

Fundamentals of Astrodynamics and Applications

Fourth Edition

David A. Vallado

with contributions by
Wayne D. McClain



**Space
Technology
Library**



Common Two-body Equations

Specific Angular Momentum $\hat{h} = \hat{r} \times \hat{v}$

$$h = \sqrt{\mu p} = rv\cos(\phi_{fpa}) = r^2\dot{\nu} = r_a v_a = r_p v_p$$

Specific Mechanical Energy $\xi = \frac{v^2}{2} - \frac{\mu}{r} = -\frac{\mu}{2a}$

Orbital Period $P = 2\pi \sqrt{\frac{a^3}{\mu}}$

Eccentricity Vector $\hat{e} = \frac{(v^2 - \frac{\mu}{r})\hat{r} - (\hat{r} \cdot \hat{v})\hat{v}}{\mu}$

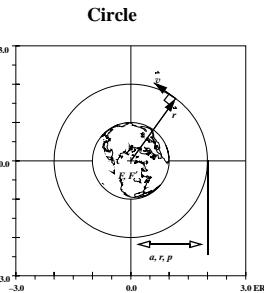
Radius $r = \frac{p}{1 + e \cos(\nu)}$

Radial Rate $\dot{r} = \frac{r\dot{\nu}e \sin(\nu)}{1 + e \cos(\nu)}$
 $\dot{\nu} = \frac{na^2}{r^2} \sqrt{1 - e^2}$

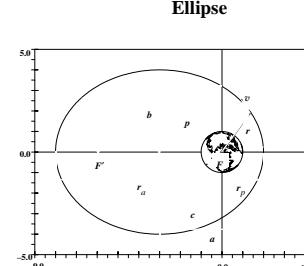
Semi-parameter $p = \frac{h^2}{\mu}$

Semi-major axis $a = \left(\mu \left(\frac{\rho}{2\pi} \right)^2 \right)^{1/3} = \sqrt[3]{\frac{\mu}{n^2}}$

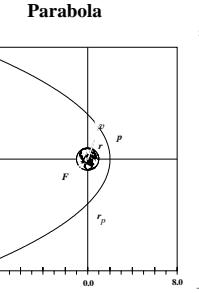
continued on next page



Eccentric Anomaly *Undefined*



$$E_{n+1} = E_n + \frac{M - E_n + e \sin(E_n)}{1 - e \cos(E_n)}$$



$$H_{n+1} = H_n + \frac{M - e \sinh(H_n) + H_n}{e \cosh(H_n) - 1}$$

$$\begin{aligned}\sin(E) &= \frac{\sin(\nu)\sqrt{1-e^2}}{1+e\cos(\nu)} \\ \cos(E) &= \frac{e+\cos(\nu)}{1+e\cos(\nu)}\end{aligned}$$

$$B = \tan\left(\frac{\nu}{2}\right)$$

$$\begin{aligned}\sinh(H) &= \frac{\sin(\nu)\sqrt{e^2-1}}{1+e\cos(\nu)} \\ \cosh(H) &= \frac{e+\cos(\nu)}{1+e\cos(\nu)}\end{aligned}$$

True Anomaly *Undefined*

$$\begin{aligned}\sin(\nu) &= \frac{\sin(E)\sqrt{1-e^2}}{1-e\cos(E)} \\ \cos(\nu) &= \frac{\cos(E)-e}{1-e\cos(E)}\end{aligned}$$

$$\sin(\nu) = \frac{pB}{r}$$

$$\begin{aligned}\sin(\nu) &= \frac{-\sinh(H)\sqrt{e^2-1}}{1-e\cosh(H)} \\ \cos(\nu) &= \frac{\cosh(H)-e}{1-e\cosh(H)}\end{aligned}$$

Flight Path Angle $\phi_{fpa} = 0^\circ$

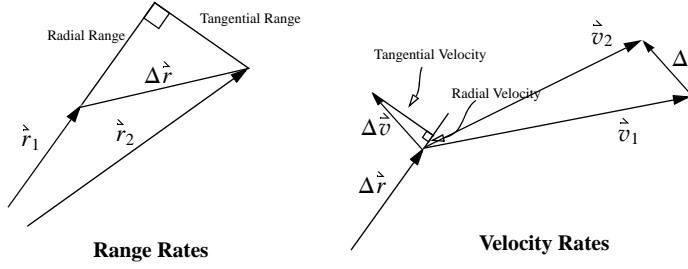
$$\begin{aligned}\sin(\phi_{fpa}) &= \frac{e \sin(E)}{\sqrt{1-e^2 \cos^2(E)}} \\ \cos(\phi_{fpa}) &= \sqrt{\frac{1-e^2}{1-e^2 \cos^2(E)}}\end{aligned}$$

$$\phi_{parabolic} = \frac{\nu}{2}$$

$$\begin{aligned}\sin(\phi_{fpa}) &= \frac{-e \sinh(H)}{\sqrt{e^2 \cosh^2(H) - 1}} \\ \cos(\phi_{fpa}) &= \sqrt{\frac{e^2 - 1}{e^2 \cosh^2(H) - 1}}\end{aligned}$$

	Circle	Ellipse	Parabola	Hyperbola
Equation	$r = a$	$r = \frac{p}{1 + e \cos(\nu)}$	$r = \frac{p}{2}(1 + B^2)$	$r = \frac{p}{1 + e \cos(\nu)}$
Perigee	$r_p = a$	$r_p = a(1 - e)$	$r_p = p/2$	$r_p = a(1 - e)$
Apogee	$r_a = a$	$r_a = a(1 + e)$	$r_a = a$	$r_a = a(1 + e)$
Semi-parameter	$p = a$	$p = a(1 - e^2)$	$p = \frac{h^2}{\mu}$	$p = a(1 - e^2)$
Semi-major axis	$a = r$	$a = -\frac{\mu}{2\xi}$	$a = \infty$	$a = -\frac{\mu}{2\xi}$
Velocity	$v = \sqrt{\frac{\mu}{r}}$	$v = \sqrt{2\left(\frac{\mu}{r} + \xi\right)} = \sqrt{\frac{2\mu}{r} - \frac{\mu}{a}} = \sqrt{\frac{\mu}{r}\left(2 - \frac{1-e^2}{1+e\cos(\nu)}\right)}$	$v_{parabola} = v_{esc} = \sqrt{\frac{2\mu}{r}}$	$v = \sqrt{\frac{2\mu}{r} - \frac{\mu}{a}}$
Eccentricity	$e = \sqrt{1 + \frac{2\xi h^2}{\mu^2}}$			
Mean Motion	$n = \frac{2\pi}{\vartheta} = \sqrt{\frac{\mu}{a^3}}$	$n = \frac{2\pi}{\vartheta} = \sqrt{\frac{\mu}{a^3}}$	$n_p = 2\sqrt{\frac{\mu}{p^3}}$	$n = \frac{2\pi}{\vartheta} = \sqrt{\frac{\mu}{a^3}}$
Mean anomaly	<i>Undefined</i>	$M = E - e\text{SIN}(E)$	$\frac{B^3}{3} + B - n_p(\Delta t) = 0$	$M = e\text{SINH}(H) - H$
Period	$\vartheta = 2\pi \sqrt{\frac{a^3}{\mu}}$	$\vartheta = 2\pi \sqrt{\frac{a^3}{\mu}}$		

Velocity and Rate Formulations



<p>Range</p> $\Delta \vec{r} = \vec{r}_2 - \vec{r}_1$ <p>Radial Range</p> $rr = \frac{\vec{r}_1 \cdot \Delta \vec{r}}{ \vec{r}_1 }$ <p>Tangential Range</p> $tr = \sqrt{(\Delta \vec{r})^2 - rr^2}$	<p>Velocity</p> $\Delta \vec{v} = \vec{v}_2 - \vec{v}_1$ <p>Radial Velocity</p> $rv = \frac{\Delta \vec{r} \times \Delta \vec{v}}{ \Delta \vec{r} }$ <p>Tangential Velocity</p> $tv = \sqrt{(\Delta \vec{v})^2 - rv^2}$ <p>Angular Rate</p> $\dot{\theta}_{vel} = \frac{\Delta \vec{r} \times \Delta \vec{v}}{\Delta \vec{r} \cdot \Delta \vec{r}}$ <p>Acceleration</p> $\ddot{\theta}_{accel} = \frac{\Delta \vec{r} \times \Delta \vec{a} - 2(\Delta \vec{r} \cdot \Delta \vec{v})(\Delta \vec{r} \times \Delta \vec{v})}{(\Delta \vec{r})^4}$ <p style="text-align: right;"><small>Angular acceleration</small></p>
<p>Acceleration</p> <p>τ must first be specified</p>	

FUNDAMENTALS OF ASTRODYNAMICS AND APPLICATIONS
Fourth Edition

THE SPACE TECHNOLOGY LIBRARY

Published jointly by Microcosm Press and Springer

- An Introduction to Mission Design for Geostationary Satellites*, J. J. Pocha
Space Mission Analysis and Design, 3rd edition, James R. Wertz and Wiley J. Larson
Space Mission Engineering, James R. Wertz
Handbook of Geostationary Orbits, E. M. Soop
Spacecraft Structures and Mechanisms, From Concept to Launch, Thomas P. Sarafin
Spaceflight Life Support and Biospherics, Peter Eckart
Reducing Space Mission Cost, James R. Wertz and Wiley J. Larson
The Logic of Microspace, Rick Fleeter
Space Marketing: A European Perspective, Walter A. R. Peeters
Orbit and Constellation Design and Management, James R. Wertz
Influence of Psychological Factors on Product Development, Eginaldo Shizuo Kamata
Essential Spaceflight Dynamics and Magnetospheric, Boris Rauschenbakh,
Michael Ovchinnikov, and Susan McKenna-Lawlor
Space Psychology and Psychiatry, Nick Kanas and Dietrich Manzey
Fundamentals of Space Medicine, Gilles Clément
Fundamentals of Space Biology, Gilles Clément and Klaus Slenzka
Microgravity Two-Phase Flow and Heat Transfer, Kamiel Gabriel
Artificial Gravity, Gilles Clément and Angelia Buckley

The Space Technology Library Editorial Board

- Managing Editor: **James R. Wertz**, *Microcosm, Inc., Hawthorne, CA*
Editorial Board: **Roland Doré**, *International Space University, Strasbourg, France*
Wiley J. Larson, *United States Air Force Academy*
Tom Logsdon, *Rockwell International (retired)*
Landis Markley, *Goddard Space Flight Center*
Robert G. Melton, *Pennsylvania State University*
Keiken Ninomiya, *Institute of Space & Astronautical Science, Japan*
Jehangir J. Pocha, *Matra Marconi Space, Stevenage, England*
Gael Squibb, *Jet Propulsion Laboratory*
Martin Sweeting, *University of Surrey, England*

About the Author

David Vallado is currently a Senior Research Astrodynamist with Analytical Graphics, Inc. in Colorado Springs, Colorado. He holds a BS degree in Astrodynamics Engineering from the U. S. Air Force Academy and MS from the Air Force Institute of Technology. He also holds an MS in Systems Management from the University of Southern California. He is a retired Air Force Lieutenant Colonel. His Air Force experience includes: Stage 1 Project Manager for MX missile, SAC Headquarters Analysis Division, U. S. Air Force instructor, Air Force Research Laboratory Scientist, and Headquarters U. S. Space Command Analysis Office. He has also worked as a principal engineer with Raytheon in Denver, Colorado.

Fundamentals of Astrodynamics and Applications

Fourth Edition

David A. Vallado
Analytical Graphics Inc.

with technical contributions by
Wayne D. McClain
Charles Stark Draper Laboratory

Space Technology Library

Published by
Microcosm Press
Hawthorne, CA

Library of Congress Cataloging-in-Publication Data

A C.I.P. Catalogue record for this book is available from
the Library of Congress

ISBN 978-1881883180 (pb) (acid-free paper)

ISBN 978-1881883197 (hb) (acid-free paper)

ISBN 978-1881883203 (ebook) (acid-free paper)

Published by

Microcosm Press

4940 West 147th Street, Hawthorne, CA 90250-6708 USA

Sold and distributed

by Microcosm, Inc.

4940 West 147th Street, Hawthorne, CA 90250-6708 USA.

Printed on acid-free paper

All Rights Reserved

© 2013 David A. Vallado

First Printing, 2013

No part of the material protected by this copyright notice may be reproduced or
utilized in any form or by any means, electronic or mechanical,
including photocopying, recording, or by any information storage and
retrieval system, without written permission from the publishers.

Printed in the United States of America

To God from whom all blessings and wisdom flow,
For the free gift of salvation through His true and only Son, Jesus Christ,
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 children,
Simone, Kathleen, and Samuel

Table of Contents

Chapter 1	Equations of Motion	1
1.1	History	1
1.1.1	Ancient Era	1
1.1.2	The Copernican Revolution	7
1.1.3	Kepler's Laws	9
1.1.4	Newton's Laws	10
1.1.5	Other Early Astrodynamical Contributions	12
1.2	Geometry of Conic Sections	12
1.2.1	Basic Parameters	12
1.3	Two-body Equation	20
1.3.1	Assumptions for the Two-Body Equation	23
1.3.2	Specific Angular Momentum	23
1.3.3	Specific Mechanical Energy	25
1.3.4	Kepler's First Law (Trajectory Equation)	27
1.3.5	Kepler's Second and Third Laws	29
1.3.6	Velocity Formulas	32
1.4	Three-body and n -body Equations	33
1.4.1	Inertial, Relative, and Barycentric Formulas	33
1.4.2	Ten Known Integrals	37
Chapter 2	Kepler's Equation and Kepler's Problem	41
2.1	Historical Background	41
2.2	Kepler's Equation	43
2.2.1	Alternate Formulation for Eccentric Anomaly	49
2.2.2	Formulation for the Parabolic Anomaly	50
2.2.3	Formulation for the Hyperbolic Anomaly	52
2.2.4	Universal Formulation	58
2.2.5	Solutions of Kepler's Equation	63
2.2.6	Summary and Related Formulas	77
2.3	Kepler's Problem	79
2.3.1	Solution Techniques	80
2.4	Satellite State Representations	95
2.4.1	Classical Orbital Elements (Keplerian)	95
2.4.2	Two-line Element Sets	105
2.4.3	Other Element Sets	108
2.4.4	Canonical Elements	110
2.5	Application: Orbital Elements from r and v	112
2.6	Application: r and v from Orbital Elements	116
2.7	Application: Groundtracks	120
2.8	Application: Find Time of Flight (FINDTOF)	123
Chapter 3	Coordinate and Time Systems	129
3.1	Historical Background	129
3.2	The Earth	131
3.2.1	Location Parameters	133
3.2.2	Shape of the Earth	134
3.2.3	Gravitational Model	141

3.3	Coordinate Systems	145
3.3.1	Interplanetary Systems	150
3.3.2	Earth-based Systems	151
3.3.3	Satellite-based Systems	154
3.4	Coordinate Transformations	159
3.4.1	Coordinate Rotation	160
3.4.2	Rotating Transformations	166
3.4.3	Common Transformations	167
3.4.4	Application: Converting IJK (ECEF) To Latitude and Longitude	169
3.5	Time	174
3.5.1	Solar Time and Universal Time	177
3.5.2	Sidereal Time	184
3.5.3	Atomic Time	190
3.5.4	Dynamical Time	190
3.5.5	Coordinate Time	193
3.5.6	Conversions	194
3.6	Time Conversions	197
3.6.1	DMS to Rad / Rad to DMS	197
3.6.2	HMS to Rad / Rad to HMS	198
3.6.3	HMS to Time of Day/ Time of Day to HMS	199
3.6.4	YMD to Day of Year/ Day of Year to YMD	200
3.6.5	YMDHMS to Days / Days to YMDHMS	201
3.6.6	Julian Date to Gregorian Date	202
3.7	Transforming Celestial and Terrestrial Coordinates	203
3.7.1	IAU-2010 Conventions	211
3.7.2	Velocity and Acceleration Transformations	222
3.7.3	IAU-76/FK5 Reduction	223
3.7.4	Other Reductions	231
3.7.5	Computational Considerations	235
3.8	Earth Models and Constants	237
3.8.1	Canonical Units	237
Chapter 4	Observations	241
4.1	Introduction	241
4.2	Obtaining Data	242
4.2.1	Quality of Data	246
4.2.2	Quantity of Data	247
4.2.3	Types of Data	248
4.2.4	Example Applications	248
4.3	Introduction to Sensor Systems	249
4.4	Observation Transformations	254
4.4.1	Geocentric Right Ascension and Declination	256
4.4.2	Topocentric Right Ascension and Declination	259
4.4.3	Azimuth-Elevation	261
4.4.4	Practical Az-El Conversions	266
4.4.5	Transformations for Ecliptic Latitude and Longitude	268
4.4.6	Practical Aspects of Obtaining Observations	270
Chapter 5	Celestial Phenomena	277
5.1	Solar Phenomena	277
5.1.1	Application: Sun Position Vector	277

5.2	5.1.2 Application: Sunrise, Sunset, and Twilight Times	281
	5.2 Lunar Phenomena	285
	5.2.1 Application: Moon Position Vector	286
	5.2.2 Application: Moon Rise and Set Times	290
	5.2.3 Phases of the Moon	294
5.3	5.3 Celestial Applications	294
	5.3.1 Application: Planetary Ephemerides	295
	5.3.2 Eclipses	299
	5.3.3 Application: Sight and Light	305
	5.3.4 Ground Illumination	309
	5.3.5 Miscellaneous Phenomena	311
Chapter 6	Orbital Maneuvering	317
6.1	Historical Background	317
6.2	Introduction	318
6.3	Coplanar Maneuvers	320
	6.3.1 Hohmann and Bi-elliptic Transfers	322
	6.3.2 Comparing Hohmann and Bi-elliptic Transfers	328
	6.3.3 Transfers Using the One-Tangent Burn	331
	6.3.4 General Transfers	337
6.4	Noncoplanar Transfers	337
	6.4.1 Introduction	337
	6.4.2 Inclination-Only Changes	342
	6.4.3 Changes in the Right Ascension of the Ascending Node	345
	6.4.4 Changes to Inclination and the Ascending Node	348
6.5	Combined Maneuvers	350
	6.5.1 Minimum-Inclination Maneuvers	351
	6.5.2 Fixed- Δv Maneuvering	354
6.6	Circular Rendezvous	357
	6.6.1 Circular Coplanar Phasing	358
	6.6.2 Circular Noncoplanar Phasing	365
6.7	Continuous-Thrust Transfers	371
	6.7.1 Introduction	372
	6.7.2 Orbit Raising	374
	6.7.3 Low-Thrust, Noncoplanar Transfers	381
6.8	Relative Motion	388
	6.8.1 Position Solutions for Nearly Circular Orbits	393
	6.8.2 Trend Analysis	398
	6.8.3 Transformation to Inertial Coordinates	411
Chapter 7	Initial Orbit Determination	423
7.1	Historical Background	423
7.2	Observations of Range, Azimuth, and Elevation	426
	7.2.1 Application: SITE-TRACK	426
7.3	Angles-only Observations	433
	7.3.1 Laplace's Method	435
	7.3.2 Gauss's Technique	439
	7.3.3 Double- r iteration	443
	7.3.4 Gooding's Method	449
7.4	Mixed Observations	450
	7.4.1 Range and Range-Rate Processing	450

7.4.2	Range-only Processing	452
7.5	Three Position Vectors and Time	454
7.5.1	Gibbs Method	454
7.5.2	Herrick-Gibbs	461
7.6	Two Position Vectors and Time—Lambert’s Problem	467
7.6.1	Lambert—Minimum Energy	470
7.6.2	Lambert—Gauss’s Solution	475
7.6.3	Lambert—Thorne’s Solution	480
7.6.4	Lambert—Universal Variables	489
7.6.5	Lambert Solution—Battin Method	493
7.7	Application: Targeting Problem	498
Chapter 8	Special Perturbation Techniques	517
8.1	Historical Background	517
8.2	Introduction to Perturbations	518
8.3	Encke’s Formulation	523
8.4	Cowell’s Formulation	525
8.5	Numerical Integration Methods	526
8.5.1	Implementing an Integrator and Determining Step Size	535
8.6	Disturbing Forces	538
8.6.1	Gravity Field of a Central Body	538
8.6.2	Atmospheric Drag	551
8.6.3	Third-Body Perturbations	574
8.6.4	Solar-Radiation Pressure	578
8.6.5	Other Perturbations	584
8.7	Forming Numerical Solutions	591
8.7.1	Application: Simplified Acceleration Model	593
8.7.2	Application: Complex Acceleration Model	596
8.8	Practical Considerations	599
8.8.1	Verifying and Validating the Propagator	599
8.8.2	Physical Data and Sources	601
Chapter 9	General Perturbation Techniques	609
9.1	Historical Background	609
9.2	Introduction	613
9.2.1	The Method of Perturbations	617
9.3	Variation of Parameters	619
9.3.1	Lagrangian VOP (Conservative Effects)	621
9.3.2	Gaussian VOP (Nonconservative and Conservative Effects)	628
9.4	Hamilton’s Formulation	637
9.5	Disturbing-Potential Formulations	641
9.5.1	Gravity Potential in Terms of the Satellite’s Orbital Elements	641
9.5.2	Third-Body Potential in Terms of the Satellite’s Orbital Elements	643
9.5.3	Tidal-Motion Potential in Terms of the Satellite’s Orbital Elements	645
9.6	Linearized Perturbations and Effects	646
9.6.1	Central-Body Analysis	647
9.6.2	Drag Analysis	671
9.6.3	Third-Body Analysis	679
9.6.4	Solar-Radiation Analysis	685
9.7	Forming Analytical Solutions	689

9.7.1 Application: Perturbed Two-body Propagation	690
9.7.2 Kozai's Method	692
9.7.3 Brouwer's Method	694
9.7.4 Operational Applications	696
9.8 Semianalytical Solutions	698
9.8.1 The Draper Semianalytical Satellite Theory (DSST)	699
9.9 Practical Considerations	704
9.9.1 Initial Data	706
9.9.2 Perturbing Force Effects on Satellites	711
9.9.3 Perturbing Force Variability	715
9.9.4 Aligning Numerical Propagation Programs	721
9.10 Summary of Perturbation Effects	724
Chapter 10 Orbit Determination and Estimation	731
10.1 Historical Background	731
10.2 Linear Least Squares	734
10.2.1 Error Analysis	739
10.2.2 Linear Weighted Least Squares	750
10.3 Nonlinear Least Squares	752
10.4 Application: Orbit Determination With Differential Correction	759
10.4.1 Calculating the Partial-Derivative Matrix (A)	762
10.4.2 Implementing Least-Squares Techniques	765
10.5 Sequential-Batch Least Squares	772
10.6 Kalman Filtering	777
10.6.1 Kalman Filter (Linear System)	785
10.6.2 Linearized Kalman Filter (LKF)	790
10.6.3 Extended Kalman Filter (EKF)	793
10.6.4 Unscented Kalman Filter (UKF)	797
10.6.5 Summary of Kalman Filters	798
10.7 Calculating the Matrices	800
10.8 Implementing Differential-Correction Solutions	813
10.8.1 Convergence Issues	816
10.8.2 Residuals	821
10.8.3 Post-processing and Prediction	821
10.8.4 Propagating the Error Covariance	825
10.8.5 Covariance Matrix Transformations	826
10.8.6 Sensor Calibration	828
10.9 Practical Considerations	828
10.9.1 Data	828
10.10 Applications	831
10.10.1 Detecting Maneuvers	832
10.10.2 Predicting Reentry	833
Chapter 11 Mission Analysis	837
11.1 Introduction	837
11.1.1 Operational Processing	839
11.1.2 Satellite Populations	845
11.2 Mission Orbits	846
11.2.1 Stationkeeping	850

11.2.2	Mission Planning	852
11.3	Geometries for Surveillance and Reconnaissance	853
11.3.1	Ground Distance on the Ellipsoid	860
11.4	Designing and Maintaining Mission Orbits	862
11.4.1	Sun-Synchronous Orbits	862
11.4.2	Repeat-Groundtrack Orbits	869
11.4.3	Minimum Altitude Variation Orbits	880
11.4.4	Frozen-Orbit Eccentricity and Argument of Perigee	885
11.4.5	Application: Designing a Specialized Orbit	888
11.5	Navigation—the Global Positioning System	895
11.5.1	Historical Background	896
11.5.2	System Introduction	899
11.5.3	Signals from GPS Satellites	901
11.6	Predicting Satellite Look Angles	908
11.6.1	PREDICT Formulation	909
11.6.2	Rise / Set	913
11.7	Performing Close Approach Analyses	919
11.7.1	Close Approach Solutions	922
Chapter 12	Interplanetary Mission Analysis	941
12.1	Historical Background	941
12.2	Patched Conic Trajectories	944
12.2.1	Sphere of Influence	945
12.2.2	Interplanetary Transfers	948
12.2.3	Departure from Earth	949
12.2.4	Arrival	952
12.2.5	Realistic Interplanetary Transfers	954
12.3	Porkchop Plots	954
12.4	Gravity Assist Trajectories	957
12.5	B-Plane Targeting	961
12.6	Planetary Capture	964
12.6.1	Aerobraking	965
12.7	Three-body Mission Design	967
12.7.1	Circular Restricted Three-body Problem (CRTBP)	967
12.7.2	Lagrange Points	971
12.7.3	Periodic Orbits	973
12.7.4	Manifolds	977
12.8	Low-energy Transfers	982
Appendix A	Dictionary of Symbols	989
Appendix B	Modeling the Atmosphere	1001
B.1	Jacchia-Roberts Atmosphere	1001
B.1.1	Evaluating Temperature	1001
B.1.2	Robert's Corrections to Temperature	1003
B.1.3	Evaluating Density	1004
B.1.4	Robert's Corrections to Density	1005
B.2	Russian GOST Atmosphere	1011

Appendix C Mathematical Fundamentals	1017
C.1 Introduction	1017
C.2 Vector Fundamentals	1018
C.3 Matrix Fundamentals	1019
C.4 Trigonometric Fundamentals	1020
C.4.1 Planar Trigonometry	1023
C.4.2 Spherical Trigonometry	1024
C.5 Numerical Techniques	1027
C.5.1 Polynomial Solutions	1027
C.5.2 Interpolation	1030
C.5.3 Blending and Splining Techniques	1031
Appendix D Constants and Expansions	1039
D.1 Gravitational Coefficients	1039
D.2 Planetary Constants	1040
D.3 Reduction Coefficients	1040
D.4 Planetary Ephemerides	1046
D.5 Data Sources	1050
D.6 Computer Programming	1056

Algorithm Summary

Algorithm 1	Find $c_2 c_3$	71
Algorithm 2	KepEqtnE	65
Algorithm 3	KepEqtnP	69
Algorithm 4	KepEqtnH	71
Algorithm 5	ν to Anomaly	85
Algorithm 6	Anomaly to ν	85
Algorithm 7	KEPLERCOE	81
Algorithm 8	KEPLER	93
Algorithm 9	RV2COE	113
Algorithm 10	COE2RV	118
Algorithm 11	FINDTOF	126
Algorithm 12	ECEF TO LATLON	172
Algorithm 13	ECEF TO LATLON	173
Algorithm 14	JULIAN DATE	183
Algorithm 15	LSTIME	188
Algorithm 16	CONVTIME	195
Algorithm 17	DMSTORAD	197
Algorithm 18	RADTODMS	197
Algorithm 19	HMSTORAD	198
Algorithm 20	RADTOHMS	198
Algorithm 21	TIMETOHMS	199
Algorithm 22	JDTOGREGORIANDATE	202
Algorithm 23	IAU-2000, CIO-Based	219
Algorithm 24	IAU-76/FK5 REDUCTION	228
Algorithm 25	Geocentric Radec	259
Algorithm 26	Topocentric	260
Algorithm 27	RAZEL	265
Algorithm 28	AZELToRADEC	268
Algorithm 29	SUN	279
Algorithm 30	SUNRISESET	283
Algorithm 31	MOON	288
Algorithm 32	MOONRISESET	290
Algorithm 33	PLANETRV	296
Algorithm 34	SHADOW	301
Algorithm 35	SIGHT	308
Algorithm 36	Hohmann Transfer	325
Algorithm 37	Bi-elliptic Transfer	326
Algorithm 38	One-Tangent Burn	333
Algorithm 39	Inclination Only	344
Algorithm 40	Change in the Ascending Node—Circular	347
Algorithm 41	Combined Changes to i and Ω —Circular	334
Algorithm 42	Minimum Combined Plane Change	353

Algorithm 43	Fixed- Δv Maneuvers	356
Algorithm 44	Circular Coplanar Phasing (Same Orbit)	362
Algorithm 45	Circular Coplanar Phasing (Different Orbit)	363
Algorithm 46	Noncoplanar Phasing	368
Algorithm 47	Low Thrust Transfer	387
Algorithm 48	Hill's Equations	396
Algorithm 49	HillEQCM to ECI	414
Algorithm 50	ECI to HillEQCM	415
Algorithm 51	SITE-TRACK	430
Algorithm 52	ANGLES-ONLY GAUSS	442
Algorithm 53	ANGLES-DOUBLE- r	444
Algorithm 54	GIBBS	460
Algorithm 55	HERRICK-GIBBS	466
Algorithm 56	Lambert's Problem—Minimum Energy	475
Algorithm 57	Lambert—Gauss's Solution	478
Algorithm 58	Lambert—Universal Variables	492
Algorithm 59	Lambert—Battin Method	494
Algorithm 60	HIT EARTH	503
Algorithm 61	TARGET	503
Algorithm 62	ENCKE	525
Algorithm 63	$AP2K_p$ CONVERSION ($K_p \Rightarrow a_p$)	558
Algorithm 64	Numerical Integration	591
Algorithm 65	PKEPLER	691
Algorithm 66	NOMINAL STATE	766
Algorithm 67	Differential Correction	768
Algorithm 68	Kalman Filter—Linear System	785
Algorithm 69	Linearized Kalman Filter	791
Algorithm 70	Extended Kalman Filter	793
Algorithm 71	Repeat Ground Track	873
Algorithm 72	Maintain Repeat Groundtrack	879
Algorithm 73	Minimum Altitude Variation	885
Algorithm 74	PREDICT	911
Algorithm 75	RISE/SET	916
Algorithm 76	CLOSE APPROACH	935
Algorithm 77	Patched Conic	953
Algorithm 78	Algorithm B-Plane I ()	962
Algorithm 79	Algorithm B-Plane II ()	963

Forward

At Microcosm, when an astrodynamics problem arises it has become standard practice to “look it up in Vallado.” In most cases, the answer is there and in a form that can be readily used and applied to the spaceflight problem at hand. I believe it is for this reason that *Fundamentals of Astrodynamics and Applications* has so rapidly become the standard astrodynamics reference for those of us in the business of spaceflight. This fourth edition will be available in softcover, hardback and eBook forms, and can be purchased at www.astrobooks.com.

Thus, it is truly a pleasure to welcome the fourth edition of “Vallado” to the Space Technology Library. While prior editions were useful and popular throughout the community, there are a number of reasons why we believe the fourth edition will be even more so. There are many reworked examples and derivations. The new edition has grown to over 1100 pages, including an entirely new chapter on interplanetary motion as well as an improved and expanded discussion of the latest coordinate systems. As with the third edition, all software routines described in the book are available for free in FORTRAN, PASCAL, C++, and MATLAB on the Microcosm Press website, www.microcosmpress.com/vallado, and the Celestrak website, <http://www.celestrak.com/>. With extensive references to both specific equations and the section of the book where the topic is discussed, this software is the perfect practical companion to Vallado’s book.

We hope that you will find this book to be as useful and productive as those of us at Microcosm have. To me, a good measure of its real utility is the fact that nearly every engineer in our Space Systems Division has purchased one for themselves with their own hard-earned money. The book is not only an excellent textbook; it is a superb reference that will be used over and over in practical astrodynamics.

Fundamentals of Astrodynamics and Applications has been a part of the Space Technology Library for over a decade now. I’m sure it will continue to be the standard reference in the field and to serve as an excellent introduction and fundamental resource for anyone working in astrodynamics.

To you, our readers, we are certain that you will enjoy the clarity and simplicity with which Dave both derives and explains modern, applied astrodynamics. To Dave Vallado, we thank you once again for all your hard work and for giving us a current and comprehensive text and reference.

James R. Wertz,
Hawthorne, CA
January, 2013
jwertz@smad.com

Preface

My goal in writing this book is to augment and consolidate 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, partly 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 scientists 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 standardization and documentation, modularity, efficiency, and maintainability. This book attempts to use the new programming capabilities to exploit each of these areas. Standardization and documentation 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 Pierre de Fermat (1601-1665) 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. 1978, Oct. 1993, Oct. 1994, and Nov. 1997, 68-73). Fermat's last theorem remained a mystery for 350 years and though proven by Andrew Wiles in 1994, the lack of Fermat's written derivation generated controversy for several hundred years! 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.

You may need some prerequisites for this book. It tries to bridge topics in astrodynamics from undergraduate courses with 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. Computer source code is available via the Internet. It details test cases and code in several languages.

This book is intended to introduce astrodynamics to the undergraduate, and to lay a solid foundation for graduate-level studies. As such, it covers a wide scope of topics, some of which are primarily for academic purposes, while others are for general interest. Topics in astronomy are introduced where relevant to astrodynamics and the study of Earth-orbiting satellites, as well as for general information. Having a degree or studying astrodynamics, you’ll invariably get questions about “what star is that?” This text probably will not prepare you for that task. However, you can consult any number of introductory astronomy books to provide an answer to that question!

The first 7 chapters discuss primarily two-body motion. In particular, the first five chapters were re-ordered from the first edition to better match undergraduate curricula. The text now begins with the two-body equations of motion, and Kepler’s solutions. This permits initial analysis of orbits, and sets the stage for the coordinate and time system discussions in Chap. 3. The original Chap. 4 was broken into two chapters and combined with information from initial orbit determination and estimation. This provides a more consolidated discussion of observations and celestial phenomena.

The last 5 chapters discuss more advanced topics, and concentrate on propagation (numerical and analytical), differential correction, and an introduction to interplanetary flight. These topics form the heart of orbit determination systems, whether they handle 20,000 objects in Earth orbit, or a single satellite. My placement of the material in this order isn’t intended to restrict first or second courses from using these chapters. In fact,

I'd expect a first course to incorporate topics such as predicting a satellite's location over a ground site (Chap. 11), or an introductory numerical propagation exercise (Chap. 8), or an initial orbit determination method (Gibbs or Herrick-Gibbs Chap. 7). I tried to keep all the similar material in the same chapter to avoid looking back several chapters to see the various approaches to accomplish a task.

There are many detailed derivations, culminating in an algorithm, suitable for programming a computer. I provide comments about computer programming because of the importance computers have become today in all disciplines. For the seasoned veteran, the algorithm may be trivial, and the computer coding comments may be unwarranted. For the student, these extra items may be the means to understanding, if not simply getting the program to work! The derivation ensures that the assumptions, methodology, and rationale are clear for each technique.

A very brief review of vectors, matrices and some numerical techniques are included in Appendix C. These do not fully explain the concepts, but rather to introduce my notation and nomenclature.

For the casual reader, instructor, etc. who wishes further information in certain areas, I've listed many references in the reference section. I include the number of the chapter in which they are cited, and the references indicate page numbers to help you find the citation quickly. I recommend a few texts that are especially good in certain areas. This list is *not* exhaustive!

Advanced Astrodynamics:	Battin (1997), Szebehely (1967), Stiefle and Scheifele (1971)
Analytical propagation:	Brouwer and Clemence (1961)
Attitude determination:	Wertz (1978)
Estimation:	Gelb (1989), Bierman (1977), Tapley et al. (2004)
General Astrodynamics:	Escobal (1985), Danby (1992), Montenbruck (2005)
Gravity, geodesy:	Lambeck (1988), Kaula (1966)
Mathematics:	Kreysig (1983), Numerical Recipes (1992), Sadler (1974)
Time, coordinates, astronomy:	Seidelmann (1992), Green (1988)

For academic courses, several pre-requisites are necessary. First, an understanding of analytical geometry, introductory calculus, differential equations, and basic computer programming are very useful. I include comments that apply to any computer programming language, and basic skills will be very useful. Data structures can also be helpful, depending on the additional computer programs that are assigned during a class. The estimation sections will benefit from introductory knowledge of statistics and probability theory. Although I explain several concepts, a single text can't explain everything!

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!

dvalldado@agi.com or davallado@gmail.com

wmcclain@adelphia.net

Questions regarding the publication of the book should contact James Wertz at

bookproject@smad.com

Acknowledgments

I credit 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 recent history: the effort to land a man on the Moon. I'll never forget that feeling I had, glued to the TV (VCRs and DVDs 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.

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 a 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, the method of ratios (also with David Negron), and conjunction probability. These works add a lot of practical insight and applications to my presentations.
- Leon Blitzer (1916-2004) 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 many years ago.
- Carole Boelitz permitted me to include sections from her thesis in the perturbations sections. She also provided valuable reviews and derivations.
- Chapter 10 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. 11. Mike Gabor helped refine this chapter into its final form.
- Jeff Parker for his contributions in writing the interplanetary chapter. Much of the chapter was adapted from his dissertation.
- 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.

- Vasiliy Yurasov provided details on the Russian Space Surveillance System, and their atmospheric models. He also submitted an update to the GOST model, which is contained in Appendix B. This represents a unique collection of data which was formerly difficult to obtain.

I received tremendous support from many people throughout the astrodynamics community. For the first edition, these folks did extensive work to mold, shape, and correct the book.

- Cheryl Arney, Gerry Baumgartner, Eric Beck, Carole Boelitz, Kitt Carlton-Wippenn, Paul Cefola, Eric Claunch, Steve Crumpton, Don Danielson (and many of his students), Gim Der, Jack Ferguson, Chuck Fosha (and his students), Michael Gabor, Jim Hall, Richard Hujšak, Karl Jensen, John Junkins, Ron Lisowski, Anne Long, Kim Luu, Ron Madler, Warner Miller, David Negron, Jr., Steve Nerem, Beny Neta, Dave Pohlen, Richard Rast, David Richardson, Chris Sabol, Ken Seidelmann, Wayne Smith, David Spencer, Sid Sridharan, David Suzuki, Jim Thorne, Nate Titus, Mark Tollefson, Scott Wallace, Gary Wirsig, Lora Worth, and Gary Yale.

Any book written mostly by an engineer needs fine tuning to become a complete work. For the first edition, 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 Air Force Research Laboratory for the first edition, including Richard Davis, Eugene Dionne, Michael Heil, Henry Pugh, and Edward Duff. Their support for, and perseverance with, this long project was invaluable.

Several people formatted the first edition in FrameMaker Ver. 4.02. 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!

The first edition also owes many thanks to Margaret Hollander and Wiley Larson for getting the book published and on the streets.

In the time since the first edition, I've enjoyed numerous insightful comments and corrections from people throughout the astrodynamics community. I list them here, alphabetically, as their attention to detail makes this book far better than I could have ever imagined!

- Salvatore Alfano, Terry Alfriend, Gary Allen, Trygve Andreassen, Nick Avlonitis, Jeff Back, David Baker, Lee Barker, Richard Battin, Robert Bell, Matthew Berry, Chris Bessette, Beth Blakney, Daryl Boden, George Born, Matthew Bourassa, Ben Bradley, Sam Brett Gilbert Phil Bridges, Kent Bryan, Bjorn Buckwalter, Irene Budianto, Richard Burns, Bryce Carpenter, David Carter, Scott Carter, Doug Cather, Paul Cefola, Ching-pyang Chang, Phil Charlesworth, Yang Chen, Vince Chioma, Chris Sabol Shannon Coffey, Camilla Columbo, Darrel Conway, Dave Cowdin, Michael Croon (and his detailed look throughout the text), Steve Crumpton (and his many proof-reading efforts!), Michael Cupples, Scott Dahlke, Don Danielson (and his colleagues), David Spencer, Lee Dewald, Marc Di Prinzio, David Diaz, Don-

ald Dichmann, John Douglass, Ted Driver, Michael Earl, Nicholas Filipski, David Finkleman, Jack Fisher, Robert Flack, Jim Fletcher, Chuck Fosha, Julian Francisco Perez Jimenez, Mark Frank, David Garza, Robert Gay, Shuting Goh, Robert Good, Jesse Gossner, Ignacio Grande, James Griggs, Tony Hagar Liam Healy (and many of his students!), Chris Heldon, Michael Hendrson, Darren Hiebert, Seth Hieronymus, Keric Hill, Keric Hill, Lane Hobson, Roy Hogstedt, Tom Homsley, Felix Hoots, Steven Hughes, Lynn Huh, Najmus Ibrahim, Haisam Ido, Egeman Imre, Randy Jackson, Tom Jerardi, Paul Jermyn, TJ Johnson, Brandon Jones, Eric Jones (and the Apollo Lunar Surface Journal), Ben Joseph, Dave Kaslow, Philippe Kehl, Clark Keith, TS Kelso, Greg Kiedrowski, Doug Kirkpatrick, Quang Lam, Casey Lambert, John Lansberry, Hungming Li, Jer-Chyi Liou, Alan Lovell, Kim Luu, Raymond Macon, Meysam Mahooti, Benjamin Mains, John Manser, Pat Marshall, Giuseppe Matarazzo, Mark Matney, Craig McLaughlin, Jeroen Melman, William Michaux, David Mikolajczak, Ted Molczan, Kurt Motekew, Andrey Nazarenko, Chris Nebelecky, Steven Nerem, Ricky O'Brien, Dan Oltrogge, Michael Owen, David Palmer, Jeff Parker, Bahvesh Patel, Donald Phillion, Brady Porter, Mehdi Raoofian, Richard Rast, Karthik Ravendar, Herb Reynolds, Robbie Robertson, Mike Rodehorst, David Rysdam, Jizhang Sang, Wayne Scheel, Conrad Schiff, Robert Scott, John Seago, Steve Seivers, Dale Shell, Sean Sherrard, Kirk Sorenson, Kevin Stedman, Kevin Stefanik, Matthew Steine, Ted Stodgell, Steve Stogsdill, Frank Stoner, Susan Strain, Shannon Sturtevant, Charles Suprin, Jim Thorne, Wang Ting, Wang Ting, Mark Tollefson, Jim Torley, Gary Turner, Bogdan Udrea, Rebecca Vallado, Jerry Vetter, Darren Wade, Sebastien Wailliez, Andrew Ward, Russ Wenzel, Martin Werner, John Westerhoff, Carlos Westholle, Paul Wilmott, Jason Yingst, Hokhan Yuksel, Vasiliy Yurasov, George Zabriskie, and Michael Zedd.

I'd like to mention John Seago and Mike Gabor in particular as they helped refine several technical sections with many detailed comments and suggestions. Their insight, in-depth knowledge, and attention to detail were invaluable in making the later editions a reality.

This fourth edition used FrameMaker Ver 10.0. While some difficulties exist with its ability to handle extended metafiles, I was able to work around most of them. Finally, and certainly not least, Jim Wertz helped get this edition published. His counsel has been important throughout this effort, and I am always very grateful for his assistance!

David A. Vallado

Colorado Springs, Colorado

February 2013

web: www.centerforspace.com
www.agi.com
email: dvallado@agi.com
davallado@gmail.com

CHAPTER 1 EQUATIONS OF MOTION

- 1.1 History
- 1.2 Geometry of Conic Sections
- 1.3 Two-body Equation
- 1.4 Three-body and n -body Equations

1.1 History

Let's begin by defining *astrodynamics*. Most dictionaries don't help—we 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 in space. 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.*

This definition combines features of celestial mechanics with orbital and attitude dynamics. Although discussions of celestial concepts are sometimes necessary, this book will focus on the behavior of artificial objects in space. From a historical perspective, true astrodynamics have only existed since the 1950's, 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 (those before about 200 A.D.) accomplished many technical feats, despite 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 what we call 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 Barclay (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 many events in nature are random (such as earthquakes and storms), astrologers sought explanations in events that were consistent and predictable, such as the movements of the Sun, Moon, and stars. In the absence of any other knowledge, astrology took on almost a religious importance. The significance of determining heavenly motions meant 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 *accurate* observations difficult. Early people studied solar eclipses in great detail because these events terrified them. 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 of the previous event. 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!

Observing the cycles of the Moon was slightly easier than for the Sun, and it permitted more precise observations. The early Babylonians, who had reasonably good weather conditions for observing the Moon's phases, developed the lunar month (about $29\frac{1}{2}$ days) (Pannekoek, 1989:21).

The 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, some of their predictions were based on incorrect phenomena or intervals from erroneously copied values and rough averages. So their mathematical understanding is suspect.

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 taught about 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!

Pythagoras (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. 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 Pythagoras. 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 didn't gain immediate acceptance because it couldn't predict the position of Mars and didn't accommodate the expected angular separation from different viewing locations. It was almost 1800 years later when Copernicus used some of Aristarchus's results 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). By assuming the light rays were parallel when they reached the Earth, he measured the Sun's rays at an angle of 7.2°, or about 1/50th of a circle (Defense Mapping Agency, 1983:2). From this meager data and simple geometry, he was able to calculate 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 our

modern technological 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 scholars 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 categorized about 1000 stars by brightness. He did this by separating the stars into six categories of brightness, each of which was separated by about 2.5 times the brightness. The brightest was first magnitude, whereas the dimmest, about 100 times dimmer, was sixth magnitude. It's remarkable 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 accurate observations, which presented some problems when trying to describe the orbital motion. As mentioned earlier, ancient 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 adequately. 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 called 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. Figure 1-1 shows the two theories.

Julius Caesar introduced the ***Julian calendar*** in 46 B.C. as a sophisticated way to track time. That particular year had 445 days due to the large error which had accumulated (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 non-integral motion about the Sun. (the actual period is 365.242 189 7). A ***Julian century*** has exactly 36,525 days. Unfortunately, the Julian year doesn't account for the remainder (about 11 minutes and 14 seconds per year, or about one day in 128 years). Thus, the system was fairly accurate over a few centuries but accumulated appreciable errors over longer periods. The Council of Trent (1545–1563) ultimately proposed changing the calendar. In 1582, Pope

* The astronomer Sosigenes actually recommended the change to Caesar after the Roman calendar (10 months with 304 days and 12 months with 355 days) had produced a three month shift in the seasons. 46 B.C. is sometimes called the year of confusion.

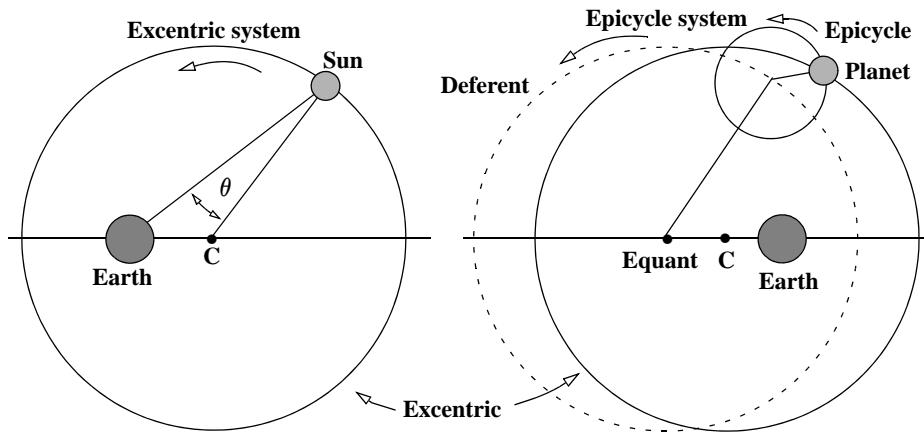


Figure 1-1. Excentric and Epicycle Systems. Both the excentric and epicycle theories adhered to an Earth-centered philosophy and used combinations of circular motion to approximate observations. The excentric system consisted of the Sun traveling around a larger circle, the excentric, with the Earth 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).

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 excess in the length of the year was partially corrected by establishing a *Gregorian calendar* in which century years have a leap day *only* if they are evenly divisible by 400 (1900 has no leap day, 2000 does, 2100 doesn't). These years are called *leap years*. The additional requirement for century years is from the insight of a librarian, Aloysius Lilius (Pannekoek, 1989:220). The length of a year for this approach is still about 27^s too long. Although we use the Gregorian calendar today, be careful in conducting historical research because some countries didn't adopt this calendar right away (Great Britain, 1752; Turkey, 1927).

Lastly, Claudius Ptolemaeus (usually called Ptolemy) (100–170 A.D.) published a 13-volume work, called *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. He separated the problem for the *inferior planets* (Mercury, Venus, and the Sun), and the *superior planets* (Mars, Jupiter, and Saturn). The difficulty that *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* (c. 150), Ptolemy described the construction of an *astrolabe*, a fundamental tool of the time, and the basis of some tools still used today. The instrument permitted angular measurements of the motions of the planets and stars.

The Ptolemaic tables in the *Almagest* were the last major astronomical endeavor for many centuries. Indeed, the fall of the Roman Empire in 476 eliminated financial support for many scientific efforts. Because of the prevalence of Catholicism after the fall of the Roman empire, the correct dates of religious festivals were required by Church officials well in advance. In particular, the correct date of Easter was of 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 difficult for most scholars and continued to perplex scientists until almost 1800 when Karl Frederick Gauss (1777-1855) found a solution. Pannekoek (1989:219) records the finding:

Dividing the year by 19, by 4 and by 7, the remainders are called a , b , and c ; put $19a + 15 =$ multiple of $30 + d$ ($d < 30$); put $2b + 4c =$ multiple of $7 + e$ ($e < 7$). Then the date of Easter is March 22nd + $d + e$.

In the 7th century, the rise of Islam demanded more accessible star tables so worshippers could determine the direction of Mecca; thus, the *Almagest* was eventually translated into Arabic.*

* In fact, the 9th-century title (*Almagest*) was created from the Arabic superlative *mageste* (meaning largest or biggest) with the added *al* prefix.

1.1.2 The Copernican Revolution

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, change was imminent.

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 1-2 shows the usual depiction.

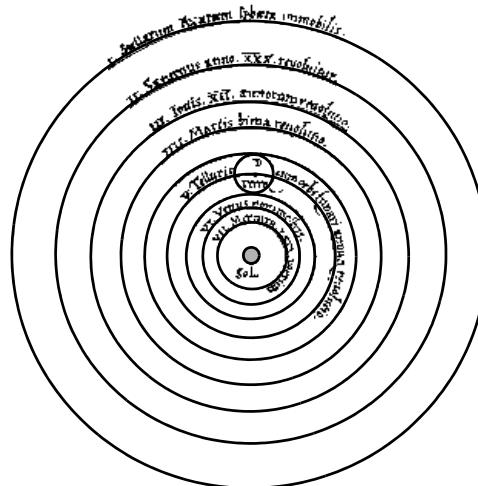


Figure 1-2. Copernicus's Solar System. (Pannekoek, 1989:191) In addition to defining a Sun-centered system, Copernicus also introduced smaller epicycles than the previous Ptolemaic system.

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.

If a motion of the Earth is assumed, this must appear, although in the opposite direction, in all that is outside, as if things filed 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 Ptolemy's theories. 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. History 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.

1.1.3 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*:

“I give thee thanks Creator and God that thou hast given me joy in Thy creation and I rejoice in the works of Thy hands. 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 problem in Chap. 2. These laws mark the beginning of a remarkable period in the history of mathematics and lead directly to the accomplishments of Isaac Newton.

1.1.4 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.

1.1.5 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!

1.2 Geometry of Conic Sections

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. A **conic section** is the intersection of a plane and a right circular cone.

Figure 1-3 illustrates 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 special cases of elliptical, parabolic, and hyperbolic motion.* Later chapters treat both standard and special orbits.

1.2.1 Basic Parameters

A few geometrical concepts, particularly conic sections, underlie many derivations throughout this book. Every conic section has two **foci**. From a geometrical standpoint, the foci provide the locations from which to construct a conic section, as illustrated in

* Rectilinear orbits 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.

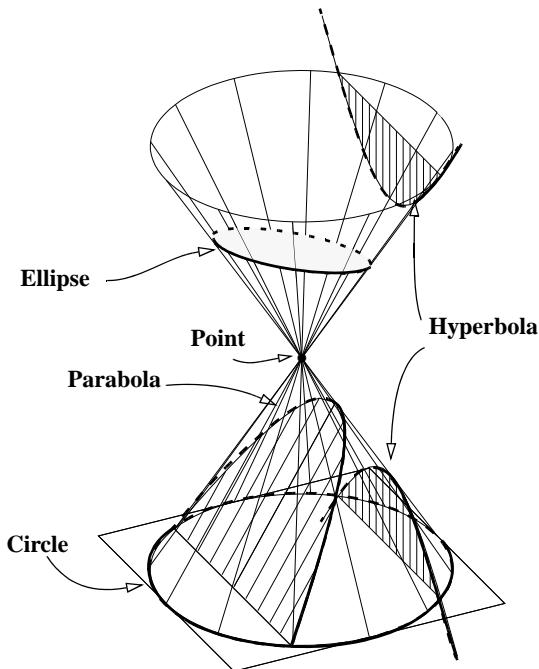


Figure 1-3. Conic Sections. Slicing a cone with a plane forms a conic section. When the plane is perpendicular to the axis of revolution, a circle results. Planes that are parallel to the axis of revolution yield hyperbolas (a pair as shown), and planes parallel to the outer surface yield parabolas. 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).

Fig. 1-4. In astrodynamics, the gravitational center of attraction coincides with one focus for all orbital motion, called the *primary focus*.

Figure 1-5 shows the two distinct foci for an ellipse, and Fig. 1-6 shows the special case of a circle in which both foci coincide. The parabola has one focus at infinity (Fig. 1-7). The hyperbola has a branch associated with each focus, as shown in Fig. 1-8. 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.

We scale the conic section by the major axis. Other parameters that describe the size of 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, infinite for a parabola, and negative for a hyperbola.

Some basic relations help us determine conic sections. The sum or difference of the distance from each focus to a point on an elliptical or hyperbolic orbit is a constant. A

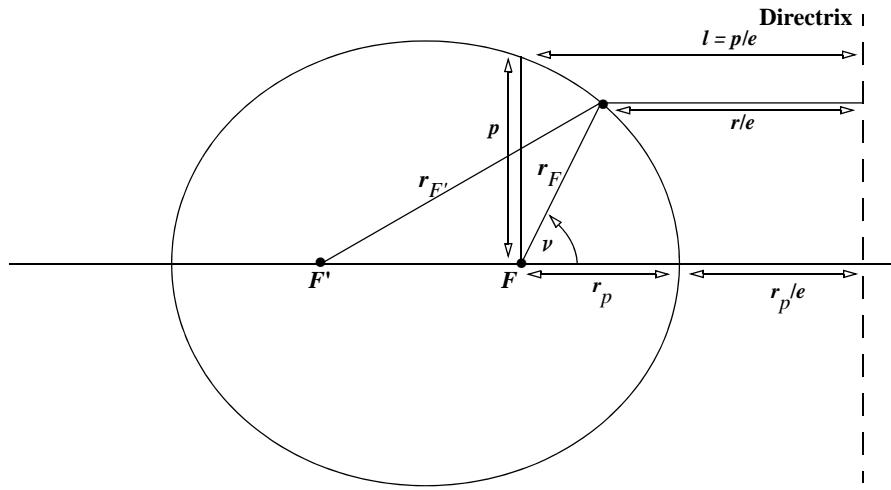


Figure 1-4. Geometry for Conic Sections. We create conic sections by recognizing that the sum of the distance from both foci to *any* point on the orbit is constant. The ratio of the distance from a focus to the orbit and the distance from that point to the directrix is also a constant called the eccentricity, 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.

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 [See Fig. 1-4]. As the pencil stretches the string and moves, the length from each focus changes, and the pencil traces out the ellipse. Prussing and Conway (1993:62) show this mathematically for any point on the conic section,

$$\begin{aligned} r_{F'} + r_F &= \text{constant} = 2a \\ |r_F - r_{F'}| &= \text{constant} = 2c \end{aligned} \quad (1-1)$$

Each focus of a conic section has a corresponding stationary line called a *directrix*. The ratio of the distance from a focus to a point on the orbit and the distance from that point to the focus' corresponding directrix is a constant defined as 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 *never negative*, and its value determines the type of conic section for common orbits. It's unity for parabolic and rectilinear orbits, less than unity for ellipses, zero for circles, and greater than unity for hyperbolas. The eccentricity is

$$e = \frac{c}{a} \quad (1-2)$$

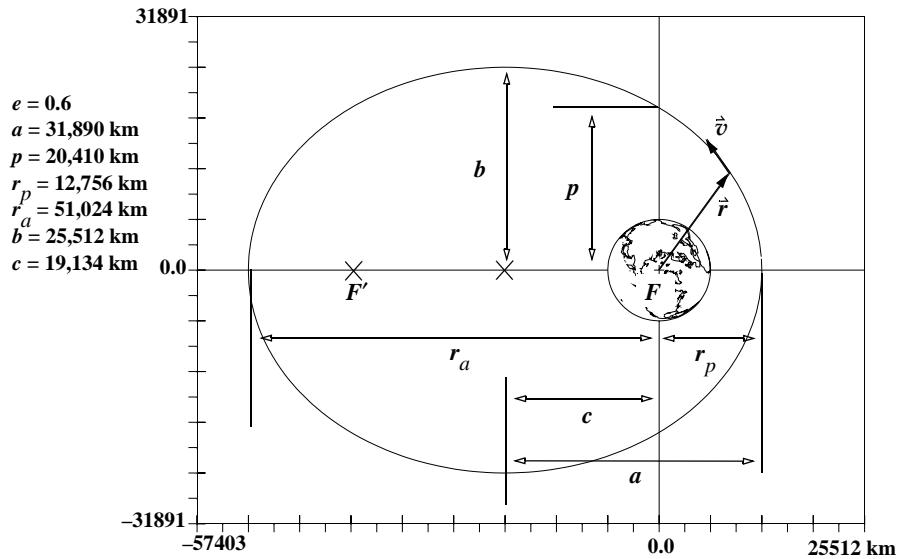


Figure 1-5. Elliptical Orbits. An elliptical orbit has two distinct foci with the primary focus, F , at the center of the Earth (or central body). The radius of apoapsis, r_a , and perapsis, r_p , denote the extreme points of the ellipse. The semimajor axis, a , and the semiminor axis, b , describe the shape of the orbit. Half the distance between the foci is c , and the semiparameter, p , locates the orbit distance normal from the semimajor axis at the focus.

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

$$f = \frac{a - b}{a} \quad (1-3)$$

Let's find an expression to convert between the flattening and eccentricity notations. Using Eq. (1-1) and Fig. 1-5, the distance from either focus to the end of the semiminor axis is given by a . Solve Eq. (1-2) 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

$$b = a\sqrt{1 - e^2} \quad (1-4)$$

This expression is useful because it relates any vertical part of an ellipse, y_e , to the corresponding part of a circle, y_c , using the semimajor axis, a . 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.

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

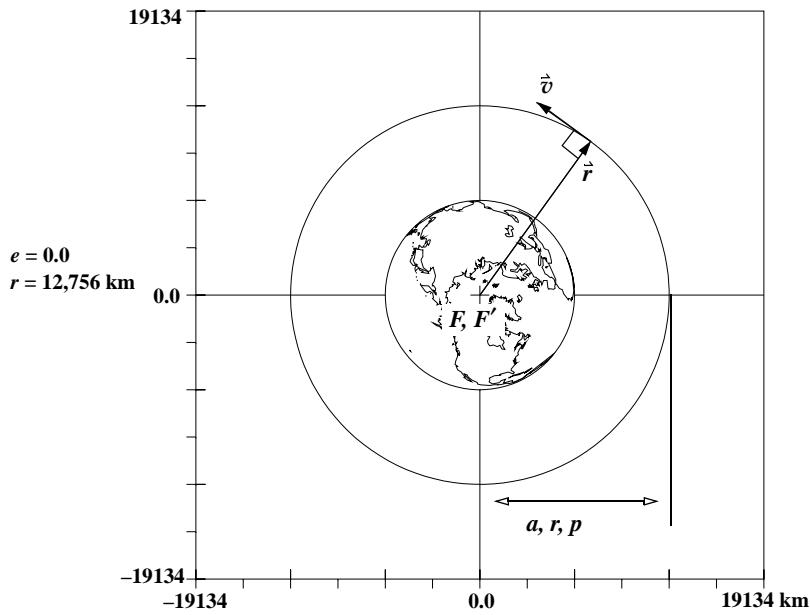


Figure 1-6. Circular Orbits. For circular orbits, the semimajor axis, semiparameter, and radius are always equal and constant. This is a special case of the ellipse, in which both foci coincide.

If we solve Eq. (1-4) for the square of the eccentricity,

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

we get an alternate equation for the eccentricity:

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

We take only the positive value for eccentricity because eccentricity is *defined* as a positive value. Continuing to develop an expression for flattening, use Eq. (1-3) 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. (1-6),

$$e^2 = 2f - f^2 \quad (1-7)$$

Another expression, derived from Eq. (1-7), can also relate the two notations:

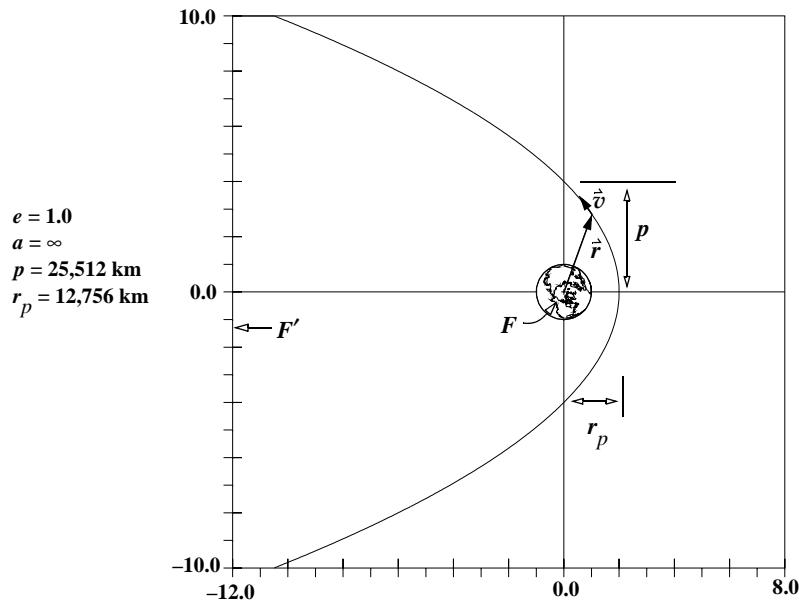


Figure 1-7. Parabolic Orbits. Parabolic orbits are considered open because they don't repeat. The semimajor axis is infinite. The second focus is at infinity.

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

The *semiparameter* is 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, where 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:

$$p = \frac{b^2}{a} \quad (1-9)$$

Using the scaling relation in Eq. (1-4), and simplifying gives us

$$p = a(1 - e^2) \quad (1-10)$$

The extreme points of an elliptical orbit are *apoapsis* and *periapsis*, representing the farthest and nearest points in the orbit, respectively, from the center of attraction. We change the endings of these words to indicate a particular planet or central body attracting the satellite—for the Sun, they are *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 orbit to the primary focus as the *radius*

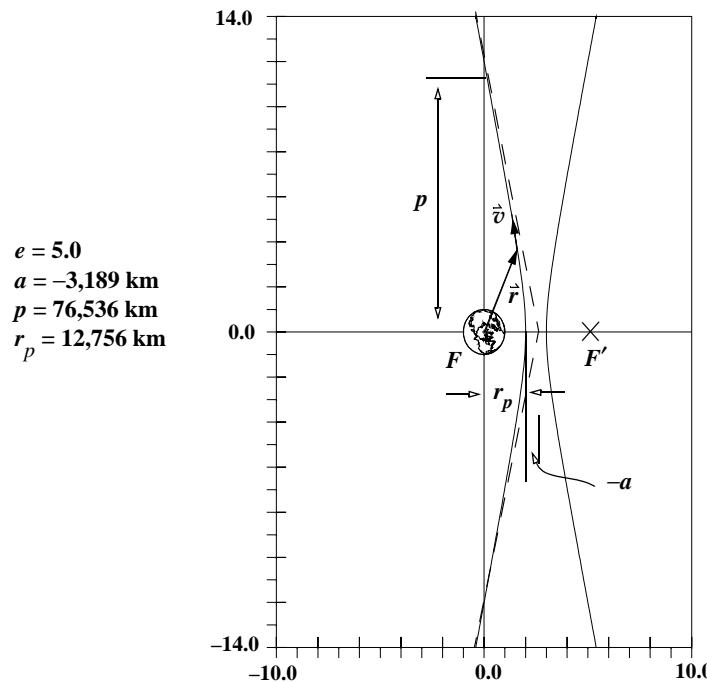


Figure 1-8. Hyperbolic Orbits. For a hyperbolic orbit, the semimajor axis is negative, and the eccentricity is larger than unity. Notice the rapid departure from the Earth with a *modest* eccentricity. The orbit approaches the dashed-line asymptotes. Hyperbolas occur in pairs called branches. The second branch is a mirror image of the first branch about the conjugate axis, shown as the dashed vertical line.

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. 1-9.

Throughout the orbit, the radius simply represents 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 , locates 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. 1-9, 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 direction in the orbit to replace periapsis as the location for the initial measurement. Computer-software routines must account for this special case.

The *flight-path angle*, ϕ_{fpa} , is the angle from the *local horizontal* (the line perpendicular to the radius vector) to the velocity vector. Figure 1-10 shows the geometry for ϕ_{fpa} . You may see other definitions.

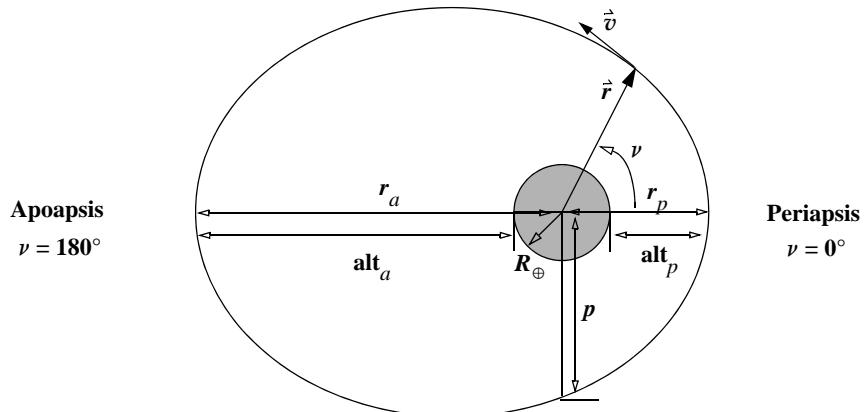


Figure 1-9. Radius of Apoapsis and Periapsis, as well as True Anomaly. Apoapsis occurs farthest from the planet; periapsis is closest. Notice the difference between altitudes and radius values. True anomaly, ν , locates the satellite's current position. We measure it in the direction of satellite motion from periapsis.

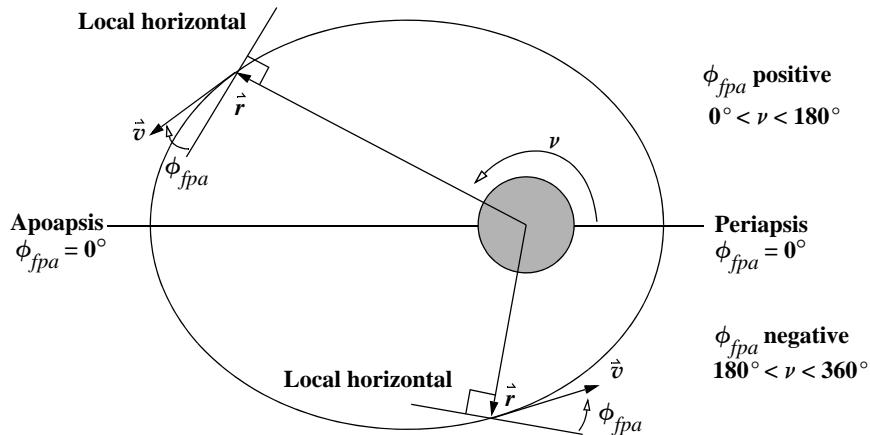


Figure 1-10. Geometry for the Flight-path Angle. The flight-path angle is always measured from the local horizontal to the velocity vector. It's always positive while the satellite travels from periapsis to apoapsis and negative for travel from apoapsis to periapsis. I've exaggerated the diagram for clarity.

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 $\phi_{fpa\text{-circle}} = 0^\circ$ at all times. In the ellipse, ϕ_{fpa} may be positive or negative. In Fig. 1-10, notice ϕ_{fpa} is positive from periapsis to apoapsis and negative on the return because we measure the angle in the *opposite* direction as the radius vector while the satellite travels

from 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. 1.3.

TABLE 1-1. Values for Conic Sections. This table shows characteristic values for the basic parameters. Sec. 2.2 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 Two-body Equation

This section 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,

$$\sum \vec{F} = \frac{d(m\vec{v})}{dt} = m\vec{a} \quad (1-11)$$

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. (3-1)] 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 1-11 shows the geometry for a fixed-mass system, in which we consider only the gravity from a spherically symmetric central body (mass M). The coordinate system is at the center of mass of the entire system.

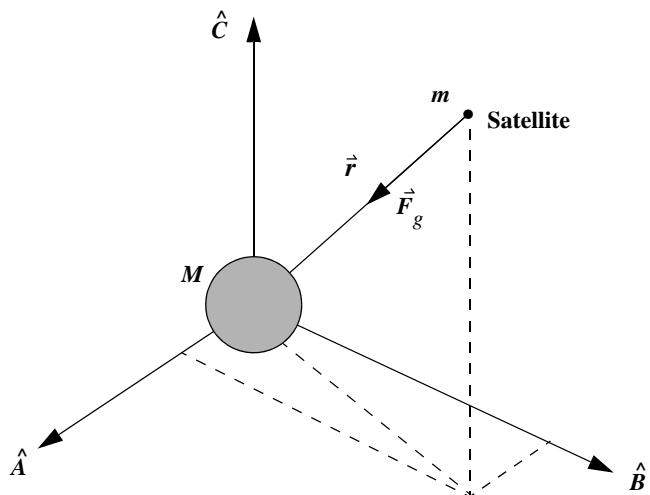


Figure 1-11. Gravitational Forces Acting on a Satellite. In a perfectly inertial system, the ideal coordinate system ABC doesn't rotate or accelerate and is often at the center of mass. Gravity acts on each object as though it were a spherically symmetrical point mass.

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 1-12 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

$$\vec{F}_g = -\frac{Gm_{\oplus}m_{sat}}{r^2}\left(\frac{\vec{r}}{r}\right) \quad (1-12)$$

To determine the equation of motion, we now use Fig. 1-12 and Eq. (1-12). 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

$$\dot{\vec{r}}_{\oplus \text{sat}} = \dot{\vec{r}}_{\text{sat}} - \dot{\vec{r}}_{\oplus}$$

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:

$$\dot{\vec{F}}_{g_{\text{sat}}} = m_{\text{sat}} \ddot{\vec{r}}_{\text{sat}} = -\frac{Gm_{\oplus}m_{\text{sat}}}{r^2} \left(\frac{\vec{r}}{r} \right) \quad \dot{\vec{F}}_{g_{\oplus}} = m_{\oplus} \ddot{\vec{r}}_{\oplus} = \frac{Gm_{\oplus}m_{\text{sat}}}{r^2} \left(\frac{\vec{r}}{r} \right)$$

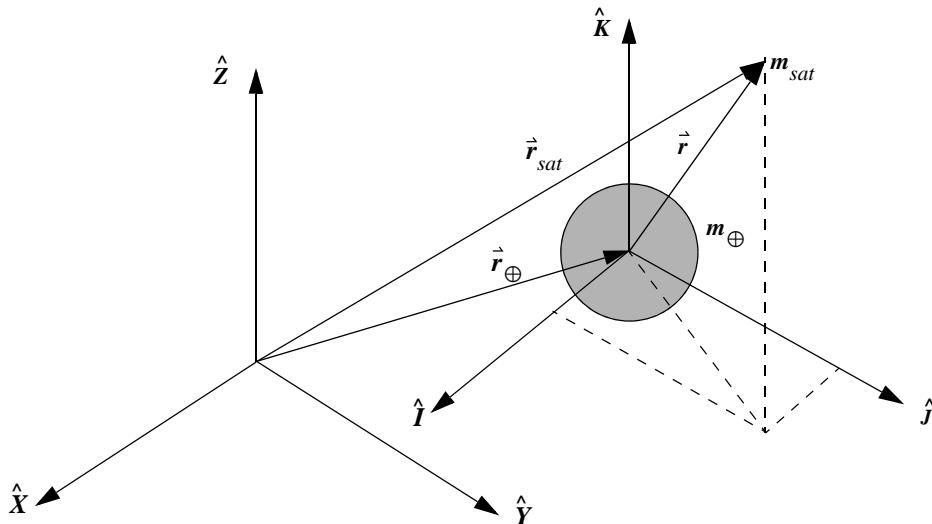


Figure 1-12. Geometry for Two Bodies in an Inertial Reference Frame. XYZ is assumed to be an inertial coordinate system. IJK is displaced from XYZ, but does not rotate or accelerate with respect to XYZ.

