(2k)!} = 1 - \frac{\alpha^2}{2!} + \frac{\alpha^4}{4!} - \frac{\alpha^6}{6!} + \frac{\alpha^8}{8!} - \frac{\alpha^{10}}{10!} + \frac{\alpha^{12}}{12!} - \dots$$

For hyperbolic functions, the series expressions are

$$\sinh(\alpha) = \sum_{k=0}^{\infty} \frac{\alpha^{(2k+1)}}{(2k+1)!} = \alpha + \frac{\alpha^3}{3!} + \frac{\alpha^5}{5!} + \frac{\alpha^7}{7!} + \frac{\alpha^9}{9!} + \frac{\alpha^{11}}{11!} + \frac{\alpha^{13}}{13!} + \dots\tag{C-17}$$

$$\cosh(\alpha) = \sum_{k=0}^{\infty} \frac{\alpha^{2k}}{(2k)!} = 1 + \frac{\alpha^2}{2!} + \frac{\alpha^4}{4!} + \frac{\alpha^6}{6!} + \frac{\alpha^8}{8!} + \frac{\alpha^{10}}{10!} + \frac{\alpha^{12}}{12!} + \dots$$

A useful miscellaneous conversion is

$$\begin{aligned}\cos(\alpha) &= \cos^2\left(\frac{\alpha}{2}\right) - \sin^2\left(\frac{\alpha}{2}\right) \\ &= 2\cos^2\left(\frac{\alpha}{2}\right) - 1\end{aligned}\tag{C-18}$$

C.3.1 Planar Trigonometry

To solve planar triangles, we usually must determine which information is given—sides, angles, or a combination—and then apply the appropriate formula. Figure C-1 shows the notation for the general equations.

For the right triangle, the *Pythagorean Theorem* is a convenient relation permits us to determine the sides:

$$a^2 + b^2 = c^2$$

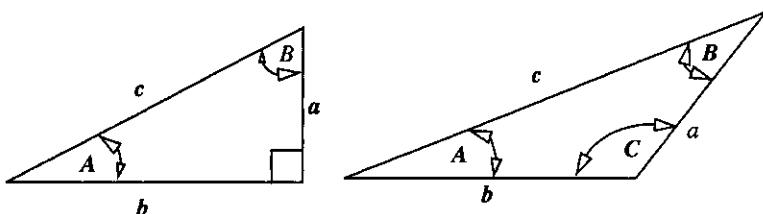


Figure C-1. Planar triangles. The most commonly used triangle is the right triangle (left), although the oblique triangle (right) appears often.

For the general (oblique) triangle, the cosine law relates the sides to the included angle. Of course, each set of known sides and angles has variations.

$$\begin{aligned}c^2 &= a^2 + b^2 - 2ab \cos(C) \\a^2 &= b^2 + c^2 - 2bc \cos(A) \\b^2 &= c^2 + a^2 - 2ac \cos(B)\end{aligned}\tag{C-19}$$

Notice that the relation reduces for the case of the right triangle when $C = 90^\circ$.

The sine law also permits expressions relating the sides and the angles:

$$\frac{a}{\sin(A)} = \frac{b}{\sin(B)} = \frac{c}{\sin(C)}\tag{C-20}$$

Finally, we can form combinations:

$$\begin{aligned}a &= b \cos(C) + c \cos(B) \\b &= c \cos(A) + a \cos(C) \\c &= a \cos(B) + b \cos(A)\end{aligned}\tag{C-21}$$

C.3.2 Spherical Trigonometry

Astroynamics is filled with problems we can solve best by using spherical trigonometry. Spherical triangles are the result of projecting plane triangles onto a sphere. The simplest example is visualizing a satellite above the Earth in three dimensions. As with plane geometry, both right and oblique spherical triangles exist. A common concept with all spherical triangles is the relation between an angle and a distance on the perimeter of the wedge. We refer to an angle, θ , and a distance, s , by using a fact from plane geometry:

$$s = r\theta\tag{C-22}$$

The radius of the circle, r , allows the relation to work, and the units of the radius and the distance will be the same.

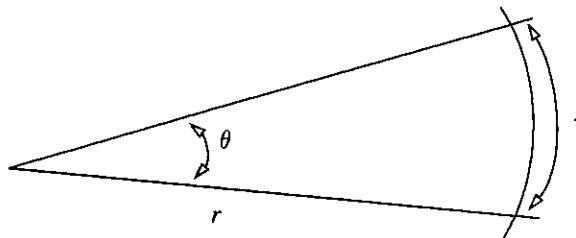


Figure C-2. Range and Angle Geometry. The ground range is related to the central angle by the simple relation $s = r\theta$.

Another important class of analysis deals with right spherical triangles, shown in Fig. C-3. Notice that lower-case letters denote all sides, and capital letters signify angles.

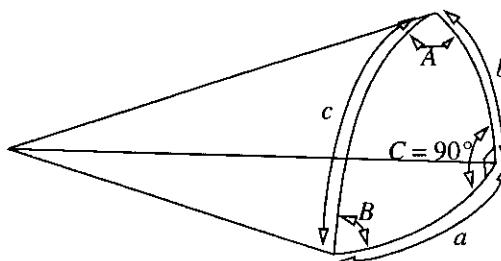


Figure C-3. Right Spherical Triangle. A right spherical triangle has one angle (C in the figure), that equals 90° .

Napier's rules apply for all right spherical triangles.

$$\begin{array}{ll}
 \cos(c) = \cos(a)\cos(b) & \tan(a) = \cos(B)\tan(c) \\
 \sin(a) = \sin(A)\sin(c) & \tan(b) = \tan(B)\sin(a) \\
 \tan(b) = \cos(A)\tan(c) & \cos(c) = \cot(A)\cot(B) \\
 \tan(a) = \tan(A)\sin(b) & \cos(A) = \cos(a)\sin(B) \\
 \sin(b) = \sin(B)\sin(c) & \cos(B) = \cos(b)\sin(A)
 \end{array} \quad (C-23)$$

The geometry of oblique triangles is very similar to the right triangles, but with no 90° angles, the formulas are very different. Four main laws apply to oblique spherical triangles. Notice I've given each formula with a primary notation and all the possible combinations of known angles and sides. Each formula also has side and angle variations. The lists may seem long, but they're helpful for a student trying to find an angle or side not in the standard configuration. Figure C-4 shows the geometry.

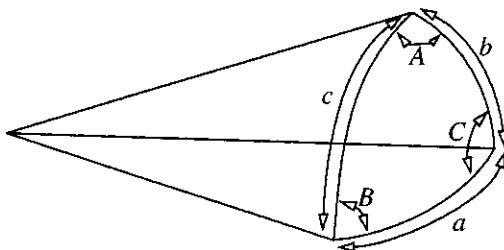


Figure C-4. Oblique Spherical Triangle. Oblique spherical triangles have no 90° angles. Also, angles and distance are interchangeable when dealing with the triangle.

1. Cosine Law:

For Sides

$$\cos(a) = \cos(b)\cos(c) + \sin(b)\sin(c)\cos(A)$$

$$\cos(b) = \cos(c)\cos(a) + \sin(c)\sin(a)\cos(B)$$

$$\cos(c) = \cos(a)\cos(b) + \sin(a)\sin(b)\cos(C)$$

(C-24)

For Angles

$$\cos(A) = -\cos(B)\cos(C) + \sin(B)\sin(C)\cos(a)$$

$$\cos(B) = -\cos(C)\cos(A) + \sin(C)\sin(A)\cos(b)$$

$$\cos(C) = -\cos(A)\cos(B) + \sin(A)\sin(B)\cos(c)$$

2. Sine Law

$$\frac{\sin(a)}{\sin(A)} = \frac{\sin(b)}{\sin(B)} = \frac{\sin(c)}{\sin(C)}$$
(C-25)

3. Four Parts

For Sides

$$\cot(A)\sin(B) = \sin(c)\cot(a) - \cos(c)\cos(B)$$

$$\cot(B)\sin(C) = \sin(a)\cot(b) - \cos(a)\cos(C)$$

$$\cot(C)\sin(A) = \sin(b)\cot(c) - \cos(b)\cos(A)$$

(C-26)

For Angles

$$\cot(a)\sin(b) = \sin(C)\cot(A) + \cos(C)\cos(b)$$

$$\cot(b)\sin(c) = \sin(A)\cot(B) + \cos(A)\cos(c)$$

$$\cot(c)\sin(a) = \sin(B)\cot(C) + \cos(B)\cos(a)$$

4. Analog to the cosine formula

For Sides

$$\sin(a)\cos(B) = \cos(b)\sin(c) - \sin(b)\cos(c)\cos(A)$$

$$\sin(b)\cos(C) = \cos(c)\sin(a) - \sin(c)\cos(a)\cos(B)$$

$$\sin(c)\cos(A) = \cos(a)\sin(b) - \sin(a)\cos(b)\cos(C)$$

(C-27)

For Angles

$$\sin(A)\cos(b) = \cos(B)\sin(C) + \sin(B)\cos(C)\cos(a)$$

$$\sin(B)\cos(c) = \cos(C)\sin(A) + \sin(C)\cos(A)\cos(b)$$

$$\sin(C)\cos(a) = \cos(A)\sin(B) + \sin(A)\cos(B)\cos(c)$$

C.4 Numerical Techniques

Any astrodynamist must know various numerical techniques to solve problems. Some astrodynamical problems don't have analytical solutions. For this book, we're mainly interested in interpolation techniques to determine precise values of coefficients. I'm assuming you're familiar with introductory series, and reduction techniques, so I won't present them here.

C.4.1 Polynomial Solutions

The first area we'll explore is solving polynomials. Explicit expressions exist for the second, third, and fourth order, but we need numerical techniques for higher orders.

Quadratic Roots

The general form is

$$y = \alpha_2x^2 + \alpha_1x + \alpha_0$$

The roots are determined analytically as

$$x_{root_{1,2}} = \frac{-\alpha_1 \pm \sqrt{\alpha_1^2 - 4\alpha_2\alpha_0}}{2\alpha_2} \quad (C-28)$$

Cubic Roots

Astrodynamical problems often require solutions to cubic polynomials. For computer programs, we want to store the *real* answers as the *first* values in the formal parameter list. This significantly eases future programming with the routine because we can drop checks to determine if a root is real.

Let's follow Escobal ([1965] 1985, 433), and Hill (1981, 182) to solve Cardan's equation. Given the general form of a cubic,

$$y = \alpha_3x^3 + \alpha_2x^2 + \alpha_1x + \alpha_0$$

Rearrange so the highest exponent has a coefficient of one.

$$y = x^3 + Px^2 + Qx + R$$

Now find the constants,

$$a = \frac{1}{3}(3Q - P^2) \quad b = \frac{1}{27}(2P^3 - 9PQ + 27R)$$

which we use to determine the type of solution.

$$\Delta = \frac{a^3}{27} + \frac{b^2}{4}$$

If the delta value is greater than zero ($\Delta > 0.0$), use a form of Cardan's solution. There will be one real root,

$$x_{root_1} = \sqrt[3]{-\frac{b}{2} + \sqrt{\Delta}} + \sqrt[3]{-\frac{b}{2} - \sqrt{\Delta}} \quad (C-29)$$

and two imaginary roots. Be careful to evaluate negative cube roots correctly:

$$x_{root_{2i}} = \sqrt{\frac{-3}{2}} \left(\sqrt[3]{-\frac{b}{2} + \sqrt{\Delta}} - \sqrt[3]{-\frac{b}{2} - \sqrt{\Delta}} \right) \quad (C-30)$$

$$x_{root_{3i}} = -x_{root_{2i}}$$

If the delta value (Δ) is equal to zero, all three roots are real, so

$$x_{root_1} = 2 \sqrt[3]{-\frac{b}{2}}$$

$$x_{root_2} = x_{root_3} = \sqrt[3]{\frac{b}{2}} \quad (C-31)$$

If the delta value is less than zero ($\Delta < 0.0$), use a trigonometric solution as follows. Let E_0 be an intermediate quantity:

$$E_0 = 2 \sqrt{\frac{-a}{3}}$$

It's prudent to use ATAN2 to resolve quadrants using both the sine and cosine values.

$$\cos(\phi) = \frac{-b}{2 \sqrt{\frac{-a^3}{27}}} \quad \sin(\phi) = \sqrt{1 - \cos^2(\phi)}$$

We determine the three roots (for $\Delta < 0.0$) through circular relations:

$$Z_1 = E_0 \cos\left(\frac{\phi}{3}\right) \quad Z_2 = E_0 \cos\left(\frac{\phi}{3} + 120^\circ\right) \quad Z_3 = E_0 \cos\left(\frac{\phi}{3} + 240^\circ\right)$$

$$x_{root_i=1\dots 3} = Z_i - \frac{P}{3} \quad (C-32)$$

All answers are real [Eq. (C-29), Eq. (C-31), and Eq. (C-32)], except for two imaginary roots in Eq. (C-30). It's computationally convenient to place the answers for the real roots in a predetermined location in a computer program's formal parameter list. This saves future effort because we don't have to test which roots are real or imaginary.

Quartic Roots

Solving fourth-order polynomial roots is very similar to the procedure for third-order equations. The derivation again follows Escobal ([1965] 1985, 430–433). Assume the general form of

$$y = \alpha_4 x^4 + \alpha_3 x^3 + \alpha_2 x^2 + \alpha_1 x + \alpha_0$$

Rearrange for a unity coefficient on the first term:

$$y = x^4 + bx^3 + cx^2 + dx + e$$

and define h as

$$h = -\frac{b}{4}$$

Doing so permits several more definitions, including a new polynomial form:

$$y = x^4 + Px^2 + Qx + R$$

$$P = 6h^2 + 3bh + c$$

$$Q = 4h^3 + 3bh^2 + 2ch + d$$

$$R = h^4 + bh^3 + ch^2 + dh + e$$

Determining the solution now rests on the value of Q . If $Q = 0.0$, let $Z = y^2$ and solve the quadratic ($Z^2 + PZ + R = 0$). This will yield two roots having real and imaginary components ($Z_{root1}, Z_{root1j}, Z_{root2}, Z_{root2j}$). For each root, solve the complex square root to find the four roots of the quadratic, y_i . Remember that for a complex square root, we need a magnitude $r = |Z|$, and a polar angle θ that is defined by the complex number. Thus with $j = 1, 2$, we obtain four intermediate roots of y ($k = 1\dots 4$)

$$r_j = \sqrt{Z_{root_j}^2 + Z_{root_{ji}}^2} \quad \theta = \text{ATAN2}\left(Z_{root_j}/r_j, Z_{root_{ji}}/r_j\right) \quad (C-33)$$

$$y_k = \sqrt{r_j} \left(\cos\left(\frac{\theta}{2}\right) \pm i \sin\left(\frac{\theta}{2}\right) \right)$$

If Q is non-zero, we can find parameters similar to the cubic case, along with a new parameter, s :

$$a = \frac{1}{3} \left\{ 3(P^2 - 4R) - 4P^2 \right\} \quad s = -\frac{2}{3}P$$

$$b = \frac{1}{27} (16P^3 - 18P(P^2 - 4R) - 27Q^2)$$

This allows us to form $\Delta = a^3/27 + b^2/4$ as in the cubic case. The solution is identical at this point, and it yields three roots, Z_i , from Eqs. (C-29), (C-30), (C-31) and (C-32). We find R' as the largest value of $(Z_i + s)$ and then

$$\zeta = \frac{1}{2}(P + R' - \sqrt{R'})$$

$$\beta = \frac{1}{2} \left(P + R' + \frac{Q}{\sqrt{R'}} \right)$$

Solving the quadratics gives us

$$\begin{aligned} y_i^2 + \sqrt{R'}y_i + \zeta &= 0 & i &= 1, 2 \\ y_j^2 + \sqrt{R'}y_j + \beta &= 0 & j &= 3, 4 \end{aligned} \tag{C-34}$$

The final roots for both cases of Q are found using Eq. (C-33) or Eq. (C-34) and

$$x_{rooti} = y_i + h \quad i = 1 \dots 4$$

C.4.2 Interpolation

For centuries, interpolation has been a mainstay of all intensive numerical operations because it enables complex evaluations with little computation. Before modern computers, interpolation was crucial to determining logarithms, trigonometric values, etc. But we can still benefit from this technique's immense power. Sadler (1974) gives an excellent summary of methods and notations. We also have the advantage of routines available through the Internet. An excellent source of data is the *Guide to Mathematical Software* (GAMS) maintained by the National Institute of Standards (<http://gams.nist.gov>)

In this book, we will illustrate a useful form for use in astrodynamics. Thus, we'll consider only a simple linear interpolation. Linear interpolation is the easiest method to use and understand. The underlying principle is to assume a linear relation between two end points and use a simple mathematical ratio to evaluate the desired point between them. The value of the function at t is

$$y(t) = y(t_o) + \frac{y(t_1) - y(t_o)}{t_1 - t_o} (t - t_o) \quad (\text{C-35})$$

Refer to Meeus (1991, 23–24) or GAMS for more complex interpolation schemes.

C.4.3 Blending and Splining Techniques

These especially interesting techniques differ slightly from interpolation. *Blending techniques* attempt to combine similar geometric shapes to find an *average* that best approximates the desired function. On the other hand, *splining techniques* try to fit a certain polynomial to represent the shape of a desired function. These techniques are mainly used to approximate zero points and roots of complex functions in simulation studies. Numerical methods often produce the data for these studies. Don't confuse integration techniques which are designed to predict or extrapolate a curve, whereas these techniques interpolate *within* the polynomial. Getting enough accuracy often requires very small step sizes and several iterative operations to determine the zero point to a desired tolerance. These techniques give us the chance to determine the actual zero points very precisely using generated data from a numerical technique with much larger step sizes. We must have high-quality, reliable techniques so we can increase step sizes and therefore reduce computational processing. Although the literature covers many numerical techniques for curve fitting and interpolation, we'll talk only about those blending and splining functions needed for algorithms in this book. In addition, we won't explore error characteristics because we're only using these functions to estimate the data.

Although I present equations for blending and splining techniques to approximate values of functions, we often require the time of an event. The techniques are general because they will operate on any distribution of initial data (even or uneven). If we use a parameterized time, τ , or a uniform time interval, T , we can recover the original input time, t , by applying the same blending and splining processes to the time values.

Parabolic Blending

This computer-graphics technique handles problems associated with Fourier series, trigonometric polynomial fits, or spline functions and their various end-point strategies. Curve fitting by *parabolic blending* uses four points to create two second-order curve functions, for which the points must be consecutive but need not be uniformly spaced. The first curve is created from the first three data points; the second curve is defined by the last three points. The curve segments passing through the central points are then linearly blended into a single third-order polynomial. We choose the equation for the blending so the blended curve matches the slope of the first parabola at the second point and also matches the slope of the second parabola at the third point. We repeat this blending process until we've used up the data, resulting in a continuous curve created by piecing together many localized cubic polynomials. This method was originally developed by Overhauser (1968) and then refined by Brewer and Anderson (1977, 132–137). I present an independent derivation below.

Let's define a quadratic polynomial $P_{qd}(T)$ passing through three consecutive points (p_1, p_2, p_3) in Eq. (C-36). T uniformly spans the interval of p_1 to p_3 , thus allowing the function to exactly equal the three points at $T = 0.0, 1.0$, and 2.0 .

$$P_{qd}(T) = \alpha_{qd2}T^2 + \alpha_{qd1}T + \alpha_{qd0} \quad (C-36)$$

We'll define the quadratic coefficients as

$$\begin{aligned}\alpha_{qd0} &= p_1 \\ \alpha_{qd1} &= \frac{-3p_1 + 4p_2 - p_3}{2} \\ \alpha_{qd2} &= \frac{p_1 - 2p_2 + p_3}{2}\end{aligned}$$

The position and slope at the central point ($T = 1$) become

$$\begin{aligned}P_{qd}(1) &= p_2 \\ \dot{P}_{qd}(1) &= \frac{p_3 - p_1}{2}\end{aligned}$$

Now consider Fig. C-5. Given four consecutive points p_1, p_2, p_3, p_4 , construct two parabolas using the f function above. The P_{qd1} parabola passes through the points p_1, p_2, p_3 , and the P_{qd2} parabola passes through points p_2, p_3, p_4 . On the central interval, p_2, p_3 , define the $f_{pb}(\tau)$ curve so it linearly weights P_1 and P_2 through the blending function:

$$\begin{aligned}f_{pb}(\tau) &= \alpha_{pb3}\tau^3 + \alpha_{pb2}\tau^2 + \alpha_{pb1}\tau + \alpha_{pb0} & 0.0 \leq \tau \leq 1.0 \\ \dot{f}_{pb}(\tau) &= 3\alpha_{pb3}\tau^2 + 2\alpha_{pb2}\tau + \alpha_{pb1}\end{aligned} \quad (C-37)$$

based on end conditions

$$f_{pb}(0) = P_{qd1}(1) \text{ and } \dot{f}_{pb}(0) = \dot{P}_{qd1}(1)$$

$$f_{pb}(1) = P_{qd2}(1) \text{ and } \dot{f}_{pb}(1) = \dot{P}_{qd2}(1)$$

Equating the above expressions for the f_{pb} curve gives us

$$f_{pb}(0) = \alpha_{pb0} = P_{qd1}(1) = p_2$$

$$\dot{f}_{pb}(0) = \alpha_{pb1} = \dot{P}_{qd1}(1) = \frac{p_3 - p_1}{2}$$

$$f_{pb}(1) = \alpha_{pb3} + \alpha_{pb2} + \alpha_{pb1} + \alpha_{pb0} = P_{qd2}(1) = p_3$$

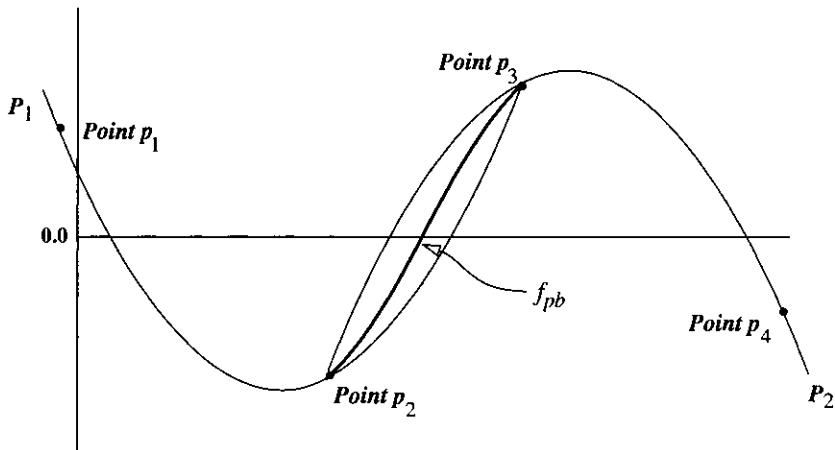


Figure C-5. Parabolic Blending. Parabolic blending involves taking two parabolas (P_i) and matching their curves to approximate the blended curve (f_{pb}). The points don't have to be equally spaced.

$$\dot{f}_{pb}(1) = 3\alpha_{pb3} + 2\alpha_{pb2} + \alpha_{pb1} = \dot{P}_{qd2}(1) = \frac{p_4 - p_2}{2}$$

In terms of the original points, the α_{pb} coefficients reduce to

$$\begin{aligned}\alpha_{pb0} &= p_2 \\ \alpha_{pb1} &= -\frac{1}{2}p_1 + \frac{1}{2}p_3 \\ \alpha_{pb2} &= p_1 - \frac{5}{2}p_2 + 2p_3 - \frac{1}{2}p_4 \\ \alpha_{pb3} &= -\frac{1}{2}p_1 + \frac{3}{2}p_2 - \frac{3}{2}p_3 + \frac{1}{2}p_4\end{aligned}\tag{C-38}$$

For the final answer, find the coefficients from Eq. (C-38) and the polynomial in Eq. (C-37), which we can solve analytically for the roots by the methods in Sec. C.4.1.

Cubic Splining

To alleviate the problems in modeling Fourier series and polynomial wiggle, we use cubic and quintic spline functions. Curve fitting by **cubic splining** uses function values and their time derivatives at the beginning and end of a time interval to create a third-order polynomial. We repeat this process until we've used up the set of sequential time intervals. The result is curve created by piecing together the localized cubic polynomials. An alternate method uses four consecutive data points.

Given a function $f(t)$ and its time derivative $\dot{f}(t)$ at t_n and t_{n+1} , define a cubic polynomial $P_c(\tau)$ as

$$\begin{aligned} P_c(\tau) &= \alpha_{c3}\tau^3 + \alpha_{c2}\tau^2 + \alpha_{c1}\tau + \alpha_{c0} \\ \alpha_{c0} &= f(t_n) \\ \alpha_{c1} &= \dot{f}(t_n)\Delta t \\ \alpha_{c2} &= -3f(t_n) - 2\dot{f}(t_n)\Delta t + 3f(t_{n+1}) - \dot{f}(t_{n+1})\Delta t \\ \alpha_{c3} &= 2f(t_n) + \dot{f}(t_n)\Delta t - 2f(t_{n+1}) + \dot{f}(t_{n+1})\Delta t \\ \Delta t &= t_{n+1} - t_n \quad (t_{n+1} > t_n) \end{aligned} \tag{C-39}$$

with τ uniformly spanning the interval $(0.0 \leq \tau \leq 1.0)$. Define the α coefficients so the cubic polynomial P_c exactly represents the function and its time-scaled derivative (\dot{x}_x) at the interval boundaries.

An alternate approach to cubic splining involves fitting four data points on the τ interval. Now, consider Fig. C-6. Given four data points (p_1, p_2, p_3, p_4) occurring at the respective times $(0, \tau_1, \tau_2, 1)$, where $0.0 < (\tau_1, \tau_2) < 1.0$, the Eq. (C-39) α_c coefficients are shown in Eq. (C-40). We define the α_c coefficients so the polynomial P_c exactly represents the function at the four data points. Variations can occur elsewhere. Thus,

$$\begin{aligned} \alpha_{c0} &= p_1 \\ \alpha_{c1} &= \frac{(\tau_2^3 - \tau_2^2)(p_2 - p_1) + (\tau_1^2 - \tau_1^3)(p_3 - p_1) + (\tau_1^3\tau_2^2 - \tau_1^2\tau_2^3)(p_4 - p_1)}{DET} \\ \alpha_{c2} &= \frac{(\tau_2 - \tau_2^3)(p_2 - p_1) + (\tau_1^3 - \tau_1)(p_3 - p_1) + (\tau_1\tau_2^3 - \tau_1^3\tau_2)(p_4 - p_1)}{DET} \\ \alpha_{c3} &= \frac{(\tau_2^2 - \tau_2)(p_2 - p_1) + (\tau_1 - \tau_1^2)(p_3 - p_1) + (\tau_1^2\tau_2 - \tau_1\tau_2^2)(p_4 - p_1)}{DET} \end{aligned} \tag{C-40}$$

$$DET = \tau_1^3\tau_2^2 + \tau_1^2\tau_2 + \tau_1\tau_2^3 - \tau_1^3\tau_2 - \tau_1^2\tau_2^3 - \tau_1\tau_2^2$$

Cubic splining uses Eq. (C-39) or Eq. (C-40) to produce the polynomial which best fits the data. We find the three roots (real and imaginary) using analytical expressions in Sec. C.4.1. Because we're only interested in real roots, we can perform a quick check to determine if any exist before using the analytical techniques of Sec. C.4.1. Expressing the cubic equation with nested multiplication yields

$$(\alpha_{c3}\tau + \alpha_{c2})\tau + \alpha_{c1}\} \tau + \alpha_{c0} = 0 \tag{C-41}$$

Each subcomponent of the nested equation is linear, having a maximum or minimum at $\tau = 0.0$ or $\tau = 1.0$. No root exists on the interval 0.0 to 1.0. Thus,

$$\text{IF } \alpha_{c0} > 0.0 \text{ and } \text{MIN}(\alpha_{c1}, \alpha_{c1} + \alpha_{c2}, \alpha_{c1} + \alpha_{c2} + \alpha_{c3}) > -\alpha_{c0}$$

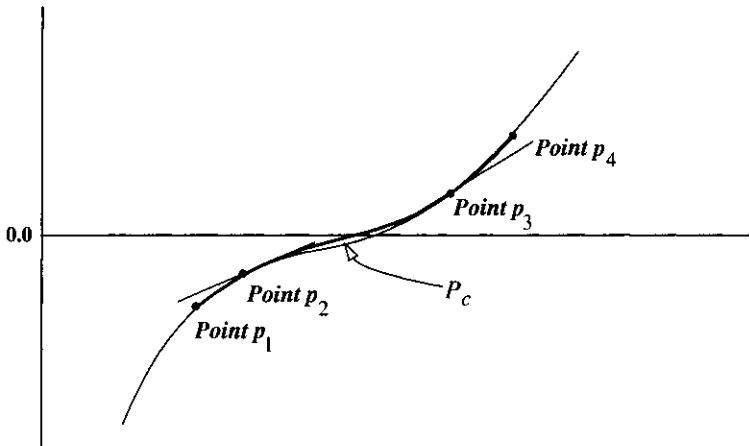


Figure C-6. Cubic Splining. We fit the curve with cubic splining by either examining the function and derivative at the start and end (p_2 and p_3), or by using four points along the curve. The resulting polynomials will perfectly match the point and slope of the function at p_2 and p_3 .

or

$$\text{IF } \alpha_{c0} < 0.0 \text{ and } \text{MAX}(\alpha_{c1}, \alpha_{c1} + \alpha_{c2}, \alpha_{c1} + \alpha_{c2} + \alpha_{c3}) < -\alpha_0$$

Remember that MIN/MAX are functions that determine the minimum or maximum value from a set of data. If $\alpha_{qr0} = 0.0$, zero is a root.

Quartic Blending

Problems associated with restrictive sinusoidal fits or high-order generalized polynomials may use a numerical technique called *quartic blending*. This method extends the work in Alfano, Negron, and Moore (1992). A series of polynomial functions, each defined by five data points, approximates a waveform. Curve fitting by quartic blending uses six consecutive points to create two fourth-order curve functions. The first curve is created from the first five data points; the second curve is created from data points two through six. As in Fig. C-5, blend the curve segments passing through the central points into a single fifth-order polynomial. Choose the equation for the blending to match the first quartic's slope and its derivative at the third point and match the second quartic's slope and its derivative at the fourth point. Repeat the blending until the data set is exhausted, resulting in a second-order continuous curve created from many localized quintic polynomials. A quartic polynomial $P_{qr}(T)$ passing through five consecutive points p_1, p_2, p_3, p_4, p_5 on a uniform interval that corresponds to the index is defined as

$$P_{qr}(T) = \alpha_{qr4}T^4 + \alpha_{qr3}T^3 + \alpha_{qr2}T^2 + \alpha_{qr1}T + \alpha_{qr0} \quad (0.0 \leq T \leq 4.0) \quad (\text{C-42})$$

where

$$\alpha_{qr0} = p_0$$

$$\alpha_{qr1} = (-50p_0 + 96p_1 - 72p_2 + 32p_3 - 6p_4)/24$$

$$\alpha_{qr2} = (35p_0 - 104p_1 + 114p_2 - 56p_3 + 11p_4)/24$$

$$\alpha_{qr3} = (-10p_0 + 36p_1 - 48p_2 + 28p_3 - 6p_4)/24$$

$$\alpha_{qr4} = (p_0 - 4p_1 + 6p_2 - 4p_3 + p_4)/24$$

The point, slope, and its derivative with respect to T at mid-point are

$$\begin{aligned} P_{qr}(2) &= 16\alpha_{qr4} + 8\alpha_{qr3} + 4\alpha_{qr2} + 2\alpha_{qr1} + \alpha \\ \dot{P}_{qr}(2) &= 32\alpha_{qr4} + 12\alpha_{qr3} + 4\alpha_{qr2} + \alpha_{qr1} \\ \ddot{P}_{qr}(2) &= 48\alpha_{qr4} + 12\alpha_{qr3} + 2\alpha_{qr2} \end{aligned} \quad (\text{C-43})$$

Quartic blending uses six consecutive points $p_1, p_2, p_3, p_4, p_5, p_6$. The first quartic, P_{qr1} , ranges from p_1 to p_5 , whereas the second quartic, P_{qr2} , ranges from p_2 to p_6 . The quintic curve, P_{qr} , ranging between p_3 and p_4 , combines the two quartics to match the P_{qr1} curve to second order at the p_3 and to match the P_{qr2} curve to second order at p_4 . This process uses Eq. (C-42) and Eq. (C-43) to determine the function and derivatives at p_3 and p_4 . After some algebraic manipulation, the equation for $f_{qr}(\tau)$ becomes

$$f_{qr}(\tau) = \alpha_{qr5}\tau^5 + \alpha_{qr4}\tau^4 + \alpha_{qr3}\tau^3 + \alpha_{qr2}\tau^2 + \alpha_{qr1}\tau + \alpha_{qr0} \quad 0.0 \leq \tau \leq 1.0 \quad (\text{C-44})$$

$$\alpha_{qr0} = p_3$$

$$\alpha_{qr1} = (2p_1 - 16p_2 + 16p_4 - 2p_5)/24$$

$$\alpha_{qr2} = (-p_1 + 16p_2 - 30p_3 + 16p_4 - p_5)/24$$

$$\alpha_{qr3} = (-9p_1 + 39p_2 - 70p_3 + 66p_4 - 33p_5 + 7p_6)/24$$

$$\alpha_{qr4} = (13p_1 - 64p_2 + 126p_3 - 124p_4 + 61p_5 - 12p_6)/24$$

$$\alpha_{qr5} = (-5p_1 + 25p_2 - 50p_3 + 50p_4 - 25p_5 + 5p_6)/24$$

Solving the quintic requires iteration because no method exists to solve the fifth-order polynomial in closed form. We can reduce the search times using several steps, including some techniques to determine if any real roots exist. This approach helps when running many operations using this technique.

Roots occur whenever a quintic function changes sign—whenever $f_{qr}(\tau)$ equals zero. The roots of $f_{qr}(\tau)$ are ignored if they are imaginary, repeating, or outside the range of τ . Any remaining roots are ignored between points p_3 and p_4 , that is,

$$0 \leq \tau \equiv \frac{t - t_3}{t_4 - t_3} \leq 1$$

Because the roots of interest are in the interval (0.0 to 1.0), a quick test determines if further processing is necessary. Reorganizing the polynomial like Eq. (C-41),

$$(((\alpha_{qr5}\tau + \alpha_{qr4})\tau + \alpha_{qr3})\tau + \alpha_{qr2})\tau + \alpha_{qr1})\tau + \alpha_{qr0} = 0$$

no roots exist

IF $\alpha_{qr6} > 0$ and

$$\text{MIN}(\alpha_{qr5}, \alpha_{qr5} + \alpha_{qr4}, \alpha_{qr5} + \alpha_{qr4} + \alpha_{qr3}, \alpha_{qr5} + \alpha_{qr4} + \alpha_{qr3} + \alpha_{qr2}, \alpha_{qr5} + \alpha_{qr4} + \alpha_{qr3} + \alpha_{qr2} + \alpha_{qr1}) > -\alpha_{qr6},$$

or

IF $\alpha_{qr6} < 0$ and

$$\text{MAX}(\alpha_{qr5}, \alpha_{qr5} + \alpha_{qr4}, \alpha_{qr5} + \alpha_{qr4} + \alpha_{qr3}, \alpha_{qr5} + \alpha_{qr4} + \alpha_{qr3} + \alpha_{qr2}, \alpha_{qr5} + \alpha_{qr4} + \alpha_{qr3} + \alpha_{qr2} + \alpha_{qr1}) < -\alpha_{qr6},$$

Using a Newton-Raphson method, let's determine if any real unique roots exist in the interval (0.0 to 1.0). Obtain a first estimate by approximating the quintic polynomial on the interval as a quartic. This step will help determine the *existence* of a root. Use Eq. (C-42) and Eq. (C-43) for the points $(f_{qr}(0), f_{qr}(0.25), f_{qr}(0.5), f_{qr}(0.75), f_{qr}(1))$, with the quartic solved in closed form. If no unique real roots exist in the quartic interval, assume no such roots exist for the quintic interval. If a root does exist in the quartic interval (0.0 to 4.0), divide it by 4 to correspond to the quintic interval (0.0 to 1.0) and use it to start the Newton-Raphson search. Once you find a single root, factor it out and solve the resulting quartic in closed form.

Quintic Splining

To do curve fitting by ***quintic splining***, use function values and their first and second time derivatives at the beginning and end of a time interval to create a fifth-order polynomial. Repeat this process until you exhaust the set of sequential time intervals, resulting in a second-order continuous curve created by piecing together many localized quintic polynomials. Given $f(t)$, $f'(t)$, and $f''(t)$ at t_n and t_{n+1} , define a quintic polynomial $P_{qr}(\tau)$ as

$$\begin{aligned}
 f_{qi}(\tau) &= \alpha_{qi5}\tau^5 + \alpha_{qi4}\tau^4 + \alpha_{qi3}\tau^3 + \alpha_{qi2}\tau^2 + \alpha_{qi1}\tau + \alpha_{qi0} & 0.0 \leq \tau \leq 1.0 \\
 \alpha_{qi0} &= f(t_n) \\
 \alpha_{qi1} &= \dot{f}(t_n)\Delta t \\
 \alpha_{qi2} &= 0.5\ddot{f}(t_n)\Delta t^2 \\
 \alpha_{qi3} &= -10f(t_n) - 6\dot{f}(t_n)\Delta t - 1.5\ddot{f}(t_n)\Delta t^2 \\
 &\quad + 10f(t_{n+1}) - 4\dot{f}(t_{n+1})\Delta t + 0.5\ddot{f}(t_{n+1})\Delta t^2 & (C-45) \\
 \alpha_{qi4} &= 15f(t_n) + 8\dot{f}(t_n)\Delta t + 1.5\ddot{f}(t_n)\Delta t^2 - 15f(t_{n+1}) + 7\dot{f}(t_{n+1})\Delta t - \ddot{f}(t_{n+1})\Delta t^2 \\
 \alpha_{qi5} &= -6f(t_n) - 3\dot{f}(t_n)\Delta t - 0.5\ddot{f}(t_n)\Delta t^2 \\
 &\quad + 6f(t_{n+1}) - 3\dot{f}(t_{n+1})\Delta t + 0.5\ddot{f}(t_{n+1})\Delta t^2
 \end{aligned}$$

with τ uniformly spanning the interval (0.0 to 1.0). Define the coefficients so the polynomial P_{qi} exactly represents the function and its time-scaled derivatives at the interval boundaries. Use numerical techniques to solve Eq. (C-45), as with Eq. (C-44).

APPENDIX D

CONSTANTS AND EXPANSIONS

- D.1 Gravitational Coefficients
- D.2 Planetary Constants
- D.3 1980 IAU Reduction Coefficients
- D.4 Planetary Ephemerides
- D.5 Data Sources
- D.6 Computer Programming

D.1 Gravitational Coefficients

The use of a real-world aspherical-potential model requires coefficients for analysis. I've selected the JGM-2 model for several reasons. First, it's a highly accurate model derived from numerous satellite measurements. Vetter (1994) compares gravitational models in detail and concludes JGM-2 is the best for general-purpose and high-accuracy use.

TABLE D-1. JGM-2 Zonal Coefficients (Unnormalized). Notice the relative magnitude past about ten to fifteen terms. (NASA, *Gravitational Models*, Sec. D.5)

<i>t</i>	<i>C_t</i>	×10	<i>t</i>	<i>C_t</i>	×10	<i>t</i>	<i>C_t</i>	×10
2	-0.108 262 692 563 881 5	-02	25	0.329 794 899 797 176 9	-07	48	0.445 131 528 064 147 4	-07
3	0.253 230 781 819 177 4	-05	26	0.597 335 666 375 466 2	-07	49	-0.154 788 325 014 295 1	-08
4	0.162 042 999 000 000 0	-05	27	0.303 539 568 003 334 4	-07	50	-0.389 655 516 210 581 7	-07
5	0.227 071 104 392 034 3	-06	28	-0.941 623 882 079 879 8	-07	51	-0.630 978 094 103 009 7	-07
6	-0.540 843 616 139 963 1	-06	29	-0.268 492 882 663 014 6	-07	52	0.943 144 298 800 088 9	-08
7	0.352 232 250 452 579 4	-06	30	0.734 975 509 004 279 3	-07	53	0.584 662 435 321 919 7	-07
8	0.203 289 475 484 741 2	-06	31	0.397 328 700 775 360 3	-07	54	0.991 672 023 054 462 6	-08
9	0.116 398 319 982 126 4	-06	32	-0.372 882 331 333 855 1	-07	55	0.237 971 245 051 701 7	-07
10	0.246 965 814 497 234 7	-06	33	-0.372 373 879 898 489 5	-08	56	-0.302 650 041 652 849 0	-07
11	-0.236 487 675 075 079 1	-06	34	-0.534 313 296 381 862 9	-07	57	-0.251 453 641 769 933 1	-07
12	0.178 027 005 000 000 0	-06	35	0.447 908 498 020 595 0	-07	58	-0.412 072 758 501 764 1	-07
13	0.203 318 258 021 709 9	-06	36	-0.516 840 790 176 396 0	-07	59	-0.187 455 614 421 839 3	-08
14	-0.118 455 817 016 941 9	-06	37	-0.463 895 029 227 107 8	-07	60	-0.334 986 135 000 000 0	-07
15	0.253 885 717 656 610 4	-07	38	0.408 251 655 487 877 1	-08	61	0.642 616 794 136 262 1	-08
16	-0.318 569 794 012 583 3	-07	39	0.183 203 276 677 626 2	-07	62	0.182 273 081 185 895 4	-07
17	0.104 962 839 036 402 5	-06	40	-0.158 421 771 000 000 0	-07	63	-0.293 400 338 237 573 1	-07
18	0.404 930 280 235 556 4	-07	41	-0.169 748 450 012 326 9	-07	64	-0.253 386 791 242 044 8	-07
19	-0.154 315 187 010 012 1	-07	42	0.153 143 566 533 066 7	-07	65	0.716 732 473 914 728 7	-09
20	0.126 723 252 105 025 2	-06	43	0.509 575 070 011 335 8	-07	66	-0.570 212 149 989 871 0	-08
21	0.505 938 710 321 824 1	-07	44	0.202 518 027 676 349 7	-07	67	0.151 705 505 156 730 3	-08
22	-0.725 138 397 542 367 5	-07	45	-0.521 771 746 418 092 3	-07	68	0.973 582 225 765 870 7	-08
23	-0.154 134 619 754 353 6	-06	46	-0.140 991 514 593 172 8	-07	69	0.188 956 887 306 312 4	-07
24	-0.198 613 583 000 000 0	-07	47	-0.826 548 385 038 065 6	-08	70	-0.965 588 440 349 553 1	-08

The values for the sectorial and tesseral coefficients are more numerous because the model is complete to order 70. See “Data Sources” on page 886 for methods to obtain electronic files.

TABLE D-2. Selected JGM-2 Sectorial and Tesseral Coefficients (Unnormalized). Sectorial and tesseral coefficients occur for all other gravitational coefficients where m is not equal to zero. (NASA, *Gravitational Models*, Sec. D.5)

l	m	C_{lm}	$\times 10$	S_{lm}	$\times 10$
2	1	-0.241 400 005 222 209 3	-09	0.154 309 997 378 437 9	-08
2	2	0.157 442 175 835 099 4	-05	-0.903 766 666 961 687 4	-06
3	1	0.219 092 208 140 471 6	-05	0.268 741 886 313 685 5	-06
3	2	0.308 914 353 381 648 8	-06	-0.211 507 512 283 537 1	-06
3	3	0.100 560 104 062 658 6	-06	0.197 178 025 045 693 7	-06
4	1	-0.508 843 315 774 593 0	-06	-0.449 128 170 460 647 0	-06
4	2	0.783 404 895 390 826 6	-07	0.148 221 992 057 051 0	-06
4	3	0.591 792 417 824 845 5	-07	-0.120 126 397 595 865 8	-07
4	4	-0.398 254 644 355 990 0	-08	0.652 554 840 627 475 5	-08
5	1	-0.506 285 045 457 373 0	-07	-0.818 093 569 395 849 0	-07
5	2	0.105 740 780 674 042 0	-06	-0.523 986 138 108 110 0	-07
5	3	-0.149 282 899 331 568 0	-07	-0.710 558 079 072 205 0	-08
5	4	-0.229 757 518 096 428 0	-08	0.386 744 752 635 903 0	-09
5	5	0.430 820 139 662 621 0	-09	-0.164 883 741 330 739 0	-08
50	1	0.101 905 786 271 621 8	-08	-0.524 800 237 352 456 7	-09
50	2	-0.382 104 712 463 244 7	-10	-0.148 310 637 747 191 5	-10
50	3	0.835 349 500 054 993 4	-13	-0.108 871 185 335 747 5	-12
50	9	0.143 339 862 999 158 7	-23	0.246 808 161 499 846 7	-22
50	10	-0.701 741 711 794 887 1	-24	-0.102 054 552 062 054 8	-24
70	61	0.111 399 974 313 095 1	-115	-0.289 388 291 021 155 9	-115
70	62	0.527 221 721 669 076 8	-117	0.533 116 187 804 438 8	-117
70	63	-0.278 248 718 777 118 6	-118	0.127 996 934 314 434 4	-118
70	69	-0.410 849 353 222 469 7	-127	0.960 634 071 978 369 8	-127
70	70	-0.295 284 361 059 312 2	-128	-0.761 103 119 457 069 9	-129

D.2 Planetary Constants

Besides defining the Earth's physical parameters, it's also useful to define constants for the remaining planets and the Sun. Table D-3 shows data for the Sun.

TABLE D-3. Solar Constants. These solar constants are from Seidelmann (1992, 696–697, 700). The value for the semimajor axis, a_{\oplus} , of the Earth's orbit and the time-unit value for the Sun are derived quantities.

Distance			Speed		
(R_{\odot})	1.0 SR	696,000,000 km	1.0 AU/ TU _⊕	29.784 691 674 9	km /solar s
1.0 AU		149,597,870.0 km			
a_{\oplus}		149,598,023.0 km			
Shape			Gravitational Parameter		
Mass	332,946 Earth masses		(μ_{\odot})	1.0 AU ³ / TU ² _⊕	1.327 124 28×10 ¹¹ km ³ / (solar s) ²
		1.9891×10 ³⁰ kg			
Rotation			Time		
			1.0 TU	58.132 440 906	solar days

Tables D-4 and Table D-5 list data for the planets. Notice that using canonical units allows identical routines for the Earth, as well as other planets. Once we've found the four defining physical parameters, the equations and routines are exactly the same, except for the gravitational potential. I discuss a way to obtain planetary positions in Sec. 3.7. Knowing the orbital elements for each planet is important when analyzing interplanetary problems. Many algorithms require knowing the physical characteristics of the Sun, Moon, and planets.