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 $\dot{\vec{r}}$ (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

$$\ddot{\vec{r}} = -\frac{G(m_{\oplus} + m_{sat})}{r^2} \frac{\vec{r}}{r} \quad (1-13)$$

It's convenient now to use the gravitational constant μ derived in Sec. 3.8. 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 .

$$\ddot{\vec{r}} = -\frac{\mu}{r^2} \frac{\vec{r}}{r} \quad (1-14)$$

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.

1.3.1 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 the defining statements. 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.

1.3.2 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*, h , begin by cross-multiplying the two-body equation with the position vector, \vec{r} .

$$\dot{\vec{r}} \times \ddot{\vec{r}} + \ddot{\vec{r}} \times \frac{\mu}{r^3} \vec{r} = 0$$

Because $\dot{\vec{r}} \times \dot{\vec{r}} = 0$, the second term vanishes. Now, consider the following differential:

$$\frac{d}{dt}(\dot{\vec{r}} \times \dot{\vec{r}}) = \dot{\vec{r}} \times \ddot{\vec{r}} + \ddot{\vec{r}} \times \dot{\vec{r}} = \dot{\vec{r}} \times \ddot{\vec{r}}$$

We can substitute this differential into the original equation, but because this results in the derivative being equal to zero, the integral quantity must be a constant, \hat{h} . If we rearrange and substitute the position derivative with the velocity, the original expression becomes

$$\hat{h} = \vec{r} \times \vec{v} = \text{constant} \quad (1-15)$$

Notice the lack of mass from the usual connotation of momentum, hence the *specific* qualifier. Because \hat{h} is the vector cross product between \vec{r} and \vec{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. Thus, *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 1-13 illustrates this situation. The standard cross product allows us to investigate the angular momentum using the auxiliary angle, ϕ_{fpav} , 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:

$$h = rv \cos(\phi_{fpa}) \quad (1-16)$$

At apoapsis and periapsis, the flight-path angle is zero, and the respective angular momenta are given by

$$\begin{aligned} h &= r_a v_a \\ h &= r_p v_p \end{aligned} \quad (1-17)$$

We find another useful expression for calculating angular rate as follows. Using the cosine expression of the flight-path angle in Fig. 1-13,

$$\cos(\phi_{fpa}) = \frac{r \dot{v}}{v}$$

Substituting the result of Eq. (1-16) solved for $\cos(\phi_{fpa})$ into this equation gives us

$$h = r^2 \dot{v} \quad (1-18)$$

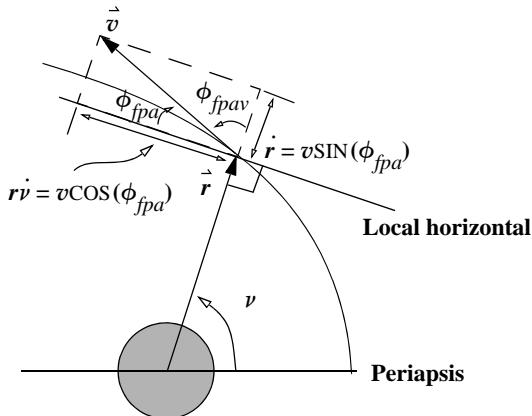


Figure 1-13. Geometry of the Flight-Path Angle. The flight-path angle, ϕ_{fpa} , is measured from the local horizontal to the velocity vector, which we can break into radial and transverse components. The vertical angle ϕ_{fpav} is sometimes used in derivations.

We can also get an expression relating the semiparameter and the angular momentum by combining the trajectory equation [Eq. (1-24)] and appropriate formulas from analytical geometry. The standard quantity in the numerator of a polar form of a conic (or trajectory equation) is given as le , where the eccentricity is known and l represents the distance from the directrix to the primary focus (see Fig. 1-4). 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. (1-6)]. When we replace it by the quantity ae ,

$$le = a - ae^2$$

By Eq. (1-10), $le = p$. If the numerator of the polar-trajectory equation is the semiparameter,

$$p = \frac{h^2}{\mu} = a(1 - e^2) \quad h = \sqrt{\mu p} \quad (1-19)$$

1.3.3 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. (1-14)] with the velocity vector. Recognize that $\dot{\vec{v}} = \dot{\vec{r}}$.

$$\dot{\vec{v}} \cdot \dot{\vec{r}} + \vec{v} \cdot \frac{\mu}{r^3} \vec{r} = 0$$

Next, eliminate the vectors. The general form of a dot product is $\dot{\vec{r}} \cdot \vec{v} = r v \cos(\theta)$, which is a scalar quantity. However, from Fig. 1-13, 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 $\dot{\vec{r}} \cdot \vec{v} = r \dot{r}$. 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 a form of the *vis-viva equation*. Rearranging then,

$$\xi = \frac{v^2}{2} - \frac{\mu}{r} + c \quad (1-20)$$

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.” A term called $c_3 = v_\infty^2$ for hyperbolic trajectories, and $c_3 = v^2 - 2\mu/r$ for general applications, is also sometimes used. Using Eq. (1-22) 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. (1-17) for the velocity at periapsis and substituting into Eq. (1-20) 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

$$\xi = -\frac{\mu}{2a} \quad (1-21)$$

The previous equations are sometimes combined to give an alternate formulation of the energy interval (Eq. (1-31) squared). This is the traditional form of the *vis-viva equation*.

$$v^2 = \mu \left(\frac{2}{r} - \frac{1}{a} \right) \quad (1-22)$$

1.3.4 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 cross-ing the two-body equation [Eq. (1-14)] into the specific angular momentum vector, \hat{h} :

$$\ddot{\vec{r}} \times \hat{h} + \frac{\mu \dot{\vec{r}} \times \hat{h}}{r^3} = 0$$

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-7)] allows us to expand the equation above. This identity, along with the definition of the angular momentum vector [Eq. (1-15)], permits

$$\frac{\mu}{r^3}(\dot{\vec{r}} \times \hat{\vec{h}}) = \frac{\mu}{r^3}\dot{\vec{r}} \times (\dot{\vec{r}} \times \hat{\vec{v}}) = \frac{\mu}{r^3}((\dot{\vec{r}} \cdot \hat{\vec{v}})\hat{\vec{r}} - (\dot{\vec{r}} \cdot \hat{\vec{r}})\hat{\vec{v}}) = \frac{\mu}{r^3}((\dot{\vec{r}} \cdot \hat{\vec{v}})\hat{\vec{r}} - r^2\hat{\vec{v}})$$

From our discussion of the vector dot product while deriving ξ , we know we can simplify the dot product to get

$$\frac{\mu}{r^3}(\dot{\vec{r}} \times \hat{\vec{h}}) = \frac{\mu}{r^2}\dot{\vec{r}} \cdot \hat{\vec{r}} - \frac{\mu}{r}\hat{\vec{v}}$$

The key is to recognize this quantity as the following time derivative:

$$-\mu \frac{d}{dt}\left(\frac{\dot{\vec{r}}}{r}\right) = \frac{\mu}{r^2}\dot{\vec{r}} \cdot \hat{\vec{r}} - \frac{\mu}{r}\hat{\vec{v}}$$

We can substitute into the starting equation and reverse the equation:

$$\frac{d}{dt}(\dot{\vec{r}} \times \hat{\vec{h}}) = \mu \frac{d}{dt}\left(\frac{\dot{\vec{r}}}{r}\right)$$

Integrating both sides,

$$\dot{\vec{r}} \times \hat{\vec{h}} = \mu \frac{\dot{\vec{r}}}{r} + \vec{B} \quad (1-23)$$

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 $\dot{\vec{r}}$ to get a scalar relation:

$$\dot{\vec{r}} \cdot (\dot{\vec{r}} \times \hat{\vec{h}}) = \dot{\vec{r}} \cdot \left(\mu \frac{\dot{\vec{r}}}{r} + \vec{B}\right)$$

Using another vector identity, rearrange the first term as $(\dot{\vec{r}} \times \dot{\vec{r}}) \cdot \hat{\vec{h}} = h^2$ 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

$$h^2 = \mu r + rB \cos(\nu)$$

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 periapsis 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. (1-19). The *trajectory equation* is then

$$r = \frac{p}{1 + e \cos(\nu)} \quad (1-24)$$

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. (1-24),

$$\dot{r} = \frac{p \dot{\nu} e \sin(\nu)}{(1 + e \cos(\nu))^2}$$

which simplifies to

$$\dot{r} = \frac{r \dot{\nu} e \sin(\nu)}{1 + e \cos(\nu)} \quad (1-25)$$

Remember, this value *isn't* equal to the magnitude of the velocity vector!

1.3.5 Kepler's Second and Third Laws

We now derive the *period*, 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 1-14 shows the geometries used to develop the period equation.

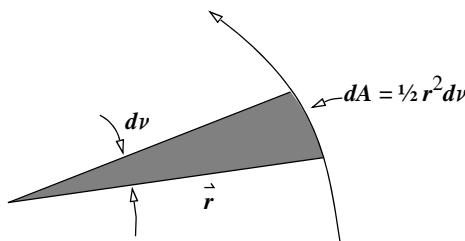


Figure 1-14. Area Swept Out by a Satellite. According to Kepler's second law, satellites sweep out equal areas in equal time intervals.

Recall from Fig. 1-13 the transverse component of velocity is simply $v \cos(\phi_{fpa})$, which we can also express in terms of simple rotational velocity, $r\dot{\nu}$. Rewriting Eq. (1-18) 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 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, ρ , 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:

$$\rho = \frac{2\pi ab}{h}$$

where πab is the area of the ellipse (and for a circle, πa^2). Recall the scaling formulas of Eq. (1-4) and Eq. (1-5). These formulas allow us to obtain

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

and using Eq. (1-19), we find an expression of Kepler's third law:

$$\rho = 2\pi \sqrt{\frac{a^3}{\mu}} \quad (1-26)$$

The period often categorizes the orbit. For instance, the U.S. Air Force's Space Command (AFSPC) 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. 9.10). 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 Posi-

tioning 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 about semi-synchronous altitude.

An alternate way to express Kepler's third law proves very useful for analysis. Let

$$n = \frac{2\pi}{\rho} = \sqrt[3]{\frac{\mu}{a^3}} \quad (1-27)$$

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. (1-26) squared (i.e., Kepler's third law) and rearranging yields an alternate form:

$$n^2 a^3 = \mu \quad \mu = G(m_1 + m_2) \quad (1-28)$$

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. (1-26) permits rapid determination of the size of a particular orbit (semimajor axis). If we solve Eq. (1-26) for the semimajor axis,

$$a = \left(\mu \left(\frac{\rho}{2\pi} \right)^2 \right)^{1/3} = \sqrt[3]{\frac{\mu}{n^2}} \quad (1-29)$$

Or we can solve for mean motion, n , which I'll discuss later. Consider the following example.

▼ Example 1-1: Determining the Semimajor Axis from the Period.

GIVEN: A geosynchronous orbit
FIND: Semimajor axis, a

ρ is equal to 24 sidereal hours ($86,400/1.002\ 737\ 909\ 350\ 795 = 86,164.090\ 518$ sec, 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.

$$\begin{aligned} a &= \left(\mu \left(\frac{\rho}{2\pi} \right)^2 \right)^{1/3} = \left(3.986\ 004\ 418 \times 10^5 \left(\frac{86164.0905}{2\pi} \right)^2 \right)^{1/3} = 42,164.1696 \text{ km} \\ &= \left(1 \left(\frac{106.795\ 863.3}{2\pi} \right)^2 \right)^{1/3} = 6.610\ 734\ 393 \text{ ER} \end{aligned}$$

Determine the semimajor axis in metric units by using the conversion from the table on the inside back cover for the Earth's radius (6378.137 km).

$$a = 6.610\ 734\ 393 (6378.137) = 42,164.1696 \text{ km}$$

From these results, we can derive several important points:

- (1) Using canonical units permits a savings in the calculation because μ is $1.0 \text{ ER}^3/\text{TU}^2$ versus $398,600.44 \text{ km}^3/\text{s}^2$.
- (2) The period in the numerator is the number of TUs in one *sidereal* day ($1436.068\ 174\ 815$ min), and it comes from the table on the inside back cover ($1.0/0.009\ 363\ 658\ 561\ 01 = 106.795\ 863\ 3$). Don't confuse this quantity with the TUs in one *solar* day ($107.088\ 260\ 7$).
- ▲ (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 ($1/\omega_{\oplus}$).

1.3.6 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. (1-20),

$$\xi = \frac{v^2}{2} - \frac{\mu}{r}$$

Solving for the velocity gives us

$$v = \sqrt{2\left(\frac{\mu}{r} + \xi\right)} \quad (1-30)$$

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. (1-21):*

$$v = \sqrt{\frac{2\mu}{r} - \frac{\mu}{a}} \quad (1-31)$$

Using the trajectory equation [Eq. (1-24)] and the definition of the semiparameter yields

$$r = \frac{p}{1 + e \cos(\nu)} = \frac{a(1 - e^2)}{1 + e \cos(\nu)}$$

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

Substituting this expression into the velocity equation [Eq. (1-31)] gives us

$$v = \sqrt{\frac{\mu}{r} \left(2 - \frac{1 - e^2}{1 + e \cos(\nu)} \right)} \quad (1-32)$$

* When writing computer software with Eq. (1-31), 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 [Eq. (1-32)] and not the semimajor axis.

The circle is just a special case of the ellipse, so we can derive the relevant formulas from the general case [Eq. (1-31)]. Because the radius and the semimajor axis are identical, let's replace the semimajor axis with the position and simplify:

$$v_{circle} = \sqrt{\frac{\mu}{r}} \quad (1-33)$$

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. (1-20); however, because the semimajor axis is infinite, Eq. (1-21) 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, find

$$v_{parabola} = v_{esc} = \sqrt{\frac{2\mu}{r}} \quad (1-34)$$

Eq. (1-34) gives the velocity on a parabolic orbit at any given radius, r .

1.4 Three-body and n -body Equations

The standard two-body relations underlie much of the general 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. 1.3. 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 nonspherical 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., $\vec{r}_{\oplus\odot} = \vec{r}_\odot - \vec{r}_\oplus$.

1.4.1 Inertial, Relative, and Barycentric Formulas

Eq. (1-11) still applies when determining the equations of motion for multiple bodies, except that additional forces arise from the interactions between the bodies. Figure 1-15 shows the general geometry.

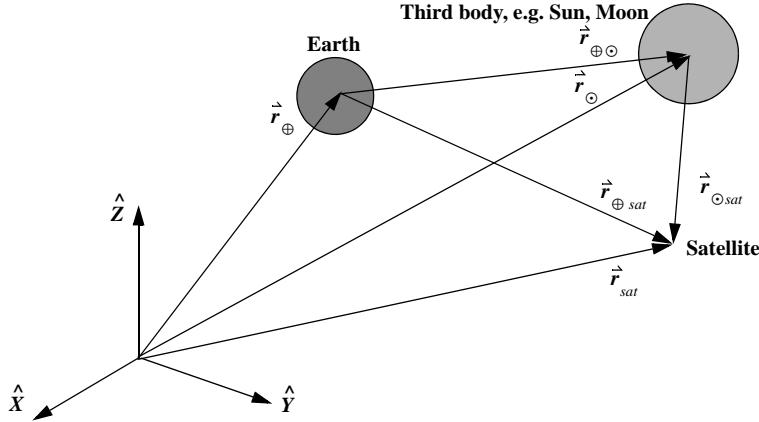


Figure 1-15. Three-body Geometry. The inertial coordinate system XYZ is used to measure all the accelerations. The double subscripts are necessary to completely define the direction of each vector. The satellite is intended to be orbiting the Earth, but for clarity, I've greatly exaggerated the figure.

For a three-body system, (Sun, Earth, and satellite, for instance), we examine the individual forces. The relationship between the vectors in Fig. 1-15 allows us to solve for the satellite's acceleration *relative to the Earth*:

$$\ddot{\vec{r}}_{\oplus sat} = \ddot{\vec{r}}_{sat} - \ddot{\vec{r}}_{\oplus}$$

Remember, this is an inertial system XYZ. Successive differentiation yields

$$\dddot{\vec{r}}_{\oplus sat} = \dddot{\vec{r}}_{sat} - \dddot{\vec{r}}_{\oplus}$$

Based on Newton's second law and the law of gravitation, the sum of the forces acting on the Earth is

$$\sum \vec{F}_{g\oplus} = m_{\oplus} \ddot{\vec{r}}_{\oplus} = \frac{Gm_{\oplus}m_{sat}\vec{r}_{\oplus sat}}{r_{\oplus sat}^3} + \frac{Gm_{\oplus}m_{\odot}\vec{r}_{\oplus\odot}}{r_{\oplus\odot}^3}$$

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 an observer would see from the origin of the inertial XYZ system.* For the satellite, the sum of the gravitational forces is

$$\sum \vec{F}_{gsat} = m_{sat} \ddot{\vec{r}}_{sat} = -\frac{Gm_{\oplus}m_{sat}\vec{r}_{\oplus sat}}{r_{\oplus sat}^3} - \frac{Gm_{\odot}m_{sat}\vec{r}_{\odot sat}}{r_{\odot sat}^3}$$

* The standard inertial system is the International Celestial Reference System (Sec. 3.3.2).

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

$$\ddot{\vec{r}}_{\oplus sat} = -\frac{G(m_{\oplus} + m_{sat})\vec{r}_{\oplus sat}}{r_{\oplus sat}^3} + Gm_{\odot} \left(\frac{\vec{r}_{sat\odot}}{r_{sat\odot}^3} - \frac{\vec{r}_{\oplus \odot}}{r_{\oplus \odot}^3} \right) \quad (1-35)$$

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 distinct 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. 8, 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 in Eq. (1-36). The three-body problem is a special case where $n = 3$.

$$\ddot{\vec{r}}_i = -G \sum_{\substack{j=1 \\ j \neq i}}^n \frac{m_j}{r_{ji}^3} \vec{r}_{ji} \quad \vec{r}_{ji} = \vec{r}_i - \vec{r}_j \quad i = 1, \dots, n \quad (1-36)$$

The inertial form is more useful if viewed from the center of mass of the primary body [as in Eq. (1-35)]. The general series representation of this relative acceleration (1 is the primary body, 2 is replaced with satellite) is

$$\ddot{\vec{r}}_{1sat} = -\frac{G(m_1 + m_{sat})\vec{r}_{1sat}}{r_{1sat}^3} + G \sum_{j=3}^n m_j \left(\frac{\vec{r}_{satj}}{r_{satj}^3} - \frac{\vec{r}_{1j}}{r_{1j}^3} \right) \quad (1-37)$$

Although the form of Eq. (1-36) is desirable to explore the ten known integrals of the motion (see Sec. 1.4.2), Eq. (1-37) 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. 1-16 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. (3-26).

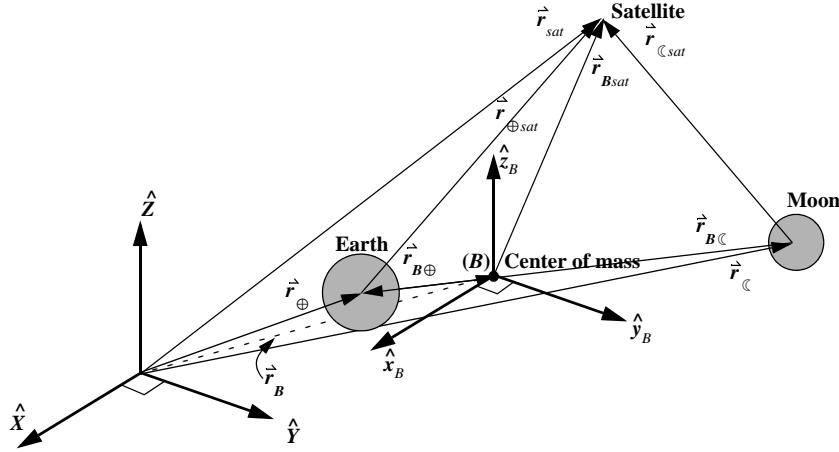


Figure 1-16. Geometry for the Barycentric Equations of Motion. Remember that the barycentric coordinate system is aligned with the inertial XYZ system. The Z and z_B axes are aligned.

Our ultimate goal is to obtain an expression for the satellite's acceleration as seen from the barycenter [center of mass in Eq. (1-39)]. The position of the i^{th} body in the barycentric frame is $\vec{r}_{Bi} = \vec{r}_i - \vec{r}_B$, so by Eq. (1-36), $\ddot{\vec{r}}_{ji}$ in the barycentric frame is

$$\ddot{\vec{r}}_{Bji} = \ddot{\vec{r}}_i - \ddot{\vec{r}}_B - (\ddot{\vec{r}}_j - \ddot{\vec{r}}_B) = \ddot{\vec{r}}_i - \ddot{\vec{r}}_j = \ddot{\vec{r}}_{ji}$$

Therefore, the acceleration of the i^{th} body in the barycentric frame is (see also Fig. 1-16)

$$\ddot{\vec{r}}_{Bi} = \ddot{\vec{r}}_i - \ddot{\vec{r}}_B = \ddot{\vec{r}}_i$$

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. (1-36) also applies to the barycentric formulation. The generalized barycentric equation of motion is

$$\ddot{\vec{r}}_{Bsat} = \ddot{\vec{r}}_{sat} = -G \sum_{\substack{j=1 \\ j \neq 3}}^n \frac{m_j \vec{r}_{jsat}}{r_{jsat}^3} \quad (1-38)$$

1.4.2 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. (1-39) 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. (1-36), the total is zero (a consequence of Newton's third law):

$$\sum_{i=1}^n m_i \ddot{\vec{r}}_i = -G \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \frac{m_i m_j \hat{\vec{r}}_{ji}}{r_{ji}^3} = 0$$

The zero result is also seen mathematically because ($\hat{\vec{r}}_{ij} = -\hat{\vec{r}}_{ji}$). We can integrate the left-hand term to yield position- and velocity-vector constants for the system:

$$\sum_{i=1}^n m_i \dot{\vec{r}}_i = m_{total} \dot{\vec{v}}_{sys} \quad \sum_{i=1}^n m_i \dot{\vec{r}}_i = (\dot{\vec{v}}_{sys} t + \dot{\vec{r}}_{sys}) m_{total}$$

From the definition of the system's center of mass, let's solve for the position, $\dot{\vec{r}}_{cm}$:

$$\dot{\vec{r}}_{cm} = \frac{\sum_{i=1}^n m_i \dot{\vec{r}}_i}{\sum_{i=1}^n m_i} = \dot{\vec{v}}_{sys} t + \dot{\vec{r}}_{sys} \quad (1-39)$$

The first six constants (two vectors) of the motion are now defined by $(\dot{\vec{r}}_{sys}/m_{total})$ and $(\dot{\vec{v}}_{sys}/m_{total})$, 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. 1.3.2). Crossing Eq. (1-36) 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

$$\sum_{i=1}^n m_i \ddot{\vec{r}}_i \times \dot{\vec{r}}_i = -G \sum_{i=1}^n \sum_{j=1, j \neq i}^n \frac{m_i m_j \dot{\vec{r}}_{ji} \times \dot{\vec{r}}_i}{r_{ji}^3} = 0$$

The right-hand side vanishes from Eq. (1-36), $\dot{\vec{r}}_{ji} \times \dot{\vec{r}}_i = \dot{\vec{r}}_i \times \dot{\vec{r}}_i - \dot{\vec{r}}_j \times \dot{\vec{r}}_i = -\dot{\vec{r}}_j \times \dot{\vec{r}}_i$, and because there is a double summation, we'll have $\dot{\vec{r}}_i \times \dot{\vec{r}}_j$ and $-\dot{\vec{r}}_i \times \dot{\vec{r}}_j$ terms, which will all cancel. For the left-hand side, we can use the relation

$$\ddot{\vec{r}}_i \times \dot{\vec{r}}_i = \frac{d}{dt}(\dot{\vec{r}}_i \times \dot{\vec{r}}_i)$$

(assuming the masses, m_i , are constant) to obtain

$$\sum_{i=1}^n m_i \ddot{\vec{r}}_i \times \dot{\vec{r}}_i = \sum_{i=1}^n m_i \frac{d}{dt}(\dot{\vec{r}}_i \times \dot{\vec{r}}_i) = \frac{d}{dt} \left[\sum_{i=1}^n m_i (\dot{\vec{r}}_i \times \dot{\vec{r}}_i) \right] = 0$$

Integrating this result yields the total angular momentum of the system:

$$\hat{\vec{h}}_{sys} = \sum_{i=1}^n m_i \dot{\vec{r}}_i \times \dot{\vec{r}}_i = \text{constant} \quad (1-40)$$

As with the two-body equations, the fact that angular momentum is conserved means that a plane perpendicular to \hat{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. 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. 1.3.3, but with a slight variation. Multiply Eq. (1-36) by m_i to find the force acting on each mass.

$$m_i \ddot{\vec{r}}_i = -G \sum_{\substack{j=1 \\ j \neq i}}^n \frac{m_i m_j \hat{r}_{ji}}{r_{ji}^3}$$

Because we have many bodies, the potential function (μ/r) is now

$$U = \frac{1}{2} G \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \frac{m_i m_j}{r_{ji}}$$

The gradient of the potential function gives the force acting on the i^{th} body (Sec. 8.2),

$$\frac{\partial U}{\partial \vec{r}_i} = -G \sum_{\substack{j=1 \\ j \neq i}}^n \frac{m_i m_j \hat{r}_{ji}}{r_{ji}^3}$$

and the equations of motion are

$$m_i \ddot{\vec{r}}_i = \frac{\partial U}{\partial \vec{r}_i}$$

Multiplying through by $\dot{\vec{r}}_i$ and summing over the individual bodies ($i = 1, 2, \dots, n$) gives

$$\sum_{i=1}^n m_i \dot{\vec{r}}_i \ddot{\vec{r}}_i = \sum_{i=1}^n \frac{\partial U}{\partial \vec{r}_i} d\vec{r}_i = \frac{dU}{dt}$$

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. (1-20). We can now write the system energy as

$$\xi_{sys} = \frac{1}{2} \sum_{i=1}^n m_i v_i^2 - \frac{1}{2} G \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \frac{m_i m_j}{r_{ji}} = \text{constant} \quad (1-41)$$

Problems

1. 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?
2. Eq. (1-5) relates the vertical distance of a circle and an ellipse. What is the corresponding relation for a hyperbola?
3. Identify the type of orbits given the following data:
 - a. $\hat{r} = 6372.0\hat{I} + 6372.0\hat{J}$ km, $\hat{v} = -4.7028\hat{I} + 4.7028\hat{J}$ km/s
 - b. $\hat{e} = 0.07\hat{I} + 0.021\hat{J} + 0.021\hat{K}$
 - c. $\hat{h} = 115969.46\hat{J}$ km²/s
 - d. $\hat{h} = -176475.27\hat{K}$ km²/s

CHAPTER 2

KEPLER'S EQUATION AND KEPLER'S PROBLEM

- 2.1 Historical Background
- 2.2 Kepler's Equation
- 2.3 Kepler's Problem
- 2.4 Satellite State Representations
- 2.5 Application: Orbital Elements from r and v
- 2.6 Application: r and v from Orbital Elements
- 2.7 Application: Groundtracks
- 2.8 Application: Find Time of Flight (*FINDTOF*)

2.1 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)

Kepler's first two laws (elliptical motion and equal areas in equal times) provide the necessary conditions for all two-body motion. We'll examine Kepler's first two laws in a little greater detail because they're the basis for Kepler's equation [Eq. (2-4)], 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 Koestler, 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):

$$r = 1 + e \cos(\nu) + Ae^2 \sin^2(\nu) + Be^3 \sin^2(\nu) \cos(\nu) + (O)e^4 \quad (2-1)$$

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.

2.2 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 2-1 shows the relationship between the area swept out, A_1 , and the associated angular displacement.

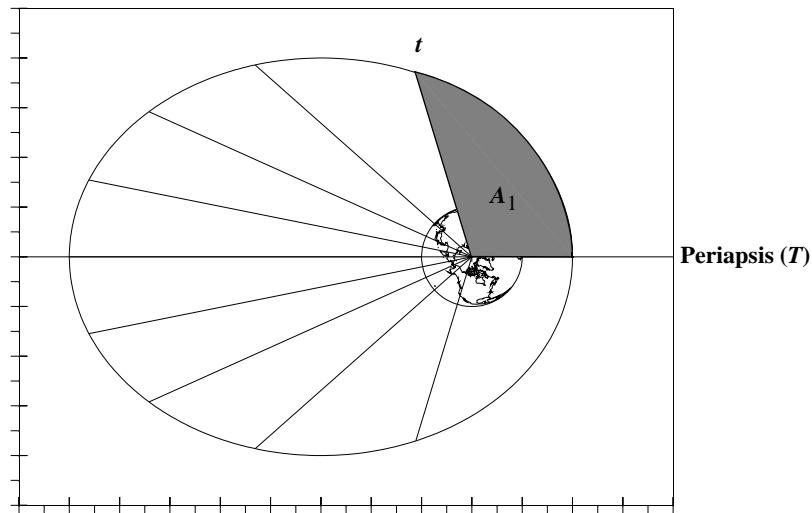


Figure 2-1. Kepler's Equation—Equal Areas. According to Kepler's second law, a satellite will sweep out equal areas in equal time. This figure shows positions for ten equal spacings in an orbit. For this particular orbit, $a = 31,890$ km, $e = 0.6$, and the space between the lines is about 94.4 minutes.

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

$$\frac{t - T}{A_1} = \frac{\rho}{\pi ab} \quad (2-2)$$

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. 2-2. 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. (1-5)] a bit differently for this derivation. Let

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

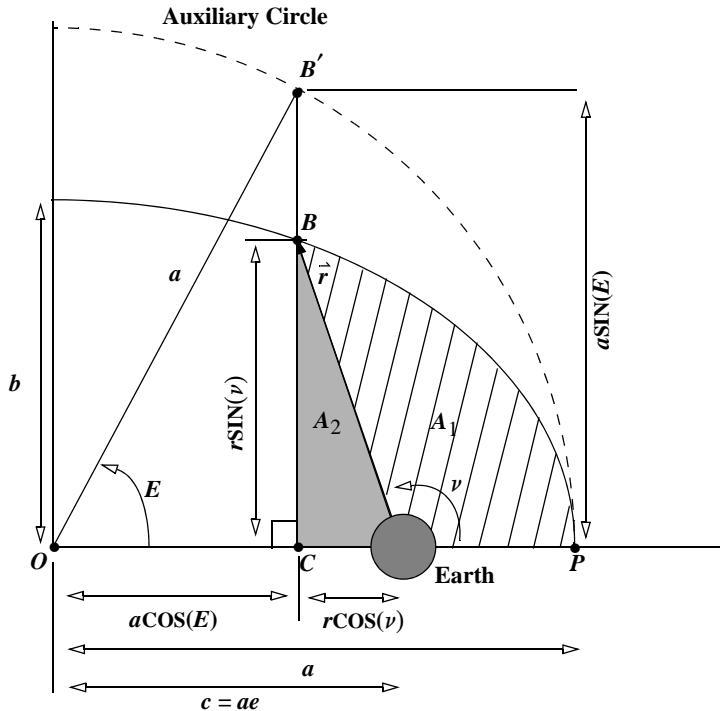


Figure 2-2. Geometry of Kepler's Equation. The eccentric anomaly uses an auxiliary circle as shown. The ultimate goal is to determine the area, A_1 , which allows us to calculate the time.

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

$$A_1 = \text{Area } PCB - \text{Triangle } A_2 \quad (2-3)$$

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. (1-5) (multiply by b/a), which relates the vertical distances of an ellipse and a circle. Thus,

$$\text{Area } PCB = \left[\frac{a^2 E}{2} - \frac{1}{2} (a \cos(E)) (a \sin(E)) \right] \frac{b}{a}$$

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. (2-3), you'll see the area

$$A_1 = \frac{ab}{2} [E - e \sin(E)]$$

Substitute back into Eq. (2-2) 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. (1-26), 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 periaxis to the eccentric anomaly, semimajor axis, and eccentricity. Kepler then introduces notation for the *mean anomaly*, M , as

$$M = E - e \sin(E) = \sqrt{\frac{\mu}{a^3}}(t-T) \quad (2-4)$$

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. (1-27). Kepler's third law relates ρ to the semimajor axis, so n is a function of a .

$$n \equiv \sqrt{\frac{\mu}{a^3}} \quad (2-5)$$

For instance, a satellite completing 14 revolutions per day has $n = (14)(360) = 5040^\circ/\text{day} = (14)(2\pi)/(86400) = 0.001\ 018\ 109 \text{ rad/sec}$. 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. 2.3 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

$$M = n(t - T) = E - e \sin(E) \quad (2-6)$$

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 2-3 shows the general situation.

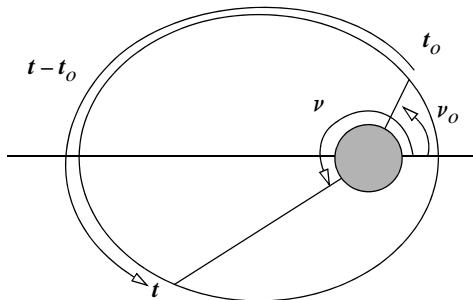


Figure 2-3. General Time of Flight. The time of flight between two specified locations has two basic forms: those of less than one period (as shown) and those with more than one period. Multiple periods include one or more complete revolutions in addition to the displacement shown here.

Again, t and t_0 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 $t - t_0$. 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,

$$\Delta t = k\varpi + t - t_0$$

which allows us to develop a completely general form of Kepler's equation (shown here in radians):

$$\frac{M - M_0}{n} = t - t_0 = \sqrt{\frac{a^3}{\mu}} \left\{ 2\pi k + E - e \sin(E) - (E_0 - e \sin(E_0)) \right\} \quad (2-7)$$

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. 2-2, we can directly obtain expressions relating eccentric anomaly and true anomaly.* These are considered basic relations because they provide a basis for many of the following derivations.

$$\sin(E) = \frac{r \sin(\nu)}{a \sqrt{1 - e^2}} \quad \cos(E) = \frac{ae + r \cos(\nu)}{a} \quad (2-8)$$

We want to find relations that are independent of the position magnitude. Begin by substituting the trajectory equation for the radius, r , into Eq. (2-8). Doing so results in

$$\begin{aligned} \sin(E) &= \frac{a(1 - e^2) \sin(\nu)}{a(1 + e \cos(\nu)) \sqrt{1 - e^2}} \\ \cos(E) &= e + \frac{\frac{a(1 - e^2)}{(1 + e \cos(\nu))} \cos(\nu)}{a} = \frac{e(1 + e \cos(\nu)) + (1 - e^2) \cos(\nu)}{(1 + e \cos(\nu))} \end{aligned}$$

Reducing gives you

$$\sin(E) = \frac{\sin(\nu) \sqrt{1 - e^2}}{1 + e \cos(\nu)} \quad \cos(E) = \frac{e + \cos(\nu)}{1 + e \cos(\nu)} \quad (2-9)$$

I've included the sine and cosine expressions so computer programs can run ATAN2 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. (2-9) for ν yields an immediate expression:

$$\cos(\nu) = \frac{\cos(E) - e}{1 - e \cos(E)} \quad (2-10)$$

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. (2-8) gives you

$$r = \frac{a \cos(E) - ae}{\cos(\nu)}$$

Using Eq. (2-10) above, note that

* Remember to divide by the scaling factor when converting from elliptical quantities to corresponding circular ones [Eq. (1-5)]. Eq. (2-8) is sometimes solved for $r \sin(\nu)$ or $r \cos(\nu)$.

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

and

$$r = a(1 - e\cos(E)) \quad (2-11)$$

You can also determine this quantity directly from Fig. 1-4.

Find the sine expression in terms of E and ν by solving Eq. (2-8) for $\sin(\nu)$:

$$\sin(\nu) = \frac{a\sin(E)\sqrt{1-e^2}}{r}$$

Now replace the position, r , with Eq. (2-11) and simplify:

$$\sin(\nu) = \frac{\sin(E)\sqrt{1-e^2}}{1 - e\cos(E)} \quad (2-12)$$

A single equation for true anomaly is useful. Divide Eq. (2-12) by Eq. (2-10) and use the tangent half-angle formula to expand the result:

$$\begin{aligned} \tan\left(\frac{\nu}{2}\right) &= \frac{\frac{\sin(E)\sqrt{1-e^2}}{1 - e\cos(E)}}{\frac{\cos(E) - e}{1 - e\cos(E)} + 1} \\ \tan\left(\frac{\nu}{2}\right) &= \frac{\sin(E)\sqrt{1-e^2}}{\cos(E) - e + 1 - e\cos(E)} \end{aligned}$$

Now recognize the denominator as the product of $(1 - e)(\cos(E) + 1)$; thus,

$$\tan\left(\frac{\nu}{2}\right) = \frac{\sin(E)\sqrt{1-e^2}}{(1 - e)(\cos(E) + 1)}$$

Reduce the terms containing eccentric anomaly with a formula for the tangent half-angle,

$$\tan\left(\frac{\nu}{2}\right) = \sqrt{\frac{1+e}{1-e}} \tan\left(\frac{E}{2}\right) \quad (2-13)$$

Then, solving for eccentric anomaly yields

$$\tan\left(\frac{E}{2}\right) = \sqrt{\frac{1-e}{1+e}} \tan\left(\frac{\nu}{2}\right) \quad (2-14)$$

2.2.1 Alternate Formulation for Eccentric Anomaly

Let's derive Kepler's equation [Eq. (2-4)] 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. (1-18)]. This equation defines the constant value for angular momentum:

$$h = \frac{r^2 d\nu}{dt} \quad (2-15)$$

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,

$$\int_T^t h dt = \int_0^\nu r^2 d\nu \quad (2-16)$$

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. (2-10), 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. (2-8) for $\sin(\nu)$ and substitute Eq. (2-11) for the squared term in the denominator:

$$d\nu = \frac{\sin(E)(1 - e^2)}{a \sqrt{1 - e^2} \sin(E) r^2 / a^2} dE$$

Simplify to get

$$d\nu = \frac{a \sqrt{1 - e^2}}{r} dE \quad (2-17)$$

Eq. (2-17) is very important when developing analytical and semianalytical routines. It's the key relation that allows us to change variables.

Substitute Eq. (2-17) into the integral [Eq. (2-16)] 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. (2-11) 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. (2-5) to get

$$h = na^2 \sqrt{1-e^2}$$

so the solution reduces to

$$(t-T) = \sqrt{\frac{a^3}{\mu}} (E - e \sin(E)) \quad (2-18)$$

2.2.2 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. (2-15):

$$h = \frac{r^2 d\nu}{dt} = \sqrt{\mu p}$$

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,

$$r = \frac{p}{2} \left[1 + \tan^2 \left(\frac{\nu}{2} \right) \right] \quad (2-19)$$

Substituting this result into Eq. (2-15) 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:

$$\int_T^t \frac{4}{\sqrt{p^3}} dt = \int_0^\nu \left(1 + 2 \tan^2 \left(\frac{\nu}{2} \right) + \tan^4 \left(\frac{\nu}{2} \right) \right) d\nu$$

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:

$$2 \sqrt{\frac{\mu}{p^3}} (t - T) = \tan \left(\frac{\nu}{2} \right) + \frac{\tan^3 \left(\frac{\nu}{2} \right)}{3} \quad (2-20)$$

Now you can define *parabolic anomaly*, B , and mean motion, n_p :

$$B \equiv \tan \left(\frac{\nu}{2} \right) \quad n_p \equiv 2 \sqrt{\frac{\mu}{p^3}} \quad (2-21)$$

Especially note the form of mean angular velocity, n_p , or mean motion, and parabolic anomaly, B .*

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

* 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. Herrick (1971:191) echoes this idea and cites the IAU adoption of B to honor Barker. This book uses M for all occurrences of mean anomaly and reserves B for parabolic orbits.

mean-motion expression, we maintain the relationship between constant angular momentum and equal areas of Kepler's second law. Finally, we usually use radian units for B because it's not associated with trigonometric functions.

We can now simplify the time equation by writing

$$n_p(t - T) = B + \frac{B^3}{3} \quad (2-22)$$

As in the elliptical case, we generalize the results to include initial and final locations which aren't directly at periapsis:

$$\frac{M - M_o}{n_p} = t - t_o = \frac{1}{2} \sqrt{\frac{p^3}{\mu}} \left\{ B + \frac{B^3}{3} - \left(B_o + \frac{B_o^3}{3} \right) \right\} \quad (2-23)$$

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. (2-24) is useful for determining the position magnitude and may be expressed in terms of B . From Eq. (2-19),

$$r = \frac{p}{2}[1 + B^2] \quad (2-24)$$

Find sine and cosine expressions using Eq. (2-22) and the formula for the tangent half-angle:

$$\tan\left(\frac{\nu}{2}\right) = \frac{\sin(\nu)}{\cos(\nu) + 1} \quad (2-25)$$

Using the trajectory equation (with $e = 1$) and solving for $\cos(\nu)$ gives you

$$\cos(\nu) = \frac{p - r}{r} \quad (2-26)$$

Substituting Eq. (2-26) into Eq. (2-25) and rearranging terms leads to

$$\sin(\nu) = \frac{pB}{r} \quad (2-27)$$

Mathematically, ν may change from -180° to 180° ; however, in practice, the angle is limited for parabolic orbits. Consider solving Eq. (2-21) 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° .

2.2.3 Formulation for the Hyperbolic Anomaly

The hyperbolic anomaly is similar to the eccentric anomaly (see Fig. 2-4).

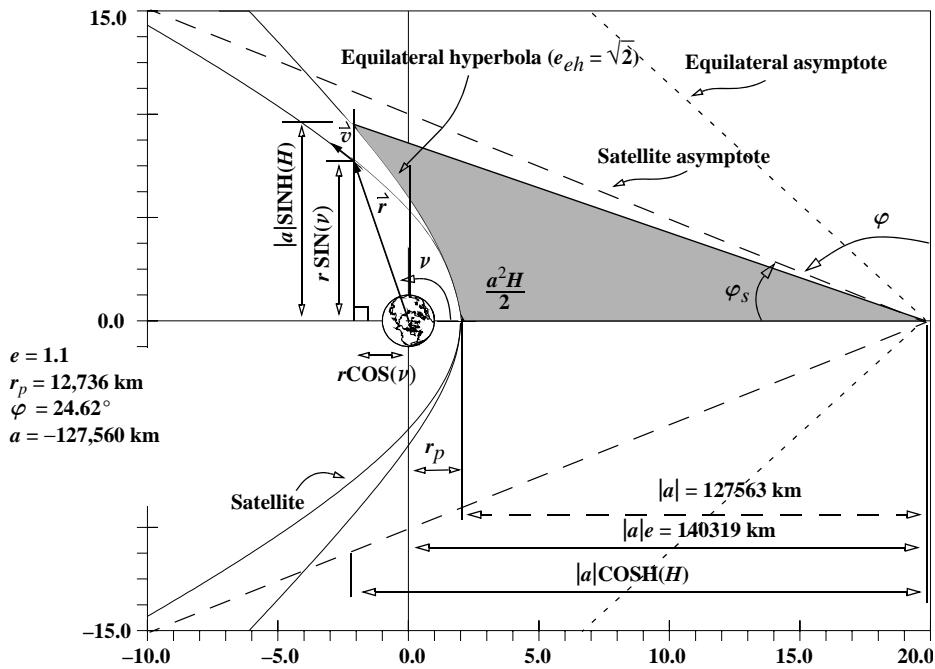


Figure 2-4. Hyperbolic Anomaly. The hyperbolic anomaly, H , is actually a measure of area, shown as the shaded region in the figure. Notice we use an equilateral hyperbola to determine H , just as we use a circle for elliptical orbits. Semimajor axis quantities are shown as absolute values to keep distances positive. The satellite and equilateral hyperbola have the same semimajor axis value. Thus, the focus for the equilateral hyperbola does not coincide with that of the satellite and its $r_p = 52811 \text{ km}$.

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 the angle between the two asymptotes. Some sources use the supplementary angle φ_s from the horizontal (Battin 1987:166). Kaplan (1976:93) shows the equation for the turning angle.

$$\sin\left(\frac{\varphi}{2}\right) = \frac{1}{e} \quad (2-28)$$

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. (1-5)]; 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

$$-180^\circ + \cos^{-1}\left(\frac{1}{e}\right) < \nu < 180^\circ - \cos^{-1}\left(\frac{1}{e}\right) \quad (2-29)$$

The following table shows some representative values of the limitations on the true anomaly of a hyperbolic orbit whose eccentricity varies.

TABLE 2-1. Hyperbolic Limitations on True Anomaly. The true anomaly may assume 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. 2-4, 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.

$$\text{SINH}(H) = \frac{-r \sin(\nu)}{a\sqrt{e^2 - 1}} \quad \text{COSH}(H) = \frac{ae + r \cos(\nu)}{a} \quad (2-30)$$

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 :

$$\begin{aligned}\text{SINH}(H) &= \frac{-a(1-e^2)\text{SIN}(\nu)}{a(1+e\cos(\nu))\sqrt{e^2-1}} \\ \text{COSH}(H) &= \frac{ae + \frac{a(1-e^2)}{(1+e\cos(\nu))}\cos(\nu)}{a} = \frac{e(1+e\cos(\nu))+(1-e^2)\cos(\nu)}{(1+e\cos(\nu))}\end{aligned}$$

Reducing and distributing the minus sign in the hyperbolic sine term gives us

$$\begin{aligned}\text{SINH}(H) &= \frac{\text{SIN}(\nu)\sqrt{e^2-1}}{1+e\cos(\nu)} \\ \text{COSH}(H) &= \frac{e+\cos(\nu)}{1+e\cos(\nu)}\end{aligned}\tag{2-31}$$

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. (2-31) for an expression in terms of ν :

$$\cos(\nu) = \frac{\text{COSH}(H)-e}{1-e\text{COSH}(H)}\tag{2-32}$$

To find an alternate expression for the position magnitude, begin with Eq. (2-30), and solve the cosine expression for the position magnitude:

$$r = \frac{a\text{COSH}(H)-ae}{\cos(\nu)}$$

Now use Eq. (2-32) to substitute for $\cos(\nu)$:

$$r = \frac{(a\text{COSH}(H)-ae)(1-e\text{COSH}(H))}{\text{COSH}(H)-e}$$

Simplifying results in

$$r = a(1-e\text{COSH}(H))\tag{2-33}$$

Find the sine expression for true anomaly by solving the sine expression of Eq. (2-30) and then using the position magnitude from Eq. (2-33):

$$\text{SIN}(\nu) = \frac{-a\text{SINH}(H)\sqrt{e^2-1}}{r} = \frac{-a\text{SINH}(H)\sqrt{e^2-1}}{a(1-e\text{COSH}(H))}$$

The result is

$$\text{SIN}(\nu) = \frac{-\text{SINH}(H)\sqrt{e^2-1}}{1-e\text{COSH}(H)}\tag{2-34}$$

A single equation for true anomaly is useful. Substituting Eq. (2-31) into the hyperbolic tangent half angle formula gives:

$$\begin{aligned}\tanh\left(\frac{H}{2}\right) &= \frac{\sinh(H)}{1 + \cosh(H)} = \frac{\frac{\sin(\nu)\sqrt{e^2 - 1}}{1 + e\cos(\nu)}}{1 + \frac{e + \cos(\nu)}{1 + e\cos(\nu)}} \\ &= \frac{\sin(\nu)\sqrt{e^2 - 1}}{1 + e\cos(\nu) + e + \cos(\nu)}\end{aligned}$$

By substitution, notice the denominator is the product of $(1 + e)(1 + \cos(\nu))$, so

$$\tanh\left(\frac{H}{2}\right) = \frac{\sin(\nu)\sqrt{e^2 - 1}}{(1 + e)(1 + e\cos(\nu))}$$

Then, reduce the terms containing H with a formula for the tangent half-angle:

$$\tanh\left(\frac{H}{2}\right) = \sqrt{\frac{e-1}{e+1}} \tan\left(\frac{\nu}{2}\right) \quad (2-35)$$

Solving for the true anomaly results in

$$\tan\left(\frac{\nu}{2}\right) = \sqrt{\frac{e+1}{e-1}} \tanh\left(\frac{H}{2}\right) \quad (2-36)$$

With the conversions identified, solve Kepler's equation. Differentiate Eq. (2-32) 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. (2-30) for $\sin(\nu)$ and substitute Eq. (2-33) for the squared term in the denominator:

$$d\nu = \frac{\sinh(H)(1 - e^2)}{\frac{a\sinh(H)\sqrt{e^2 - 1}}{r} \frac{r^2}{a^2}}dH$$

Simplify to get

$$d\nu = \frac{a(1-e^2)}{r\sqrt{e^2-1}} dH = \frac{-a\sqrt{e^2-1}}{r} dH \quad (2-37)$$

Remember that a is intrinsically negative here, so be careful of the signs while simplifying. Now, substitute Eq. (2-37) into the integral [Eq. (2-16)] 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. (2-33) 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

$$(t-T) = \frac{\sqrt{a^4(e^2-1)}}{\sqrt{\mu a(1-e^2)}} (e \sinh(H) - H)$$

Reduce and keep track of the negative sign to write

$$(t-T) = \sqrt{\frac{-a^3}{\mu}} (e \sinh(H) - H) \quad (2-38)$$

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

$$\frac{M-M_o}{n} = t-t_o = \sqrt{\frac{-a^3}{\mu}} \left\{ e \sinh(H) - H - (e \sinh(H_o) - H_o) \right\} \quad (2-39)$$

2.2.4 Universal Formulation

Developing the preceding anomalies is straight forward when processing data from one satellite. But building applications for an orbit determination system for hundreds or thousands of satellites including all orbit types indicates a need for 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. (1-18)],

$$\xi = \frac{v^2}{2} - \frac{\mu}{r} = -\frac{\mu}{2a}$$

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

$$\dot{\chi} = \frac{\sqrt{\mu}}{r} \quad (2-40)$$

Now you need to develop a general equation for position, r , in terms of this universal variable. Squaring Eq. (2-40) 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. (2-79), 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:

$$r = a \left(1 + e \sin \left(\frac{\chi + c_o}{\sqrt{a}} \right) \right) \quad (2-41)$$

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. (2-41) into Eq. (2-40).

$$\frac{d\chi}{dt} = \frac{\sqrt{\mu}}{a \left(1 + e \sin \left(\frac{\chi + c_o}{\sqrt{a}} \right) \right)}$$

Separate and integrate the equation to find the result, assuming that $\chi = 0$ at the initial time:

$$\sqrt{\mu} \Delta t = a \chi - ae \sqrt{a} \left\{ \cos \left(\frac{\chi + c_o}{\sqrt{a}} \right) - \cos \left(\frac{c_o}{\sqrt{a}} \right) \right\} \quad (2-42)$$

Notice that these equations [Eq. (2-42) and Eq. (2-41)] 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. (2-42) above with trigonometric identities for the cosine of a sum:

$$\sqrt{\mu}\Delta t = a\chi - ae\sqrt{a} \left\{ \cos\left(\frac{\chi}{\sqrt{a}}\right) \cos\left(\frac{c_o}{\sqrt{a}}\right) - \sin\left(\frac{\chi}{\sqrt{a}}\right) \sin\left(\frac{c_o}{\sqrt{a}}\right) - \cos\left(\frac{c_o}{\sqrt{a}}\right) \right\} \quad (2-43)$$

Now, determine expressions for the terms having the integration constant c_o . To eliminate these terms, use the position equation [Eq. (2-41)] and separate as follows:

$$e \sin\left(\frac{\chi + c_o}{\sqrt{a}}\right) = \frac{r}{a} - 1$$

However, remember at the initial time, $\chi = t = 0$, and the position is the magnitude of the original position vector. So

$$e \sin\left(\frac{c_o}{\sqrt{a}}\right) = \frac{r_o}{a} - 1 \quad (2-44)$$

Next, differentiate the position equation [Eq. (2-41)] to see the effect of the universal variable on the velocity. Be sure to remember the derivative of the universal variable in Eq. (2-40). 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 $r\dot{r} = \dot{r} \cdot \vec{v}$ developed in Sec. 1.3.3, note that

$$e \cos\left(\frac{c_o}{\sqrt{a}}\right) = \frac{\dot{r}_o \cdot \vec{v}_o}{\sqrt{\mu a}} \quad (2-45)$$

Now, substitute these two expressions containing c_o , Eq. (2-44) and Eq. (2-45), into the expanded time equation [Eq. (2-43)]:

$$\sqrt{\mu}\Delta t = a\chi - a\sqrt{a} \left\{ \cos\left(\frac{\chi}{\sqrt{a}}\right) \frac{\dot{r}_o \cdot \vec{v}_o}{\sqrt{\mu a}} - \sin\left(\frac{\chi}{\sqrt{a}}\right) \left(\frac{r_o}{a} - 1 \right) - \frac{\dot{r}_o \cdot \vec{v}_o}{\sqrt{\mu a}} \right\}$$

Although not obvious now, it's useful to rearrange the equation:

$$\sqrt{\mu} \Delta t = a \left\{ \chi - \sqrt{a} \sin \left(\frac{\chi}{\sqrt{a}} \right) \right\} + \frac{\dot{r}_o \cdot \vec{v}_o}{\sqrt{\mu}} a \left\{ 1 - \cos \left(\frac{\chi}{\sqrt{a}} \right) \right\} + r_o \sqrt{a} \sin \left(\frac{\chi}{\sqrt{a}} \right) \quad (2-46)$$

Use a trigonometric identity to expand the position equation [Eq. (2-41)]:

$$r = a + a e \left\{ \sin \left(\frac{\chi}{\sqrt{a}} \right) \cos \left(\frac{c_o}{\sqrt{a}} \right) + \cos \left(\frac{\chi}{\sqrt{a}} \right) \sin \left(\frac{c_o}{\sqrt{a}} \right) \right\}$$

Make the same substitutions for the expressions containing c_o , [Eq. (2-44) and Eq. (2-45)]:

$$r = a + a \left\{ \sin \left(\frac{\chi}{\sqrt{a}} \right) \frac{\dot{r}_o \cdot \vec{v}_o}{\sqrt{\mu a}} + \cos \left(\frac{\chi}{\sqrt{a}} \right) \left(\frac{r_o}{a} - 1 \right) \right\} \quad (2-47)$$

At this point, we have the time [Eq. (2-46)] and position [Eq. (2-47)] 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:

$$\psi \equiv \frac{\chi^2}{a} \quad (2-48)$$

Solve Eq. (2-48) for the semimajor axis and substitute it into the time and position equations [Eq. (2-46) and Eq. (2-47)] to get

$$\begin{aligned} \sqrt{\mu} \Delta t &= \frac{\chi^2}{\psi} \left\{ \chi - \frac{\chi}{\sqrt{\psi}} \sin(\sqrt{\psi}) \right\} + \frac{\dot{r}_o \cdot \vec{v}_o \chi^2}{\sqrt{\mu}} \frac{1 - \cos(\sqrt{\psi})}{\psi} + \left(\frac{r_o \chi}{\sqrt{\psi}} \right) \sin(\sqrt{\psi}) \\ r &= \frac{\chi^2}{\psi} + \frac{\chi^2}{\psi} \left\{ \sin(\sqrt{\psi}) \frac{\dot{r}_o \cdot \vec{v}_o \sqrt{\psi}}{\chi \sqrt{\mu}} + \cos(\sqrt{\psi}) \left(\frac{r_o \psi}{\chi^2} - 1 \right) \right\} \end{aligned}$$

With a little rearranging, find

$$\begin{aligned} \sqrt{\mu} \Delta t &= \chi^3 \left\{ \frac{\sqrt{\psi} - \sin(\sqrt{\psi})}{\sqrt{\psi^3}} \right\} + \frac{\dot{r}_o \cdot \vec{v}_o}{\sqrt{\mu}} \chi^2 \left\{ \frac{1 - \cos(\sqrt{\psi})}{\psi} \right\} + \frac{r_o \chi \sin(\sqrt{\psi})}{\sqrt{\psi}} \\ r &= \chi^2 \left\{ \frac{1 - \cos(\sqrt{\psi})}{\psi} \right\} + \frac{\dot{r}_o \cdot \vec{v}_o}{\sqrt{\mu}} \chi \left\{ 1 - \psi \frac{\sqrt{\psi} - \sin(\sqrt{\psi})}{\sqrt{\psi^3}} \right\} + r_o \left\{ 1 - \psi \frac{1 - \cos(\sqrt{\psi})}{\psi} \right\} \end{aligned}$$

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. For now, let's separate and calculate the common terms with the following definitions (Battin, 1987:181):

$$c_2 \equiv \frac{1 - \cos(\sqrt{\psi})}{\psi} \quad c_3 \equiv \frac{\sqrt{\psi} - \sin(\sqrt{\psi})}{\sqrt{\psi^3}} \quad (2-49)$$

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.

$$\begin{aligned} \sqrt{\mu}\Delta t &= \chi^3 c_3 + \frac{\overset{\circ}{r}_o \cdot \overset{\circ}{v}_o}{\sqrt{\mu}} \chi^2 c_2 + r_o \chi (1 - \psi c_3) \\ r &= \chi^2 c_2 + \frac{\overset{\circ}{r}_o \cdot \overset{\circ}{v}_o}{\sqrt{\mu}} \chi (1 - \psi c_3) + r_o (1 - \psi c_2) \end{aligned} \quad (2-50)$$

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. (1-27) for $\sqrt{\mu}$, we can divide through and obtain

$$n\Delta t = c_3 \left(\frac{\chi}{\sqrt{a}} \right)^3 + \frac{c_2 \overset{\circ}{r}_o \cdot \overset{\circ}{v}_o}{\sqrt{\mu a}} \left(\frac{\chi}{\sqrt{a}} \right)^2 + \frac{r_o}{a} \left(\frac{\chi}{\sqrt{a}} \right) (1 - \psi c_3)$$

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

$$\begin{aligned} c_2 &= \frac{1 - \cos(\sqrt{\psi})}{\psi} \text{ for } \psi \geq 0 \quad c_2 = \frac{1 - \cosh(\sqrt{-\psi})}{\psi} \text{ for } \psi < 0 \\ &= \sum_{k=0}^{\infty} \frac{(-\psi)^k}{(2k+2)!} = \frac{1}{2!} - \frac{\psi}{4!} + \frac{\psi^2}{6!} - \frac{\psi^3}{8!} + \frac{\psi^4}{10!} - \frac{\psi^5}{12!} + \dots \\ c_3 &= \frac{\sqrt{\psi} - \sin(\sqrt{\psi})}{\sqrt{\psi^3}} \text{ for } \psi \geq 0 \quad c_3 = \frac{\sinh(\sqrt{-\psi}) - \sqrt{-\psi}}{\sqrt{-\psi^3}} \text{ for } \psi < 0 \\ &= \sum_{k=0}^{\infty} \frac{(-\psi)^k}{(2k+3)!} = \frac{1}{3!} - \frac{\psi}{5!} + \frac{\psi^2}{7!} - \frac{\psi^3}{9!} + \frac{\psi^4}{11!} - \frac{\psi^5}{13!} + \dots \end{aligned}$$

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 motion 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 2-5 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 1 (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 1: *Find* $c_2 c_3 (\psi \Rightarrow c_2, c_3)$

$$\left[\begin{array}{ll} \text{IF } \psi > 1 \times 10^{-6} & \\ c_2 = \frac{1 - \cos(\sqrt{\psi})}{\psi} & c_3 = \frac{\sqrt{\psi} - \sin(\sqrt{\psi})}{\sqrt{\psi^3}} \\ \text{ELSE} & \\ \text{IF } \psi < -1 \times 10^{-6} & \\ c_2 = \frac{1 - \cosh(\sqrt{-\psi})}{\psi} & c_3 = \frac{\sinh(\sqrt{-\psi}) - \sqrt{-\psi}}{\sqrt{(-\psi)^3}} \\ \text{ELSE} & \\ c_2 = \frac{1}{2} & c_3 = \frac{1}{6} \end{array} \right]$$

2.2.5 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. (2-6) and Eq. (2-7). 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 most applications, which is extremely desirable. The bottom line is to use a method which gives correct results for your problem.

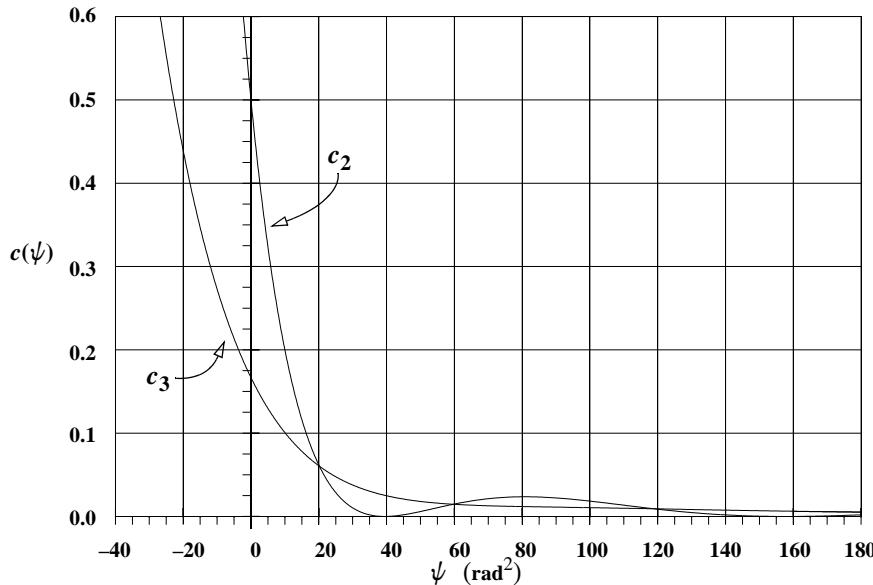


Figure 2-5. $c(\psi)$ Functions. Notice how quickly the variables tend towards zero. The zero points for c_2 occur where $\cos(\sqrt{\psi})$ equals one, or when ψ equals squared even multiples of 2π — $(2\pi)^2, (4\pi)^2, (6\pi)^2\dots$

You may encounter equinoctial formulations solving the eccentric anomaly. This particular arrangement is useful for analytical theories using equinoctial elements, however, I will not develop those equations here. See Pon (1973:86-88) for additional information.

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

$$x_{n+1} = x_n + \delta_n = x_n - \frac{f(x_n)}{f'(x_n)} \quad (2-51)$$

Note from Eq. (2-6) 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 2 presents the results.

ALGORITHM 2: *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. 2-6 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 74, for techniques.

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 2-7 plots the same eccentricities as those in Fig. 2-6. For now, consider an example:

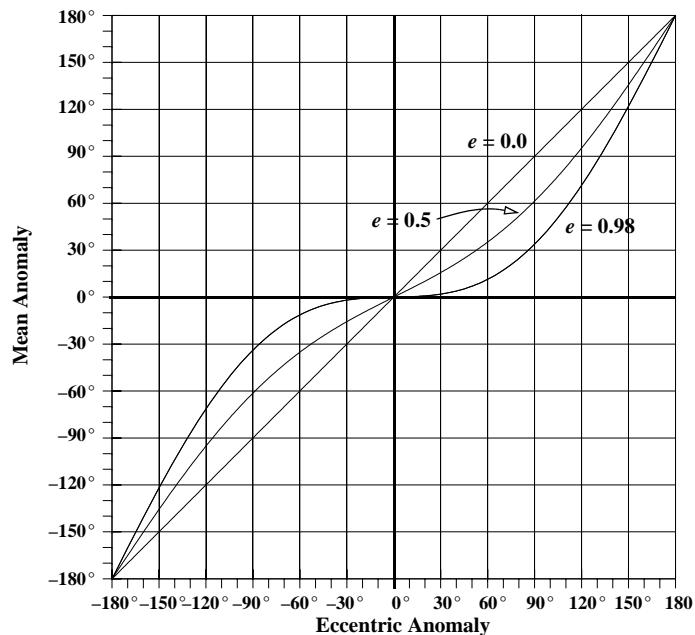


Figure 2-6. Mean Anomaly vs. Eccentric Anomaly. Although the difference between the eccentric and mean anomaly isn't great, as the eccentricity increases, the differences become larger.

▼ **Example 2-1: Using Kepler's Equation.**

GIVEN: $M = 235.4^\circ$, $e = 0.4$

FIND: E

Begin by letting $E_n = M - e = 4.108\ 505\ 06 - 0.4 = 3.708\ 505\ 06$ radians. 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

$$\begin{aligned} E_{n+1} &= E_n + \frac{M - E_n + e \sin(E_n)}{1 - e \cos(E_n)} \\ &= 3.708\ 505\ 06 + \frac{4.108\ 505\ 06 - 3.708\ 505\ 06 + (0.4) \sin(3.708\ 505\ 06)}{1 - 0.4 \cos(3.708\ 505\ 06)} \\ &= 3.848\ 697\ 1 \text{ rad} \end{aligned}$$

After the first iteration, the difference between E_n and E_{n+1} is still large. Therefore, reassign E_n and repeat the process:

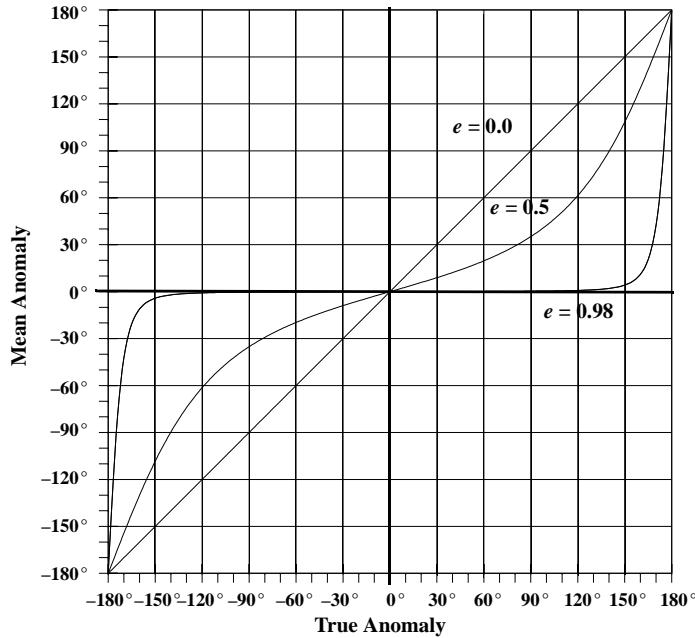


Figure 2-7. Mean Anomaly vs. True Anomaly. To obtain this plot, we first solve Kepler's equation and obtain the eccentric anomaly, E . We can then convert E to the true anomaly, ν . Notice the rapid slope changes as the eccentricity approaches 1.0. Because E , M , and ν are related through Eq. (2-7) and Eq. (2-9), these rapid changes translate to increased difficulty in solving Kepler's equation whenever e is near 1.0.

$$\begin{aligned}
 E_{n+1} &= E_n + \frac{M - E_n + e \sin(E_n)}{1 - e \cos(E_n)} \\
 &= 3.848\,697\,11 + \frac{4.108\,505\,06 - 3.848\,697\,11 + 0.4 \sin(3.848\,697\,11)}{1 - 0.4 \cos(3.848\,697\,11)} \\
 &= 3.848\,661\,46 \text{ rad}
 \end{aligned}$$

▲ Notice that the values of E_n and E_{n+1} are much closer. Subsequent iterations will yield the results in Table 2-2. When the process is complete, $E = 220.512\,074\,767\,522^\circ$.

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

TABLE 2-2. 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	$8.051\ 068\ 5 \times 10^{-15}$

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-degree polynomial:

$$\frac{B^3}{3} + B - n_p(t - T) = 0$$

An analytical solution of this third-degree polynomial will result in one real root, which is the answer. See "Cubic Roots" on page 1027.

Another solution uses trigonometric substitutions to solve for the parabolic anomaly. This formula is more common, but it's about 10-20% 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.* 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. (2-22). Substituting the identity into Eq. (2-22) 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

* 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.

$$\begin{aligned}\frac{1}{3} \left(\lambda^3 - \frac{1}{\lambda^3} \right) &= \frac{1}{3} \left(-\tan(s) + \frac{1}{\tan(s)} \right) = \frac{1}{3} \left(\frac{-1}{\cot(s)} + \cot(s) \right) \\ &= \frac{1}{3} \frac{\cot^2(s) - 1}{\cot(s)} = \frac{2}{3} \cot(2s)\end{aligned}$$

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 3: *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 2-2. Using Kepler's Equation for Parabolic Orbits.**

GIVEN: $\Delta t = 53.7874$ min, $p = 25,512$ km, $e = 1$

FIND: B

Begin by finding the mean motion:

$$n_p = 2 \sqrt[3]{\frac{\mu}{p^3}} = 2 \sqrt[3]{\frac{398600.4418}{25512^3}} = 0.000\ 309\ 9 \text{ rad/s}$$

Now, use the time equation:

$$\cot(2s) = \frac{3}{2}n_p\Delta t = \frac{3}{2}(0.000\ 309\ 9)(53.7874)(60) = 1.500\ 048 \text{ rad}$$

which means $s = 16.844\ 61^\circ$

$$\tan^3(w) = \tan(s) = \tan(16.844\ 61^\circ)$$

$$w = 33.880\ 776^\circ$$

Finally, find $B = 2\cot(2w) = 2\cot(33.880\ 776^\circ)$

▲ $B = 0.817\ 751 \text{ rad}$

Figure 2-8 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. 2-8 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° .

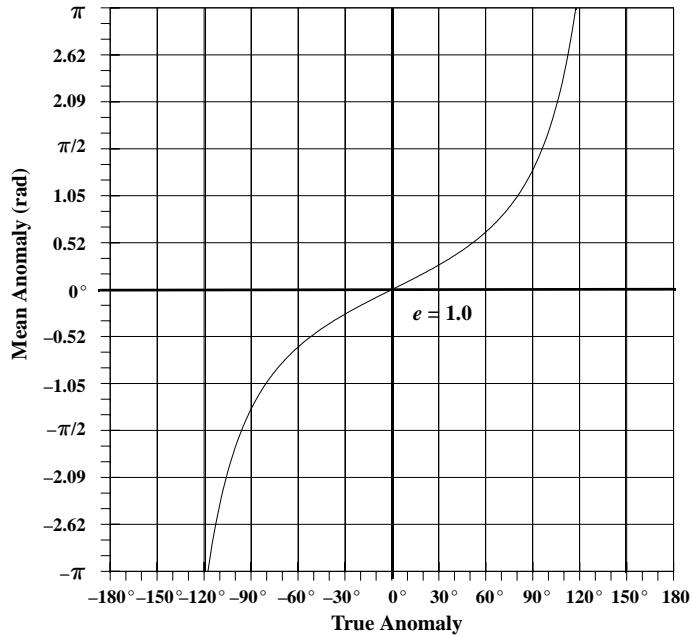


Figure 2-8. Mean Anomaly vs. True Anomaly for Parabolic Orbits. For a parabolic orbit, the mean anomaly is best described as a radian measure. Although the true anomaly appears to be constrained to certain values, at infinite distance true anomaly will equal $\pm 180^\circ$.

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, $f(H) = M - e\text{SINH}(H) + H$, and $f'(H) = -e\text{COSH}(H) + 1$. The result, as shown in Algorithm 4, is very similar to the elliptical case except for the initial guess. See *Hybrid Solutions*, page 74, for details on the initial guess.

ALGORITHM 4: KepEqtnH ($M, e \Rightarrow H$)

```

IF  $e < 1.6$ 
  IF  $\pi - M < M < 0$  or  $M > \pi$ 
     $H = M - e$ 
  ELSE
     $H = M + e$ 
ELSE
  IF  $e < 3.6$  and  $|M| > \pi$ 
     $H = M - \text{SIGN}(M) e$ 
  ELSE
     $H = \frac{M}{e - 1}$ 
LOOP

$$H_{n+1} = H_n + \frac{M - e \text{SINH}(H_n) + H_n}{e \text{COSH}(H_n) - 1}$$

UNTIL  $|H_{n+1} - H_n| < \text{tolerance}$ 

```

▼ Example 2-3. 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 ,

$$H_n = M - \text{SIGN}(M)e = (4.108\ 505\ 059) - (+1)2.4 = 1.708\ 505\ 059 \text{ rad}$$

Then, making sure you use radians, you'll find the hyperbolic anomaly is

$$H_{n+1} = H_n + \frac{M - e \text{SINH}(H_n) + H_n}{e \text{COSH}(H_n) - 1} = 1.607\ 435\ 562\ 6 \text{ rad}$$

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 2-3 summarizes the results.

TABLE 2-3. Values for Kepler's Equation—Hyperbolic Example 2-3. Each iterative value is in radians—from the starting values of $e = 2.4$, $M = 235.4^\circ = 4.108\ 505\ 059\ 194\ 65$ rad. The initial guess is $H_n = 1.708\ 505\ 059$.

Iteration, n	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. 2-7 gave us insight into solving Kepler's equation. Because M , H , and ν are related via Eq. (2-31) and Eq. (2-39), 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. 2-9) 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, some computer programs still use series solutions to the problem. Modern computers make these solutions less useful, but we 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 $k < 0, j = -k, -k+1, -k+2, \dots$

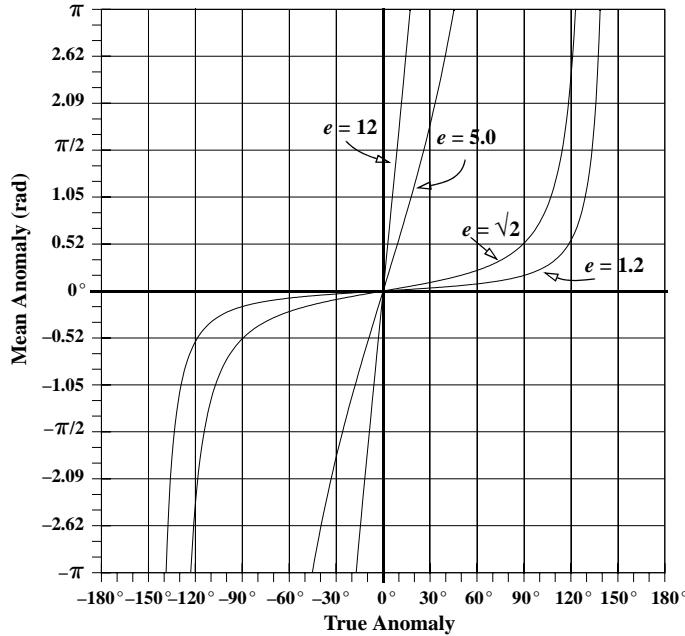


Figure 2-9. Mean Anomaly vs. True Anomaly for Hyperbolic Orbits. Notice the effective limits on true anomaly. Unlike the parabola, these limits are real because all hyperbolic orbits are limited by the asymptotes. High eccentricities produce nearly straight lines.

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:

$$\begin{aligned}
 E &= M + \left(e - \frac{e^3}{8} + \frac{e^5}{192} \right) \sin(M) + \left(\frac{e^2}{2} - \frac{e^4}{6} \right) \sin(2M) + \left(\frac{3e^3}{8} - \frac{27e^5}{128} \right) \sin(3M) + \left(\frac{e^4}{3} \right) \sin(4M) + \dots \\
 E &= \nu - \left(e + \frac{e^3}{4} \right) \sin(\nu) + \left(\frac{e^2}{4} + \frac{e^4}{8} \right) \sin(2\nu) - \left(\frac{e^3}{12} \right) \sin(3\nu) + \left(\frac{e^4}{32} \right) \sin(4\nu) + \dots \\
 \nu &= E + \left(e + \frac{3e^3}{12} \right) \sin(E) + \left(\frac{e^2}{4} + \frac{e^4}{24} \right) \sin(2E) + \left(\frac{e^3}{12} \right) \sin(3E) + \left(\frac{e^4}{96} \right) \sin(4E) + \dots \quad (2-52) \\
 \nu &= M + \left(2e - \frac{e^3}{4} + \frac{5e^5}{96} \right) \sin(M) + \left(\frac{5e^2}{4} - \frac{11e^4}{24} \right) \sin(2M) + \left(\frac{13e^3}{12} - \frac{43e^5}{64} \right) \sin(3M) \\
 &\quad + \left(\frac{103e^4}{96} \right) \sin(4M) + \dots \\
 M &= \nu - 2e \sin(\nu) + \left(\frac{3e^2}{4} + \frac{e^4}{8} \right) \sin(2\nu) - \left(\frac{e^3}{3} + \frac{e^5}{8} \right) \sin(3\nu) + \left(\frac{5e^4}{32} \right) \sin(4\nu) + \dots
 \end{aligned}$$

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 $v - M$.

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. (2-52) 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 2-10 shows the number of iterations for these cases. The time it requires is the baseline for comparison to the other methods.

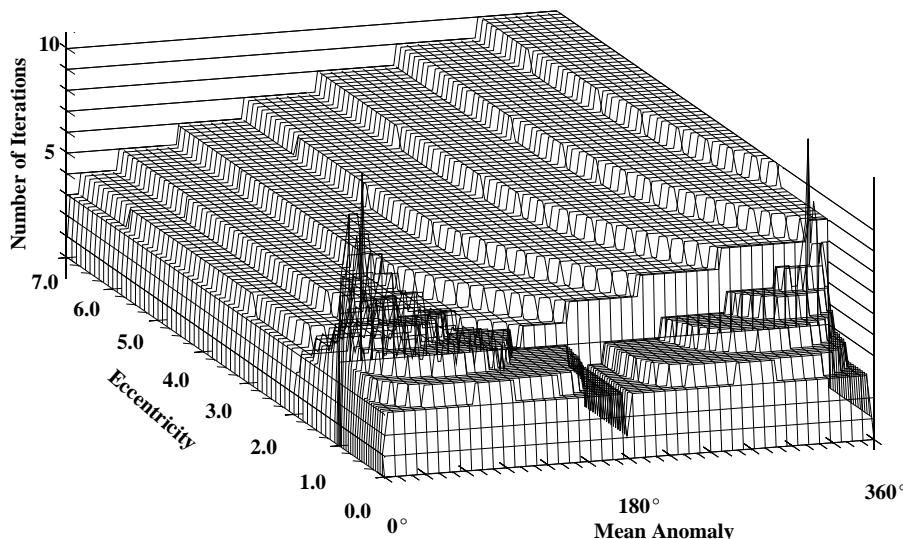


Figure 2-10. Solution of Kepler's Equation, $E, H = M$. This figure shows a complete spectrum of values through $\pm 360^\circ$, which enables correct sign determination and is crucial to effective computations. Notice the difficulty this method has at low values of mean anomaly and eccentricity near 1.0. Although we may not need a robust solution for everyday operations, some situations may require it.

Because Eq. (2-52) 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 eccentricity. Examination of Fig. 2-11 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,

```
IF  $-\pi < M < 0$  or  $M > \pi$   
let  $E, H = M - e$   
ELSE  
let  $E, H = M + e$ 
```

Figure 2-11 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.

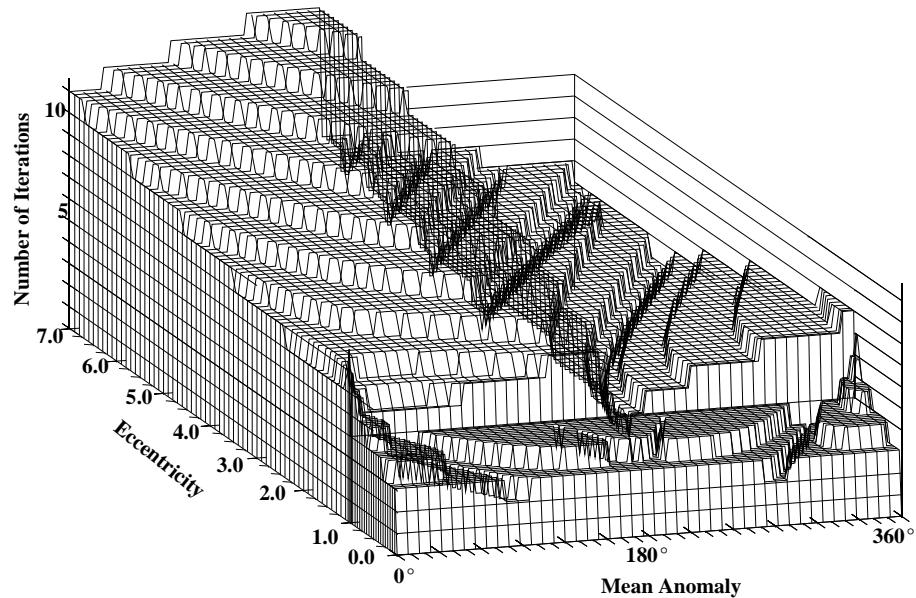


Figure 2-11. Solution of Kepler's Equation, $E, H = M + e$. Note the improvement in the eccentric case by simply adding the eccentricity when $0^\circ < M < 360^\circ$ and subtracting elsewhere. The hyperbolic case shows some improvement over Fig. 2-10, but only from 180° to 360° .

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 2-12 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. 2-9, 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 4, so consider them only an initial baseline.

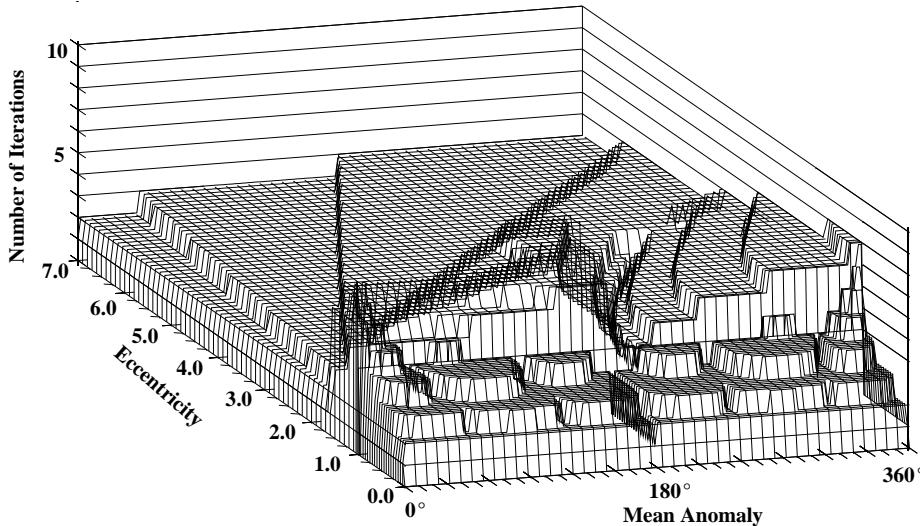


Figure 2-12. Solution of Kepler's Equation, $E = M + e \sin(M) + \frac{e^2}{2} \sin(2M)$. Including more terms in the initial guess improves the performance, but not significantly. The hyperbolic solution is actually a hybrid approach combining several desirable features of solution from the linearity of the problem. Notice that the hyperbolic solution actually improves as eccentricity increases.

The results in Fig. 2-12 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 4 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. 2-11 (Algorithm 2), the cubic solution for parabolic cases (Algorithm 3), and the technique for solving hyperbolic cases in Fig. 2-12 (Algorithm 4).

2.2.6 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—a key input 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 5: *ν to Anomaly ($e, \nu \Rightarrow E, B, H$)*

$$\begin{array}{lll} \text{if } e < 1.0 & \text{if } e = 1.0 & \text{if } e > 1.0 \\ \sin(E) = \frac{\sin(\nu)\sqrt{1-e^2}}{1+e\cos(\nu)} & B = \tan\left(\frac{\nu}{2}\right) & \sinh(H) = \frac{\sin(\nu)\sqrt{e^2-1}}{1+e\cos(\nu)} \\ \cos(E) = \frac{e+\cos(\nu)}{1+e\cos(\nu)} & & \cosh(H) = \frac{e+\cos(\nu)}{1+e\cos(\nu)} \end{array}$$

ALGORITHM 6: *Anomaly to ν ($e, E, B (p,r), H \Rightarrow \nu$)*

$$\begin{array}{lll} \text{if } e < 1.0 & \text{if } e = 1.0 & \text{if } e > 1.0 \\ \sin(\nu) = \frac{\sin(E)\sqrt{1-e^2}}{1-e\cos(E)} & \sin(\nu) = \frac{pB}{r} & \sin(\nu) = \frac{-\sinh(H)\sqrt{e^2-1}}{1-e\cosh(H)} \\ \cos(\nu) = \frac{\cos(E)-e}{1-e\cos(E)} & \cos(\nu) = \frac{p-r}{r} & \cos(\nu) = \frac{\cosh(H)-e}{1-e\cosh(H)} \end{array}$$

The derivation for the cosine of the flight-path angle begins with Eq. (2-94). Substitute Eq. (2-11).

$$\cos(\phi_{fpa}) = \sqrt{\frac{a^2(1-e^2)}{a(1-e\cos(E))\{2a-a(1-e\cos(E))\}}}$$

Rearrange the denominator and simplify the equation to get

$$\cos(\phi_{fpa}) = \sqrt{\frac{1-e^2}{(1-e\cos(E))(1+e\cos(E))}}$$

After expanding the denominator, you'll have the final result:

$$\cos(\phi_{fpa}) = \sqrt{\frac{1-e^2}{1-e^2\cos^2(E)}} \quad (2-53)$$

Find the sine expression of the flight-path angle using the Pythagorean relation of Eq. (C-10), which requires you to solve

$$\frac{1-e^2}{1-e^2\cos^2(E)} + \sin^2(\phi_{fpa}) = 1$$

Once you expand the $\cos^2(E)$ term with Eq. (C-10), you can quickly get

$$\sin(\phi_{fpa}) = \frac{e\sin(E)}{\sqrt{1-e^2\cos^2(E)}} \quad (2-54)$$

The solution for parabolic orbits is a little different. First, find the tangent expression for ϕ_{fpa} by using Eq. (2-95) and Eq. (2-96):

$$\tan(\phi_{fpa}) = \frac{e\sin(\nu)}{1+e\cos(\nu)}$$

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:

$$\phi_{parabolic} = \frac{\nu}{2} \quad (2-55)$$

Finally, hyperbolic orbits are similar to the elliptical case. Begin with Eq. (2-95) and substitute Eq. (2-34) for ($\sin(\nu)$):

$$\sin(\phi_{fpa}) = \frac{-e\sinh(H)\sqrt{e^2-1}}{\sqrt{1+2e\cos(\nu)+e^2(1-e\cosh(H))}}$$

Now, use Eq. (2-32) 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

$$\text{SIN}(\phi_{fpa}) = \frac{-e \text{SINH}(H)}{\sqrt{e^2 \text{COSH}^2(H) - 1}} \quad (2-56)$$

Develop the cosine expression using Eq. (C-10). For the solution, you'll need

$$\frac{e^2 \text{SINH}^2(H)}{e^2 \text{COSH}^2(H) - 1} + \text{COS}^2(\phi_{fpa}) = 1$$

Now use Eq. (C-11) to expand the hyperbolic sine term and get the final result, remembering the denominator must be the same as for the sine relation:

$$\text{COS}(\phi_{fpa}) = \sqrt{\frac{e^2 - 1}{e^2 \text{COSH}^2(H) - 1}} \quad (2-57)$$

2.3 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. 9) which aren't rigorous but are very useful. Fortunately, the modern computer allows calculations of these quantities with almost arbitrary accuracy. Figure 2-13 shows the general situation.

Although at first glance this seems to be an easy problem, it's a foundation in 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. 2.2.

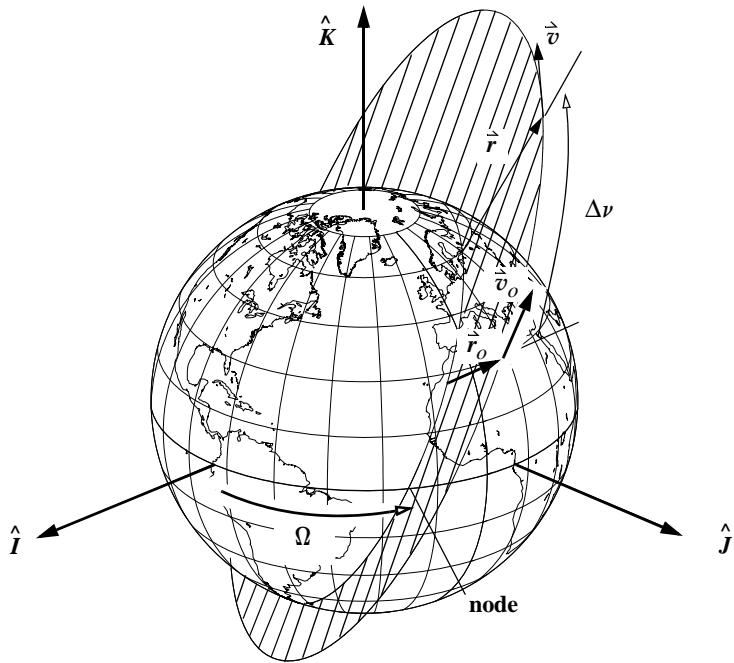


Figure 2-13. Geometry for Kepler's Problem. The concept behind Kepler's problem is the propagation of the satellite's state vector from an initial epoch to a future time.

2.3.1 Solution Techniques

We'll discuss three broad ways to solve Kepler's problem for two-body motion: an orbital-element technique, the series method using f and g functions, and the universal-variable techniques. Chapter 9 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 , ω_{true}). Knowing the conversion between the anomalies and the mean motion 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 9 (*RV2COE*), 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. 2.2 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 10, *COE2RV*).

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 7, below.

ALGORITHM 7: *KEPLERCOE* ($\hat{r}_o, \hat{v}_o, \Delta t \Rightarrow \hat{r}, \hat{v}$)

$$\begin{aligned} & \text{RV2COE} (\hat{r}_o, \hat{v}_o \Rightarrow p, a, e, i, \Omega, \omega, \nu(u, \lambda_{true}, \tilde{\omega}_{true})) \\ & \left[\begin{array}{l} \text{IF } e \neq 0 \\ \quad \text{vtoAnomaly} (e, \nu \Rightarrow E_o, B_o, H_o) \\ \text{ELSE} \\ \quad E_o = u \text{ or } E_o = \lambda_{true} \end{array} \right. \\ & \quad \begin{array}{lll} \text{if } e < 1.0 & \text{if } e = 1.0 & \text{if } e > 1.0 \\ M_o = E_o - e \sin(E_o) & \hat{h} = \hat{r}_o \times \hat{v}_o & M_o = e \sinh(H_o) - H_o \\ M = M_o + n \Delta t & p = \frac{h^2}{\mu} & M = M_o + n \Delta t \\ \text{KEPEQTNE} (M, e \Rightarrow E) & M_o = B_o + \frac{B_o^3}{3} & \text{KEPEQTNH} (M, e \Rightarrow H) \\ & \text{KEPEQTNP} (\Delta t, p \Rightarrow B) & \end{array} \\ & \left. \begin{array}{l} \text{IF } e \neq 0 \\ \quad \text{Anomalytov} (e, E, B, H \Rightarrow \nu) \\ \text{ELSE} \\ \quad u = E \text{ or } \lambda_{true} = E \end{array} \right] \\ & \quad \text{COE2RV} (a, e, i, \Omega, \omega, \nu(u, \lambda_{true}, \tilde{\omega}_{true}) \Rightarrow \hat{r}, \hat{v}) \end{aligned}$$

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. 2-14), write an expression for the position and velocity vectors simply from the components x and y in each axis:

$$\vec{r} = x\hat{P} + y\hat{Q} \quad \vec{v} = \dot{x}\hat{P} + \dot{y}\hat{Q}$$

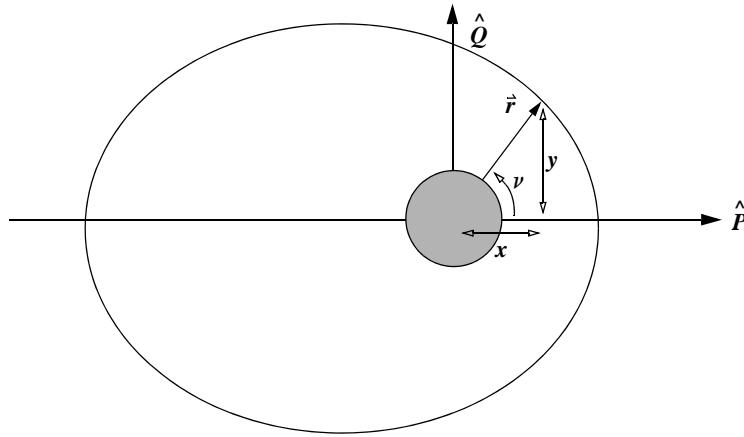


Figure 2-14. Geometry for the Perifocal Coordinate System. We can break a satellite's position into two components (x, y) in the orbital plane.

If you assume a linear combination of the initial position and velocity vectors and use arbitrary functions, f and g , you'll get*

* We can also adopt a matrix notation and write the update in Eq. (2-58) as a state vector of position and velocity vectors, and with Φ_s as a matrix containing scalar 3×3 matrices (See Sec. C.3) of the f and g values,

$$\bar{X}_{k+1} = \begin{bmatrix} f & g \\ \cdot & \cdot \\ f & g \end{bmatrix} \bar{X}_k = \Phi_s \bar{X}_k$$

The Φ_s matrix is called the **state-transition matrix** because it moves the state through time. Be aware that Φ_s will *not* accurately propagate the state errors that we'll discuss in Chapter 10—the *error* state transition matrix, Φ , does. I'll use the "s" subscript to differentiate the two. Recognize that the f and g values change over time so you cannot use a single value of Φ_s for multiple transformations. See Chap. 10 for additional discussion.

$$\begin{aligned}\hat{\vec{r}} &= f\hat{\vec{r}}_o + g\hat{\vec{v}}_o \\ \hat{\vec{v}} &= \dot{f}\hat{\vec{r}}_o + \dot{g}\hat{\vec{v}}_o\end{aligned}\tag{2-58}$$

To find the specific f and g functions, start by crossing $\hat{\vec{r}}$ into $\hat{\vec{v}}_o$ to isolate f :

$$\hat{\vec{r}} \times \hat{\vec{v}}_o = (f\hat{\vec{r}}_o + g\hat{\vec{v}}_o) \times \hat{\vec{v}}_o = f(\hat{\vec{r}}_o \times \hat{\vec{v}}_o) + g(\hat{\vec{v}}_o \times \hat{\vec{v}}_o)$$

Because the last cross product is zero and the first has a value only in the \hat{W} direction, you can consider the term multiplying f as a magnitude. Also, recognize that this term $(\hat{\vec{r}}_o \times \hat{\vec{v}}_o)$ is the angular-momentum. Therefore,

$$x\dot{y}_o - \dot{x}_o y = fh$$

Solving results in

$$f = \frac{x\dot{y}_o - \dot{x}_o y}{h}\tag{2-59}$$

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,

$$\dot{f} = \frac{\dot{x}\dot{y}_o - \dot{x}_o \dot{y}}{h}\tag{2-60}$$

Next, to find the g expression, expand $\hat{\vec{r}}_o \times \hat{\vec{r}}$:

$$\begin{aligned}\hat{\vec{r}}_o \times \hat{\vec{r}} &= \hat{\vec{r}}_o \times (f\hat{\vec{r}}_o + g\hat{\vec{v}}_o) \\ &= f(\hat{\vec{r}}_o \times \hat{\vec{r}}_o) + g(\hat{\vec{r}}_o \times \hat{\vec{v}}_o)\end{aligned}$$

Examine only the component in the \hat{W} direction (the first term is zero):

$$x_o y - x y_o = gh$$

Solving yields

$$g = \frac{x_o y - x y_o}{h}\tag{2-61}$$

Differentiating gives the dot term (x_o and y_o are initial constants):

$$\dot{g} = \frac{x_o \dot{y} - \dot{x} y_o}{h}\tag{2-62}$$

Eq. (2-58) 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:

$$\begin{aligned}
\dot{\hat{h}} &= \dot{\hat{r}} \times \dot{\hat{v}} = (f\hat{r}_o + g\hat{v}_o) \times (\dot{f}\hat{r}_o + \dot{g}\hat{v}_o) \\
&= \dot{f}f\hat{r}_o \times \dot{\hat{r}}_o + f\dot{g}\hat{r}_o \times \dot{\hat{v}}_o + \dot{f}g\hat{v}_o \times \dot{\hat{r}}_o + g\dot{g}\hat{v}_o \times \dot{\hat{v}}_o \\
&= f\dot{g}\hat{h} - \dot{f}g\hat{h}
\end{aligned}$$

This can only be true (for $\hat{h} \neq 0$) if

$$1 = f\dot{g} - \dot{f}g \quad (2-63)$$

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 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. (1-18) and Eq. (1-19), 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. (2-59) to Eq. (2-62). 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

$$\begin{aligned}
f &= 1 - \left(\frac{r}{p}\right)(1 - \cos(\Delta\nu)) & g &= \frac{rr_o\sin(\Delta\nu)}{\sqrt{\mu p}} \\
\dot{f} &= \sqrt{\frac{\mu}{p}}\tan\left(\frac{\Delta\nu}{2}\right)\left(\frac{1 - \cos(\Delta\nu)}{p} - \frac{1}{r} - \frac{1}{r_o}\right) & \dot{g} &= 1 - \left(\frac{r_o}{p}\right)(1 - \cos(\Delta\nu))
\end{aligned} \quad (2-64)$$

If the eccentric anomaly, E , is known, use relations developed in the solution to Kepler's equation [Eq. (2-8)] in Sec. 2.2 [see also Battin (1987:162)]:

$$\begin{aligned}
x &= a(\cos(E) - e) & y &= a\sin(E)\sqrt{1 - e^2} & \dot{E} &= \frac{1}{r}\sqrt{\frac{\mu}{a}} \\
\dot{x} &= -a\dot{E}\sin(E) & \dot{y} &= a\dot{E}\cos(E)\sqrt{1 - e^2}
\end{aligned}$$

To find the overall expressions for the eccentric anomaly, use the same procedure we used for the true anomaly.

$$\begin{aligned} f &= 1 - \frac{a}{r_o}(1 - \cos(\Delta E)) & g &= (t - t_o) - \sqrt{\frac{a^3}{\mu}}(\Delta E - \sin(\Delta E)) \\ \dot{f} &= \frac{-\sin(\Delta E)\sqrt{\mu a}}{r_o r} & \dot{g} &= 1 - \frac{a}{r}(1 - \cos(\Delta E)) \end{aligned} \quad (2-65)$$

Assume the parabolic anomaly, B , is known. Then, by Eqs. (2-24), (2-26), and (2-27), 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

$$\begin{aligned} f &= \frac{1 - B^2 + 2BB_o}{1 + B_o^2} & g &= \frac{p^2 \Delta B(1 + BB_o)}{2h} \\ \dot{f} &= \frac{4h \Delta B}{p^2(1 + B^2)(1 + B_o^2)} & \dot{g} &= \frac{1 - B_o^2 + 2BB_o}{1 + B^2} \end{aligned} \quad (2-66)$$

If the hyperbolic anomaly, H , is known, you can use Eq. (2-30), Eq. (2-33), and Eq. (2-38) [see also Battin (1987:170)] to write

$$\begin{aligned} x &= a(\cosh(H) - e) & y &= -a \sinh(H) \sqrt{e^2 - 1} & \dot{H} &= -\frac{1}{r} \sqrt{\frac{\mu}{-a}} \\ \dot{x} &= \frac{-\sqrt{-\mu a}}{r} \sinh(H) & \dot{y} &= \frac{\sqrt{\mu p}}{r} \cosh(H) \end{aligned}$$

The resulting functions are

$$\begin{aligned} f &= 1 - \frac{a}{r_o}(1 - \cosh(\Delta H)) & g &= (t - t_o) - \sqrt{\frac{(-a)^3}{\mu}}(\sinh(\Delta H) - \Delta H) \\ \dot{f} &= \frac{-\sinh(\Delta H)\sqrt{-\mu a}}{r_o r} & \dot{g} &= 1 - \frac{a}{r}(1 - \cosh(\Delta H)) \end{aligned} \quad (2-67)$$

Finally, if universal variables χ and ψ are known, you have (Bate, Mueller, and White, 1971:200–201) (c_{int} is a constant of integration)

$$\begin{aligned}x &= -ae - a \sin\left(\frac{\chi + c_{int}}{\sqrt{a}}\right) & y &= a\sqrt{1-e^2} \cos\left(\frac{\chi + c_{int}}{\sqrt{a}}\right) \\ \dot{x} &= -\frac{\sqrt{\mu a}}{r} \cos\left(\frac{\chi + c_{int}}{\sqrt{a}}\right) & \dot{y} &= -\frac{h}{r} \sin\left(\frac{\chi + c_{int}}{\sqrt{a}}\right)\end{aligned}$$

Substituting to determine the f and g functions yields (see Battin, 1987:179–180)

$$\begin{aligned}f &= 1 - \frac{\chi^2}{r_o^2} c_2 \\ g &= (t - t_o) - \frac{\chi^3}{\sqrt{\mu}} c_3 \\ \dot{f} &= \frac{\sqrt{\mu}}{r_o r} \chi (\psi c_3 - 1) \\ \dot{g} &= 1 - \frac{\chi^2}{r} c_2\end{aligned}\tag{2-68}$$

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. 7, 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 $t = 0$.

$$\ddot{\vec{r}}(\tau) = \ddot{\vec{r}}(0) + \dot{\vec{r}}(0)\tau + \frac{\ddot{\vec{r}}(0)\tau^2}{2!} + \frac{\ddot{\vec{r}}(0)\tau^3}{3!} + \dots$$

Notice this requires additional derivatives, evaluated at $t = 0$.

$$\begin{aligned}\dot{\vec{r}} &= \vec{v} \\ \ddot{\vec{r}} &= -\frac{\mu}{r^3} \vec{r} = -u \vec{r} \\ \ddot{\vec{r}} &= -\dot{u} \vec{r} - u \dot{\vec{r}}\end{aligned}$$

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 posi-

tion 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:

$$\begin{aligned} f = & 1 - \frac{u}{2}\tau^2 - \frac{\dot{u}}{6}\tau^3 - \frac{\ddot{u} - u^2}{24}\tau^4 - \frac{\ddot{u} - 4u\dot{u}}{120}\tau^5 - \frac{\overset{\text{iv}}{u} - 7u\ddot{u} + u^3 - 4u^2}{720}\tau^6 \\ & - \frac{\overset{\text{v}}{u} - 15\dot{u}\ddot{u} - 11u\ddot{u} + 9u^2\dot{u}}{5040}\tau^7 \\ & - \frac{\overset{\text{vi}}{u} - 15\ddot{u}^2 - 26\dot{u}\ddot{u} - 16u\overset{\text{iv}}{u} + 28u\dot{u}^2 + 22u^2\ddot{u} - u^4}{40,320}\tau^8 \quad (2-69) \\ g = & \tau - \frac{u}{6}\tau^3 - \frac{\dot{u}}{12}\tau^4 - \frac{3\ddot{u} - u^2}{120}\tau^5 - \frac{2\ddot{u} - 3u\dot{u}}{360}\tau^6 \\ & - \frac{5\overset{\text{iv}}{u} - 13u\ddot{u} - 10\dot{u}^2 + u^3}{5040}\tau^7 - \frac{6\overset{\text{v}}{u} - 48u\dot{u} - 24u\ddot{u} + 12u^2\dot{u}}{40,320}\tau^8 \end{aligned}$$

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. (2-58) 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 between orbit types make computation difficult. Although this section is rather long, I've included the steps so you'll 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 9,567,205.5 km, the solution (using the elliptical formula) is

$$\begin{aligned} E &= \cos^{-1} \left(\frac{e + \cos(\nu)}{1 + e \cos(\nu)} \right) = \cos^{-1} \left(\frac{0.999 + \cos 60^\circ}{1 + 0.999 \cos 60^\circ} \right) = 1.479\,658\,4^\circ \\ t - t_o &= \frac{1}{n} \{E - e \sin(E)\} \\ &= \sqrt{\frac{9567205.5^3}{398600.4418}} (0.0258\,249\,11 - 0.999 \sin(1.479\,658\,4)) = 22.414\text{m} \end{aligned}$$

This answer is actually reasonable because an elliptical orbit with a radius of perigee of 9567.2 km [9,567,205.5 (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 (9567.2 km), the semiparameter is 19134.4 km for the parabolic orbit, so

$$B = \tan\left(\frac{\nu}{2}\right) = \tan\left(\frac{60^\circ}{2}\right) = 0.577\ 350\ 27$$

$$(t - T) = \frac{1}{2} \sqrt{\frac{p^3}{\mu}} \left(B + \frac{B^3}{3} \right) = \frac{1}{2} \sqrt{\frac{19134.4^3}{398600.4418}} \left(0.577\ 35 + \frac{0.577\ 35^3}{3} \right) = 22.411^m$$

This answer is slightly more accurate and proves to be correct when verified through numerical techniques.

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 can 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— involving the various anomalies, E , B , and H —don't work very well when the orbits are nearly parabolic. Using the universal variable overcomes this problem.

The first equation in Eq. (2-50) 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. (2-50). From Eq. (2-40), 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. (2-50). Differentiating with respect to χ gives $f'(\chi) = -r$. Then, using Eq. (2-51), we find subsequent values of χ are

$$\chi_{n+1} = \chi_n - \frac{\sqrt{\mu} \Delta t - \chi_n^3 c_3 - \frac{\dot{r}_o \cdot \dot{\vec{v}}_o}{\sqrt{\mu}} \chi_n^2 c_2 - r_o \chi_n (1 - \psi c_3)}{-r}$$

I've replaced $\int r d\chi$ with the right-hand side of Eq. (2-50). You can also verify this relation by differentiating Eq. (2-50) with respect to χ and using the derivatives

$$\begin{aligned} \frac{dc_2}{d\psi} &= \frac{1}{2\psi} \{1 - \psi c_3 - 2c_2\} \\ \frac{dc_3}{d\psi} &= \frac{1}{2\psi} \{c_2 - 3c_3\} \end{aligned} \tag{2-70}$$

The result is r in Eq. (2-50), as shown in Eq. (2-40). The final update equation is

$$\chi_{n+1} = \chi_n + \frac{\sqrt{\mu} \Delta t - \chi_n^3 c_3 - \frac{\dot{r}_o \cdot \dot{\vec{v}}_o}{\sqrt{\mu}} \chi_n^2 c_2 - r_o \chi_n (1 - \psi c_3)}{r}$$

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. 2-15. 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. 2-15 suggests the Newton-Raphson iteration will perform well on all orbit types.

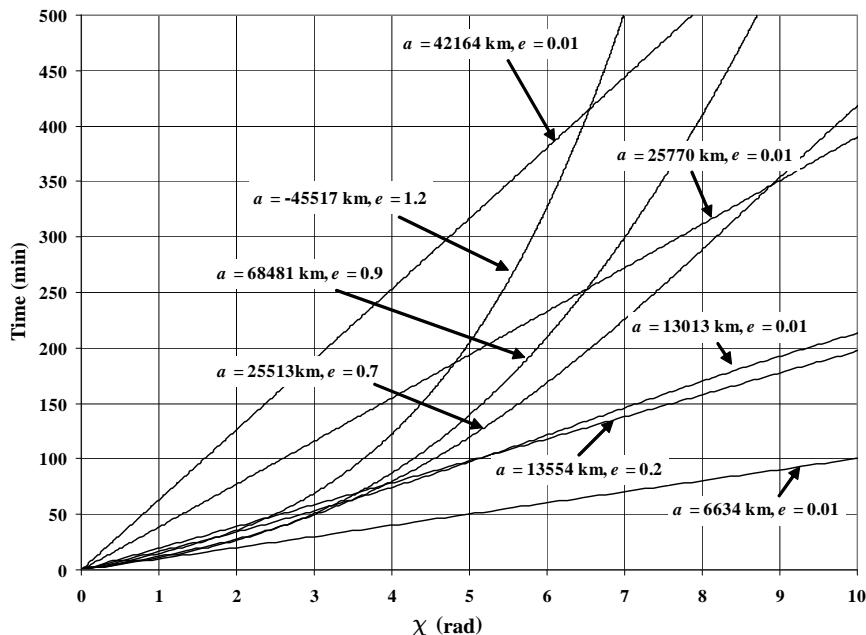


Figure 2-15. Plot of Time versus χ . Several different orbits are shown with their equivalent values of χ . Because the orbit type is always known in advance, the iteration will not have discontinuities. Notice that elliptical orbits ($e > 0.01$) have a slight curve compared to circular orbits, and it becomes more pronounced as the eccentricity increases. This result is analogous to the elliptical results in Figs. 2-6 and 2-7.

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 $\chi_o = 0$:

$$E - E_o = \frac{\pi}{2} + \frac{\chi + c_o}{\sqrt{a}} - \left\{ \frac{\pi}{2} + \frac{c_o}{\sqrt{a}} \right\} = \frac{\chi}{\sqrt{a}}$$

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. (2-50):

$$r = \chi^2 c_2 + \frac{\dot{\vec{r}}_o \cdot \dot{\vec{v}}_o}{\sqrt{\mu}} \chi (1 - \psi c_3) + r_o (1 - \psi c_2)$$

Recall the parabola has special parameters: $a = \infty$, so $\psi = 0$, $c_2 = 1/2$, and $c_3 = 1/6$. Thus,

$$r = \frac{\chi^2}{2} + \frac{\dot{\vec{r}}_o \cdot \dot{\vec{v}}_o}{\sqrt{\mu}} \chi + r_o$$

Use interim results from the **COE2RV** derivation and the trajectory equation to find an expression for the dot product:

$$\dot{\vec{r}}_o \cdot \dot{\vec{v}}_o = r \dot{r} = \frac{p}{1 + \cos(\nu)} \sqrt{\frac{\mu}{p}} \sin(\nu) = \sqrt{\mu p} \tan\left(\frac{\nu}{2}\right) = \sqrt{\mu p} B$$

Now, recall Eq. (2-24) and substitute into the universal-variable relation for the position above, being careful in identifying initial and final time values:

$$\frac{p + pB^2}{2} = \frac{\chi^2}{2} + \frac{\sqrt{\mu p} B_o}{\sqrt{\mu}} \chi + \frac{p + pB_o^2}{2}$$

$$\frac{\chi^2}{2} + \sqrt{\mu p} B_o \chi + \frac{p + pB_o^2}{2} - \frac{p + pB^2}{2} = \frac{\chi^2}{2} + \sqrt{\mu p} B_o \chi + \frac{p(B_o^2 - B^2)}{2} = 0$$

Because this equation is quadratic in χ , solve and find

$$\chi = \frac{-\sqrt{p}B_o \pm \sqrt{pB_o^2 - 2\left\{\frac{p(B_o^2 - B^2)}{2}\right\}}}{2\left(\frac{1}{2}\right)}$$

$$\chi = -\sqrt{p}B_o \pm \sqrt{pB_o^2 - pB_o^2 + pB^2} = -\sqrt{p}B_o \pm \sqrt{pB^2}$$

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. (1-26):

$$\frac{\chi}{2\pi\sqrt{a}} = \frac{\Delta t}{\varphi}$$

This relation is an equality for circular orbits. We can see this in Fig. 2-15 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

$$\chi_o \approx \sqrt{\mu}(\Delta t)\alpha \quad (2-71)$$

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

$$\chi_o \approx \sqrt{p} \tan\left(\frac{v_o}{2}\right) \quad (2-72)$$

Selecting the hyperbolic initial guess is a little more complicated, but Bate, Mueller, and White (1971:206–207) provide the template. By Eq. (2-48), 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. 2-5, 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. (2-50), 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

$$t - t_o \approx -\frac{\dot{\vec{r}}_o \cdot \dot{\vec{v}}_o}{2\mu} ae^{\sqrt{-\psi}} - \text{SIGN}(t - t_o) \left\{ 1 - \frac{r_o}{a} \right\} \frac{a\sqrt{-a}}{2\sqrt{\mu}} e^{\sqrt{-\psi}}$$

Solving for $e^{\sqrt{-\psi}}$ results in

$$e^{\sqrt{-\psi}} \approx \frac{-2\mu(t - t_o)}{a(\dot{\vec{r}}_o \cdot \dot{\vec{v}}_o) + a\text{SIGN}(t - t_o)\sqrt{-\mu a}\left(1 - \frac{r_o}{a}\right)}$$

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,

$$\chi_o \approx \text{SIGN}(t - t_o) \sqrt{-a} \ln \left\{ \frac{-2\mu\alpha(t - t_o)}{\dot{\vec{r}}_o \cdot \dot{\vec{v}}_o + \text{SIGN}(t - t_o)\sqrt{-\mu a}(1 - r_o\alpha)} \right\} \quad (2-73)$$

Implementing Kepler's Problem

Algorithm 8 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 simplify some 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 can 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. The formula for α is valid for all orbits. Finally, for parabolic orbits, the initial guess requires we find the semiparameter as h^2/μ .

ALGORITHM 8: *KEPLER* ($\hat{r}_o, \hat{v}_o, \Delta t \Rightarrow \hat{r}, \hat{v}$)

$$\xi = \frac{v_o^2}{2} - \frac{\mu}{r_o}$$

$$\alpha = \frac{-v_o^2}{\mu} + \frac{2}{r_o}$$

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 ($|\alpha| < 0.000\ 001$):

$$\begin{cases} \hat{h} = \hat{r}_o \times \hat{v}_o \\ p = \frac{h^2}{\mu} \\ \text{COT}(2s) = 3 \sqrt{\frac{\mu}{p^3}} (\Delta t) \\ \text{TAN}^3(w) = \text{TAN}(s) \end{cases} \quad \chi_o \approx \sqrt{p} 2 \text{COT}(2w)$$

Hyperbola ($\alpha < -0.000\ 001$):

$$a = \frac{1}{\alpha}$$

$$\chi_o \approx \text{SIGN}(\Delta t) \sqrt{-a} \text{LN} \left\{ \frac{-2\mu\alpha(\Delta t)}{\hat{r}_o \cdot \hat{v}_o + \text{SIGN}(\Delta t) \sqrt{-\mu a} (1 - r_o \alpha)} \right\}$$

LOOP

$$\psi = \chi_n^2 \alpha$$

Find $c_2 c_3(\psi \Rightarrow c_2, c_3)$

$$r = \chi_n^2 c_2 + \frac{\dot{\vec{r}}_o \cdot \vec{v}_o}{\sqrt{\mu}} \chi_n (1 - \psi c_3) + r_o (1 - \psi c_2)$$

$$\chi_{n+1} = \chi_n + \frac{\sqrt{\mu} \Delta t - \chi_n^3 c_3 - \frac{\dot{\vec{r}}_o \cdot \vec{v}_o}{\sqrt{\mu}} \chi_n^2 c_2 - r_o \chi_n (1 - \psi c_3)}{r}$$

$$\chi_n \Leftarrow \chi_{n+1}$$

$$\text{UNTIL } |\chi_n - \chi_{n-1}| < 1 \times 10^{-6}$$

$$f = 1 - \frac{\chi_n^2}{r_o} c_2 \quad g = \Delta t - \frac{\chi_n^3}{\sqrt{\mu}} c_3$$

$$\dot{g} = 1 - \frac{\chi_n^2}{r} c_2 \quad \dot{f} = \frac{\sqrt{\mu}}{r r_o} \chi_n (\psi c_3 - 1)$$

$$\dot{\vec{r}} = f \dot{\vec{r}}_o + g \dot{\vec{v}}_o$$

$$\dot{\vec{v}} = \dot{f} \dot{\vec{r}}_o + \dot{g} \dot{\vec{v}}_o$$

Check $\dot{f}g - \dot{g}f = 1$ for success.

An example will demonstrate Kepler's problem.

▼ Example 2-4. Solving Kepler's Problem.

GIVEN: $\hat{\vec{r}}_{IJK} = 1131.340 \hat{I} - 2282.343 \hat{J} + 6672.423 \hat{K}$ km

$$\hat{\vec{v}}_{IJK} = -5.643 \ 05 \hat{I} + 4.303 \ 33 \hat{J} + 2.428 \ 79 \hat{K}$$
 km/s

$$\Delta t = 40 \text{ min}$$

FIND: Position and velocity vectors at the future time

Use part of Algorithm 9 to find a .

$$\xi = \frac{v_o^2}{2} - \frac{\mu}{r_o} = -27.678 \ 777 \text{ km}^2/\text{s}^2$$

$$a = -\frac{\mu}{2\xi} = 7200.4706 \text{ km} \quad \alpha = \frac{1}{a} = \frac{-v_o^2}{\mu} + \frac{2}{r_o} = 0.000 \ 138 \ 879 \text{ /km}$$

The orbit is elliptical, ($\alpha > 0.000 \ 001$), so approximate the first guess as

$$\chi_o \approx \sqrt{\mu}(\Delta t) \alpha = 210.435 \ 618 \text{ km}$$

Iterate to get the values shown in Table 2-4:

TABLE 2-4. Values for Kepler's Problem in Example 2-4. Iteration values are shown. Notice how quickly we arrive at the correct answer.

Iteration	$\chi_n \sqrt{\text{km}}$	ψ	$r \text{ km}$	$\chi_{n+1} \sqrt{\text{km}}$
1	210.453 618	6.150 035	7246.4895	210.855 173
2	210.855 173	6.174 583	7246.6661	210.855 167

$$f = 1 - \frac{\chi_n^2}{r_o} c_2 = -0.806 632 \quad g = \Delta t - \frac{\chi_n^3}{\sqrt{\mu}} c_3 = 586.061 95 \text{ s}$$

$$\dot{g} = 1 - \frac{\chi_n^2}{r} c_2 = -0.780 574 \quad \dot{f} = \frac{\sqrt{\mu}}{rr_o} \chi_n (\psi c_3 - 1) = -0.000 631 95 \text{ /s}$$

$$\hat{\vec{r}}_{IJK} = f\hat{\vec{r}}_o + g\hat{\vec{v}}_o = -4219.7527 \hat{I} + 4363.0292 \hat{J} - 3958.7666 \hat{K} \text{ km}$$

$$\hat{\vec{v}}_{IJK} = \dot{f}\hat{\vec{r}}_o + \dot{g}\hat{\vec{v}}_o = 3.689 866 \hat{I} - 1.916 735 \hat{J} - 6.112 511 \hat{K} \text{ km/s}$$



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. 10.

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. 9.2. Several other element sets have been developed for convenience or to avoid the difficulties the classical orbital elements suffer for certain orbital geometries: two-line, equinoctial, Delaunay, and Poincaré.

2.4.1 Classical Orbital Elements (Keplerian)

Figure 2-16 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. (1-22)], but we can also use the extreme points of the orbit—the radii of apoapsis and periapsis.

$$a = \left(\frac{2}{r} - \frac{v^2}{\mu} \right)^{-1} = \frac{r_a + r_p}{2} \quad (2-74)$$

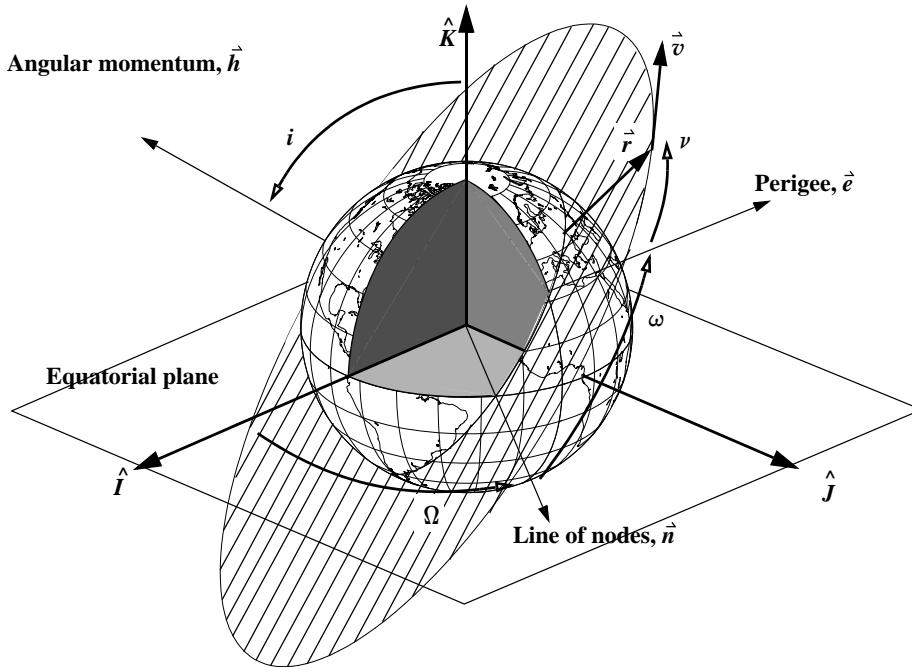


Figure 2-16. Classical Orbital Elements. The six classical orbital elements are the *semimajor axis*, a ; *eccentricity*, e ; *inclination*, i ; *right ascension of the ascending node*, Ω —often referred to as simply the *node*; *argument of perigee*, ω ; and *true anomaly*, ν . I haven't shown scale 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. (2-74) is indeterminate. For this reason, the *semiparameter* is sometimes used as the first orbital element. 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. (1-9) and Eq. (1-10).

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

The most general form of the semiparameter is Eq. (1-19) 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. (1-24)] to yield

$$\begin{aligned} \nu = 0^\circ \quad r_p &= \frac{p}{1 + e \cos(\nu)} = \frac{p}{1 + e} \\ \nu = 180^\circ \quad r_a &= \frac{p}{1 + e \cos(\nu)} = \frac{p}{1 - e} \end{aligned}$$

For *most* orbits, we can simplify to get useful results:

$$\begin{aligned} r_p &= \frac{a(1-e^2)}{1+e} = a(1-e) \\ r_a &= \frac{a(1-e^2)}{1-e} = a(1+e) \end{aligned} \tag{2-75}$$

Many of the preceding formulas are indeterminate for a parabola because the semi-major axis is infinite and $e = 1$. Obviously, a parabola has a value for the semiparameter and radius of perigee. From Eq. (1-19), we can find the semiparameter. If we substitute the result into the trajectory equation [Eq. (1-24)], the perigee radius may also determine the semiparameter.

$$p_{\text{parabola}} = 2r_p \tag{2-76}$$

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 [Eq. (1-27)] as

$$n = \sqrt{\frac{\mu}{a^3}} \tag{2-77}$$

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. (1-23) 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,

$$\hat{e} = \frac{\hat{v} \times (\hat{r} \times \hat{v})}{\mu} - \frac{\hat{r}}{r}$$

Then, expand the vector triple cross product to get

$$\hat{\vec{e}} = \frac{(\hat{\vec{v}} \cdot \hat{\vec{v}})\hat{\vec{r}} - (\hat{\vec{r}} \cdot \hat{\vec{v}})\hat{\vec{v}}}{\mu} - \frac{\hat{\vec{r}}}{r}$$

Simplify the final equation:

$$\hat{\vec{e}} = \frac{v^2 \hat{\vec{r}} - (\hat{\vec{r}} \cdot \hat{\vec{v}})\hat{\vec{v}}}{\mu} - \frac{\hat{\vec{r}}}{r}$$

Combine to get

$$\hat{\vec{e}} = \frac{\left(v^2 - \frac{\mu}{r}\right)\hat{\vec{r}} - (\hat{\vec{r}} \cdot \hat{\vec{v}})\hat{\vec{v}}}{\mu} \quad (2-78)$$

The eccentricity vector always points to periapsis. We can show $\hat{\vec{e}}$ points to perigee mathematically by substituting position and velocity vectors in the PQW system into Eq. (2-78) 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. (2-78) 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 find a more general formula using the semiparameter equation and solving it for the eccentricity:

$$e = \sqrt{1 - \frac{p}{a}} \quad (2-79)$$

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:

$$e = \sqrt{1 + \frac{2\xi h^2}{\mu^2}} \quad (2-80)$$

Solving Eq. (2-75) for eccentricity yields useful results.

$$e = \frac{a - r_p}{a} = \frac{r_a - a}{a} \quad (2-81)$$

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 circular 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 options 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,

$$\cos(i) = \frac{\hat{K} \cdot \hat{h}}{|\hat{K}| |\hat{h}|} \quad (2-82)$$

Because the arccosine function will always return an angle between 0° and 180° , no further checks are necessary for computer applications.

The ***right ascension of the ascending node***, Ω , RAAN, or simply the ***node***,* is the angle in the equatorial plane measured positive eastward from the 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

$$\hat{n} = \hat{K} \times \hat{h} \quad (2-83)$$

Equatorial orbits have no nodes, and thus the right ascension 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:

$$\cos(\Omega) = \frac{\hat{I} \cdot \hat{n}}{|\hat{I}| |\hat{n}|} \quad (2-84)$$

$$\text{IF } (n_J < 0) \text{ THEN } \Omega = 360^\circ - \Omega$$

But this time we must check quadrants to ensure we have the correct answer. From Fig. 2-16, 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° .†

* The term ***longitude of the ascending node*** is sometimes used interchangeably, but remember that it is measured from Greenwich, not the vernal equinox. If the longitude is intended, it implies an Earth Fixed (ECEF) frame and the value changes continuously over time. Using Fig. 3-23 and Eq. (3-43), we can establish $\lambda_{AN} = \Omega - \theta_{GMST}$.

The **argument of perigee**, ω , measured from the ascending node in the direction of satellite motion, locates the closest point of the orbit (periapsis). Again, this quantity varies from 0° to 360° and is obtained by extending a dot product:

$$\cos(\omega) = \frac{\hat{n} \cdot \hat{e}}{|\hat{n}| |\hat{e}|} \quad (2-85)$$

IF($e_K < 0$) THEN $\omega = 360^\circ - \omega$

This angle is undefined for perfectly circular ($\hat{e} = 0$) or equatorial ($\hat{n} = 0$) orbits because periapsis and the node (respectively) don'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:

$$\cos(\nu) = \frac{\hat{e} \cdot \hat{r}}{|\hat{e}| |\hat{r}|} \quad (2-86)$$

IF($\hat{r} \cdot \hat{v} < 0$), 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 cir-

† 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 ATAN2 process.

cular 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: non-circular equatorial ($\tilde{\omega}_{true}$, which includes parabolic and hyperbolic), circular inclined ($\tilde{\omega}$), 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_M) for real-world cases which approach the theoretical circular and equatorial orbits.

A non-circular 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

$$\cos(\tilde{\omega}_{true}) = \frac{\hat{I} \cdot \hat{e}}{|\hat{I}| |\hat{e}|} \quad \text{Non-circular Equatorial} \quad (2-87)$$

IF ($e_J < 0$) 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:

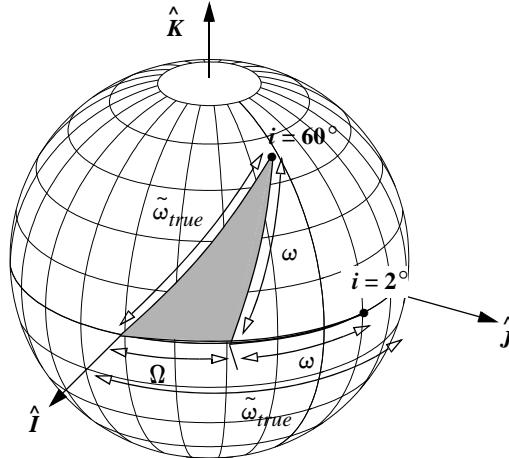
$$\tilde{\omega} = \Omega + \omega \quad (2-88)$$

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. (2-87) *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. 2-17. 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. 2-17, 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 inclina-

* I've chosen to measure $\tilde{\omega}_{true}$ positive eastward. You could also switch the sign of the e_J component test for retrograde equatorial orbits. This would also apply to λ_{true} in Eq. (2-92).

tion 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.



$$\begin{aligned} i &= 2^\circ, \quad e = 0.000\ 01 \\ \Omega &= 30^\circ, \quad \omega = 40^\circ \\ \tilde{\omega}_{true} &= 69.988\ 27^\circ \end{aligned}$$

$$\begin{aligned} i &= 60^\circ, \quad e = 0.000\ 01 \\ \Omega &= 30^\circ, \quad \omega = 40^\circ \\ \tilde{\omega}_{true} &= 59.820\ 08^\circ \end{aligned}$$

Figure 2-17. 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 in the direction of satellite motion. A moon of Neptune provides an example of this type of orbit. Triton is inclined 157.345° to the equator of Neptune and has an eccentricity of about 0.000 016 (*Astronomical Almanac*, 1992:F2). Thus,

$$\cos(u) = \frac{\hat{n} \cdot \hat{r}}{|\hat{n}| |\hat{r}|} \quad \text{Circular Inclined} \quad (2-89)$$

$$\text{IF } (r_K < 0) \text{ THEN } u = 360^\circ - u$$

Note that this equation is also valid for elliptical inclined orbits. A relation that's always valid exists between classical orbital elements and the argument of latitude:

$$u = \omega + \nu \quad (2-90)$$

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. (2-89) 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. 2.2:

$$u_M = \omega + M \quad (2-91)$$

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:

$$\cos(\lambda_{true}) = \frac{\hat{I} \cdot \hat{r}}{|\hat{I}| |\hat{r}|} \quad \text{Circular Equatorial} \quad (2-92)$$

$$\text{IF } (r_J < 0) \text{ THEN } \lambda_{true} = 360^\circ - \lambda_{true}$$

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 the vernal equinox. The expression uses the mean anomaly, M (pgs 53-54). Thus,

$$\lambda_M = \Omega + \omega + M = \tilde{\omega} + M \quad (2-93)$$

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. 1-13, notice that

$$\cos(\phi_{fpa}) = \frac{r\dot{\nu}}{v}$$

But we've already developed several alternate formulas. In addition, if we use the velocity formula for an ellipse [Eq. (1-31) and Eq. (1-18)],

$$\cos(\phi_{fpa}) = \frac{\frac{h}{r}}{\sqrt{\frac{2\mu}{r} - \frac{\mu}{a}}}$$

Because $h = \sqrt{\mu p}$, using Eq. (1-10) permits

$$\cos(\phi_{fpa}) = \sqrt{\frac{a^2(1-e^2)}{r(2a-r)}} \quad (2-94)$$

We can similarly develop an expression for the sine of the flight-path angle because, from Fig. 1-13,

$$\sin(\phi_{fpa}) = \frac{\dot{r}}{v} = \frac{\dot{r}}{\sqrt{\frac{2\mu}{r} - \frac{\mu}{a}}}$$

Now substitute \dot{r} from Eq. (1-25):

$$\sin(\phi_{fpa}) = \frac{\frac{r\dot{\nu}e \sin(\nu)}{1+e \cos(\nu)}}{\sqrt{\frac{2\mu}{r} - \frac{\mu}{a}}}$$

Notice the dependence on the true anomaly, which makes developing a relation for the sine easier. First, use Eq. (1-18) to replace the derivative of the true anomaly in the numerator and find

$$\sin(\phi_{fpa}) = \frac{\frac{h}{r} \left(\frac{e \sin(\nu)}{1+e \cos(\nu)} \right)}{\sqrt{\frac{2a-r}{\mu \frac{ra}{ra}}}}$$

Substitute definitions for the trajectory equation, Eq. (1-24) and Eq. (1-19):

$$\sin(\phi_{fpa}) = e \sin(\nu) \sqrt{\left(\frac{1}{p}\right) \left(\frac{ra}{2a-r}\right)}$$

Next, substitute the definition of the semiparameter, Eq. (1-10):

$$\sin(\phi_{fpa}) = e \sin(\nu) \sqrt{\frac{r}{(1-e^2)(2a-r)}}$$

Substitute the trajectory equation again and simplify to get the final result:

$$\sin(\phi_{fpa}) = \frac{e \sin(\nu)}{\sqrt{1 + 2e \cos(\nu) + e^2}} \quad (2-95)$$

To determine the cosine expression in terms of true anomaly, use Eq. (2-95) and Eq. (C-10). 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

$$\cos(\phi_{fpa}) = \frac{1 + e \cos(\nu)}{\sqrt{1 + 2e \cos(\nu) + e^2}} \quad (2-96)$$

We use both sine and cosine expressions to aid computer programs and employ *ATAN2* to help find the correct quadrant. The expressions in Eq. (2-95) and Eq. (2-96) permit a graphical analysis of the flight-path angle. Figure 2-18 shows various values for the flight-path angle. Notice the relatively small values for orbits with an eccentricity of 0.5 or less.

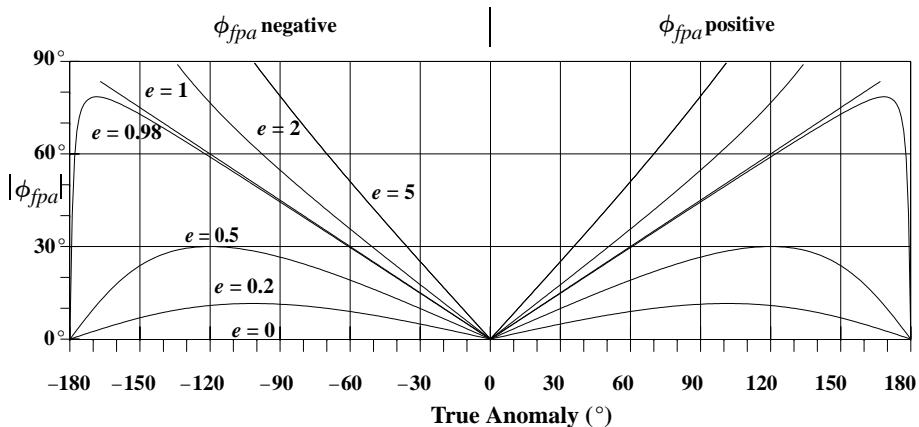


Figure 2-18. Values for the Flight-path Angle. Notice how the behavior of the flight-path angle changes dramatically as the eccentricity increases. I've shown the absolute value for clarity. For parabolic and hyperbolic orbits, the range of true anomaly is limited. See Sec. 2.2.3 for details.

2.4.2 Two-line Element Sets

The classical orbital elements are widely used throughout the scientific community, but they don't apply universally. For instance, the U.S. military uses a *Two-line element set* (TLE), which is similar but not identical to the classical elements. Because these elsets are available to the general public through Air Force Space Command (AFSPC), we'll

describe their formation. Other sources are shown in Appendix D. 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 a TLE are shown in Eq. (2-97). 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. 9.7.1. The mean anomaly, M , is used instead of the true anomaly. The time is based on UTC and the coordinate system should be considered a true-equator, mean-equinox system. This is actually an approximate system because it lacks the formalism we use in Chap. 3.

$$\begin{aligned} \bar{n} &= \sqrt{\frac{\mu}{\bar{a}^3}} & e & i & \Omega & \omega & M \\ \dot{\bar{n}} &= \frac{\ddot{\bar{n}}}{2} & B^* & = \frac{1}{2} \frac{c_D A}{m} \rho_o R_{\oplus} & & & UTC \end{aligned} \quad (2-97)$$

Here, ρ_o is the atmospheric density at perigee of the orbit (assumed to be 2.461×10^{-5} kg/m²/ER with ER = 6375.135 km), 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. 9. In fact, all the values are means. To recover the true *ballistic coefficient*, BC = $\frac{m}{c_D A}$, from the B^* term, in the TLE, see Hoots and Roehrich (1980).

$$\frac{R_{\oplus} \rho_o}{2B^*} = BC$$

With an Earth radius of 6378.135 km defined in Hoots and Roehrich (1980), a constant conversion is[†]

$$BC = \frac{1}{12.741 \cdot 621 B^*} \frac{\text{kg}}{\text{m}^2} \quad (2-98)$$

Let's examine the format in which these elements arrive—it's rather cryptic. The following is an example of a TLE. The two numbered lines are found commonly on the Internet.

Dataset for the TLE example:

* Air Force Space Command uses n internally as a Brouwer value. However, to maintain compatibility with older theories, the published TLE set converts n to a Kozai value.