TABLE D-4. Mean Planetary Constants For Epoch J2000. All values for the top section are from Seidelmann (1992, 704). Values for the lower section are from Seidelmann (1992, 706). The values for gravitational parameters come from Seidelmann (1992, 700).

	Moon	Mercury	Venus	Earth	Mars
Semimajor axis, AU*	60.27 ER	0.387 098 309	0.723 329 82	1.000 001 017 8	1.523 679 34
km	384,400	57,909,083	108,208,601	149,598,023	227,939,186
Eccentricity	0.054 90	0.205 631 752	0.006 771 882	0.016 708 617	0.093 400 620
Inclination (°)	5.145 396	7.004 986 25	3.394 466 19	0.000 000 00	1.849 726 48
Longitude of ascending node (°)	0.0	48.330 893 04	76.679 920 19	0.000 000 00	49.558 093 21
Longitude of perihelion (°)	0.0	77.456 119 04	131.563 707 24	102.937 348 08	336.060 233 98
True longitude (°)	0.0	252.250 905 51	181.979 800 84	100.466 448 51	355.433 274 63
Orbital period (yrs)	0.0	0.240 844 45	0.615 182 57	0.999 978 62	1.880 711 05
Orbital velocity(km/s)	0.0	47.8725	35.0214	29.7859	24.1309
Equatorial radius (km)	1738.	2439.	6052.	6378.1363	3397.2
Reciprocal flattening				0.003 353 813 1	0.006 476 30
Gravitational parameter (km ³ /s ²)	4902.799	2.2032×10 ⁴	3.257×10 ⁵	3.986 004 415	4.305×10 ⁴
Mass ($m_{\oplus} = 1$)	0.012 30	0.055 274 3	0.8150	1.0	0.107 44
Mass (kg)	7.3483×10 ²²	3.3022×10 ²³	4.869×10 ²⁴	5.9742×10 ²⁴	6.4191×10 ²³
Rotation (days)	27.321 66	58.6462	-243.01	0.997 269 68	1.025 956 75
Inclination of Equator to orbit (°)	6.68	0.0	177.3	23.45	25.19
J_2	0.000 202 7	0.000 06	0.000 027	0.001 082 626 9	0.001 964
J_3				-0.000 002 532 3	0.000 036
J_4				-0.000 001 620 4	
Density (gm/cm ³)	3.34	5.43	5.24	5.515	3.94

*1 AU = 149,597,870 km.

TABLE D-5. Mean Planetary Constants For Epoch J2000. All values for the top section are from Seidelmann (1992, 704). Values for the lower section are from Seidelmann (1992, 706). The values for gravitational parameters come from Seidelmann (1992, 700).

	Jupiter	Saturn	Uranus	Neptune	Pluto
Semimajor axis, AU*	5.202 603 191	9.554 909 595	19.218 446 061	30.110 386 87	39.544 674
km	778,298,361	1,429,394,133	2,875,038,615	4,504,449,769	5,915,799,000
Eccentricity	0.0484 948 51	0.055 508 622	0.046 295 898	0.008 988 095	0.249 050
Inclination (°)	1.303 269 66	2.488 878 10	0.773 196 17	1.769 952 21	17.142 166 67
Longitude of ascending node (°)	100.464 440 64	113.665 523 70	74.005 947 23	131.784 057 02	110.297 138 89
Longitude of perihelion (°)	14.331 309 24	93.056 787 28	173.005 159 22	48.123 690 50	224.134 861 11
True longitude (°)	34.351 483 92	50.077 471 38	314.055 005 11	304.348 665 48	238.743 944 44
Orbital period (yrs)	11.856 525	29.423 519	83.747 406	163.723 2045	248.020 8
Orbital velocity(km/s)	13.0697	9.6724	6.835 24	5.4778	4.7490
Equatorial radius (km)	71,492	60,268	25,559.	24,764	1151.
Reciprocal flattening	0.064 874 4	0.097 962 4	0.022 927 3	0.0171	0.0
Gravitational parameter (km ³ /s ²)	1.268×10 ⁸	3.794×10 ⁷	5.794×10 ⁶	6.809×10 ⁶	9.00×10 ²
Mass ($m_{\oplus} = 1$)	318.0	95.159	14.4998	17.203	0.00251
Mass (kg)	1.8988×10 ²⁷	5.685×10 ²⁶	8.6625×10 ²⁵	1.0278×10 ²⁶	1.5×10 ²²
Rotation (days)	0.413 54	0.4375	-0.65	0.768	-6.3867
Inclination of Equator to orbit (°)	3.12	26.73	97.86	29.56	118.0
J_2	0.014 75	0.016 45	0.012	0.004	
J_3	-0.000 58	-0.001			
J_4					
Density (gm/cm ³)	1.33	0.70	1.30	1.76	1.1

*1 AU = 149,597,870 km.

D.3 1980 IAU Reduction Coefficients

Trigonometric series are used extensively when forming the nutation coefficients for the 1980 IAU theory. Table D-6 lists the values.

TABLE D-6. 1980 IAU Theory of Nutation Coefficients, Epoch J2000. The units for A_i , C_i are $0.0001''$ and for B_i , D_i are $0.0001''$ per Julian century. Data is from Seidelmann (1991, 112–113).

i	a_1	a_2	a_3	a_4	a_5	A_i	B_i	C_i	D_i	i	a_1	a_2	a_3	a_4	a_5	A_i	B_i	C_i	D_i
1	0	0	0	0	1	-171,996	-174.2	92,025	8.9	54	1	0	2	2	2	-8	0	3	0
2	0	0	0	0	2	2062	0.2	-895	0.5	55	1	0	0	2	0	6	0	0	0
3	-2	0	2	0	1	46	0	-24	0	56	2	0	2	-2	2	6	0	-3	0
4	2	0	-2	0	0	11	0	0	0	57	0	0	0	2	1	-6	0	3	0
5	-2	0	2	0	2	-3	0	1	0	58	0	0	2	2	1	-7	0	3	0
6	1	-1	0	-1	0	-3	0	0	0	59	1	0	2	-2	1	6	0	-3	0
7	0	-2	2	-2	1	-2	0	1	0	60	0	0	0	-2	1	-5	0	3	0
8	2	0	-2	0	1	1	0	0	0	61	1	-1	0	0	0	5	0	0	0
9	0	0	2	-2	2	-13,187	-1.6	5736	-3.1	62	2	0	2	0	1	-5	0	3	0
10	0	1	0	0	0	1426	-3.4	54	-0.1	63	0	1	0	-2	0	-4	0	0	0
11	0	1	2	-2	2	-517	1.2	224	-0.6	64	1	0	-2	0	0	4	0	0	0
12	0	-1	2	-2	2	217	-0.5	-95	0.3	65	0	0	0	1	0	-4	0	0	0
13	0	0	2	-2	1	129	0.1	-70	0	66	1	1	0	0	0	-3	0	0	0
14	2	0	0	-2	0	48	0	1	0	67	1	0	2	0	0	3	0	0	0
15	0	0	2	-2	0	-22	0	0	0	68	1	-1	2	0	2	-3	0	1	0
16	0	2	0	0	0	17	-0.1	0	0	69	-1	-1	2	2	2	-3	0	1	0
17	0	1	0	0	1	-15	0	9	0	70	-2	0	0	0	1	-2	0	1	0
18	0	2	2	-2	2	-16	0.1	7	0	71	3	0	2	0	2	-3	0	1	0
19	0	-1	0	0	1	-12	0	6	0	72	0	-1	2	2	2	-3	0	1	0
20	-2	0	0	2	1	-6	0	3	0	73	1	1	2	0	2	2	0	-1	0
21	0	-1	2	-2	1	-5	0	3	0	74	-1	0	2	-2	1	-2	0	1	0
22	2	0	0	-2	1	4	0	-2	0	75	2	0	0	0	1	2	0	-1	0
23	0	1	2	-2	1	4	0	-2	0	76	1	0	0	0	2	-2	0	1	0
24	1	0	0	-1	0	-4	0	0	0	77	3	0	0	0	0	2	0	0	0
25	2	1	0	-2	0	1	0	0	0	78	0	0	2	1	2	2	0	-1	0
26	0	0	-2	2	1	1	0	0	0	79	-1	0	0	0	2	1	0	-1	0
27	0	1	-2	2	0	-8	0	0	0	80	1	0	0	-4	0	-1	0	0	0
28	0	1	0	0	2	1	0	0	0	81	-2	0	2	2	2	1	0	-1	0
29	-1	0	0	1	1	1	0	0	0	82	-1	0	2	4	2	-2	0	1	0
30	0	1	2	-2	0	-1	0	0	0	83	?	0	0	-4	0	-1	0	0	0
31	0	0	2	0	2	-2274	-0.2	977	-0.5	84	1	1	2	-2	2	1	0	-1	0

TABLE D-6. 1980 IAU Theory of Nutation Coefficients, Epoch J2000. The units for A_i C_i are 0.0001" and for B_i D_i are 0.0001" per Julian century. Data is from Seidelmann (1991, 112–113).

i	a_1	a_2	a_3	a_4	a_5	A_i	B_i	C_i	D_i	i	a_1	a_2	a_3	a_4	a_5	A_i	B_i	C_i	D_i
32	1	0	0	0	0	712	0.1	-7	0	85	1	0	2	2	1	-1	0	1	0
33	0	0	2	0	1	-386	-0.4	200	0	86	-2	0	2	4	2	-1	0	1	0
34	1	0	2	0	2	-301	0	129	-0.1	87	-1	0	4	0	2	1	0	0	0
35	1	0	0	-2	0	-158	0	-1	0	88	1	-1	0	-2	0	1	0	0	0
36	-1	0	2	0	2	123	0	-53	0	89	2	0	2	-2	1	1	0	-1	0
37	0	0	0	2	0	63	0	-2	0	90	2	0	2	2	2	-1	0	0	0
38	1	0	0	0	1	63	0.1	-33	0	91	1	0	0	2	1	-1	0	0	0
39	-1	0	0	0	1	-58	-0.1	32	0	92	0	0	4	-2	2	1	0	0	0
40	-1	0	2	2	2	-59	0	26	0	93	3	0	2	-2	2	1	0	0	0
41	1	0	2	0	1	-51	0	27	0	94	1	0	2	-2	0	-1	0	0	0
42	0	0	2	2	2	-38	0	16	0	95	0	1	2	0	1	1	0	0	0
43	2	0	0	0	0	29	0	-1	0	96	-1	-1	0	2	1	1	0	0	0
44	1	0	2	-2	2	29	0	-12	0	97	0	0	-2	0	1	-1	0	0	0
45	2	0	2	0	2	-31	0	13	0	98	0	0	2	-1	2	-1	0	0	0
46	0	0	2	0	0	26	0	-1	0	99	0	1	0	2	0	-1	0	0	0
47	-1	0	2	0	1	21	0	-10	0	100	1	0	-2	-2	0	-1	0	0	0
48	-1	0	0	2	1	16	0	-8	0	101	0	-1	2	0	1	-1	0	0	0
49	1	0	0	-2	1	-13	0	7	0	102	1	1	0	-2	1	-1	0	0	0
50	-1	0	2	2	1	-10	0	5	0	103	1	0	-2	2	0	-1	0	0	0
51	1	1	0	-2	0	-7	0	0	0	104	2	0	0	2	0	1	0	0	0
52	0	1	2	0	2	7	0	-3	0	105	0	0	2	4	2	-1	0	0	0
53	0	-1	2	0	2	-7	0	3	0	106	0	1	0	1	0	1	0	0	0

D.4 Planetary Ephemerides

From Meeus (1991, 202–204), the orbital elements of the planets, referenced to the mean equator and mean equinox of J2000, are as follows:

Reference J2000 in Degrees

Mercury

$$\begin{aligned} a &= 0.387\,098\,310 \\ e &= 0.205\,631\,75 + 0.000\,020\,406 T_{UT1} - 0.000\,000\,028\,47 T_{UT1}^2 - 0.000\,000\,000\,17 T_{UT1}^3 \\ i &= 7.004\,986^\circ - 0.005\,951\,6 T_{UT1} + 0.000\,000\,81 T_{UT1}^2 + 0.000\,000\,041 T_{UT1}^3 \\ \Omega &= 48.330\,893^\circ - 0.125\,422\,9 T_{UT1} - 0.000\,088\,33 T_{UT1}^2 - 0.000\,000\,196 T_{UT1}^3 \\ \tilde{\omega} &= 77.456\,119^\circ + 0.158\,864\,3 T_{UT1} - 0.000\,013\,43 T_{UT1}^2 + 0.000\,000\,039 T_{UT1}^3 \\ \lambda_M &= 252.250\,906^\circ + 149,472.674\,635\,8 T_{UT1} - 0.000\,005\,35 T_{UT1}^2 + 0.000\,000\,002 T_{UT1}^3 \end{aligned}$$

Venus

$$\begin{aligned} a &= 0.723\,329\,820 \\ e &= 0.006\,771\,88 - 0.000\,047\,766 T_{UT1} + 0.000\,000\,097\,5 T_{UT1}^2 + 0.000\,000\,000\,44 T_{UT1}^3 \\ i &= 3.394\,662^\circ - 0.000\,856\,8 T_{UT1} - 0.000\,032\,44 T_{UT1}^2 + 0.000\,000\,010 T_{UT1}^3 \\ \Omega &= 76.679\,920^\circ - 0.278\,008\,0 T_{UT1} - 0.000\,142\,56 T_{UT1}^2 - 0.000\,000\,198 T_{UT1}^3 \\ \tilde{\omega} &= 131.563\,707^\circ + 0.004\,864\,6 T_{UT1} - 0.001\,382\,32 T_{UT1}^2 - 0.000\,005\,332 T_{UT1}^3 \\ \lambda_M &= 181.979\,801^\circ + 58,517.815\,676\,0 T_{UT1} + 0.000\,001\,65 T_{UT1}^2 - 0.000\,000\,002 T_{UT1}^3 \end{aligned}$$

Earth

$$\begin{aligned} a &= 1.000\,001\,018 \\ e &= 0.016\,708\,62 - 0.000\,042\,037 T_{UT1} - 0.000\,000\,123\,6 T_{UT1}^2 + 0.000\,000\,000\,04 T_{UT1}^3 \\ i &= 0.000\,000\,0^\circ + 0.013\,054\,6 T_{UT1} - 0.000\,009\,31 T_{UT1}^2 - 0.000\,000\,034 T_{UT1}^3 \\ \Omega &= 0.0^\circ \\ \tilde{\omega} &= 102.937\,348^\circ + 0.322\,555\,7 T_{UT1} + 0.000\,150\,26 T_{UT1}^2 + 0.000\,000\,478 T_{UT1}^3 \\ \lambda_M &= 100.466\,449^\circ + 35,999.372\,851\,9 T_{UT1} - 0.000\,005\,68 T_{UT1}^2 + 0.000\,000\,000 T_{UT1}^3 \end{aligned}$$

Mars

$$\begin{aligned} a &= 1.523\,679\,342 \\ e &= 0.093\,400\,62 + 0.000\,090\,483 T_{UT1} - 0.000\,000\,080\,6 T_{UT1}^2 - 0.000\,000\,000\,35 T_{UT1}^3 \end{aligned}$$

$$\begin{aligned}
 i &= 1.849\,726^\circ - 0.008\,147\,9T_{UT1} - 0.000\,022\,55T_{UT1}^2 - 0.000\,000\,027T_{UT1}^3 \\
 \Omega &= 49.558\,093^\circ - 0.294\,984\,6T_{UT1} - 0.000\,639\,93T_{UT1}^2 - 0.000\,002\,143T_{UT1}^3 \\
 \tilde{\omega} &= 336.060\,234^\circ + 0.443\,889\,8T_{UT1} - 0.000\,173\,21T_{UT1}^2 + 0.000\,000\,300T_{UT1}^3 \\
 \lambda_M &= 355.433\,275^\circ + 19,140.299\,331\,3T_{UT1} + 0.000\,002\,61T_{UT1}^2 - 0.000\,000\,003T_{UT1}^3
 \end{aligned}$$

Jupiter

$$\begin{aligned}
 a &= 5.202\,603\,191 + 0.000\,000\,191\,3T_{UT1} \\
 e &= 0.048\,494\,85 + 0.000\,163\,244T_{UT1} - 0.000\,000\,471\,9T_{UT1}^2 - 0.000\,000\,001\,97T_{UT1}^3 \\
 i &= 1.303\,270^\circ - 0.001\,987\,2T_{UT1} + 0.000\,033\,18T_{UT1}^2 + 0.000\,000\,092T_{UT1}^3 \\
 \Omega &= 100.464\,441^\circ + 0.176\,682\,8T_{UT1} + 0.000\,903\,87T_{UT1}^2 - 0.000\,007\,032T_{UT1}^3 \\
 \tilde{\omega} &= 14.331\,309^\circ + 0.215\,552\,5T_{UT1} + 0.000\,722\,52T_{UT1}^2 - 0.000\,004\,590T_{UT1}^3 \\
 \lambda_M &= 34.351\,484^\circ + 3034.905\,674\,6T_{UT1} - 0.000\,085\,01T_{UT1}^2 + 0.000\,000\,004T_{UT1}^3
 \end{aligned}$$

Saturn

$$\begin{aligned}
 a &= 9.554\,909\,596 - 0.000\,002\,138\,9T_{UT1} \\
 e &= 0.055\,508\,62 - 0.000\,346\,818T_{UT1} - 0.000\,000\,645\,6T_{UT1}^2 + 0.000\,000\,003\,38T_{UT1}^3 \\
 i &= 2.488\,878^\circ + 0.002\,551\,5T_{UT1} - 0.000\,049\,03T_{UT1}^2 + 0.000\,000\,018T_{UT1}^3 \\
 \Omega &= 113.665\,524^\circ - 0.256\,664\,9T_{UT1} - 0.000\,183\,45T_{UT1}^2 + 0.000\,000\,357T_{UT1}^3 \\
 \tilde{\omega} &= 93.056\,787^\circ + 0.566\,549\,6T_{UT1} + 0.000\,528\,09T_{UT1}^2 + 0.000\,004\,882T_{UT1}^3 \\
 \lambda_M &= 50.077\,471^\circ + 1222.113\,794\,3T_{UT1} + 0.000\,210\,04T_{UT1}^2 - 0.000\,000\,019T_{UT1}^3
 \end{aligned}$$

Uranus

$$\begin{aligned}
 a &= 19.218\,446\,062 - 0.000\,000\,037\,2T_{UT1} + 0.000\,000\,000\,98T_{UT1}^2 \\
 e &= 0.0462\,959\,0 - 0.000\,027\,337T_{UT1} + 0.000\,000\,079\,0T_{UT1}^2 + 0.000\,000\,000\,25T_{UT1}^3 \\
 i &= 0.773\,196^\circ - 0.001\,686\,9T_{UT1} + 0.000\,003\,49T_{UT1}^2 + 0.000\,000\,016T_{UT1}^3 \\
 \Omega &= 74.005\,947^\circ + 0.074\,146\,1T_{UT1} + 0.000\,405\,40T_{UT1}^2 + 0.000\,000\,104T_{UT1}^3 \\
 \tilde{\omega} &= 173.005\,159^\circ + 0.089\,320\,6T_{UT1} - 0.000\,094\,70T_{UT1}^2 + 0.000\,000\,413T_{UT1}^3 \\
 \lambda_M &= 314.055\,005^\circ + 428.466\,998\,3T_{UT1} - 0.000\,004\,86T_{UT1}^2 + 0.000\,000\,006T_{UT1}^3
 \end{aligned}$$

Neptune

$$\begin{aligned}
 a &= 30.110\,386\,869 - 0.000\,000\,166\,3T_{UT1} + 0.000\,000\,000\,69T_{UT1}^2 \\
 e &= 0.008\,988\,09 + 0.000\,006\,408T_{UT1} - 0.000\,000\,000\,8T_{UT1}^2 \\
 i &= 1.769\,952^\circ + 0.000\,225\,7T_{UT1} + 0.000\,000\,23T_{UT1}^2 - 0.000\,000\,000T_{UT1}^3 \\
 \Omega &= 131.784\,057^\circ - 0.006\,165\,1T_{UT1} - 0.000\,002\,19T_{UT1}^2 - 0.000\,000\,078T_{UT1}^3 \\
 \tilde{\omega} &= 48.123\,691^\circ + 0.029\,158\,7T_{UT1} + 0.000\,070\,51T_{UT1}^2 - 0.000\,000\,000T_{UT1}^3 \\
 \lambda_M &= 304.348\,665^\circ + 218.486\,200\,2T_{UT1} + 0.000\,000\,59T_{UT1}^2 - 0.000\,000\,002T_{UT1}^3
 \end{aligned}$$

The following elements for Pluto are from the *Astronomical Almanac* of 1995.

Pluto

$$\begin{aligned}
 a &= 39.537\,58 \\
 e &= 0.250\,877 \\
 i &= 17.132\,33^\circ \\
 \Omega &= 110.406\,5^\circ \\
 \tilde{\omega} &= 224.6148^\circ \\
 \lambda_M &= 218.887\,35^\circ
 \end{aligned}$$

Reference Mean Equinox of J2000 in Radians

Remember that the semimajor axis and eccentricity will be identical to the degree expressions already shown.

Mercury

$$\begin{aligned}
 i &= 0.122\,260\,07 - 0.000\,103\,875T_{UT1} + 0.000\,000\,014\,1T_{UT1}^2 + 0.000\,000\,000\,72T_{UT1}^3 \\
 \Omega &= 0.843\,533\,21 - 0.002\,189\,043T_{UT1} - 0.000\,001\,541\,6T_{UT1}^2 - 0.000\,000\,003\,42T_{UT1}^3 \\
 \tilde{\omega} &= 1.351\,864\,30 + 0.002\,772\,705T_{UT1} - 0.000\,000\,234\,4T_{UT1}^2 + 0.000\,000\,000\,68T_{UT1}^3 \\
 \lambda_M &= 4.402\,608\,85 + 2608.790\,314\,157T_{UT1} - 0.000\,000\,093\,4T_{UT1}^2 + 0.000\,000\,000\,03T_{UT1}^3
 \end{aligned}$$

Venus

$$\begin{aligned}
 i &= 0.059\,248\,03 - 0.000\,014\,954T_{UT1} - 0.000\,000\,566\,2T_{UT1}^2 + 0.000\,000\,000\,17T_{UT1}^3 \\
 \Omega &= 1.338\,317\,07 - 0.004\,852\,155T_{UT1} - 0.000\,002\,488\,1T_{UT1}^2 - 0.000\,000\,003\,46T_{UT1}^3 \\
 \tilde{\omega} &= 2.296\,219\,86 + 0.000\,084\,903T_{UT1} - 0.000\,024\,126\,0T_{UT1}^2 - 0.000\,000\,093\,06T_{UT1}^3 \\
 \lambda_M &= 3.176\,146\,70 + 1021.328\,554\,621T_{UT1} + 0.000\,000\,028\,8T_{UT1}^2 - 0.000\,000\,000\,03T_{UT1}^3
 \end{aligned}$$

Earth

$$i = 0.000\ 000\ 00 + 0.000\ 227\ 846T_{UT1} - 0.000\ 000\ 162\ 5T_{UT1}^2 \sim 0.000\ 000\ 000\ 59T_{UT1}^3$$

$$\Omega = 0.0$$

$$\tilde{\omega} = 1.796\ 595\ 65 + 0.005\ 629\ 659T_{UT1} + 0.000\ 002\ 622\ 5T_{UT1}^2 + 0.000\ 000\ 008\ 34T_{UT1}^3$$

$$\lambda_M = 1.753\ 470\ 32 + 628.307\ 584\ 919T_{UT1} - 0.000\ 000\ 099\ 1T_{UT1}^2 + 0.000\ 000\ 000\ 007T_{UT1}^3$$

Mars

$$i = 0.032\ 283\ 81 - 0.000\ 142\ 208T_{UT1} - 0.000\ 000\ 393\ 6T_{UT1}^2 \sim 0.000\ 000\ 000\ 47T_{UT1}^3$$

$$\Omega = 0.864\ 951\ 89 - 0.005\ 148\ 453T_{UT1} - 0.000\ 011\ 168\ 9T_{UT1}^2 - 0.000\ 000\ 037\ 40T_{UT1}^3$$

$$\tilde{\omega} = 5.865\ 357\ 57 + 0.007\ 747\ 339T_{UT1} - 0.000\ 003\ 023\ 1T_{UT1}^2 + 0.000\ 000\ 005\ 24T_{UT1}^3$$

$$\lambda_M = 6.203\ 480\ 92 + 334.061\ 243\ 148T_{UT1} + 0.000\ 000\ 045\ 6T_{UT1}^2 - 0.000\ 000\ 000\ 057T_{UT1}^3$$

Jupiter

$$i = 0.022\ 746\ 35 - 0.000\ 034\ 683T_{UT1} + 0.000\ 000\ 579\ 1T_{UT1}^2 + 0.000\ 000\ 001\ 61T_{UT1}^3$$

$$\Omega = 1.753\ 435\ 28 + 0.003\ 083\ 697T_{UT1} + 0.000\ 015\ 775\ 5T_{UT1}^2 - 0.000\ 000\ 122\ 73T_{UT1}^3$$

$$\tilde{\omega} = 0.250\ 128\ 53 + 0.003\ 762\ 101T_{UT1} + 0.000\ 012\ 610\ 4T_{UT1}^2 - 0.000\ 000\ 080\ 11T_{UT1}^3$$

$$\lambda_M = 0.599\ 546\ 50 + 52.969\ 096\ 509T_{UT1} - 0.000.001\ 483\ 7T_{UT1}^2 + 0.000\ 000\ 000\ 077T_{UT1}^3$$

Saturn

$$i = 0.043\ 439\ 12 + 0.000\ 044\ 532T_{UT1} - 0.000\ 000\ 855\ 7T_{UT1}^2 + 0.000\ 000\ 000\ 31T_{UT1}^3$$

$$\Omega = 1.983\ 837\ 64 - 0.004\ 479\ 648T_{UT1} - 0.000\ 003\ 201\ 8T_{UT1}^2 + 0.000\ 000\ 006\ 23T_{UT1}^3$$

$$\tilde{\omega} = 1.624\ 147\ 32 + 0.009\ 888\ 156T_{UT1} + 0.000\ 009\ 216\ 9T_{UT1}^2 + 0.000\ 000\ 085\ 21T_{UT1}^3$$

$$\lambda_M = 0.874\ 016\ 75 + 21.329\ 909\ 545T_{UT1} + 0.000\ 003\ 665\ 9T_{UT1}^2 - 0.000\ 000\ 000\ 337T_{UT1}^3$$

Uranus

$$i = 0.013\ 494\ 82 - 0.000\ 029\ 442T_{UT1} + 0.000\ 000\ 060\ 9T_{UT1}^2 + 0.000\ 000\ 000\ 28T_{UT1}^3$$

$$\Omega = 1.291\ 647\ 44 + 0.001\ 294\ 094T_{UT1} + 0.000\ 007\ 075\ 6T_{UT1}^2 + 0.000\ 000\ 001\ 82T_{UT1}^3$$

$$\tilde{\omega} = 3.019\ 509\ 65 + 0.001\ 558\ 939T_{UT1} - 0.000\ 001\ 652\ 8T_{UT1}^2 + 0.000\ 000\ 007\ 21T_{UT1}^3$$

$$\lambda_M = 5.481\ 293\ 87 + 7.478\ 159\ 856T_{UT1} - 0.000\ 000\ 084\ 8T_{UT1}^2 + 0.000\ 000\ 000\ 10T_{UT1}^3$$

Neptune

$$\begin{aligned} i &= 0.030\,891\,49 + 0.000\,003\,939T_{UT1} + 0.000\,000\,004\,0T_{UT1}^2 - 0.000\,000\,000\,007T_{UT1}^3 \\ \Omega &= 2.300\,065\,70 - 0.000\,107\,601T_{UT1} - 0.000\,000\,038\,2T_{UT1}^2 - 0.000\,000\,001\,36T_{UT1}^3 \\ \tilde{\omega} &= 0.839\,916\,86 + 0.000\,508\,915T_{UT1} + 0.000\,001\,230\,6T_{UT1}^2 - 0.000\,000\,000\,007T_{UT1}^3 \\ \lambda_M &= 5.311\,886\,28 + 3.813\,303\,564T_{UT1} + 0.000\,000\,010\,3T_{UT1}^2 - 0.000\,000\,000\,03T_{UT1}^3 \end{aligned}$$

Pluto

$$\begin{aligned} i &= 0.299\,015\,6 \\ \Omega &= 1.926\,957 \\ \tilde{\omega} &= 3.920\,268 \\ \lambda_M &= 3.820\,304\,9 \end{aligned}$$

D.5 Data Sources*Gravitational Models:*

The Joint Gravitational Model is available through the Center for Space Research at the University of Texas at Austin. An anonymous ftp node is available at

[ftp.csr.utexas.edu](ftp://ftp.csr.utexas.edu) (128.83.154.16)

Other sources include NASA Goddard's anonymous ftp site at
[geodesy.gsfc.nasa.gov](ftp://geodesy.gsfc.nasa.gov)

Polar Motion and Timing Coefficients:

Coefficients of polar motion and timing are available through the National Earth Orientation Service (NEOS) at the U.S. Naval Observatory (USNO). The International Earth Rotation Service (IERS) Bulletins A and B are both available. You must first contact Mr. Schmidt for a password.

USNO, Attn: Mr. Richard Schmidt

3450 Massachusetts Ave., N.W.

Washington D.C. 20392-5420

Phone: (202) 653-0487 or E-mail: res@uttle.usno.navy.mil.

The number for the modem is (202) 653-0597, with 8 bits, 1 stop, no parity.

Internet anonymous ftp sites include

[maia.usno.navy.mil](ftp://maia.usno.navy.mil) (/ser7/finals.data)

Atmospheric Constants:

The National Oceanic and Atmospheric Administration (NOAA) is the best source for actual and predicted atmospheric constants.

ftp through
gopher.sel.noaa.gov
Phone (303) 497-5000
300, 1200, 2400 Baud
8 Bit, 1 stopbit, no parity
Viola Hill-Raben
(303) 497-5691
hill@sel.blrdoc.gov

Some anonymous ftp sites are
cddis.gsfc.nasa.gov.
ftp.ngdc.noaa.gov (/STP/GEOMAGNETIC_DATA/INDICES/Kp-Ap)

Planetary Ephemerides:

Planetary ephemerides are available through the Jet Propulsion Laboratory. You can get them in binary or ASCII form on your PC through an anonymous ftp server:

navigator.jpl.nasa.gov [128.149.23.82]

Send requests and current information to the Jet Propulsion Laboratory.

Dr. Myles Standish, Jr.
4800 Oak Drive
Pasadena, CA, 91109
Phone: (818) 354-3959

Initial Conditions:

Send requests for 2-line element sets to Space Command. If you're an authorized user, you may use a bulletin board at (719) 474-3390.

USSPACECOM/J36COS
1 Norad Road, Suite 9-102
Cheyenne Mtn AFB, CO 80914-6022
Phone: (719) 474-3670

They are also available from Bulletin Boards (205) 409-9280, and (513) 427-0674, 24 hours/day, 300, 1200, 2400, 4800, or 9600 bps, 8 data bits, 1 stop bit, no parity.

D.6 Computer Programming

In this book we focus mainly on presenting workable astrodynamical routines through computer programming. Many organizations shy away from computer programming and therefore do little with it. This section documents *sound* and *tested* programming practices which, if applied rigorously in *any* language, will result in code that is more efficient, faster running, and more accurate. They will also greatly reduce the lifecycle costs for maintaining this computer code.

Engineers must firmly grasp good programming techniques to correct problems in existing code. Much of today's code was developed decades ago when documentation and efficiency weren't as important. In fact, many of the original programmers are no longer alive, survived only by their code. Fortunately (or unfortunately), much of the code still works, but continual fixes to huge, ill-conceived programs have begun to create problems. Examples include inadvertent launches of bomber aircraft in anticipation of nuclear war (NORAD 1979), loss of life in accidental operations, and economic and financial disasters like the Vancouver Stock exchange in 1974. Whatever the case, computer code is very important, and it demands knowledge and care.

Most astrodynamics code is in some form of FORTRAN. Although the code usually works, you'll quickly see the horrors of programming legacy whenever you need to change it. You'll need to learn every version of FORTRAN, tricks to make a program think it is using characters, and so on. The most common excuse for this mess is that we don't have enough time and money to fix the problem. Weak as this argument is, it has stood up, and only a huge loss of life may be able to change it. Competent engineers will be prepared for that time, so I've tried to help by developing the companion computer code.

I'm providing the following checklist as a brief review of considerations for using the computer to program the algorithms in this book. They're general and can apply to any programming language.

Number, order, type. This is the *number one* computer programming error. The formal parameters (arguments passed back and forth in the procedure or subroutine call) must have the identical number of variables. They must be in the same order (for example, position, velocity, time) each time the routine is accessed, and they must be the same type (for example, vector, vector, real in the last example).

Proper control structures. There are three broad categories of control structures: assignment, selection, and iteration. It's very important to use the correct structure. {perhaps the most common abuse in this area is the misuse of iteration constructs. For iterations that execute a known number of times, you should use a FOR construct. When you don't know the number of iterations, use a WHILE or a REPEAT-UNTIL construct. Never use a jump to exit a control structure, it is perhaps the worst programming mistake that can be made.

Documentation. The programmer has an important responsibility to document their code. In times of tight budgets, documentation is often deleted or left out altogether. Although this may seem reasonable in the short run, it will always haunt the program or system later on. Documenting your work is difficult, and can be time consuming, but you'll convince yourself the first time you or someone else reuses some of your code after 10 or 15 years, and doesn't have to spend weeks of effort trying to remember why the program was built the way it was.

Variable names. Many programs still adhere to a six to eight maximum length for a variable name. This practice originated decades ago when computers didn't have sufficient memory to store longer names. Unfortunately, this practice persists today, and is a serious impediment to internal code documentation. On the other hand, some programmers have gone to the other extreme, creating variable names that are 20 and 30 characters

long. Use of common sense is encouraged. Finally, you should try to use variable names that are identical to the mathematical documentation. For instance, a ballistic coefficient could be *BC*.

Modularity. Some sources suggest a maximum line number for a routine. Following this practice often causes an algorithm to become fragmented. I suggest that an algorithm should be kept intact unless breaking out certain sections is necessary for use with other applications. Keeping the algorithm intact has the potential to create some “long” routines, making documentation even more important.

Indentation. Several decades ago, computer memory and disk space were limited, and extra lines and spaces were deleted to make the code compact. Unfortunately, this compaction hinders readability and maintenance. Be sure to indent all control structures.

Testing. The final aspect of computer programming receives the least emphasis, dollars, and discussion. Testing and validating code are extremely important because once a program is operating, all results are assumed to be correct. Unfortunately, constraints on budgets create many programs with insufficient testing. But we should try to test “enough,” *especially* when using older, undocumented code, by

1. Testing each occurrence for a different dynamic occurrence. For instance, if the routine is designed to propagate orbits, include a test for each orbit type (circular, elliptical, parabolic, and hyperbolic). This type of testing is common but often not exercised fully and limited by undocumented systems. Avoid these shortcomings at all costs.
2. Testing each computational method. Exercise each WHILE, IF, and so on, in the code. Instead of relying on testing of type one, above, fully exercise the *code itself*. Also extend the process to test each tolerance employed. Although correct data structures can eliminate many of these types of tests, adequate tolerance checking is necessary to ensure smooth performance over the range of possible values.
3. As more situations and tests arise, develop new test cases to match them. Actual operation over time is a valuable tool in refining computer code. Limited time usually causes the omission of this step, despite its tremendous potential.

The bottom line in testing—be thorough and write each test case down with its rationale to ease future operations.

Commercial Computer Programs. Finally, a wide variety of programs exist for commercial use. One study lists features of several hundred computer programs that provide astrodynamics capability, Boelitz (1995). You should consult this important and comprehensive study to determine if any software exists that will solve your problems. The report undergoes periodic revision at the Air Force Phillips Laboratory.

The Internet also provides access to computer software. A wide variety of sophisticated routines is available on Netlib (<http://www.netlib.com>).

References

- Numbers in brackets after each reference indicate the chapter in which references occur.
- 6570th Aerospace Medical Research Laboratory. 1962. *Relative Motion between an Interceptor and a Target in a Circular Orbit*. AMRL-TDR-62-124. Aerospace Medical Division. Wright-Patterson AFB, Ohio. [5]
- Abramowitz, M., and I. Stegun. 1970. *Handbook of Mathematical Functions*. New York: Dover Publications. [5]
- Alfano, Salvatore. 1982. Low Thrust Orbit Transfer. Master's thesis, Department of Aeronautics and Astronautics. Air Force Institute of Technology, Ohio. [5]
- _____. 1991. Personal discussions. U.S. Air Force Academy. [3]
- _____. 1993. Rapid Generation of Site/Satellite Timetables. *Journal of Spacecraft and Rockets*. 30(6): 760–764. [10, C]
- _____. 1994. Determining Satellite Close Approaches. Part II. *Journal of the Astronautical Sciences*. 42(2): 143–152. [10, C]
- Alfano, Salvatore, David Negron, Jr., and Jennifer L. Moore. 1992. Rapid Determination of Satellite Visibility Periods. *Journal of the Astronautical Sciences*. 40(2): 281–296. [10, C]
- Alfano, Salvatore, David Negron, Jr., and Daniel D. Wright. 1992. The Method of Ratios. *Journal of the Astronautical Sciences*. 40(2): 297–309. [10]
- Alfano, Salvatore, and David Negron, Jr. 1993. The Complete Method of Ratios. *Journal of Spacecraft and Rockets*. 30(6): 374–379. [10, C]
- _____. 1993. Determining Satellite Close Approaches. *Journal of the Astronautical Sciences*. 41(2): 217–225. [10]
- Alfano, Salvatore, and James D. Thorne. 1994. Circle to Circle Constant-thrust Orbit Raising. *Journal of the Astronautical Sciences*. 42(1): 35–45. [5]
- Allan, R. R., and G. E. Cook. 1964. The Long Period Motion of the Plane of a Distant Circular Orbit. *Proceedings of the Royal Society London*. 280(1380): 97. [8]
- Almanac for Computers 1990*. Washington, DC: U.S. Government Printing Office. [3]
- Amelina, T., et al. 1996. Comparison of Atmospheric Density Models. Presented at the US-Russia Space Surveillance Workshop. Poznan, Poland. [7, B]
- Aoki, S. et al. 1982. The New Definition of Universal Time. *Astronomy and Astrophysics*. Vol. 105: 359–361. [1]
- Asimov, Isaac. 1971. *The Stars in their Courses*. New York: ACE Books. [1]
- Astronomical Almanac*. Annual printing. Washington, DC: U.S. Government Printing Office. [1, 2, 3]
- Baker, Robert M., and Maud W. Makemson. 1967. *An Introduction into Astrodynamics*. New York: Academic Press. [6, 7]
- Baker, Robert M. 1967. *Astrodynamics, Applications and Advanced Topics*. New York: Academic Press. [6, 7, 8, 9]
- Barclay, William. 1975. *The Letters to the Phillipians, Colossians, and Thessalonians*. Louisville, KY: Westminster Press. [1]

- Barker, William N., S. J. Casali, and R. N. Wallner. 1995. The Accuracy of General Perturbations and Semianalytic Satellite Ephemeris Theories. Paper AAS-95-432 presented at the AAS/AIAA Astrodynamics Conference, Halifax, Nova Scotia, Canada. [8]
- Barlier, F., et al. 1978. A Thermospheric Model based on Satellite Drag Data. *Annales de Geophysicae*. 34(1): 9-24. [7]
- Bashforth, F., and J. C. Adams. 1883. *An Attempt to Test the Theories of Capillary Action... with an Explanation of the Method of Integration Employed*. New York: Cambridge University Press. [7]
- Bate, Roger R., D. D. Mueller, and J. E. White. 1971. *Fundamentals of Astrodynamics*. New York: Dover Publications. [4, 5, 6, 7, 8]
- Battin, Richard H. 1987. *An Introduction to the Mathematics and Methods of Astrodynamics*. New York: AIAA Education Series. [4, 6, 7, 8, 9]
- Beer, Arthur and K. A. Strand. 1975a. *Copernicus*. Oxford: Pergamon Press. [2]
- . 1975b. *Kepler*. Oxford: Pergamon Press. [1, 2, 4, 6]
- Bell, Eric T. 1965. *Men of Mathematics*. New York: Simon & Schuster Inc. [6, 7, 8, 9]
- Berry, Arthur. 1961. *A Short History of Astronomy*. New York: Dover Publications. [1, 2, 4]
- Bierman, Gerald. 1977. *Factorization Methods for Discrete Sequential Estimation*. New York: Academic Press. [9]
- Blitzer, Leon. 1970. *Handbook of Orbital Perturbations*. Unpublished report. TRW, Inc. [7, 8]
- Boelitz, Carole. 1992. Confidence in Atmospheric Density Forecasts. Paper AAS-92-181 presented at the AAS/AIAA Spaceflight Mechanics Meeting, Colorado Springs, CO. [7]
- . 1995. Orbit Analysis Software Survey. Phillips Laboratory Technical Report. PL-TR-95-1139. [D]
- Borkowski, K. M. 1989. Accurate Algorithms to Transform Geocentric to Geodetic Coordinates. *Bulletin Geodesique* Vol. 63: 50-56. [3]
- Boyce, William E., and Richard C. DiPrima. 1977. *Elementary Differential Equations and Boundary Value Problems*. New York: John Wiley & Sons. [7]
- Bretagnon, P. 1982. Théorie du mouvement de l'ensemble des planètes Solution VSOP82. *Astronomy and Astrophysics*. Vol. 114: 278-288. [3]
- Bretagnon, P., and G. Francou. 1988. Planetary Theories in Rectangular and Spherical Variables, VSOP87 Solutions. *Astronomy and Astrophysics*. Vol. 114: 309-315. [3]
- Brewer, J.A., and D. C. Anderson. 1977. Visual Interaction with Overhauser Curves and Surfaces. *Computer Graphics*, SIGGRAPH-ACM. 11(2): 132-137. [C]
- Brouwer, Dirk. 1959. Solutions of the Problem of Artificial Satellite Theory Without Drag. *Astronomical Journal*. 64(1274): 378-397. [8]
- Brouwer, Dirk, and G. M. Clemence. 1961. *Methods of Celestial Mechanics*. New York: Academic Press, Inc. [1, 7, 8]
- Brouwer, Dirk, and G. I. Hori. 1961. Theoretical Evaluation of Atmospheric Drag Effects on the Motion of an Artificial Satellite. *Astronomical Journal*. 66(5): 193-225. [8]
- Bryson, Arthur E., and Y. Ho. 1987. *Applied Optimal Control*. Washington, DC: Hemisphere Publishing Corporation. [5, 9]

- Buhler, Walter K. 1981. *Gauss, A Biographical Study*. New York: Springer-Verlag. [6, 10]
- Burton, James R., and Robert G. Melton. 1992. Variation of Parameters using Battin's Universal Functions. Paper AAS-92 125 presented at the AAS Astrodynamics Conference. [8]
- Cangahuala, Laureano et al. 1995. Topex/POSEIDON Orbit Determination: Quick Look Operations with GPS and Laser Tracking. Paper AAS 95-368 presented at the AAS/AIAA Astrodynamics Conference, Halifax, Nova Scotia, Canada. [6]
- Caprara, Giovanni. 1986. *The Complete Encyclopedia of Space Satellites*. New York: Portland House. [10]
- Caputo, Michele. 1967. *The Gravity Field of the Earth*. New York: Academic Press. [7]
- Carlton-Wippern, Kitt C. 1990. Statistical Analysis of Elliptical Keplerian Orbits, with Applications to Search and Surveillance Algorithms. *Space Power*, 9(4): 349–364. [10]
- Carter, David, Wayne D. McClain, and Paul J. Cefola. 1987. LANDSAT Orbit Determination Study Technical Report. #CSDL-R-1952. Cambridge, MA: Charles Stark Draper Laboratory. [7, B]
- Carter, David. 1991. When is the Groundtrack drift rate zero? CSDL Memorandum ESD-91-020. Cambridge, MA: Charles Stark Draper Laboratory. [10]
- . 1996. Personal correspondence. [6]
- Cassato, Stefano. 1989. Nominal Ocean Tide Models for TOPEX Precise Orbit Determination. Ph.D. Dissertation. University of Texas at Austin. [7, 8]
- Cayley, A. 1861. Tables of the Developments of Functions in the Theory of Elliptic Motion. *Memoirs of the Royal Astronomical Society*. Vol. 29: 191–306. [8]
- Cefola, Paul J. 1972. Equinoctial Orbit Elements—Application to Artificial Satellite Orbits. AIAA paper 72-937 presented at the AIAA/AAS Astrodynamics Conference, Palo Alto, CA. [8]
- . 1997. Personal Discussions. [8]
- Cefola, Paul J., Anne C. Long, and G. Holloway. 1974. The Long-Term Prediction of Artificial Satellite Orbits. Paper AIAA-74-170 presented at the AIAA Aerospace Sciences Meeting, Washington, DC. [8]
- Cefola, Paul J., and Roger A. Broucke. 1975. On the Formulation of the Gravitational Potential in Terms of Equinoctial Variables. AIAA Paper 75-9 presented at the AIAA Aerospace Sciences Meeting, Pasadena, CA. [8]
- Cefola, Paul J., and Wayne D. McClain. 1978. A Recursive Formulation of the Short-Periodic Perturbations in Equinoctial Variables. Paper AIAA-78-1383 presented at the AIAA/AAS Astrodynamics Specialist Conference, Palo Alto, CA. [8]
- Cefola, Paul J., D. J. Fonte, and N. Shah. 1996. The Inclusion of the Naval Space Command Theory PPT2 in the R&D GTDS Orbit Determination System. Paper AAS-96-142 presented at the AAS Spaceflight Mechanics Meeting, Austin, Texas. [8]
- Champion, K. S. W. 1963. Atmospheric Structure and its Variations in the Lower Atmosphere. Technical Report AD-417-201. Air Force Cambridge Research Laboratory: Hanscom AFB, MA. [7]
- Chetty, P. R. K. 1988. *Satellite Technology and its Applications*. Blue Ridge Summit, PA: Tab Books. [6]
- Chilulli, Roy M. 1994. *International Launch Site Guide*. The Aerospace Press. El Segundo, CA. [5]