† Be aware that the value of B^* is *always* modified. It's really an arbitrary free parameter in differential correction. Chapter 10 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.

```
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 the TLE 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 2-19 shows the format.

Cont#	Satellite Number	Class	International Designator	Yr	Epoch Day of Year (plus fraction)	Mean motion derivative (rev/day /2)	Mean motion second derivative (rev/day2 /6)	Bstar (/ER)	\bar{a}	Elem num	Chk Sum
	Year	Lch#	Piece		S.	S.	S.	S.	E		
1	16609	U	86017A	93	352.53502934	.00007889	00000 0	10529-3	34		
					9 3 3 5 2 1 5 3 5 0 2 9 3 4	.	0 0 0 0 7 8 8 9	0 0 0 0 0 0 0	1 0 5 2 9 -3 0	1 3 4 2	
					Inclination (deg)	Right Ascension of the Node (deg)	Eccentricity	Arg of Perigee (deg)	Mean Anomaly (deg)	Mean Motion (rev/day)	Epoch Rev
2	16609		51.6190	13.3340	0 0 0 5 7 7 0	1 0 2 1 5 6 8 0	2 5 7 1 5 9 5 0	1 5 1 5 9 1 1 4 0 7 0 4 4 7 8 6 9			

Figure 2-19. Transmission Format for the Two-Line Element Set. This example TLE set uses data from the previous example in the text. Note the use of implied decimal points. S is the sign of the values, and E is the exponent.

By applying the translation to the dataset, we can get a usable result:

Epoch December 18, 1993, 12^h 50^m 26.5350^s UTC

$$\bar{n} = 15.591 \ 140 \ 70 \text{ rev / day} \Rightarrow$$

$$\bar{a} = 6768.3568 \text{ km}$$

$$\frac{\dot{n}}{2} = 7.889 \times 10^{-5} \frac{\text{rev}}{\text{day}^2} \quad \frac{\ddot{n}}{6} = 0.0 \frac{\text{rev}}{\text{day}^3}$$

$$B^* = 0.000 \ 105 \ 29 / \text{ER}$$

$$e = 0.000 \ 577 \ 0 \quad M = 257.5950^\circ$$

$$i = 51.6190^\circ \quad \Omega = 13.3340^\circ \quad \omega = 102.5680^\circ$$

We find the estimated ballistic coefficient using Eq. (2-98) and $BC = 745.398 \text{ kg/m}^2$.* Notice also that the B^* term uses canonical units.

Many people often incorrectly assume these element sets have no errors. In fact, all element sets contain errors based on their mathematical formation. Figure 9-15 suggests

* You may notice that the satellite is always near the equator at the epoch time. That's because the epoch is chosen to coincide with the ascending node *prior* to the last observation used in the differential correction (Chap. 10). Most operations place the epoch at the last observation time. Additional notes concerning the accuracy and use of TLE's are in Sec 10.10.

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 TLE 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 only eight decimal places given in the epoch time value, it 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, or 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 Other Element Sets

Equinoctial Elements

Equinoctial elements are often 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, ω . 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 all direct orbits, and -1 for nearly retrograde orbits. Many programs ignore f_r because retrograde equatorial orbits are virtually non-existent.

$$\begin{aligned}
 a_f &= k_e = e \cos(\omega + f_r \Omega) \\
 n & \\
 a_g &= h_e = e \sin(\omega + f_r \Omega) \\
 L &= \lambda_M = M + \omega + f_r \Omega \\
 \chi &= p_e = \tan^{f_r} \left(\frac{i}{2} \right) \sin(\Omega) = \frac{\sin(i) \sin(\Omega)}{1 + \cos^{f_r}(i)} \\
 \psi &= q_e = \tan^{f_r} \left(\frac{i}{2} \right) \cos(\Omega) = \frac{\sin(i) \cos(\Omega)}{1 + \cos^{f_r}(i)}
 \end{aligned} \tag{2-99}$$

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. The mean motion is often used in place of a and the tangent half-angle formula can be replaced by the sine and cosine.

Flight Elements

Because there is some variability in the literature over what constitutes flight parameters, I'll introduce the elements for some common groupings under the generic flight elements name. Remember that the selection of elements must represent the complete state of the satellite. **Flight elements** consist of the right ascension (α) and declination (δ), flight-path angle from the local horizontal (ϕ_{fpa}), azimuth (β), and the position and velocity magnitudes. The flight path angle is sometimes taken from local vertical (ϕ_{fpav})—be sure of which convention is assumed for an operation. Chobotov (1996:30) uses the term **ADBDRV Elements** for this set (the name is taken using an alternate nomenclature for the variables).

$$\begin{aligned} \sin(\delta) &= \frac{r_K}{r} & \cos(\delta) &= \frac{\sqrt{r_I^2 + r_J^2}}{r} \\ \sin(\alpha) &= \frac{r_J}{\sqrt{r_I^2 + r_J^2}} & \cos(\alpha) &= \frac{r_I}{\sqrt{r_I^2 + r_J^2}} \\ r &= \sqrt{r_I^2 + r_J^2 + r_K^2} \\ v &= \sqrt{v_I^2 + v_J^2 + v_K^2} \end{aligned} \quad (2-100)$$

Note that the azimuth and flight path angle are found using the velocity vector, transformed to the SEZ system (See Sec. 4.4.3). Some sources define different planes from which to determine these angles. The selection often is based on subsequent partial derivatives that are taken for various operations.

$$\begin{aligned} \begin{bmatrix} \text{SEZ} \\ \text{ECI} \end{bmatrix} &= \text{ROT2}(90^\circ - \delta) \text{ROT3}(\alpha) = \begin{bmatrix} \cos \alpha \cos \delta & \sin \alpha \cos \delta & -\sin \delta \\ -\sin \alpha & \cos \alpha & 0 \\ \sin \delta \cos \alpha & \sin \alpha \sin \delta & \cos \delta \end{bmatrix} \\ \dot{\vec{v}}_{\text{SEZ}} &= \begin{bmatrix} \text{SEZ} \\ \text{ECI} \end{bmatrix} \dot{v} \\ \sin(\phi_{fpa}) &= \frac{\sqrt{v_S^2 + v_E^2}}{v} & \cos(\phi_{fpa}) &= \frac{v_Z}{v} \\ \sin(\beta) &= \frac{v_E}{\sqrt{v_S^2 + v_E^2}} & \cos(\beta) &= \frac{-v_S}{\sqrt{v_S^2 + v_E^2}} \end{aligned}$$

The reverse process requires first finding the ECEF position vector [Eq. (3-14)]. Remember to convert the vectors back to inertial (ECI) if needed.

$$\overset{\curvearrowright}{r}_{ECEF}, yr, mo, day, UTC, \Delta UT1, \Delta AT, x_p, y_p \Rightarrow \overset{\curvearrowright}{r}_{ECI}$$

To find the velocity vector, you must modify the velocity equation in inertial coordinates [Eq. (4-2)]. Using the flight-path angle from the vertical (ϕ_{fpav}) and the azimuth (β), the velocity $\overset{\curvearrowright}{v}_{ECI} =$

$$\begin{bmatrix} v\left(\cos(\alpha)\left(-\cos(\alpha)\cos(\beta)\sin(\phi_{fpav})\sin(\delta) + \cos(\phi_{fpav})\cos(\delta)\right) - \sin(\beta)\sin(\phi_{fpav})\sin(\alpha)\right) \\ v\left(\sin(\alpha)\left(-\cos(\alpha)\cos(\beta)\sin(\phi_{fpav})\sin(\delta) + \cos(\phi_{fpav})\cos(\delta)\right) + \sin(\beta)\sin(\phi_{fpav})\cos(\alpha)\right) \\ v\left(\cos(\alpha)\cos(\delta)\sin(\phi_{fpav}) + \cos(\phi_{fpav})\sin(\delta)\right) \end{bmatrix}$$

We can define ***Geographic elements*** which use the geocentric latitude (ϕ_{gc}) and longitude (λ) instead of right ascension and declination. Remember that when you use Earth fixed parameters (latitude and longitude), the state vectors must be in Earth-fixed (ECEF) coordinates. See O'Conner (1983:1-29) for additional information.

$$\begin{aligned} \phi_{gc} &= \text{SIN}^{-1}\left(\frac{r_K}{\sqrt{r_I^2 + r_J^2 + r_K^2}}\right) \\ \lambda &= \text{TAN}^{-1}\left(\frac{r_J}{r_I}\right) \end{aligned}$$

Geosynchronous orbits sometimes refer to the satellite longitude instead of the orbital elements. To convert from longitude (λ) to true anomaly, first find the longitude along the equator (λ_u), and then find the true anomaly. Note that λ_u should be within 0-360°, and u increases by 180° if λ_u is within 90-270°.

$$\lambda_u = \theta_{GMST} + \lambda - \Omega$$

$$u = \text{TAN}^{-1}\left(\frac{\text{TAN}(\lambda_u)}{\cos(i)}\right) \quad (2-101)$$

$$v = u - \omega$$

2.4.4 Canonical Elements

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 elements, the 6×6 matrix is purely diagonal, which simplifies the construction of perturbation solutions (Chap. 9).

Delaunay Elements

The Delaunay elements 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 (note that some sources use l, g, h, L, G, H respectively, but this can cause confusion with angular momentum, h , and other element notations)

$$\begin{aligned} M \\ \omega \\ \Omega \\ L_d &= \sqrt{\mu a} \\ h \\ H_d &= \sqrt{\mu a(1 - e^2)} \cos(i) \end{aligned} \tag{2-102}$$

Recognize H_d is really the component of angular momentum, h , projected onto the polar (K) axis. Don't confuse the Delaunay elements with quantities in the motion of the coordinate system [Eq. (3-88)]. The subscripts are used to distinguish similar elements. These formulas permit some rather elegant perturbation analysis, discussed in Chap. 9. In this book, we'll address these elements mainly with Brouwer's theory concerning treatment of orbital perturbations (Sec. 9.7.3).

Poincaré Elements

As the canonical counterpart of equinoctial elements, Poincaré elements 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 elements are (Brouwer and Clemence, 1961:540 and Plummer, 1960:153)

$$\begin{aligned} \lambda_M &= M + \omega + \Omega \\ g_p &= \sqrt{2(\sqrt{\mu a}(1 - \sqrt{1 - e^2}))} \cos(\omega + \Omega) \\ h_p &= \sqrt{2\sqrt{\mu a}(1 - e^2)(1 - \cos(i))} \cos(\Omega) \\ L_p &= \sqrt{\mu a} \\ G_p &= -\sqrt{2(\sqrt{\mu a}(1 - \sqrt{1 - e^2}))} \sin(\omega + \Omega) \\ H_p &= -\sqrt{2\sqrt{\mu a}(1 - e^2)(1 - \cos(i))} \sin(\Omega) \end{aligned} \tag{2-103}$$

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. Although we develop this section in a generic IJK coordinate system, we actually require an inertial coordinate system (see Sec. 3.7) because Earth-fixed coordinate systems modify the velocity, thereby altering the orbital elements. We'll refer to this problem as the ***RV2COE*** 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. (2-78). 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 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. (12-18), 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. (1-19).

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. (2-82) and (2-84) to (2-86). We determine whether we must evaluate special cases based on the particular needs of the program.

Implementing RV2COE

Several substitutions permit computational efficiency. Using canonical units could eliminate 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.

RV2COE 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 9: **RV2COE** ($\hat{r}_{IJK}, \hat{v}_{IJK} \Rightarrow p, a, e, i, \Omega, \omega, \nu(u, \lambda_{true}, \tilde{\omega}_{true})$)

$$\begin{aligned} \hat{h} &= \hat{r} \times \hat{v} & h &= |\hat{h}| \\ \hat{n} &= \hat{K} \times \hat{h} \\ \hat{e} &= \frac{\left(v^2 - \frac{\mu}{r}\right)\hat{r} - (\hat{r} \cdot \hat{v})\hat{v}}{\mu} & e &= |\hat{e}| \end{aligned}$$

$$\xi = \frac{v^2}{2} - \frac{\mu}{r}$$

$$\left[\begin{array}{l} \text{IF } e \neq 1.0 \text{ THEN} \\ \quad a = -\frac{\mu}{2\xi} \\ \quad p = a(1 - e^2) \\ \text{ELSE} \\ \quad p = \frac{h^2}{\mu} \text{ and } a = \infty \end{array} \right]$$

$$\cos(i) = \frac{h_K}{|\hat{h}|}$$

$$\cos(\Omega) = \frac{n_I}{|\hat{n}|} \quad \text{IF}(n_J < 0) \text{THEN } \Omega = 360^\circ - \Omega$$

$$\cos(\omega) = \frac{\hat{n} \cdot \hat{e}}{|\hat{n}| |\hat{e}|} \quad \text{IF}(e_K < 0) \text{THEN } \omega = 360^\circ - \omega$$

$$\cos(\nu) = \frac{\hat{e} \cdot \hat{r}}{|\hat{e}| |\hat{r}|} \quad \text{IF}(\hat{r} \cdot \hat{v} < 0) \text{THEN } \nu = 360^\circ - \nu$$

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:

$$\cos(\lambda_{true}) = \frac{r_I}{|\hat{r}|} \quad \text{IF}(r_J < 0) \text{THEN } \lambda_{true} = 360^\circ - \lambda_{true}$$

An example shows how to convert position and velocity vectors into orbital elements.

▼ Example 2-5. Finding the Orbital Elements with *RV2COE*.

GIVEN: $\hat{r}_{IJK} = 6524.834 \hat{I} + 6862.875 \hat{J} + 6448.296 \hat{K}$ km

$$\hat{\vec{v}}_{IJK} = 4.901\ 327 \hat{I} + 5.533\ 756 \hat{J} - 1.976\ 341 \hat{K} \text{ km/s}$$

FIND: Classical orbital elements

The magnitudes are $r = 11456.57$ km and $v = 7.651\ 888$ km/s

Begin by finding the specific angular momentum:

$$\begin{aligned}\hat{h} &= \hat{r} \times \hat{\vec{v}} = \begin{bmatrix} 6524.834 & 6862.875 & 6448.296 \\ 4.901\ 327 & 5.533\ 756 & -1.976\ 341 \end{bmatrix} \text{ km}^2/\text{s} \\ &= -49246.71 \hat{I} + 44500.5 \hat{J} + 2469.6 \hat{K}\end{aligned}$$

The magnitude of angular momentum is 66420.10 km²/s.

Find the node vector (not the mean motion) using a cross product:

$$\begin{aligned}\hat{n} &= \hat{K} \times \hat{h} = -44500.5 \hat{I} - 49246.7 \hat{J} \quad |\hat{n}| = 66374.17 \text{ km}^2/\text{s} \\ \hat{e} &= \frac{1}{\mu} \left[\left(v^2 - \frac{\mu}{r} \right) \hat{r} - (\hat{r} \cdot \hat{v}) \hat{v} \right] = -0.3146 \hat{I} - 0.385\ 23 \hat{J} + 0.668\ 04 \hat{K} \\ e &= 0.832\ 853 \\ \xi &= \frac{v^2}{2} - \frac{\mu}{r} = \frac{7.651\ 888^2}{2} - \frac{398600.4418}{11456.67} = -5.516\ 604 \text{ km}^2/\text{s}^2\end{aligned}$$

Because the orbit isn't parabolic (the eccentricity is not 1.0), find the semimajor axis by $a = -\frac{\mu}{2\xi}$:
 $a = 36,127.343$ km

Then, find the semiparameter using the specific angular momentum:

$$p = \frac{h^2}{\mu} = \frac{66420.10^2}{398600.4418} = 11,067.790 \text{ km}$$

The next task is to determine the angles. Find the inclination with a cosine expression. No quadrant check is necessary.

$$\cos(i) = \frac{h_K}{|\hat{h}|} = \frac{2469.6}{66420.10}$$

$i = 87.870^\circ$

Find the right ascension of the ascending node. Notice the quadrant check here affects the final value.

$$\cos(\Omega) = \frac{n_I}{|\hat{n}|} = \frac{-44500.5}{66374.17} \quad \text{Initially, } \Omega \text{ is } 132.102^\circ$$

IF($n_J < 0$) THEN $\Omega = 360^\circ - \Omega$ $\Omega = 227.898^\circ$

Find the argument of perigee similarly but without modification for quadrants:

$$\cos(\omega) = \frac{\hat{n} \cdot \hat{e}}{|\hat{n}| |\hat{e}|} = \frac{32970.953 \mu u}{(66374.17)(0.832\ 853)} \quad \omega = 53.38^\circ$$

IF($e_K < 0$) THEN $\omega = 360^\circ - \omega$

Finally, find the true anomaly:

$$\cos(\nu) = \frac{\hat{e} \cdot \hat{r}}{|\hat{e}| |\hat{r}|} = \frac{-388.773}{(0.832\ 853)(11456.67)} \quad \nu = 92.335^\circ$$

IF($\hat{r} \cdot \hat{v} < 0$) THEN $\nu = 360^\circ - \nu$

Evaluate the special angles recalling the limitations imposed for planes in each calculation. Determine the elements relating to the location of perigee. Notice the difference in the values of the true longitude of perigee and the longitude of perigee!

$$\cos(\tilde{\omega}_{true}) = \frac{e_I}{|e|} = \frac{-0.3146}{0.832853} , \text{ and initially, } \tilde{\omega}_{true} \text{ is } 112.194^\circ$$

$$\text{IF}(e_J < 0) \text{THEN } \tilde{\omega}_{true} = 360^\circ - \tilde{\omega}_{true} , \quad \tilde{\omega}_{true} = 247.806^\circ$$

$$\tilde{\omega} = \Omega + \omega = 227.89^\circ + 53.38^\circ = 281.27^\circ$$

Find the argument of latitude:

$$\cos(u) = \frac{\hat{n} \cdot \hat{r}}{|\hat{n}| |\hat{r}|} \quad \text{IF}(r_K < 0) \text{THEN } u = 360^\circ - u$$

$$u = \cos^{-1}\left(\frac{(-44500.5)(6524.834) + (-49246.7)(6862.875)}{(66374.17)(11456.67)}\right) = 145.60549^\circ$$

Finally, determine the true longitude:

$$\cos(\lambda_{true}) = \frac{r_I}{|\hat{r}|} = \frac{6524.834}{11456.67} \quad \lambda_{true} = 55.282587^\circ$$

$$\text{IF}(r_J < 0) \text{THEN } \lambda_{true} = 360^\circ - \lambda_{true}$$



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 *COE2RV* 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.

$$\hat{r}_{PQW} = \begin{bmatrix} \frac{p \cos(\nu)}{1 + e \cos(\nu)} \\ \frac{p \sin(\nu)}{1 + e \cos(\nu)} \\ 0 \end{bmatrix} \quad (2-104)$$

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:

IF Circular Equatorial
let $\omega = 0.0, \Omega = 0.0$, and $\nu = \lambda_{true}$ (2-105)
IF Circular Inclined
let $\omega = 0.0$ and $\nu = u$

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. 1-13, solve Eq. (1-18) 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. (1-25) 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:

$$\dot{\vec{v}}_{PQW} = \begin{bmatrix} -\sqrt{\frac{\mu}{p}} \sin(\nu) \\ \sqrt{\frac{\mu}{p}}(e + \cos(\nu)) \\ 0 \end{bmatrix} \quad (2-106)$$

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:

$$\begin{aligned} &\text{IF Elliptical Equatorial} \\ &\text{set } \Omega = 0.0 \text{ and } \omega = \tilde{\omega}_{true} \end{aligned} \quad (2-107)$$

The assumptions remain intact because Ω is undefined for elliptical equatorial orbits.

We can now do the coordinate transformations using Eq. (3-30). 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 COE2RV

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 **RV2COE** algorithm, we may run many test cases to verify the routine. Because **COE2RV** is simply designed to be a mirror calculation of the **RV2COE** routine, we can use the same set of test reference data. But we must test several limiting cases. Algorithm 10 summarizes the process.

ALGORITHM 10: COE2RV ($p, e, i, \Omega, \omega, \nu, u, \lambda_{true}, \tilde{\omega}_{true}$) $\Rightarrow \hat{r}_{IJK} \hat{v}_{IJK}$

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 $\omega = \tilde{\omega}_{true}$

$$\hat{r}_{PQW} = \begin{bmatrix} \frac{p \cos(\nu)}{1 + e \cos(\nu)} \\ \frac{p \sin(\nu)}{1 + e \cos(\nu)} \\ 0 \end{bmatrix} \quad \hat{v}_{PQW} = \begin{bmatrix} -\sqrt{\frac{\mu}{p}} \sin(\nu) \\ \sqrt{\frac{\mu}{p}}(e + \cos(\nu)) \\ 0 \end{bmatrix}$$

$$\vec{r}_{IJK} = [\text{ROT3}(-\Omega)][\text{ROT1}(-i)][\text{ROT3}(-\omega)]\vec{r}_{PQW} = \left[\frac{IJK}{PQW}\right]\vec{r}_{PQW}$$

$$\vec{v}_{IJK} = [\text{ROT3}(-\Omega)][\text{ROT1}(-i)][\text{ROT3}(-\omega)]\vec{v}_{PQW} = \left[\frac{IJK}{PQW}\right]\vec{v}_{PQW}$$

$$\left[\frac{IJK}{PQW}\right] = \begin{bmatrix} \cos(\Omega)\cos(\omega) - \sin(\Omega)\sin(\omega)\cos(i) & -\cos(\Omega)\sin(\omega) - \sin(\Omega)\cos(\omega)\cos(i) & \sin(\Omega)\sin(i) \\ \sin(\Omega)\cos(\omega) + \cos(\Omega)\sin(\omega)\cos(i) & -\sin(\Omega)\sin(\omega) + \cos(\Omega)\cos(\omega)\cos(i) & -\cos(\Omega)\sin(i) \\ \sin(\omega)\sin(i) & \cos(\omega)\sin(i) & \cos(i) \end{bmatrix}$$

An example demonstrates the technique.

▼ **Example 2-6. Finding Position and Velocity Vectors (COE2RV Test Case).**

GIVEN: $p = 11,067.790 \text{ km}$, $e = 0.83285$, $i = 87.87^\circ$, $\Omega = 227.89^\circ$, $\omega = 53.38^\circ$, $\nu = 92.335^\circ$

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:

$$\vec{r}_{PQW} = \begin{bmatrix} \frac{p \cos(\nu)}{1 + e \cos(\nu)} \\ \frac{p \sin(\nu)}{1 + e \cos(\nu)} \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{11067.790 \cos(92.335)^\circ}{1 + 0.83285 \cos(92.335)^\circ} \\ \frac{11067.790 \sin(92.335)^\circ}{1 + 0.83285 \cos(92.335)^\circ} \\ 0 \end{bmatrix} = \begin{bmatrix} -466.7639 \\ 11447.0219 \\ 0 \end{bmatrix} \text{ km}$$

$$\vec{v}_{PQW} = \begin{bmatrix} -\sqrt{\frac{\mu}{p}} \sin(\nu) \\ \sqrt{\frac{\mu}{p}}(e + \cos(\nu)) \\ 0 \end{bmatrix} = \begin{bmatrix} -\sqrt{\frac{398600.4418}{11067.790}} \sin(92.335) \\ \sqrt{\frac{398600.4418}{11067.790}}(0.83285 + \cos(92.335)) \\ 0 \end{bmatrix} = \begin{bmatrix} -5.996222 \\ 4.753601 \\ 0 \end{bmatrix} \frac{\text{km}}{\text{s}}$$

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

$$\left[\frac{IJK}{PQW}\right] = \begin{bmatrix} -0.37786007 & 0.55464179 & -0.74134625 \\ -0.46252560 & 0.58055638 & 0.67009280 \\ 0.80205476 & 0.59609293 & 0.03716695 \end{bmatrix}$$

Finally, multiply each vector to apply the transformation:

$$\begin{aligned}\hat{\vec{r}}_{IJK} &= \begin{bmatrix} IJK \\ PQW \end{bmatrix} \hat{\vec{r}}_{PQW} = \begin{bmatrix} -0.377\,860\,07 & 0.554\,641\,79 & -0.741\,346\,25 \\ -0.462\,525\,60 & 0.580\,556\,38 & 0.670\,092\,80 \\ 0.802\,054\,76 & 0.596\,092\,93 & 0.037\,166\,95 \end{bmatrix} \begin{bmatrix} -466.7639 \\ 11447.0219 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} 6525.344 \\ 6861.535 \\ 6449.125 \end{bmatrix} \text{km}\end{aligned}$$

$$\begin{aligned}\hat{\vec{v}}_{IJK} &= \begin{bmatrix} IJK \\ PQW \end{bmatrix} \hat{\vec{v}}_{PQW} = \begin{bmatrix} -0.377\,860\,07 & 0.554\,641\,79 & -0.741\,346\,25 \\ -0.462\,525\,60 & 0.580\,556\,38 & 0.670\,092\,80 \\ 0.802\,054\,76 & 0.596\,092\,93 & 0.037\,166\,95 \end{bmatrix} \begin{bmatrix} -5.996\,222 \\ 4.753\,601 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} 4.902\,276 \\ 5.533\,124 \\ -1.975\,709 \end{bmatrix} \text{km/s}\end{aligned}$$



2.7 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. 2-20.

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 2-21 shows the projection on a rotating Earth. 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*

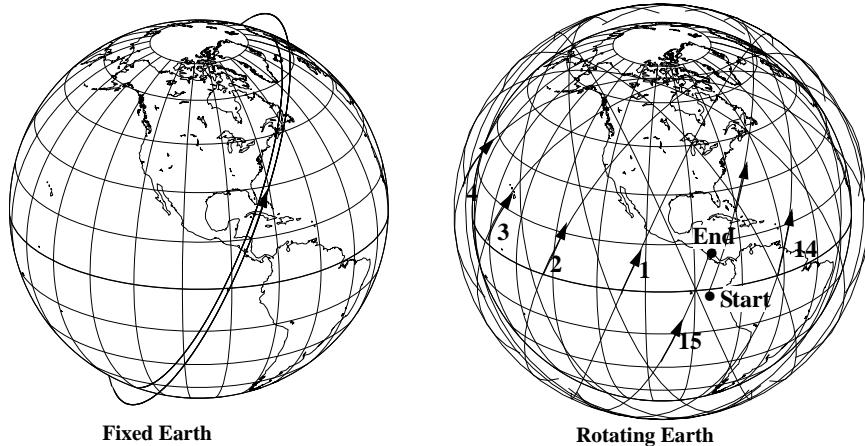


Figure 2-20. Three-D Groundtracks. With the Earth fixed, the satellite’s orbital plane is visible, but this figure doesn’t represent the satellite as it travels through time. In reality, the Earth moves under the “fixed” orbital plane. When a groundtrack is displayed over the positions it would occupy 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 reference.

groundtrack, but we won’t discuss this until Chap. 11. 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. 9.6.1. The satellite in Fig. 2-21 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 2-22 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 and the location at the end of one revolution shifts to the west.* The symmetry about lines of longitude indicates the circular nature of the orbits, and we call this *line*

* At geosynchronous altitude, the location after one revolution is the same. For geostationary orbits, the satellite always appears to be stationary over one location. The figure “8” occurs because the satellites relative velocity is less and greater, than locations on the Earth as it travels from the ascending node. Thus, the satellite appears to travel west, then east, and so on.

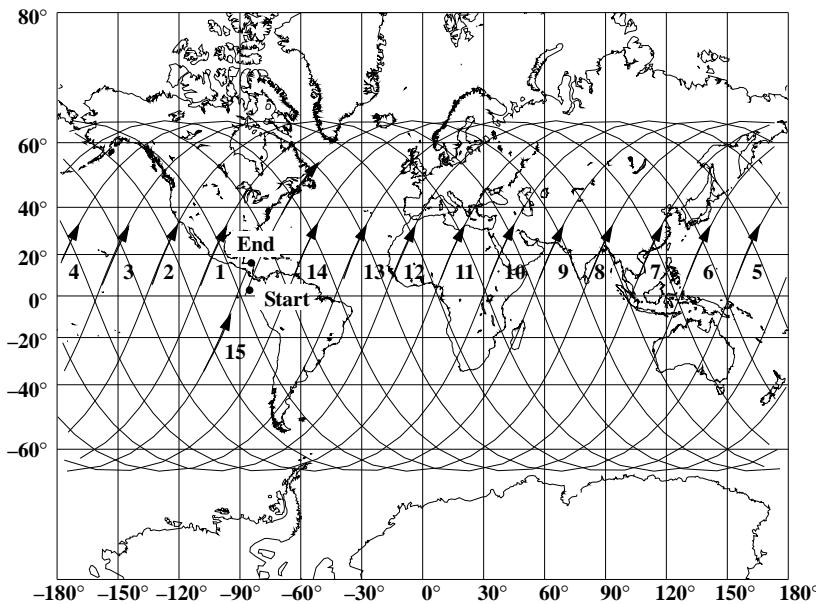


Figure 2-21. Groundtrack of a Satellite in Circular Orbit on a Rotating Earth. This form of the two-dimensional groundtrack is very useful because it shows the actual locations over a rotating Earth. Also notice the mismatch in the starting and ending points after 15 revolutions.

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. 2-20.

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. (1-29). The change in longitude, $\Delta\lambda$ measured from the initial starting location, in one revolution gives us the satellite's period,*

$$\rho = \frac{360^\circ - \Delta\lambda}{15^\circ/\text{sidereal hr}}$$

Be careful of the signs when you examine retrograde orbits.

Looking at changes in eccentricity and argument of perigee, consider Fig. 2-23. 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

* $\Delta\lambda$ could be interpreted differently. Here, it is the total longitude traveled in one orbital revolution. Thus for a LEO satellite, $\Delta\lambda$ will be about 340° , and about 380° for a similar retrograde orbit. 360° is in the equation to permit GEO orbit calculations where $\Delta\lambda = 0^\circ$. Finally, solar time may be used, in which case the denominator is $15.041\ 067^\circ/\text{solar hr}$.

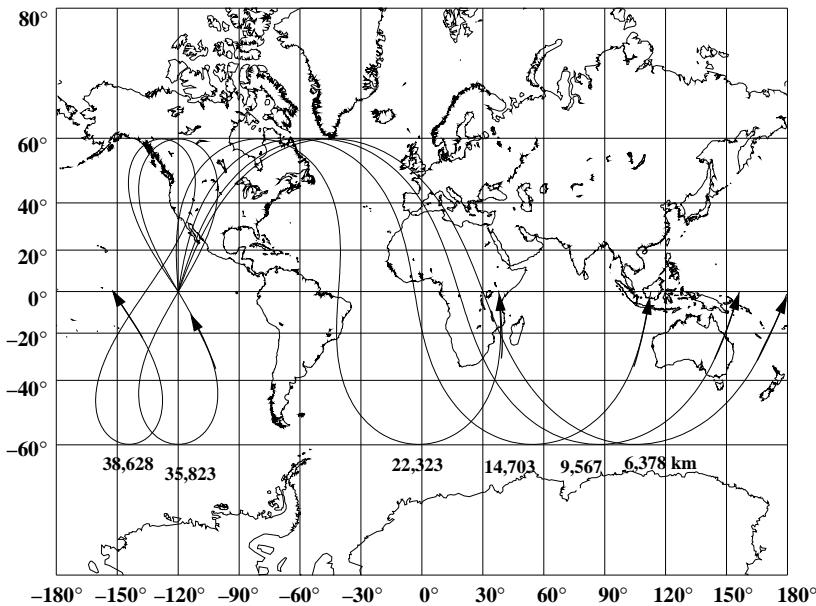


Figure 2-22. Effects of Increasing Altitude for Groundtracks on a Rotating Earth. Notice how the span of the groundtrack shrinks as the altitude increases. All groundtracks shown are for circular orbits. Altitudes are given in kilometers above the Earth's surface.

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. 2-23, 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. The nearly vertical line in the groundtrack characterizes this motion. 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.

Changes in inclination aren't as dramatic but do produce some interesting results, as Fig. 2-24 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° .

2.8 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

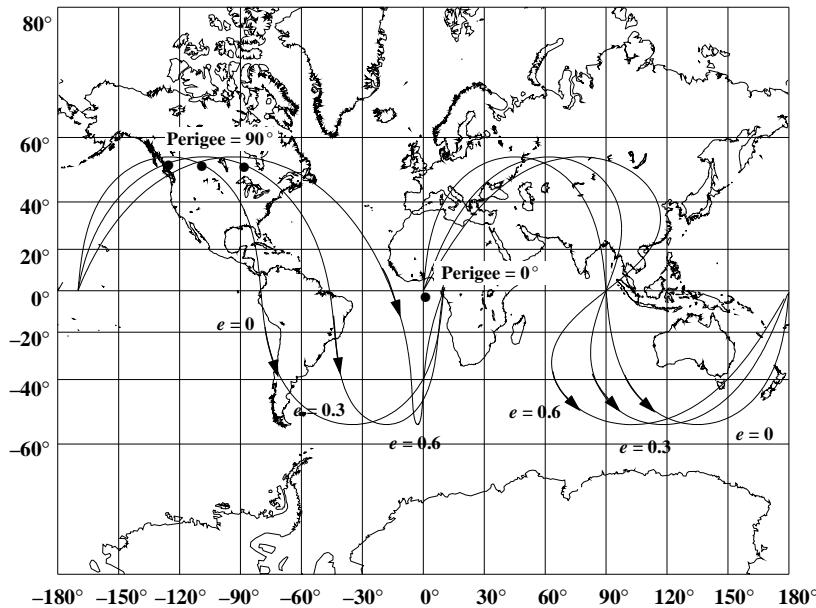


Figure 2-23. Effects of Eccentricity and Perigee Location for Groundtracks on a Rotating Earth. This figure shows two sets of orbits with different perigees. Each set contains three orbits consisting of eccentricity values of 0.0, 0.3, and 0.6. Notice how distorted each groundtrack becomes with increased eccentricity. Perigee locations at $\pm 180^\circ$ result in equal but opposite images.

slightly different because we already know the two position vectors. At first glance, this problem may even resemble Lambert's problem (Chap. 7); 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 82). Eq. (2-64) to Eq. (2-68) will ultimately be part of any solution. Notice that only Eq. (2-64) does *not* include time. Thus, if we know the change in eccentric, parabolic, 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. (2-64) and Eq. (2-65) 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

$$a = \frac{rr_o}{p} \frac{(1 - \cos(\Delta\nu))}{(1 - \cos(\Delta E))} \quad (2-108)$$

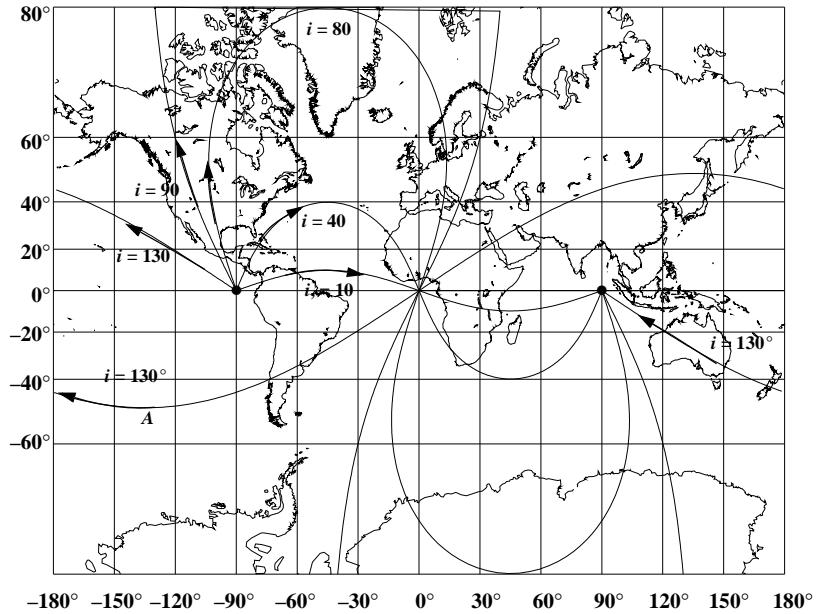


Figure 2-24. Effects of Inclination for Groundtracks on a Rotating Earth. We determine inclination by the maximum latitude on the groundtrack. Notice how the retrograde orbits ($i > 90^\circ$) tend to move to the left. Don't confuse this motion with very high, direct orbits (as in Fig. 2-22 altitude = 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".

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. (2-64) and Eq. (2-65):

$$\sqrt{\frac{\mu}{p}} \tan\left(\frac{\Delta\nu}{2}\right) \left(\frac{1 - \cos(\Delta\nu)}{p} - \frac{1}{r} - \frac{1}{r_o} \right) = -\frac{\sqrt{\mu a} \sin(\Delta E)}{r_o r}$$

Use Eq. (C-14), 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 and using Eq. (2-108) 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-14)]:

$$\frac{1 - \cos(\Delta\nu)}{p} = -\frac{1}{\sqrt{r_o r}} \left\{ \sqrt{2} \cos\left(\frac{\Delta\nu}{2}\right) \sqrt{2} \cos\left(\frac{\Delta E}{2}\right) \right\} + \frac{r_o + r}{r_o r}$$

Finally, find the semiparameter,

$$p = \frac{\{1 - \cos(\Delta\nu)\} r_o r}{r_o + r - 2\sqrt{r_o r} \cos\left(\frac{\Delta\nu}{2}\right) \cos\left(\frac{\Delta E}{2}\right)} \quad (2-109)$$

We now have several options. Fundamentally, if we know the semimajor axis or semiparameter, we can use Eq. (2-64) 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. 7.6.1) to solve the parabolic case. Notice that we only need the two position magnitudes and the change in true anomaly. The sign of $\sin(\Delta\nu)$ is chosen by the direction of flight as shown in Algorithm 53, page 444. The complete solution is in Algorithm 11, below:

ALGORITHM II: FINDTOF ($\hat{r}_o, \hat{r}, p \Rightarrow \text{TOF}$)

$$\cos(\Delta\nu) = \frac{\hat{r}_o \cdot \hat{r}}{r_o r}$$

$$k = r_o r (1 - \cos(\Delta\nu))$$

$$l = r_o + r$$

$$m = r_o r (1 + \cos(\Delta\nu))$$

$$a = \frac{mkp}{(2m - l^2)p^2 + 2klp - k^2}$$

$$f = 1 - \frac{r}{p} (1 - \cos(\Delta\nu))$$

$$g = \frac{r_o r \sin(\Delta\nu)}{\sqrt{\mu p}}$$

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 ($a < 0.0$):

$$\cosh(\Delta H) = 1 + (f-1) \frac{r_o}{a}$$

$$TOF = g + \sqrt{\frac{(-a)^3}{\mu}} (\sinh(\Delta H) - \Delta H)$$

Problems

1. Calculate the eccentric anomaly, true anomaly, and universal variable for $M = 235.4^\circ$ and $e = 0.4$. For χ , use $a = 8000$ km.
2. Calculate the Delaunay, Poincaré, flight, and equinoctial orbital elements for the satellite in Example 2-5.
3. 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.
4. Can Algorithm 8 handle multiple-revolution transfers? If so, how; and if not, why not?
5. 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

$$a = 6678.137 \text{ km}, e = 0.000096, i = 28.5^\circ, M = 278.946 88^\circ$$

Is your analysis correct?

6. A certain groundtrack is noticed with a regular 25° westward displacement on each revolution. What is the semimajor axis? How can you be sure?
7. You have received two position vectors for a satellite. How much time has elapsed between these two positions? ($p = 6681.571$ km)

$$\hat{r}_o = -2574.9533 \hat{i} + 4267.0671 \hat{j} + 4431.5026 \hat{k} \text{ km}$$

$$\hat{r} = 2700.6738 \hat{i} - 4303.5378 \hat{j} - 4358.2499 \hat{k} \text{ km}$$

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. (2-64) change?
9. You're briefing a general about a satellite constellation having six orbital planes and flying at 800 km altitude with an inclination of 60° . He asks how many of the orbital planes will cross over the United States. Wishing to remain employed, how do you answer?
10. Figure 2-15 shows time vs the universal variable χ . If two-body motion is used, the orbit will repeat in time intervals that are multiples of the period. Generate a figure that reduces time to be within one period.

CHAPTER 3 COORDINATE AND TIME SYSTEMS

- 3.1 Historical Background
- 3.2 The Earth
- 3.3 Coordinate Systems
- 3.4 Coordinate Transformations
- 3.5 Time
- 3.6 Time Conversions
- 3.7 Transforming Celestial and Terrestrial Coordinates
- 3.8 Earth Models and Constants

3.1 Historical Background

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

In Chap. 1, we derived two-body motion and then looked 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.

Newton's Universal Law of Gravitation enables us to model planetary and satellite motion. In the *Philosophiae Naturalis Principia Mathematica* (*Mathematical Principles of Natural Philosophy*, or simply *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 in vector notation as

$$\hat{\vec{F}}_g = -\frac{Gm_{\oplus}m_{sat}}{r^2} \frac{\hat{r}}{|\hat{r}|} \quad (3-1)$$

where $\hat{\vec{F}}_g$ is the force on the satellite (of mass m_{sat}) due to the Earth (of mass m_{\oplus}), and \hat{r} is the vector from m_{\oplus} to m_{sat} . As a consequence of Newton's Universal Law of Gravitation, Kepler's elliptical motion model is extended to *any* conic section. This gives us tools to define the motion of an object in space. Eq. (3-1) 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. (3-1) to account for nonuniform mass distributions in finite bodies. This is called a nonspherical gravity field. The difficulty in the 1600^s was finding the *gravitational constant*, G , which is really a proportionality constant.

In 1798, Henry Cavendish (1731–1810) measured 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 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 complex 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 EGM-96 (See Appendix D) values for μ and G are

$$\begin{aligned}\mu &\equiv G(m_{\oplus} + m_{sat}) \approx Gm_{\oplus} = 3.986\,004\,415 \times 10^5 \frac{\text{km}^3}{\text{s}^2} \\ G &= 6.673 \times 10^{-20} \pm 0.001 \times 10^{-20} \frac{\text{km}^3}{\text{kg} \cdot \text{s}^2} \quad m_{\oplus} \approx 5.973\,332\,0 \times 10^{24} \text{ kg}\end{aligned}$$

Although the precise definition includes the masses of the Earth and the satellite, we neglect the satellite's mass because it's so small relative to the Earth's. Whenever the objects of interest are more nearly the same size (the Sun and Jupiter, for instance), both masses contribute to μ . Although μ and G are experimentally measured, m_{\oplus} is a derived quantity.

Another approach to mitigate the difficulty in finding G involved the Gaussian constant of gravitation. At the time of Cavendish, measuring the mass of the Sun and Earth was difficult. 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, IAU, adopted a value of κ for the Sun at their 1939 meeting and proclaimed that it couldn't be revised (Taff, 1988:94–95).

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

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 defined it to be unity and gave it the name ***astronomical unit***, AU. We can examine the orbital period relation [Kepler's third law and Eq. (1-26)].

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

Substituting the Gaussian constant of gravitation and solving gives us

$$\begin{aligned}\kappa_{\odot} &\equiv \frac{2\pi\sqrt{a_{\oplus}^3}}{P_{\oplus}\sqrt{1 + \frac{m_{\oplus}}{m_{\odot}}}} \\ P_{\oplus} &= 365.256\ 383\ 5 \text{ days} \\ a_{\oplus} &= 1.0 \text{ AU} \\ m_{\oplus} &= \frac{1}{354,710} \text{ solar masses} \\ m_{\odot} &= 1.0\end{aligned}$$

The introduction of the Gaussian constant didn't solve the problem of measuring the mass of the Sun and the Earth. Instead, it normalized the orbit determination problem by making mass an issue of a scale factor and units. 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 κ_{\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 κ_{\odot} , is 1.0000003 a.u. Much confusion has been caused by careless usage of the words semimajor axis and mean distance in the astronomical literature.

3.2 The Earth

Because satellites orbiting the Earth are a primary concern of this book, we must establish certain parameters for the Earth. In fact, we can generalize a set of fundamental parameters to specify the shape of any planet. These values allow us to specify locations, shape, the planet's precise size, and the gravity field—a process known as ***geodesy***.

Four 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 rotational velocity, ω_{\oplus}
4. The gravitational parameter, μ

The Earth's radius has been investigated for literally thousands of years. Modern efforts began to converge on an accepted value in the mid-1800s. The JGM-3 model (Sec. 3.8) defines the *mean equatorial radius of the Earth*, R_{\oplus} , as

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

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 (JGM-3)*

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

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.

Using Eq. (1-3) and R_{\oplus} as the semimajor axis, we can also derive the *flattening* and the *eccentricity of the Earth*, e_{\oplus} :

$$f = 0.003\ 352\ 813\ 178 = \frac{1}{298.257} \quad e_{\oplus} = 0.081\ 819\ 221\ 456$$

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. 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 in the 1930's, observers could confirm small variations in the value of the Earth's rotational velocity.[†] 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

* We must carefully convert between constant values so we don't exceed the actual number of significant digits. Taff (1985:71–73) shows several examples of this error. Unfortunately, the lack of number of digits retained can introduce errors when trying to reverse a conversion. To preserve the original accuracy, it's usually good practice to keep as many decimal places as possible. Although this “rule” 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.

[†] We don't actually measure the rotational velocity. Rather, we measure the Earth's position with respect to an inertial direction and differentiate to obtain the rotational velocity (Seidelmann, 1992:76–77).

recognize several types of variability in the Earth's rotation with secular, irregular, and seasonal attributes.*

The final parameter is the gravitational parameter, which we've already discussed. For precise applications, even the mass of the atmosphere contributes to μ .

3.2.1 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 and describe locations on the Earth. Consider Fig. 3-1, which shows these parameters.

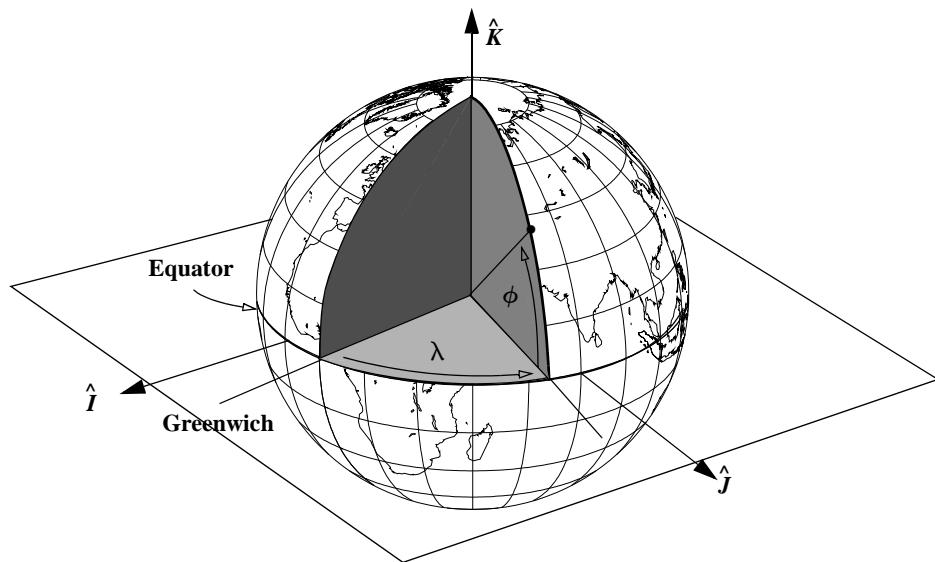


Figure 3-1. Using Latitude and Longitude to Measure Location. Terrestrial latitude, ϕ , and longitude, λ , reference the Earth's equator. The Greenwich meridian is the zero point for longitude, and the Earth's equator is the reference point for latitude.

Terrestrial latitude, ϕ , and longitude, λ , are familiar to us through common use. **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 a 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 in the *Astronomical Almanac* until 1981 so we must carefully check older programs. A

* The rotation is actually a combination of the inertial rate ($7.292\ 115\ 146\ 7 \times 10^{-5}$ rad/s) and a secular rate of precession ($7.086 \times 10^{-12} T_{UT1} + 4.3 \times 10^{-15} T_{UT1}^2$ rad/s) [NIMA (2000, 3-4 to 3-5)]. You can find this result using Eq. (3-40) and multiplying by $2\pi/86400$. From 1967–1985 the combined values ranged from $7.292\ 114\ 832 \times 10^{-5}$ to $7.292\ 115\ 099 \times 10^{-5}$ rad/s.

meridian is the intersection of a plane passing through the Earth's axis and the Earth's surface. The **prime meridian** for the Earth is the meridian that the Royal Observatory at Greenwich lies on. Longitude may take on values from 0° to 360° when measured east or from 0° to $\pm 180^\circ$ if measured east and west. The **pole** is defined by the axis of rotation of the Earth. The **equator** is the locus of points on the Earth's surface resulting from a perpendicular plane to the axis of rotation, passing through the center of mass of the Earth. The Earth's **equatorial plane** extends the equator from the Earth. It's the reference plane for measuring latitude. **Latitude** is the north-south measurement from this reference plane (the Earth's equator), with values from 0° to $\pm 90^\circ$. It's positive in the Northern Hemisphere.

3.2.2 Shape of the Earth

Knowledge of the Earth's shape is essential to many astrodynamical studies—including 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. 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. 3-2. Keep in mind the true scale of the ellipsoidal Earth model as we work through this section; I'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, a, b) with its semimajor axis equal to the equatorial radius and semiminor axis equal to the polar radius.

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—because the gravity potential is equal at all points and the direction of gravity is always perpendicular to the geoid surface.

Although an oblate Earth model introduces no unique problems with longitude, it creates problems with latitude. Consider the Earth's cross-section in Fig. 3-3 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

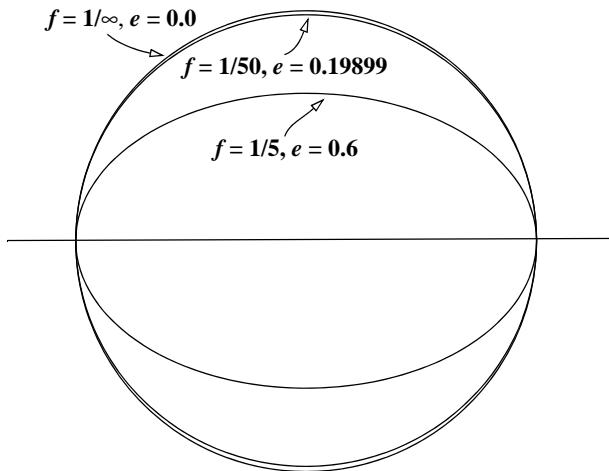


Figure 3-2. Ellipsoidal Representation of the Earth. We can approximate the Earth's shape with a flattening of about 1/298.3 to account for the 21 km difference in equatorial and polar radii. This flattening would be indistinguishable from the outer circle at this scale. Notice how closely the ellipse approaches a circle even with $f = 1/50$. For the JGM-3 model, the eccentricity of the Earth is 0.081 819 221 456.

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 a plumb-bob perpendicular to the geoid; thus it accounts for all local irregularities. It's measured using the angle between the plumb-bob direction and known star locations. This angle is very close to geodetic latitude and differs by less than $\pm 40''$. 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. It's a measured quantity. Appendix D shows sources for more information. 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.

Application: Converting Latitude Values

A basic 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 gravitational potential operations (Sec. 8.6.1). 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. 3-3). The **reduced latitude**, ϕ_{rd} , is the geocentric angle measured to a point on the auxiliary circle resulting

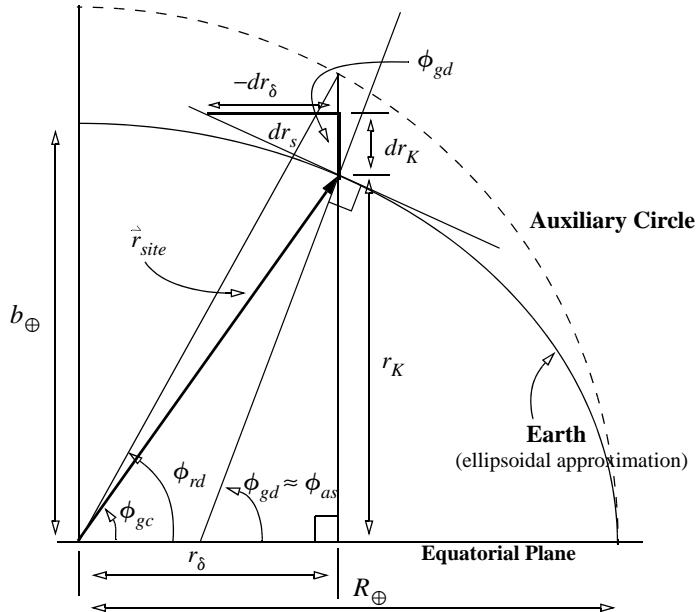


Figure 3-3. Latitude Geometry (exaggerated scale). The geodetic latitude, ϕ_{gd} , makes an angle perpendicular to the surface and the equatorial plane, whereas the geocentric latitude, ϕ_{gc} , is referenced to the center of the Earth. Astronomic latitude, ϕ_{as} , is very close to ϕ_{gd} . The center of the ellipsoid differs slightly for ϕ_{gd} and ϕ_{as} . Reduced latitude, ϕ_{rd} , is used only to derive relations.

about a line perpendicular to the equatorial plane, through the site, and intersecting the auxiliary circle. 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 in the relationship between latitudes, we can consider the position as a two-dimensional problem independent of λ . In Eq. (3-2), we use r_δ as the horizontal component (this is the equatorial component in Fig. 3-1) along the semimajor axis and r_K as the vertical component parallel to the semiminor axis.

$$\begin{aligned} r_\delta &= \sqrt{r_I^2 + r_J^2} = r_{site} \cos(\phi_{gc}) \\ r_K &= r_{site} \sin(\phi_{gc}) \end{aligned} \quad (3-2)$$

We can also find the station coordinates in terms of the reduced latitude. We find the horizontal component similar to Eq. (3-2). To discover the vertical component, we use

the scaling transformation [Eq. (1-5)], which allows us to determine the vertical component on an ellipse or a circle. The results are

$$\begin{aligned} r_\delta &= R_\oplus \cos(\phi_{rd}) \\ r_K &= R_\oplus \sqrt{1 - e_\oplus^2} \sin(\phi_{rd}) \end{aligned} \quad (3-3)$$

We can express the site position's magnitude as

$$r_{site} = \sqrt{r_\delta^2 + r_K^2} = R_\oplus \sqrt{1 - e_\oplus^2 \sin^2(\phi_{rd})} \quad (3-4)$$

The geodetic latitude is also related to the site position. If we take the differentials of Eq. (3-3), 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.

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

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. The connection between the components of the differentials and the geodetic latitude drawn at the site in Fig. 3-3 is

$$\begin{aligned} \sin(\phi_{gd}) &= -\frac{dr_\delta}{dr_s} = \frac{\sin(\phi_{rd})}{\sqrt{1 - e_\oplus^2 \cos^2(\phi_{rd})}} \\ \cos(\phi_{gd}) &= \frac{dr_K}{dr_s} = \frac{\sqrt{1 - e_\oplus^2} \cos(\phi_{rd})}{\sqrt{1 - e_\oplus^2 \cos^2(\phi_{rd})}} \end{aligned} \quad (3-5)$$

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

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

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:

$$1 - e_{\oplus}^2 \cos^2(\phi_{rd}) = \frac{1 - e_{\oplus}^2}{1 - e_{\oplus}^2 \sin^2(\phi_{gd})} \quad (3-6)$$

From the equations of the geodetic latitude above [Eq. (3-5)], we'll solve for the reduced latitude term in the denominator and substitute it into the quantity Eq. (3-6):

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

Finally, substituting these expressions for the reduced latitude into Eq. (3-3), 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 an ellipsoidal surface, but surface features 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. These variations don't cause too much difficulty because we can add the altitude into the component equations. We use the term h_{ellp} for the height above the ellipsoid; the subscript ($ellp$) specifies that the elevation is related to the ellipsoid.

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

$$C_{\oplus} = \frac{R_{\oplus}}{\sqrt{1 - e_{\oplus}^2 \sin^2(\phi_{gd})}} \quad S_{\oplus} = \frac{R_{\oplus}(1 - e_{\oplus}^2)}{\sqrt{1 - e_{\oplus}^2 \sin^2(\phi_{gd})}} \quad (3-7)$$

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

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

Notice the two coefficients differ only by the numerator, thus enabling some computational efficiency when programming on the computer. 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. 3-3 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 [combining Eq. (3-3) and Eq. (3-4)].

$$\begin{aligned}\sin(\phi_{gc}) &= \frac{r_K}{r_{site}} = \frac{\sin(\phi_{rd})\sqrt{1-e_\oplus^2}}{\sqrt{1-e_\oplus^2 \sin^2(\phi_{rd})}} \\ \cos(\phi_{gc}) &= \frac{r_\delta}{r_{site}} = \frac{\cos(\phi_{rd})}{\sqrt{1-e_\oplus^2 \sin^2(\phi_{rd})}}\end{aligned}\quad (3-8)$$

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 using Eq. (C-10):

$$\sqrt{\frac{1}{1-e_\oplus^2 \cos^2(\phi_{gc})}} = \sqrt{\frac{1-e_\oplus^2 \sin^2(\phi_{rd})}{1-e_\oplus^2}} \quad (3-9)$$

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. (3-3) to find the site's vertical component. Using Eq. (3-8) and Eq. (3-9) 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. (3-3) to complete the result: (note the approximate result because h_{ellp} is not measured directly along the geocentric radius)

$$\begin{aligned}
 r_\delta &\equiv \left(\frac{R_\oplus \sqrt{1 - e_\oplus^2}}{\sqrt{1 - e_\oplus^2 \cos^2(\phi_{gc})}} + h_{ellp} \right) \cos(\phi_{gc}) \\
 r_K &\equiv \left(\frac{R_\oplus \sqrt{1 - e_\oplus^2}}{\sqrt{1 - e_\oplus^2 \cos^2(\phi_{gc})}} + h_{ellp} \right) \sin(\phi_{gc}) \\
 r_{site} &= \sqrt{r_\delta^2 + r_K^2}
 \end{aligned} \quad (3-10)$$

For locations on the ellipsoid, we 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. 3-3 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 the two latitude types, depending on the given information. Thus,

$$\begin{aligned}
 \tan(\phi_{gd}) &= \frac{\tan(\phi_{gc})}{1 - e_\oplus^2} \\
 \tan(\phi_{gc}) &= (1 - e_\oplus^2)\tan(\phi_{gd})
 \end{aligned} \quad (3-11)$$

Remember that Eq. (3-11) is valid only for locations on the Earth's surface (ellipsoid). Section 3.4.4 shows a way to convert between latitudes above the Earth's surface. The following example will tie together this section's concepts.

▼ Example 3-1: Determining Site Components.

GIVEN: Mt. Evans, Colorado. $\phi_{gd} = 39.586667^\circ$, $\lambda = -105.640^\circ$, $H_{MSL} = 4347.667$ 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.

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

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

$$r_{\delta} = 4925.429\,802\,6 \text{ km and } r_K = 4045.493\,742\,6 \text{ km}$$

As an aside, we can also find the reduced latitude ($\phi_{rd} = \tan^{-1}(\tan(\phi_{gd})\sqrt{1 - e_{\oplus}^2})$) and then use Eq. (3-3) to find the components directly.

3.2.3 Gravitational Model

If we assume a spherical Earth of uniform density, a point mass at a given position above the Earth's surface, and a satellite of negligible mass, Newton's inverse square law [Eq. (3-1)] predicts that 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 3-4 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 NIMA (2000:4-1), the closed-gravity formula of Somigliana is

$$g_{th} = g_{equator} \left\{ \frac{1 + k_g \sin^2(\phi_{gd})}{\sqrt{1 - e_{\oplus}^2 \sin^2(\phi_{gd})}} \right\} \quad g_{equator} = 9.780\,325\,335\,9 \text{ m/s}^2$$

$$k_g = \frac{b_{\oplus} g_{pole}}{R_{\oplus} g_{equator}} - 1 = 0.001\,931\,852\,652\,41 \quad g_{pole} = 9.832\,184\,937\,8 \text{ m/s}^2$$
(3-12)

For gravity calculations at locations above the surface of the Earth, see NIMA (2000:4-2 to 4-8). For general calculations, use the mean value of theoretical gravity at the surface of the Earth as $g = 9.797\,643\,222 \text{ m/s}^2$.

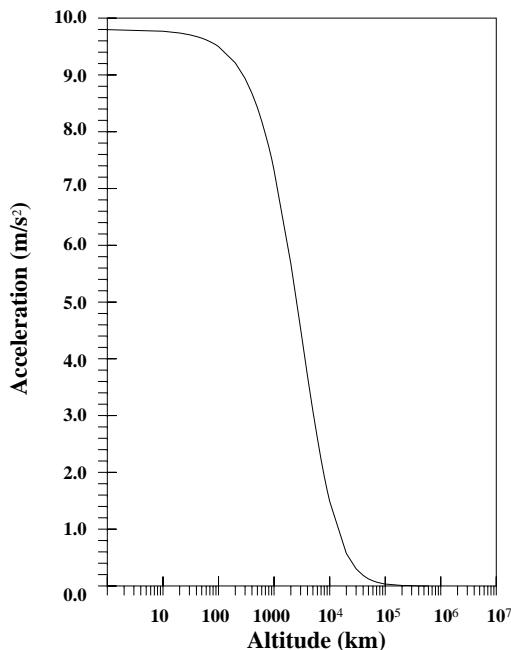


Figure 3-4. Magnitude of Gravitational Acceleration. Notice how high a satellite must be before it is in “zero gravity.” It may seem confusing that objects float on board a Space Shuttle (altitude ≈ 250 km), which has almost 90% of normal gravity, but the gravity is balanced with centrifugal forces, thus giving the perception of zero gravity.

Geopotential Surfaces

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. Altimetry is used for open sea locations. 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. Figure 3-5 shows the relations between the different types of geopotential surfaces.

The *geoid’s undulation*, N_{\oplus} , is the geoid’s height above the ellipsoid, whereas the actual height [of the topography] 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. These heights are referenced to a starting set of values called a *regional*, or *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

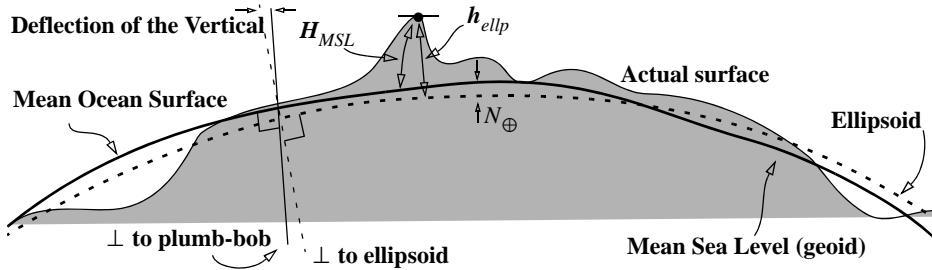


Figure 3-5. Geopotential Reference Surfaces. The height above sea level, H_{MSL} , and the geoid's undulation, N_{\oplus} , determine the ellipsoidal height, h_{ellp} . N_{\oplus} is typically very small. Notice that the ellipsoid is “smooth” whereas the geoid has many small variations. Also, H_{MSL} is measured along a curved surface as shown.

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, the *South American Datum* (SAD-69), the *European Datum* (ED-50), the *Tokyo Datum* (TD), and the *Australian Geodetic Datum* [AGD-66 (84)]. Unfortunately, having several datums introduces complexity when we tie them together. If we simply combine the datums, their centers may not coincide, and we need a transformation. The WGS-84 model ties many horizontal and vertical datums together. The conversion from a certain datum to WGS-84 often uses a form of the *Molodensky formulas* (See Defense Mapping Agency, 1987b:7-5 to 7-28 and NIMA, 2000:App. D). The **ellipsoidal height**, h_{ellp} , is similar to the orthometric height, but it's measured perpendicular to the ellipsoid (as shown in Fig. 3-5) and differs by the geoid undulation. The geoid's undulation at any location (NIMA, 2000:6-2) is

$$N_{\oplus} \cong h_{ellp} - H_{MSL}$$

$$N_{\oplus} = \frac{\mu}{rg_{th}} \sum_{l=2}^{\infty} \sum_{m=0}^l \left(\frac{R_{\oplus}}{r}\right)^l \bar{P}_{l,m} [\sin(\phi_{gc})] \left\{ \bar{C}_{l,m}^* \cos(m\lambda) + \bar{S}_{l,m} \sin(m\lambda) \right\} \quad (3-13)$$

Note here that r is the distance to the ellipsoid from the Earth's center and $\bar{C}_{l,m}^*$ indicates the even zonal harmonics, modified by subtracting their geometric (nonrotating) value [NIMA (2000:6-3)]. We usually limit the number of harmonics used. $\bar{P}_{l,m}$ is a normalized Legendre polynomial (Sec. 8.6.1). Eq. (3-13) is similar to the gravitational potential [Eq. (8-19)], but we add a centrifugal term and use *Brun's theorem* [Lambeck (1988:20)] to find the actual values of N_{\oplus} . Eq. (3-13) uses the gravitational coefficients of the Earth and the theoretical gravity, g_{th} , in Eq. (3-12). Because the geoid's undulation takes on values between +85 and -107 m, you use this formula only for precise analysis. See NIMA (2000:6-6) 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 pro-

cess is time consuming, N_{\oplus} is tabulated in contour charts for grids of $10^{\circ} \times 10^{\circ}$, $1^{\circ} \times 1^{\circ}$, and so on. Locations not contained on a grid are found using interpolation techniques.

Application: Finding Terrestrial Coordinates

We often require the precise location of a sensor or station. For general applications, the geodetic values provide satisfactory results. These representations require spherical coordinates and use either geodetic or geocentric inputs. However, we can also use geocentric rectangular coordinates and reference the location of the site from the Earth's center. Remember that the location should always be the same!

$$\begin{bmatrix} r_I \\ r_J \\ r_K \end{bmatrix} = |\hat{r}_{IJK}| \begin{bmatrix} \cos(\phi_{gc})\cos(\lambda) \\ \cos(\phi_{gc})\sin(\lambda) \\ \sin(\phi_{gc}) \end{bmatrix} = \begin{bmatrix} (C_{\oplus} + h_{ellp})\cos(\phi_{gd})\cos(\lambda) \\ (C_{\oplus} + h_{ellp})\cos(\phi_{gd})\sin(\lambda) \\ (S_{\oplus} + h_{ellp})\sin(\phi_{gd}) \end{bmatrix} \quad (3-14)$$

A range, $|\hat{r}_{IJK}|$, is included in the formulation with geocentric values. This value comes from the magnitude of the geodetic formulation. The reverse process is in Sec 3.4.4.

▼ Example 3-2. Finding Station Coordinates.

GIVEN: Ascension Island coordinates: $\phi_{gd} = -7^{\circ} 54' 23.886''$, $\lambda = 345^{\circ} 35' 51.000''$, $h_{ellp} = 56$ m

FIND: Position vectors

First, we must find the decimal latitude and longitude values, and convert units. Using Algorithm 17, $\phi_{gd} = -7.906\ 635\ 7^{\circ}$, $\lambda = 345.5975^{\circ}$. Using Eq. (3-7) and $R_{\oplus} = 6378.137$ km, find the interim values for $C_{\oplus} = 6378.541\ 012\ 74$ and $S_{\oplus} = 6335.840\ 635\ 42$ km. The geodetic position vector is

$$\begin{aligned} \hat{r}_{IJK} &= \begin{bmatrix} (C_{\oplus} + h_{ellp})\cos(-7.906)\cos(345.5975) \\ (C_{\oplus} + h_{ellp})\cos(-7.906)\sin(345.5975) \\ (S_{\oplus} + h_{ellp})\sin(-7.906) \end{bmatrix} \\ &= 6,119.400\ 269\ 32 \hat{I} - 1,571.479\ 555\ 45 \hat{J} - 871.561\ 180\ 90 \hat{K} \text{ km} \end{aligned}$$

The magnitude is 6,377,791.666 51 m.

Now, using the geocentric latitude (which equals $-7.854\ 367\ 75^{\circ}$ using Eq. (3-11)) and the magnitude from the geodetic calculation, the results are

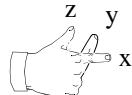
$$\hat{r}_{IJK} = 6,119.400\ 276\ 66 \hat{I} - 1,571.479\ 557\ 34 \hat{J} - 871.561\ 125\ 98 \hat{K} \text{ km}$$

▲ The small difference between the two answers is from truncation errors.

You can obtain station locations from a variety of sources. The data comes from regional datums (NAD-83 for instance) in which the defining parameters of the datum are specified. Another issue arises from a particular reference system (WGS, ITRF, etc.). These coordinate systems may include translation, rotation, and scaling to transform between systems. McCarthy (1996:12) shows the transformation between previous ITRF realizations. NIMA (2000:App. D) lists parameters to convert vectors from certain datums to WGS-84.

3.3 Coordinate Systems

One of the first requirements for describing an orbit is to define a suitable reference system. We define a rectangular coordinate system* 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 have 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—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 helps the naming of each type. In general, there are three common designations: an object's center of mass, 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 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. 3-6). **Great circles** are the intersections 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** projects the Earth's equator onto the celestial sphere. **Hour circles** are great circles that are perpendicular to the celestial equator.

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. One coordinate describes the angular distance of the point of interest above or below the fundamental

* In reality, there is a difference between a system and a frame. Feissel (1997) defines a **system** as a “set of prescriptions and conventions together with the modeling requirements to define, at any time, a triad of axes.” A **frame** is the realization of a certain coordinate set within the overall description of the system. Perhaps the best information on various reference frames comes from the publications of the International Earth Rotation Service (McCarthy and Petit, 2003). They distinguish between *celestial* and *terrestrial* reference frames. **Celestial reference frames** are barycentric frames determined from VLBI measurements, whereas **terrestrial frames** are bound to the Earth. The suggested Earth reference frame is the *International Terrestrial Reference Frame* (ITRF) model described by McCarthy and Petit (2003). Each revision is designated with a two-digit year (ITRF-94, ITRF-96). We'll mainly use terrestrial frames for the satellite motions in this book. As mentioned earlier, many datums further define terrestrial reference frames. McCarthy and Petit (2003:Chap. 5) show ways to convert between terrestrial reference frames (e.g., ITRF-90) and datums (e.g., WGS-84).

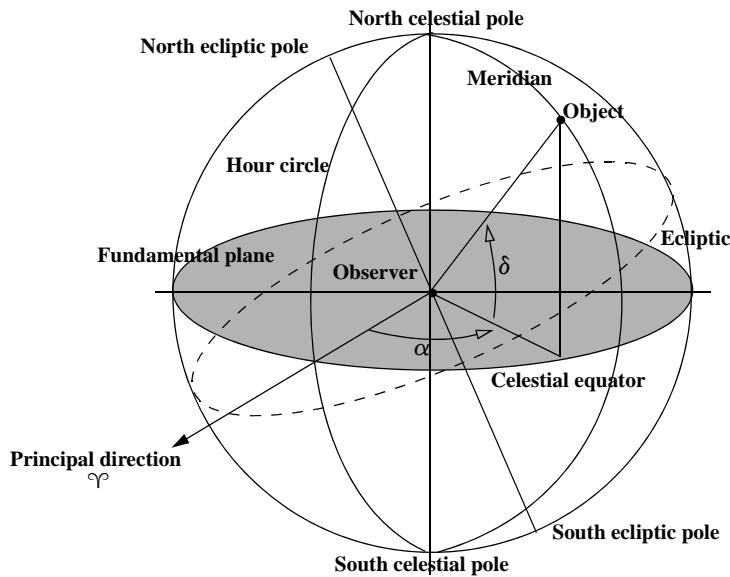


Figure 3-6. Geometry of the Celestial Sphere. The celestial sphere is based on an observer's perceived view of objects in space. A meridian, or hour circle, is any circle that passes through the observer at the center. I've shown the principal direction to help us define right ascension, α , and declination, δ .

plane, and the other describes the angle from the reference point (sometimes called a *fiducial* direction) along the celestial equator to the hour circle passing through the object. These coordinates are analogous to latitude and longitude and are called declination and right ascension, respectively.

The Earth and its orbit around the Sun form the basis for celestial coordinate systems. The *ecliptic* is the plane of the Earth's mean 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. 3-7. It doesn't exactly follow the ecliptic because this path is defined as the plane of the Earth's mean orbit. 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. The intersection of the two planes helps us fix a principal direction. The line of intersection is called the *line of nodes*. The Sun crosses this intersection twice a year. 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 (*aequinoc-tium*) meaning equal night.

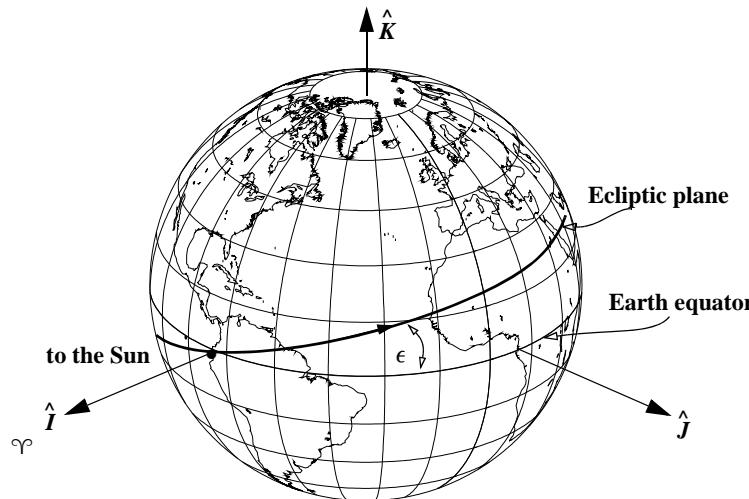


Figure 3-7. Geometry of the Vernal Equinox. The Earth's mean orbit around the Sun forms the ecliptic plane. The Earth's equatorial plane is inclined about 23.5° to the ecliptic. When the Sun is at the intersection of the two planes (has zero declination) and is at the ascending node, as viewed from Earth, it's the first day of spring.

A formal definition for the *vernal equinox* is that it occurs when the Sun's declination (Fig. 3-6) 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 226), 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 reference the ecliptic as the fundamental plane and measure ecliptic longitude from the vernal equinox positive to the east, as shown in Fig. 3-7 and Fig. 3-8.

Measurements of right ascension and declination are similar to those using longitude and latitude, but the reference point is the vernal equinox. Because the vernal equinox moves slightly, we must specify its exact location at a certain time (Fig. 3-8). **Right ascension**, α , is measured positive to the east in the plane of the equator from the vernal equinox direction. Similar to longitude, the range of values may be from 0^h to 24^h (0° to 360°). 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°). Right ascension and declination define only the object's direction. Astronomers use this

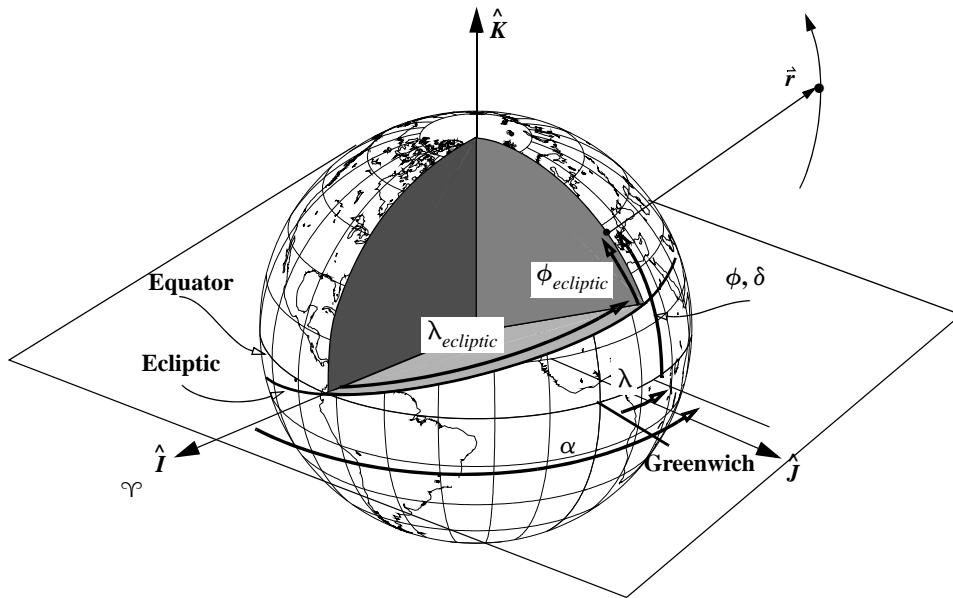


Figure 3-8. Using Right Ascension and Declination or Longitude and Latitude to Measure Location. Notice that both terrestrial latitude and longitude, ϕ and λ , and right ascension and declination, α and δ , reference the Earth's equator, whereas ecliptic latitude and longitude are celestial coordinates that reference the ecliptic. Declination and terrestrial latitude are the same, but right ascension and terrestrial longitude reference different starting locations.

system to catalog star positions. The extreme distance of stars from the Earth means that we often don't have accurate range information and thus specify only the direction. We'll examine these angles in more detail in Chap. 4.

Hour Angle

We use hour angles to describe 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, with a sign convention in which angles are measured positively *westward* to the object. 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) for local observers (the **Local Hour Angle**, LHA) and observers at Greenwich (the **Greenwich Hour Angle**, GHA). 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. 3-9, notice that $LHA = GHA + \lambda$. Calculations are simply additions and subtractions of

GHA, LHA, and terrestrial longitude; with Eq. (3-37), degrees and hours are interchangeable, although the common units are hours.

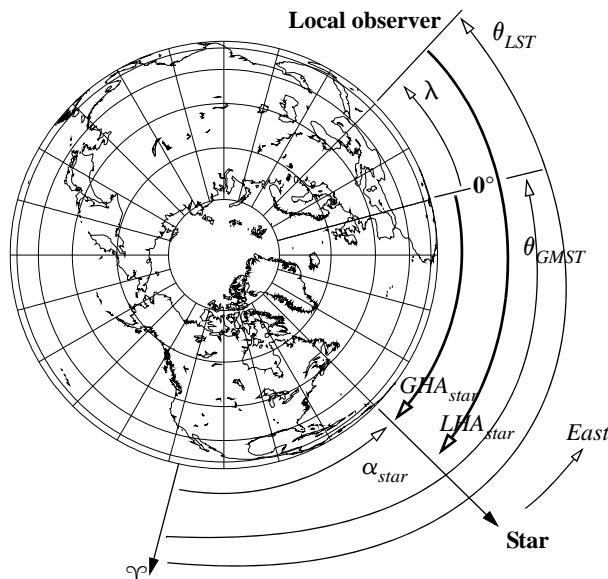


Figure 3-9. Geometry for LHA and GHA. GHA, and LHA are always measured positive to the west from the observer to the object of interest. Longitude, λ , is positive to the east. Both GHA_{star} and LHA_{star} are positive here. We use subscripts to avoid confusion when referring to specific objects. θ_{LST} and θ_{GMST} will be discussed shortly.