- Chobotov, Val A. 1991. *Spacecraft Attitude Dynamics and Control*. Malabar FL: Krieger Publishing. [8, 10]
- Clohessy, W. H., and R. S. Wiltshire. 1960. Terminal Guidance System for Satellite Rendezvous. *Journal of the Aerospace Sciences*. 27(9): 653–674. [5]
- Coffey, Shannon, and A. Deprit. 1980. Short Period Elimination in Satellite Theory. Paper AIAA 80-1657. *Journal of Guidance and Control*. [8]
- Colliers Encyclopedia*. 1992. Vol. 19. New York: MacMillan Educational Co. [1]
- Cook, G. E. 1962. Luni-Solar Perturbations of the Orbit of an Earth Satellite. *Geophysical Journal of the Royal Astronomical Society*. Vol. 6: 271. [8]
- Cook, G. E., and D. W. Scott. 1967. Lifetimes of Satellites in Large Eccentricity Orbits. *Planetary Space Science*. Vol. 15: 1549-1556. [8]
- COSPAR Working Group. 1965. *COSPAR International Reference Atmosphere*. Amsterdam: North Holland Pub. Co. [7]
- Curtis, Anthony R. 1994. *Space Satellite Handbook*. Houston, Texas: Gulf Publishing Co. [10]
- Cutting, E., G.H. Born, J. C. Frautnick. 1978. Orbit Analysis for SEASAT-A. *Journal of the Astronautical Sciences*. 26(4): 315–342. [10]
- Danby, John M. A. 1992. *Fundamentals of Celestial Mechanics*. Richmond VA: Willmann Bell Inc. [2, 4, 7]
- Danielson, Donald A. 1996. *Vectors and Tensors in Engineering and Physics*. Reading, MA: Addison-Wesley Publishing. [1]
- Danielson, Donald A., Beny Neta, and Leo W. Early. 1994. *Semianalytical Satellite Theory (SST): Mathematical Algorithms*. Technical Report NPS-MA-94-001. Monterey, CA: Naval Postgraduate School. [8]
- Danielson, Donald A. et al. 1995. *Semianalytical Satellite Theory (SST): Mathematical Algorithms*. Technical Report NPS-MA-95-002. Monterey, CA: Naval Postgraduate School. [8]
- Davis, G. W., John C. Ries, and Byron D. Tapley. 1993. Preliminary Accuracy Assessment of the DORIS Tracking System on TOPEX/Poseidon. Paper AAS-93-574 presented at the AAS/AIAA Specialist Conference, Victoria, BC, Canada. [6]
- Defense Mapping Agency. 1983. *Geodesy for the Layman*. DMA-TR 80-003. Washington, DC: Headquarters, Defense Mapping Agency. [1]
- . 1987a. *Department of Defense World Geodetic System 1984 DMA*. DMA-TR 8350.2. Washington, DC: Headquarters, Defense Mapping Agency. [1]
- . 1987b. *Supplement to Department of Defense World Geodetic System 1984 DMA Technical Report: Part I - Methods, Techniques, and Data Used in WGS 84 Development*. DMA-TR 8350.2-A. Washington, DC: Headquarters, Defense Mapping Agency. [1]
- de Lafontaine, Jean. 1986. Orbital Dynamics in a Stochastic Atmosphere and a Nonspherical Gravity Field. Ph.D. thesis. University of Toronto Canada: Institute for Aerospace Studies. [7]
- Deprit, Andre. 1969. Canonical Transformations Depending on a Small Parameter. *Celestial Mechanics*. 1(1): 12–30. [8]
- . 1981. The Main Problem in the Theory of Artificial Satellites to Order Four. *Journal of Guidance and Control*. 4(2): 201–206. [8]

- Deprit, Andre, and Shannon Coffey. 1982. Third-Order Solution to the Main Problem in Satellite Theory. *Journal of Guidance, Control, and Dynamics*. 5(4): 366-371. [8]
- Der, Gim. 1995. Runge-Kutta Integration Methods for Trajectory Propagation Revisited. Paper AAS 95-420 presented at the AAS/AIAA Astrodynamics Specialist Conference, Halifax, Nova Scotia, Canada. [7]
- Der, Gim, and Roy Danicheck. 1996. Conversion of Osculating Orbital Elements to Mean Orbital Elements. Internal TRW report. Publication Pending. [9]
- Dreyer, J. L. E. 1953. *A History of Astronomy from Thales to Kepler*. New York: Dover Publications. [1, 2]
- Dyar, Walter. 1993. Comparison of Orbit Propagators in the Research and Develop Goddard Trajectory Determination System (R&D GTDS). M.S. thesis. Monterey, CA: Naval Postgraduate School, Department of Applied Mathematics. [8]
- Early, Leo W. 1986. A Portable Orbit Generator Using Semianalytical Satellite Theory. Paper AIAA 86-2164-CP presented at the AIAA/AAS Astrodynamics Conference. Williamsburg, VA. [8]
- Edelbaum, Theodore N. 1961. Propulsion Requirements for Controllable Satellites. *American Rocket Society Journal*. Vol. 31: 1079-1089. [5]
- Edelbaum, Theodore N., L. L. Sackett, and H. L. Malchow. 1973. Optimal Low Thrust Geocentric Transfer. Paper AIAA-73-1074 presented at the AIAA 10th Electric Propulsion Conference. Lake Tahoe, Nevada. [5]
- Eller, Thomas J. et al. 1994. Possible Correlation Between Geomagnetic Planetary Index and Satellite Catalog Accuracy. Paper AIAA-94-3727 presented at the AIAA/AAS Conference, Scottsdale, AZ. [7]
- Escobal, Pedro R. [1965] 1985. *Methods of Orbit Determination*. New York: John Wiley & Sons. (Reprint edition. Malabar FL: Krieger Publishing Co.) [2, 3, 4, 6, 7, 8, 10, C]
- _____. [1968] 1979. *Methods of Astrodynamics*. New York: John Wiley & Sons. Reprint: Malabar FL: Krieger Publishing Co. [5, 6, 7]
- Explanatory Supplement to the Astronomical Ephemeris and the American Ephemeris and Nautical Almanac*. 1961. London: Her Majesty's Stationery Office. [1]
- Fehlberg, Erwin. 1968. Classical Fifth-, Sixth-, Seventh-, and Eighth-Order Runge-Kutta Formulas with Stepsize Control. NASA Technical Report, TR-R-287. [7]
- _____. 1969. Low-Order Classical Runge-Kutta Formulas with Stepsize Control and their Application to some Heat Transfer Problems. NASA Technical Report, TR-R-315. [7]
- Ferguson, Jack. 1978. *Statistical Orbit Determination*. Astro 422 reading. Colorado Springs, CO: US Air Force Academy. [9]
- Fieger, Martin. 1987. An Evaluation of Semianalytical Satellite Theory Against Long Arcs of Real Data for Highly Eccentric Orbits. M. S. thesis, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology. CSDL-T-938. [8]
- Fisher, R. A. 1912. On an Absolute Criterion for Fitting Frequency Curves. *Messenger of Mathematics*. Vol. 41: 155. [9]
- Fitzpatrick, Philip M. 1970. *Principles of Celestial Mechanics*. New York: Academic Press. [7, 8]

- Fonte, Daniel J. Jr. 1993. Implementing a 50×50 Gravity Field Model in an Orbit Determination System. M.S. Thesis, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology. CSDL-T-1169. [8]
- _____. 1994. PC Based Orbit Determination. Paper AIAA 94-3776 presented at the AIAA/AAS Astrodynamics Conference, Scottsdale AZ. [9]
- _____. 1994. Evaluation of Orbit Propagators for the HI-CLASS Program. Technical Report PL-94-1017. Kirtland Air Force Base, NM: Phillips Laboratory. [8]
- _____. 1995 (March 30). HI-CLASS. Paper STK-235, Vol. 1 presented at the Space Surveillance Workshop at MIT Lincoln Laboratory, MA [9]
- Fonte, Daniel J. Jr., and Chris Sabol. 1995. Optimal DSST Input Decks for Various Orbit Types. Phillips Laboratory Technical Report. PL 95-1072. Kirtland AFB, NM: Phillips Laboratory. [9]
- Frank, Mark. 1984. *The New FK5 Mean of J2000 Time and Reference Frame Transformations*. DVO Technical Report 84-2. [1]
- Fraser-Smith, A. C. 1972. Spectrum of the Geomagnetic Activity Index A_p . *Journal of Geophysical Research*. 77(22): 4209-4220. [7]
- Fricke, W. 1982. Determination of the Equinox and Equator of the FK5. *Astronomy and Astrophysics*, Vol. 107: L13-L16. [1]
- Gabor, Michael J. 1993. Orbit Determination for the GEOSAT Exact Repeat Mission Utilizing TRANET Doppler Data. M. S. Thesis. University of Texas at Austin. [9]
- Gammon, Richard H. *Chemistry Between the Stars*. U.S. Government Printing Office: September 1976. [6]
- Gaposchkin, E. M., and A. J. Coster. 1987. Evaluation of New Parameters for Use in Atmospheric Models. Paper AAS-87-555. *Proceedings of 1987 AAS/AIAA Astrodynamics Conference*. San Diego, CA: AAS Publications Office. [8]
- _____. 1987. Evaluation of Recent Atmospheric Density Models. Paper AAS-87-557. *Proceedings of 1987 AAS/AIAA Astrodynamics Conference*. San Diego, CA: AAS Publications Office. [8]
- Gauss, Karl F. [1809] 1963. *Theory of the Motion of the Heavenly Bodies Moving about the Sun*. Translation and Reprint. New York: Dover Publications. [1, 7]
- Gedeon, G. S. 1968. *Luni-Solar Perturbations*. TRW Report No. 3414.5-59. Redondo Beach, CA: TRW, Inc. [8]
- Gelb, Arthur. 1989. *Applied Optimal Estimation*. Cambridge, MA: MIT Press. [9]
- Geyling, Franz T., and R. H. Westerman. 1971. *Introduction to Orbital Mechanics*. Reading, MA: Addison-Wesley Publishing. [7, 8]
- Gibbs, Josiah W. 1889. On the Determination of Elliptic Orbits from Three Complete Observations. *Memoires National Academy of Science*: 4 (2): 79-104. [6]
- Gooding, R. H. 1990. A Procedure for the Solution of Lambert's Orbital Boundary-Value Problem. *Celestial Mechanics and Dynamical Astronomy*. 48(2): 145-165. [6]
- Green, Robin M. 1988. *Spherical Astronomy*. New York: Cambridge University Press. [1, 3]
- Griffin, Michael D., and James R. French. 1991. *Space Vehicle Design*. Washington DC: AIAA Press. [1]

- Gutzwiller, Martin C., and Dieter S. Schmidt. 1986. *The Motion of the Moon as Computed by Hill, Brown, and Eckert*. Washington, DC: U.S. Government Printing Office. [8]
- Hansen, P. A. 1855. Expansion of the Power of the Radius Vector with the Sine or Cosinus of a Multiple of the True Anomaly in Terms of a Series. *Abhandlungen der Königlichen Sächsischen Gesellschaft für Wissenschaft*. 2(3): 183–281. [8]
- Harris, I., and W. Priester. 1962. Theoretical Models for the Solar Cycle Variations of the Upper Atmosphere. *Journal of Geophysical Research*. 67(12): 4585–4591. [7]
- Hart, Andrew. 1991. Application of the Homotopy Continuation Method to Low-Eccentricity Preliminary Orbit Determination. M.S. thesis. Massachusetts Institute of Technology. [6]
- Hedin, A. E., et al. 1977. A Global Thermospheric Model Based on Mass Spectrometer and Incoherent Scatter Data. MSIS-1 and 2, N₂ Density and temperature, Composition. *Journal of Geophysical Research*. Vol. 82: 2139–2156. [7]
- Hedin, A. E. 1987. MSIS-86 Thermospheric Model. *Journal of Geophysical Research*. Vol. 92: 4649–4662. [7]
- Herrick, Samuel. 1971. *Astrodynamicics: Orbit Determination, Space Navigation, Celestial Mechanics* (Vol. I). London: Van Nostrand Reinhold Co. [4, 6, 7]
- . 1972. *Astrodynamicics: Orbit Correction, Perturbation Theory, Integration* (Vol. II). London: Van Nostrand Reinhold Co. [7, 8]
- Herriges, Darrell. 1988. NORAD General Perturbation Theories: An Independent Analysis. M.S. thesis. CSDL-T-972. Massachusetts Institute of Technology, Department of Aeronautics and Astronautics. [8]
- Hildebrand, F. B. 1987. *Introduction to Numerical Analysis*. New York: Dover Publications. [C]
- Hill, George W. 1878. Researches in Lunar Theory. *American Journal of Mathematics*. Vol. 1: 5–26. [5]
- Hill, Louis A., Jr. 1981. *Structured Programming in FORTRAN*. Englewood Cliffs, NJ: Prentice Hall. [C]
- Hobson, E. W. [1931] 1965. *The Theory of Spherical and Ellipsoidal Harmonics*. New York: Chelsea Publishing. [7]
- Hoffmann-Wellenhof, B., H. Lichtenegger, and J. Collins. 1994. *GPS Theory and Practice*. New York: Springer-Verlag. [10]
- Hohmann, Walter. 1925. *Die Erreichbarkeit der Himmelskörper* (The Attainability of Heavenly Bodies). NASA Technical Translation TTF44. Nov 1960 Washington, DC: National Aeronautics and Space Administration. [5]
- Hoots, Felix. R. 1982. An Analytical Satellite Theory using Gravity and a Dynamic Atmosphere. Paper AIAA-82-1409 presented at the AIAA/AAS Astrodynamics Conference. San Diego, CA. [8]
- Hoots, Felix R., and Ronald L. Roehrich. 1980. *Models for Propagation of NORAD Element Sets*. Spacetrack Report #3. U.S. Air Force: Aerospace Defense Command. [2, 8]
- Hoots, Felix R., Linda L. Crawford, and Ronald L. Roehrich. 1984. An Analytic Method to Determine Future Close Approaches Between Satellites. *Celestial Mechanics*. 33(2): 143–158. [10]

- Hoots, Felix R., and R. G. France. 1987. An Analytic Satellite Theory using Gravity and a Dynamic Atmosphere. *Celestial Mechanics*. 40(1): 1–18. [8]
- Hovanessian, S. A. 1988. *Introduction to Sensor Systems*. Norwood, MA: Artech House. [6]
- Hujšak, Richard. 1979. A Restricted Four-body Solution for Resonating Satellites with an Oblate Earth. AIAA Paper 79-136. [8]
- . 1995. Using a Single SLR Site to Validate GPS Receiver Navigation. Paper AAS-95-369 presented at the AAS/AIAA Astrodynamics Conference, Halifax, Nova Scotia, Canada. [9, 10]
- Jablonski, [Boelitz] Carole A. 1992. Confidence in Atmospheric Density Forecasts. Paper AAS-92-181 presented at the AAS/AIAA Spaceflight Mechanics Conference, Colorado Springs, CO. [7]
- . 1991. *Application of Semianalytic Satellite Theory to Maneuver Planning*. M.S. thesis. Massachusetts Institute of Technology. #CSDL-T-1086. Cambridge, MA: Charles Stark Draper Laboratory, Inc. [7, 10]
- Jacchia, L. G. 1970. *New Static Models for the Thermosphere and Exosphere with Empirical Temperature Profiles*. SAO Special Report No. 313. Cambridge, MA: Smithsonian Institution Astrophysical Observatory. [7, B]
- . 1971. *Revised Static Models for the Thermosphere and Exosphere with Empirical Temperature Profiles*. SAO Special Report No. 332. Cambridge, MA: Smithsonian Institution Astrophysical Observatory. [7, B]
- Jane's Space Directory*. 1994. Alexandria, VA: Jane's Information Group. [5, 10]
- Junkins, John L. 1975. Optimal Estimation Theory. NSWC/DL Technical Note. TN-K-45/74. Dahlgren, VA: US Naval Weapons Laboratory. [9, C]
- . 1977. *An Introduction to Optimal Estimation of Dynamical Systems*. Sijthoff-Noordhoff. [9]
- Junkins, John L., Maruthi R. Akella, and Kyle T. Alfriend. 1996. Non-Gaussian Error Propagation in Orbital Mechanics. *Journal of the Astronautical Sciences*. (forthcoming). [8, 9]
- Jursa, Adolph S., ed. 1985. *Handbook of Geophysics and the Space Environment*. USAF Document ADA 167000. U.S. Air Force Geophysics Laboratory. [7, 8]
- Kalman, Rudolf E. 1960. A New Approach to Linear Filtering and Prediction Problems. *ASME Transactions - D, Journal of Basic Engineering*. 82D(1): 35–45. [9]
- Kane, Thomas R., Peter Likens, and David Levinson. 1983. *Spacecraft Dynamics*. New York: McGraw Hill. [8]
- Kaplan, Marshall H. 1976. *Modern Spacecraft Dynamics and Control*. New York: John Wiley & Sons. [5, 6, 7, 8]
- Kaula, William M. 1959. Statistical and Harmonic Analysis of Gravity. *Journal of Geophysical Research*. 64(12): 2401–2421. [7, 9]
- . 1966. *Theory of Satellite Geodesy*. Waltham MA: Blaisdell Publishing Co. [7, 8]
- . 1969. Tidal Friction and Longitude-Dependant Amplitude and Phase Angle. *Astronomical Journal*. 74(9): 1108–1114. [7, 8]
- Kennedy, J. 1993. Debris gets in Your Space. *Guardian, The Magazine of Air Force Space Command*. 1(1): 4–5. [10]

- Khutorovsky, Z. N. 1996. Low-perigee Satellite Catalog Maintenance: Issues of Methodology. Paper presented at U.S.-Russian Space Surveillance Workshop, Poland. [9]
- King-Hele, D. 1964. *Theory of Satellite Orbits in an Atmosphere*. London: Butterworths. [1]
- . 1987. *Satellite Orbits in an Atmosphere*. London: Blackie. [8, 9]
- King-Hele, D., G. E. Cook, and J. M. Rees. 1963. Determination of the Even Harmonics in the Earth's Gravitational Potential. *Geophysical Journal*. Vol. 8: 119. [8]
- Knowles, Stephen H. 1995. A Comparison of Geocentric Propagators for Operational Use. Paper AAS-95-429 presented at the AAS/AIAA Astrodynamics Conference, Halifax, Nova Scotia, Canada. [7, 8]
- Koestler, Arthur. 1959. *The Sleepwalkers*. New York: MacMillan Co. [4]
- Kozai, Yoshiohide. 1959. The Motion of a Close Earth Satellite. *Astronomical Journal*. 64(1274): 367–377. [8]
- . 1962. Second-Order Solution of Artificial Satellite Theory Without Drag. *Astronomical Journal*. Vol. 67: 446. [8]
- Kreyszig, Erwin. 1983. *Advanced Engineering Mathematics*. 5th ed. New York: John Wiley Publishing. [1, 7, 8]
- Lagrange, J. L. 1873. *Collected Works*. Vol. 6. Paris: Gauthier-Villars. [8]
- Lambeck, Kurt. 1988. *Geophysical Geodesy*. Oxford: Clarendon Press. [1, 7, 8]
- Lane, Mark T. 1991. A Numerical Approach to the Angles-Only Initial Orbit Determination Problem. Paper AAS-91-356 presented at the AAS/AIAA Astrodynamics Specialist Conference, Durango, CO. [6]
- Larson, Wiley J., and J. R. Wertz. 1992. *Space Mission Analysis and Design*. 2nd ed. Norwell, MA and Torrance, CA: Kluwer Academic Publications and Microcosm, Inc. [5, 10]
- Lawden, D. F. 1952. Orbital Transfer via Tangential Ellipses. *British Interplanetary Society Journal*. 11(6): 278–289. [5]
- Lawton, J. A. 1987. Numerical Method for Rapidly Determining Satellite-Satellite and Satellite-Ground Station In-View Periods. *Journal of Guidance, Navigation, and Control*. Vol. 10: 32–36. [10]
- Lisowski, Ronald. 1995. Personal discussions. U.S. Air Force Academy. [5]
- Liu, Joseph. 1973. *A Second-order Theory of an Artificial Satellite under the Influence of the Oblateness of the Earth*. M-240-1203. Northrop Services. [8]
- . 1974. Satellite Motion about an Oblate Earth. *AIAA Journal*. Vol. 12: 1511–1516. [8]
- Logsdon, Tom. 1992. *The Navstar Global Positioning System*. New York, NY: Van Nostrand-Rinehold. [10]
- Long, Anne C. et al. 1978. *System Description for the GTDS R&D Averaged Orbit Generator*. CSC/SD-78/6020. Goddard Space Flight Center: National Aeronautics and Space Administration. [8, 9]
- Long, Anne C. et al. 1989. *Goddard Trajectory Determination System (GTDS) Mathematical Theory (Revision 1)*. FDD/552-89/001 and CSC/TR-89/6001. Goddard Space Flight Center: National Aeronautics and Space Administration. [6, 7, 8, 9, 10, B]

- Lundberg, J. 1995. Mitigation of Orbit Errors Resulting from the Numerical Integration across Shadow Boundaries. Paper AAS-95-408 presented at the AAS/AIAA Astrodynamics Conference, Halifax, Nova Scotia, Canada. [7]
- Lundberg, J. et al. 1985. Surfaces of Zero Velocity in the Restricted Problem of Three Bodies. *Celestial Mechanics*. 36(2): 191–205. [2]
- Lyddane, Robert H. 1963. Small Eccentricities or Inclinations in the Brouwer Theory of the Artificial Satellite. *Astronomical Journal*. 68 (8): 555–558. [8]
- MacCullagh, J. 1855. On the Attraction of Ellipsoids with a New Demonstration of Clairut's Theorem. *Transactions of the Royal Irish Academy*. Vol. 22. Dublin. [7]
- Marcos, Frank A. 1991. Development and Validation of New Satellite Drag Models. Paper AAS-91-491 presented at the AAS/AIAA Astrodynamics Specialist Conference, Durango, CO. [8]
- Marcos, Frank A. et al. 1993. Satellite Drag Models: Current Status and Prospects. Paper AAS-93-621 presented at the AAS/AIAA Conference, Victoria BC, Canada. [7, 8]
- Marshall, J. A. et al. 1995. Precision Orbit Determination and Gravity Field Improvements Derived from TDRSS. Paper AAS-95-312 presented at the AAS/AIAA Astrodynamics Conference, Halifax, Nova Scotia, Canada. [6]
- Matthews, John H. 1987. *Numerical Methods*. New Jersey: Prentice Hall Inc. [7]
- Maury, Jesse L. Jr., and Gail P. Segal. 1969. Cowell Type Numerical Integration as Applied to Satellite Orbit Computation. Goddard Space Flight Center. NASA Technical Report TM-X-63542, X-553-69-46. [7]
- Maybeck, Peter S. 1979. *Stochastic Models, Estimation, and Control*. New York: Academic Press. [9]
- McCarthy, Dennis. 1992. *IERS Technical Note #13*. U.S. Naval Observatory. [1]
- McClain, Wayne D. 1978. *A Recursively Formulated First-Order Semianalytic Artificial Satellite Theory Based on the Generalized Method of Averaging*. Vol. 2. CSC/TR-78/6001; Computer Sciences Corporation. [8]
- . 1987. Eccentricity Control and the Frozen Orbit Concept. AAS/AIAA Paper No. AAS 87-516. Presented at Kalispell, Montana, Aug. 10–13. [10]
- . 1990. *Single Station Orbit Determination for Landsat 6*. Paper AIAA-90-2923 presented at the AIAA Astrodynamics Conference, Portland OR. [7, 10]
- . 1990. The Nodal Period. Unpublished interoffice memo. [10]
- . 1992. *Semianalytic Artificial Satellite Theory*. Vol. 1. Charles Stark Draper Laboratory. [8]
- Meeus, Jean. 1991. *Astronomical Algorithms*. Richmond, VA: Willmann Bell Inc. [1, 3, C, D]
- Merson, Robert H. 1961. The Motion of a Satellite in an Axi-Symmetric Gravitation Field. *Geophysical Journal of the Royal Astronomical Society*. Vol. 4: 17. [8]
- Metzinger Richard. 1993. *Validation of the Workstation Version of R&D GTDS*. Cambridge, MA: Charles Stark Draper Laboratory. [7]
- Moe, M. M. 1960. Solar-Lunar Perturbations of the Orbit of an Earth Satellite. *ARS Journal*. Vol. 30: 485. [8]