Barycentric Coordinate Systems

We often form equations of motion using a system's center of mass, or **barycenter**, as the origin. The remaining characteristics are derived from the "parent" system—for instance, the heliocentric and geocentric systems. 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 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 system. This rate is the mean motion of the secondary object about the 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 the orbital plane defined by two-body motion. I'll use the letters of the "parent"

system with a subscripted S . 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. The Z_S axis is normal to the invariable plane (parallel to the angular momentum vector). This system is especially useful in studying three- and four-body problems, introduced in Chap. 1.

3.3.1 Interplanetary Systems

The ***Heliocentric Coordinate System***, XYZ, as shown in Fig. 3-10, is a generic system for interplanetary missions (e.g., patched-conic approximations). 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. Because the principal direction is the vernal equinox, the Earth's aphelion and perihelion don't match any particular axis of the coordinate system. There is also a rotating synodic system ($X_S Y_S Z_S$).

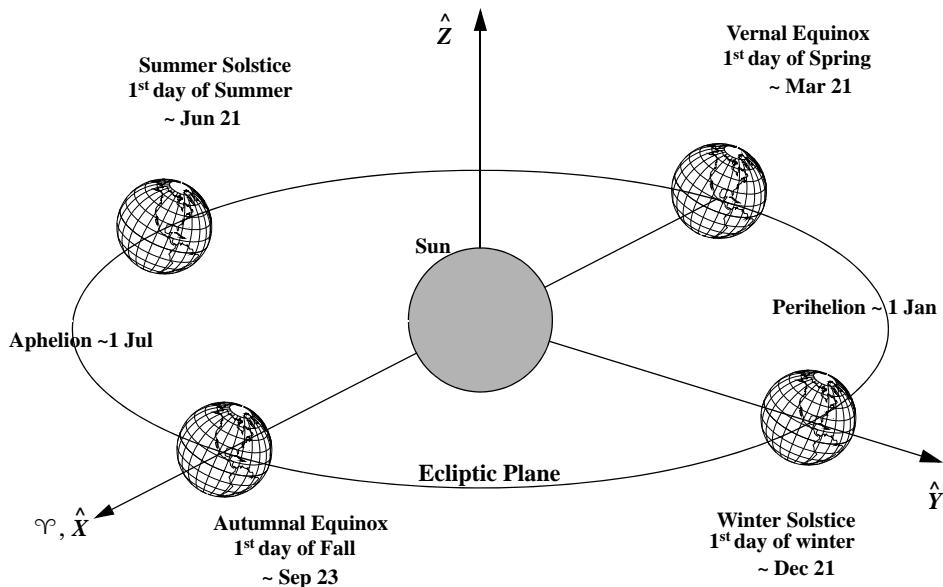


Figure 3-10. Heliocentric Coordinate System, XYZ. The primary direction for this system is the intersection of the ecliptic and equatorial planes. The positive sense (denoted γ°) is given by the location of the Sun with respect to the Earth as it crosses the equatorial plane on the first day of spring. All seasonal dates are for the Northern Hemisphere. The revolution of the Earth about the Sun doesn't match the Gregorian calendar exactly. Only after about 28 years are the dates approximately the same. This affects the dates for the seasons. Seidelmann (1992:477) lists the earliest seasonal dates that occur in 2096 (March 19, June 20, September 21, and December 20). The latest seasonal dates occurred in 1903 (March 21, June 22, September 24, and December 23).

Beginning January 1, 1998, the IAU adopted the ***International Celestial Reference System*** (ICRS) as the standard reference system. The origin is at the barycenter of the solar system. The frame (ICRF) is realized by observations of 3414 extragalactic radio sources from Very Long Baseline Interferometry (VLBI) measurements. The primary direction uses the IAU-76/FK5 J2000 value adopted from the radio source 3C 273. Although the frame of the ICRF is periodically re-evaluated (ICRF1 in 1995, ICRF Ext1 in 1999 and Ext2 in 2002, and ICRF2 in 2009), the solutions introduce no rotation from previous realizations. Numerical operations can benefit from the axes remaining fixed. In 1991, the IAU introduced the Barycentric (BCRS) and Geocentric (GCRS) Celestial Reference Systems. These systems are intended for applications within the framework of General Relativity. The IAU-2000 Resolutions use both the celestial (GCRF) and terrestrial (ITRF) frames.*

3.3.2 Earth-based Systems

The origin of Earth-based systems may be at the Earth's center (geocentric), or at a site on the Earth's surface (topocentric), or off the Earth's surface. 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. Remember the Earth-Moon barycenter is about 4671 km from the geocenter, in the direction of the Moon (See Prob. 8 in Chap. 9 for numerical examples).

Geocentric Equatorial Coordinate System, IJK

This system originates at the center of the Earth, as the name implies, and is generically designated with the letters IJK. The fundamental plane is the Earth's equator, as shown in Fig. 3-11. 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.

It's one of the most common systems in astrodynamics, but it's also potentially the most confusing. The geocentric frame, IJK, is often used interchangeably with an *Earth-Centered Inertial* (ECI) or *Conventional Inertial System* (CIS) nomenclature. The equinox and plane of the equator move very slightly over time, so the term "inertial" can cause confusion. We can achieve a "pseudo" Newtonian inertial system if we refer to the equator and equinox at a particular epoch. For many years, the IAU-76/FK5 system was the standard pseudo-inertial system for geocentric coordinates. The ***IAU-76/FK5 system*** is based on the *Fundamental Katalog*, FK5 star catalog and J2000 is a quasi-inertial frame realized in the IAU-76/FK5 system. The equator and equinox's motion is modeled using the IAU-1976 Precession Model and the IAU-1980 Theory of Nutation, so inertial frames at other times can be transformed to the IAU-76/FK5 system. I'll use ECI for generic applications.

* Be aware that some centers use ICRF interchangeably with GCRF. This may be confusing because the ICRF references the solar system barycenter, while the GCRF references the Earth's center. See Chap. 8 Problem 9 for numerical values of the barycenter and planetary centers.

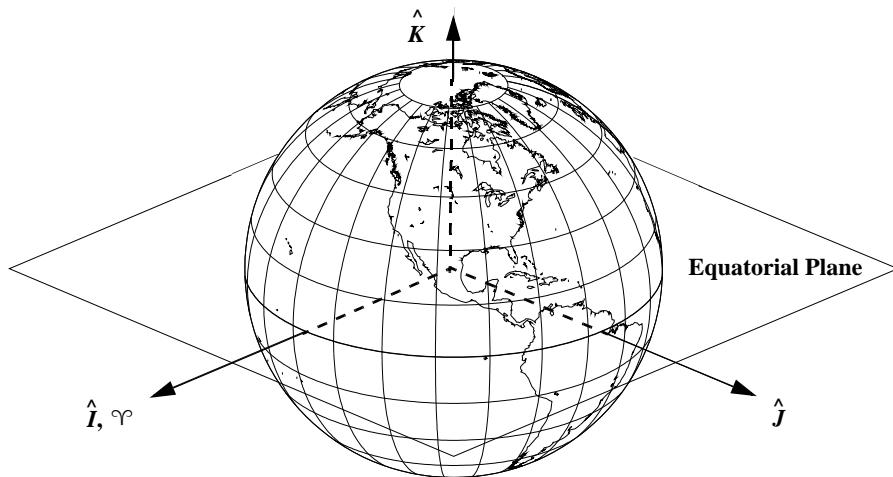


Figure 3-11. Geocentric Equatorial System (IJK). This system uses the Earth's equator and the axis of rotation to define an orthogonal set of vectors. The vernal equinox direction is fixed at a specific epoch for most applications.

Geocentric Celestial Coordinate System, GCRF

The *Geocentric Celestial Reference Frame* is the standard inertial coordinate system for the Earth and it is the geocentric counterpart of the ICRF. It has been the adopted celestial reference frame for the IERS since January 1, 1997. The axes of the GCRF are closely aligned to the frame of IAU-76/FK5 (See J2000 in Sec 3.7) to provide continuity with former IAU systems. Because there was no official nutation theory compatible with the ICRF in 1997, the IERS maintains tabulated corrections for the IAU-76/FK5 theory to relate it to the GCRF. Applying these corrections effectively defines a more accurate (but different) conventional theory than the IAU-76/FK5. The IAU-2000 Resolutions reference the GCRF directly.

Body-Fixed Coordinate System, ITRF

A geocentric coordinate system fixed to the rotating Earth, results in the *Body-Fixed* (BF), or *International Terrestrial Reference Frame* (ITRF), coordinate system. Its origin is at the center of the Earth and the axes are realized by the adopted coordinates of defining stations on the Earth's surface.* Because these stations are affected by plate tectonic motion (~cm/yr), the ITRF is regularly re-estimated as a weighted solution combining

* Confusion may exist because the ITRF system is frequently called the *Earth-Centered, Earth-Fixed* (ECEF) coordinate frame. The term “Earth-fixed” describes a terrestrial reference system whose net global orientation remains unchanged over time with respect to the crust of the Earth. However, imprecise use of this term invites confusion. For example, the U.S. Air Force Space Command has historically reserved the term “*Earth-Fixed Greenwich* frame, EFG” for the (rotating) pseudo-body fixed frame (no polar motion), and used “*Earth-Centered Rotating*, ECR” to describe the Earth-fixed ITRF!

several analysis center solutions so there is no net rotation with previous ITRF realizations.* The ITRF solutions are usually identified with the year of their publication (ITRF-08). Conversions between ITRF solutions involves translation, scale, and rotation. Typically, we use this system to process actual observations and acceleration calculations during orbit propagation. In some applications, a different meridian is chosen, but I will specify these cases. I'll use ECEF for generic references to the ITRF.

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. 3-12. 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. The Z axis (zenith) points radially outward from the site, along the site's local vertical. We state that the vertical lies along the geographic latitude, however sites generally choose geodetic or astronomical latitude.

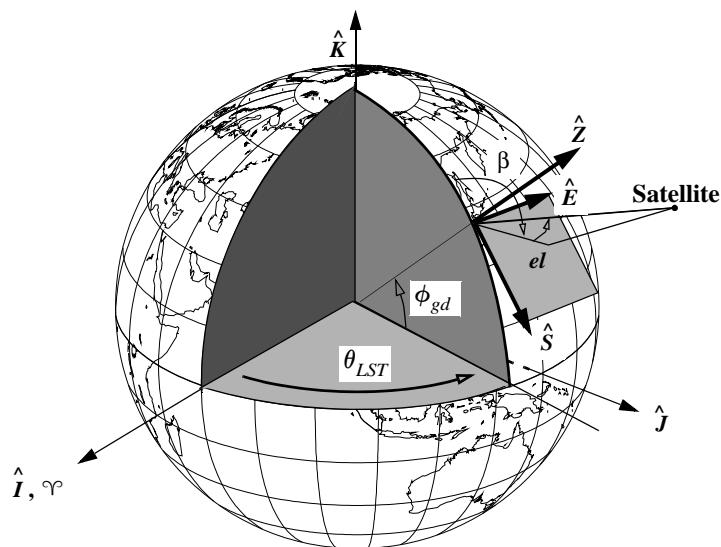


Figure 3-12. Topocentric Horizon Coordinate System, SEZ. This system moves with the Earth, so it's not fixed like the ITRF. θ_{LST} is required to orient the SEZ system to a fixed location. Note that for precise applications, θ_{AST} and longitude replace θ_{LST} . The SEZ system is used extensively in radar observations and often involves azimuth, β , and elevation, el . Geodetic latitude is common, ϕ_{gd} , but astronomical latitude is also used in many systems.

* The US DoD uses the WGS-84 terrestrial frame. It's practically identical to the ITRF, but it's realized through GPS observations. Because the positions of many fundamental stations in the WGS-84 solutions are not adjusted, but defined *a-priori* by their adopted ITRF positions, modern realizations of the WGS-84 frame are now closely aligned with the ITRF.

The SEZ system lets us 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). We usually reconcile values to be 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 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. 4 and 7.

A subtlety exists in the definition for azimuth. You might expect β to be measured positive, counter-clockwise from south (in the SEZ frame). We get the same result using the previous definition (positive, clockwise from north) because a North-East-Zenith (NEZ) system is left-handed. This NEZ 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. The direction of the principal axis (North or South) and the sign of the azimuth (clockwise or counter-clockwise) are necessary for interfacing with any particular system. These conventions must be clear for initial orbit determination (Chap. 7).

Topocentric Equatorial Coordinate System, $I_tJ_tK_t$

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 observer (see Fig. 3-13). The orientation of the axes is the same as the IJK system; the I_t axis is parallel to the vernal equinox direction. This system is used mainly for highly accurate optical systems (for instance the GEODSS and other optical telescopes).

This system’s usefulness may not yet be apparent. As discussed earlier, right ascension and declination reference the geocentric equatorial system. The $I_tJ_tK_t$ 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, observations of satellites orbiting near Earth are performed in the topocentric equatorial system. In Fig. 3-13, 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.

3.3.3 Satellite-based Systems

Although many 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 use true anomaly and the classical orbital elements (Sec. 2.4.1) to describe object locations. We need many systems to cover various satellite operations and missions.

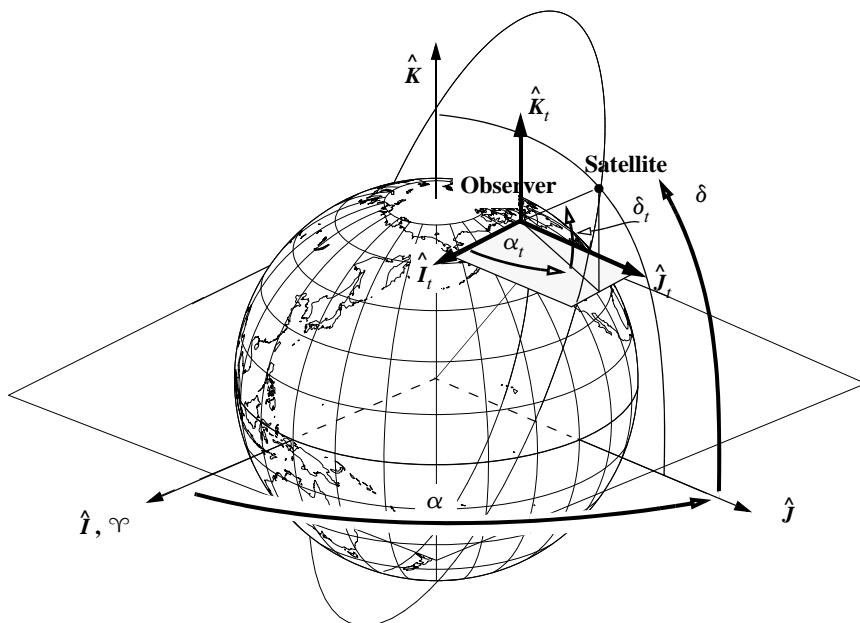


Figure 3-13. Topocentric Equatorial Coordinate System, $I_t J_t K_t$. This system uses the orientation axes of the IJK system, but it's translated to a particular site. The origin rotates with the Earth but maintains the fixed orientation. The difference in numerical values (δ_t and δ) is apparent in the figure. The topocentric declination, δ_t , in this figure is about 20° , whereas the declination, δ , is about 50° .

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 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 normal to the orbit. Figure 3-14 shows the PQW system.

The PQW system 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.

We sometimes use a variant of this system in which the principal axis is in the direction of the orbit's ascending node (where it crosses the equator to the Northern hemisphere)—called the *nodal coordinate system*. Although it permits evaluation of low-eccentricity orbits, it's ill-defined for equatorial orbits.

Satellite Coordinate System, RSW

The *satellite coordinate system* applies to studies of relative motion (see the discussion of relative motion in Sec. 6.8). The system moves with the satellite and is sometimes

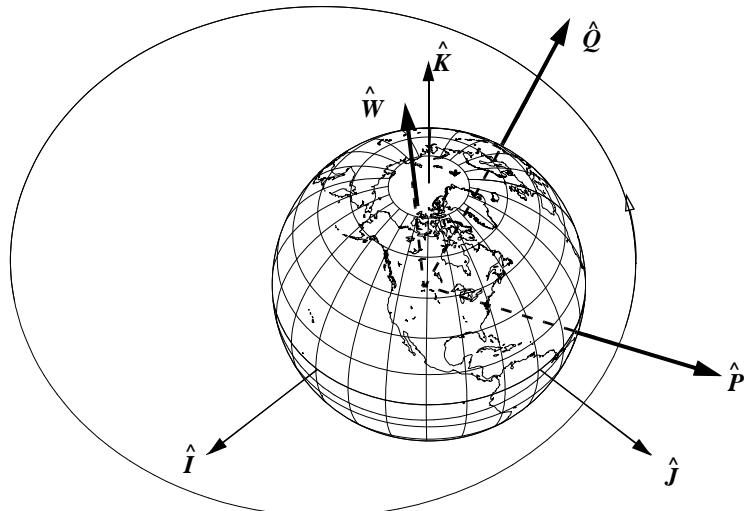


Figure 3-14. Perifocal Coordinate System, PQW. This system points towards perigee. Satellite motion is in the P - Q plane. The W axis is normal to the orbit plane. Although this particular orbit is near equatorial (the J axis is 10° up from the P - Q plane), it doesn't have to be. We often view this system looking down from the W axis.

called the *Gaussian coordinate system* and is sometimes given the letters RTN (radial, transverse, and normal) or LVLH (local vertical, local horizontal). The R axis always points out from the satellite along the Earth's radius vector to the satellite as it moves through the orbit. The S axis points in the direction of (but not necessarily parallel to) the velocity vector and is *perpendicular* to the radius vector—an important additional requirement. The W axis is normal to the orbital plane. Figure 3-15 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. The coordinate system applies to all orbit types.

Relative positions or displacements of satellite orbits use the RSW coordinate system. **Radial** positions and displacements are parallel to the position vector (along the R axis). **Along-track** or **transverse** displacements are normal to the position 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. 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 and are often used to describe orbital errors.

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 *yaw* axis is opposite to the position vector, points toward the center of the Earth, and also lies in the orbital plane. Finally, the

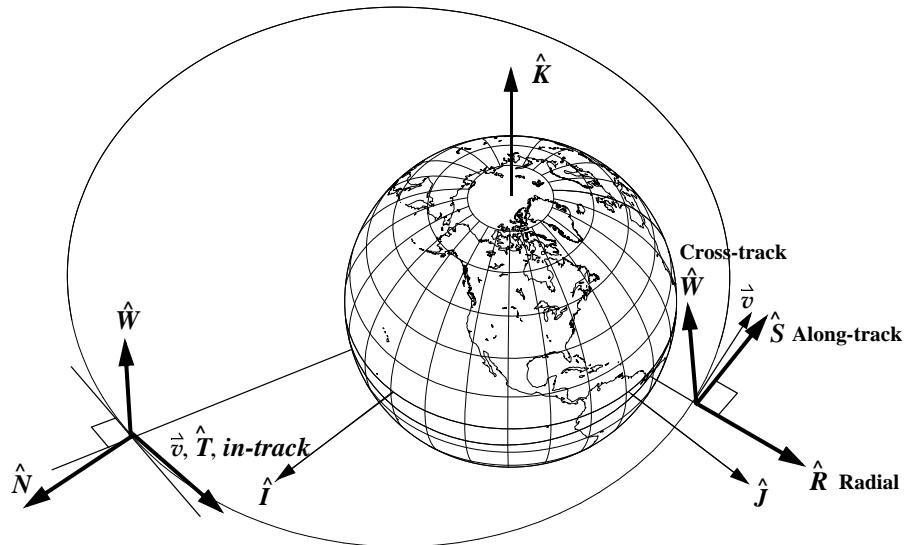


Figure 3-15. Satellite Coordinate Systems, RSW and NTW. These coordinate systems move with the satellite. The R axis points out from the satellite along the geocentric radius vector, the W axis is normal to the orbital plane (*not* usually aligned with the K axis), and the S axis is normal to the position vector and positive in the direction of the velocity vector. The S axis is aligned with the velocity vector *only* for circular orbits. In the NTW system, the T axis is always parallel to 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 at apogee and perigee in elliptical orbits.

pitch axis is opposite to the angular-momentum vector [Eq. (1-15)]; it completes the orthogonal set of vectors. Equivalent axis directions are R —yaw, S —roll, and W —pitch. Many books reference the *roll* axis as being along the velocity vector, but this is true *only* for circular orbits.

Some organizations define the LVLH frame differently placing the primary axis perpendicular to the radius vector, an axis opposite the angular momentum vector, and an axis pointing to the Earth's center. Be sure which convention is assumed!

Satellite Coordinate System, NTW

Figure 3-15 shows this system, which is valid for all orbit types. In this system, the T axis is tangential to the orbit and always points to the velocity vector. The N axis lies in the orbital plane, normal to the velocity vector. The W axis is normal to the orbital plane (as in the RSW system). We define *in-track* or *tangential* 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 to analyze drag effects on the orbit (Sec. 9.6.2) because drag always acts along the velocity vector. It's sometimes referred to as the *Frenet system*.

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. 3-16. We find the principal axis, E , by rotating in the equatorial 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, some 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). Equatorial orbits can still cause difficulties. Remember, when we consider perturbations, the system actually moves over time.

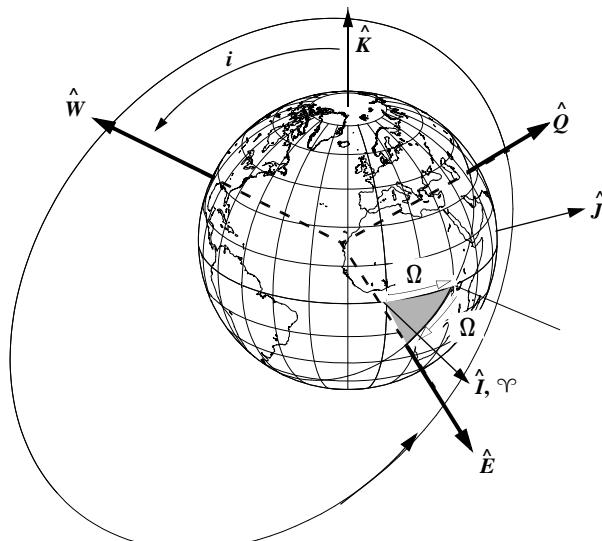


Figure 3-16. Equinoctial System, EQW. This system is formed starting at the IJK system (I axis). Rotating through the right ascension of the ascending node (Ω), tilting through the inclination (i), and then rotating back through Ω places us in the correct orientation and defines the E axis. The Q axis is perpendicular to the E axis in the orbital plane. The W axis is normal to the orbit.

Table 3-1 summarizes the types of coordinate systems.

TABLE 3-1. Summary of Coordinate Systems. With a few exceptions, most coordinate systems are simply variations of the ones in this table. The IJK letters indicate a generic Earth centered coordinate system. I'll use subscripts to indicate variations of the IJK coordinate system. The ITRF is defined for certain years, indicated by the ##—ITRF-94.

System	Symbol	Origin	Fundamental Plane	Principal Direction	Example Use
Interplanetary systems					
Heliocentric	XYZ	Sun	Ecliptic	Vernal equinox	Patched conic
Solar system*	(XYZ) _{ICRF}	Barycenter	varies	varies	Planetary motion
Earth-based systems					
Geocentric [†]	IJK	Earth	Earth equator	Vernal equinox	General
Earth	(IJK) _{GCRF}	Earth	varies	varies	Perturbations
Body-fixed	(IJK) _{ITRF-##}	Earth	Earth equator	Greenwich meridian	Observations
Earth-Moon (synodic)	(IJK) _S	Barycenter	Invariable plane	Earth	Restricted three-body
Topocentric horizon	SEZ	Site	Local horizon	South	Radar observations
Topocentric equatorial	(IJK) _t	Site	Parallel to Earth equator	Vernal equinox	Optical observations
Satellite-based systems					
Perifocal**	PQW	Earth	Satellite orbit	Periapsis	Processing
Satellite radial	RSW ^{††}	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

*The ICRF is our best estimate of an inertial frame located at the barycenter of the solar system. The fundamental plane and the principal direction are defined by the adopted coordinates of extra-galactic VLBI measurements. The GCRF is the geocentric counterpart of the ICRF.

[†]We specify a particular epoch for the IAU-76/FK5, or ECI system. This applies to all frames using Υ .

^{**}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.

3.4 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 (2003). 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 with each problem that needs them because they are simple vector additions and subtractions and are independent of rotations.

3.4.1 Coordinate Rotation

We can express a vector's components in any coordinate system, but often, we must convert between systems. 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 3-17 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** and the rotation is called an Euler angle rotation.

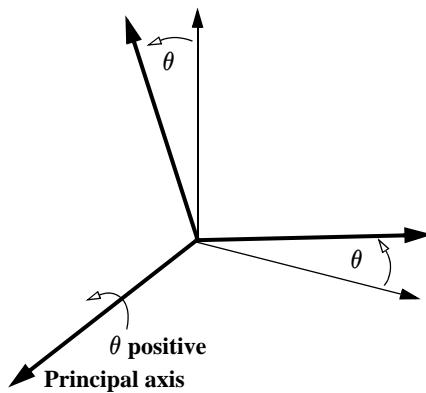


Figure 3-17. Coordinate Rotation. A coordinate rotation is a series of single-axis rotations. The sign of a rotation is considered positive if the right-hand-rule applies, i.e., if the angular motion is about the positive direction of the principal axis, as shown.

I'll look at this process by working 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. 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 3-18 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 (*local*) between the SEZ frame and the IJK frame. This *local* system is aligned with a particular meridian. We do so by rotating about the E axis to align \hat{Z} with the \hat{K}_{local} axis. The negative sign appears because the direction of the motion from \hat{Z} to \hat{K}_{local} opposes the positive direction of the right-hand rule about the E axis.* Mathematically, the idea is to represent each component of the position vector in the intermediate frame. Symbolically, the vector \hat{r}_{local} is

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

* Because we are only expressing a vector in another coordinate frame, we don't address translation here. We'll solve this issue later in Eq. (4-7) and Eq. (7-3).

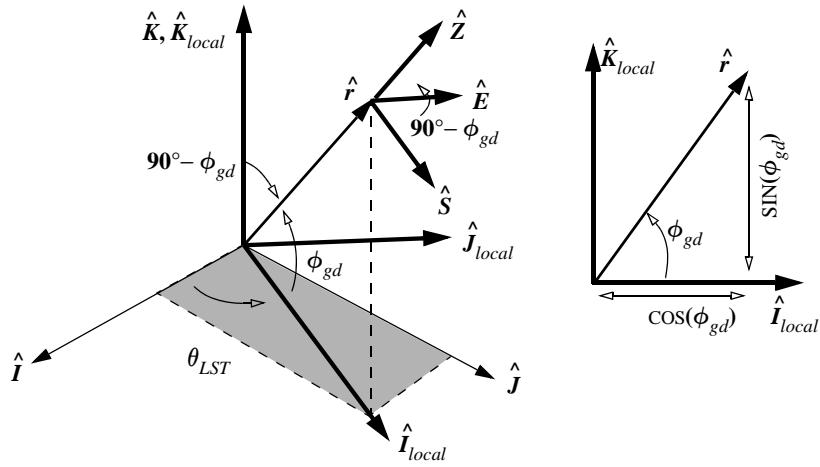


Figure 3-18. Topocentric to Geocentric Rotation #1. The first step in transforming coordinates is to rotate the vector from the SEZ system to an intermediate (*local*) system. The rotation occurs about the *E* axis, as shown, through the angle $(90^\circ - \phi_{gd})$. The right-hand figure shows the I_{local} - K_{local} plane. Depending on your application, you should know which latitude you need. Geodetic latitude is the most common.

where again the unit vector permits a general formula.

Because the K and K_{local} axes are the same, the difficulty is the component that comprises the I_{local} axis direction. Breaking this component into the \hat{I} and \hat{J} axes forces the second step, which is to rotate about the K_{local} axis to align \hat{I}_{local} with \hat{I} . The sign is again due to the direction of motion from \hat{I}_{local} to \hat{I} . Figure 3-19 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}_{local} component,

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

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. (3-15). 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.

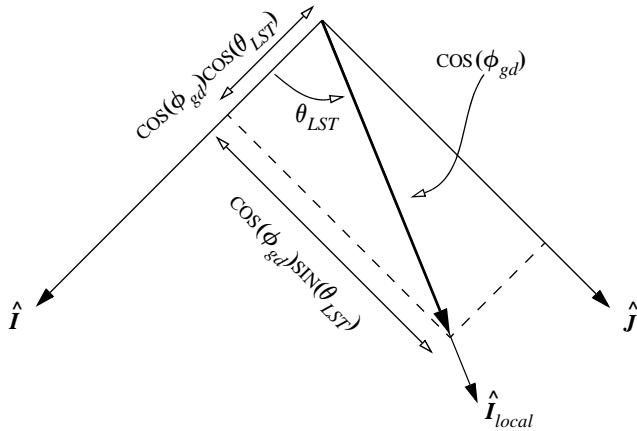


Figure 3-19. Topocentric to Geocentric Rotation #2. The final conversion involves changing from the intermediate system to the IJK system, which requires rotating by θ_{LST} about the K axis.

$$\begin{aligned} \text{ROT1}(\alpha) &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha) & \sin(\alpha) \\ 0 & -\sin(\alpha) & \cos(\alpha) \end{bmatrix} \\ \text{ROT2}(\alpha) &= \begin{bmatrix} \cos(\alpha) & 0 & -\sin(\alpha) \\ 0 & 1 & 0 \\ \sin(\alpha) & 0 & \cos(\alpha) \end{bmatrix} \\ \text{ROT3}(\alpha) &= \begin{bmatrix} \cos(\alpha) & \sin(\alpha) & 0 \\ -\sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{aligned} \quad (3-15)$$

The matrix form is especially useful for computer software; it allows easy processing of multiple rotations. Note that these vectors [Eq. (3-15)] assume a particular convention of rotation—other conventions exist (e.g., Battin, 1987:86). The result for the example is (including the vectors, or 3×1 matrices)

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

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. 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 matrix rotation 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. (3-15) are special. A 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 the transpose, and the transpose is calculated by simply 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,

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

For transformations, the inverse process is simply the reverse order of rotations with negative signs.

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

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 resultant system. 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.

We orient the SEZ coordinate system for a given site with geodetic latitude, ϕ_{gd} , and at a particular local sidereal time, θ_{LST} . For the vector approach, we must first find the site position vector because it defines the direction of \hat{Z} :

$$\begin{aligned}\hat{\vec{r}}_{site} &= r[\cos(\phi_{gd})\cos(\theta_{LST})\hat{I} + \cos(\phi_{gd})\sin(\theta_{LST})\hat{J} + \sin(\phi_{gd})\hat{K}] \\ \hat{Z} &= \frac{\hat{\vec{r}}_{site}}{|\hat{\vec{r}}_{site}|} \quad \hat{E} = \frac{\hat{K} \times \hat{Z}}{|\hat{K} \times \hat{Z}|} \quad \hat{S} = \hat{E} \times \hat{Z}\end{aligned}\tag{3-16}$$

Notice the use of unit vectors eliminates 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 oper-

ation, 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. (3-28).

$$\hat{\vec{r}}_{IJK} = [\hat{S} : \hat{E} : \hat{Z}] \hat{\vec{r}}_{SEZ} \quad (3-17)$$

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

$$\hat{\vec{r}}_{SEZ} = \begin{bmatrix} \hat{S}^T \\ \cdots \\ \hat{E}^T \\ \cdots \\ \hat{Z}^T \end{bmatrix} \hat{\vec{r}}_{IJK} \quad (3-18)$$

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 have derived 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.

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

The actual formula is

$$\hat{\vec{r}}_{IJK} = [\hat{P} : \hat{Q} : \hat{W}] \hat{\vec{r}}_{PQW} \quad (3-19)$$

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

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

which allow the transformation

$$\hat{\vec{r}}_{IJK} = [\hat{R} : \hat{S} : \hat{W}] \hat{\vec{r}}_{RSW} \quad (3-20)$$

The NTW coordinate system has the following equations:

$$\hat{T} = \frac{\hat{\vec{v}}}{|\hat{\vec{v}}|} \quad \hat{W} = \frac{\hat{\vec{r}} \times \hat{\vec{v}}}{|\hat{\vec{r}} \times \hat{\vec{v}}|} \quad \hat{N} = \hat{T} \times \hat{W} \quad (3-21)$$

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 function evaluations 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.

Rotations Using Quaternions

Quaternions are used primarily with satellite attitude determination. Satellite attitude can be very important when determining the cross-sectional area of a satellite exposed to atmospheric drag, or solar radiation pressure. In general, a ***quaternion*** is a four-dimension vector that describes the rotation of one coordinate frame to another. There are two conventions, depending on which element contains the magnitude. I'll use the convention where the magnitude is stored in the fourth component. Thus,

$$\mathbf{q} = [q_1, q_2, q_3, q_4]$$

where q_4 is the magnitude.

Although the use of quaternions has declined somewhat over the years as computers have become faster, they are very fast and are an attractive alternative to Euler angles for many applications. Two basic operations are of interest.

Finding a quaternion given two vectors (\vec{r}_1 and \vec{r}_2) expressed in different coordinate frames. Once this quaternion is found, you can transform additional vectors by simple quaternion multiplication.

$$\begin{aligned}\hat{\vec{v}}_c &= \vec{r}_1 \times \vec{r}_2 \\ \hat{\vec{v}}_{nc} &= \frac{\hat{\vec{v}}_c}{|\hat{\vec{v}}_c|} \\ c &= \frac{\vec{r}_1 \cdot \vec{r}_2}{|\vec{r}_1||\vec{r}_2|} \\ s &= -\sqrt{\frac{1-c}{2}} \\ q_i &= \hat{\vec{v}}_{nc}s \quad i = 1 \dots 3 \\ q_4 &= -s\end{aligned}\tag{3-22}$$

Multiplying a vector by a quaternion. This operation quickly finds the vector in an alternate coordinate frame. Note there are two directions ($+q_4$ and $-q_4$)

$$\begin{aligned}t_1 &= q_3 v_{in}(2) - q_2 v_{in}(3) \\ t_2 &= q_1 v_{in}(3) - q_3 v_{in}(1) \\ t_3 &= q_2 v_{in}(1) - q_1 v_{in}(2) \\ v_{out}(1) &= v_{in}(1) + 2(t_1 q_4 + t_2 q_3 - t_3 q_2)\end{aligned}\tag{3-23}$$

$$v_{out}(2) = v_{in}(2) + 2(t_2 q_4 + t_3 q_1 - t_1 q_3)$$

$$v_{out}(3) = v_{in}(3) + 2(t_3 q_4 + t_1 q_2 - t_2 q_1)$$

There are two other common operations—that of converting a quaternion to euler angles or direction cosines, and the reverses (See also Wertz, 1978:758 and Shuster 2008 for details).

3.4.2 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 any motion of the frames. The terms *inertial* and *rotating* specify the stability or motion (respectively) 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

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

Of course $\hat{r}_{Inertial} = \hat{r}_{Rot}$ 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 can translate the origin, \hat{r}_{org} , so both frames coincide for complete generality. Differentiation permits analysis in each coordinate system as follows:

$$\begin{aligned} \hat{v}_{Inertial} &= \dot{\eta}_1 \hat{I} + \dot{\eta}_2 \hat{J} + \dot{\eta}_3 \hat{K} & \hat{v}_{Rot} &= \dot{\xi}_1 \hat{I}' + \dot{\xi}_2 \hat{J}' + \dot{\xi}_3 \hat{K}' \\ \hat{a}_{Inertial} &= \ddot{\eta}_1 \hat{I} + \ddot{\eta}_2 \hat{J} + \ddot{\eta}_3 \hat{K} & \hat{a}_{Rot} &= \ddot{\xi}_1 \hat{I}' + \ddot{\xi}_2 \hat{J}' + \ddot{\xi}_3 \hat{K}' \end{aligned} \quad (3-24)$$

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,

$$\hat{v}_{Inertial} = \dot{\xi}_1 \hat{I}' + \dot{\xi}_2 \hat{J}' + \dot{\xi}_3 \hat{K}' + \xi_1 \dot{\hat{I}}' + \xi_2 \dot{\hat{J}}' + \xi_3 \dot{\hat{K}}'$$

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.

$$\dot{\vec{v}}_{Inertial} = \dot{\vec{v}}_{Rot} + \dot{\vec{\omega}}_{R/I} \times \dot{\vec{r}}_{Rot} + \dot{\vec{v}}_{org} \quad (3-25)$$

where $\dot{\vec{\omega}}_{R/I}$ 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

$$\ddot{\vec{a}}_{Inertial} = \ddot{\vec{a}}_{Rot} + \dot{\vec{\omega}}_{R/I} \times \dot{\vec{r}}_{Rot} + \dot{\vec{\omega}}_{R/I} \times (\dot{\vec{\omega}}_{R/I} \times \dot{\vec{r}}_{Rot}) + 2\dot{\vec{\omega}}_{R/I} \times \dot{\vec{v}}_{Rot} + \ddot{\vec{a}}_{org} \quad (3-26)$$

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 ($\dot{\vec{\omega}}_{R/I} \times \dot{\vec{r}}_{Rot}$) is called the *tangential acceleration* because ω changes. For the satellite problems, it's zero only for a true circular orbit. Next are the *centripetal acceleration*, ($\dot{\vec{\omega}}_{R/I} \times (\dot{\vec{\omega}}_{R/I} \times \dot{\vec{r}}_{Rot})$), and the *Coriolis acceleration*, ($2\dot{\vec{\omega}}_{R/I} \times \dot{\vec{v}}_{Rot}$). A final term covers an accelerating origin.

Another method 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,

$$\begin{aligned} \dot{\vec{r}}_{Inertial} &= [\text{ROT}] \dot{\vec{r}}_{SEZr} \\ \dot{\vec{v}}_{Inertial} &= [\text{ROT}] \dot{\vec{v}}_{SEZr} + [\dot{\text{ROT}}] \dot{\vec{r}}_{SEZr} \\ \ddot{\vec{a}}_{Inertial} &= [\text{ROT}] \ddot{\vec{a}}_{SEZr} + 2[\dot{\text{ROT}}] \dot{\vec{v}}_{SEZr} + [\ddot{\text{ROT}}] \dot{\vec{r}}_{SEZr} \end{aligned} \quad (3-27)$$

Here, the derivatives of the rotation matrices transform the vector to the inertial system and account for centripetal and Coriolis effects. We often use the first method of differentiating vectors because we don't have to determine the rates of change of the transformation matrix. Both notations allow you to work with any system.

3.4.3 Common Transformations

Because the geocentric equatorial reference frame, IJK, is most common, the transformations listed here often relate to this system. However, precise applications must specify inertial or Earth-fixed vectors (Sec. 3.7), so I've included these distinctions in this section. The following equations highlight some of these transformations—in different forms. Special care is sometimes required for the velocity vectors because the acceleration and velocity of rotating frames can affect the magnitude and orientation of velocity vectors. Also, systems using a specific site often require a translation of the origin. Sec. 2.4 covers orbital elements included in the transformations. I've included complete rota-

tion 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

$$\text{final} = \begin{bmatrix} \text{final} \\ \text{initial} \end{bmatrix} \text{initial}$$

Transformations between SEZ and IJK (ECEF): We must include the translation between the coordinate systems. Note that $\vec{\rho}$, $\dot{\vec{\rho}}$ are the slant range vectors from a site to the satellite.

$$\begin{aligned} \vec{\rho}_{ECEF} &= [\text{ROT3}(-\lambda)][\text{ROT2}(-(90^\circ - \phi_{gd}))]\vec{\rho}_{SEZ} = \left[\frac{ECEF}{SEZ} \right] \vec{\rho}_{SEZ} \\ \dot{\vec{\rho}}_{ECEF} &= [\text{ROT3}(-\lambda)][\text{ROT2}(-(90^\circ - \phi_{gd}))]\dot{\vec{\rho}}_{SEZ} = \left[\frac{ECEF}{SEZ} \right] \dot{\vec{\rho}}_{SEZ} \\ \left[\frac{ECEF}{SEZ} \right] &= \begin{bmatrix} \sin(\phi_{gd})\cos(\lambda) & -\sin(\lambda) & \cos(\phi_{gd})\cos(\lambda) \\ \sin(\phi_{gd})\sin(\lambda) & \cos(\lambda) & \cos(\phi_{gd})\sin(\lambda) \\ -\cos(\phi_{gd}) & 0 & \sin(\phi_{gd}) \end{bmatrix} \quad (3-28) \\ \vec{r}_{ECEF} &= \vec{\rho}_{ECEF} + \vec{r}_{siteECEF} \\ \dot{\vec{v}}_{ECEF} &= \dot{\vec{\rho}}_{ECEF} \end{aligned}$$

Transformations between $I_t J_t K_t$ and IJK. Notice that no rotations are required because the two systems are parallel, but we must account for the translation of the origin.

$$\begin{aligned} \vec{r}_{IJK} &= \vec{r}_{I_t J_t K_t} + \vec{r}_{siteIJK} & \vec{v}_{IJK} &= \vec{v}_{I_t J_t K_t} + \vec{v}_{siteIJK} \\ \vec{r}_{I_t J_t K_t} &= \vec{r}_{IJK} - \vec{r}_{siteIJK} & \dot{\vec{v}}_{I_t J_t K_t} &= \dot{\vec{v}}_{IJK} - \dot{\vec{v}}_{siteIJK} \end{aligned} \quad (3-29)$$

Transformations between PQW and IJK (ECI):

$$\begin{aligned} \vec{r}_{ECI} &= \text{ROT3}(-\Omega)\text{ROT1}(-i)\text{ROT3}(-\omega)\vec{r}_{PQW} \\ \vec{r}_{PQW} &= \text{ROT3}(\omega)\text{ROT1}(i)\text{ROT3}(\Omega)\vec{r}_{ECI} \end{aligned} \quad (3-30)$$

$$\left[\frac{ECI}{PQW} \right] = \begin{bmatrix} \cos(\Omega)\cos(\omega) - \sin(\Omega)\sin(\omega)\cos i & -\cos(\Omega)\sin(\omega) - \sin(\Omega)\cos(\omega)\cos(i) & \sin(\Omega)\sin(i) \\ \sin(\Omega)\cos(\omega) + \cos(\Omega)\sin(\omega)\cos i & -\sin(\Omega)\sin(\omega) + \cos(\Omega)\cos(\omega)\cos(i) & -\cos(\Omega)\sin(i) \\ \sin(\omega)\sin(i) & \cos(\omega)\sin(i) & \cos(i) \end{bmatrix}$$

Transformations between PQW and RSW:

$$\begin{aligned}
 \dot{\vec{r}}_{RSW} &= \text{ROT3}(\nu) \dot{\vec{r}}_{PQW} \\
 \dot{\vec{r}}_{PQW} &= \text{ROT3}(-\nu) \dot{\vec{r}}_{RSW} \\
 [\frac{RSW}{PQW}] &= \begin{bmatrix} \cos(\nu) & \sin(\nu) & 0 \\ -\sin(\nu) & \cos(\nu) & 0 \\ 0 & 0 & 1 \end{bmatrix}
 \end{aligned} \tag{3-31}$$

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

$$\begin{aligned}
 \dot{\vec{r}}_{EQW} &= \text{ROT3}(-f_r\Omega) \text{ROT1}(i) \text{ROT3}(\Omega) \dot{\vec{r}}_{ECI} \\
 \dot{\vec{r}}_{ECI} &= \text{ROT3}(-\Omega) \text{ROT1}(-i) \text{ROT3}(f_r\Omega) \dot{\vec{r}}_{EQW}
 \end{aligned} \tag{3-32}$$

$$\begin{bmatrix} \frac{EQW}{ECI} \end{bmatrix} = \begin{bmatrix} \cos^2(\Omega) + f_r \cos(i) \sin^2(\Omega) & \cos(\Omega) \sin(\Omega) \{1 - f_r \cos(i)\} & -f_r \sin(i) \sin(\Omega) \\ f_r \cos(\Omega) \sin(\Omega) \{1 - f_r \cos(i)\} & f_r \sin(\Omega)^2 + \cos(i) \cos^2(\Omega) & \sin(i) \cos(\Omega) \\ \sin(i) \sin(\Omega) & -\sin(i) \cos(\Omega) & \cos(i) \end{bmatrix}$$

The next three transformations are illustrative only, so I won't show combined matrices.

Transformations between NTW and IJK (ECI):

$$\dot{\vec{r}}_{ECI} = \text{ROT3}(-\Omega) \text{ROT1}(-i) \text{ROT3}(-u) \text{ROT3}(\phi_{fpa}) \dot{\vec{r}}_{NTW} \tag{3-33}$$

Transformations between RSW and IJK (ECI):

$$\dot{\vec{r}}_{ECI} = \text{ROT3}(-\Omega) \text{ROT1}(-i) \text{ROT3}(-u) \dot{\vec{r}}_{RSW} \tag{3-34}$$

Transformations between SEZ and satellite body:

$$\dot{\vec{r}}_{Body} = \text{ROT3}(Yaw) \text{ROT1}(Roll) \text{ROT2}(Pitch) \dot{\vec{r}}_{SEZ} \tag{3-35}$$

The last few examples point out some problems with coordinate transformations. Recall 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. The angles in the last three transformations aren't all independent. Although these series of angles help us see the transformations between frames, 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. They are often used with satellite attitude determination.

3.4.4 Application: Converting IJK (ECEF) To Latitude and Longitude

A procedure to convert a position vector for a satellite to the corresponding latitude and longitude is the core technique in determining ground tracks (Sec. 2.7). For this problem, we know the orbital elements of the satellite's position vector in the geocentric equatorial system (Earth fixed) but must find the geocentric or geodetic latitude, the lon-

gitude, and the height of the satellite above the reference ellipsoid. The difference between geocentric and geodetic latitude can cause errors up to about 20 km. This procedure is also a basis for surveillance calculations, which I'll discuss in Chap. 11.

We have two ways to do the transformation of position to latitude and longitude: one iterative and one analytical. Figure 3-20 below shows the situation. If the Earth were a perfect sphere, the algebraic relationship between spherical and cartesian coordinates would yield a direct solution. 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. 3-20 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. First, we find the right ascension through sine and cosine expressions:

$$\sin(\alpha) = \frac{r_J}{r_{\delta_{sat}}} \quad \cos(\alpha) = \frac{r_I}{r_{\delta_{sat}}}$$

Because the vector is Earth fixed, the right ascension is also equal to the longitude. In Sec 3.7 we'll discuss the reduction from ECEF to ECI. Some references show $\lambda = \alpha - \theta_{GMST}$, however this is only an approximation to the rigorous techniques of Sec 3.7.

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,

$$\sin(\delta) = \frac{r_{K_{sat}}}{r} \quad (3-36)$$

At this point, we can use various techniques to perform the conversion. 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 coordinates. Assume for the first guess that $\phi_{gd} \approx \delta$. Using Eq. (3-7), find the sine and cosine expressions:

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

Solving the sine expression for h_{ellp} gives you

$$h_{ellp} = \frac{r_K}{\sin(\phi_{gd})} - S_{\oplus}$$

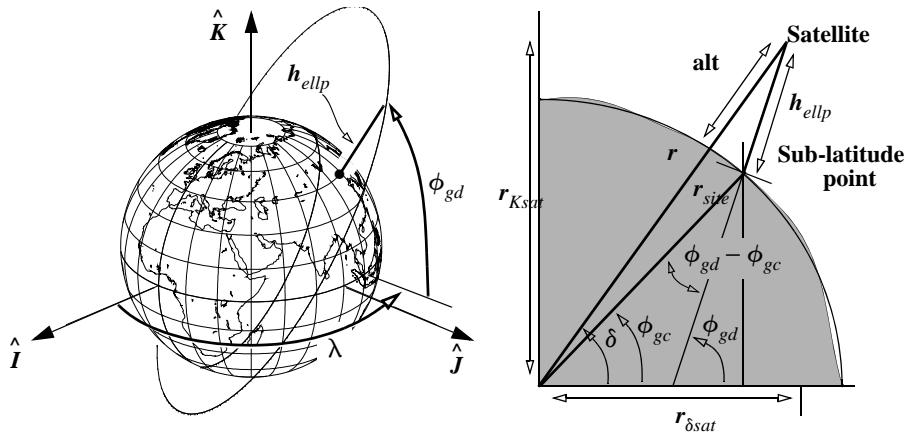


Figure 3-20. Determining a Satellite's Sub-latitude Point. Notice h_{ellp} is used for the altitude, and not H_{MSL} . The difference between the geodetic and geocentric angles is helpful in setting up the iteration for the problem.

Now determine a tangent expression and substitute h_{ellp} to get

$$\tan(\phi_{gd}) = \frac{\sin(\phi_{gd})}{\cos(\phi_{gd})} = \frac{r_K \left(C_{\oplus} + \frac{r_K}{\sin(\phi_{gd})} - S_{\oplus} \right)}{\left(S_{\oplus} + \frac{r_K}{\sin(\phi_{gd})} - C_{\oplus} \right) r_{\delta}}$$

Also remember from Eq. (3-7) 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:

$$\begin{aligned}\text{TAN}(\phi_{gd}) &= \frac{r_K(1 - e_\oplus^2) + (1 - e_\oplus^2)e_\oplus^2 C_\oplus \sin(\phi_{gd})}{r_\delta(1 - e_\oplus^2)} \\ \text{TAN}(\phi_{gd}) &= \frac{r_K + e_\oplus^2 C_\oplus \sin(\phi_{gd})}{r_\delta}\end{aligned}$$

The notation in this section is specific. We have a solution by rearranging equations, but for this particular problem, we don't have the site coordinates (r_δ and r_K). Careful examination of Fig. 3-20 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 $r_{K_{sat}}$ in the formula. We find the height above mean sea level by solving Eq. (3-7) for h_{ellp} . The algorithm appears below.

ALGORITHM 12: ECEF TO LATLON ($\vec{r}_{ECEF} \Rightarrow \phi_{gd}, \lambda, h_{ellp}$)

$$r_{\delta_{sat}} = \sqrt{r_I^2 + r_J^2}$$

$$\text{SIN}(\alpha) = \frac{r_J}{r_{\delta_{sat}}} \quad \text{COS}(\alpha) = \frac{r_I}{r_{\delta_{sat}}}$$

$$\lambda = \alpha$$

$$\text{SIN}(\delta) = \frac{r_{K_{sat}}}{r} \quad \text{or} \quad \text{TAN}(\delta) = \frac{r_{K_{sat}}}{r_{\delta_{sat}}}$$

$$\text{Let } \phi_{gd} = \delta \quad r_\delta = r_{\delta_{sat}} \quad r_K = r_{K_{sat}}$$

LOOP

$$C_\oplus = \frac{R_\oplus}{\sqrt{1 - e_\oplus^2 \text{SIN}^2(\phi_{gd})}}$$

$$\text{TAN}(\phi_{gd}) = \frac{r_K + C_\oplus e_\oplus^2 \text{SIN}(\phi_{gd})}{r_\delta}$$

UNTIL $\phi_{gd} - \phi_{gd_{old}} < \text{Tolerance}$

$$h_{ellp} = \frac{r_\delta}{\text{COS}(\phi_{gd})} - C_\oplus \quad \text{if near poles } (\sim 1^\circ) \quad h_{ellp} = \frac{r_K}{\text{SIN}(\phi_{gd})} - S_\oplus$$

This method often requires an extra iteration over that of Escobal, but the formula's simplicity actually results in quicker execution. Sensor coordinates are often given in ECEF coordinates. To find the latitude and longitude, be sure to use the same R_\oplus and e_\oplus .

▼ Example 3-3. Converting ECEF to Lat Lon.

GIVEN: $\vec{r}_{ECEF} = 6524.834 \hat{I} + 6862.875 \hat{J} + 6448.296 \hat{K}$ km

FIND: Latitude and longitude

$$r_{\delta_{sat}} = \sqrt{r_I^2 + r_J^2} = 9469.6 \text{ km}$$

$$\sin(\alpha) = \frac{r_J}{r_{\delta_{sat}}} \quad \alpha = 46.4464^\circ \quad \tan(\delta) = \frac{r_K}{r_{\delta_{sat}}} \quad \delta = 34.2527^\circ$$

The right ascension requires a quadrant check (using the cosine). It's correct as shown. Determine the longitude without iteration:

$$\lambda = \alpha = 46.4464^\circ$$

Now the iteration begins. Three iterations are shown:

$$C_{\oplus} = \frac{R_{\oplus}}{\sqrt{1 - e_{\oplus}^2 \sin^2(\phi_{gd})}} = 6384.911\,021, 6384.945\,592, 6384.945\,681 \text{ km}$$

$$\tan(\phi_{gd}) = \frac{r_K + C_{\oplus} e_{\oplus}^2 \sin(\phi_{gd})}{r_{\delta}} \Rightarrow \phi_{gd} = 34.352\,242, 34.352\,495, 34.352\,496$$

Finally, the geodetic latitude is

$$\phi_{gd} = 34.352\,496^\circ$$

You can use Eq. (3-11) to find the geocentric latitude: $\phi_{gc} = 34.173\,429^\circ$. The height is

▲ $h_{ellp} = \frac{r_{\delta}}{\cos(\phi_{gd})} - C_{\oplus} = \frac{1.4847}{\cos(34.352\,496)} - 6384.945\,681 = 5085.22 \text{ km}$

A direct method provides the geodetic latitude and ellipsoidal height but requires no iteration. The method is from Borkowski (1989) and it's given in the following algorithm. Essentially, the procedure relies on a fourth-degree polynomial to evaluate the results. Although this method provides only the geodetic latitude, we can find the geocentric latitude in one step using Eq. (3-11). Seidelmann (1992:206–207) shows the algorithm. I use r_b as the polar radius and e_b as the eccentricity of an arbitrary body.

ALGORITHM 13: ECEF TO LATLON ($\vec{r}_{ECEF} \Rightarrow \phi_{gd}, \lambda, h_{ellp}$)

$$r_{\delta_{sat}} = \sqrt{r_I^2 + r_J^2}$$

$$a = R_{\oplus}$$

$$b = \sqrt{r_b(1 - e_b^2)} \text{SIGN}(r_K)$$

$$E = \frac{br_K - (a^2 - b^2)}{ar_{\delta_{sat}}}$$

$$\begin{aligned}
\sin(\alpha) &= \frac{r_J}{r_{\delta_{sat}}} \\
\cos(\alpha) &= \frac{r_I}{r_{\delta_{sat}}} \\
\lambda &= \alpha \\
F &= \frac{br_K + (a^2 - b^2)}{ar_{\delta_{sat}}} \\
P &= \frac{4(EF + 1)}{3} \\
Q &= 2(E^2 - F^2) \\
D &= P^3 + Q^2 \\
&\text{IF } (D > 0) \\
\nu &= \sqrt[3]{\sqrt{D} - Q} - \sqrt[3]{\sqrt{D} + Q} \\
&\text{IF } (D < 0) \\
\nu &= 2\sqrt{-P} \cos\left(\frac{1}{3} \cos^{-1}\left(\frac{Q}{P\sqrt{-P}}\right)\right) \\
G &= \frac{1}{2}(\sqrt{E^2 + \nu} + E) \\
t &= \sqrt{G^2 + \frac{F - \nu G}{2G - E}} - G \\
\tan(\phi_{gd}) &= \frac{a(1 - t^2)}{2bt} \\
h_{ellp} &= (r_{\delta_{sat}} - at) \cos(\phi_{gd}) + (r_K - b) \sin(\phi_{gd})
\end{aligned}$$

3.5 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 origin.

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 phe-

nomenon 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 scales now provide timekeeping for scientific, engineering, and general purposes: sidereal time, solar (universal time), dynamical time, and atomic time. Sidereal time and solar time are based on the Earth's rotation and are related through mathematical relationships. Dynamical and atomic time are independent from the other forms.

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. (3-37). Time divides into hours (h), minutes (m), and seconds (s); angular measure has three similar distinctions: degrees ($^\circ$), arcminutes ($'$), and arcseconds ($''$). Although the names are similar, the symbols and magnitudes aren't:

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

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

$$\begin{array}{ll} 1^h = 15^\circ & 1^\circ = \frac{1^h}{15} = 4^m \\ 1^m = 15' & 1' = \frac{1^m}{15} = 4'' \\ 1^s = 15'' & 1'' = \frac{1^s}{15} \end{array} \quad (3-37)$$

In modern times, the second has replaced the day as the basic interval. Ephemeris time (the precursor to Universal 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.

More than anything else, the Sun governs our daily activity, and because its motion is apparently regular, we base solar time on that motion. ***Solar time*** is loosely defined by successive transits of the Sun over a local meridian. 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 3-21 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.

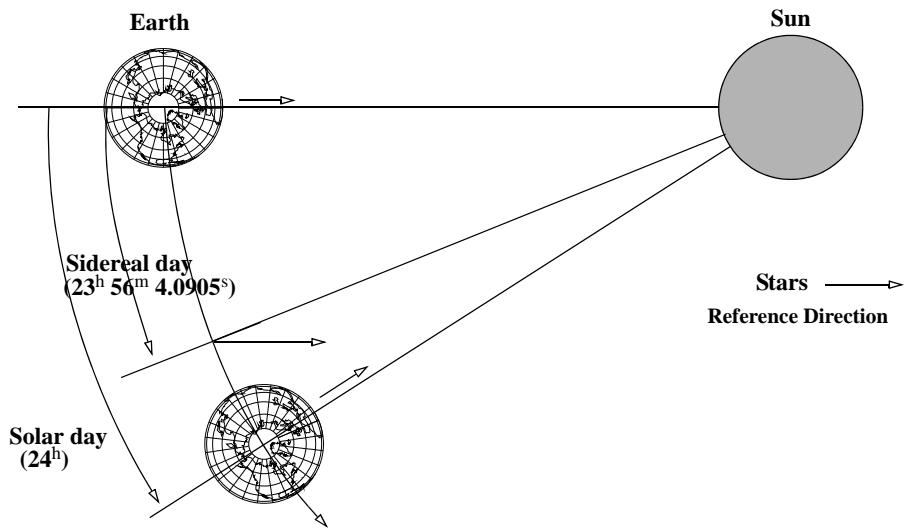


Figure 3-21. Measuring Solar and Sidereal Time. Distance and angular motion are greatly exaggerated for clarity. One solar day is the time required for an observer on the Earth to revolve once *and* observe the Sun at the same location. A sidereal day uses observations of the stars instead of the Sun. Thus, the sidereal day is slightly shorter than the solar day because the stars remain essentially in the same location over time.

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.

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. 3-21).

However, other irregularities in the Sun's apparent 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 now 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 1930's. These variations demonstrated the need for a more stable reference for time intervals. Thus, other time systems have been developed to provide more accurate timing measurements.

We can also measure time by the 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 object's motion. By observing the motion of the body from one point in time to another, we deduce the elapsed time from the mathematical description of the motion. The precise determination of dynamical time requires relativistic corrections in modeling the Earth's motion.

Atomic time is the 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.

3.5.1 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 solar day. This concept has served adequately for ages. The Sun's apparent motion results from a combination of 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 (as viewed from the Earth). 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 when projected on 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. 3-10).

Apparent Solar Time

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

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

The 12-hour increment ensures that 0^{h} corresponds to midnight and 12^{h} to noon, respectively. Greenwich apparent solar time is formed similarly ($\text{GHA}_{\odot} + 12^{\text{h}}$). 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 reference the Sun (See Fig. 3-9). Now if we incorporate

the right ascension of the Sun:

$$\text{Greenwich apparent solar time} = \theta_{GMST} - \alpha_{\odot} + 12^{\text{h}} \quad (3-38)$$

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

A *Fictitious Mean Sun* was 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 now

$$\alpha_{FMS} = 18^{\text{h}} 38^{\text{m}} 45.836^{\text{s}} + 8,640,184.542^{\text{s}} T + 0.0929^{\text{s}} T^2$$

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. (3-38) 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 GMST (Sec. 3.5.2) and corrections for the Earth's rotation. **Mean solar time** is derived from measurements of the Earth orientation. These measurements allow GMST to be found, and then mean solar time is calculated through a numerical formula (Seidelmann, 1992:50). 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. 5.1.1:

$$\begin{aligned} EQ_{time} = & -1.914\,666\,471^{\circ} \sin(M_{\odot}) - 0.019\,994\,643 \sin(2M_{\odot}) \\ & + 2.466 \sin(2\lambda_{ecliptic}) - 0.0053 \sin(4\lambda_{ecliptic}) \end{aligned} \quad (3-39)$$

During the year, the difference in apparent and mean solar time varies from minus 14 minutes to plus 16 minutes (Seidelmann, 1992:6). Figure 3-22 shows the variations.

There are three distinct realizations of UT. For precise applications, we must distinguish between UT0, UT1, and UT2. The differences are small. *Greenwich mean time*, (GMT), 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 found by reducing the observations of stars from many ground stations. Determining UT0 depends on the apparent diurnal motion of the stars.

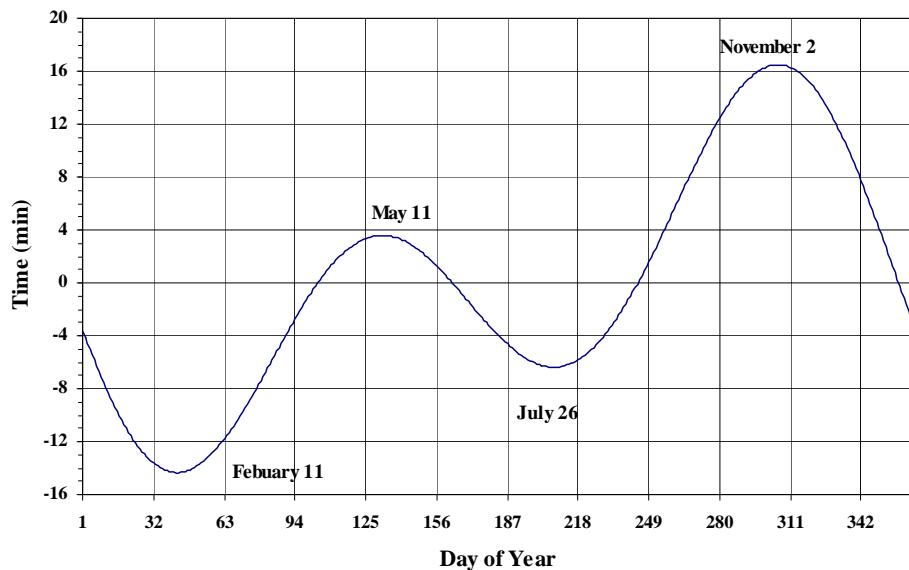


Figure 3-22. Equation of Time Variation. The difference in mean and apparent (or true) solar time varies throughout the year from about plus 16 minutes, to about minus 14 minutes.

At first this seems to contradict the concept of solar time already introduced; however, instruments can't measure 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 12:00 P.M. local time (Green, 1988:37). Supposing $GHA = 2$ hours, UT0 is then 2:00 P.M., 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 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 angles describe the instantaneous positions of the pole and are discussed on page 207. 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

(0.030^s). Finally, UT1 can be corrected for seasonal variations to yield **UT2**. UT2 is considered obsolete.

Converting between Solar Time and Sidereal Time

Recall Fig. 3-21 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).

$$\begin{aligned} 1 \text{ solar day} &= 1.002\,737\,909\,350\,795 \text{ sidereal day} \\ 1 \text{ sidereal day} &= 0.997\,269\,566\,329\,084 \text{ solar day} \\ 1 \text{ solar day} &= 24^h\,3^m\,56.555\,367\,8^s \text{ sidereal time} \\ 1 \text{ sidereal day} &= 23^h\,56^m\,4.090\,524^s \text{ solar time} \end{aligned}$$

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

$$\omega_{\oplus_{prec}} = 1 \text{ solar day} = 1.002\,737\,909\,350\,795 + 5.9006 \times 10^{-11} T_{UT1} - 5.9 \times 10^{-15} T_{UT1}^2 \text{ sidereal day} \quad (3-40)$$

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. The units are rev/day. I use the multiple subscripts because the value also approximates the mean rotation of the Earth with respect to a precessing equinox. The length of the solar day increases slightly—an effect of tidal acceleration.

Coordinated Universal Time

The most commonly used time system is **Coordinated Universal Time**, UTC, which is derived from an ensemble of atomic clocks. It's designed to follow UT1 within $\pm 0.9^s$ ($\Delta UT1 = UT1 - UTC$). UTC is the basis of civil time systems and is on ordinary clocks. This definition of UTC was introduced in January, 1972, as a convenient approximation of 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 insert **leap seconds** into UTC to keep the two time scales in close agreement. This minimizes navigation errors (when using $UTC = UT1$) and maximizes the time between leap second insertions. 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 (Sec. 3.5.3). 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 changed in 1972 when the current UTC was adopted.

For civilian timekeeping, we define *local time** as offsets from UTC. This creates time zones for particular regions. 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 3-2 lists common time zones in the United States and their conversion to UTC.[†] Be aware that UTC does *not* change for daylight savings time. Many countries use *Daylight Savings Time*, DST, to adjust the work hours within the available daylight time throughout the year. Benjamin Franklin (1706-1790) suggested the concept of DST, but the U.S. didn't use it until World War I to save energy used to produce electricity. Today, DST begins on the second Sunday in March in the United States, and ends on the first Sunday in November.^{**} The actual switch occurs at 2:00 A.M. local time. To determine the date for DST in a year, we use a method that finds the *day of the week*. For any JD (at noon), the day of the week is

$$DW = \text{TRUNC}\left(JD - 7\text{TRUNC}\left(\frac{JD+1}{7}\right) + 2\right) \quad (3-41)$$

The day of the week is an integer that ranges from 1 (Sunday) to 7 (Saturday). We can use Eq. (3-41) to search for the beginning day in April and the ending day in October, to find the interval of DST in a given year.

GPS Time

Satellite often use GPS time. This time system began with the introduction of the GPS system operational capability in 1980—the GPS epoch is January 6, 1980, 0 hr UTC. GPS is not adjusted so it differs from UTC by the number of leap seconds. GPS time is steered on a daily basis to keep it within one microsecond of UTC, although it is usually much closer.

* WWV is the radio call sign for a short-wave radio station in Fort Collins, Colorado that broadcasts timing information on 2.5, 5, 10, 15, and 20 MHz channels. The station also broadcasts at 60 kHz, on station WWVB. There are similar transmissions on WWVH in Kauai, Hawaii.

* Local time is sometimes abbreviated LST for Local Solar Time, however, this could be confused with Local Sidereal Time, so I'll always use LT for local solar time.

† Time zones were originally established by the railroads on November 18, 1883 for the United States and Canada. Congress made these definitions into law in 1918, and in the late 1960's, they transferred authority for the determination of time zones to the U.S. Department of Transportation (DOT). The principal standard for accepting a time zone change is the “convenience of commerce.” There are 9 time zones in the U.S. today.

** The first Congressional bill introducing DST was signed in 1918. However, complaints caused its repeal just the next year. Its use was sporadic from WWI to WWII and beyond. In 1966, Congress established the starting and ending dates for DST as the last Sunday in April, and the last Sunday in October. However by 1986, summer industries persuaded Congress to change the start date to the first Sunday in April. In 2007, DST was changed to go from the second Sunday in March to the first Sunday in November. Other countries use similar, but not identical conventions.

TABLE 3-2. UTC Offsets for Selected Time Zones. The central meridian defines the difference in UTC from the Greenwich meridian. Daylight Savings Time (DST) runs from April to October. Some countries don't use DST (column is left blank). Central meridian values are approximate.

Time Zone	Standard	Daylight Savings	Central Meridian
Atlantic	UTC - 4 ^h	UTC - 3 ^h	-60°
Eastern	UTC - 5 ^h	UTC - 4 ^h	-75°
Mountain	UTC - 7 ^h	UTC - 6 ^h	-105°
Alaska	UTC - 9 ^h	UTC - 8 ^h	-135°
Hawaii	UTC - 10 ^h		-150°
Central European	UTC + 1 ^h	UTC + 2 ^h	10°
Moscow	UTC + 4 ^h		40°
China	UTC + 8 ^h		100°
Japan	UTC + 9 ^h		140°
Australian Eastern	UTC + 10 ^h	UTC + 11 ^h	146°

Julian Date

An essential concept in astrodynamics is the Julian date. The **Julian date**, JD, is the interval of time measured in days from the epoch January 1, 4713 B.C., 12:00.* Joseph Scaliger made the actual determination 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 14 (*use 61 seconds if the day contains a leap second). Notice that the year, month, day, hour, minute, and second are known. The INT relation denotes *real* truncation. 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 (except for JD_{UTC}) whereas the calendar contains periodic steps through the addition of leap years and seconds. Algorithm 14 is valid for any of the time systems

* Because the JD is precisely 365.25 days per year (leap year every 4 years), and the Gregorian calendar is 365.2425 (leap years every 4 years except for evenly divisible years of 400), a difference of 0.0075 days per year occurs. Some references cite the Gregorian proleptic (using the Gregorian calendar before it was actually established) date for the 0 year as November 24, 4713 BC. ($4713 + 1582) * 0.0075 - 10$ days that were corrected in 1582.

we've discussed and should be identified to avoid confusion (JD_{UT1} , JD_{TDT} , JD_{TDB} , ...). Unless specified, JD usually implies a time based on UT1.

ALGORITHM 14: *JULIAN DATE* ($yr, mo, d, h, min, s \Rightarrow JD$) {1900 to 2100}

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

The values are typically quite large, so we may be tempted to use only a few decimal places. However, because the units are days, retaining at least 8 decimal digits provides reasonable accuracy (about $4^s \times 10^{-4}$). The most precise approach is to split the JD into a day and fractional day part. Finally, the IAU recommends using a **Modified Julian Date**, **MJD**, commonly calculated as follows:

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

This adjustment reduces the size of the date (about two significant digits), and it can reduce potential confusion because it begins each day at midnight instead of noon.

Meeus (1991:61) gives a general formula for JD without restriction. We treat January and February as months 13 and 14, respectively. Thus, if $mo = 1$ or 2, then we assign $yr = yr - 1$ and $mo = mo + 12$. The daily time is added as in the final term of Algorithm 14. For Julian calendar dates, set $B = 0$ (* use 61 seconds for days with a leap second).

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

$$C = \frac{\left(\frac{s}{60^*} + min\right)}{60} + h$$

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

$$+ B - 1524.5 + C$$

Because Julian Centuries are often used as the time scale, a shorthand notation such as J1900.0 or J2000.0 is common. To be precise, I've shown the J2000.0 epoch using TT, although we often use UT1 for most problems. Some commonly used epochs are

GPS epoch	= 2,444,244.5	= January 6, 1980 00:00:00.000 UTC
J2000.0	= 2,451,545.0	= January 1, 2000 12:00:00.000 TT
J1900.0	= 2,415,021.0	= January 1, 1900 12:00:00.000 UT1
B1900.0	= 2,415,020.313 52	= December 31, 1899 19:31:28.128 UT1
B1950.0	= 2,433,282.423 459 05	= December 31, 1949 22:09:46.862 UT1

The notations of B1900 and B1950 are for *Besselian years*.^{*} Besselian years are some-

times 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 3-4 shows how to find the Julian date.

▼ **Example 3-4. Determining the Julian Date.**

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

FIND: JD

From Algorithm 14,

$$JD = 367(1996)$$

$$JD = 732,532.0 - \frac{3494.0 + 305.0 + 26 + 1,721,013.5 + 0.597\overline{222}}{4} + \frac{\left(\frac{0}{60} + 20\right)}{24} + 14$$

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

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 (i.e. J2000.0). These relations require Julian centuries that reference a particular time scale, so I'll use subscripts (T_{UT1} , T_{TT} , ...) for clarity. The general formula referencing J2000.0 is

$$T_{xxx} = \frac{JD_{xxx} - 2,451,545.0_{xxx}}{36,525} \quad (3-42)$$

Day of the Year

The **day of the year** has integer values from 1 to 365 (366 in a leap year). Table 3-3 shows the values throughout the year, as well as elapsed time from January 1.

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 later.

3.5.2 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 observa-

* Besselian years depend on the tropical year—the time it takes for the Earth to complete one revolution. Seidelmann (1992:80) gives the definition as the “interval during which the Sun’s mean longitude, referred to the mean equinox of date, increases by 360° .”

TABLE 3-3. Day of the Year. This table lists numbers for days of the year. The elapsed days from the beginning of the year are often useful. 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)

tions of a star would suffice for determining sidereal time. The changing instantaneous axis of rotation, causes station locations to continually change. 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. 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 Mean Sidereal Time**, θ_{GMST} or GMST. 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 3-23 shows the relation between the GMST and LST.

Be careful when comparing GHA and GMST 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 GMST may be expressed in degrees or hours and may have positive or negative values. It helps to make a simple sketch like Fig. 3-9 or Fig. 3-23 when working problems to ensure the answer is correct. Also, you may need to check quadrants to correct the angle.

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

$$\theta_{LST} = \theta_{GMST} + \lambda \quad (3-43)$$

This formula requires a convention for east and west longitudes. The convention for this book is positive for east longitudes, negative for west longitudes. The 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 different longitude definitions cause small deviations in the exact longitudes of sites.

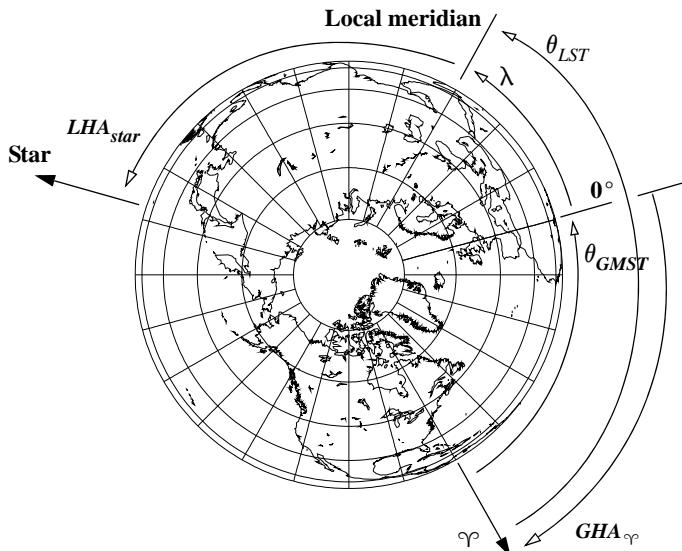


Figure 3-23. Geometry for LST and GMST. Sidereal time is measured positive to the east from the vernal equinox to the location of interest. If we try to locate a star, we can also use hour angles with an appropriate subscript (LHA_{star}). In this example, both sidereal times are positive in a right-handed system, whereas GHA_φ is positive and LHA_{star} is negative in a left-handed system.

Recall that the equinox results from the intersection of the Earth's equator and the ecliptic, and both planes are moving. This permits a distinction 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 [in the plane of the true equator] is the **equation of the equinoxes**. It gives us the difference between the mean and apparent sidereal times, and is expressed in Eq. (3-79). For this book, I'll consider all LST and GMST values to be mean sidereal times and will identify occurrences of apparent sidereal time.

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. 3-9, we can develop relations between systems. Remember that θ_{LST} always references the vernal equinox.

$$\theta_{LST} = \alpha_{star} + GHA_{star} + \lambda = \alpha_{star} + LHA_{star} \quad (3-44)$$

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

* The literature commonly uses GST for Greenwich apparent sidereal time. This could be confusing with GMST, so I'll use AST. Some texts also use Greenwich Apparent Sidereal Time, GAST.

Finally, I'll introduce two techniques to determine sidereal time when we know universal time, UT1. One uses an adopted epoch such as the J2000.0 epoch; the other uses an epoch at the beginning of each year. Both methods yield the mean sidereal time. This mean time is referenced to the mean equinox, and apparent sidereal time is referenced to the true equinox. (We'll explore this more in Sec. 3.7.)

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 GMST calculations for the period from 1901–2199 and are ideally suited for the computer. The limitation on applicable time is because the constants are set for leap days occurring every four years, but not for centuries evenly divisible by 400.

Simon Newcomb developed a mathematical expression for the fictitious mean Sun and he defined sidereal time as the hour angle of this fictitious mean Sun. A single expression references Greenwich, providing the Greenwich mean sidereal time at midnight, $\theta_{GMST\ 0^h}$, ($0^h\ 0^m\ 0^s$) UT1. Expressions for $\theta_{GMST\ 0^h}$ at a desired time in seconds and degrees are (*Astronomical Almanac*, 1984:S15),

$$\begin{aligned}\theta_{GMST\ 0^h} &= 24,110.548\ 41^s + 8,640,184.812\ 866T_{UT1} + 0.093\ 104\ T_{UT1}^2 - 6.2 \times 10^{-6} T_{UT1}^3 \\ \theta_{GMST\ 0^h} &= 100.460\ 618\ 4^\circ + 36,000.770\ 053\ 61T_{UT1} + 0.000\ 387\ 93\ T_{UT1}^2 - 2.6 \times 10^{-8} T_{UT1}^3\end{aligned}\quad (3-45)$$

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

Sometimes the last term is ignored 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:

$$\theta_{GMST} = \theta_{GMST\ 0^h} + \omega_{\oplus_{prec}} UT1 \quad (3-46)$$

where $\omega_{\oplus_{prec}}$ is the Earth's mean angular rotation in degrees per second [Eq. (3-40)] and UT1 is the universal time in seconds past 0^h . (units need to be consistent with $\theta_{GMST\ 0^h}$.

The most useful relation for computer software is one that uses only the JD (UT1) (Seidelmann, 1992:162).*

* Capitaine and Gontier (1993) relate Eq. (3-40) to Eq. (3-47) by first taking the time derivative of Eq. (3-47) (with respect to T_{UT1}) and dividing by 36,525. The result gives the number of sidereal seconds in one solar day ($86,636.555\ 367\ 908\ 73^s + 5.098\ 097 \times 10^{-6} T_{UT1} - 5.0924 \times 10^{-10} T_{UT1}^2$). Dividing this result by 86,400 yields Eq. (3-40). The ratio, $\omega_{\oplus_{prec}}$, is sometimes called $\dot{\theta}_{GMST}$, and it's also used as a generic value for the Earth's rotation, ω_{\oplus} .

$$\begin{aligned}\theta_{GMST} = & 67,310.548\ 41^s + (876,600^h + 8,640,184.812\ 866^s)T_{UT1} \\ & + 0.093\ 104\ T_{UT1}^2 - 6.2 \times 10^{-6}T_{UT1}^3\end{aligned}\quad (3-47)$$

The following algorithm calculates GMST and LST and an example illustrates each.*

ALGORITHM 15: *LSTIME* ($JD_{(UT1)}$, $UT1$, $\lambda \Rightarrow \theta_{LST}$, θ_{GMST})

$$T_{UT1} = \frac{JD_{(UT1)} - 2,451,545.0}{36,525}$$

$$\theta_{GMST} = 67,310.548\ 41^s + (876,600^h + 8,640,184.812\ 866^s)T_{UT1} + 0.093\ 104\ T_{UT1}^2 - 6.2 \times 10^{-6}T_{UT1}^3$$

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

▼ **Example 3-5. Finding GMST and LST (Method 1).**

GIVEN: August 20, 1992, 12:14 P.M. UT1 at 104° west longitude

FIND: GMST (θ_{GMST}) and LST (θ_{LST})

Start by finding the Julian date.

$$JD = 367(1992) - \text{INT} \left\{ \frac{7 \left\{ 1992 + \text{INT} \left(\frac{8+9}{12} \right) \right\}}{4} \right\} + \text{INT} \left(\frac{275(8)}{9} \right) + 20 + 1,721,013.5 + \frac{\frac{14}{60} + \frac{12}{24}}{24}$$

$$JD = 731,064.0 - 3487 + 244 + 20 + 1,721,013.5 + 0.509\ 722 = 2,448,855.009\ 722$$

$$T_{UT1} = \frac{JD - 2,451,545.0}{36,525} = -0.073\ 647\ 919$$

We prefer Eq. (3-47) to find GMST because the operation is complete with one formula, but for computer applications, you may require extended precision.

$$\begin{aligned}\theta_{GMST} = & 67,310.548\ 41^s + (876,600^h + 8,640,184.812\ 866^s)T_{UT1} \\ & + 0.093\ 104\ T_{UT1}^2 - 6.2 \times 10^{-6}T_{UT1}^3\end{aligned}$$

$$\theta_{GMST} = -232,984,181.090\ 925\ 5^s$$

Reduce this quantity to a result within the range of 86400^s ($-49,781.090\ 925\ 5^s$). Then convert to degrees by dividing by 240 ($1^s = 1/240^\circ$ using Eq. (3-37)).

$$\theta_{GMST} = -207.421\ 212\ 189\ 5^\circ$$

Add this answer to 360° to get the final answer.

$$\theta_{GMST} = 152.578\ 787\ 810^\circ$$

Find LST given $\theta_{GMST} + \lambda = 152.578\ 787\ 810^\circ + (-104.0^\circ) = 48.578\ 787\ 810^\circ$

* Notice that UT1 is input to the algorithm. Because UTC is generally available and not UT1, you must first calculate UT1 for precise applications [See Eq. (3-49)].

▲ $\theta_{LST} = 48.578\ 787\ 810^\circ$

The second method determines 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 ($\theta_{GMST\ Jan\ 1,\ 0^h}$) of the year of interest but doesn't provide the means to find it. We can find values for $\theta_{GMST\ Jan\ 1,\ 0^h}$ in the *Astronomical Almanac* for a given year or calculate them for *any* year by the first method [Algorithm 15]. Then, by knowing the day of the year and the fraction of a day, we find the GMST using the following formula (an approximation of Eq. (3-40)) in radians:

$$\theta_{GMST} = \theta_{GMST\ 0^h} + 1.002\ 737\ 909\ 350\ 795(2\pi)(ED) \quad (3-48)$$

where ED is the total elapsed time in solar days from the epoch January 1, 0^h UT1. See Table 3-3 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 $\theta_{GMST\ Jan\ 1,\ 0^h}$ 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. This method neglects the contributions of the quadratic and cubic terms in Eq. (3-45). These contributions are usually small but measurable for precise applications. Consider Example 3-6.

▼ **Example 3-6. Finding GMST and LST (Method 2).**

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

FIND: GMST (θ_{GMST}) and LST (θ_{LST})

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

$$\theta_{GMST\ Jan\ 1,\ 0^h} = 6^h\ 39^m\ 37.4863^s = 99.906\ 192\ 9^\circ$$

Next find ED = 232 from Table 3-3 (1992 is a leap year). Then, update ED with 12/24 + 14/1440 for UT.

$$ED = 232.509\ 722\ 222 \text{ elapsed days from January 1}$$

Find the GMST and be sure to eliminate whole revolutions.

$$\theta_{GMST} = \theta_{GMST\ Jan\ 1,\ 0^h} + 1.002\ 737\ 909\ 350\ 795 (360^\circ) (ED) = 84,032.578\ 788\ 27 \text{ MOD } 360^\circ$$

$$\theta_{GMST} = 152.578\ 788\ 27^\circ = 10^h\ 10^m\ 18.909^s$$

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

▲ $\theta_{LST} = 48.578\ 788\ 27^\circ = 3^h\ 14^m\ 18.909^s$

Although the computations are less complex than Example 3-5, notice the similarity in results, which suggests we can compare them for accuracy. As we would expect, 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.

3.5.3 Atomic Time

We need a highly accurate time system which is independent of the average rotation of the Earth. 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. However, its rate is affected by relativistic effects, such as the clock's location to the geoid. One SI ***second***, (*Système International*) equals the duration of 9,192,631,770 periods of the wavelength associated with the radiation emitted by the electron transition between two hyperfine levels of the ground state of cesium-133 at 0° K. The official introduction of TAI dates to January 1972, but it's been available since 1955.

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 distributed 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 accelerated by a 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.

3.5.4 Dynamical Time

In the 1950s, *ephemeris time*, ET, was created to provide a more stable time reference than those based on the Earth's variable rotation. As additional observations were taken, it became apparent that relativistic effects *were* significant. Many astronomical equations of motion now reference the barycenter of the solar system. ***Terrestrial dynamical time***, TDT, and ***barycentric dynamical time***, TDB (the acronyms for dynamical times are ordered after the French translation, *Temps Dynamic Barycentric*), were introduced in 1979 and 1982 to replace ET and provide a link to the barycentric-referenced equations of motion. In 1991, an IAU resolution defined ***terrestrial time***, TT as the new name for TDT.

Terrestrial time, TT, is the “*theoretical timescale of apparent geocentric ephemerides of bodies in the solar system*” (Seidelmann, 1992:42). TT is independent of equation of motion theories* and uses the SI second as the fundamental interval. It is related to other times as follows:

$$\begin{aligned} UTC &= UT1 - \Delta UT1 \\ TAI &= UTC + \Delta AT \quad TAI = GPS + 19.0^s \\ TT &= TAI + 32.184^s \end{aligned} \tag{3-49}$$

* Within General Relativity, the IAU Colloquim 180 (October 2000) recognized that TT differs from TCG by a constant rate, $1 - L_G = 1 - 6.969\ 290\ 134 \times 10^{-10}$.

where ΔAT and ΔUT1 are the accumulated differences. These parameters are the ***Earth Orientation Parameters*** (EOP), that describe the motion of the Earth. There are several quantities: polar motion angles (x_p, y_p), ΔUT1 timing values, nutation correction terms ($\Delta\Psi, \Delta\epsilon$), and excess length of day (LOD).^{*} The USNO provides long term predicted values that are useful in prediction applications. NGA also provides weekly EOP prediction values that enable the user to find x_p, y_p , and ΔUT1 —see Appendix D.5. The 32.184-second difference with the atomic time is the predefined constant offset between the two systems. The difference arises because it was needed to set the origin of TT equal to ET. Representative values are in Table 3-4.

TABLE 3-4. Timing Coefficients. This table shows sample values for the difference in atomic time ($\Delta\text{AT} = \text{TAI} - \text{UTC}$) as well as coordinated and universal time ($\Delta\text{UT1} = \text{UT1} - \text{UTC}$). ΔAT remains constant until changed, whereas ΔUT1 changes continuously. Offsets to the GPS satellites are needed when determining navigational information. $\text{GPS} - \text{UTC} = \Delta\text{AT} - 19^\circ$. (Source: *Astronomical Almanac*, 2006:K9, and USNO)

		$\Delta\text{AT} (\text{s})$ (After Jan 1)		ΔUT1			$\Delta\text{AT} (\text{s})$ (After Jan 1)		ΔUT1
		Date	UT1 – UTC (s)		Date	UT1 – UTC (s)		UT1 – UTC (s)	
1988	Jan 1	24	0.364 300		1997	Jan 1	30	-0.111 026 0	
1989	Jan 1	24	-0.116 600		1997	Jul 1	31	0.526 894 0	
1989	Jul 1	24	-0.386 170		1998	Jan 1	31	0.218 143 0	
1990	Jan 1	25	0.328 640		1999	Jan 1	32	0.716 637 0	
1991	Jan 1	26	0.618 670		2000	Jan 1	32	0.355 500 0	
1992	Jan 1	26	-0.125 300		2001	Jan 1	32	0.093 227 6	
1992	Jul 1	27	0.442 910 0		2002	Jan 1	32	-0.115 822 3	
1993	Jan 1	27	0.062 020 0		2003	Jan 1	32	-0.289 318 0	
1993	Jul 1	28	0.598 830 0		2004	Jan 1	32	-0.389 590 8	
1994	Jan 1	28	0.199 280 0		2005	Jan 1	32	-0.503 902 4	
1994	Jul 1	29	0.782 670 0		2005	Dec 31	32	-0.661 123 0	
1995	Jan 1	29	0.398 500 0		2006	Jan 1	33	0.338 833 5	
1996	Jan 1	30	0.555 290 0		2009	Jan 1	34	0.407 172 8	

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 TT 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 TT, UT1, gravitational constants, and many other parameters. Green (1988:253–254) shows the original development. Begin with a conversion between barycentric and terrestrial time:

* There are subdaily variations that are not part of the ΔUT1 and LOD values. They arise primarily from tidal and libration effects.

$$TDB = TT + \frac{2r_s}{a_e n_e} e_e \sin(E) + \text{other} \quad (3-50)$$

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, and it's often encountered in computations. See Eq. (3-58). I've shown a way to convert between the two values [Eq. (2-52)], but we can use Eq. (2-52) now to find a “second-order” result:

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

Be careful to use radians in this expression! We also know from Eq. (2-4) that

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

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

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

Substituting into Eq. (3-50), we get

$$\begin{aligned} TDB &= TT + 0.001\,657^s \sin(M_{\oplus}) + 0.000\,013\,85 \sin(2M_{\oplus}) \\ &\quad + \text{lunar/planetary terms} + \text{daily terms} \end{aligned} \quad (3-51)$$

$$M_{\oplus} \cong 357.527\,723\,3^\circ + 35,999.050\,34 T_{TDB}$$

The mean anomaly (Sec. 2.2) from Eq. (3-58) is used interchangeably for the Earth and the Sun (*Astronomical Almanac*, 1984:S26), and T_{TDB} is the Julian centuries of TDB.

In preparing ephemerides of the solar system today, as is done by the JPL, the relativistic equations of motion for the numerical integrations must use an independent variable called *time*. This must be a form of barycentric coordinate time, but it does not necessarily have a perfectly defined relation to an Earth-defined time. When an integrated ephemeris is fit to planetary observations, appropriate relativistic transformations are made to reconcile the different coordinate systems involved. After several iterations, the result is a barycentric ephemeris whose originating epoch and rate of barycentric coordinate time can differ from both TCB and TT. JPL adjusts the rate and epoch of this time scale to match TT over the span of the ephemeris, and calls it T_{eph} (not to be confused with ephemeris time). T_{eph} thus becomes almost the same as TT over the time span of the ephemeris.

T_{eph} is functionally equivalent to TDB. TDB was originally intended to be an independent argument of barycentric ephemerides and equations of motion, and was defined by an IAU 1976 resolution to differ from TT only by periodic terms (M_{\oplus} is the same as

in Eq. (3-51))

$$TDB \cong TT + 0.001\ 657^s \sin(M_{\oplus}) + 0.000\ 014 \sin(2M_{\oplus})$$

Later it became clear that this condition cannot be rigorously satisfied, so IAU 2006 resolutions recommended TDB be defined as linearly related to TCB. Note that TT is usually used in place of TDB, and $TDB_0 = -6.55 \times 10^{-5}$ sec.

$$TCB - TDB = 1.550\ 519\ 767\ 72 \times 10^{-8} \times \{(JD_{TCB} - t_0) \times 86400\} + TDB_0 \text{ sec} \quad (3-52)$$

Kaplan (2005:15) uses the Julian centuries of TT and shows

$$\begin{aligned} T_{eph} &\cong TDB \cong TT + 0.001\ 657^s \sin(628.3076T_{TT} + 6.2401) \\ &+ 0.000\ 022 \sin(575.3385T_{TT} + 4.2970) + 0.000\ 014 \sin(1256.6152T_{TT} + 6.1969) \\ &+ 0.000\ 005 \sin(606.9777T_{TT} + 4.0212) + 0.000\ 005 \sin(52.9691T_{TT} + 0.4444) \\ &+ 0.000\ 002 \sin(21.3299T_{TT} + 5.5431) + 0.000\ 010 T_{TT} \sin(628.3076T_{TT} + 4.2490) + \dots \end{aligned} \quad (3-53)$$

The Astronomical Almanac (2012:B7) shows the following (accurate to about 30 μ s during 1980-2050, and $(M_{\oplus}$ as in Eq. (3-51))

$$\begin{aligned} TDB &= TT + 0.001\ 657^s \sin(M_{\oplus}) + 0.000\ 022 \sin(\Delta\lambda_{mean}) \\ \Delta\lambda_{mean} &= 246.11^\circ + 0.902\ 517\ 92(JD_{TT} - 2451545.0) \end{aligned} \quad (3-54)$$

where $\Delta\lambda_{mean}$ is the difference in the mean longitudes of the Earth and Jupiter.

In practice, the time difference $TT - T_{eph}$ never exceeds 1.7 ms. Thus, any error resulting from the use of TT instead of TDB or T_{eph} in the JPL planetary ephemerides is insignificant for almost all near-Earth orbit-determination applications. Since T_{eph} remains an awkwardly subscripted label, and since there is a need for a barycentric time scale for star catalogs, the label TDB is being revived for those purposes in the Astronomical Almanac, to be used as in the past. However, note that this text uses TT and Julian centuries of TT (T_{TT}) in many relations where prior literature indicates TDB. According to Kaplan (2005:43), TDB and TT are sufficiently close so as not to introduce any noticeable error in the solutions.*

3.5.5 Coordinate Time

Seidelmann and Seago (2005) discuss general time-scale usage in the context of relativistic transformations between the geocentric and barycentric coordinate systems, which can be important for some astrodynamical applications. An IAU 1991 resolution introduced *coordinate time* (TCG and TCB) for space-time applications within the framework of general relativity. Consult Seidelmann (1992:41–48) and McCarthy (1996) concerning the development and use of these highly accurate representations of time.

* It is correct to say that each ephemeris realizes its own version of TDB, but this level of distinction is inconsequential for practically every astrodynamical purpose.

Geocentric coordinate time, TCG is a theoretical coordinate time, and **Barycentric coordinate time**, TCB is the solar-system barycentric coordinate time, and TCB differs from TCG by both secular and periodic terms. The complicated relationships between TT, TCG, and TCB are based on relativistic Lorentz transformations, although approximate expressions for (TCB – TCG) exist (Fairhead and Bertagnon, 1990).

The TT, TCG, and TCB timescales were defined to be equivalent at the TAI epoch of $t_0 = \text{Jan 1, 1977, 00:00:00.000}$ ($JD_{\text{TAI}} = 2443144.5$ precisely). TT, TCG, TCB at t_0 equal Jan 1, 1977 00:00:32.184 ($JD = 2443144.5003725$ precisely); that is to say, the epoch of these three (3) timescales read ahead of TAI by exactly 32.184 seconds at t_0 by convention. With an originating epoch ($t_0 = 2443144.5003725$) prescribed,

$$\text{TT} - \text{TCG} = -L_G \times \{ (JD_{\text{TCG}} - t_0) \times 86400 \} \text{ sec of TT} \quad (3-55)$$

with $L_G \equiv 6.969290134 \times 10^{-10}$.* The elapsed epoch difference in seconds is indicated by TT–TCG, while the Julian date epoch is indicated with JD_{TCG} (due to our use of Julian date, a scale factor of 86400 converts elapsed Julian-date interval from t_0 into units of seconds). This expression is consistent with Kaplan (2005:13), but the form may be inconvenient, because TCG is a coordinate time representing the independent argument of the equations of motion of bodies in its frame and will not be ordinarily kept by any physically real clock. It is more convenient to have a relationship in terms of TT, which is realized via TAI and UTC. Petit and Luzum (2010) provide a slightly different recommendation having TT as the independent variable.

$$\text{TT} - \text{TCG} = -\frac{L_G}{1 - L_G} \times \{ (JD_{\text{TT}} - t_0) \times 86400 \} \text{ sec} \quad (3-56)$$

The precision differences in using TAI versus TCG and TT as independent variables appears on the order of tens of nanoseconds. This difference is within the uncertainty of broadcast UTC itself, so TAI tends to be a sufficiently accurate proxy when TT comes from an atomic clock. Barycentric Coordinate time may be found using Eq. (3-52).

3.5.6 Conversions

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 J2000.0. This equation is similar to Eq. (3-42); 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.

* L_G is a scale constant accounting for the Earth's gravitational and rotational potential affecting the rate of clocks according to the IAU-specified relativistic metric. IAU Resolution B1.9 (2000) recommends L_G as a defining constant, so the relationship cannot change.

ALGORITHM 16: CONVTIME

$$(yr, mo, day, UTC, \Delta UT1, \Delta AT \Rightarrow UT1, TAI, TT, TDB, T_{UT1}, T_{TT}, T_{TDB})$$

$$UT1 = UTC + \Delta UT1$$

$$TAI = UTC + \Delta AT \quad GPS = UTC + \Delta AT - 19^s$$

$$TT = TAI + 32.184^s$$

$$JULIAN\ DATE (yr, mo, day, \{h, min, s\}_{TT} \Rightarrow JD_{TT}), T_{TT} = \frac{JD_{TT} - 2,451,545.0}{36,525}$$

$$JULIAN\ DATE (yr, mo, day, \{h, min, s\}_{TAI} \Rightarrow JD_{TAI})$$

$$JULIAN\ DATE (yr, mo, day, \{h, min, s\}_{TDB} \Rightarrow JD_{TDB}), T_{TDB} = \frac{JD_{TDB} - 2,451,545.0}{36,525}$$

$$\begin{aligned} TDB \equiv & TT + 0.001\ 657^s \sin(628.3076T_{TT} + 6.2401) \\ & + 0.000\ 022 \sin(575.3385T_{TT} + 4.2970) + 0.000\ 014 \sin(1256.6152T_{TT} + 6.1969) \\ & + 0.000\ 005 \sin(606.9777T_{TT} + 4.0212) + 0.000\ 005 \sin(52.9691T_{TT} + 0.4444) \\ & + 0.000\ 002 \sin(21.3299T_{TT} + 5.5431) + 0.000\ 010 T_{TT} \sin(628.3076T_{TT} + 4.2490) + \dots \end{aligned}$$

$$TCB = TT + 1.550\ 519\ 767\ 72 \times 10^{-8} \times \{ (JD_{TAI} - t_0) \times 86400 \}$$

$$TCG = TT + \frac{L_G}{1 - L_G} \times \{ (JD_{TT} - t_0) \times 86400 \}$$

We can visualize the various time conversions as in Fig. 3-24.

▼ **Example 3-7. Calculating Dynamical Time.**

GIVEN: May 14, 2004, 10:43 Mountain Standard Time (UTC based)

FIND: $UT1, TAI, GPS, TT, TDB, T_{UT1}, T_{TT}, T_{TDB}$

Obtain ΔAT (32.0^s) directly from the *Astronomical Almanac*. Unfortunately, $\Delta UT1$ is observed and isn't listed directly. Appendix D lists sources to find values. You must find correct values as $\Delta UT1$ varies about $\pm 0.9^s$. From the USNO's IERS final data, $\Delta UT1 = -0.463\ 326^s$ for May 14, 2004. Remember that in May, Daylight Savings Time exists in the Mountain Standard Time Zone. Therefore, from Table 3-2, we must add six hours to get UTC. We find $UT1$ as

$$UT1 = UTC + \Delta UT1 = 16:43 - 0.463\ 326 = 16:42:59.5367$$

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

$$TAI = UTC + \Delta AT = 16:43:0.00 + 32^s = 16:43:32.0000$$

Notice that TAI and GPS time are always within an integer number of seconds of UTC.

$$GPS = UTC + \Delta AT - 19^s = 16:43:0.00 + 32^s - 19^s = 16:43:13.000$$

Dynamical time is then

$$TT = TAI + 32.184^s = 16:43:32.0 + 32.184^s = 16:44:04.1840$$

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

$$DMY\ HMS \Rightarrow JD_{TT} = 2,453,140.197\ 270\ 65$$

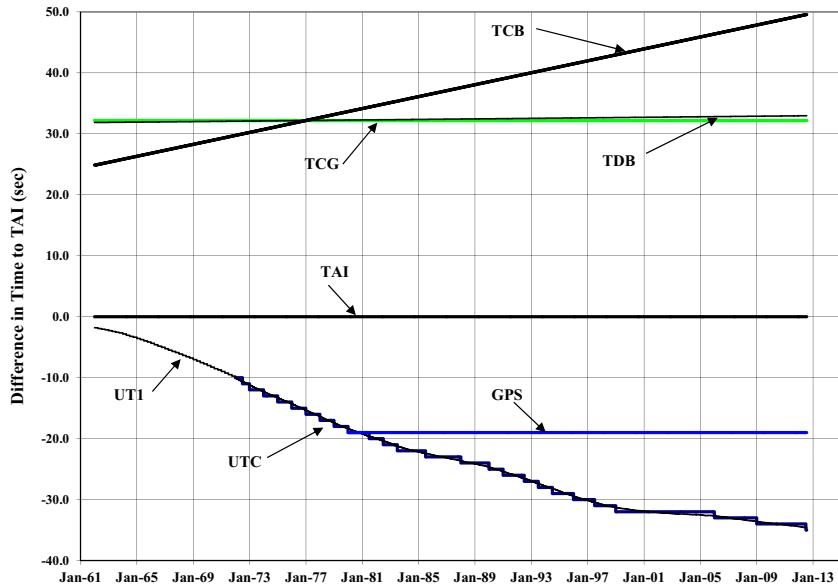


Figure 3-24. Time Conversions. The difference in each time is shown with respect to TAI. Note that TT and TDB are indistinguishable at this scale.

Now find the Julian centuries of TT using Eq. (3-42):

$$T_{TT} = \frac{JD_{TT} - 2,451,545.0}{36,525} = \frac{2,453,140.197 - 2,451,545.0}{36,525} = 0.043\ 674\ 121\ 031$$

If TDB is needed,

$$T_{DB} \approx TT + 0.001\ 657^s \sin(628.3076T_{TT} + 6.2401) + \dots = 16:44:04.185\ 640$$

Finally, determine additional Julian centuries for specific applications. Remember to first find the Julian date for each time before determining the centuries. Using the appropriate JD with Eq. (3-42), ($JD_{TDB} = 2,453,140.197\ 270\ 667$; $JD_{UT1} = 2,453,140.196\ 522\ 415$)

$$T_{TDB} = \frac{JD_{TDB} - 2,451,545.0}{36,525} = 0.043\ 674\ 121\ 031 \quad T_{UT1} = 0.043\ 671\ 100\ 545$$

In summary, the results are as follows.

$UT1 = 16:42:59.5367$	$T_{UT1} = 0.043\ 671\ 100\ 545$
$UTC = 16:43:00.0000$	
$TAI = 16:43:32.0000$	
$GPS = 16:43:13.0000$	
$TT = 16:44:04.1840$	$T_{TT} = 0.043\ 674\ 121\ 031$
$TDB = 16:44:04.1856$	$T_{TDB} = 0.043\ 674\ 121\ 031$
$TCB = 16:44:17.5255$	
$TCG = 16:44:04.7836$	

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

3.6 Time Conversions

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 your accuracy requirements, this error could be significant.

Formats of time vary widely, and many “standards” exist. Some computer programs 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 formats may be useful in particular applications, they can introduce errors, so we must carefully convert between formats. Especially for computer software, we’d like to have a *common* time format to pass between all routines, and we must ensure that truncation, roundoff, and computer word length don’t adversely affect our accuracy.

The algorithms in this section are intended primarily for computer software applications. They fall into two main areas: those to convert between time formats in hours or degrees and those relating to the time of an event expressed in days or Julian dates.

3.6.1 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, 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 17: *DMSToRAD* ($^{\circ}$, ' , " \Rightarrow α)

$$\alpha = \left(deg + \frac{'}{60} + \frac{"}{3600} \right) \frac{\pi}{180^{\circ}}$$

The reverse procedure (**RADtoDMS**) requires more steps to separate each of the components from the given α . The first step is to convert α to degrees because most computer software will carry this value in radians. We do the actual separation using a truncation function (taking the integer part of a real value). Also, notice that you can store $(Temp - deg)60$ as a temporary variable and save some computational operations.

ALGORITHM 18: *RADtoDMS* ($\alpha \Rightarrow ^{\circ}$, ' , ")

$$deg = \text{TRUNC}(Temp)$$

$$Temp = \alpha \left(\frac{180^{\circ}}{\pi} \right) \quad ' = \text{TRUNC}((Temp - deg)60)$$

$$" = \left(Temp - deg - \frac{'}{60} \right) 3600$$

An example illustrates both processes. To use this in computer software, be sure you understand how the TRUNC function works for negative numbers.

▼ **Example 3-8. Finding DMS.**

GIVEN: $-35^\circ - 15' - 53.63''$

FIND: α

First, determine the degrees in a fractional form:

$$\alpha = -35 - \frac{15}{60} - \frac{53.63}{3600} = -35.264\ 897^\circ = -0.615\ 488\ 6 \text{ rad}$$

The reverse process is

$$^\circ = \text{TRUNC}(-35.264\ 897^\circ) = -35$$

$$' = \text{TRUNC}(-35.264\ 897 - -35)60 = -15$$

$$'' = (-35.264\ 897 - -35 - -15/60)3600 = -53.6299$$

3.6.2 HMS to Rad / Rad to HMS

These routines convert *times* in an hour-minute-second (HMS) format to one in radians and vice versa. As with the DMS routines, **HMSRAD** contains only one equation which carries out the operation; however, recall the basic relations in Eq. (3-37):

ALGORITHM 19: HMSRAD ($h, min, s \Rightarrow \tau$)

$$\tau = 15\left(h + \frac{min}{60} + \frac{s}{3600}\right) \frac{\pi}{180}^\circ$$

The reverse procedure (**RADTOHMS**) requires more steps to separate each of the components from the given τ . We convert τ to degrees, because most computer software carries this value in radians, and then to hours by dividing by 15. We can also convert from radians to hours recognizing that 24 hours = 2π radians. Don't confuse this generic conversion value with one for the Earth's rotation, which is quite close: $0.250\ 684\ 4\ %/\text{min}$ (60) = $15.041\ %/\text{h}$. We do the actual separation using truncation function.

ALGORITHM 20: RADTOHMS ($\tau \Rightarrow h, min, s$)

$$\begin{aligned} h &= \text{TRUNC}(\text{Temp}) \\ \text{Temp} &= \tau \frac{180^\circ}{15\pi} = \tau \frac{24^h}{2\pi} & min &= \text{TRUNC}((\text{Temp} - h)60) \\ && s &= \left(\text{Temp} - h - \frac{min}{60}\right)3600 \end{aligned}$$

An example shows the operation of these two routines. Be careful of negative signs.

▼ **Example 3-9. Finding HMS.**

GIVEN: $15^{\text{h}} 15^{\text{m}} 53.63^{\text{s}}$

FIND: τ

First, determine the time in hours:

$$\tau = 15 + \frac{15}{60} + \frac{53.63}{3600} = 15.26490^{\text{h}}$$

Now recall the conversion to degrees involves multiplying by $15^{\circ}/\text{h}$.

$$\tau = 228.9735^{\circ}$$

To use this value in computer software, convert to radians:

▲ $\tau = 228.9735 \left(\frac{\pi}{180} \right) = 3.996341 \text{ rad}$

3.6.3 HMS to Time of Day/ Time of Day to HMS

We often use the time of day in calculations, although it may be inconvenient to send three separate variables (h-min-s) throughout a computer program. In addition, as we convert Julian Dates, we often need the h-min-sec information that's contained within the overall date. To convert the time to elapsed seconds from the beginning of the day,

$$\tau = 3600h + 60min + s$$

The reverse procedure uses truncation routines. Be sure to use negative values correctly.

ALGORITHM 21: TIMEToHMS ($\tau \Rightarrow h, min, s$)

$$\begin{aligned} h &= \text{TRUNC}(Temp) \\ Temp &= \frac{\tau}{3600} & min &= \text{TRUNC}((Temp - h)60) \\ && s &= \left(Temp - h - \frac{min}{60} \right) 3600 \end{aligned}$$

Consider the following example.

▼ **Example 3-10. Converting HMS to Time of Day.**

GIVEN: $13^{\text{h}} 22^{\text{m}} 45.98^{\text{s}}$

FIND: τ

Begin by converting the given time to seconds:

$$\tau = 3600(13) + 60(22) + 45.98 = 48,165.98$$

In reverse, we carefully use the truncation routines and start with $Temp = 13.37943889$.

$$h = \text{TRUNC}\left(\frac{48,165.98}{3600}\right) = 13$$

$$min = \text{TRUNC}((13.37943889 - 13)60) = 22$$

$$s = \left(13.379\ 438\ 89 - 13 - \frac{22}{60} \right) 3600 = 45.979\ 999$$

▲ Notice the small errors in regeneration that occur with all computer routines. Computer implementations can be tricky in shifting the values into the appropriate decimal locations.

3.6.4 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 occur in years evenly divisible by 4, and in centuries evenly divisible by 400 (the year 2000 is a leap year while 1900 and 2100 are not). Because we usually don't account for the even century requirement, the formula is valid only for years between the century years that do not have leap years—in this case, 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 3-11. 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,

{Temp = Temp + LMonth} 5 times.

▲ $\text{Temp} = 121 + 31$ for June = 152 which is greater than 129. Thus the month is May, and the day is $129 - 121 = 8$.

3.6.5 YMDHMS to Days / Days to YMDHMS

The year-month-day-hour-minute-second (YMD HMS) 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 procedure 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 additional computer operations. The approach below is slightly longer and it's limited to the years 1900-2100 depending on the method of determining leap years, but it's much easier in terms of calculations.

This process follows the YMD to day of the year method (Sec. 3.6.4) but adds time information to the final result. At this point, add the HMS information:

$$\begin{aligned} \text{YMDTODAYOFYEAR } & (YMD \Rightarrow Day\ of\ Yr) \\ days = Day\ of\ Yr + h/24 + min/1440 + s/86400 \end{aligned}$$

We can develop test cases from Table 3-3. The reverse process is more difficult because it requires many truncation operations. 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:

$$\begin{aligned} Day\ of\ Yr &= \text{TRUNC}(Days) \\ \text{YMDTODAYOFYEAR } & (YMD \Leftarrow Day\ of\ Yr) \end{aligned}$$

Find the hours, minutes, and seconds for the UT once you've converted to seconds:

$$\begin{aligned} \tau &= (Days - Day\ of\ Yr)86,400 \\ \text{TIMEToHMS } & (\tau \Rightarrow h, min, s) \end{aligned}$$

An example demonstrates each approach.

▼ Example 3-12. Converting YMD HMS to Days.

GIVEN: March 18, 2001 12:14 P.M.

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:

$$\text{YMDTODAYOFYEAR } (2001, 3, 18 \Rightarrow Day\ of\ Yr)$$

$$Day\ of\ Yr = 77$$

Then, correct for the UT:

$$Days = Day\ of\ Yr + 12/24 + 14/1440 = 77.509\ 722\ 2$$

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 **YMDTODAYOFYEAR** routine, discover the 18th of March. Finally, convert back to *T*:

$$\tau = (Days - Day\ of\ Yr)24 = (77.509\ 722\ 2 - 77)24 = 12.233\ 332\ 8^h$$

▲ TIMEToHMS ($\tau \Rightarrow h, min, s$) and $h = 12, min = 13, s = 59.998\ 08$.

3.6.6 Julian Date to Gregorian Date

The process for finding the Julian date is straightforward in Algorithm 14. But reversing it to find the Gregorian calendar date is also necessary. The first step is to find the year of the observation by determining the number of Julian *years* from an epoch. I use the 1900 epoch, T_{1900} , because it isn't a leap year. However, this epoch limits the time span for which the algorithm is valid (1900-2100). Don't confuse this formula with the Julian centuries used in determining sidereal time and other formulas.

The following algorithm shows the complete process. I've included the details from *YMDTODAYOFYEAR* and *TIMETOHMS* to provide a complete overview of this problem. You may find that breaking the routine into smaller steps will be useful for some computer applications. It's necessary to check for the case at the beginning of a year ($Days < 1.0$). The algorithm also uses the *LMonth* array discussed in Sec. 3.6.4.

ALGORITHM 22: *JDtoGregorianDate* ($JD \Rightarrow Year, Mon, Day, h, min, s$)

$$T_{1900} = \frac{JD - 2,415,019.5}{365.25}$$

$$Year = 1900 + \text{TRUNC}(T_{1900})$$

$$LeapYrs = \text{TRUNC}((Year - 1900 - 1)(0.25))$$

$$Days = (JD - 2,415,019.5) - ((Year - 1900)(365.0) + LeapYrs)$$

IF $Days < 1.0$ THEN

$$Year = Year - 1$$

$$LeapYrs = \text{TRUNC}((Year - 1900 - 1)(0.25))$$

$$Days = (JD - 2,415,019.5) - ((Year - 1900)(365.0) +$$

LeapYrs)

IF (*Year MOD 4*) = 0 THEN

$$LMonth[2] = 29$$

$$DayofYr = \text{TRUNC}(Days)$$

Sum days in each month until *LMonth* + 1 summation > *DayofYr*

Mon = # of months in summation

Day = *DayofYr* - *LMonth* summation

$$\tau = (Days - Day\ of\ Yr) 24$$

$$h = \text{TRUNC}(\tau)$$

$$min = \text{TRUNC}((\tau - h)60)$$

$$s = \left(\tau - h - \frac{min}{60} \right) 3600$$

▼ **Example 3-13. Finding the Gregorian Calendar Date from the Julian Date.**

GIVEN: $JD = 2,449,877.345\ 876\ 2$

FIND: YMD, HMS

Begin by finding the Julian centuries (times 100) from 1900:

$$T_{1900} = (2,449,877.345\ 876\ 2 - 2,415,019.5) / 365.25 = 34,857.845\ 876\ 2 / 365.25 = 95.435\ 580$$

To find the year, use $Yr = 1900 + \text{TRUNC}(T_{1900}) = 1995$. 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

$$\text{YMD} \Rightarrow \text{DAYOFYEAR} (1995, Day\ of\ Yr \Rightarrow YMD)$$

to find $Yr = 1995$, Mo = 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 τ :

$$\tau = (\text{Days} - \text{Day of Yr})24 = (159.845\ 876\ 2 - 159)24 = 20.301\ 028\ 8^h$$

$$\text{TIME} \Rightarrow \text{HMS} (\tau \Rightarrow h, \text{min}, \text{s}) \text{ and } h = 20, \text{min} = 18, s = 3.703\ 68.$$

Notice that if we needed only the h-min-sec information, we could take the fractional part of the JD , add 0.5, and multiply by 24 to get τ . 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!

3.7 Transforming Celestial and Terrestrial Coordinates

Orbit determination requires both celestial reference frames (which define the Newtonian-inertial space in which differential equations of satellite motion are valid) and terrestrial reference frames (in which satellite observations are taken). The conventional terrestrial frame is the ITRF, and this is the frame used with most observations. The WGS-84 terrestrial frame is used primarily by the U.S. DoD. It's realized through GPS observations, although the GPS monitor stations are usually constrained by their adopted ITRF coordinates during gravity field development. Thus, the modern WGS-84 and ITRF terrestrial frames agree at the cm level. In fact, within the uncertainty of the WGS-84 frame, they are equivalent.

With the advent of the IAU-2006/2000 Resolutions, the fundamental approach to coordinate systems was changed, but not the need for transforming coordinates. Seidelmann and Kovalevsky (2003) succinctly summarize this as follows:

In the past, the fixed reference frame was defined in terms of a moving equator and a moving equinox corresponding to a certain date (B1950.0, J2000.0). The practical difficulty was that the definition of the motions was theory-dependent, and sometimes replaced by extrapolation of observations (FK5 and dynamical equinoxes). When observations of different stars from various epochs have been reduced to a fixed reference frame using precession values with errors, the resulting positions on this fixed frame contain systematic errors; these errors cause a rotation in catalogs based upon these stars. The observational realization of the fixed frame was based on nearby bright stars with proper motions.

In the new system, the ICRS is realized by the ICRF, based on distant radio sources, with its fixed coordinate axes and fiducial point. It is independent of the solar system dynamics.

Thus it is truly a kinematically fixed frame. It is aligned to the FK5 at J2000.0 within the FK5 errors.

In both cases, the time-dependent reference frames do not coincide with the fixed ones, so one has to apply a rotation matrix to go from the fixed to the moving frame or the reverse.

The celestial frame is related to a time-dependent terrestrial frame through an Earth orientation model, calculated by the standard matrix-multiplication sequence of transformational rotations:

$$\hat{\vec{r}}_{GCRF} = [\mathbf{P}(t)][\mathbf{N}(t)][\mathbf{R}(t)][\mathbf{W}(t)]\hat{\vec{r}}_{ITRF} \quad (3-57)$$

where $\hat{\vec{r}}_{GCRF}$ is location with respect to the GCRF, \mathbf{P} and \mathbf{N} are the precession-nutation matrices of date t , \mathbf{R} is the sidereal-rotation matrix of date t , \mathbf{W} is the polar-motion matrix of date t , and $\hat{\vec{r}}_{ITRF}$ is location with respect to the ITRF. A frame bias matrix $[\mathbf{B}]$ is needed with the equinox based approach discussed later. A combined \mathbf{PN} matrix may be formed as a single operator, depending on the theory adopted.

The rotations in Eq. (3-57) are collectively known as **reduction formulas** (Seidelmann, 1992:20) or an Earth orientation model. There are two approaches for the reduction formulae: the classical transformation and the **Celestial Intermediate Origin** (CIO). The latter theory is defined by IAU Resolutions. The IAU-2006/2000 Resolutions were officially released over a period of time (McCarthy and Petit: 2003, and Kaplan: 2005). These resolutions stated that beginning January 1, 2003, the IAU-1976 Precession Model and the IAU-1980 Theory of Nutation were replaced by the IAU-2000A Precession-Nutation theory (accurate to 0.0002"). The IAU-2010 Conventions combined and updated these resolutions (Petit and Luzum: 2010). As of January 1, 2009, the IAU-2000A Precession-nutation theory was separated into the IAU-2000 Nutation theory and the IAU-76 Precession was replaced with the IAU-2006 Precession, also called the **P03 model** (Capitaine et al. 2003, and Wallace and Capitaine 2006). Before 2003, the recommended reduction to GCRF coordinates was the IAU 1980 Theory of Nutation, plus the corrections ($\delta\Delta\Psi_{1980}$, $\delta\Delta\epsilon_{1980}$) given in the EOP data (the corrections still exist in 2012). Although some nomenclature exists for the IAU-2000, I'll use IAU-2006/2000 to clarify the IAU-2006 precession and IAU-2000 Nutation respectively.

A truly inertial coordinate frame is difficult to realize. 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 realized that inertial frames can move 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 solution relies on finding **pseudoinertial** reference frames, which we'll define as a specific system for which Newtonian mechanics hold. Sec. 3.3 discusses the existence of many different coordinate systems, some of which may be considered pseudoinertial. Depending on the problem, we can consider most of the coordinate systems in Sec. 3.3 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. Proper understanding of these pseudoinertial coordinate systems is needed for highly accurate simulations.

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} .

For precise applications, 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 GCRF is sufficient.

This section introduces concepts necessary for the IAU-2006/2000 and classical transformations. We can perform these reductions with the current inertial system (GCRF) or its predecessors, IAU-76/FK5 (J2000.0),^{*} FK4 (B1950.0), or any other systems of choice.[†] The classical transformation exists in numerous computer programs, so it's also necessary to explain its formation. I won't discuss translations (topocentric to geocentric to barycentric) in detail because they are simply vector operations. Sec. 4.4.4 describes the complete sequence of steps to reduce observations. I separate polar motion and the fundamental arguments because they are common to both transformations.

To understand the motion of a coordinate system, we begin with the geometry of the Earth's orbit. The equatorial plane of the Earth and the ecliptic plane are not actually fixed relative to the stars. Mathematically, these discrepancies have been divided into

^{*} Previously, the terms “FK5” and “J2000” have been loosely applied to the inertial coordinate system. Unfortunately, these descriptors cannot precisely convey what is meant given the new IAU reference systems. Nomenclature of a “J2000 reference system” is ambiguous now because the IAU-2000/2006 resolutions introduced an entirely new reference system relative to that epoch. Thus, I refer to the former reduction operations as “IAU-76/FK5” and the inertial coordinate system as the “Dynamical Mean of J2000”. This should alleviate some of the confusion between a star catalog (FK5) and an epoch (J2000.0), as well as clarify which reduction operation is referenced.

[†] The FK4 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. I'll briefly introduce the FK4 system in an upcoming section.

precessional and nutational parts. First, the gravitational forces of the planets affect the Earth's orbit through ***precession of the ecliptic*** (formerly *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 (dominated by a significant equatorial bulge), the Sun 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. 3-25 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 pre-

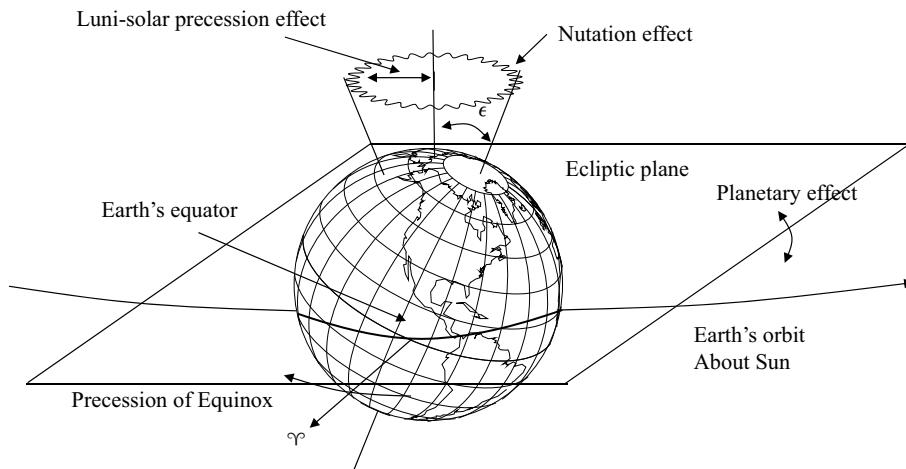


Figure 3-25. Precession and Nutation of the Earth's Equatorial Plane. This figure shows the combined effects of perturbing forces on the Earth. The general motion of the equinox moves west. Planetary and nutation effects are small compared to the luni-solar effects.

cession 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 $0.013\ 889^\circ(50'')$ per year in longitude.

The Moon produces 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 caused by the Moon has a primary period of about 18.6 years. This is more commonly called the ***precession of the equator*** (formerly *regression of the [lunar] node*) because the motion is in a negative direction. Another 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 frequencies, and appears as a superimposed motion with a maximum amplitude of about $0.0025^\circ(9'')$ in the obliquity of the ecliptic, and about $0.004^\circ 72''(17'')$ in longitude. Figure 3-25 shows the resulting motions.

The motion of the ecliptic plane due to general precession causes the equinox to move along [in the plane of] 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 nearly secular 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 reference frame for all measurements (Green, 1988:54). We can also specify each of these frames as *of epoch* where the epoch is an external date that fixes the orientation of the coordinate system, rather than the current date.

There are three primary methods currently available to perform the transformations: the newer CIO-based approach, an equivalent equinox-based approach we'll call IAU-2006/2000, and a corrected equinox-based approach using the former IAU-76/FK5 based technique. See Figure 3-26. The challenge is to transform between these coordinate frames in a consistent and accurate manner.

Polar Motion

Seidelmann (1982) defines polar motion as the “movement of the rotation axis with respect to the crust of the Earth”. The *Celestial Intermediate Pole* (CIP)* is the axis of Earth rotation (formerly the *Celestial Ephemeris Pole*, CEP), which is normal to the true equator. The CIP motion is defined in the GCRS “by the motion of the Tisserand mean axis of the Earth with periods greater than two days”. The Tisserand mean axis is defined in Seidelmann (1982) as the mean surface geographic axis. The axes of the ITRF define one set of spherical coordinates, and the CIP describes a slightly different set for the same location on the Earth. Seidelmann and Kovalevsky (2003) comment on the old and new systems.

The changes [in pole nomenclature] were made necessary by the improvement of the nutation theory and by the requirement to clearly separate nutation from polar motion, but the basic idea remained the same, and there were generally no changes in notation. Similarly, the origin on the celestial equator defined by these poles was the mean equinox of date, whose position depended on the theory of the motion of the Earth and on the precession constant. The important feature is that both the pole and origin were time-dependent.

In the new system, rather than a theoretically defined equinox of date, the origin is a point called the Celestial Ephemeris Origin (CEO) on the true equatorial plane, perpendicular

* The term intermediate is now reserved for use with “moving” frames.

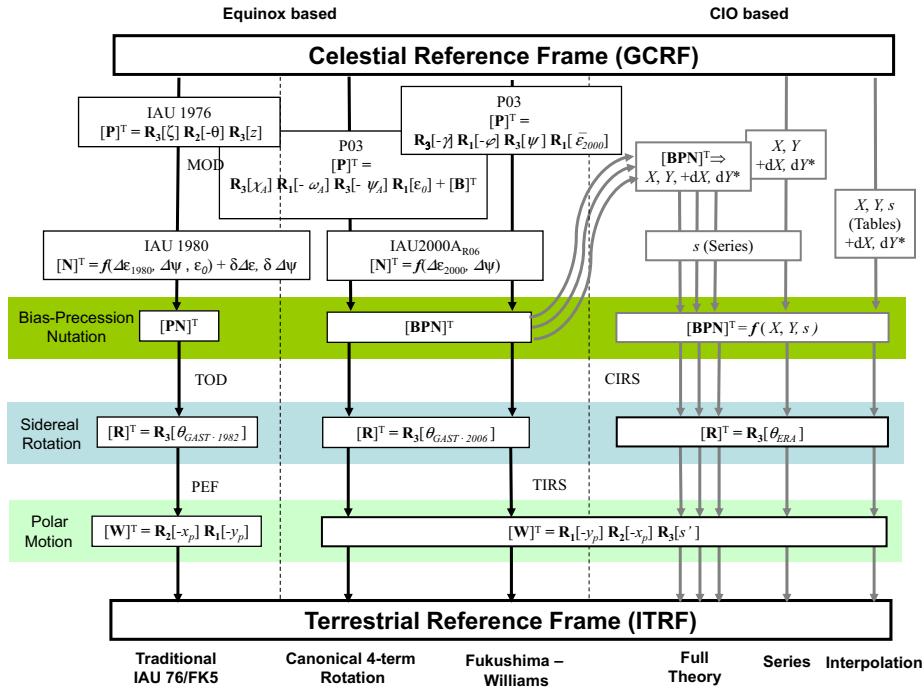


Figure 3-26. Celestial to Terrestrial Transformation. The approaches to transform vectors between terrestrial (ITRS) and celestial coordinate systems (GCRS) are shown. Although names (e.g., mean-of-date, MOD, and operations such as polar motion, etc.) are the same, the formulae are different, as indicated by subscripts and functional dependence on variables for each operation. The IAU-76/FK5 approach is similar to the traditional approach, but with different formulae. The CIO based BPN matrix is composed of many equations. Note that the IAU-76/FK5 and CIO based precession/nutation corrections are from the EOP data, and are available only for historical dates.

to the CIP, and the origin is determined by a formula based uniquely on observations of the CIP motion. The origin does not move along the equator at the precession rate. Thus, the new system is based on the kinematics of the Earth, rather than the dynamics of the solar system.

Because Earth orientation observations are sensitive to the CIP, and not the ITRF (the ITRF is defined by convention, not observation), the recovered latitude and longitude are with respect to the CIP. Unfortunately, this is not the Earth-fixed system we would desire for performing accurate surveys or navigation. The change of basis between these systems is **polar motion**, and it uses the angles given in the EOP data. The angles are updated online by the **International Polar Motion Service** (IPMS) and published by **International Earth Rotation Service** (IERS*). The *Astronomical Almanac* (2006:K10, B60) publishes predicted values, but not the corrections needed for real-time operations. Appendix D lists other sources including the National Geospatial-Intelligence Agency

(NGA) coefficients used to recreate weekly EOP values. Table 3-5 shows some sample values. The actual displacements are given in arcseconds in the ***IERS Reference Meridian*** (IRM) direction (“ x_p ” measured positive south along the 0° longitude meridian) and the y_p meridian (90°W or 270°E meridian). The ***International Reference Pole*** (IRP*) is the location of the terrestrial pole agreed upon by international committees. It results from observed values between 1900 and 1905. Deflections locate the IRP from the CIP. The motion is roughly a circular spiral about the IRP. The maximum amplitude of the variations is about 9 m in any direction. Figure 3-27 shows the geometry for the effect of polar motion.

TABLE 3-5. Pole Positions. This table shows the coordinates of the North Pole over about 5 years. Get current values for highly precise studies. 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
2002	-0.1770	0.2940	-0.0307	0.5406	0.2282	0.4616	0.1985	0.2005
2003	-0.0885	0.1880	-0.1333	0.4356	0.1306	0.5393	0.2585	0.3040
2004	0.0313	0.1538	-0.1401	0.3206	-0.0075	0.5101	0.1985	0.4315
2005	0.1492	0.2380	-0.0287	0.2425	-0.0398	0.3968	0.0588	0.4169
2006	0.0527	0.3834	0.1028	0.3741	0.1285	0.2997	0.0327	0.2524

Fundamental Arguments

The fundamental arguments have remained relatively consistent through the various IAU publications for reduction (1980, 1996, 2000, 2006, 2010). They are an integral part of the IAU-2010 Conventions, and are presented here. Be aware that slightly different values were used with previous theories. There are arguments for the nutation, as well as for the planetary nutation.

Note that TT and Julian centuries of TT (T_{TT}) are used throughout this section. As published, many equations show the use of TDB. According to Kaplan (2005:43), these two quantities are sufficiently close so as not to introduce any noticeable error in the solutions.

The nutation argument uses the mean anomalies (Sec. 2.2) 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

* IERS was formerly the *Bureau International de l'Heure*, BIH, and the IRP was formerly the *Conventional International Origin* (CIO).

* The angles ($M_{\mathbb{C}}$, M_{\odot}) are actually found using the Mean longitude (of the Sun and Moon) minus the longitude of periaxis of each object.

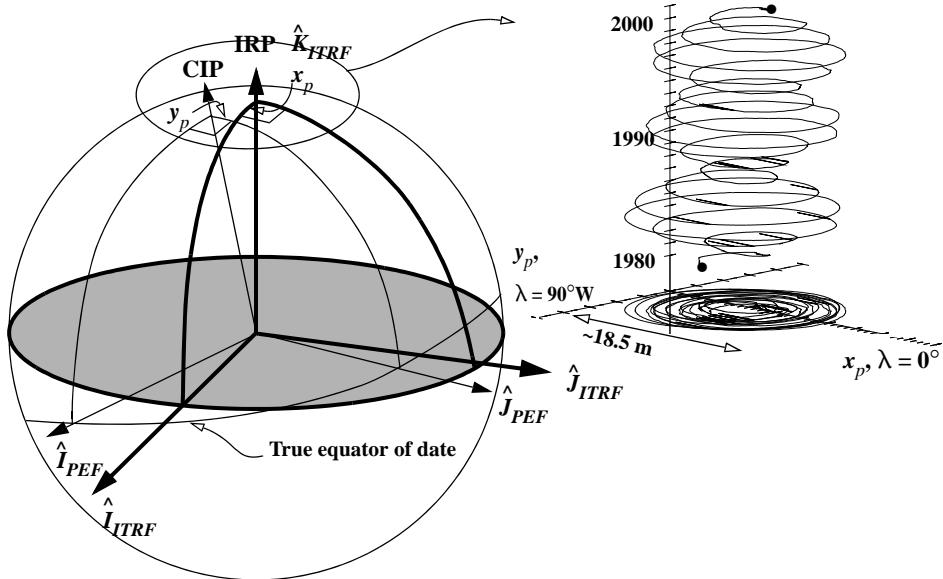


Figure 3-27. Transformation Geometry Due to Polar Motion. This transformation takes into account the actual location of the Celestial Intermediate Pole (CIP) over time. It relates a Pseudo Earth fixed (PEF) system about the CIP, and the ITRF system and the IERS Reference Pole (IRP). The inset plot shows the IRP motion for about 20 years. Notice the slight “structure” to the variations over time. These motions are partially due to the annual (365 d) and Chandler (435 d) periods. The PEF is virtually identical to the TIRS (See Table 3-6), but it’s convenient to separate the names as the exact formulae differ.

from the mean equinox of date, the mean **elongation*** from the Sun, D_{\odot} , and the right ascension of the ascending node of the mean lunar orbit, $\Omega_{\mathbb{Q}}$. $\Omega_{\mathbb{Q}}$ is measured along the ecliptic from the mean equinox of date. These parameters are sometimes known as the **Delaunay arguments**, or the fundamental arguments, and they account for luni-solar nutation. From Kaplan (2005: 46),

$$\begin{aligned}
 M_{\mathbb{Q}} &= 485868.249\,036'' + 1717915923.2178T_{TT} + 31.8792T_{TT}^2 + 0.051\,635T_{TT}^3 - 0.000\,244\,70T_{TT}^4 \\
 M_{\odot} &= 1287104.79305'' + 129596581.0481T_{TT} - 0.5532T_{TT}^2 + 0.000\,136T_{TT}^3 - 0.000\,011\,49T_{TT}^4 \\
 u_{M_{\mathbb{Q}}} &= 335779.526\,232'' + 1739527262.8478T_{TT} - 12.7512T_{TT}^2 - 0.001\,037T_{TT}^3 \\
 &\quad + 0.000\,004\,17T_{TT}^4 \\
 D_{\odot} &= 1072260.703\,69'' + 1602961601.2090T_{TT} - 6.3706T_{TT}^2 + 0.006\,593T_{TT}^3 - 0.000\,031\,69T_{TT}^4 \\
 \Omega_{\mathbb{Q}} &= 450160.398\,036'' - 6962890.5431T_{TT} + 7.4722T_{TT}^2 + 0.007702T_{TT}^3 - 0.000\,059\,39T_{TT}^4
 \end{aligned} \tag{3-58}$$

* The elongation is the central angle of a satellite, east or west (0 - 180°) of a planet (or of a planet from the Sun).

We can also include corrections for planetary effects on the nutation and the obliquity of the ecliptic. These arguments are the mean Heliocentric longitudes of the planets (λ_i), and the general precession in longitude (p_λ). Note that the longitude of Neptune used by SOFA differs from Eq. (3-59) and all values are in radians (Petit and Luzum 2010:68).

$$\begin{aligned}
 \lambda_{M\text{♀}} &= 4.402\,608\,842 + 2608.790\,314\,157\,4T_{TT} \\
 \lambda_{M\text{♀}} &= 3.176\,146\,697 + 1021.328\,554\,621\,1T_{TT} \\
 \lambda_{M\oplus} &= 1.753\,470\,314 + 628.307\,584\,999\,1T_{TT} \\
 \lambda_{M\odot} &= 6.203\,480\,913 + 334.061\,242\,670\,0T_{TT} \\
 \lambda_{M\text{☿}} &= 0.599\,546\,497 + 52.969\,096\,264\,1T_{TT} \\
 \lambda_{M\text{☿}} &= 0.874\,016\,757 + 21.329\,910\,496\,0T_{TT} \\
 \lambda_{M\text{⊕}} &= 5.481\,293\,872 + 7.478\,159\,856\,7T_{TT} \\
 \lambda_{M\Psi} &= 5.311\,886\,287 + 3.813\,303\,563\,8T_{TT} \\
 p_\lambda &= 0.024\,381\,75T_{TT} + 0.000\,005\,386\,91T_{TT}^2
 \end{aligned} \tag{3-59}$$

Epoch Descriptors

There is an additional nuance sometimes associated with precise transformation applications—the “of date” and “of epoch” descriptors. Usually, this is more applicable to the analytical theories (TEME and SGP4, in Sec. 3.7.4), but the terminology is sometimes applied to other operations.

- “of Date”—With this option, the epoch of the coordinate frame is always the same as the epoch of the associated ephemeris generation time. Thus, you determine the nutation parameters at each time step. This is the most accurate interpretation, and is the most common use.
- “of Epoch”—In this approach, the epoch of the coordinate frame is held constant. Subsequent rotation matrices use a static nutation transformation that is applied at each time requested in an ephemeris generation. This is computationally easier, but introduces error into the subsequent solutions depending on how much time has elapsed from the epoch.

The “of epoch” approach is not recommended because of the additional error that is inserted. In addition, if the precision for an operation warrants the distinction of these two approaches, the slight increase in computations to rigorously determine the complete nutation transformation at each time is likely also warranted.

3.7.1 IAU-2010 Conventions

The IAU-2010 conventions summarize and update the IAU-2006/2000 Resolutions that address transformations [Eq. (3-57)] between the body-fixed (ITRF), and the inertial (GCRF) frames (McCarthy and Petit: 2003, Kaplan: 2005, and Petit and Luzum: 2010). The original basis was the non-rotating origin (now the CIO). Guinot (1979) introduces

the non-rotating sphere by describing its attachment to the stars, and its relation to a terrestrial sphere attached to the Earth. The GCRF is realized by a great circle and a reference point. The motion of the pole is constrained so that “when [the pole] moves on the non-rotating sphere, the instantaneous system has no component of rotation around [the non-rotating pole]” Guinot (1979). Thus, the transformation is kinematically correct because it segregates the terrestrial and celestial motion of the pole from the Earth’s rotation (Seago and Vallado, 2000). The CIO approach of the IAU-2006/2000 Resolutions is significant in that it eliminates some of the older distinctions of MOD and TOD. This will be advantageous for numerical operations as less potential confusion will arise from the use of undocumented, intermediate coordinate frames.

Polar Motion (IAU-2006/2000, CIO based)

The IAU-2006/2000 has an additional rotation from the IAU-76/FK5 approach to account for the instantaneous prime meridian called the TIO locator, s' —after January 1, 2003, it’s “necessary to provide an exact realization of the ‘instantaneous prime meridian’” (McCarthy and Petit, 2003:38). Notice that the Terrestrial Intermediate Reference System (TIRS) rotations are the opposite order from IAU-76/FK5 approach, Eq. (3-77) (Kaplan, 2005:64).

$$\dot{\vec{r}}_{TIRS} = \text{ROT3}(-s')\text{ROT2}(x_p)\text{ROT1}(y_p)\dot{\vec{r}}_{ITRF} = [W]_{ITRF-TIRS}\dot{\vec{r}}_{ITRF} \quad (3-60)$$

The main component of s' is found using the average values for the **Chandler wobble** ($a_c = 0.26''$) and the **annual wobble** ($a_a = 0.12''$) of the pole. It is less than $0.0004''$ over the next century (Ast Almanac, 2006:B76).

$$s' = -0.0015'' \left(\frac{a_c^2}{1.2} + a_a^2 \right) T_{TT} \approx -0.000\ 047'' T_{TT} \quad (3-61)$$

The usual x_p and y_p values are sometimes augmented by tidal and nutation (libration) contributions. The combined rotation matrix from ITRF to TIRS is ($[W]_{ITRF-TIRS}$):

$$\begin{bmatrix} \cos(x_p)\cos(s') & -\cos(y_p)\sin(s') + \sin(y_p)\sin(x_p)\cos(s') & -\sin(y_p)\sin(s') - \cos(y_p)\sin(x_p)\cos(s') \\ \cos(x_p)\sin(s') & \cos(y_p)\cos(s') + \sin(y_p)\sin(x_p)\sin(s') & \sin(y_p)\cos(s') - \cos(y_p)\sin(x_p)\sin(s') \\ \sin(x_p) & -\sin(y_p)\cos(x_p) & \cos(y_p)\cos(x_p) \end{bmatrix}$$

Earth Rotation Angle (IAU-2006/2000, CIO based)

The rotation of the Earth is taken into account by the **Earth Rotation Angle**, θ_{ERA} —the angle between the CIO and the *Terrestrial Intermediate Origin*, TIO (a reference meridian on the Earth that’s about 100 m offset from the Greenwich meridian (Kaplan, 2005:54) along the equator of the CIP. θ_{ERA} uses UT1 (McCarthy and Petit, 2003:39). If additional precision is needed, the fractional UT1 date and $0.002737\dots(0.985^\circ)$ may be used instead of $1.002737\dots(360.985^\circ)$.

$$\begin{aligned}\theta_{ERA} &= 280.460\,618\,375\,04^\circ + 360.985\,612\,288\,808(JD_{UT1} - 2,451,545.0) \\ &= 2\pi(0.779\,057\,273\,264\,0 + 1.002\,737\,811\,911\,354\,48(JD_{UT1} - 2,451,545.0))\end{aligned}\quad (3-62)$$

The transformation yields the *intermediate reference frame of epoch*. Notice the transformation for velocity, and the use of the rotation of the Earth.

$$\begin{aligned}\dot{\vec{r}}_{CIRS} &= \text{ROT3}(-\theta_{ERA})\dot{\vec{r}}_{TIRS} \\ \dot{\vec{v}}_{CIRS} &= \text{ROT3}(-\theta_{ERA})\dot{\vec{v}}_{TIRS} + \vec{\omega}_{\oplus} \times \dot{\vec{r}}_{TIRS}\end{aligned}$$

Precession-Nutation (IAU-2006/2000, CIO based)

The coordinates of the CIP in the ICRS (X and Y) are used to locate the GCRF position and velocity vectors. Notice this approach does not require a frame bias modification (Kaplan, 2005:48 and McCarthy and Petit 2003:36).

$$\begin{aligned}\dot{\vec{r}}_{GCRF} &= [PN]\dot{\vec{r}}_{CIRS} \\ &= \begin{bmatrix} 1 - aX^2 & -aXY & X \\ -aXY & 1 - aY^2 & Y \\ -X & -Y & 1 - a(X^2 + Y^2) \end{bmatrix} \text{ROT3}(s)\dot{\vec{r}}_{CIRS} \\ a &= \frac{1}{1 + \cos(d)} \approx \frac{1}{2} + \frac{1}{8}(X^2 + Y^2)\end{aligned}$$

d^* is an argument that defines the CIP in the ICRS and a takes the units of X and Y .

McCarthy and Petit (2003) show how to find the X and Y coordinates of the CIP with respect to the GCRS. Both expressions use the fundamental arguments to form the trigonometric argument.