- Montenbruck, Oliver, and T. Pfleger. 1994. *Astronomy on the Personal Computer*. New York: Springer Verlag. [6, 7]
- Morgen, Walter L., and G. D. Gordon. 1989. *Communications Satellite Handbook*. New York: John Wiley & Sons. [6]
- Moritz, Helmut, and I. Mueller. 1987. *Earth Rotation—Theory and Observation*. New York: Ungar Publishing Company. [1, 7]
- Moulton, Forest R. 1914. *An Introduction to Celestial Mechanics*. New York: Dover Publications. [1, 6]
- Muolo, Michael J. 1993. *Space Handbook, A Warfighter's Guide to Space*. Vol. I. AU-18. Maxwell AFB, AL: Air University Press. [9]
- Musen, P. 1960. On the Motion of a Satellite in an Asymmetrical Gravitational Field. *Journal of Geophysical Research*. 65(9): 2783-2792. [8]
- National Geophysical Data Center. *Geomagnetic Indices Bulletin*. Boulder, CO: National Geophysical Data Center. [7]
- National Geophysical Data Center. *Solar Indices Bulletin*. Boulder, CO: National Geophysical Data Center. [7]
- Nayfeh, A. 1973. *Perturbation Methods*. New York: Wiley-Interservice Publications. [8]
- Nerem, R. S. et al. 1994. Gravity Model Developments for TOPEX / POSEIDON: Joint Gravity Models 1 and 2. *Journal of Geophysical Research*. 99 (C12): 24,421-24,447. [1, 7, D]
- Nerem, R. S. et al. 1995. Preliminary Results from the Joint GSFC/DMA Gravity Model Project. Paper AAS-95-310 presented at the AAS/AIAA Astrodynamics Conference, Halifax, Nova Scotia, Canada. [1, 8]
- Neta, Beny. 1996. Personal discussions. [7, 8, 9]
- Neter, John, William Wasserman, and G. A. Whitmore. 1982. *Applied Statistics*. Boston, MA: Allyn and Bacon Inc. [9]
- Newcomb, Simon. [1906] 1960. *A Compendium of Spherical Astronomy*. Reprint edition. New York: Dover Publications. [1]
- Newton, Sir Isaac. [1687] 1962. *Mathematical Principles of Natural Philosophy*. Reprint edition. Berkeley, CA: University of California Press. [1, 2, 5]
- Nostrand, Philip M. 1984. *Forecast Verification of the 10.7 Centimeter Solar and the Ap Daily Geomagnetic Activity Indices*. M.S. thesis. AFIT/GSO/PH-OS/84D-2. Wright-Patterson AFB: Air Force Institute of Technology. [7]
- Overhauser, A. W. 1968. *Analytic Definition of Curves and Surfaces by Parabolic Blending*. Tech Report No. SL68-40. Dearborn, MI: Ford Motor Company Scientific Laboratory. [C]
- Oza, D. H., and R. J. Friertag. 1995. Assessment of Semi-empirical Atmospheric Density Models for Orbit Determination. Paper AAS 95-101 presented at the AAS/AIAA Spaceflight Mechanics Conference. Austin TX. [7]
- Pannekoek, Anton. 1989. *A History of Astronomy*. New York: Dover Publications. [1, 2, 3, 4, 8]
- Phillips, Charles L., and H. T. Nagle Jr. 1984. *Digital Control System Analysis and Design*. Englewood Cliffs NJ: Prentice Hall Inc. [7]

- Phillips Laboratory. 1995. 1995 *Success Stories*. Kirtland Air Force Base, NM: Phillips Laboratory History Office. [9]
- Plummer, H. C. [1918] 1960. *An Introductory Treatise on Dynamical Astronomy*. Reprint edition. New York: Dover Publications. [4, 7]
- Pohlen, David J., and N. A. Titus. 1995. Orbit Propagation Using Bulirsch-Stoer Integration. Paper AAS-95-224 presented at the AAS/AIAA Spaceflight Mechanics Conference. Albuquerque, NM. [7]
- Press, William H. et al. 1992. *Numerical Recipes in Fortran*. Cambridge, England: Cambridge University Press. [5, 7]
- Proulx, Ron J. et al. 1981. A Theory for the Short-Periodic Motion Due to the Tesseroid Harmonic Gravity Field. Paper AAS-81-180 presented at the AAS/AIAA Astrodynamics Specialist Conference, Lake Tahoe, NV. [8]
- Prussing, John E. and B. A. Conway. 1993. *Orbital Mechanics*. New York: Oxford Press. [1, 6]
- Reid, Gary J. 1983. *Linear System Fundamentals*. New York: McGraw-Hill, Inc. [C]
- Roberts, Charles E., Jr. 1971. An Analytic Model for Upper Atmosphere Densities Based upon Jacchia's 1970 Models. *Celestial Mechanics*. 4(314): 368–377. [7, B]
- Roy, Archie E. 1988. *Orbital Motion*. New York: John Wiley & Sons. [3, 7, 8]
- Sabol, Chris. 1994. Applications of Sun-Synchronous Critically Inclined Orbits to Global Personal Communication Systems. MIT M.S. Thesis. #CSDL-T-1235. Cambridge, MA: Charles Stark Draper Laboratory, Inc. [8]
- Sadler, D. H. 1974. *Interpolation and Allied Tables*. London: Her Majesty's Nautical Almanac Office. [7, C]
- Schatten, Kenneth H., and W. D. Pesnell. 1993. An Early Solar Dynamo Prediction: Cycle 23~Cycle 22. *Geophysics Research Letters*. Vol. 20: 2275–2278. [7]
- Schatten, Kenneth H., and S. Sofia. 1988. A Model for Solar Constant Secular Changes. *Geophysics Research Letters*. Vol. 15: 121–124. [7]
- Schatten, Kenneth H., and S. Sofia. 1987. Forecast of an Exponentially Large Even-numbered Solar Cycle. *Geophysics Research Letters*. Vol. 14: 632–635. [7]
- Schumacher, Paul W., Jr., and R. A. Glover. 1995. Analytical Orbit Model for U.S. Naval Space Surveillance—an Overview. Paper AAS-95-427 presented at the AAS/AIAA Specialist Conference, Halifax, Nova Scotia, Canada. [9]
- Seidelmann, Kenneth. 1992. *Explanatory Supplement to the Astronomical Almanac*. California: University Science Books. [1, 3, D]
- Selby, Samuel M., ed. 1975. *Standard Mathematical Tables*. 23rd ed. Cleveland, OH: CRC Press, Inc. [4, 6, 7, 9, C]
- Sellers, Jerry J. 1994. *Understanding Space: An Introduction to Astronautics*. New York: McGraw-Hill, Inc. [2]
- Shaver, Jeffrey S. 1980. Formulation and Evaluation of Parallel Algorithms for the Orbit Determination Problem. MIT Ph.D. Dissertation. #CSDL-T-709. Cambridge, MA: Charles Stark Draper Laboratory, Inc. [9]

- Siewert, C. E. and E. E. Burniston. 1972. An Exact Analytical Solution of Kepler's Equation. *Celestial Mechanics*. 6(3): 294-304. [4]
- Slutsky, Mark. 1983. First Order Short Periodic Motion of an Artificial Satellite due to Third Body Perturbations. IOC AOD/SD 021-15Z-MS. Charles Stark Draper Lab: Cambridge, MA. [8]
- Small, Robert. [1804] 1963. *An Account of the Astronomical Discoveries of Kepler*. Reprint edition. Madison, WI: University of Wisconsin Press. [2, 4]
- Smith, Richard L., and C. Y. Huang. 1986. *Study of a Homotopy Continuation Method for Early Orbit Determination with the Tracking and Data Relay Satellite System (TDRSS)*. NASA Technical Memorandum 86230. [6]
- Smith, Richard L., and C. Y. Huang. 1985. *A Homotopy Continuation Method for General Preliminary Orbit Determination and Special Application to the Tracking and Data Relay Satellite System*. New York: American Institute of Aeronautics and Astronautics. [6]
- Smith, D. E. 1962. The Perturbation of Satellite Orbits by Extra-Terrestrial Gravitation. *Planetary and Space Science*. Vol. 9: 659-674. [8]
- Sorenson, Harold W. 1970. Least-squares Estimation: from Biases to Kalman. *IEEE Spectrum*. 7(7): 63-68. [9]
- Standish, Myles. 1990. The Observational Basis for JPL's DELOD, the Planetary Ephemerides of the Astronomical Almanac. *Astronomy and Astrophysics*. Vol. 233: 252-271. [7]
- Stephenson, Bruce. 1987. *Kepler's Physical Astronomy*. New York: Springer-Verlag. [2, 4, 6]
- Sterne, 1960. *An Introduction to Celestial Mechanics*. New York: Interscience Publications Inc. [8]
- Stiefel, Eduard L., and G. Scheifele. 1971. *Linear and Regular Celestial Mechanics*. New York: Springer Verlag. [7]
- Sundman, Karl F. 1913. Mémoire sur le problème des trois corps. *Acta Mathematica*. Vol. 36: 105-179. [4]
- Swerling, Peter. 1959. First Order Propagation in a Stage-wise Smoothing Procedure for Satellite Observations. *Journal of the Astronautical Sciences*. 6(3): 46-62. [9]
- Szebehely, Victor. 1967. *Theory of Orbits*. New York: Academic Press. [2, 7]
- . 1989. *Adventures in Celestial Mechanics*. Austin, TX: University of Texas Press. [7, 9]
- Taff, Laurence G. 1985. *Celestial Mechanics, A Computational Guide for the Practitioner*. NY: John Wiley & Sons. [1, 3, 4, 7, 8]
- Thompson, William T. 1986. *Introduction to Spacecraft Dynamics*. New York: Dover Publications. [5]
- Thorne, James D. 1989. Series Reversion/Inversion of Lambert's Time Function. M.S.Thesis. GA/ENY/89-D-6. Columbus, OH: Air Force Institute of Technology. [6]
- Thorne, James D., and R. D. Bain. 1995. Series Reversion/Inversion of Lambert's Time Function. *Journal of the Astronautical Sciences*. 43(3): 277-287. [4, 6]
- Thorne, James D., and Christopher D. Hall. 1996. Approximate Initial Lagrange Costates for Continuous-Thrust Spacecraft. *Journal of Guidance, Control, and Dynamics*. 19(2): 283-288. [5]

- Thuillier, G., J. L. Falin, and F. Barlier. 1977. Global Experimental Model of the Exospheric Temperature using Optical and Incoherent Scatter Measurements. *Journal of Atmospheric and Terrestrial Physics*. Vol. 39: 1195. [7]
- Tisserand, F. 1889. *Traité de mécanique céleste*, 4 vols. Paris: Gauthier-Villars. [8]
- U.S. Standard Atmosphere. 1976. Washington, DC: U.S. Government Printing Office. [7]
- Vallado, David A. 1991. Orbital Phasing. USAFA Astro-321 course reading. July. [5]
- _____. 1991a. Reconnaissance Mission Planning. Astro 422 course reading. Colorado Springs, CO: U.S. Air Force Academy. [10]
- _____. 1992a. Statistical Orbit Determination. Astro 422 Course Reading. Colorado Springs, CO: U.S. Air Force Academy. [7, 8, 9]
- _____. 1992b. Methods of Astrodynamics—A Computer Approach Ver 4.0. Colorado Springs: U.S. Air Force Academy. [all]
- _____. 1992c. Software Standard, *PASCAL and FORTRAN*, U.S. Air Force Academy. [D]
- _____. 1993a. A Study in Relative Motion. Phillips Laboratory, Kirtland AFB, NM. [5]
- _____. 1993b. A Study in Intercept and Rendezvous Techniques. Phillips Laboratory, Kirtland AFB, NM. [5, 6]
- Vallado, David A., and John M. Hanson. 1989. *Satellite Ephemeris Error Modeling (SEEM) Implementation on Government Microcomputers*. USSPACECOM/AN Technical Note, CAA-TN-89-010. Colorado Springs, CO: U.S. Space Command. [9]
- Vetter, Jerome R. 1994. The Evolution of Earth Gravity Models used in Astrodynamics. *APL Technical Digest*, John Hopkins. 15(4): 319–335. [1, 7]
- Voiskovskii, M.I., et al. 1973. An Aspherical Model for the Upper-Atmosphere Density. *Cosmic Research*, 11(1): 70–79. [7, B]
- Volkov, I.I., Ye.I. Knyazeva, and B. V. Kugayenko. 1982. Refining a Model of Atmospheric Density for Ballistic Calculations. Nablyudeniya Iskusstvennykh Nebesnykh Tel (Moskva), No. 80, Soviet Union. pp. 126–135. [7]
- Von Zeipel, H. 1916. *Recherches sur le mouvement des petites planètes*. Arkiv. Mat. Astron. Fys Col 11(1), 11(7), 12(9). [8]
- Wallace, Scott T. 1995. Parallel Orbit Propagation and the Analysis of Satellite Constellations. MIT M.S. Thesis. #CSDL-T-1245. Cambridge, MA: Charles Stark Draper Laboratory, Inc. [8]
- Walpole, Ronald E. 1974. *Introduction to Statistics*. New York: Macmillan Publishing. [9]
- Wertz, James R. 1978. *Spacecraft Attitude Determination and Control*. Dordrecht, Holland: D. Reidel Publishing Company. [1, 7, 8]
- Wheeler, J. A. 1990. *A Journey into Gravity and Spacetime*. San Francisco, CA: W. H. Freeman & Company. [10]
- Wiesel, William E., and Salvatore Alfano. 1985. Optimal Many-revolution Orbit Transfer. *Journal of Guidance*. 8(1): 155–157. [5]
- Williams, Trevor, and David Baughman. 1994. *Self-Rescue Strategies for EVA Crewmembers Equipped with SAFER Backpack*. Paper 28 presented at the Goddard Flight Mechanics/Estimation Theory Symposium. [5]

- Wright, James R. 1981. Sequential Orbit Determination with Auto-Correlated Gravity Modeling Errors. *AIAA Journal of Guidance and Control*. 4(2): 304. [9]
- _____. 1994a. Orbit Determination Solution to Non-Markov Gravity Error Problem. Paper AAS-94-176 presented at the AAS/AIAA Spaceflight Mechanics Meeting. Cocoa Beach, FL. [9]
- _____. 1994b. Analytical Expressions for Orbit Covariance due to Gravity Errors. Paper AAS-94-3722 presented at the AIAA/AAS Astrodynamics Conference, Scottsdale, AZ. [9]
- Yeomans, Donald K. 1991. *Comets*. New York: John Wiley & Sons. [4, 6]
- Yurasov, V., and A. Moscovsky. 1996. Geostationary Orbit Determination and Prediction. Paper presented at U.S.-Russian Space Surveillance Workshop, Poland. [9]
- Zarchan, Paul. 1994. *Tactical and Strategic Missile Guidance*. 2nd ed. Washington, DC: AIAA Press. [9]

2-line element sets, 140–142, 640
 translating, 141
 3-hourly index, 500
 4-simplex geometry, 789

A

Aberration
 annual, 90
 correction for, 90
 diurnal, 90
 light time, 90
 planetary, 90
 secular and stellar, 90

Acceleration
 centripetal, 53
 Coriolis, 53
 inertial, 104
 in three-body system, 116, 118
 tangential, 53

Acceleration, inertial
 of interceptor satellite, 346

Acronyms, 838–842

Adams, John Couch, 539, 541

Adams-Bashforth-Moulton method *See Integrators, numerical*

Air Force Satellite Control Network (AFSCN), 653

Airy, Sir George Biddell, 541

d'Alembert, Jean le Rond, 539

Alfano transfer, 337, 341
 definition of, 329

Alfano-Negron Close Approach Software (ANCAS), 805–814

Almagest, 5

Along-track variations, 42

Alphonsine tables, 97

Altitude
 in coordinate system, 39
 of an orbit, 15

Ambiguity, in signal coordination, 795

Anderle, Richard, 650

Angular momentum, specific, 106

Annular eclipse, definition of, 194

Anti-spoofing, definition of, 794

a_p , planetary amplitude, 500

Apoapsis, definition of, 14

Apollonius, 3

Apollo-Soyuz program, 274

Apsidal rotation, 593

Argument of latitude, definition of, 136

Argument of perigee
 definition of, 134
 frozen, 779

Aries, first point of, 34

Aristarchus, 3

Aristotle, concepts of motion in, 103

Ascending node *See Node, ascending*

Aspherical-potential function
 common notations for, 492–493
 deriving, 485–497
 MacCullagh's method, 488–489
 partial derivatives for, 496
 spherical harmonics method, 494
 orbital elements method, 572–574

Associated Legendre Functions, 490

Astrodynamic, definition of, 1

Astrolabe, description of, 6

Astrology, 2

Astronomical unit, definition of, 8

Asymptotic expansions
 definition of, 544
 Theory of, 549

Atmospheric density, determining, 843–854

Atmospheric models, 502–514, 843–854
 static, 503
 time-varying, 504
 types of
 CIRA, 511
 DTM, 514
 exponential, 508
 Harris-Priester, 511
 Jacchia-Roberts, 512, 843–851
 MSIS, 514
 Soviet Cosmos, 514, 852–854
 Standard atmosphere, 511
 variations in, 503–506

Atomic time, international, 69

Attitude dynamics, definition of, 1

Attitude, of a satellite, 137, 533

Australian Geodetic Datum, 28

Auxiliary circle, for eccentric anomaly, 210

Azimuth-elevation
 conversions for, 170–176

Azimuth, definition of, 39

B

B1950 system, 92
 Babylonians, 2
 Ballistic coefficient, 140, 499
 Barker, Thomas, 218, 235
 Barycenter, definition of, 32, 35
 Basic kinematic equation, 52
 Batch processing, definition of, 658
 Beamwidth, definition of, 377
 Bessel functions, 240, 606–607
 Bessel, Frederick, 157
 Besselian year, definition of, 92
 Bias, definition of, 664
 Bifurcation point, definition of, 594
 Blending techniques, 867–874

 parabolic, 867
 quartic, 871

Bode's law, 373
 Boresight angle
 definition of, 752
 solving for, 754

Borgi, Jorst, 465

Brahe, Tycho, 99, 156, 207, 208

Bretagnon, method of, 190

Brouwer, Dirk, 542

Brouwer's method, 626–628

Brown, Ernest, 185, 540

Brun's theorem, 29

Buran, 369–370

Bureau International de l'Heure (BIH), 82

C

$c(\psi)$ functions, 229–231

 solutions for, 229

Canonical transformations, in equations of motion, 569–571

Canonical units, development of, 94–96

Canonical variables, 143–144

Cavendish, Henry, 7

Cayley-Hamilton theorem, 710

Cefola, Paul, 542

Celestial

 equator, 32

 latitude and longitude, 34

 mechanics, definition of, 1

 poles, 32

reference frames, 93

 sphere, geometry of, 32

Celestial Ephemeris Pole, 82, 536

Central differencing, 687

Central limit theorem, definition of, 668

Central-body effects, 579–603

 categories of, 589

 long- and short-periodic, 595

m-daily, 596

 resonance, 597–602

 secular, 590–595

Ceres, 373, 649

Chaldeans, 2

Clairaut, Alexis Claude, 539

Classical orbital elements, 130–139

 illustration of, 130

Clohessy-Wiltshire equations *See Hill's equations*

Close-approach determination, 805–814

 definition of, 743

 statistical probability in, 814

Coarse-Acquisition code, from GPS, 793

Colinear Lagrange points, 126

Colosseum, 781

Condition of osculation, 551

Conic sections

 basic parameters, 9–17

 definition of, 9

 directrix, 11

 foci of, 9

 geometry of, 9–17

Conjugate coordinates *See Canonical variables*

Consider parameters, 671

Constants, physical, 93–96, 877

 planetary and solar, 877

Continued fractions, 441, 600

Contour grids, 29

Control law, for short-time orbital transfer, 338

Conventional International Origin, 82

Coordinate rotations, 45–51

 using vectors, 50

Coordinate systems, 31–46

 barycentric, 35

 common transformations, 53–55

 Conventional Inertial, 38

 Earth-based, 37–40

 Earth-Centered Inertial (ECI), 37

 Earth-centered, Earth-fixed (ECEF), 39

- Earth-Fixed Greenwich (EFG), 39
 equinoctial, 44
 Fundamental Katalog (FK5), 38
 geocentric equatorial (IJK), 37
 heliocentric (XYZ), 36
 inertial, 104
 interplanetary, 36
 J2000, 38
 mean of date, 77
 motion of, 74–93
 perifocal (PQW), 41
 pseudoinertial, 75
 quasi-inertial, 75
 roll-pitch-yaw (RPY), 43
 rotating transformations, 51–53
 satellite (NTW), 43
 satellite (RSW), 42
 satellite-based, 40–46
 summary of, 46
 synodic, 35
 topocentric equatorial, 39
 topocentric horizon, *SEZ*, 39
 true of date, 81
 Coordinate transformations, 45–55
 Coordinated universal time (UTC), 66
 Copernicus, Nicholas, 97–99, 157
 Corner-cube retro-reflectors, use of, 375, 639
 Correlation coefficients, 674
 Cost functions, definition of, 328, 659
 Council of Trent, and Gregorian calendar, 156
 Covariance matrix, 673
 Cowell, Phillip Herbert, 466
 Cowell's formulation, 466, 473–474
 Cowell's method, 474
 Critical inclination, 593
 Cross-sectional area, of a satellite, 498
 Cross-track variations, 43
 Cubic roots *See Roots*
 Culmination, definition of, 731
- D**
- Data batches, definition of, 658
 Data, for orbit determination, 730
 Daylight Savings Time, 66
 Day of the year
 deriving, 161
 values for, 69
 Debris, orbital, 746, 747
 Declination, 35
 geocentric, 164–167
 topocentric, 167–170
 transformation with
 azimuth and elevation, 174–176
 ecliptic latitude and longitude, 174–178
 topocentric and geocentric, 169
 Deep Space Network (DSN), 653
 Deep-space satellites
 definition of, 113
 Diferent, definition of, 4
 Deflection of light, correction for, 91
 Deflection of the vertical, 22
 Del operator (gradients), use of, 470
 Delaunay variables, 143, 626
 Delaunay, Charles E., 540
 Density, atmospheric, 610
 definition of, 498
 magnetic field models for, 499
 Deprit, Andre, 540, 594
 Differential correction *See also Estimation of orbits*, 406, 658, 679
 Dilution of precision (DOP), 799
 Direct effect, in three-body motion, 118
 Directrix, of a conic section, 11
 Displacements *See also Relative positions, rendezvous*, 42, 353–366
 along-track, 42
 cross-track, 43
 in-track, 43
 radial, 42
 Disturbing functions, 470–471
 Doppler shift, 376, 377
 second-order, relativistic correction for, 791
 DORIS, 379, 653
 Drag, aerodynamic, 498, 603–611
 coefficient of, 498, 611
 drag paradox, 608, 609
 effects on orbits, 766, 768
 equations for, 604
 spacecraft attitude effects, 533
 Draper Semianalytical Satellite Theory (DSST), 629–634
 Drift (data), definition of, 664
 DSN *See Deep Space Network*
 DSST *See Draper Semianalytical Satellite Theory*

Dynamical time, 70–74
 barycentric, 70
 calculating, 73
 terrestrial, 70

E

Earth
 eccentricity of, 18
 fundamental parameters, 17–31
 location parameters, 19
 mean equatorial radius, 18
 orbital eccentricity, 182
 radius
 definition of, 95
 estimation of, 3
 rotational velocity, 19
 semiminor axis, 18
 shape of, 20–27
 models for, 20
 zonal gravitational coefficient, 18
 Earth-Moon system, regions of motion in,
 127, 129
 Easter, date of, 155–156
 East-west stationkeeping, 748
 Eccentric anomaly, 210
 alternate formulation, 215–216
 Eccentricity
 in computer applications, 132
 of a conic section, 11
 of an orbit
 definition of, 131
 frozen, 779
 of Earth, 18
 of Earth's orbit, 182
 ECHO 1, 744
 Eckert, Wallace, 540
 Eclipses, 2, 192–194
 lunar, 193
 penumbra region, 192
 solar, 193
 umbra region, 192
 Ecliptic latitude and longitude
 transformations for, 176–180
 Ecliptic of date, definition of, 76
 Ecliptic system
 transformation to equatorial, 176
 Ecliptic, obliquity of, 33

Einstein, Albert, 790
 Electromagnetic spectrum, 375
 Element set, definition of, 129
 Elements
 2-line element sets, 140–142, 623, 640
 classical orbital, 130–139
 converting orbital, 642–644
 equinoctial, 142
 mean, 546, 626, 633, 641
 orbital
 converting with position and velocity
 vectors, 144–148
 osculating, 129, 546, 633, 641
 two-body, 129, 641
 Elevation, definition of, 39
 Ellipse, fundamental, definition of, 426
 ELLIPSO, 628
 Ellipsoidal function, for satellite orbits, 808
 Ellipsoidal models, 20, 21
 Encke, Johann Franz, 466
 Encke's formulation, 472–473
 Energy integral, 109
 Ephemerides
 definition of, 534
 planetary, 188–192
 Epicycle system, 4, 5
 Epoch, definition of, 55
 Equant, 6
 Equation of the center
 for satellite motion, 240
 Equation of the equinoxes, 59, 81
 Equation of time, 64
 Equations of motion
 n-body, 116–127
 polar form, 329
 three-body, 116–127
 barycentric form, 119
 center of mass in, 120
 equilibrium solutions, 126
 general series representation, 118
 mass ratio, 125
 phase space, 126
 relative formula, 117
 restricted, 122–128
 ten known integrals for, 120
 two-body, 103–116
 assumptions for, 105–106
 basic, 105

- derivation of, 103
 - relative form, 105
 - specific angular momentum, 106
 - specific mechanical energy, 108
 - Equator crossings, nodal drift effect on, 765
 - Equatorial plane, definition of, 33
 - Equatorial system
 - transformation to ecliptic, 177
 - Equilateral hyperbola, use of, 220
 - Equilibrium solutions, three-body problem, 126
 - Equinoctial elements, 142
 - Equinox
 - definition of, 33
 - vernal, 34
 - Equinoxes, equation of, 59, 81
 - Eratosthenes, 3–4
 - Error
 - analysis, in data estimation, 664–671
 - ellipsoid, 805, 808
 - instrument/measurement, 664–665
 - time-varying, 665
 - Error budget, for perturbation analysis, 635
 - Escape velocity, definition of, 115
 - Estimation of orbits
 - availability, quantity of data, 730
 - convergence issues, 737
 - deterministic, 657
 - EKF, nonlinear system, 724–727
 - error analysis in, 664–671
 - geometric constraints, 730
 - Kalman filtering, 707–729
 - summary, 728–729
 - KF, linear system, 717–722
 - linear least squares, 658–675
 - unweighted, 658
 - weighted, 671–675
 - LKF, nonlinear system, 722–724
 - mathematical modeling for, 665
 - Maximum likelihood, 650
 - nonlinear least squares, 675–682
 - post-processing vs. prediction, 733
 - predicting, filtering, smoothing, 656–657
 - probability techniques for, 666–671
 - propagation models, 732
 - propagator selection, 732
 - sequential-batch least-squares techniques, 702–707
 - stochastic, 657
 - Euclid, 3
 - Euler angles, 45
 - Euler integrator, 476
 - Euler, Leonhard, 102, 372, 539, 540
 - European Datum, 28
 - Excentric system, 4, 5
 - moveable, 6
 - Extreme ultra-violet (EUV) radiation, 506
- ## F
- Fading-memory filters, definition of, 737
 - Fast variables, of orbital elements, 546
 - Fictitious mean Sun
 - right ascension of, 64
 - use of, 57, 64
 - Field of view
 - calculating, 754
 - definition of, 751
 - geometry for, 751, 752, 755
 - Filters
 - fading-memory, 737
 - Kalman, 707–729
 - square-root, 738
 - tuning, 718
 - types of, 717
 - Finite differencing, 686, 711
 - Fisher, Ronald A., 650
 - Fit spans, definition of, 734–735
 - Fixed- Δv *See Orbital transfers*
 - FK5 inertial frame, 86
 - Flamsteed, John, 372
 - Flattening
 - of a conic section, 11
 - of the Earth, 18
 - Flicker noise, 665
 - Flight-path angle, 137–139
 - geometry of, 16, 107
 - Forces, disturbing *See Perturbations*
 - Free-return trajectories, 128
 - Frenet system *See Coordinate systems, satellite, NTW*
 - Frozen orbits, 745, 747, 773, 778–780
 - Functions, disturbing-potential, 471
 - Functions, potential, 470
 - Fundamental ellipse, 426, 441
 - Fundamental Katalog system, 38
 - Fundamental period, 590

G

Galilei, Galileo, 99, 101
 Galle, Johann Gottfried, 541
 Gauss, Karl Friedrich, 8, 373, 649, 789
 Gaussian coefficients, 502
 Gaussian constant of gravitation, 8
 Gaussian coordinate system *See Coordinate systems, satellite, RSW*
 Gaussian variation of parameters, 275, 559–568
 Gemini program, U.S., 274
Generalized momenta *See Canonical variables*
 General Perturbation Techniques *See Propagation of Orbits, analytical*
 definition of, 543
 Geocentric values, from topocentric values, 169
 Geodesy, 28–29
 Geoid
 definition of, 21, 28
 undulation, 28
 Geops, definition of, 28
 Geostationary satellites, definition of, 113
 Geosynchronous orbits
 semimajor axis for, 114
 Geosynchronous satellites
 definition of, 113
 orbit error propagation, 641
 Gibbs method, 408–415
 Gibbs, Josiah, 374
 Global Positioning System (GPS), 727,
 788–800
 determining position from, 798
 differential techniques for, 799
 dilution of precision, 799
 double-difference processing, 800
 signals, 793–800
 system description, 791–793
 Goddard Earth Model, 31
 Goddard Trajectory Determination System
 (GTDS), 542
 Goddard, Robert, 273
GPS *See Global Positioning System*
 Gradient
 definition of, 470
 in synodic acceleration, 124
 Grand tour of the planets, 153
 Gravitation, law of, 7
 Gravitational acceleration, 30

Gravitational coefficients, 30–31, 875, 876

normalizing, 493
 example of, 493–494
 using, 535

Gravitational constant, 7

Gravitational model, 29–31

Gravitational parameter, measurement of, 7

Gravitational potential, deriving, 486–494

Gravity, theoretical, 30

Great circle, definition of, 32

Greenwich Mean Time, definition of, 64

Greenwich meridian, definition of, 19

Greenwich Sidereal Time, formula, 60

Gregorian calendar, 5, 156

Grissom, Virgil, 274

Ground-based Electro-Optical Deep Space Surveillance (GEODSS), 653

Groundtracks

 definition of, 264
 determining period, 266
 eccentricity and perigee location effects, 267
 examples of, 264–268
 inclination effects, 268
 repeat, 763–773, 780–787
 shift parameter, 765

H

Halley, Edmond, 101, 371
 Halley's comet, 371–372, 466
 Hamilton, William Rowen, 569
 Hansen coefficients, 574, 576
 Hansen, Peter A., 540
 Harmonics, spherical, 494–496
 Harmony of the spheres, definition of, 3
 Harrison, John, 788
 Height
 ellipsoidal, 29
 orthometric, 28
 Heliocentric system, parallax in, 157
 Herrick, Samuel, 374
 Herrick-Gibbs method, 415–421
 Hill, George W., 540
 Hill's equations, 343–368
 solutions for, 343–346, 351, 365
 Hinge symmetry for groundtracks, 266
 Hipparchus, 4
 Hohmann transfer *See Orbital transfers*

Hohmann, Walter, 278
 Homotopy continuation method, 406
 Hooke, Robert, 101
 Hour angle
 Greenwich, 35
 local, 35
 Hour circles, definition of, 32
 Hubble Space Telescope, 96, 352, 365
 Hyperbolic anomaly
 definition of, 220
 formulation for, 219–224
 Hyperbolic excess velocity, 109

I

Ideal rocket equation, 330
 Ignorable coordinate, definition of, 571
 Inclination
 changing, 298–301, 304, 307–309
 critical, 593
 definition of, 133
 Indirect effect, in three-body motion, 118
 Inertial acceleration, for three-body problem, 123
 Inertial formula, for three-body motion, 116
 Inferior planets, definition of, 6
 Initial Orbit Determination *See Orbit Determination, Initial*
 Integrals of the Motion, 119
 Integration, numerical
 determining step size, 483
 implementing, 482–485
 methods of, 475–485
 order of, 477
 ordinate, summed-ordinate forms, 481
 regularization, 484
 single-step, multi-step, 475
 time-regularized Cowell, 484
 Integrators, numerical
 Adams-Basforth-Moulton, 475, 479
 Burlisch-Stoer, 475
 Euler, 476
 Runge-Kutta, 475, 476
 Runge-Kutta-Fehlberg, 477–478
 Sum-squared, Cowell-Störmer, Gauss-Jackson), 475, 481–482
 Intercept orbits, 446
 International Polar Motion Service (IPMS), 82

International Terrestrial Reference Frame, 93
 Interpolation, linear, 866
 In-track variations, 43
 Invariable plane of Laplace, 121
 Inverse Julian date, 163

J

J2000 system *See Coordinate systems, J2000*
 Jacchia-Roberts atmospheric model, 512, 843–852
 Jacobi integral equation, 124
 Jacobi, Carl, 102
 Jacobi's constant, 125
 Jacobian equation, 125
 Jet Propulsion Laboratory, 181, 184
 Joint Gravity Model, 31, 94, 875–876
 Joseph algorithm, 716
 Julian calendar, 156
 Julian century, 5
 Julian date, 67–69
 formula, 67
 inverse, 163
 modified, 67
 Julian period, 67
 Julian year, 5, 67, 163

K

Kalman filtering *See also Estimation of orbits*, 707–729
 applications, 727–728
 extended, EKF, 724–727
 linear, KF, 717–722
 linearized, LKF, 722–724
 nonlinear system use, 722–727
 summary, 728–729
 Kalman gain matrix, 715
 Kalman, Rudolf E., 650
 Kalman-Bucy method *See Kalman filtering, extended*
 Kaula gravitational coefficient, 588
 Kepler, Johann, 99
 calculating Mars orbit, 207–209
 Rudolphine tables, 371
 Kepler's equation, 211, 213
 solutions for, 209–247
 elliptical, 231–234

- hybrid, 241–243
- hyperbolic, 237–239
- Newton's method, 231
- Newton-Raphson iteration, 231–232, 241, 406
- parabolic, 235–237
- series, 239–240
- summary of formulas, 244–247
- universal formulation, 224–230
- Kepler's Laws**, 7, 100, 209
 - first, 110–112
 - second and third, 112–114
- Kepler's problem**, 209, 247–263
 - solutions for
 - first-order approximation, 622
 - orbital-element technique, 247–249
 - series forms of f and g , 254
 - universal variables technique, 255–263
 - using f and g functions, 249–255
- Keplerian elements** *See Classical orbital elements*
- Keplerian period** *See Period of a satellite, anomalistic*
- Kolmogorov, Andry N., 650
- Kozai, Yoshihide, 541
- Kozai's method, 624–626
- k_p , planetary index, 500
- Kutta, Wilhelm, 476

L

- LACE, propagation errors for, 637
- LAGEOS** *See Laser Geodynamics Satellites*
- Lagrange brackets, 552–553
- Lagrange multipliers, 331
- Lagrange planetary equations, 552, 557
 - Gaussian form, 567
 - Equinoctial form, 559
- Lagrange points, 126
- Lagrange, Joseph, 102, 373, 540
- Lambert, Johann H., 372, 422
- Lambert's problem, 275, 372, 421–445
 - Battin method, 441–445
 - equation, 427
 - Gaussian solution, 430–435
 - first and second equations of, 433
 - intercept and rendezvous, 446–461
 - minimum eccentricity, 424, 426
- minimum energy, original formulation, 423–428
 - fundamental ellipse, 426
- Thorne's solution, 429–430
 - universal-variable solution, 435–440
- Landgrave William IV, 156
- Landsat, 598, 742
- de Laplace, Pierre-Simon, 373, 465, 539
- Laplace transform, 348
- Laser clearinghouse problem, 823
- Laser Geodynamics Satellites (LAGEOS), 639, 653, 732
 - propagation errors for, 639
- Latitude**
 - astronomical, 21
 - definition of, 20
 - ecliptic, 34
 - geocentric
 - definition of, 21
 - determining, 205
 - geodetic
 - definition of, 21
 - determining, 24, 202–205
 - geometry of, 22
 - reduced, 22
 - transformations between, 26
- Launches**
 - direct, 294
 - locations for, 294, 295–298
 - velocity of, 295
 - windows for, 296–298
- Law of gravitation, 7, 130
- Le Verrier, Urbain-Jean-Joseph, 541
- Lead angle, definition of, 317
- Leap seconds, use of, 66
- Leap year, 5
- Least squares techniques, implementing, 696–702
- Legendre functions, 491
- Legendre polynomials, 487
- Legendre, Adrian Marie, 466, 649
- Leonardo of Pisa, 156
- Leonov, Alexei, 274
- Light, correcting for deflection of, 91
- Lighting conditions, 198–201
- Lilius, Aloysius, 156
- Line of nodes, definition of, 33, 133
- Line of sight

- determining, 198–201
- geometry for, 199
- with Sun, 200
- Line symmetry for groundtracks**, 266
- Linear least squares**, 658–675
 - unweighted, 658
 - weighted, 671–675
- Local horizontal**, definition of, 15
- Longitude**
 - definition of, 19
 - ecliptic, 34
 - reduced, 59
- Longitude of periaxis**, 135
- Look angles**, predicting satellite, 800–804
- Love numbers**, 578
- Luna-1 and Luna-2**, 273–274

- M
- MacCullagh's formula**, 489
- Macro models**, 533
- Main Problem**, theory of, 549, 593–594
- Maneuvers**
 - combined, 306–313
 - coplanar, 275–293
 - fixed- Δv , 310–313
 - intercept and rendezvous, 314–327, 446–461
 - relative motion in, 343–368
 - maximum efficiency, 292–293
 - minimum-inclination, 307–309
 - noncoplanar, 293–306
 - orbital, definition of, 273
 - orbital-maintenance, 768–773
- Maneuvers** *See also Orbital transfers*
- Manifold techniques**, for analyzing regions of motion, 128
- Mathematical modeling**
 - in data processing, 665
- Matrices**, fundamentals of, 856
- Matrix inversion lemma**, 714
- Maximum likelihood**, 650
- M-daily variations**, 596–597
- Mean anomaly**, definition of, 211, 212
- Mean elements**, 546, 626, 633, 641
- Mean equator of date**, 77
- Mean equinox of date**, 77
- Mean latitude, longitude at epoch**, 137

- Mean longitude**, 137, 142
- Mean motion**, of a satellite, 131
 - definition of, 212
 - distribution of, 746
- Mean Sea Level**, determining, 28
- Mean sidereal day**, definition of, 65
- Mean, statistical**, definition of, 667
- Mechanized Algebraic Operations**, MAO, 540
- Meridian**, definition of, 19
- Minkowski**, Herman, 790
- MIR**, 274, 645
- Molniya orbit**
 - definition of, 746
 - propagation errors for, 640
- Molodensky formulas**, 29
- Monte-Carlo methods**, 667
- Motion, relative** *See Relative motion of satellites*
- Multipath signals**, definition of, 797

- N
- Nadir point**, definition of, 752
- Napier**, John, 465
- Navigation**, space-based, history of, 788
- NAVSPASUR**, 653
- Navstar** *See Global Positioning System*
- n-body equations** *See Equations of motion*
- Near-identity transform**, 630
- Newcomb**, Simon, 540
- Newton**, Isaac, 6–7, 207, 273, 539
 - developing his laws, 101–102
 - method for orbit determination, 371–372
- Newton's Laws**, 102
 - law of gravitation, 7, 103
 - second, 103
- Newton-Raphson iteration**, 231–232, 241, 406
- Nodal drift**, effect of, 760–763
- Nodal regression**, 76, 591
- Node, ascending**
 - definition of, 133
 - longitude of, 133
 - changing, 301–304
- Node, descending**, 133
- Noise (data)**, definition of, 664
- Nonlinear least squares**, 675–682
- Normal equations**, 662
- North American Datum**, 28

North-south stationkeeping, 748
Numerical Integration *See Integration, numerical*
 Nutation
 coefficients, 880
 definition of, 76
 transformations for, 79–80
 Nyquist critical frequency, 483

O

Oberon, orbit of, 137
 Oberth, Herman, 273
 Oblate models, definition of, 20
 Oblate spheroid, as Earth model, 21
 Oblers, Heinrich Wilhelm, 373
 Obliquity of the ecliptic, 33, 79
 Observability, definition of, 732
 Observables
 carrier-phase, in GPS, 795
 ionospheric correction for, 797
 linearly combined, 797
 pseudorange, in GPS, 794
 Observations
 angles-only, 386–403
 double- r iteration, 397–403
 Gaussian method, 392–397
 Laplace's method, 388–392
 mixed, 403–407
 range and range-rate processing, 403
 range-only processing, 405
 range, azimuth, and elevation, 379–386
 short- and long-arc, 378
 Ohio State University model, 31
 Orbit determination
 close-approach, 743
 differential correction for, 682–702
 initial, 371–461
 data for, 374–379
 definition of, 371
 Gibbs method, 408–415
 Herrick-Gibbs method, 415–421
Lambert's problem *See also Lambert's problem*, 421–445
 range-only processing, 405–407
 range, range-rate processing, 403–405
 three position vectors and time, 408–421
 partial derivative matrix for, 685–687

sensitivity matrix for, 688
 state-transition matrix, 709–712
 statistical, 649–733
 update intervals, 732
 variational equations for, 689–695
 Orbit raising, 331–337
 accumulated velocity change for, 335
 control law for, 332
 λ dynamics in, 331
 Orbital dynamics, definition of, 1
 Orbital elements
 definition of, 129
 fast and slow variables, 546
 from position and velocity vectors, 144–148
 mean, 546, 626, 633, 641
 osculating, 546
 periodic and secular changes, 545
 single- and double-averaged, 547
 Orbital transfers
 Alfano, 337
 bi-elliptic, 279, 282, 283–287, 292
 changing inclination and longitude of ascending node, 304–306
 continuous-thrust, 327–342
 coplanar, 275–293
 comparison of, 292
 critical limits, 287
 fixed- Δv , 307
 general, 293
 Hohmann, 278–287, 292
 comparison to bi-elliptic, 284–287
 critical limits, 287
 inclination-only, 298–301
 low-thrust, noncoplanar, 337–342
 node-only, 301–304
 noncoplanar, 293–306
 nontangential, 277
 one-tangent burn, 287–293
 ratio for, 285
 tangential, 275, 277
 Orbits
 circular, 13
 closed and open, 10
 direct or prograde, 133
 elliptical, 12
 apoapsis and periapsis, 14
 equatorial, 133
 frozen, 745, 747, 779–780

frozen-eccentricity, 773
 geostationary, 113
 geosynchronous, 113, 744
 hyperbolic, 15
 inclined, 133
 low-Earth, 113, 744
 maneuvering for, 273–368
 minimum-altitude variation, 747,
 773–778, 787
 mission design for, 757–787
 mission types, 743–748
 Molniya, 746
 parabolic, 14
 phasing
 use of, 323
 polar, 133
 rectilinear, 9
 repeat-groundtrack, 745, 747, 763–773
 example of, 780–787
 initial elements for, 781–785
 retrograde, 133
 semi-synchronous, 113, 744
 Sun-synchronous, 745, 747, 757–760
 transfers between *See Orbital transfers*
 Osculating elements, 129, 546, 633
 Osiander, Andreas, 98

P

Parabolic anomaly
 definition of, 218
 flight time, 429
 formulation for, 217–219
 Parabolic mean point radius, 441
 Parallax
 annual, 90
 correction for, 90
 definition of, 157
 diurnal, 90
 example values, 157
 geocentric and heliocentric, 90
 horizontal, 90
 Moon's, 186
 Parameter estimation, 663
 Pass, satellite, definition of, 265, 802
 Patched-conic approximation, 153–154
 Payload angle, definition of, 311
 Payload Assist Module, 369

Penalty functions, definition of, 328
 Penumbral eclipse, definition of, 193
 Performance indices, definition of, 328
 Periapsis
 definition of, 14, 134
 longitude of, 135
 true longitude of, 135
 Period, of a satellite, 112
 anomalistic, 764
 equation for, 113
 nodal, 764
 Periodic effects, in orbital elements, 545
 Perturbation analysis, definition of, 466
 Perturbation effects
 central-body, 586–603
 categories of, 589
 zonal, secular, 590–595
 zonal, long-periodic, 595
 zonal, short-periodic, 595
 zonal, LP/SP/beat, 595–596
 sectorial and tesseral, m-daily, 596–597
 sectorial and tesseral, LP/m-daily, 597
 sectorial and tesseral, LP/m-daily/SP linear combination, 597
 sectorial and tesseral, resonance, 597–602
 drag, 607–611
 solar-radiation, 619–621
 summary, 646
 third-body, 614–617
 Perturbation techniques
 accuracy in, 635–639
 analytical *See also Perturbation techniques, general*, 543
 Cowell's method, 474
 Encke's method, 472–473
 general, 539–647
 Brouwer's method, 626–628
 Kozai's method, 624–626
 two-body propagation, 247–263
 PKEPLER, 622–624
 initial conditions for, 640–642
 numerical, 475–485
 complex acceleration model, 530–533
 order of, 548
 simplified acceleration model, 527–530
 Method of, 547–549
 physical data sources, 535, 886–887