$$\begin{aligned}a_{p_i} &= a_{0x1_i}M_{\mathbb{C}} + a_{0x2_i}M_{\odot} + a_{0x3_i}u_{M\mathbb{C}} + a_{0x4_i}D_{\odot} + a_{0x5_i}\Omega_{\mathbb{C}} \\ &\quad + a_{0x6_i}\lambda_{M\mathfrak{Q}} + a_{0x7_i}\lambda_{M\mathfrak{Q}} + a_{0x8_i}\lambda_{M\oplus} + a_{0x9_i}\lambda_{M\mathfrak{S}} \\ &\quad + a_{0x10_i}\lambda_{M\mathfrak{U}} + a_{0x11_i}\lambda_{M\mathfrak{L}} + a_{0x12_i}\lambda_{M\mathfrak{D}} + a_{0x13_i}\lambda_{M\mathfrak{P}} + a_{0x14_i}p_{\lambda}\end{aligned}\quad (3-63)$$

Be careful with the order of these coefficients and terms. Notice that the coefficients of the fundamental arguments have “0”, “ x ”, “1”, and “ i ” subscripts. The “0” indicates the index of the terms which corresponds to a summation quantity (0 - 5), “ x ” indicates the parameter under consideration (X , Y , s), “1” indicates the term that the coefficient

* E and d are parameters which may be derived from X and Y . From Capitaine et al. (2003)

$$E = \text{ATAN2}(Y, X) \text{ and } d = \text{ATAN}\left(\sqrt{\frac{X^2 + Y^2}{1 - X^2 - Y^2}}\right)$$

applies to (1 - 14), and “ i ” is the index for the number of terms in the summation (1..1306, 1..253, etc.). The main coefficients (X, Y, s) have “ X ”, “ s ”, “0”, and “ i ” subscripts. The “ X ” indicates the parameter under consideration (X, Y, s), “ s ” indicates if it’s for a sine or cosine, “0” and “ i ” are as before. Also, although the last expression is a summation of only one term, future revisions may include additional terms, and the existing values may change slightly (Kaplan 2005:39). The constants are found in online files (Maia 2011 in tab5.2a.txt and tab5.2b.txt). The latest expressions (Petit and Luzum, 2010:54) are as follows.

$$\begin{aligned}
 X = & -0.016\,617'' + 2004.191\,898T_{TT} - 0.429\,782\,9T_{TT}^2 - 0.198\,618\,34T_{TT}^3 \\
 & + 0.000\,007\,578T_{TT}^4 + 0.000\,005\,928\,5T_{TT}^5 \\
 & + \sum_{i=1}^{1306} [A_{xs0_i} \sin(a_{p_i}) + A_{xc0_i} \cos(a_{p_i})] + \sum_{i=1}^{253} [A_{xs1_i} \sin(a_{p_i}) + A_{xc1_i} \cos(a_{p_i})]T_{TT} \\
 & + \sum_{i=1}^{36} [A_{xs2_i} \sin(a_{p_i}) + A_{xc2_i} \cos(a_{p_i})]T_{TT}^2 + \sum_{i=1}^4 [A_{xs3_i} \sin(a_{p_i}) + A_{xc3_i} \cos(a_{p_i})]T_{TT}^3 \\
 & + \sum_{i=1}^1 [A_{xs4_i} \sin(a_{p_i}) + A_{xc4_i} \cos(a_{p_i})]T_{TT}^4 \\
 Y = & -0.006\,951'' - 0.025\,896T_{TT} - 22.407\,274\,7T_{TT}^2 + 0.001\,900\,59T_{TT}^3 \\
 & + 0.001\,112\,526T_{TT}^4 + 0.000\,000\,135\,8T_{TT}^5 \\
 & + \sum_{i=1}^{962} [A_{ys0_i} \sin(a_{p_i}) + A_{yc0_i} \cos(a_{p_i})] + \sum_{i=1}^{277} [A_{ys1_i} \sin(a_{p_i}) + A_{yc1_i} \cos(a_{p_i})]T_{TT} \\
 & + \sum_{i=1}^{30} [A_{ys2_i} \sin(a_{p_i}) + A_{yc2_i} \cos(a_{p_i})]T_{TT}^2 + \sum_{i=1}^5 [A_{ys3_i} \sin(a_{p_i}) + A_{yc3_i} \cos(a_{p_i})]T_{TT}^3 \\
 & + \sum_{i=1}^1 [A_{ys4_i} \sin(a_{p_i}) + A_{yc4_i} \cos(a_{p_i})]T_{TT}^4
 \end{aligned}$$

The Earth Orientation Parameter (EOP) data includes dX and dY terms that account for the free-core nutation and time dependent effects. The final values are then $X = X + dX$ and $Y = Y + dY$.

The value of s provides the position of the [CIO] on the equator of the CIP corresponding to the kinematical definition of the [non-rotating origin] in the GCRS when the CIP is moving with respect to the GCRS between the reference epoch and the epoch due to precession and nutation (McCarthy and Petit 2003). The complete expression and coefficients are found online (Maia 2011, tab5.2d.txt).

$$\begin{aligned}
s = & -\frac{XY}{2} + 0.000\ 094'' + 0.003\ 808\ 65T_{TT} - 0.000\ 122\ 68T_{TT}^2 - 0.072\ 574\ 11T_{TT}^3 \\
& + 0.000\ 027\ 98T_{TT}^4 + 0.000\ 015\ 62T_{TT}^5 \\
& + \sum_{i=1}^{33} [A_{ss0_i} \sin(a_{p_i}) + A_{sc0_i} \cos(a_{p_i})] + \sum_{i=1}^3 [A_{ss1_i} \sin(a_{p_i}) + A_{sc1_i} \cos(a_{p_i})] T_{TT} \\
& + \sum_{i=1}^{25} [A_{ss2_i} \sin(a_{p_i}) + A_{sc2_i} \cos(a_{p_i})] T_{TT}^2 + \sum_{i=1}^4 [A_{ss3_i} \sin(a_{p_i}) + A_{sc3_i} \cos(a_{p_i})] T_{TT}^3 \\
& + \sum_{i=1}^1 [A_{ss4_i} \sin(a_{p_i}) + A_{sc4_i} \cos(a_{p_i})] T_{TT}^4
\end{aligned} \tag{3-64}$$

Petit and Luzum (2010:59) show a simplification for the summations which only includes terms larger than $0.5''$. It's valid for the period 1975-2025. Notice that the T_{TT}^4, T_{TT}^5 terms are removed from Eq. (3-64).

$$\begin{aligned}
s = & -\frac{XY}{2} + 0.000\ 094'' + 0.003\ 808\ 65T_{TT} - 0.000\ 122\ 68T_{TT}^2 - 0.072\ 574\ 11T_{TT}^3 \\
& - 0.002\ 640\ 73 \sin(\Omega_{\zeta}) - 0.000\ 063\ 53 \sin(2\Omega_{\zeta}) \\
& - 0.000\ 011\ 75 \sin(2u_{M\zeta} - 2D_{\odot} + 3\Omega_{\zeta}) - 0.000\ 011\ 21 \sin(2u_{M\zeta} - 2D_{\odot} + \Omega_{\zeta}) \\
& + 0.000\ 004\ 57 \sin((2u_{M\zeta} - 2D_{\odot} + 2\Omega_{\zeta})) - 0.000\ 002\ 02 \sin(2u_{M\zeta} + 3\Omega_{\zeta}) \\
& - 0.000\ 001\ 98 \sin((2u_{M\zeta} + \Omega_{\zeta}) + 0.000\ 001\ 72 \sin(3\Omega_{\zeta})) \\
& + 0.000\ 001\ 41 \sin((M_{\odot} + \Omega_{\zeta}) + 0.000\ 001\ 26 \sin(M_{\odot} - \Omega_{\zeta})) \\
& + 0.000\ 000\ 63 \sin((M_{\zeta} + \Omega_{\zeta}) + 0.000\ 000\ 63 \sin(M_{\zeta} - \Omega_{\zeta})) \\
& + 0.000\ 001\ 73 T_{TT} \sin(\Omega_{\zeta}) + 0.000\ 003\ 57 T_{TT} \cos(2\Omega_{\zeta}) + 0.000\ 743\ 52 T_{TT}^2 \sin(\Omega_{\zeta}) \\
& + 0.000\ 056\ 91 T_{TT}^2 \sin(2(u_{M\zeta} - D_{\odot} + \Omega_{\zeta})) + 0.000\ 009\ 84 T_{TT}^2 \sin(2(u_{M\zeta} + \Omega_{\zeta})) \\
& - 0.000\ 008\ 85 T_{TT}^2 \sin(2\Omega_{\zeta})
\end{aligned} \tag{3-65}$$

Classical Equinox Based Transformations with IAU-2006/2000

The previous reduction may also be expressed in a classical equinox-based notation. I present the parameters used in this approach so you can compare with the former IAU-76/FK5 theory discussed later, but the CIO approach is usually preferred. Polar motion is handled as in Eq. (3-60).

The formation of the Greenwich Apparent Sidereal Time is different from the former theories. It now uses the Earth Rotation Angle, θ_{ERA} [Eq. (3-62)], and the **Equation of the Origins** (EO). The constants and complete equation for EO are found online (Maia, 2011 tab5.2e.txt).

$$\begin{aligned}
EO = & -0.014\ 506'' - 4612.156\ 534T_{TT} - 1.391\ 581\ 7T_{TT}^2 + 0.000\ 000\ 44T_{TT}^3 \\
& - 0.000\ 299\ 56T_{TT}^4 - 3.68 \times 10^{-8}T_{TT}^5 - \Delta\Psi_{2000}\cos(\bar{\epsilon}_{2000}) \\
& - \sum_{i=1}^{33} [A_{gs0_i}\sin(a_{p_i}) + A_{gc0_i}\cos(a_{p_i})] - \sum_{i=1}^1 [A_{gs1_i}\sin(a_{p_i}) + A_{gc1_i}\cos(a_{p_i})]T_{TT}
\end{aligned} \tag{3-66}$$

An approximate expression for EO is given by Petit and Luzum (2010:61). The approach is similar to the approximation in Eq. (3-65). Beware that the trigonometric terms are negative from Eq. (3-65) and a few coefficients are slightly different.

$\theta_{GAST2000}$ uses both the ERA and EO angles.

$$\begin{aligned}
\theta_{GAST2000} &= \theta_{ERA} - EO \\
[\mathbf{R}] &= \text{ROT3}(\theta_{GAST2000})
\end{aligned}$$

Applications may require the Greenwich Mean Sidereal Time and the equation of the equinoxes. Petit and Luzum (2010:60-61) show the expressions.*

$$\begin{aligned}
Eq_{equinox2000} &= \Delta\Psi_{2000}\cos(\bar{\epsilon}_{2000}) \\
\theta_{GMST2000} &= \theta_{ERA} + 0.014\ 506'' + 4612.156\ 534T_{TT} + 1.391\ 581\ 7T_{TT}^2 - 0.000\ 000\ 44T_{TT}^3 \\
&\quad - 0.000\ 299\ 56T_{TT}^4 - 3.68 \times 10^{-8}T_{TT}^5
\end{aligned} \tag{3-67}$$

The transformations use the obliquity of the ecliptic, ϵ . Earth precession causes the obliquity to move some. Thus, a mean obliquity of the ecliptic is (Kaplan, 2005:44)

$$\begin{aligned}
\bar{\epsilon}_{2000} &= 84,381.406'' - 46.836\ 769T_{TT} - 0.000\ 183\ 1T_{TT}^2 + 0.002\ 003\ 40T_{TT}^3 \\
&\quad + 5.76 \times 10^{-7}T_{TT}^4 + 4.34 \times 10^{-8}T_{TT}^5 \\
\bar{\epsilon}_{2000} &= 23.439\ 279^\circ - 0.013\ 010\ 2T_{TT} - 5.086 \times 10^{-8}T_{TT}^2 + 5.565 \times 10^{-7}T_{TT}^3 \\
&\quad + 1.6 \times 10^{-10}T_{TT}^4 + 1.21 \times 10^{-11}T_{TT}^5
\end{aligned} \tag{3-68}$$

where T_{TT} is the Julian centuries from J2000.0 (TT) and $\epsilon_o = 84,381.406''$.†

The nutation ($\Delta\Psi_{2000}$) and obliquity ($\Delta\epsilon_{2000}$) corrections are of similar form to previous reduction theories. The classical IAU-76/FK5 transformation, there are contributions from nutation and planetary terms. Be careful to use the correct formulae, coefficients, and units. The arguments (nutation and planetary) are different from Eq. (3-63). The argument subscript is now n for nutation. The out-of-phase contributions are not used in the equinox approaches (Kaplan 2005:45-46). The planetary terms contrib-

* Note that although Kaplan (2005:48) shows the true obliquity in $Eq_{equinox2000}$, the narrative suggests it's more appropriately the mean value.

† The coefficient of the first term in Eq. (3-68) is the approximate decrease in the obliquity of the ecliptic resulting from precession of the ecliptic ($0.013\ 011^\circ$ [$\sim 47''$] per century)

ute only in the first summation, and the angle does not include them in the second summation. The constants are in IERS documents (Kaplan 2005), and online (Maia, 2011b, tab5.3a.txt and tab5.3b.txt). Notice this is the former IAU-2000A nutation theory. Corrections are added later. Newer forms do not require these corrections.

$$\begin{aligned}
 \Delta\Psi_{2000n} &= \sum_{\substack{i=1 \\ 38}}^{1320} \{A_{ps0_i}\} \sin(a_{p_i}) + \{A_{pc0_i}\} \cos(a_{p_i}) \\
 &\quad + \sum_{\substack{i=1 \\ 1037}}^{19} \{A_{ps1_i} T_{TT}\} \sin(a_{p_i}) + \{A_{pc1_i} T_{TT}\} \cos(a_{p_i}) \\
 \Delta\epsilon_{2000n} &= \sum_{\substack{i=1 \\ 19}} \{A_{ec0_i}\} \cos(a_{p_i}) + \{A_{es0_i}\} \sin(a_{p_i}) \\
 &\quad + \sum_{\substack{i=1 \\ 19}} \{A_{ec1_i} T_{TT}\} \cos(a_{p_i}) + \{A_{es1_i} T_{TT}\} \sin(a_{p_i}) \\
 a_{p_i} &= a_{p1_i} M_{\mathbb{C}} + a_{p2_i} M_{\odot} + a_{p3_i} u_{M\mathbb{C}} + a_{p4_i} D_{\odot} + a_{p5_i} \Omega_{\mathbb{C}} + a_{p6_i} \lambda_{M\mathbb{C}} + a_{p7_i} \lambda_{M\odot} \\
 &\quad + a_{p8_i} \lambda_{M\oplus} + a_{p9_i} \lambda_{M\mathcal{S}} + a_{p10_i} \lambda_{M\mathcal{U}} + a_{p11_i} \lambda_{M\mathbb{H}} + a_{p12_i} \lambda_{M\mathbb{D}} + a_{p13_i} \lambda_{M\mathbb{P}} + a_{p14_i} p_{\lambda}
 \end{aligned} \tag{3-69}$$

The planetary contribution series expression is also included (the IAU-2000B Theory uses constant values, $\Delta\Psi_{2000pl} = -0.000\ 135''$ and $\Delta\epsilon_{2000pl} = 0.000\ 388''$).

There are additional corrections for **Free Core Nutation** (FCN) which models the effect of a liquid Earth core. These corrections are provided directly in the EOP data, but you can also convert the corrections to X and Y (dX, dY) to find an estimate for these quantities (Vallado et al. 2006). They are very small and often omitted (Kaplan, 2005:47). The complete nutation matrix includes a small correction for the IAU2000 to IAU2006 approaches (J_{2d}) using these expressions.

$$\begin{aligned}
 J_{2d} &= -2.7774 \times 10^{-6} / \text{century} \\
 \Delta\Psi_{2000} &= \Delta\Psi_{2000} + \Delta\Psi_{2000}(0.4697 \times 10^{-6} J_{2d}) \\
 \Delta\epsilon_{2000} &= \Delta\epsilon_{2000} + \Delta\epsilon_{2000} J_{2d}
 \end{aligned} \tag{3-70}$$

The true obliquity of the ecliptic is found by summing the corrections.

$$\epsilon = \bar{\epsilon}_{2000} + \Delta\epsilon_{2000} \tag{3-71}$$

The complete nutation matrix is formed via rotations. Note that I leave the year subscript out of the complete matrix for generality with the IAU-76/FK5 approach discussed later on.

$$[N] = \text{ROT1}(-\bar{\epsilon}_{2000}) \text{ROT3}(\Delta\Psi_{2000}) \text{ROT1}(\bar{\epsilon}_{2000} + \Delta\epsilon_{2000}) \tag{3-72}$$

$$\begin{bmatrix} \cos(\Delta\Psi) & \sin(\Delta\Psi)\cos(\epsilon) & \sin(\epsilon)\sin(\Delta\Psi) \\ -\sin(\Delta\Psi)\cos(\bar{\epsilon}) & \cos(\epsilon)\cos(\Delta\Psi)\cos(\bar{\epsilon}) + \sin(\epsilon)\sin(\bar{\epsilon}) & \sin(\epsilon)\cos(\Delta\Psi)\cos(\bar{\epsilon}) - \sin(\bar{\epsilon})\cos(\epsilon) \\ -\sin(\Delta\Psi)\sin(\bar{\epsilon}) & \sin(\bar{\epsilon})\cos(\epsilon)\cos(\Delta\Psi) - \sin(\epsilon)\cos(\bar{\epsilon}) & \sin(\epsilon)\sin(\bar{\epsilon})\cos(\Delta\Psi) + \cos(\epsilon)\cos(\bar{\epsilon}) \end{bmatrix}$$

Kaplan (2005:43) gives the precession formulae for the P03 precession theory (Capitaine et al. 2003). The P03 equations are found using $\epsilon_o = 84,381.406''$. The original values were used until the P03 became the official standard on January 1, 2009 (Wallace and Capitaine, 2006).

$$\begin{aligned} \Psi_a &= 5038.481\ 507'' T_{TT} - 1.079\ 006\ 9 T_{TT}^2 - 0.001\ 140\ 45 T_{TT}^3 + 1.328\ 51 \times 10^{-4} T_{TT}^4 - 9.51 \times 10^{-8} T_{TT}^5 \\ \omega_a &= \epsilon_o - 0.025\ 754'' T_{TT} + 0.051\ 262\ 3 T_{TT}^2 - 0.007\ 725\ 03 T_{TT}^3 - 4.67 \times 10^{-7} T_{TT}^4 + 3.337 \times 10^{-7} T_{TT}^5 \\ \chi_a &= 10.556\ 403'' T_{TT} - 2.381\ 429\ 2 T_{TT}^2 - 0.001\ 211\ 97 T_{TT}^3 + 1.706\ 63 \times 10^{-4} T_{TT}^4 - 5.60 \times 10^{-8} T_{TT}^5 \\ [\mathbf{P}] &= \text{ROT3}(\chi_a) \text{ROT1}(-\omega_a) \text{ROT 3}(-\bar{\Psi}_a) \text{ROT1}(\bar{\epsilon}_{2000}) \end{aligned} \quad (3-73)$$

Fukushima-Williams (2003/1994) provide an equivalent form which includes the frame bias (Petit and Luzum: 2010). Wallace and Capitaine (2006) provide important details on the use as well.

$$\begin{aligned} \gamma &= -0.052\ 928'' + 10.556\ 378 T_{TT} + 0.493\ 204\ 4 T_{TT}^2 - 0.000\ 312\ 38 T_{TT}^3 - 2.788 \times 10^{-6} T_{TT}^4 \\ &\quad + 2.60 \times 10^{-8} T_{TT}^5 \\ \phi &= 84381.412\ 819'' - 46.811\ 016 T_{TT} + 0.051\ 126\ 8 T_{TT}^2 + 0.000\ 532\ 89 T_{TT}^3 - 4.40 \times 10^{-7} T_{TT}^4 \\ &\quad - 1.76 \times 10^{-8} T_{TT}^5 \\ \psi &= -0.041\ 775'' + 5038.481\ 484 T_{TT} + 1.558\ 417\ 5 T_{TT}^2 - 0.000\ 185\ 22 T_{TT}^3 - 2.645\ 2 \times 10^{-5} T_{TT}^4 \\ &\quad - 1.48 \times 10^{-8} T_{TT}^5 \\ [\mathbf{P}] &= \text{ROT3}(-\gamma) \text{ROT1}(-\phi) \text{ROT 3}(\psi) \text{ROT1}(\bar{\epsilon}_{2000}) \end{aligned} \quad (3-74)$$

The rotations for this approach replace the [PN] and [ERA] matrices in the CIO based approach. A frame bias matrix (\mathbf{B}) enables conversion from one conventional frame to another using the offsets from the ICRS pole (ξ_o and η_o) and the shift in the origin ($\delta\alpha_o$) (Astronomical Almanac 2006:B27, Kaplan, 2005:27). The frame bias matrix that permits conversion from the J2000.0 dynamical equinox to the GCRF is:^{*}

$$\begin{aligned} \delta\alpha_o &= -0.0146'' \\ \xi_o &= \delta\psi \sin(\epsilon_o) = -0.041\ 775'' \sin(84381.406) = -0.166\ 17'' \end{aligned}$$

* The frame bias is not a precise transformation for all times. If it were, there would be no advantage to a non-rotating origin! The frame bias is applicable only to the equinox based approach, and is only an approximation.

$$\eta_o = -0.006\ 819\ 2''$$

$$[\mathbf{B}] = \text{ROT3}(-\delta\alpha_o) \text{ROT2}(-\xi_o) \text{ROT1}(\eta_o)$$

$$[\mathbf{B}] = \begin{bmatrix} 0.9999\ 9999\ 9999\ 9942 & 0.0000\ 0007\ 0782\ 7948 & -0.0000\ 0008\ 0562\ 1738 \\ -0.0000\ 0007\ 0782\ 7974 & 0.9999\ 9999\ 9999\ 9969 & -0.0000\ 0003\ 3060\ 4088 \\ 0.0000\ 0008\ 0562\ 1715 & 0.0000\ 0003\ 3060\ 4145 & 0.9999\ 9999\ 9999\ 9962 \end{bmatrix}$$

Implementing the IAU-2010 Conventions

The preceding equations are complicated, partially compounded by the incremental development of the theories. I focus on the CIO approach because there are fewer ambiguities, and it has less reliance on EOP values. In addition, it is easier to generate long series for interpolation from the CIO series. We can visualize the transformation in Fig. 3-28. The distinctions of “date” and “epoch” are removed with the new theory.

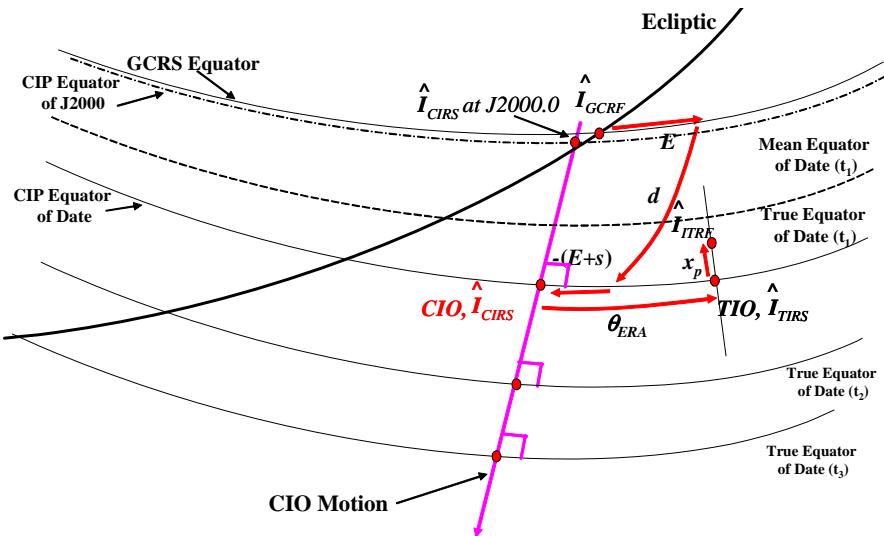


Figure 3-28. Celestial to Terrestrial Transformation. This exaggerated figure shows the CIO based approach to transform between celestial (GCRF) and terrestrial coordinate systems (ITRF). The motion of the CIO (CIRS) is perpendicular to the CIP equator at each time. The rotation for the y_p component of polar motion is not shown. The contribution of s' to polar motion is small and is therefore not shown. E and d are alternate variables for X and Y .

ALGORITHM 23: IAU-2000, CIO-Based

$$(\dot{\vec{r}}_{ITRF}, \dot{\vec{v}}_{ITRF}, yr, mo, day, UTC, \Delta UT1, \Delta AT, x_p, y_p \Rightarrow \dot{\vec{r}}_{GCRF}, \dot{\vec{v}}_{GCRF})$$

$$\text{CONVTIME} (yr, mo, day, UTC, \Delta UT1, \Delta AT \Rightarrow UT1, TAI, TT, T_{UTI}, T_{TT})$$

Find values for $a, X, Y, s', s, \theta_{ERA}$

$$[PN] = \begin{bmatrix} 1 - aX^2 & -aXY & X \\ -aXY & 1 - aY^2 & Y \\ -X & -Y & 1 - a(X^2 + Y^2) \end{bmatrix} \text{ROT3}(s)$$

$$[R] = \text{ROT3}(-\theta_{ERA})$$

$$[W] = \text{ROT3}(-s') \text{ROT2}(x_p) \text{ROT1}(y_p)$$

$$\begin{cases} \dot{\vec{r}}_{ITRF} = [W]^T [R]^T [PN]^T \dot{\vec{r}}_{GCRF} \\ \dot{\vec{v}}_{ITRF} = [W]^T \left\{ [R]^T [PN]^T \dot{\vec{v}}_{GCRF} - \vec{\omega}_{\oplus} \times \dot{\vec{r}}_{TIRS} \right\} \end{cases} \text{ or}$$

$$\begin{cases} \dot{\vec{r}}_{GCRF} = [PN][R][W]\dot{\vec{r}}_{ITRF} \\ \dot{\vec{v}}_{GCRF} = [PN][R] \left\{ [W]\dot{\vec{v}}_{ITRF} + \vec{\omega}_{\oplus} \times \dot{\vec{r}}_{TIRS} \right\} \end{cases}$$

▼ Example 3-14. Performing an IAU-2000 Reduction

$$\text{GIVEN: } \dot{\vec{r}}_{ITRF} = -1033.479\ 383\ 0 \hat{I} + 7901.295\ 275\ 4 \hat{J} + 6380.356\ 595\ 8 \hat{K} \text{ km}$$

$$\dot{\vec{v}}_{ITRF} = -3.225\ 636\ 520 \hat{I} - 2.872\ 451\ 450 \hat{J} + 5.531\ 924\ 446 \hat{K} \text{ km/}$$

$$\text{FIND: } \dot{\vec{r}}_{GCRF}, \dot{\vec{v}}_{GCRF} \text{ on April 6, 2004, 07:51:28.386 009 UTC}$$

Calculate the time quantities before starting the problem. Get ΔAT (32.0^s) directly from the *Astronomical Almanac* (2004:K9). From the IERS's EOPCO4 final data, find $\Delta\text{UT1} = -0.439\ 961\ 9^s$, $x_p = -0.140\ 682''$ and $y_p = 0.333\ 309''$, $dX = -0.000\ 205''$ and $dY = -0.000\ 136''$. Although general practice interpolates these values with a spline, don't interpolate these values for this example.

$$\text{UT1} = \text{UTC} + \Delta\text{UT1} = 07:51:28.386\ 009 - 0.439\ 961\ 9 = 07:51:27.946\ 047$$

$$\text{TAI} = \text{UTC} + \Delta\text{AT} = 07:51:28.386\ 009 + 32^s = 07:52: 0.386\ 009$$

$$\text{TT} = \text{TAI} + 32.184^s = 07:52: 0.386\ 009 + 32.184^s = 07:52:32.570\ 009$$

$$\text{DMY HMS} \Rightarrow \text{JD}_{\text{TT}} = 2,453,101.828\ 154\ 745 \text{ (Be sure to keep all digits for accuracy.)}$$

$$T_{\text{TT}} = \frac{\text{JD}_{\text{TT}} - 2,451,545.0}{36,525} = \frac{2,453,101.828\ 154\ 745 - 2,451,545.0}{36,525} = 0.042\ 623\ 631\ 9$$

First, include the effects of polar motion. Note that s' is -9.712×10^{-12} rad.

After doing the rotation, you'll get these PEF vectors:

$$\dot{\vec{r}}_{TIRS} = -1033.475\ 031\ 2 \hat{I} + 7901.305\ 585\ 6 \hat{J} + 6380.344\ 532\ 7 \hat{K} \text{ km}$$

$$\dot{\vec{v}}_{TIRS} = -3.225\,632\,747 \hat{I} - 2.872\,442\,511 \hat{J} + 5.531\,931\,288 \hat{K} \text{ km/s}$$

To find the Earth rotation angle, we use Eq. (3-62). $JD_{UT1} = 2,453,101.827\,406\,783$.

$$\theta_{ERA} = 312.755\,282\,9^\circ$$

The rotations produce the celestial intermediate reference position.

$$\dot{\vec{r}}_{CIRS} = 5100.018\,404\,7 \hat{I} + 6122.786\,364\,8 \hat{J} + 6380.344\,532\,7 \hat{K} \text{ km}$$

$$\dot{\vec{v}}_{CIRS} = -4.745\,380\,330 \hat{I} + 0.790\,341\,453 \hat{J} + 5.531\,931\,288 \hat{K} \text{ km/s}$$

Finally, determine the precession nutation matrix. Remember that the CIO formulation includes the frame bias, so no additional rotations are needed. Determine the angles for the Earth's nutation ($r = 360^\circ$).

$$M_{\zeta} = 485868.249\,036'' + 1717915923.2178T_{TT} + 31.8792T_{TT}^2 + 0.051\,635T_{TT}^3 \\ - 0.000\,244\,70T_{TT}^4$$

$$M_{\odot} = 1287104.79305'' + 129596581.0481T_{TT} - 0.5532T_{TT}^2 + 0.000\,136T_{TT}^3 \\ - 0.000\,011\,49T_{TT}^4$$

$$u_{M\zeta} = 335779.526\,232'' + 1739527262.8478T_{TT} - 12.7512T_{TT}^2 - 0.001\,037T_{TT}^3 \\ + 0.000\,004\,17T_{TT}^4$$

$$D_{\odot} = 1072260.703\,69'' + 1602961601.2090T_{TT} - 6.3706T_{TT}^2 + 0.006\,593T_{TT}^3 \\ - 0.000\,031\,69T_{TT}^4$$

$$\Omega_{\zeta} = 450160.398\,036'' - 6962890.5431T_{TT} + 7.4722T_{TT}^2 + 0.007702T_{TT}^3 - 0.000\,059\,39T_{TT}^4$$

$$M_{\zeta} = 314.912\,287\,3^\circ, M_{\odot} = 91.939\,376\,9^\circ, u_{M\zeta} = 169.097\,004\,3^\circ, D_{\odot} = 196.751\,642\,8^\circ,$$

$$\Omega_{\zeta} = 42.604\,646\,7^\circ$$

Next, the planetary nutation values are found.

$$\lambda_{M\varphi} = 252.250\,905\,494^\circ + 149,472.674\,635\,8T_{TDB} = 143.319\,167^\circ$$

$$\lambda_{M\vartheta} = 181.979\,800\,853^\circ + 58,517.815\,674\,8T_{TDB} = 156.221\,635^\circ$$

$$\lambda_{M\oplus} = 100.466\,448\,494^\circ + 35,999.372\,852\,1T_{TDB} = 194.890\,465^\circ$$

$$\lambda_{M\ominus} = 355.433\,274\,605^\circ + 19,140.299\,314T_{TDB} = 91.262\,347^\circ$$

$$\lambda_{M\varpi} = 34.351\,483\,900^\circ + 3,034.905\,674\,64T_{TDB} = 163.710\,186^\circ$$

$$\lambda_{M\mathrm{h}} = 50.077\,471\,399\,8^\circ + 1,222.113\,794\,04T_{TDB} = 102.168\,400^\circ$$

$$\lambda_{M\mathrm{d}} = 314.055\,005\,137^\circ + 428.466\,998\,313T_{TDB} = 332.317\,825^\circ$$

$$\lambda_{M\Psi} = 304.348\,665\,499^\circ + 218.486\,200\,208T_{TDB} = 313.661\,341^\circ$$

$$p_{\lambda} = 1.396\,971\,372\,14^\circ T_{TDB} + 0.000\,308\,6T_{TDB}^2 = 0.059\,545^\circ$$

Now find the coordinates X , Y , and s . I show only the final answers due to the length of the calculations. Be sure to add in the EOP corrections dX and dY .

$$X = 80.531\,880'', Y = 7.273\,921'', s = -0.003\,027'', \text{ and } a = 28.647\,891^\circ.$$

Convert to radians and find the final inertial coordinates.

$$\dot{\vec{r}}_{GCRF} = 5102.508\,959 \hat{i} + 6123.011\,403 \hat{j} + 6378.136\,925 \hat{k} \text{ km}$$

$$\dot{\vec{v}}_{GCRF} = -4.743\,220\,16 \hat{i} + 0.790\,536\,50 \hat{j} + 5.533\,755\,73 \hat{k} \text{ km/s}$$



3.7.2 Velocity and Acceleration Transformations

The velocity transformation is similar to that already presented, except we can change the sidereal time rotation, $[R]$, to reflect the angular velocity of the Earth with respect to a Newtonian-inertial frame. We use the time derivative of the stellar angle [Eq. (3-62)] to define the rotation rate of the Earth

$$\omega_{\oplus} = 7.292\,115\,146\,706\,979 \times 10^{-5} \left\{ 1 - \frac{LOD}{86400} \right\} \quad (3-75)$$

where **Length of Day**, LOD is the instantaneous rate of change (in seconds) of UT1 with respect to a uniform time scale (UTC or TAI). LOD is maintained by the IERS. Although $\omega_{\oplus,prec}$ doesn't reference an inertial equinox, the difference is small for most satellite applications. However, the difference between ω_{\oplus} and $\omega_{\oplus,prec}$ [Eq. (3-40)] is about as large as the average effect of LOD ($\sim 10^{-11}$ rad/sec). Thus, applications using LOD will likely require ω_{\oplus} in lieu of $\omega_{\oplus,prec}$. Abbreviating the rotations, we have (see Eq. (3-26) and note the frame bias is only for the IAU-2000 equinox based approaches)

$$\begin{aligned} \dot{\vec{v}}_{GCRF} &= [\mathbf{B}][\mathbf{P}][\mathbf{N}][\mathbf{R}] \left\{ [\mathbf{W}] \dot{\vec{v}}_{ITRF} + \dot{\vec{\omega}}_{\oplus} \times \dot{\vec{r}}_{TIRS} \right\} \\ \dot{\vec{v}}_{ITRF} &= [\mathbf{W}]^T \left\{ [\mathbf{R}]^T[\mathbf{N}]^T[\mathbf{P}]^T[\mathbf{B}]^T \dot{\vec{v}}_{GCRF} - \dot{\vec{\omega}}_{\oplus} \times \dot{\vec{r}}_{TIRS} \right\} \\ \dot{\vec{a}}_{GCRF} &= [\mathbf{B}][\mathbf{P}][\mathbf{N}][\mathbf{R}] \left\{ [\mathbf{W}] \dot{\vec{a}}_{ITRF} + \dot{\vec{\omega}}_{\oplus} \times \dot{\vec{\omega}}_{\oplus} \times \dot{\vec{r}}_{TIRS} + 2\dot{\vec{\omega}}_{\oplus} \times \dot{\vec{v}}_{TIRS} \right\} \end{aligned} \quad (3-76)$$

By using $\dot{\vec{r}}_{TIRS}$, we correct the velocity before 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. The literature contains alternate notations. Using the IAU-76/FK5 position vector and the time rate of change of the rotation matrix for sidereal time (because it's the only parameter that varies significantly) gives us the same results [see also Eq. (3-27)]:

$$\begin{aligned} \dot{\vec{v}}_{ITRF} &= [\mathbf{W}]^T[\dot{\mathbf{R}}]^T[\mathbf{N}]^T[\mathbf{P}]^T[\mathbf{B}]^T \dot{\vec{r}}_{GCRF} + [\mathbf{W}]^T[\mathbf{R}]^T[\mathbf{N}]^T[\mathbf{P}]^T[\mathbf{B}]^T \dot{\vec{v}}_{GCRF} \\ \dot{\vec{v}}_{GCRF} &= [\mathbf{B}][\mathbf{P}][\mathbf{N}][\dot{\mathbf{R}}][\mathbf{W}] \dot{\vec{r}}_{ITRF} + [\mathbf{B}][\mathbf{P}][\mathbf{N}][\mathbf{R}][\mathbf{W}] \dot{\vec{v}}_{ITRF} \end{aligned}$$

$$\dot{[\mathbf{R}]} = \begin{bmatrix} -\omega_{\oplus} \sin(\theta_{GAST}) & -\omega_{\oplus} \cos(\theta_{GAST}) & 0 \\ \omega_{\oplus} \cos(\theta_{GAST}) & -\omega_{\oplus} \sin(\theta_{GAST}) & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Although this approach yields identical answers, it doesn't show the dependence on the rotating coordinate systems. If the acceleration is also needed, use Eq. (3-26) similar to the use of Eq. (3-25) in Eq. (3-76) above. Finally to reverse the transformation, be careful of the order of operations and remember to use transposes of rotation matrices.

3.7.3 IAU-76/FK5 Reduction

Although this is no longer the current theory, I've included a discussion adapted from McCarthy (1996), Frank (1984), Seidelmann (1992:99-145), the *Astronomical Almanac* (1992:B18-B19), NIMA (2000), and the Defense Mapping Agency (1987b:A15-A31) because many programs still use this approach. The EOP corrections permit continued use of the IAU-76/FK5 approach.

Polar Motion (IAU-76/FK5)

To do this transformation rigorously involves two rotations. For ITRF to PEF,

$$\hat{r}_{PEF} = \text{ROT1}(y_p)\text{ROT2}(x_p)\hat{r}_{ITRF} \quad [W] = \begin{bmatrix} \cos(x_p) & 0 & -\sin(x_p) \\ \sin(x_p)\sin(y_p) & \cos(y_p) & \cos(x_p)\sin(y_p) \\ \sin(x_p)\cos(y_p) & -\sin(y_p) & \cos(x_p)\cos(y_p) \end{bmatrix} \quad (3-77)$$

Because these values are very small—typically only a few meters in each direction—we often 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

$$\hat{r}_{PEF} \approx \begin{bmatrix} 1 & 0 & -x_p \\ 0 & 1 & y_p \\ x_p & -y_p & 1 \end{bmatrix} \hat{r}_{ITRF} \quad (3-78)$$

Be aware of two observations: (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 very long periods because the motion isn't fully understood.

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

$$\Delta\phi_{gd} \approx x_p \cos(\bar{\lambda}) - y_p \sin(\bar{\lambda}) \quad \Delta\lambda \approx \{x_p \sin(\bar{\lambda}) - y_p \cos(\bar{\lambda})\} \tan(\bar{\phi})$$

Sidereal Time (IAU-76/FK5)

The second transformation is responsible for most of the positional difference of the vector components. It transforms the rotating coordinate system to a nonrotating true of date frame. This motion addresses the variability in sidereal time and requires the

Greenwich *apparent* sidereal time because the rotating frame references the Earth's actual equator of date—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 true equator relative to the mean equinox. The difference between these two times is called the equation of the equinoxes. This quantity is really the accumulated precession and nutation in right ascension. It's obtained by projecting the difference between the mean and true equinoxes onto the true equator—thus a simple formula for the *equation of the equinoxes* is $\Delta\Psi_{1980}\cos(\epsilon_{1980})$, which we will see 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_{1980} . Be aware that before February 27, 1997 (IAU 1994 Resolutions), the last two terms (using Ω_{\odot}) were not used (an effect of less than about 1 m). These terms make the (newer) kinematic and (older) geometric definitions more consistent. From Fig. 3-29, we find the Greenwich apparent sidereal time, $\theta_{GAST1982}$, using the Greenwich mean sidereal time [Eq. (3-47) or Eq. (3-49) and Eq. (3-46)], plus the equation of the equinoxes (McCarthy, 1992:30):

$$\begin{aligned} Eq_{equinox1982} &= \Delta\Psi_{1980}\cos(\epsilon_{1980}) + 0.002\ 64''\sin(\Omega_{\odot}) + 0.000\ 063\sin(2\Omega_{\odot}) \\ \theta_{GAST1982} &= \theta_{GMST1982} + Eq_{equinox1982} \end{aligned} \quad (3-79)$$

Notice the additional term for velocity transformations. The rotations are as follows.

$$\begin{aligned} \dot{\vec{r}}_{TOD} &= \text{ROT3}(-\theta_{GAST1982})\dot{\vec{r}}_{PEF} \\ \dot{\vec{v}}_{TOD} &= \text{ROT3}(-\theta_{GAST1982})\dot{\vec{v}}_{PEF} + \dot{\vec{\omega}}_{\oplus} \times \dot{\vec{r}}_{PEF} \end{aligned} \quad (3-80)$$

Nutation (IAU-76/FK5)

The next transformation accounts for the periodic effects contributed primarily by the Moon. The primary model is from the **1980 IAU Theory of Nutation**. We determine the primary variation from a trigonometric series of 106 terms. The number of terms was driven by the arc-second accuracy requirements at the development time. The result is periodic, rather than secular like the precession effect. Adding these effects transforms the *true equator of date*, TOD, to the *mean* equator of date. Figure 3-29 shows both the mean and true obliquity of the ecliptic values.

The first step is to determine the mean obliquity of the ecliptic. IAU-76/FK5 defines the obliquity of the ecliptic (at J2000.0) as $\epsilon_o = -23^{\circ}26'21.448'' = 23.439\ 291^{\circ}$. We use a form similar to Eq. (3-68) and the angles from Eq. (3-58). From McCarthy (1992:31) and USNO Circ 163:A8.

* Because the angle is actually measured along the mean equator and then projected onto the true equator, we use the mean obliquity of the ecliptic.

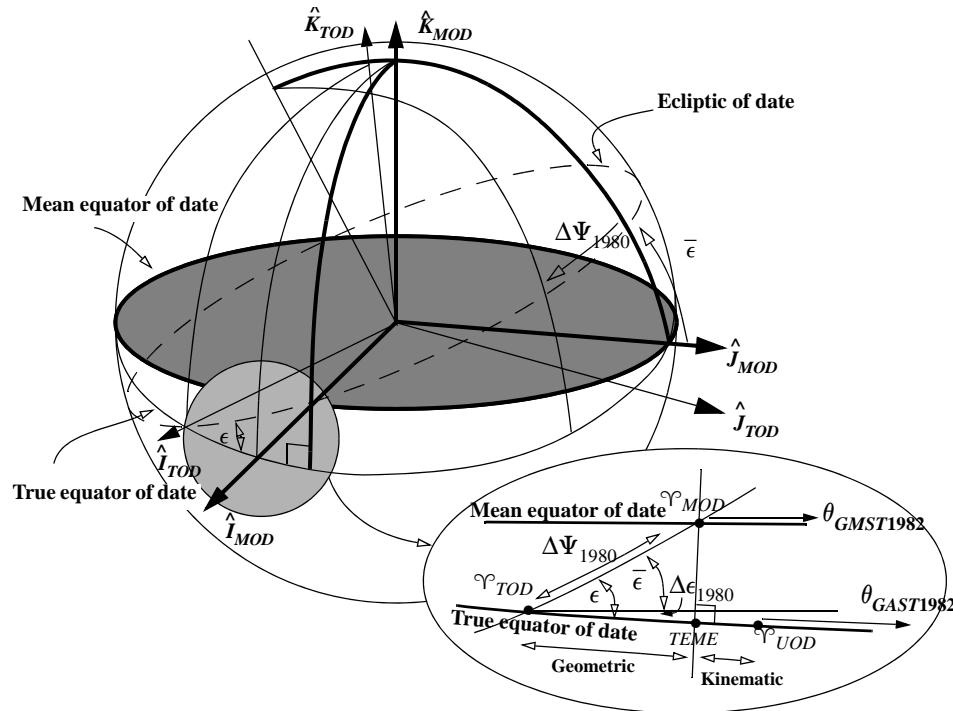


Figure 3-29. Transformation Geometry Due to Nutation. Nutation moves the state from the true equinox to the mean equinox of date. The starting point is the true equator and ecliptic of date found from the sidereal time calculations. The geometric (before 1997) and kinematic portions (afterwards) of the equation of the equinoxes are shown. The uniform of date (UOD) frame is sometimes used.

$$\begin{aligned}\bar{\epsilon}_{1980} &= 84,381.448'' - 46.8150T_{TT} - 0.000\,59T_{TT}^2 + 0.001\,813T_{TT}^3 \\ \bar{\epsilon}_{1980} &= 23.439\,291^\circ - 0.013\,004\,2T_{TT} - 1.64 \times 10^{-7}T_{TT}^2 + 5.04 \times 10^{-7}T_{TT}^3\end{aligned}\quad (3-81)$$

Once we find these values, we evaluate the Delaunay parameters because they are different from the IAU-2000 formulae. From McCarthy (1992:32) with $r = 360^\circ$ in ",

$$\begin{aligned}M_{\mathbb{C}} &= 134.963\,402\,51^\circ + (1325r + 198.867\,560\,5)T_{TT} + 0.008\,855\,3T_{TT}^2 + 1.4343 \times 10^{-5}T_{TT}^3 \\ M_{\odot} &= 357.529\,109\,18^\circ + (99r + 359.050\,291\,1)T_{TT} - 0.000\,153\,7T_{TT}^2 + 3.8 \times 10^{-8}T_{TT}^3 \\ u_{M\mathbb{C}} &= 93.272\,090\,62^\circ + (1342r + 82.017\,457\,7)T_{TT} - 0.003\,542\,0T_{TT}^2 - 2.88 \times 10^{-7}T_{TT}^3 \\ D_{\odot} &= 297.850\,195\,47^\circ + (1236r + 307.111\,446\,9)T_{TT} - 0.001\,769\,6T_{TT}^2 + 1.831 \times 10^{-6}T_{TT}^3 \\ \Omega_{\mathbb{C}} &= 125.044\,555\,01^\circ - (5r + 134.136\,185\,1)T_{TT} + 0.002\,075\,6T_{TT}^2 + 2.139 \times 10^{-6}T_{TT}^3\end{aligned}\quad (3-82)$$

Later updates to the original IAU-76/FK5 theory included corrections for planetary effects on the nutation and the obliquity of the ecliptic. However, to strictly adhere to the original definition (USNO Circ 163), they are not included.

We then evaluate a trigonometric series to find the nutation in longitude, $\Delta\Psi_{1980}$, and the nutation in obliquity, $\Delta\epsilon_{1980}$. Table D-6 lists sample coefficients for the integers ($a_{n1i} - a_{n5i}$) and the real coefficients (A_i, B_i, C_i, D_i).

$$\begin{aligned}\Delta\Psi_{1980} &= \sum_{i=1}^{106} (A_i + B_i T_{TT}) \sin\{a_{p_i}\} & \Delta\epsilon_{1980} &= \sum_{i=1}^{106} (C_i + D_i T_{TT}) \cos\{a_{p_i}\} \\ a_{p_i} &= a_{n1i} M_{\mathbb{C}} + a_{n2i} M_{\odot} + a_{n3i} u_M M_{\mathbb{C}} + a_{n4i} D_{\odot} + a_{n5i} \Omega_{\mathbb{C}}\end{aligned}\quad (3-83)$$

Despite the effort to model nutation with these calculations, discrepancies exist when processing *Very Long Baseline Interferometry* (VLBI) and *Lunar Laser Ranging* (LLR) data. Thus, if compatibility with the GCRF is needed, you use the 106 term series, and add EOP corrections ($\delta\Delta\epsilon_{1980}, \delta\Delta\Psi_{1980}$) to be consistent with the GCRS. The combined update for the nutation and the obliquity of the ecliptic are

$$\Delta\Psi_{1980} = \Delta\Psi_{1980} + \delta\Delta\Psi_{1980} \quad \Delta\epsilon_{1980} = \Delta\epsilon_{1980} + \delta\Delta\epsilon_{1980} \quad (3-84)$$

The EOP corrections are *not* used for applications referencing the original IAU-76/FK5 theory. Beware that the differences can be several meters.

Finally, the true obliquity of the ecliptic is needed.

$$\epsilon_{1980} = \bar{\epsilon}_{1980} + \Delta\epsilon_{1980} \quad (3-85)$$

Figure 3-29 shows the transformation. The rotations needed to rigorously transform coordinates for nutation are (velocity is transformed the same as position) (McCarthy, 1992:29)

$$\hat{\vec{r}}_{mod} = \text{ROT1}(-\bar{\epsilon}_{1980}) \text{ROT3}(\Delta\Psi_{1980}) \text{ROT1}(\epsilon_{1980}) \hat{\vec{r}}_{tod} \quad (3-86)$$

The complete matrix for transformations from TOD to MOD is the same as Eq. .

Precession (IAU-76/FK5)

The **IAU 1976 Precession** theory models the transformation of systems with certain angles shown in Fig. 3-30. The transformation is obtained by recognizing that a common point exists between the mean equator at J2000.0 and the ecliptic of date. This permits a motion along the mean equator at J2000.0, a translation from the mean equator of date, and a final motion to the location of the equinox of date. The process converts a vector having the **mean equinox of date** (MOD), to a vector in the IAU-76/FK5 system, having a mean equinox at J2000.0. The angles ζ, Θ , and z represent the combined effects of general precession. Fortunately, these angles are secular values and are easily calculated (Lieske et al. 1977, and Seidelmann, 1992:104). Eq. (3-87) presents the formulas necessary to determine the precession angles. We use a base epoch (J2000.0 in this

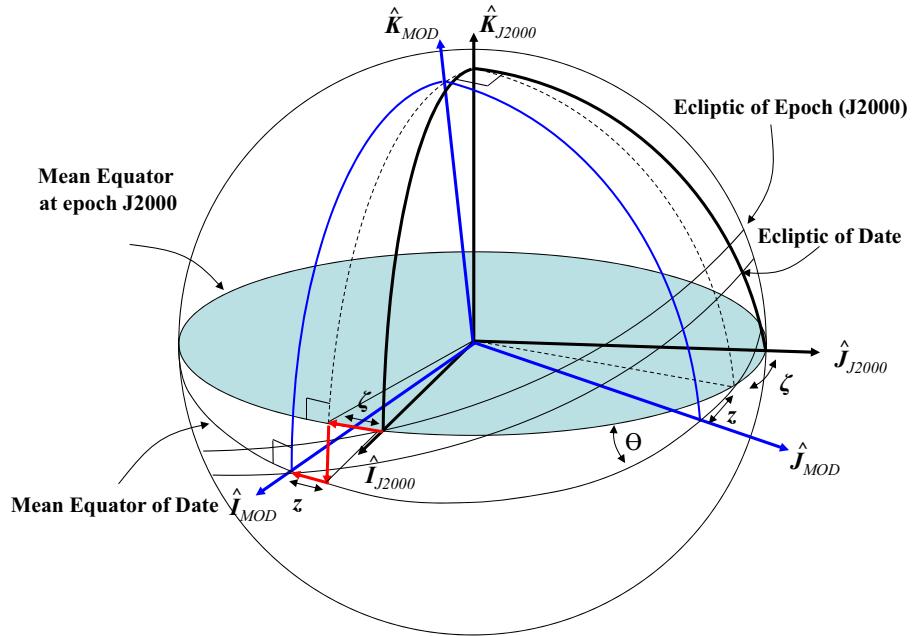


Figure 3-30. Transformation Geometry Due to Precession. Accounting for precession moves from the mean of date values to mean of epoch values. Typically, the epoch is J2000.0. The diagram is greatly exaggerated. In order of application, the angles are ζ , Θ , and z .

case), an arbitrary fixed epoch, JD_{start} , and the epoch of date, JD_{end} . Then, we use these Julian dates (TDB \equiv TT is the suggested practice after 1994) to determine the number of Julian centuries required for the formulas. (USNO Circ 163:A2)

$$\begin{aligned}
 T_o &= \frac{JD_{start} - J2000.0}{36,525} & T_\Delta &= \frac{JD_{end} - JD_{start}}{36,525} \\
 \zeta &= (2306.2181'' + 1.396\,56T_o - 0.000\,139T_o^2)T_\Delta + (0.301\,88 - 0.000\,344T_o)T_\Delta^2 + 0.017\,998T_\Delta^3 & (3-87) \\
 \Theta &= (2004.3109'' - 0.853\,30T_o - 0.000\,217T_o^2)T_\Delta + (-0.426\,65 - 0.000\,217T_o)T_\Delta^2 - 0.041\,833T_\Delta^3 \\
 z &= (2306.2181'' + 1.396\,56T_o - 0.000\,139T_o^2)T_\Delta + (1.094\,68 + 0.000\,066T_o)T_\Delta^2 + 0.018\,203T_\Delta^3
 \end{aligned}$$

Eq. (3-87) allows direct transformation from a starting to ending date, without having to do an intermediate transformation to the J2000.0 epoch. In many cases, however, the fixed epoch and the base epoch are the same ($T_o = 0$), and we can simplify the formulas (McCarthy, 1992:31):

$$\begin{aligned}\xi &= 2306.2181''T_{TT} + 0.301\ 88T_{TT}^2 + 0.017\ 998T_{TT}^3 \\ \Theta &= 2004.3109''T_{TT} - 0.426\ 65T_{TT}^2 - 0.041\ 833T_{TT}^3 \\ z &= 2306.2181''T_{TT} + 1.094\ 68T_{TT}^2 + 0.018\ 203T_{TT}^3\end{aligned}\quad (3-88)$$

T_{TT} represents the number of Julian centuries of TT from the base epoch (J2000.0) [Eq. (3-42)]. Equations also exist using the “four-term” precession similar to Eq. (3-73). The order of the rotations is important, and for precession, there are identical transformations for position and velocity.

$$\dot{\vec{r}}_{GCRF} = \text{ROT3}(\xi)\text{ROT2}(-\Theta)\text{ROT3}(z)\dot{\vec{r}}_{mod} \quad (3-89)$$

The complete rotation matrix for transformations from MOD to GCRF is*

$$\begin{bmatrix} \cos(\Theta)\cos(z)\cos(\xi) - \sin(z)\sin(\xi) & \sin(z)\cos(\Theta)\cos(\xi) + \sin(\xi)\cos(z) & \sin(\Theta)\cos(\xi) \\ -\sin(\xi)\cos(\Theta)\cos(z) - \sin(z)\cos(\xi) & -\sin(z)\sin(\xi)\cos(\Theta) + \cos(z)\cos(\xi) & -\sin(\Theta)\sin(\xi) \\ -\sin(\Theta)\cos(z) & -\sin(\Theta)\sin(z) & \cos(\Theta) \end{bmatrix}$$

Summary (IAU-76/FK5)

We can apply the transformation equations in either direction, recognizing that the order of rotations is important for specific steps of the transformation. To transform completely between the celestial (GCRF) and the terrestrial (ITRF) frames, we’ll need to do four rotations and conversions: precession, nutation, sidereal time, and polar motion. Be sure to include the EOP corrections as they are required to be compatible with the GCRF. Less accurate calculations sometimes use just the transformation for sidereal time. This option is often exercised but seldom documented!

Implementing IAU-76/FK5 Reductions

Algorithm 24, defines IAU-76/FK5 reductions to and from the GCRF. If you program these routines on a computer, you should be aware of a couple of points. First, the summation for the series should be done in reverse order to preserve accuracy during the operation. The equations should use the original published units. For instance, many of the equations are given in arcseconds or radians. These units should be maintained and not converted for ease of programming. This will also ease future comparisons if the coefficients change slightly. The computer program can convert the values during program execution as necessary. The complete process is shown in Fig. 3-31.

ALGORITHM 24: IAU-76/FK5 REDUCTION

$$(\dot{\vec{r}}_{GCRF}, \dot{\vec{v}}_{GCRF}, yr, mo, day, UTC, \Delta UT1, \Delta AT, x_p, y_p \Rightarrow \dot{\vec{r}}_{ITRF}, \dot{\vec{v}}_{ITRF})$$

* Note that if the EOP corrections are not used, the transformation will reference the Mean of J2000.0 epoch.

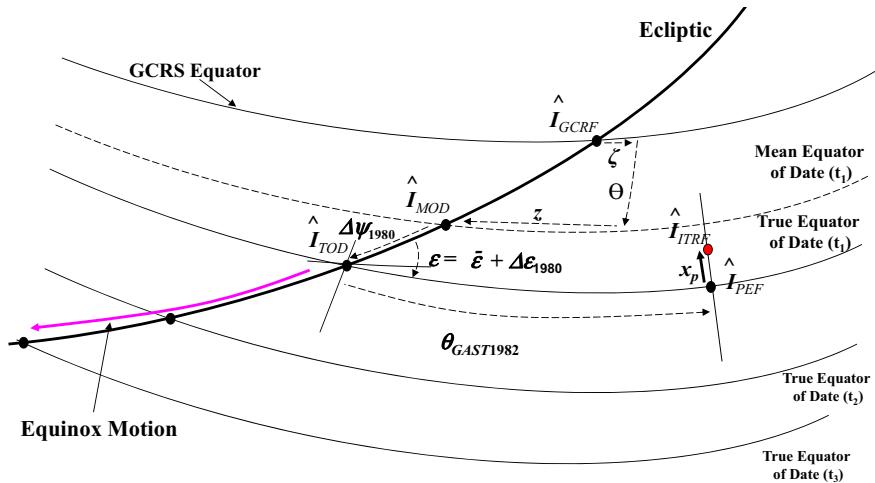


Figure 3-31. IAU-76/FK5 Classical Transformation. This exaggerated figure depicts the transformation of a state vector in the inertial (GCRF) frame to the body fixed (ITRF) frame. This two-way conversion is necessary for many orbit determination problems. There are several intermediate frames. Note the similarities to Fig. 3-28.

CONVTIME (yr, mo, day, UTC, ΔUT1, ΔAT ⇒ UT1, TAI, TT, T_{UT1}, T_{TT})

Find precession and nutation angles $z, \Theta, \xi, \epsilon, \Delta\Psi_{1980}, \bar{\epsilon}_{1980}, \theta_{GAST1982}$

$$[P] = \text{ROT3}(\xi)\text{ROT2}(-\Theta)\text{ROT3}(z)$$

$$[N] = \text{ROT1}(-\bar{\epsilon}_{1980})\text{ROT3}(\Delta\Psi_{1980})\text{ROT1}(\epsilon_{1980})$$

$$[R] = \text{ROT3}(-\theta_{GAST1982})$$

$$[W] = \text{ROT1}(y_p)\text{ROT2}(x_p) \approx \begin{bmatrix} 1 & 0 & -x_p \\ 0 & 1 & y_p \\ x_p & -y_p & 1 \end{bmatrix}$$

$$\left\{ \begin{array}{l} \dot{\vec{r}}_{GCRF} = [P][N][R][W]\dot{\vec{r}}_{ITRF} \\ \dot{\vec{v}}_{GCRF} = [P][N][R]\left\{ [W]\dot{\vec{v}}_{ITRF} + \dot{\vec{\omega}}_{\oplus} \times \dot{\vec{r}}_{PEF} \right\} \end{array} \right. \text{ or}$$

$$\left\{ \begin{array}{l} \dot{\vec{r}}_{ITRF} = [W]^T[R]^T[N]^T[P]^T\dot{\vec{r}}_{GCRF} \\ \dot{\vec{v}}_{ITRF} = [W]^T\left\{ [R]^T[N]^T[P]^T\dot{\vec{v}}_{GCRF} - \dot{\vec{\omega}}_{\oplus} \times \dot{\vec{r}}_{PEF} \right\} \end{array} \right.$$

An example is useful to tie all the concepts together.

▼ **Example 3-15. Performing a IAU-76/FK5 Reduction.**

$$\text{GIVEN: } \vec{r}_{ITRF} = -1033.479\ 383\ 0 \hat{I} + 7901.295\ 275\ 4 \hat{J} + 6380.356\ 595\ 8 \hat{K} \text{ km}$$

$$\vec{v}_{ITRF} = -3.225\ 636\ 520 \hat{I} - 2.872\ 451\ 450 \hat{J} + 5.531\ 924\ 446 \hat{K} \text{ km/s}$$

$$\text{FIND: } \vec{r}_{GCRF}, \vec{v}_{GCRF} \text{ on April 6, 2004, 07:51:28.386 009UTC}$$

Calculate the time quantities before starting the problem. Get ΔAT (32.0^s) directly from the *Astronomical Almanac* (2004:K9). From the IERS's EOPCO4 final data, find ΔUT1 = -0.439 961 9^s, $x_p = -0.140\ 682''$ and $y_p = 0.333\ 309''$, $\delta\Delta\Psi_{1980} = -0.052\ 195''$ and $\delta\Delta\epsilon_{1980} = -0.003\ 875''$, and LOD = 0.001 556 3s. Although general practice interpolates these values with a spline, don't interpolate these values for this example.

$$UT1 = UTC + \Delta UT1 = 07:51:28.386 009 - 0.439 961 9 = 07:51:27.946 047$$

$$TAI = UTC + \Delta AT = 07:51:28.386 009 + 32^s = 07:52: 0.386 009$$

$$TT = TAI + 32.184^s = 07:52: 0.386 009 + 32.184^s = 07:52:32.570 009$$

DMY HMS ⇒ $JD_{TT} = 2,453,101.828\ 154\ 745$ (Be sure to keep all digits for accuracy.)

$$T_{TT} = \frac{JD_{TT} - 2,451,545.0}{36,525} = \frac{2,453,101.828\ 154\ 745 - 2,451,545.0}{36,525} = 0.042\ 623\ 631\ 9$$

First, determine the effects of polar motion and you'll get these initial PEF vectors:

$$\vec{r}_{PEF} = -1033.475\ 031\ 3 \hat{I} + 7901.305\ 585\ 6 \hat{J} + 6380.344\ 532\ 8 \hat{K} \text{ km}$$

$$\vec{v}_{PEF} = -3.225\ 632\ 747 \hat{I} - 2.872\ 442\ 511 \hat{J} + 5.531\ 931\ 288 \hat{K} \text{ km/s}$$

Next, determine the $\theta_{GMST1982}$ using the JD_{UT1} , Eq. (3-45), and Eq. (3-42):

$$\theta_{GMST1982} = 312.809\ 894\ 3^\circ$$

Find the AST using Eq. (3-79) with the last two terms because the example occurs in 2004

$$\theta_{GAST1982} = \theta_{GMST1982} + \Delta\Psi_{1980}\cos(\epsilon_{1980}) = 312.806\ 765\ 4^\circ$$

You'll need the velocity correction from

$$\vec{\omega}_\oplus \times \vec{r}_{PEF} = -0.576\ 172\ 29 \hat{I} - 0.075\ 362\ 19 \hat{J} + 0.000 \hat{K} \text{ km/s}$$

After doing the rotation for sidereal time, you'll find the vectors for true equinox and true equator of date:

$$\vec{r}_{tod} = 5094.514\ 780\ 4 \hat{I} + 6127.366\ 461\ 2 \hat{J} + 6380.344\ 532\ 8 \hat{K} \text{ km}$$

$$\vec{v}_{tod} = -4.746\ 088\ 567 \hat{I} + 0.786\ 077\ 222 \hat{J} + 5.531\ 931\ 288 \hat{K} \text{ km/s}$$

Determine the angles for the Earth's nutation ($r = 360^\circ$). Notice these are similar to those in Example 3-14.

$$M_{\mathbb{C}} = 314.911\ 859\ 0^\circ, M_{\odot} = 91.937\ 993\ 1^\circ, u_{M_{\mathbb{C}}} = 169.096\ 827\ 2^\circ,$$

$$D_{\odot} = 196.751\ 811\ 6^\circ, \Omega_{\mathbb{C}} = 42.604\ 614\ 0^\circ$$

Now find the obliquity of the ecliptic [Eq. (3-68), $\bar{\epsilon}_{1980} = 23.438\ 736\ 8^\circ$] and the nutation corrections for the true of date system, using the 106 coefficients from the IAU 1980 Theory of Nutation [Eq. (3-83)].

$$\Delta\Psi_{1980} = -0.003\ 410\ 8^\circ, \Delta\epsilon_{1980} = 0.002\ 031\ 6^\circ$$

$$\epsilon_{1980} = \bar{\epsilon}_{1980} + \Delta\epsilon_{1980} = 23.438\ 736\ 8 + 0.002\ 031\ 6 = 23.440\ 768\ 5^\circ$$

Check that the true and mean obliquities are very close. If not, recheck your calculations! Because we wish to reference the GCRF frame, we add the corrections ($\delta\Delta\Psi_{1980}$ and $\delta\Delta\epsilon_{1980}$) to the nutation and obliquity values. The rotations for precession produce an intermediate vector, still in a mean-equinox, mean-equator system of date:

$$\dot{\vec{r}}_{mod} = 5094.028\ 374\ 5 \hat{I} + 6127.870\ 816\ 4 \hat{J} + 6380.248\ 516\ 4 \hat{K} \text{ km}$$

$$\dot{\vec{v}}_{mod} = -4.746\ 263\ 052 \hat{I} + 0.786\ 014\ 045 \hat{J} + 5.531\ 790\ 562 \hat{K} \text{ km/s}$$

Finally, determine the rotation angles for precession from Eq. (3-88).

$$\zeta = 0.027\ 305\ 5^\circ$$

$$\Theta = 0.023\ 730\ 6^\circ$$

$$z = 0.027\ 305\ 9^\circ$$

Convert to radians and rotate to find the final Earth-centered inertial coordinates

$$\dot{\vec{r}}_{GCRF} = 5102.508\ 958 \hat{I} + 6123.011\ 401 \hat{J} + 6378.136\ 928 \hat{K} \text{ km}$$

$$\dot{\vec{v}}_{GCRF} = -4.743\ 220\ 16 \hat{I} + 0.790\ 536\ 50 \hat{J} + 5.533\ 755\ 28 \hat{K} \text{ km/s}$$

If we had chosen not to use the $\delta\Delta\Psi_{1980}$ and $\delta\Delta\epsilon_{1980}$ corrections, we find a vector that is close to the GCRF, but about a meter different.

$$\dot{\vec{r}}_{J2000} = 5102.5096 \hat{I} + 6123.01152 \hat{J} + 6378.1363 \hat{K} \text{ km}$$

$$\dot{\vec{v}}_{J2000} = -4.743\ 2196 \hat{I} + 0.790\ 536\ 6 \hat{J} + 5.533\ 756\ 19 \hat{K} \text{ km/s}$$

The length and complexity of this example 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 3-6 shows the position and velocity vectors through each of the reduction rotations.

3.7.4 Other Reductions

You may encounter two additional frames in operations—TEME and FK4. TEME coordinates are used by AFSPC with their SGP4 analytical theory (Sec 9.7.4). The result of SGP4 propagations will be in TEME. The FK4 system precedes the IAU-76/FK5, but some programs still use it. In both cases, it is important to convert to a standard modern system for further processing and analysis.

TEME Coordinates

The AFSPC analytical theory (SGP4) produces ephemerides in a “**true equator, mean equinox**” (TEME*) (Herrick, 1971:325, 338, 341). The system is related to the “uniform equinox” (Seidelmann, 1992:116, and Atkinson and Sadler, 1951). The basic geometry is shown in the inset of Fig 3-29. The intent was to provide an efficient, if approximate, coordinate system for use with the AFSPC analytical theories.

* TEME is actually a misnomer. The true equator correctly implies the CEP as the Z-axis, but the X-axis is not the conventional mean equinox.

TABLE 3-6. Summary of Reduction Values for Example 3-14 and Example 3-15. Values are given for each transformation. Notice the difference in values with “similar” names. Difference magnitudes from the base vectors, and from the previous coordinate system are listed in meters and mm/s. The effect of precession becomes larger as the epoch date increases from January 1, 2000. If this example were in 2014, the deltas would be significantly larger. The base values are the IAU-76/FK5 results (with corrections $\delta\Delta\Psi_{1980}$ and $\delta\Delta\epsilon_{1980}$) except for the GCRF which derives from the CIO-based option.

Frame	Method	<i>I</i>	<i>J</i>	<i>K</i>	Difference within
ITRF	IAU-76/FK5	-1033.47938300	7901.29527540	6380.35659580	m
PEF	IAU-76/FK5	-1033.47503130	7901.30558560	6380.34453270	
TIRS	IAU-2006	-1033.47503120	7901.30558560	6380.34453270	0.0001
TOD 2 w corr	IAU-76/FK5	5094.51620300	6127.36527840	6380.34453270	
TOD 2 w corr	IAU-76/FK5	5094.51478040	6127.36646120	6380.34453270	1.8501
ERS	IAU-2006	5094.51462800	6127.36658790	6380.34453270	2.0483
ERS	IAU-2000 B	5094.51462780	6127.36658800	6380.34453270	2.0485
CIRS	IAU-2006 CIO	5100.01840470	6122.78636480	6380.34453270	7158.2591
TEME	--	5094.18016210	6127.64465950	6380.34453270	437.0095
MOD 2 w corr	IAU-76/FK5	5094.02837450	6127.87081640	6380.24851640	
MOD 2 w corr	IAU-76/FK5	5094.02901670	6127.87093630	6380.24788850	0.9061
MOD	IAU-2006	5094.02896110	6127.87113500	6380.24774200	1.0224
MOD	IAU-2000 B	5094.02896120	6127.87113250	6380.24774440	1.0199
J2000 2wo corr	IAU-76/FK5	5102.50960000	6123.01152000	6378.13630000	
CIRS	IAU-2006 CIO	5100.01840470	6122.78636480	6380.34453270	3336.6211
GCRF	IAU-2006 CIO	5102.50895290	6123.01139910	6378.13693380	
GCRF	IAU-2006	5102.50895780	6123.01140380	6378.13692530	0.0109
GCRF	IAU-2000 B	5102.50895790	6123.01140130	6378.13692770	0.0082
GCRF 2 w corr	IAU-76/FK5	5102.50895790	6123.01140070	6378.13692820	0.0077
FK4	FK4	5201.58611800	6065.40181800	6353.10157310	117311.2005
ITRF	IAU-76/FK5	-3.2256365200	-2.8724514500	5.5319244460	m/s
PEF	IAU-76/FK5	-3.2256327470	-2.8724425110	5.5319312880	
TIRS	IAU-2006	-3.2256327470	-2.8724425110	5.5319312880	0.0000
TOD 2 w corr	IAU-76/FK5	-4.7460883850	0.7860783240	5.5319312880	
TOD 2 w corr	IAU-76/FK5	-4.7460885670	0.7860772220	5.5319312880	1.1169
ERS	IAU-2006	-4.7460885870	0.7860771040	5.5319312880	1.2366
ERS	IAU-2000 B	-4.7460885870	0.7860771040	5.5319312880	1.2366
CIRS	IAU-2006 CIO	-4.7453803300	0.7903414530	5.5319312880	4321.5288
TEME	--	-4.7461314870	0.7858180410	5.5319312880	263.8276
MOD 2 w corr	IAU-76/FK5	-4.7462630520	0.7860140450	5.5317905620	
MOD 2 w corr	IAU-76/FK5	-4.7462624950	0.7860141490	5.5317910250	0.7317
MOD	IAU-2006	-4.7462624800	0.7860141930	5.5317910320	0.7550
MOD	IAU-2000 B	-4.7462624800	0.7860141910	5.5317910320	0.7546
J2000 2wo corr	IAU-76/FK5	-4.7432196000	0.7905366000	5.5337561900	
CIRS	IAU-2006 CIO	-4.7453803300	0.7903414530	5.5319312880	2834.9786
GCRF	IAU-2006 CIO	-4.7432201610	0.7905364950	5.5337557240	
GCRF	IAU-2006	-4.7432201560	0.7905364970	5.5337557280	0.0067
GCRF	IAU-2000 B	-4.7432201560	0.7905364950	5.5337557290	0.0071
GCRF 2 w corr	IAU-76/FK5	-4.7432201570	0.7905364970	5.5337557270	0.0054
FK4	FK4	-4.7071391680	0.8433730560	5.5567167110	67976.0778

Technically, the direction of the uniform equinox resides along the true equator “between” the origin of the intermediate Pseudo Earth Fixed (PEF) and True of Date (TOD) frames. It is found by observing that $\theta_{GAST1982}$ may be separated into its components ($Eq_{equinox1982}$ and $\theta_{GMST1982}$). Thus,

$$\begin{aligned}\hat{\vec{r}}_{TOD} &= \text{ROT3}(-Eq_{Equinox1982*})\hat{\vec{r}}_{TEME} \\ \hat{\vec{r}}_{PEF} &= \text{ROT3}(\theta_{GMST1982})\hat{\vec{r}}_{TEME}\end{aligned}\quad (3-90)$$

Notice the use of TEME and not UOD (Fig 3-29) because we're using a modified $Eq_{Equinox1982*}$ which includes just the geometric terms. Because no official definition exists for TEME,^{*} you should convert to a truly standard coordinate frame before interfacing with other external programs. The preferred approach is to rotate to PEF using $\theta_{GMST1982}$ because there is less ambiguity and you can then convert to other standard coordinate frames. This approach requires a rotation about the Z-axis by $\theta_{GMST1982}$ (using UT1). Because polar motion has been historically neglected for General Perturbation applications, we assume that the pseudo Earth-fixed frame is the closest conventional frame.[†]

If a rotation is made to TOD using the equation of the equinoxes (first equation in Eq. (3-90)), several approximations *may* be introduced with the calculation of the nutation of the longitude ($\Delta\Psi_{1980}$) and the obliquity of the ecliptic (ϵ_{1980}). There are at least three options: the number of terms to include in the nutation series, trigonometric simplifications, and the inclusion of the post-1996 “kinematic correction” terms to the equation of the equinoxes. After choosing the length of the IAU 1980 nutation series (4, 9, 10, and 106 terms are popular choices), the transformation is sometimes further reduced by assuming that $\Delta\Psi_{1980} \approx 0$, $\epsilon \approx \epsilon_{1980}$, and $\Delta\epsilon \approx 0$. This results in a nutation matrix that is significantly simpler than the complete nutation matrix (page 218), although the complete form is more common today. The equation of the equinox may be approximated by ignoring the “kinematic correction” terms starting in 1997 [such that $Eq_{equinox1982} \approx \Delta\Psi_{1980}\text{COS}(\epsilon_{1980})$]. If you choose this approach, we assume the TEME conversion uses 4 terms for nutation, no simplification of trigonometric terms, and just the geometric terms in the equation of the equinox.

There is an additional nuance, specifically the “of date” and “of epoch” formulations. Researchers generally believe the “of date” option is correct, but confirmation from official sources is uncertain, and others infer that the “of epoch” is correct. To be complete, we provide the equations and an example problem of both. Consider the following TLE data at epoch and at 3 days into the future.

1 00005U	58002B	00179.78495062	.00000023	00000-0	28098-4	0	4753
2 00005	34.2682	348.7242	1859667	331.7664	19.3264	10.82419157413667	

* Koskela (1967:21) states that “[o]nly those terms of nutation in longitude whose coefficients are greater than 0.2”, and the corresponding terms of nutation in obliquity, are included” and that “[t]he right ascension of the mean equinox is referred to the true equator and equinox is called the ‘equation of the equinoxes’, and is equal to [$\Delta\Psi\text{COS}(\epsilon)$].” Thus, using the 4 term approximation in the later IAU-1982 nutation theory is consistent with this, as is the use of the geometric terms for $Eq_{equinox}$.

† Here we assume that JSPOC orbit determination approximates the reference frames of radar and optical sensors differently, and that numerical and analytical orbit determination methods use different techniques due to the differences in TEME, ECI, and the uncertain use of polar motion in coordinate systems.

Using the sample TLE data, we find the TEME vector at a time 3 days in the future (day = 182.784 950 62) using SGP4 from Sec. 9.7.4.

$$\begin{aligned}\hat{\vec{r}}_{TEME} &= -9060.473\ 735\ 69 \quad 4658.709\ 525\ 02 \quad 813.686\ 731\ 53 \text{ km} \\ \hat{\vec{v}}_{TEME} &= -2.232\ 832\ 783 \quad -4.110\ 453\ 490 \quad -3.157\ 345\ 433 \text{ km/s}\end{aligned}$$

Next, find the precession, nutation, and equation of the equinox parameters at this time (day = 182.784 950 62).

$$\begin{aligned}\zeta &= 0.003\ 179\ 6^\circ, \Theta = 0.002\ 763\ 3^\circ, z = 0.003\ 179\ 6^\circ \\ \Delta\Psi_{1980} &= -0.004\ 250\ 260^\circ, \Delta\epsilon_{1980} = -0.001\ 260\ 854^\circ, \epsilon_{1980} = 23.439\ 226\ 57^\circ \\ Eq_{Equinox1982*} &= -0.003\ 899\ 5^\circ\end{aligned}$$

Using

$$\hat{\vec{r}}_{J2000} = [\mathbf{P}][\mathbf{N}]ROT3(-Eq_{Equinox1982})\hat{\vec{r}}_{TEME} \quad (3-91)$$

we find the inertial (IAU-76/FK5, ‘‘J2000’’) vector at day = 182.784 950 62.

$$\begin{aligned}\hat{\vec{r}}_{J2000} &= -9059.941\ 378\ 6 \quad 4659.697\ 200\ 0 \quad 813.958\ 887\ 5 \text{ km} \\ \hat{\vec{v}}_{J2000} &= -2.233\ 348\ 094 \quad -4.110\ 136\ 162 \quad -3.157\ 394\ 074 \text{ km/s}\end{aligned}$$

For the future time using the ‘‘of epoch’’ option, the transformation parameters are found at the TLE epoch time (day = 179.784 950 62) as

$$\begin{aligned}\zeta &= 0.003\ 127\ 0^\circ, \Theta = 0.002\ 717\ 6^\circ, z = 0.003\ 127\ 0^\circ \\ \Delta\Psi_{1980} &= -0.004\ 337\ 544^\circ, \Delta\epsilon_{1980} = -0.001\ 247\ 061^\circ, \epsilon_{1980} = 23.439\ 226\ 57^\circ \\ Eq_{Equinox1982*} &= -0.003\ 979\ 6^\circ\end{aligned}$$

and the resulting vector from Eq. (3-91) differs by 23.6 m in just 3 days!

$$\begin{aligned}\hat{\vec{r}}_{J2000} &= -9059.951\ 079\ 9 \quad 4659.680\ 755\ 6 \quad 813.945\ 045\ 1 \text{ km} \\ \hat{\vec{v}}_{J2000} &= -2.233\ 336\ 111 \quad -4.110\ 141\ 024 \quad -3.157\ 396\ 220 \text{ km/s}\end{aligned}$$

Seago and Vallado (2000) discuss the TEME system and provide several additional observations. The differential corrections hold $\Delta\Psi_{1980}$ and $\Delta\epsilon_{1980}$ constant during the observation time interval to reduce the number of trigonometric operations. This produces errors in the range of about 15 m per Earth radii of the satellite position, depending on the span of time for the observations. In addition, confusing MOD or TOD with the TEME equinox can insert several hundred meters per Earth radii of the satellite positional error! Thus for modern applications, this system is not recommended due to its limitations in accuracy and ambiguity of precise definition.

FK4 Reduction

The FK4 system still exists in some older computer programs even though it’s less compatible with modern constants. It relies on the **Besselian year**, defined as the time for the mean Sun’s right ascension to increase by 24 hours (Green, 1988:249). The FK4 system has expressions for precession, nutation, and sidereal time similar to the IAU-76/FK5 system. See Seidelmann (1992:167–187) for more information on the rigorous transfor-

mation. It's sometimes necessary to transform an FK4 mean equator and equinox, to the IAU-76/FK5 J2000.0 epoch. This conversion from FK4 uses a simple matrix operation. Seidelmann (1992:313) shows the process for any position or velocity vectors, $\dot{\vec{x}}$. The literature contains some variations to this equation. The difference is about 5 m between the various approaches.

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

The reverse transformation is simply the transpose.

Be aware that converting between the FK4 and IAU-76/FK5 systems introduces two important differences—UT1 has an updated secular rate in IAU-76/FK5, and their origins don't coincide.

3.7.5 Computational Considerations

It's convenient 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 for integrating the equations of motion is the mean equinox, mean equator of J2000.0. However, some space centers use other “inertial” frames to integrate the equations of motion (the AF Space Command uses a true equator and mean equinox of date, others use TOD, etc.). These other frames are a legacy of earlier computers, when saving a few matrix operations (for precession and nutation) greatly accelerated system throughput. 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 lengthy nature of the complete equations for IAU-76/FK5 reduction has resulted in the formation of numerous approximations.

It is useful to examine the behavior of the nutation and CIO parameters because they are used with precise numerical propagation activities (Ch. 8), but can require some computational resources. Remember that the effect of using approximate forms of the reduction matrices influences only the location of the satellite used for acceleration calculations. Figure 3-32 shows historical values, including the new IAU-2000 parameters.

The Naval Observatory Vector Astrometry Subroutines (NOVAS) routines are a sophisticated method to preserve the accuracy of the full nutation (1980) 106 term theory, but within the processing time of the 4-term approximation. The technique is originally due to Coffey and Deprit (1980) and computer source code is available on the USNO's homepage under the NOVAS routines. FORTRAN and C listings are available, as well as data for the reduction techniques.

http://aa.usno.navy.mil/software/novas/novas_info.html

<http://maia.usno.navy.mil/conv2000/chapter5/NU2000A.f>

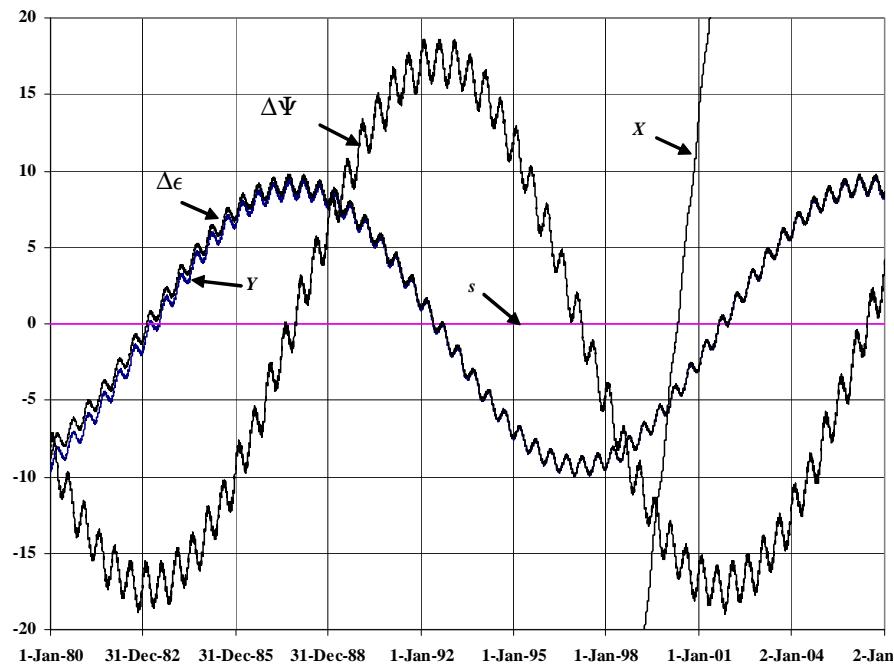


Figure 3-32. Typical Nutation Values. This figure shows the two nutation angles ($\Delta\Psi$, $\Delta\epsilon$) and the CIO coordinates (X , Y , s) over time. They contain corrections from the basic and planetary series. Notice the long and short periodic behavior.

The IAU provides similar routines with the Standards of Fundamental Astronomy (SOFA) code.

<http://www.iau-sofa.rl.ac.uk/>

Vallado et al. (2006) investigated interpolation techniques for the various reduction approaches. It appears that the best approach is to interpolate the individual CIO-based parameters (there are only 3) as this was the fastest, and it preserves the orthogonality of the resulting transformation matrices.

Some historical implementations of the IAU-76/FK5 approach use a four-term approximation discussed earlier. This is popular among users of analytical theories because it captures a portion of the nutation effect, and it requires little processing time. Examining the coefficients in the 106 term series, we can select the largest terms—be aware that list is sometimes sorted differently (See App. D). The selection of these terms is due to the fact that they are the largest terms (by at least an order of magnitude) over all other terms in the series.