- atmospheric and solar, 536, 886–887
- practical considerations, 533–536, 634–644
- semanalytical, 543, 628–634
 - Draper Semianalytical Satellite Theory, 629–634
- special *See also Perturbation techniques, numerical*, 465–536
- VOP *See Variations of Parameters*
- Perturbations, 485–525
 - atmospheric drag, 497–514
 - acceleration, 498–499
 - categories of, 589
 - central-body
 - deriving aspherical-potential, 485–493
 - partial derivatives, 496–497
 - conservative and nonconservative, 470
 - definition of, 467
 - magnetic field of the Earth, 499–502
 - solar-radiation pressure, 517–523
 - shadowing analysis, 520–523
 - summary of, 644–647
 - third-body effects, 514–517
 - acceleration, 515–516
 - partial derivatives, 516–517
 - thrust, 524
 - tidal, 525, 578
- Perturbations, linearized equations
 - central-body, 574–586
 - drag, 604–607
 - third-body, 611–614
 - solar-radiation pressure, 617–619
- Phase angle, definition of, 314, 752
- Phase lag, 578
- Phase space, in regions of motion, 126, 128
- Phasing, orbital
 - circular coplanar, 314–321
 - circular noncoplanar, 321–327
- Piazzi, Giuseppe, 373
- Pisa, leaning tower, 101
- Plane change *See Inclination, changing; Orbital transfers*
- Planetary amplitude, 500
- Planetary constants, 878, 879
- Planetary ephemerides, 536, 882–886
 - calculation of, 188–192
- Planetary index, geomagnetic, 500
- Plato, 3
- Pochammer symbol, 429
- Poincaré variables, 143
- Poincaré, Henri, 102
- Poisson brackets, 557
- Poisson series, example of, 546
- Polar motion, 58
 - transformation for, 82–84
- Polar orbits, 133
- Polynomial solutions, 863–866
- Position vector
 - converting to latitude and longitude, 202–205
 - Moon, 184–188
 - Sun, 181–184
- Potential function
 - general, 470–471
 - central-body, COE's, 572–574
 - third-body, COE's, 574–577
 - tidal, COE's, 578
- Power-density matrix, 665
- Precession
 - general, 76
 - luni-solar, 75
 - planetary, 75
 - transformation for, 77–78
- Precise conversion of elements, 643
- Precise Positioning Service, 793
- Precision code, from GPS, 793
- Precision Orbit Ephemerides, 534
- Primary hour circle, 35
- Prime meridian, definition of, 19
- Process noise
 - definition of, 665, 712
 - second moment of, 665, 713
- Prolate models, definition of, 20
- Propagation of orbits *See also perturbation techniques*
 - accuracy of, 635–639
 - analytical, 621–628
 - example, 544
 - numerical, 526–533
 - example, 469
 - weak time dependence, 530
 - semanalytical, 628–634
 - stability, 532
 - two-body, 247–263
 - validation, 534
- Propellant mass fraction, 330
- Proper motion, correcting for, 92

Pseudoinertial systems, 75

Pseudorandom noise, 793

Pseudorange

definition of, 794

effects on, 796

Ptolemaic tables, 155

Ptolemy, 4, 5, 155, 207

Pythagorean theorem, 12, 859

Pythagorous, 3

Q

Quadratic, quartic roots *See Roots*

Quasi-inertial systems, 75

R

RADARSAT, 742, 759

Radial component, of position vector, 108

Radial variations, 42

Radial velocity, correcting for, 92

Radius of apoapsis, periapsis, 15

Radius of convergence, 679

Radius of curvature in the meridian, 25

Reconnaissance from space, definition of, 749

Rectilinear orbits, 9

Recursions *See also Legendre polynomials*, 531

Redshift correction, for relativistic effects, 791

Reduction of coordinates, 38, 74–93

Reference frames

celestial, 93

inertial, 74

pseudoinertial, 75

quasi-inertial, 75

terrestrial, 93

Reflectivity, 519

Refraction, correcting for, 91

Relative formula, for three-body motion, 116

Relative motion of satellites *See also Hill's*

equations

analyzing, 343–368

coordinate system geometry for, 344

extending the results, 366–368

trend analysis for, 353–366

Relative positions

coordinate system for, 42

rendezvous, 367

Relativity, general and special, 790

Rendezvous, circular, 314–327

circular coplanar phasing, 314–321

circular noncoplanar phasing, 321–327

relative motion in, 343–368

Residual, definition of, 659

Resonance

anomalistic, 598

deep, 599

effects of, 597–602

parameter, 599

repeat groundtrack, 598

shallow, 599

Restricted three-body problem, 122–128

Retrograde factor, in equinoctial elements, 142

Retrograde motion, 6, 133

Retro-reflectors, corner cubes, 375

Revolution, satellite, definition of, 265

Rheticus, Georg, 98

Right ascension, 35

geocentric, 164–167

topocentric, 167–169

transformations with

azimuth and elevation, 174–176

ecliptic latitude and longitude, 176–178

topocentric and geocentric, 169

Right ascension of the node, 133

Rise/set problem, 815–821

definition of, 743, 815

Root mean square, RMS, definition of, 667

Roots of polynomials

cubic, 863

quadratic, 863

quartic, 865

Rotating acceleration, for three-body problem,

123

Rotational velocity of the Earth, 19

Runge, Carl, 476

Runge-Kutta methods, 476

Russian Space Surveillance System (RSSS),

653

S

Saros cycle, 2

Satellite Laser Ranging (SLR), 379, 403, 653

Satellite populations, distribution of, 743

Satellites, types of missions, 744

Scaliger, Joseph, 67

- Scheiner, Christopher, 99
 Schmidt functions and coefficients, 502
 Schur's identity, 705
 Schwarzschild geometry, 791
 Schwarzschild radius, 71
 Schwarzschild, Karl, 791
 Sectorial harmonics, 495
 Selective Availability (SA), for GPS signals, 793
Semianalytical techniques *See also Propagation of orbits, semianalytical*
 definition of, 543
Semilatus rectum *See Semiparameter*
Semimajor axis, 10, 130
 equation for, 114
Semiminor axis, 10
 of Earth, 18
Semiparameter, definition of, 14, 131
Semiperimeter, definition of, 423
Semi-synchronous satellites
 definition of, 113
Sensitivity matrix, 688
Sensor sites, geometry of, 382
Sensor systems, 651–658
Sensors
 calibration, 655, 656, 732
 data processing techniques, 656
 dedicated, collateral, contributing, 651, 654
 limitations of, 655
 optical, 386
 representative ranges, 657
 staring, scanning, 751
Sequential-batch least-squares techniques, 702–707
Sequential-batch routines, definition of, 703
Series solutions for eccentric, true, and mean anomalies, 239
Sexagesimal number system, 156
Shannon-Sampling theorem, 483
Sidereal time, 58–62
 apparent, 59
 converting to solar, 65
 definition of, 57, 58
 Greenwich, 58
 calculating, 63
 local, 58
 mean, 59
 transformation for, 81
Signals
 active, passive, retransmitted, 375
 bi-static, mono-static, 376
Similarity transformation, definition of, 669
Simple harmonic oscillator equation, 350
Single-point conversion, 643
Singular elements, definition of, 574
Singularities, definition of, 683
Site position vector, 22–27, 381
SLR *See Satellite Laser Ranging*
Smugness, in data filters, 708, 727, 729
Solar flux, 506
Solar incidence angle, 519
Solar maxima, Solar minima, 506, 508
Solar pressure, definition of, 518
Solar-radiation effects, 617–621
 equations for, 617–619
 shadow analysis, 520–523
Solar-rotation cycle, 505
Solar time, 63–69
 apparent, 63
 converting to sidereal, 65
 definition of, 57
 mean, 64
Solve-for parameters, 682
South American Datum, 28
Soviet Cosmos atmospheric model, 852–854
Space support, definition of, 653
Space surveillance
 definition of, 749
 space-based, 749–756
Space Surveillance Network (SSN), 652
Space surveillance, ground-based, 651
Special perturbation techniques *See also propagation of orbits, numerical*
 definition of, 468
Specific angular momentum
 in two-body equation of motion, 106
Specific mechanical energy
 derivation of, 108
Speed of light, 374
Sphere of influence, for a planet, 109, 153
Spherical harmonics, 494–496
 addition theorem of, 490
 zonal, 494
 sectorial, 495
 tesseral, 495–496
Splining techniques, 867–874

- cubic, 869
- quintic, 873
- SPOT**, 742
- Sputnik I, II, 273, 541, 651, 741
- Square-root filters, definition of, 738
- SSN** *See Space Surveillance Network*
- Stability, in recursion algorithms, 532
- Standard deviation
 - definition of, 668
 - probability values for, 669
- Standard Positioning Service, 793
- Star magnitudes, cataloging, 4
- State equations, definition of, 328
- State representations
 - of satellites, 129–144, 660
- State space or vector, definition of, 129, 660, 683
- State-transition matrix, 251, 709–712
 - definition of, 709
- Stationary systems, definition of, 710
- Stationkeeping maneuvers, 746–748
- Statistical Orbit Determination *See differential correction*
- Stokes, Sir George Gabriel, 466
- Sturm-Liouville theorem, 494
- Sub-latitude point, of a satellite
 - determining, 202–205
- Sum-squared method, 481
- Sundman transformation, 224
- Sundman, Karl F., 122, 224
- Sunrise, sunset, determining, 194–198
- Sun-spot cycle, 505
- Sun synchronous** *See Orbits, Sun synchronous*
- Superior planets, 6
- Super-synchronous satellites
 - definition of, 113
- Surveillance from space, definition of, 749
- Swerling, Peter, 650
- Synodic coordinate frame, in three-body problem, 123
- Synodic frame, definition of, 118
- Synodic period, definition of, 318
- Synodic system, definition of, 32
- Système International (SI) second, 56, 69

- T**
- Targeting problem, 446–461
- Taylor, Brook, 475
- Taylor series, 231, 476
- TDRS, 379
- Temperature, exospheric, 843, 844
- Ten known integrals, for equations of motion, 120
- Terminator, definition of, 801
- Terrestrial latitude, 19
- Terrestrial parameters, definition of, 19
- Tesseral harmonics, 495
- Thales, 3
- Third-body effects, 611–617
 - equations for, 611–614
 - on geosynchronous satellites, 615
 - partial derivatives for, 516–517
- Three-body equations *See Equations of motion*
- Tide-gauge measurements, use of, 28
- Tides, solid-Earth, definition of, 525
- Tides, ocean, 525
- TIMATION Program**, 788
- Time
 - coefficients of, 71
 - conversions for, 72, 158–164, 536
 - DMS to RAD/Rad to DMS, 158
 - HMS to Rad/Rad to HMS, 159
 - HMS to UT/UT to HMS, 160
 - YMD to Day of Year, 161
 - YMDHMS to Days, 162
 - equation of, 64
 - systems of, 55–74
 - apparent solar time, 63
 - atomic (TAI), 58, 69
 - barycentric dynamical (TDB), 70
 - coordinated universal (UTC), 66
 - dynamical, 58, 70–74
 - mean solar, 64
 - sidereal *See also Sidereal time*, 58
 - solar *See also Solar time*, 63
 - terrestrial dynamical (TDT), 70
 - universal (UT), 57, 64
- Time dependence, weak, definition of, 530
- Time determination, from GPS, 794
- Time of flight, calculating, 268–271
- Time unit, definition of, 95
- Time zones, UTC offsets for, 66
- TIROS 1**, 744
- Tokyo Datum, 28
- TOPEX**, 379, 597–598, 645, 653, 727

- orbit determination for, 733
- propagation errors for, 638
- Topocentric declination, 40
- Topocentric right ascension, 40
- Tracking and Data Relay Satellite *See TDRS*
- Tracking system, definition of, 652
- Tracks, uncorrelated, 738
- Trajectory equation, 111, 217
- Transfer method, 422
- Transit Satnav System, 788
- Triangular points, finding, 126
- Triaxial ellipsoid, 20
- Trigonometric relations, basics of, 857–859
- Trigonometry
 - planar, 859
 - spherical, 860–863
- Trilateration, 403
- Triton, orbit of, 136
- Trojan planets, 102, 127
- True anomaly
 - definition of, 15, 134
 - hyperbolic limitations on, 221
- True equator of date, 77
- True longitude at epoch, 137
- Tsiolkovsky, Konstantin, 273
- Tuning the filter, 718
- Turning angle, in hyperbolic orbit, 219
- Twilight, definitions of, 194, 195
- Two-body equation *See Equations of motion*

U

- Umbra and penumbra, geometry for, 192, 194
- Uncorrelated tracks, definition of, 738, 742
- Universal time, 64
- Universal variable
 - derivation for Kepler's equation, 224–225
- Update intervals, 733
- UT0, UT1, UT2, definitions of, 64–65

V

- Variance, statistical, definition of, 668
- Variation of parameters (VOP) *See also Perturbation techniques*, 549–552
 - Gaussian, 559–568
 - Hamilton's (canonical) form, 569–572
 - Lagrangian, 552–559

- equinoctial form, 559
- Variational equations, 689–695
- Vectors
 - fundamentals of, 855–856
 - position and velocity
 - from orbital elements, 149–152
 - transforming, 52
- Velocity
 - change, accumulated, 330
 - equation for, 115
 - radial, 92
- Vernal equinox, 33
- Vertical datums, 28
- v -infinity, definition of, 109
- Visibility
 - satellite to Sun, satellite to satellite, 198–201
 - to a sensor site, 800–804
- Vis-viva equation, 109
- von Braun, Wernher, 273
- von Zach, Franz Xavor, 373

W

- Wait time in orbital rendezvous, 318
- Weak time dependance, 530
- Weiner, Norbert, 650
- World Geodetic Survey, 31, 93
- Wren, Sir Christopher, 101
- WWV, 66, 536

Y

- Young, John W., 274

Z

- Zenith distance, definition of, 39
- Zonal harmonics, 494
 - effects on orbital elements, 579–586
- Zulu time *See Coordinated universal time*

UTC Offsets for Selected Time Zones. The central meridian defines the difference in *UTC* from the Greenwich meridian. Hawaii doesn't use daylight savings time. (See Table 1-4, pg 66).

Time Zone	Standard	Daylight Savings (April to October)	Central Meridian
Atlantic	<i>UTC</i> – 4 h	<i>UTC</i> – 3 h	–60°
Eastern	<i>UTC</i> – 5 h	<i>UTC</i> – 4 h	–75°
Central	<i>UTC</i> – 6 h	<i>UTC</i> – 5 h	–90°
Mountain	<i>UTC</i> – 7 h	<i>UTC</i> – 6 h	–105°
Pacific	<i>UTC</i> – 8 h	<i>UTC</i> – 7 h	–120°
Alaska	<i>UTC</i> – 9 h	<i>UTC</i> – 8 h	–135°
Hawaii	<i>UTC</i> – 10 h	<i>UTC</i> – 10 h	–150°

Day of the Year. This table lists numbers for certain days of the year. The elapsed days from the beginning of the year are useful to calculate time. Leap-year values are in parentheses. (See Table 1-5, pg 69).

Date (0 h)	Day of the Year	Elapsed Days from Jan 1, 0 h	Date (0 h)	Day of the Year	Elapsed Days from Jan 1, 0 h
January 1	1	0	July 1	182 (183)	181 (182)
January 31	31	30	July 31	212 (213)	211 (212)
February 28	59	58	August 31	243 (244)	242 (243)
March 31	90 (91)	89 (90)	September 30	273 (274)	272 (273)
April 30	120 (121)	119 (120)	October 31	304 (305)	303 (304)
May 31	151 (152)	150 (151)	November 30	334 (335)	333 (334)
June 30	181 (182)	180 (181)	December 31	365 (366)	364 (365)

Julian Date Values. This table lists some common Julian dates for selected years. It also gives the Greenwich Sidereal Time for January 1 at 0 h *UT* for each year. The years 1900, 1950, and 2000 are common for use in coordinate systems.

Year	GST (°)	JD	Jan 1	Feb 1	Mar 1	Apr 1	May 1	Jun 1	Jul 1	Aug 1	Sep 1	Oct 1	Nov 1	Dec 1
1900	100.1837764	241	5021	5052	5080	5111	5141	5172	5202	5233	5264	5294	5325	5355
1950	100.0756886	243	3283	3314	3342	3373	3403	3434	3464	3495	3526	3556	3587	3617
1960*	99.6598354	243	6935	6966	6995	7026	7056	7087	7117	7148	7179	7209	7240	7270
1970	100.2296373	244	0588	0619	0647	0678	0708	0739	0769	0800	0831	0861	0892	0922
1980*	99.8137995	244	4240	4271	4300	4331	4361	4392	4422	4453	4484	4514	4545	4575
1990	100.3836169	244	7893	7924	7952	7983	8013	8044	8074	8105	8136	8166	8197	8227
1991	100.1449049	244	8258	8289	8317	8348	8378	8409	8439	8470	8501	8531	8562	8592
1992*	99.9061929	244	8623	8654	8683	8714	8744	8775	8805	8836	8867	8897	8928	8958
1993	100.6531284	244	8989	9020	9048	9079	9109	9140	9170	9201	9232	9262	9293	9323
1994	100.4144166	244	9354	9385	9413	9444	9474	9505	9535	9566	9597	9627	9658	9688
1995	100.1757049	244	9719	9750	9778	9809	9839	9870	9900	9931	9962	9992	0023	0053
1996*	99.9369932	245	0084	0115	0144	0175	0205	0236	0266	0297	0328	0358	0389	0419
1997	100.6839290	245	0450	0481	0509	0540	0570	0601	0631	0662	0693	0723	0754	0784
1998	100.4452175	245	0815	0846	0874	0905	0935	0966	0996	1027	1058	1088	1119	1149
1999	100.2065061	245	1180	1211	1239	1270	1300	1331	1361	1392	1423	1453	1484	1514
2000*	99.96777947	245	1545	1576	1605	1636	1666	1697	1727	1758	1789	1819	1850	1880
2001	100.7147308	245	1911	1942	1970	2001	2031	2062	2092	2123	2154	2184	2215	2245
2002	100.4760196	245	2276	2307	2335	2366	2396	2427	2457	2488	2519	2549	2580	2610
2003	100.2373085	245	2641	2672	2700	2731	2761	2792	2822	2853	2884	2914	2945	2975
2004*	99.9985975	245	3006	3037	3066	3097	3127	3158	3188	3219	3250	3280	3311	3341
2005	100.7455339	245	3372	3403	3431	3462	3492	3523	3553	3584	3615	3645	3676	3706
2006	100.5068230	245	3737	3768	3796	3827	3857	3888	3918	3949	3980	4010	4041	4071
2007	100.2681122	245	4102	4133	4161	4192	4222	4253	4283	4314	4345	4375	4406	4436
2008*	100.0294015	245	4467	4498	4527	4558	4588	4619	4649	4680	4711	4741	4772	4802
2009	100.7763382	245	4833	4864	4892	4923	4953	4984	5014	5045	5076	5106	5137	5167
2010	100.5376277	245	5198	5229	5257	5288	5318	5349	5379	5410	5441	5471	5502	5532

*Leap years occur every four years, and in centuries that are evenly divisible by 400.

JGM-2 Constants & Conversions. We can use the four fundamental defining parameters (shaded cells) to recalculate this table as newer coefficients are determined. Underlined digits indicate extra precision in the conversion beyond the significant digits contained in the base values.

Mean Equatorial Radius			Shape	
(R_{\oplus}) 1.0 ER	6,378.1363	km	Mass	$5.973\,332\,8 \times 10^{24}$ kg
	20,925,644.028 871 3	ft	(b_{\oplus}) Semi-minor axis	$6,356.751\,600\,562\,9$ km
	3963.190 1 <u>56</u> 983	miles	(e_{\oplus}) Earth eccentricity	0.081 819 221 456
	3443.918 088 553	nm	e_{\oplus}^2	0.006 694 385 000
Time				
1.0 TU	13.446 849 855 11	solar min	(f) Earth flattening	1.0 / 298.257
	806.810 991 306 73	solar s	J_2	0.001 082 626 9
	0.009 338 090 177 16	solar day	J_3	-0.000 002 532 3
	0.009 363 657 021 58	sidereal days	J_4	-0.000 001 620 4
1.0 Solar Day	1.002 737 909 350 795	sidereal days	$C_{2,0}$	$-484.165\,466\,3 \times 10^{-6}$
Speed				
1.0 ER/TU	25,936.240 649 101 294	ft / solar s		
	7.905 366 149 846	km / solar s		
Gravitational Parameter			Conversions & Constants	
(μ) 1.0 ER ³ /TU ²	398,600.441 5	km ³ / (solar s) ²	1.0 ft	0.304 8 m exact
	14,076,441,746.610,709.60	ft ³ / (solar s) ²	1.0 mile	1.609 344 km exact
Rotation			1.0 mile	5280 ft exact
ω_{\oplus}	0.000 072 921 158 553 0	rad / solar sec	1.0 nm	1.852 km exact
	0.058 833 599 801 549 19	rad / TU	$\pi/2$	1.570 796 326 794 90
	0.250 684 477 337 462 15	deg / solar min	π	3.141 592 653 589 79
1.0° / solar s	1.0/0.071 015 111 031 520 1	rad / TU	2π	6.283 185 307 179 59
	6.300 388 098 665 74	rad / solar day	1.0 radian	57.295 779 513 082 30°

Fundamentals of Astrodynamics and Applications

Author: David A. Vallado

United States Air Force Phillips Laboratory

With Technical Contributions by:

Wayne B. McClain

Charles Stark Draper Laboratory

About the Book

This book integrates two-body dynamics and applications with perturbation methods and real-world applications. It cites many references and alternative methods that allow you to expand your understanding of the topics. Undergraduate courses can cover introductory sections on each topic, whereas graduate-level courses can explore the concepts in greater detail. All courses use the same standardized approach and consistent notation.

The book consolidates recent literature about coordinate systems and fundamental concepts while recognizing that the computer is how you'll implement most of the routines. Its extensive derivations give you a concrete knowledge of the basics; it thoroughly discusses maneuvers, covers initial orbit determination, perturbation techniques (propagation), and optimization algorithms. Perhaps more than any other textbook within astrodynamics, the study of perturbations has changed tremendously with the increased power of microcomputers. What used to require a room-sized mainframe computer now occurs on your laptop in an airplane. The last chapter analyzes real-world problems of specific orbit design, GPS navigation, and several useful routines for analyzing orbital problems.

From the Readers

"This is an encyclopedic collection of relevant topics, and I'm happy to see all the material collected in one place. It's particularly good to see detailed information about inertial coordinate systems, for example, since many people don't seem to realize that just specifying "ECI" isn't enough." (James F. Hall)

"The book contains a large collection of information pertaining to operational systems and qualitative insights gleaned from recent applications. The integration of these into a treatment of the traditional development of Keplerian motion, perturbed motion, and orbit determination represents a significant accomplishment." (John L. Junkins, Texas A&M)

"I found [the book] to be very comprehensive, including a lot of useful material not found elsewhere." (Anne C. Lang, Computer Sciences Corporation)

"This is a very thorough text with lots of historical comment and motivation, there is a particularly impressive collection of material on orbital maneuvering. Also, there is an excellent collection of material on modeling the perturbative forces." (Wayne E. Smith)

"The Vallado text was an added plus—a great book." (Student, Naval Postgraduate School)



McGraw-Hill

A Division of The McGraw-Hill Companies