3.8 Earth Models and Constants

Engineering demands that we correctly use constants and conversions within a certain model. 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 each value and reference frame.

Unfortunately, although we have various models for the Earth, some programming practices complicate attempts to modernize to newer models. For the average amateur astronomer, the differences between these data sets are often insignificant; however, for precise orbit determinations, we require consistent use of the most accurate constants. Perhaps the most common error is using conversion factors derived from different systems or sources, which can cause unexpected and erroneous results.

3.8.1 Canonical Units

The discussion of the Gaussian constant of gravitation is relevant 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. For rough hand-calculations, we can sometimes work faster by eliminating additional mathematical operations—replacing them with unity values. This book sometimes uses canonical units because they

1. Reduce the size of the numbers involved in calculations, so we can verify them more easily in our minds.
2. Can enable certain computer operations to have greater precision.
3. May speed up an algorithm by eliminating unnecessary mathematical operations because unity values replace many large constants.
4. Can reduce maintenance programming to incorporate future updates to constants. They can permit changes to constants used in conversion routines while using the same code for technical algorithms.

Computer-software code often contains unit conversions within each routine, 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 within 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. 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. 3.2) provide a way to develop canonical units for the planet. These constants allow us to normalize units. 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} :

$$ER \equiv R_{\oplus} \quad (3-92)$$

The *gravitational parameter*, μ_{\oplus} , is also set to unity, so

$$G(M_{\oplus} + m_{sat}) = \frac{ER^3}{TU^2} \equiv \mu_{\oplus} = 1 \quad (3-93)$$

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). To transform the TU, we can recognize 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 $r = R_{\oplus}$. We can also use the definitions above. Solving Eq. (3-93) for TU and using constant values, we find that

$$TU \equiv \sqrt{\frac{R_{\oplus}^3}{\mu_{\oplus}}} \quad (3-94)$$

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 an 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 back cover lists the defined and derived constants of WGS-84/EGM-96. Again, be aware of the need for consistency between values in this table. Don't mix conversions from different models such as EGM-08, JGM-3, 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 Postgraduate School ($\lambda = -121.53^\circ$)? What is UTC and TAI? What other information would you need?
2. The Astrodynamics division at Air Force Research Laboratory, located at $\phi_{gc} = 35.05^\circ$, $\lambda = -106.40^\circ$, $h_{ellp} = 1866.03$ m, was aligned with the vernal equinox on May 2, 1996. What are the LHA_γ , GHA_γ , LST , and $GMST$?
3. What are numerical values of canonical units for Jupiter? Use Table D-4.
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 GCRF position and velocity vectors for Example 3-15 in the PQW, SEZ, RSW, NTW, and EQW coordinate systems for a site at $\phi_{gd} = 40^\circ\text{N}$ and $\lambda = 100^\circ\text{W}$? Hint: Use **RV2COE** (Algorithm 9) to find the orbital elements.
6. Examine the Mean of B1950 position and velocity vectors (FK4) for the inertial Mean of J2000 vectors given in Table 3-6. 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 IAU-76/FK5 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. How many significant digits of decimal degrees are required to match the accuracy of $12^{\text{h}} 41^{\text{m}} 37.4567^{\text{s}}$? Convert your answer for units of degrees, arcminutes, and arcseconds.
10. What calendar date corresponds to 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?
11. If the position and velocity vectors we determined in Example 2-6 (ITRF) are taken as mean-of-date values at the same epoch, what are the mean-of-date vectors on August 20, 1994? (Hint: Use one transformation only.)
12. Last night while you were asleep, all the distances in the universe were reduced by 50%. How could you tell?

CHAPTER 4 OBSERVATIONS

- 4.1 Introduction
- 4.2 Obtaining Data
- 4.3 Introduction to Sensor Systems
- 4.4 Observation Transformations

4.1 Introduction

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.

In this chapter, I'll introduce how we obtain and transform observational data. A glance back at some historical technologies and mechanical developments will show that many modern conventions are products of the ingenuity and methods adopted at that time.

Observations have always played an important role in verifying the accuracy of the stellar tables. Although observational techniques and equipment of the 16th century had made only modest improvements 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 observatories were built, the machinery was state-of-the-art. The ability of the craftsmen of the fifteenth and sixteenth centuries was truly remarkable, as the accuracy of their machines attests.

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, 1989:199–200). The Babylonian *sexagesimal* system, used with the Arabic

numerals, effectively recorded and computed the results of observations. The *sexagesimal* system is a degree-arcminute-arcsecond format in which the interval of division is 60. It's the basis of several time and angle conventions we still use today.

These computing practices nicely matched the limited measuring devices of the time. 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, arcminutes, 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. Additional gradations beyond 60 were possible, but the operator couldn't easily interpret them. Most observatories now use digital readouts, but the answers remain in the sexagesimal format. You'll find that most published values of declination, latitude, and elevation angles are in the sexagesimal degrees, arcminutes, arcseconds format.*

4.2 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. 4-1. 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:

$$\lambda = \frac{c}{f}$$

Several effects 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. 8) interferes with the signals and may distort them. To a lesser extent, the troposphere also causes interference. Precise applications must account for these effects. The causes of the interferences 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, making the task difficult. Rain and fog limit some applications, as Hovanessian (1988:6–9) discusses in greater detail.

Most observations reference UTC as the standard time. However, there are practical difficulties with implementing UTC at an observation site. One approach uses on-site atomic clocks. Special receivers can also electronically receive broadcast UTC from WWV and WWVH. Even the Internet has several options to obtain precise timing updates. The rapid growth in the use of GPS signals offers yet another technique. Once a

* The hour-minute-second format arises similarly. As clocks became more popular, their gears could be machined to offer 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.

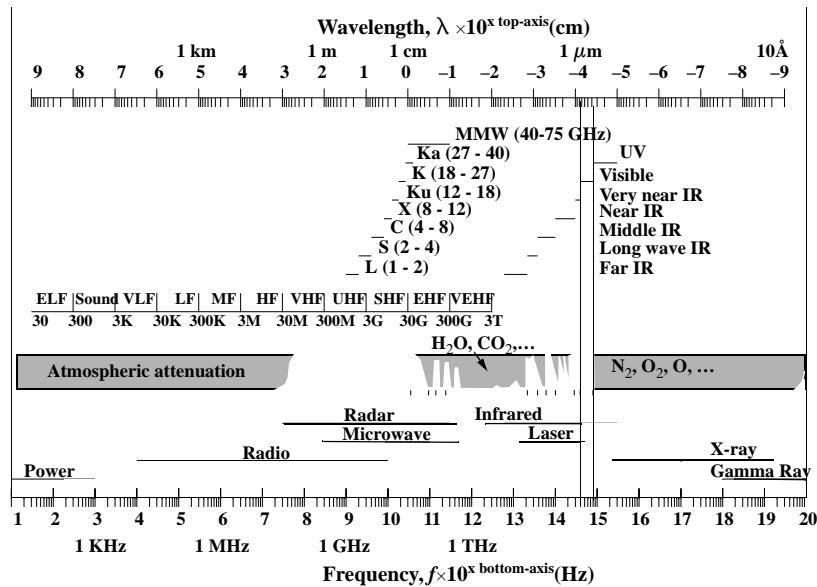


Figure 4-1. Electromagnetic Spectrum. Atmospheric attenuation partially blocks much of the incoming electromagnetic radiation, as the dark bar illustrates. Frequency values are given for various bands next to each letter descriptor. Values are in GHz, THz, and so on, as appropriate. The MMW is sometimes broken into O (40-50 GHz) and V (50-75 GHz) bands. Satellite communication occurs in the L to MMW range. [Hovanessian (1988:5-7), Chetty (1988:311), Gammon (1976:17), Morgen & Gordon (1989:233)]

site obtains UTC, disseminating it to all the local sensors can be challenging. Thus, ensuring that the time for each observation exactly matches actual UTC time can be a difficult 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. A common form of two-way signals occurs with **transponder** systems. Here, the signal is received and retransmitted back to the ground. This can give very precise range information, and is used by many satellite operators. There is a delay associated with each transponder system. **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. 4-2.

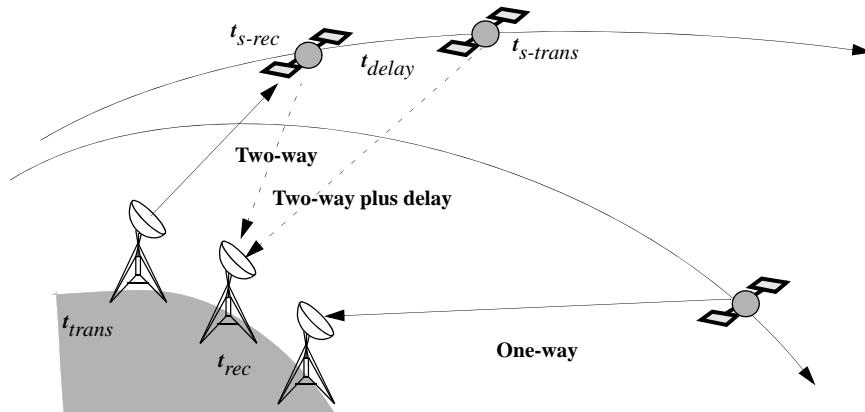


Figure 4-2. Geometry for One-way and Two-way Transmission. When a signal transmits to or from a satellite, we determine precise values of range from the speed of light and the total time (including any delays) for the trip. We must calculate physical and atmospheric distortions during each trip through 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 4-2 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. 8) 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.

The range, ρ , is found 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, $\Delta t = t_{rec} - t_{trans}$. For two-way systems that retransmit the signal, this expression becomes more complex. If we use t_{s-rec} and $t_{s-trans}$ to indicate the time that the satellite receives and transmits the signal,

$$\rho = |r_{sat}(t_{s-rec}) - r_{site}(t_{trans})| + |r_{site}(t_{rec}) - r_{sat}(t_{s-trans})|$$

For the greatest accuracy, we 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 requires 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.

Observational data may need to first be reduced from the observed place to the ITRS, or from the *Geocentric Reference System* (GCRS) to the barycentric ICRS. Figure 4-3 illustrates some of these possible operations. These operations are an extension of the applications in Sec. 3.7. ICRS to GCRS transformations are used for the reduction of stellar observations or interplanetary spacecraft measurements.

* The 2 indicates the round trip and the negative sign is chosen to give a positive doppler shift when the range rate is negative, indicating the object is coming towards the receiver.

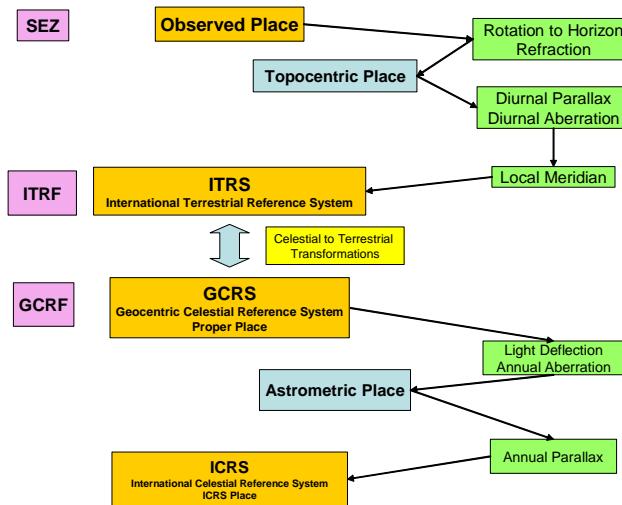


Figure 4-3. Observed Place to ITRS to GCRS to ICRS Transformations. Operations for transforming vectors between the observed place and ICRS reference systems, most important for stellar applications. Additional processing for observations (observed place to ITRS) is also shown for orbit determination and propagation operations.

4.2.1 Quality of Data

Several things affect the quality of observational data. A primary effect is any delay in the signal which increases the uncertainty of the range value. Optical data is strongly affected by atmospheric refraction. See Montenbruck (2006: Chap. 6) for additional information.

Refraction

Any fluid medium can slow down the propagation of electromagnetic energy (such as light rays), potentially diverting it from straight line motion. This effect is known as refraction, and it's the reason that underwater objects can look distorted when viewed from above. The same effect also distorts the apparent line-of-sight of a satellite observation when viewed through the Earth's atmosphere. While this effect is often appreciable, we can correct it using mathematical models. Such corrections are based on numerical models of the neutral atmosphere and the ionosphere—See “Solar Flux Data” on page 560. Both the Ionosphere and the Troposphere contribute to this phenomena.

Ionosphere

The ionosphere extends from about 80 to 300 km above the Earth's surface and contains free electrons that have been ionized from solar UV radiation. The electrons (generally termed the Total Electron Count (TEC)) tend to delay the propagation of signals as they pass through the region. Thus the ionosphere primarily affects range measurements. An ionosphere model can be applied when a satellite is tracked at a frequency that is signif-

icantly affected by the ionosphere (e.g., GPS signals). The magnitude of the ionosphere range delay is inversely proportional to the square of the frequency, thus, a higher frequency measurement yields a smaller ionosphere delay. The effect can be about 10 m if not corrected so it must be accounted for when processing observations. The current model is the International Reference Ionosphere (IRI 2007). The IRI models are self-contained; however, they require current IRI model data.* When processing GPS data, Single Differenced signals can eliminate the effects of the Ionosphere to first order.

Troposphere

The troposphere extends to about 9 km and it also affects signals and tends to delay them. There are two primary effects—the hydrostatic [dry] delay (75–95%), and the wet delay (5–25%). The dry delay depends on the orientation of the poles of the atmospheric molecules. The wet delay depends on the amount of water vapor in the air. A simple solution uses a 15 deg elevation mask to eliminate most of the problem because the troposphere is relatively thin. A better approach is to use a mathematical tropospheric model. Meteorological data consisting of the pressure, temperature, and humidity at the facility, are needed to support computations with the Marini-Murray model or Saastamoinen model (See Seidelmann (1992:24, 140–145). These data may be specified as constant values, in tables via text files, or there may be extracted from the input tracking data. Other models require only the nominal surface refractivity at each facility location, which can be modeled as a polynomial function from 0th to 10th-order spanning one year, but the results are less accurate. With sufficient tracking, a time-varying correction can be estimated to the wet component of a facility’s delay.

4.2.2 Quantity of Data

A direct 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. 10 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.

* The data is found at <http://iri.gsfc.nasa.gov/>

4.2.3 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 4-1 shows representative observation types and corresponding solution methods.

TABLE 4-1. 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). Radars provide most of this data ($\rho, \beta, el, \dot{\rho}$). Lasers yield range data, and optical devices produce α_t and δ_t . Sec. 7.4.1 discusses techniques which can process ρ and $\dot{\rho}$ data.

Data Type	Restrictions	Sec./Chap.	Solution Method
$\dot{\rho}$	None	10	Estimation
β, el	3 sets minimum	7.3	LAPLACE, GAUSS, DOUBLE-R
ρ, β, el	2 sets minimum	7.2, 7.5, 7.6	SITE-TRACK , then LAMBERT (2) or GIBBS/HGIBBS (3)
$\rho, \beta, el, \dot{\rho}$	2 or 3 sets minimum	7.2	SITE-TRACK
$\rho, \beta, el, \dot{\rho}, \dot{\beta}, \dot{el}$	None	7.2	SITE-TRACK
α_t, δ_t	3 sets minimum	7.3	LAPLACE, GAUSS, DOUBLE-R
ρ	6 simultaneous, None	7.4, 10	Trilateration, Estimation

Moritz and Mueller (1987:350-353) give a brief introduction to the **International Laser Ranging System** (ILRS) and **Satellite Laser Ranging** (SLR) data. SLR data is a product of the ILRS. The ILRS also provides laser ranging for the Moon. Analysis supports geodetic research and is an input into the formation of the ITRF. Although the primary goal is geophysical research, orbit determination has begun to embrace SLR data as a means to calibrate space surveillance systems. The Lunar measurements began in 1969 with the Lunar reflectors left by the Apollo astronauts. In 1976, progress continued with the launch of LAGEOS, and later with GEOS-A. Today, over 50 stations contribute to the collection of SLR data on a variety of satellites. Although most laser reflecting satellites are in higher altitude orbits like LAGEOS (6000 km altitude), some like GFZ, have been flown within the drag regime (400 km altitude).

4.2.4 Example Applications

I'll introduce a few satellite systems to illustrate the principles of obtaining observational data: TOPEX/Poseidon, TDRS, and GPS.

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. 10.9) 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.

The *Global Positioning System* (GPS) transmits signals that can be processed into range and timing information. I'll introduce additional details in Chap. 11, but you should understand that the processing of GPS signals is becoming a popular technique for autonomous orbit determination. Primarily, satellites in LEO (below semi-synchronous altitude) can detect these signals, and determine their position to within a few meters. Techniques also exist to determine satellite attitude. This is just another application for the GPS signals.

4.3 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 orbit* (including space-based sensors observing other satellites). Chapter 11 discusses the opposite problem of observing the Earth (space-based *surveillance* and *reconnaissance*). I'll define *space surveillance* as the general practice of observing objects in orbit.

Sensors measure various quantities on satellites. A sensor "site" can be either ground-based or space-based, and they often have one or more types of sensors at each location.

If you obtain data from the US Air Force Space Command (AFSPC), you may encounter specific categories for sensors. Muolo's (1993:97-103) lists three distinct cat-

egories: dedicated, collateral, and contributing. **Dedicated sensors** perform a primary mission, such as space surveillance, or missile warning. **Collateral sensors** have other primary missions, but they still support space surveillance. For example, the U.S. Space Command's ballistic-missile-warning sensors routinely track low-Earth satellites. Lastly, **contributing sensors** include university, commercial, and other sensors. It's possible for dedicated sensors from one organization to function as collateral or contributing sensors for another organization.

A **tracking network** of globally distributed sensors to acquire data on satellites enables development of a catalog of all objects in space. Global distribution of sites permits better observability of the satellites' dynamic motion. The network is necessary because our assumptions of the equations of motion are only approximate, so we must monitor the effects of the dynamics and continually update the model based on current observations.

There are several networks that can provide satellite observations (See Vallado and Griesbach 2011 for extensive location listings). One network is the **US Space Surveillance Network** (SSN), which developed from ballistic-missile warning assets. The SSN includes radar, optical, and interferometer systems. In 2013, JSPOC was tracking almost 17,000 objects. This number will grow as they locate and catalog smaller particles of debris. Safety and avoiding satellite collisions are the main reasons to increase the number of objects being tracked.

The former Soviet Union has a large network, the **Russian Space Surveillance System** (RSSS). It's designed to operate using geographically limited observations. Khutorovsky (1996) describes the operation of its sensors. The RSSS has evolved over the years similar to the AFSPC SSN network. Beginning as a large network of phased array radars and radars forming a missile warning system, some of the sensors were gradually transitioned to perform space surveillance roles. Coverage of geosynchronous region seems to have been limited in the past. Today, the Russian Space Agency (Roscosmos) is responsible for the safety of Russian space activities. The Ministry of Defense is responsible for Russian military operations in space and space surveillance in support of these operations. Other organizations (Russian Academy of Sciences, Foreign Ministry etc.) are taking part in the development of SSA policy, data collection and sharing, and establishing international cooperation.

The **European Space Agency** (ESA) also maintains a large collection of sensors including optical (ESA 1 meter telescope at Tenerife), the French GRAVES interferometer, a backscatter EISCAT system, and radar systems (FGAN in Germany). ESA is planning the development of a modern sensor network to accurately provide for their missions in space. As a new system, designed for accuracy and numerical operations, this system has tremendous potential and opportunity.

The French space agency, Centre National d'Études Spatiales (CNES) has developed the DORIS system to track the TOPEX/Poseidon, SPOT-4, 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 beacons and later transmits them back to the ground. Master receiver stations are in Toulouse, France, and Kourou, French Guiana.

There are also smaller governmental systems. 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 (Coster et al. 2001). Other countries are developing systems such as Canada, China, India, Japan, Korea, Ukraine, and Uzbekistan.

In about 2001, cooperation of optical observatories began, and by 2005 the **International Scientific Optical Network** (ISON) (Molotov and Agapov, 2008) was created. Additional observatories have been added to primarily study scientific and applied problems in space, notably in geosynchronous orbits. A great deal of modernization of equipment and software has taken place, and the Keldysh Institute of Applied Mathematics Russian Academy of Sciences (KIAM RAS) has been a principal scientific and organizational coordinator of ISON. By 2013, more than 50 telescopes at 33 observatories in 11 countries were operating around the world with over 90 researchers. Their tasks include regular GEO monitoring, new object discovery and tracking, and maintenance of as complete a catalog as possible. ISON was tracking 1557 objects in GEO compared to the JSPOC TLE catalog of 1016 objects, and the observers track between 150 up to 800 individual objects each night. The observations are stored at the KIAM Ballistic Center upon collection. The processing and analysis of information on space debris is also performed at the Center.

The **Deep Space Network** (DSN), managed by the Jet Propulsion Laboratory, communicates with objects *far away* from the Earth. NASA is the primary user of this network for interplanetary probes. The term *far* is obviously relative. Recall the definition of deep space from Chap. 1, 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 (35.39°N , 116.848°W , 1036m), the Spanish complex in Madrid, Spain (40.41°N , 3.685°W , 670m), and the Australian complex near Canberra (35.52°S , 149.008°E , 1036m).

Satellite owner operators use a vast international network of sensors to communicate with their satellites. During communication transmissions, transponders provide very accurate range, and sometimes angular information for orbit determination. There are literally hundreds of sites around the world. Like the AFSCN though, they are specifically for communication with the satellite, and not used for passive satellite detection.

Amateurs observers maintain reasonably current orbital elements on about 300 classified objects, including nearly all in LEO, and a large fraction of those in GEO and other orbits. Equipment ranges from binoculars and stopwatches, to telescopic video cameras with GPS time inserters. Automated and semi-automated data reduction is commonly used. In the year ending in June 2011, twenty-six observers contributed 22,700 positional observations. Observers are located in the United States, United Kingdom, Canada, France, Italy, Netherlands, and South Africa.

NASA is a contributor to an international consortium of almost 50 dedicated and contributing stations known as the SLR Network. Each of the **Satellite Laser Ranging** (SLR) sites provide very accurate satellite ranges ($\sim 1 \text{ cm}$). These sites track satellites

which have corner-cube retro-reflectors that reflect the laser light back in the same direction from which it was received. The primary mission of the SLR is ultra-precise geodetic research. These highly accurate measurements also permit calibration of radar and other sensors. The satellites used by the SLR network are often called calibration satellites, like the Laser Geodynamics Satellites (Lageos).

Table 4-2 lists some of the categories of sensors and the sites that handle them. NAVSPASUR interferometer consists of nine stations (3 transmitters and 6 receivers). The three transmitters broadcast a planar, continuous-wave vertical fan of radar energy. When objects pass through the fan, the energy is reflected back towards the Earth. Up to six receivers can detect and sample the signal. By differencing the phase, you obtain direction cosines (with respect to the fence plane) (see Schumacher et al. (1998) for additional information). A satellite passing through the fence triggers detection. The system's output is direction cosines that can be transformed to azimuth and elevation during processing. Additional processing, called triangulation, will sometimes yield a crude estimate of range, but it's usually two to three times less accurate than the angular data.

Optical systems are an important part of many networks. Large photographic cameras, called Baker-Nunn cameras, were first used during the 1960's for space surveillance by the US military. They were equatorially mounted and provided accurate photographs for many years. Comparing the resulting time-tagged photographs to star catalogs yielded very accurate angular measurements. Ground-based Electro-Optical Deep Space Surveillance (GEODSS) sites have replaced the Baker-Nunn cameras, and they give about 0.5"-1.0" accuracy. 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. Another modern improvement is the RAVEN telescope developed at the Air Force Research Laboratory. It uses technology from Maui Space Surveillance System (MSSS) and has demonstrated arc-second accuracy (Sabol, et al. 1999).

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 elevation due to the high mountain range to the west. Sensors' range limitations must be considered when evaluating observability and viewing because satellites can easily be too far away from the site. Many of the tracking resources serve specific missions (launch corridors, missile warning, and so on) and are greatly constrained; thus, all the available 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 4-3 lists representative limits for some of the sensor sites we've identified.

Tracking data is imperfect, often containing biases and noise. *Calibration* is a process that identifies the biases and deduces the noise statistics. Knowing the biases helps you correct the raw sensor data. Noise statistics help us to accurately weigh data so we can improve the accuracy of the solution. Table 4-4 lists representative noise and biases for some sensors. Later, we'll discuss how we use this data.

TABLE 4-2. Representative Space Sensors Grouped by Usage. Data is combined from Taff (1985:132–135), Muolo (1993:97–103), Vallado (1989, App. A), Vallado and Griesbach (2011), and the *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), Maui Space Surveillance Station (MSSS), and Moron Optical Space Surveillance System (MOSS).

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	398, 399
NAVSPASUR, 9 sites	Interferometer	~33.0	-87 to -117	10 to 1400	741–749
MSSS, HI	1.6 m Optical	20.71	-156.26	3058.2	952
MOTIF, HI	1.2 m Optical	20.71	-156.26	3059.0	951
HAX	Mech Tracker	42.62	-71.49	101.1	373
Spain (MOSS)	Optical	37.10	-5.37	287.0	260
SSN Collateral Sensors					
Fylingdales, UK	Phased Array	54.37	-0.67	338.9	344–346
Ascension, Atlantic	Radar	-7.91	-14.40	56.1	354, 355
Clear, AK	Radar	64.29	-149.19	213.3	349, 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 (Pave)	Phased Array	39.14	-121.35	115.7	388, 389
Shemya, AK (Cobra)	Phased Array	52.74	-174.09	89.8	392, 393
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, ALTAIR, TRA-DEX	9.39	167.48	62.7	333–335
Millstone, MA	Radar	42.62	-71.49	123.1	369, 370
Haystack, MA	Radar	42.62	-71.49	115.7	372

*The Midcourse Space Experiment (MSX) Space Based Visible (SBV) sensor produced space-based optical observations, as do the newer SBSS satellites (NORAD 34903 and 37168). See Table 9-4.

TABLE 4-3. Physical Limitations of Representative Sensors. This table includes representative range, azimuth, and elevation values. Atmospheric distortions for optical systems at elevations lower than 20° limit the accuracy of the observations. Some azimuth ranges go through 360°. (Vallado, 1989 and Spence, 2000)

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	
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	

*Observational data is right ascension and declination. Sensor acquisition of the satellite depends on optical cross-section size.

†Observational data is direction cosines in two directions and range depends on radar cross-section size.

Given the limitations of the sensors and their effective ranges, sensors must be spread throughout the world to cover a particular satellite. The primary mission usually drives the locations for sensors. 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 4-4 shows some of the sensors used to track satellites.

4.4 Observation Transformations

Observations are the fundamental data for orbit determination. We generally make observations from a site fixed to the Earth's surface (terrestrial reference frame), although satellite-based observations are becoming more routine. There are two tech-

TABLE 4-4. 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 and Spence, 2000)

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	100.0	0.0220	0.0100	50	0.0220	0.0200
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
Shemya, AK	29.0	0.0010	0.0030	2.9	0.0540	0.0530
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.

niques. Radar sensors use electromagnetic radiation, and the observation information (often topocentric azimuth and elevation) is deduced from the raw signals. Optical measurements generally refer to a star catalog (celestial reference frame) and “photograph” the satellite over a short period of time. The satellite “streak” on the photographic plate (or Charge Coupled Device, CCD) allows us to find the observational data (usually topocentric right ascension and declination) through end-point detection and further processing.

The differences between topocentric and geocentric angles are often negligible for stars but can be quite large for satellites orbiting the Earth. There are several equations that are valid in the topocentric and geocentric frames, but we must also understand any translations needed to convert between each frame. Finally, there are corrections that we apply to observations (Sec. 4.4.4). Thus, when we use the data in an orbit determination program, we must convert the answers to a common coordinate system before processing.

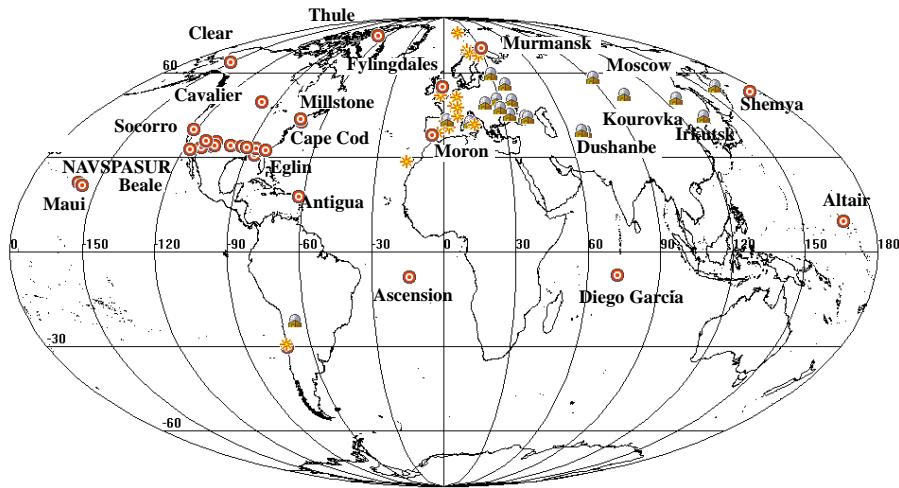


Figure 4-4. Representative Worldwide Sensors. Sensors are around the world, but coverage isn't particularly good in the Southern Hemisphere and over the oceans. The US SSN (circles), ESSS (stars), and ISON (icons) systems are shown. (Image courtesy of AGI/STK)

Notice that to use observations correctly requires us to combine many topics including location parameters (Sec. 3.2.1), various coordinate systems and angular measurements (Sec. 3.3), the motion of the coordinate system (Sec. 3.7), and corrections that we apply to observations (Sec. 4.4.4). Ultimately, we must correctly process the following coordinate system information: the site location, the observations, any reference data (i.e., star catalogs), and the necessary systems for further processing. This topic can be complex because it often involves both hardware and software. Consult Seidelmann (1992:145-164) and Roy (1988:67) for additional information. I'll introduce the stages that Seidelmann outlines in the practical conversion section (Sec. 4.4.4). Because the previously mentioned topics are described elsewhere in the text, I'll address just two operations: converting between the angles and the position and velocity vectors, and converting between angles of different systems. Although I show rate information for right ascension and declination, it's seldom used for practical applications. However, azimuth and elevation rates are usually available from radar data, so I also show how to convert this information.

4.4.1 Geocentric Right Ascension and Declination

Our problem is to determine right ascension and declination when given the position in geocentric coordinates and vice versa. This method is valid for any single coordinate system (e.g., GCRF), but may require the corrections to the observations in Sec. 4.4.4. Figure 4-5 shows the geometry.

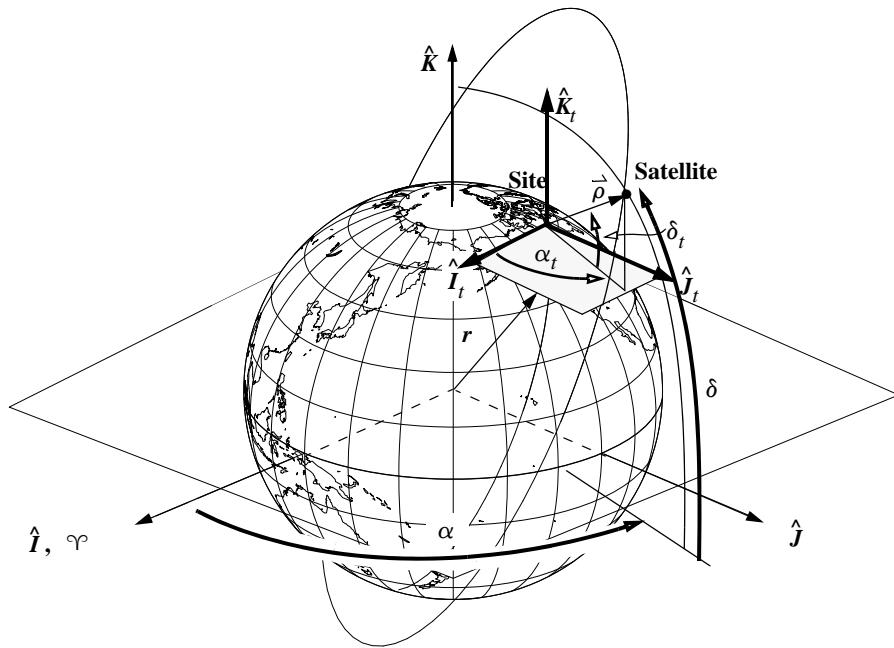


Figure 4-5. Right Ascension and Declination. Geocentric right ascension and declination use the Earth's equatorial plane. Topocentric right ascension and declination values use a plane parallel to the Earth's equator but located at a particular site. Transformations rely on determining the slant-range vector (ρ) in the IJK system, and the radius vector to the satellite, r .

This is the easiest of the transformations because we can find the position and velocity vectors directly without any translations. If we write the position vector in terms of right ascension and declination directly from Fig. 4-5 with $r = |\hat{r}_{IJK}|$,

$$\hat{r}_{IJK} = \begin{bmatrix} r \cos(\delta) \cos(\alpha) \\ r \cos(\delta) \sin(\alpha) \\ r \sin(\delta) \end{bmatrix} \quad (4-1)$$

We find the velocity vector by direct differentiation.*

* Note that, when we work with stars, the angular rates usually come from proper motion. (See “Proper Motion and Radial Velocity” on page 314.)

$$\dot{\vec{v}}_{IJK} = \begin{bmatrix} \dot{r} \cos(\delta) \cos(\alpha) - r \sin(\delta) \cos(\alpha) \dot{\delta} - r \cos(\delta) \sin(\alpha) \dot{\alpha} \\ \dot{r} \cos(\delta) \sin(\alpha) - r \sin(\delta) \sin(\alpha) \dot{\delta} + r \cos(\delta) \cos(\alpha) \dot{\alpha} \\ \dot{r} \sin(\delta) + r \cos(\delta) \dot{\delta} \end{bmatrix} \quad (4-2)$$

This transformation is the reverse of determining the angles from the position and velocity vectors. From Fig. 4-5, we can find the formulas for right ascension and declination. Singularities 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 in most cases. Algorithm 25 shows the result. Include the velocity vector if possible—even if you need the velocity only to resolve ambiguities at zenith.

Although finding the rates is less common, a few simple equations make the transformation complete. Find the range rate using results from Sec. 2.2.3. Recall that $\dot{\vec{r}} \cdot \dot{\vec{v}} = \dot{r}\dot{r}$, where \dot{r} was the radial velocity. Solving this relation gives the result:

$$\dot{r} = \frac{\dot{\vec{r}} \cdot \dot{\vec{v}}}{r}$$

Find the declination rate by solving the K -component of the velocity vector [Eq. (4-2)].

$$\dot{\delta} = \frac{v_K - \dot{r} \sin(\delta)}{r \cos(\delta)}$$

Consider the negative K -component of $\dot{\vec{r}} \times \dot{\vec{v}}$ to find the right ascension rate,

$$v_I r_J - v_J r_I = \left\{ \dot{r} \cos(\delta) \cos(\alpha) - r \sin(\delta) \cos(\alpha) \dot{\delta} - r \cos(\delta) \sin(\alpha) \dot{\alpha} \right\} r_J - \left\{ \dot{r} \cos(\delta) \sin(\alpha) - r \sin(\delta) \sin(\alpha) \dot{\delta} + r \cos(\delta) \cos(\alpha) \dot{\alpha} \right\} r_I$$

By expanding and substituting r_I and r_J from Eq. (4-1), 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 25: Geocentric Radec ($\hat{r}_{IJK}, \hat{v}_{IJK} \Rightarrow r, \alpha, \delta, \dot{r}, \dot{\alpha}, \dot{\delta}$)

$$\begin{aligned}
 r &= |\hat{r}_{IJK}| \\
 \sin(\delta) &= \frac{r_K}{r} \quad \cos(\delta) = \frac{\sqrt{r_I^2 + r_J^2}}{r} \\
 &\left[\begin{array}{ll} \text{IF } \sqrt{r_I^2 + r_J^2} \neq 0 \\ \sin(\alpha) = \frac{r_J}{\sqrt{r_I^2 + r_J^2}} \quad \cos(\alpha) = \frac{r_I}{\sqrt{r_I^2 + r_J^2}} \\ \text{ELSE} \\ \sin(\alpha) = \frac{v_J}{\sqrt{v_I^2 + v_J^2}} \quad \cos(\alpha) = \frac{v_I}{\sqrt{v_I^2 + v_J^2}} \end{array} \right. \\
 \dot{r} &= \frac{\hat{r}_{IJK} \cdot \hat{v}_{IJK}}{r} \quad \dot{\alpha} = \frac{v_I r_J - v_J r_I}{-r_J^2 - r_I^2} \quad \dot{\delta} = \frac{v_K - \dot{r} \frac{r_K}{r}}{\sqrt{r_I^2 + r_J^2}}
 \end{aligned}$$

Notice we need only the sine expression for declinations 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. Finally, the angular rate terms may be indeterminate for rare cases when there is no acceleration available.

4.4.2 Topocentric Right Ascension and Declination

Operations with the topocentric right ascension and declination present an additional problem in that the origin of the coordinate system is displaced from the Earth's center. 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. (4-1) and Eq. (4-2) to form the vectors, but remember we're in the topocentric coordinate system, so we also need a translation to get geocentric coordinates.

$$\begin{aligned}
 \hat{r}_{ECI} &= \hat{\rho}_{ECI} + \hat{r}_{SiteECI} \\
 \hat{v}_{ECI} &= \dot{\hat{\rho}}_{ECI} + \hat{v}_{SiteECI}
 \end{aligned} \tag{4-3}$$

Notice that this calculation may be done in either ECI or ECEF, but because the coordinates refer to ECI, this is preferred. Also, because the coordinate system maintains a

fixed orientation with respect to the vernal equinox but the site is moving, there is additional processing for the velocity components.

The reverse process begins by determining the slant-range vector, $\hat{\rho}_{ECI}$, from the site to the satellite. 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 25 but we use the slant-range vectors.

ALGORITHM 26: Topocentric

$$\begin{aligned}
 & (\dot{\hat{r}}_{ECI}, \dot{\hat{v}}_{ECI}, \dot{\hat{r}}_{SiteECI}, \dot{\hat{v}}_{SiteECI} \Rightarrow \rho, \alpha_t, \delta_t, \dot{\rho}, \dot{\alpha}_t, \dot{\delta}_t) \\
 & \quad \hat{\rho}_{ECI} = \hat{r}_{ECI} - \hat{r}_{SiteECI} \\
 & \quad \rho = |\hat{\rho}_{ECI}| \\
 & \quad \text{SIN}(\delta_t) = \frac{\rho_K}{\rho} \\
 & \quad \dot{\rho}_{ECI} = \dot{\hat{v}}_{ECI} - \dot{\hat{v}}_{SiteECI} \\
 & \quad \text{IF } \sqrt{\rho_I^2 + \rho_J^2} \neq 0 \\
 & \quad \left[\begin{array}{ll} \text{SIN}(\alpha_t) = \frac{\rho_J}{\sqrt{\rho_I^2 + \rho_J^2}} & \text{cos}(\alpha_t) = \frac{\rho_I}{\sqrt{\rho_I^2 + \rho_J^2}} \\ \text{ELSE} & \\ \text{SIN}(\alpha_t) = \frac{\dot{\rho}_J}{\sqrt{\dot{\rho}_I^2 + \dot{\rho}_J^2}} & \text{cos}(\alpha_t) = \frac{\dot{\rho}_I}{\sqrt{\dot{\rho}_I^2 + \dot{\rho}_J^2}} \end{array} \right] \\
 & \quad \dot{\rho} = \frac{\dot{\hat{\rho}}_{ECI} \cdot \dot{\hat{\rho}}_{ECI}}{\rho} \quad \dot{\alpha}_t = \frac{\dot{\rho}_I \rho_J - \dot{\rho}_J \rho_I}{-\rho_J^2 - \rho_I^2} \quad \dot{\delta}_t = \frac{\dot{\rho}_K - \rho \text{SIN}(\delta_t)}{\sqrt{\rho_I^2 + \rho_J^2}}
 \end{aligned}$$

A variation of this approach converts topocentric values to geocentric right ascension and declination values without the intermediate vectors from Algorithm 26. Carefully examining the problem reveals that we must know the slant range to accurately transform the angles. 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. Avoid 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 ($\rho, \alpha_t, \delta_t, r_{site}$) (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:

$$\begin{aligned}\tan(\gamma) &= \frac{\tan(\phi_{gc}) \cos\left(\frac{\alpha_t - \alpha}{2}\right)}{\cos\left(\frac{\alpha_t + \alpha}{2} - \theta_{LST}\right)} \\ \tan(\delta_t - \delta) &= \frac{\frac{r_{site}}{\rho} \sin(\phi_{gc}) \sin(\delta_t - \gamma)}{\sin(\gamma) + \frac{r_{site}}{\rho} \sin(\phi_{gc}) \cos(\delta_t - \gamma)} \\ r &= \frac{\rho}{\cos(\delta_t - \delta)} \left[1 + \frac{r_{site} \sin(\phi_{gc}) \cos(\delta_t - \gamma)}{\rho \sin(\gamma)} \right]\end{aligned}$$

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. The solution ultimately rests on choosing adequate initial guesses and successively evaluating the equations. In this case, we have four equations (including γ) and four unknowns (ρ, r, α , and δ). Many algorithms (here and in subsequent chapters) will require iteration.

4.4.3 Azimuth-Elevation

Conversions with azimuth and elevation are similar to those for geocentric right ascension and declination except for 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. (4-1), begin with the position vector (SEZ) (See Fig. 7-2 also):

$$\overset{\curvearrowleft}{\rho}_{SEZ} = \begin{bmatrix} -\rho \cos(el) \cos(\beta) \\ \rho \cos(el) \sin(\beta) \\ \rho \sin(el) \end{bmatrix} \quad (4-4)$$

Notice the negative sign in the S 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!*

As before, find the velocity as the derivative of Eq. (4-4):

$$\dot{\hat{\rho}}_{SEZ} = \begin{bmatrix} -\dot{\rho} \cos(el) \cos(\beta) + \rho \sin(el) \cos(\beta) \dot{el} + \rho \cos(el) \sin(\beta) \dot{\beta} \\ \dot{\rho} \cos(el) \sin(\beta) - \rho \sin(el) \sin(\beta) \dot{el} + \rho \cos(el) \cos(\beta) \dot{\beta} \\ \dot{\rho} \sin(el) + \rho \cos(el) \dot{el} \end{bmatrix} \quad (4-5)$$

Application: RAZEL

In this section, I'm focusing on transformations of position and velocity vectors into range, azimuth, elevation, and their rates because they're quite common and occur in several places throughout this book. This process applies mainly to orbit determination and estimation (Chap. 10), in which we must determine the predicted value of the measurements (range, azimuth, elevation, and the associated rates) as the state vector (position and velocity vectors) changes. Later we'll see the overall process is simply the reverse of the *SITE-TRACK* (Algorithm 51) equations. Begin by converting the position and velocity vectors to ECEF.^{*} Next, find the slant-range vector from the site to the satellite in ECEF:

$$\hat{\rho}_{ECEF} = \hat{r}_{ECEF} - \hat{r}_{SiteECEF} \quad (4-6)$$

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 following rotations. The velocity vector is simply equal to the slant-range velocity [See also Eq. (7-3)].

$$\dot{\hat{\rho}}_{ECEF} = \hat{v}_{ECEF} \quad (4-7)$$

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 4-6 shows this transformation along with the azimuth and elevation.

This conversion uses the relations in Eq. (3-28):

$$\begin{aligned} \hat{\rho}_{SEZ} &= [\text{ROT2}(90^\circ - \phi_{gd})][\text{ROT3}(\lambda)]\hat{\rho}_{ECEF} \\ \dot{\hat{\rho}}_{SEZ} &= [\text{ROT2}(90^\circ - \phi_{gd})][\text{ROT3}(\lambda)]\dot{\hat{\rho}}_{ECEF} \end{aligned}$$

* We determine the angles from the vectors in the SEZ coordinate system, so we may need to first rotate the vectors from another system, such as the geocentric equatorial system, IJK.

* For approximate calculations in which the distinction between ECI and ECEF is unimportant, the reduction is avoided and the equations are changed to use θ_{LST} instead of λ . Note that when using this approach, the correction for rotating coordinates no longer occurs in the sidereal time calculation and Eq. (4-7) is changed to

$$\dot{\hat{\rho}}_{IJK} = \hat{v}_{IJK} - \hat{\omega}_\oplus \times \hat{r}_{IJK}$$

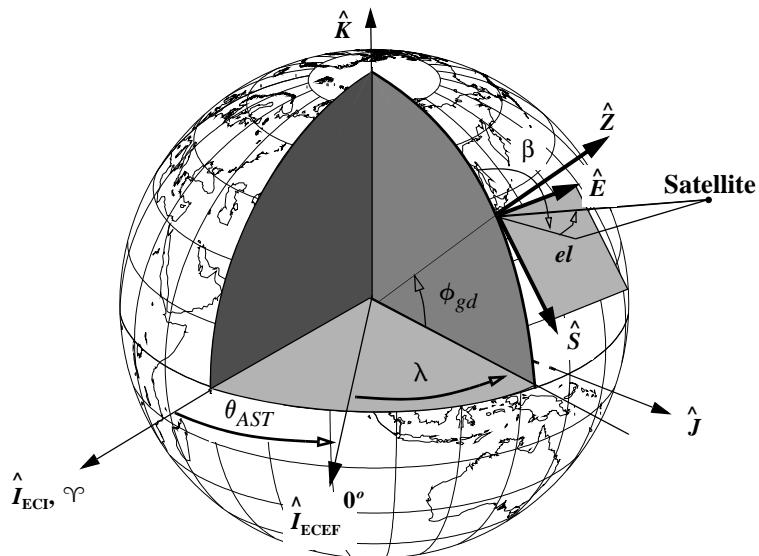


Figure 4-6. Topocentric-Horizon Coordinate System (SEZ). We measure azimuth, β , and elevation, el , in the topocentric-horizon coordinate system. Most analyses consider the differences between ECI and ECEF coordinate systems.

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. 4-6. The sine function permits unique determination because elevation is limited to -90° to 90° . Both expressions are given for completeness.

$$\sin(el) = \frac{\rho_Z}{\rho} \quad \cos(el) = \frac{\sqrt{\rho_S^2 + \rho_E^2}}{\rho}$$

Due to the direction of positive measurement (positive clockwise from north), the relation for azimuth is somewhat different from the formulas for right ascension (Algorithm 25).

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°

$$\sin(\beta) = \frac{\dot{\rho}_E}{\sqrt{\dot{\rho}_S^2 + \dot{\rho}_E^2}} \quad \cos(\beta) = \frac{-\dot{\rho}_S}{\sqrt{\dot{\rho}_S^2 + \dot{\rho}_E^2}}$$

Notice several things. First, we need both the sine and cosine expressions to correctly determine the quadrant. The case for 90° elevation permits continuous coverage through all possible slant range vectors. 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. (4-5). The range rate is the easiest, but it requires knowledge of the discussion of rates in Sec. 1.3.3.

$$\dot{\rho} = \frac{\dot{\rho}_S \cdot \dot{\rho}}{\rho} \quad (4-8)$$

We find the elevation rate by examining the *Z* component in Eq. (4-5). 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

$$\dot{el} = \frac{\dot{\rho}_Z - \dot{\rho} \sin(el)}{\sqrt{\rho_S^2 + \rho_E^2}} \quad (4-9)$$

The rate of change in azimuth is a more interesting task. Careful examination of Eq. (4-5) 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. (4-5),

$$\begin{aligned} \dot{\rho}_S \rho_E - \dot{\rho}_E \rho_S &= \{-\dot{\rho} \cos(el) \cos(\beta) + \rho \sin(el) \cos(\beta) \dot{el} + \rho \cos(el) \sin(\beta) \dot{\beta}\} \rho_E \\ &\quad - \{\dot{\rho} \cos(el) \sin(\beta) - \rho \sin(el) \sin(\beta) \dot{el} + \rho \cos(el) \cos(\beta) \dot{\beta}\} \rho_S \end{aligned}$$

If we use the expressions from Eq. (4-4), the terms involving $\dot{\rho}$ and \dot{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

$$\dot{\beta} = \frac{\dot{\rho}_S \rho_E - \dot{\rho}_E \rho_S}{\rho_S^2 + \rho_E^2} \quad (4-10)$$

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 27 shows the complete result. We can be more computationally efficient by saving trigonometric variables. The square root of the range-vector component also helps to eliminate operations.

ALGORITHM 27: RAZEL

$$(\dot{\vec{r}}_{ECI}, \dot{\vec{v}}_{ECI}, yr, mo, day, UTC, \Delta UT1, \Delta AT, x_p, y_p, \phi_{gd}, \lambda, h_{ellp} \Rightarrow \rho, \beta, el, \dot{\rho}, \dot{\beta}, \dot{el})$$

$$FK5(\dot{\vec{r}}_{ECI}, \dot{\vec{v}}_{ECI}, yr, mo, day, UTC, \Delta UT1, \Delta AT, x_p, y_p \Rightarrow \dot{\vec{r}}_{ECEF}, \dot{\vec{v}}_{ECEF})$$

$$SITE(\phi_{gd}, \lambda, h_{ellp} \Rightarrow \dot{\vec{r}}_{siteECEF})$$

$$\dot{\vec{\rho}}_{ECEF} = \dot{\vec{r}}_{ECEF} - \dot{\vec{r}}_{SiteECEF}$$

$$\dot{\vec{\rho}}_{ECEF} = \dot{\vec{v}}_{ECEF}$$

$$\dot{\vec{\rho}}_{SEZ} = [\text{ROT2}(90^\circ - \phi_{gd})][\text{ROT3}(\lambda)]\dot{\vec{\rho}}_{ECEF}$$

$$\dot{\vec{\rho}}_{SEZ} = \underbrace{[\text{ROT2}(90^\circ - \phi_{gd})][\text{ROT3}(\lambda)]}_{\text{SEZ}} \dot{\vec{\rho}}_{ECEF}$$

$$\begin{bmatrix} \text{SEZ} \\ \text{ECEF} \end{bmatrix} = \begin{bmatrix} \sin(\phi_{gd})\cos(\lambda) & \sin(\phi_{gd})\sin(\lambda) & -\cos(\phi_{gd}) \\ -\sin(\lambda) & \cos(\lambda) & 0 \\ \cos(\phi_{gd})\cos(\lambda) & \cos(\phi_{gd})\sin(\lambda) & \sin(\phi_{gd}) \end{bmatrix}$$

$$\rho = |\dot{\vec{\rho}}_{SEZ}|$$

$$\sin(el) = \frac{\rho_Z}{\rho}$$

IF Elevation $\neq 90^\circ$

$$\sin(\beta) = \frac{\rho_E}{\sqrt{\rho_S^2 + \rho_E^2}} \quad \cos(\beta) = \frac{-\rho_S}{\sqrt{\rho_S^2 + \rho_E^2}}$$

IF Elevation = 90°

$$\sin(\beta) = \frac{\dot{\rho}_E}{\sqrt{\dot{\rho}_S^2 + \dot{\rho}_E^2}} \quad \cos(\beta) = \frac{-\dot{\rho}_S}{\sqrt{\dot{\rho}_S^2 + \dot{\rho}_E^2}}$$

$$\dot{\rho} = \frac{\dot{\vec{\rho}}_{SEZ} \cdot \dot{\vec{\rho}}_{SEZ}}{\rho} \quad \dot{\beta} = \frac{\dot{\rho}_S \rho_E - \dot{\rho}_E \rho_S}{\rho_S^2 + \rho_E^2} \quad \dot{el} = \frac{\dot{\rho}_Z - \dot{\rho} \sin(el)}{\sqrt{\rho_S^2 + \rho_E^2}}$$

RAZEL is simply the reverse of **SITE-TRACK** (Algorithm 51), so the test cases are just the reverse of those for **SITE-TRACK**. See Example 4-1 following this section.

4.4.4 Practical Az-El Conversions

For precise applications, converting azimuth and elevation to right ascension and declination requires several steps. The previous sections discussed how to convert position and velocity vectors to these angles, but not how to convert between frames. As mentioned earlier, this technique combines many topics, so I'll only outline the general procedure. Consult Seidelmann (1992:145-164) and Roy (1988:67) for additional information.

Seidelmann (1992:20, 145-164) describes many steps needed to accurately convert between given angular values. The major steps for a terrestrial to celestial conversion are as follows.

1. Form the vectors from the observations (β and el : Eq. (4-4) and Eq. (4-5), α and δ : Eq. (4-1) and Eq. (4-2)) in the local frame. The local frame is usually a topocentric frame (SEZ, or $I_t J_t K_t$).
2. Rotate to the geocentric coordinate system. The SEZ frame uses Eq. (3-28).
3. Translate the origin if necessary. The translations are shown in Eq. (4-6) and Eq. (4-7), and in Eq. (4-3).
4. Perform rotations to the celestial frame (motion of the coordinate system). Sec. 3.7 provides the necessary equations. In this case, the vectors would be transformed from the ITRF to the J2000(IAU-76/FK5) or the GCRF.
5. Perform a rotation to the celestial frame if required.
6. Translate the origin to the center of the Solar System (barycenter) if required.
7. Apply corrections as needed—aberration of light, relativistic deflection of light, light time, proper motion, parallax, and radial velocity.
8. Decompose the angular values (usually α and δ) and use for a particular application.

Depending on your application, not all of the steps are required. When observations are found using a star catalog, the process flows from the last step. When you begin with observations in the SEZ frame, the steps go in order. The bottom line is that we must carefully consider what frame the observations are in, and what frame ephemerides are located in for any comparative analysis.

Historically, transformations between topocentric right ascension and declination values with azimuth and elevation were used when range information was unavailable. Although I include these equations, they are only valid for an observer in the PEF frame, and an object in the uniform equinox of date (because θ_{LST} is used). Thus, it is not accu-

rate enough for modern applications, but it is instructive for academic examples and low accuracy applications. To develop the relations, consider Fig. 4-7, in which the spherical triangle shows all the necessary angles. (See Sec. C.4.2 for a discussion of spherical trigonometry.) The type of latitude is not specified because it's implied by the observations, that in turn are measured from a particular plane. For azimuth, this plane is typically the horizontal plane, which determines the type of latitude.

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 [Eq. (3-44)]

$$LHA = \theta_{LST} - \alpha_t \quad (4-11)$$

The cosine law for sides will yield the desired result for elevation:

$$\begin{aligned} \cos(90^\circ - el) &= \cos(90^\circ - \phi) \cos(90^\circ - \delta_t) \\ &\quad + \sin(90^\circ - \phi) \sin(90^\circ - \delta_t) \cos(LHA) \end{aligned}$$

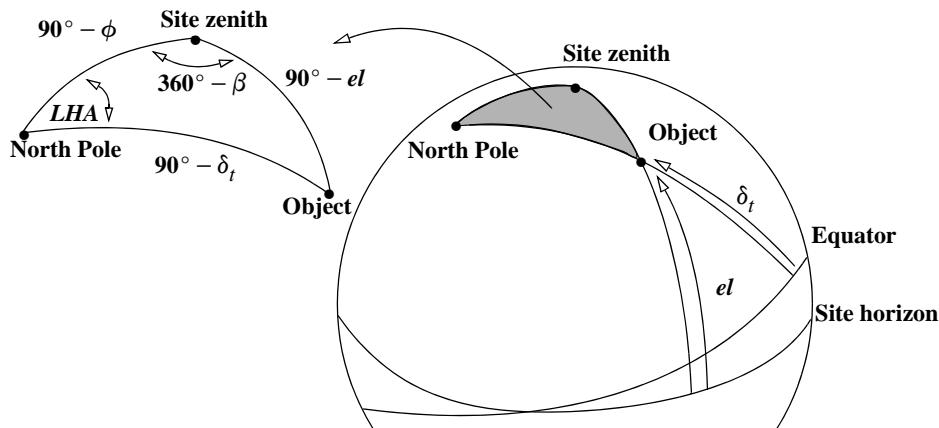


Figure 4-7. Az-el and RA-dec Geometry. The spherical triangles shown are necessary to develop relations between the two sets of angles. Don't confuse the site *horizon* with the equator shown in many diagrams.

Simplifying gives us

$$\sin(el) = \sin(\phi) \sin(\delta_t) + \cos(\phi) \cos(\delta_t) \cos(LHA) \quad (4-12)$$

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

$$\sin(\beta) = -\frac{\sin(LHA)\cos(\delta_t)}{\cos(el)} \quad (4-13)$$

Use the cosine law to find the remaining azimuth expression:

$$\cos(\beta) = \frac{\sin(\delta_t) - \sin(el)\sin(\phi)}{\cos(el)\cos(\phi)} \quad (4-14)$$

The conversion is accomplished using Eq. (4-11) to Eq. (4-14).

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 [Eq. (4-13)], keeping an extra latitude term. Find the cosine expression from the sine of the elevation in Eq. (4-12). The algorithm is

ALGORITHM 28: AZELToRADEC ($\beta, el, \phi, \theta_{LST} \Rightarrow \alpha_t, \delta_t$)

$$\begin{aligned} \sin(\delta_t) &= \sin(el)\sin(\phi) + \cos(el)\cos(\phi)\cos(\beta) \\ \sin(LHA) &= -\frac{\sin(\beta)\cos(el)\cos(\phi)}{\cos(\delta_t)\cos(\phi)} = -\frac{\sin(\beta)\cos(el)}{\cos(\delta_t)} \\ \cos(LHA) &= \frac{\sin(el) - \sin(\phi)\sin(\delta_t)}{\cos(\delta_t)\cos(\phi)} = \frac{\cos(\phi)\sin(el) - \sin(\phi)\cos(\beta)\cos(el)}{\cos(\delta_t)} \\ \alpha_t &= \theta_{LST} - LHA \end{aligned}$$

4.4.5 Transformations for Ecliptic Latitude and Longitude

Representing right ascension and declination angles in the ecliptic latitude and longitude system is the starting point for determining the position of the Sun and Moon, which are common inputs for perturbation analyses. Because the vernal equinox is common to both systems, you must be sure that the coordinate frames are the same. First, find the position vector in the heliocentric coordinate system by looking at Fig. 4-8. Because the geocentric and heliocentric (ecliptic) systems differ only by the obliquity of the ecliptic, it's simple to transform between the systems.

$$\begin{aligned} \hat{r}_{XYZ} &= \begin{bmatrix} r\cos(\phi_{ecliptic})\cos(\lambda_{ecliptic}) \\ r\cos(\phi_{ecliptic})\sin(\lambda_{ecliptic}) \\ r\sin(\phi_{ecliptic}) \end{bmatrix} & \hat{r}_{IJK} &= \text{ROT1}[-\epsilon]\hat{r}_{XYZ} \end{aligned}$$

Note the similarity to Eq. (4-1). To determine the position vector of any object with position magnitude, r , use

$$\hat{r}_{IJK} = r \begin{bmatrix} \cos(\phi_{ecliptic})\cos(\lambda_{ecliptic}) \\ \cos(\epsilon)\cos(\phi_{ecliptic})\sin(\lambda_{ecliptic}) - \sin(\epsilon)\sin(\phi_{ecliptic}) \\ \sin(\epsilon)\cos(\phi_{ecliptic})\sin(\lambda_{ecliptic}) + \cos(\epsilon)\sin(\phi_{ecliptic}) \end{bmatrix} \quad (4-15)$$

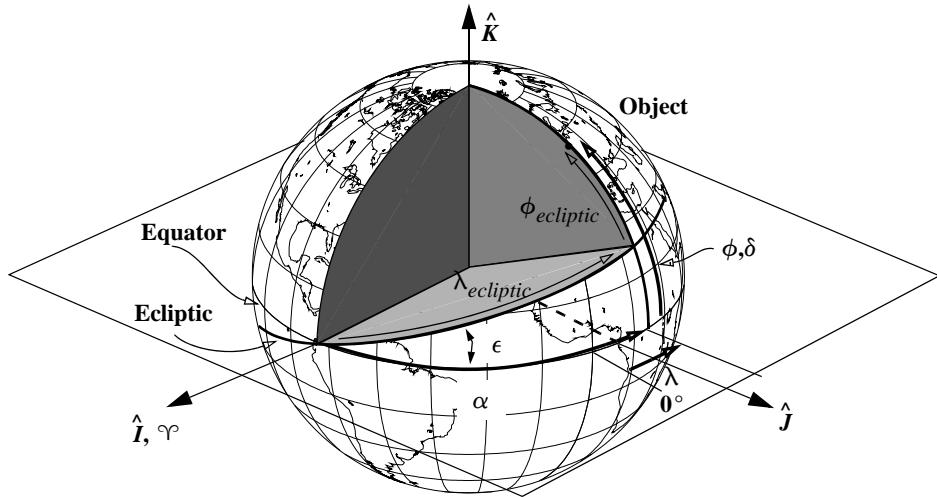


Figure 4-8. Transforming the Ecliptic and Equatorial Systems. Ecliptic coordinates and geocentric coordinates differ by the obliquity of the ecliptic, ϵ and by their origin (heliocentric vs. geocentric, respectively). This permits a simple rotation between systems about the vernal equinox direction.

Solve the problem of converting to right ascension and declination values by looking at unit vectors in each system. In the ecliptic system, we determine a unit vector in the same way as the position vector without the magnitude:

$$\hat{L}_{XYZ} = \begin{bmatrix} \cos(\phi_{ecliptic})\cos(\lambda_{ecliptic}) \\ \cos(\phi_{ecliptic})\sin(\lambda_{ecliptic}) \\ \sin(\phi_{ecliptic}) \end{bmatrix} \quad (4-16)$$

Likewise, in the geocentric system, use the right ascension and declination:

$$\hat{L}_{IJK} = \begin{bmatrix} \cos(\delta)\cos(\alpha) \\ \cos(\delta)\sin(\alpha) \\ \sin(\delta) \end{bmatrix} \quad (4-17)$$

Now rotate the geocentric system to the ecliptic system (or vice versa) depending on the desired formulation for the problem:

$$\hat{L}_{XYZ} = \text{ROT1}(\epsilon)\hat{L}_{IJK} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\epsilon) & \sin(\epsilon) \\ 0 & -\sin(\epsilon) & \cos(\epsilon) \end{bmatrix} \begin{bmatrix} \cos(\delta)\cos(\alpha) \\ \cos(\delta)\sin(\alpha) \\ \sin(\delta) \end{bmatrix} \quad (4-18)$$

This equation yields an immediate way to transform the values. Equating Eq. (4-16) and Eq. (4-18) gives us

$$\begin{aligned}\cos(\phi_{ecliptic}) \cos(\lambda_{ecliptic}) &= \cos(\delta) \cos(\alpha) \\ \cos(\phi_{ecliptic}) \sin(\lambda_{ecliptic}) &= \cos(\delta) \sin(\alpha) \cos(\epsilon) + \sin(\delta) \sin(\epsilon) \\ \sin(\phi_{ecliptic}) &= -\cos(\delta) \sin(\alpha) \sin(\epsilon) + \sin(\delta) \cos(\epsilon)\end{aligned}$$

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:

$$\begin{aligned}\sin(\phi_{ecliptic}) &= -\cos(\delta) \sin(\alpha) \sin(\epsilon) + \sin(\delta) \cos(\epsilon) \\ \sin(\lambda_{ecliptic}) &= \frac{\cos(\delta) \sin(\alpha) \cos(\epsilon) + \sin(\delta) \sin(\epsilon)}{\cos(\phi_{ecliptic})} \\ \cos(\lambda_{ecliptic}) &= \frac{\cos(\delta) \cos(\alpha)}{\cos(\phi_{ecliptic})}\end{aligned}\tag{4-19}$$

Solve the inverse problem by transforming the ecliptic unit vector, \hat{L}_{XYZ} , to the equatorial frame [ROT1($-\epsilon$)] and equating. The result is

$$\begin{aligned}\cos(\delta) \cos(\alpha) &= \cos(\phi_{ecliptic}) \cos(\lambda_{ecliptic}) \\ \cos(\delta) \sin(\alpha) &= -\sin(\phi_{ecliptic}) \sin(\epsilon) + \cos(\phi_{ecliptic}) \cos(\epsilon) \sin(\lambda_{ecliptic}) \\ \sin(\delta) &= \sin(\phi_{ecliptic}) \cos(\epsilon) + \cos(\phi_{ecliptic}) \sin(\epsilon) \sin(\lambda_{ecliptic})\end{aligned}$$

Remember to be careful of the quadrants for α ; that's why we present both sine and cosine expressions. The solution is:

$$\begin{aligned}\sin(\delta) &= \sin(\phi_{ecliptic}) \cos(\epsilon) + \cos(\phi_{ecliptic}) \sin(\epsilon) \sin(\lambda_{ecliptic}) \\ \sin(\alpha) &= \frac{-\sin(\phi_{ecliptic}) \sin(\epsilon) + \cos(\phi_{ecliptic}) \cos(\epsilon) \sin(\lambda_{ecliptic})}{\cos(\delta)} \\ \cos(\alpha) &= \frac{\cos(\phi_{ecliptic}) \cos(\lambda_{ecliptic})}{\cos(\delta)}\end{aligned}\tag{4-20}$$

4.4.6 Practical Aspects of Obtaining Observations

Sensors systems may be categorized into two broad classes, surveillance, and tracking. **Surveillance** is the process in which a search is made for objects whose location may or not be known. **Tracking** is the routine process of updating an objects location by first predicting where it will be, and then observing it. For a complete catalog of space objects, both processes are necessary. Tracking is much more common, and networks like the SSN have far greater numbers of tracking sensors. A true surveillance sensor in the SSN is the former Navy Fence Interferometer, now under control of the Air Force. Note that phased array radars can sometimes act in a surveillance mode by spreading the radiated beam. This reduces the power, and hence, the accuracy and ability to detect is

reduced. The surveillance problem involves many aspects which are beyond the present scope, however I will mention a few basics. Montenbruck (2006: Chap. 6) has additional information.

Many sensor systems average or smooth the data before reporting the observations. This will have a significant effect in the orbit determination and resulting covariance (the estimates will be overly optimistic as we'll see in Chap. 10). Radar systems routinely take periodic outputs from a Kalman filter which processes the nearly continuous stream of observational data. Transponder systems often perform similar averaging techniques on the data.

The volume of space is a limitation for surveillance. Consider observing satellites in geosynchronous orbit. If we plot the longitudes of the satellites at the same time (Fig. 4-10), we observe a relatively diffuse distribution.

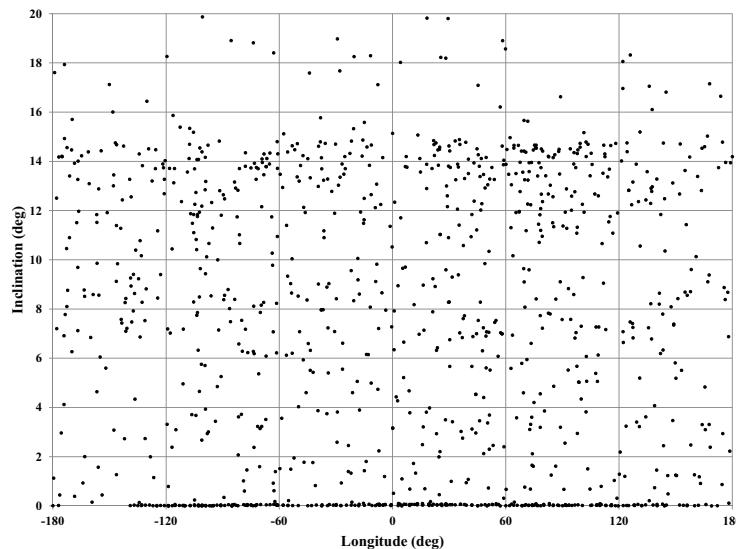


Figure 4-9. Geosynchronous Object Location. Inclination and longitude values of all objects passing within ± 300 km of geosynchronous values. Note the relatively few gaps in the low inclination region.

Burnham and Sridharan (1995) showed that a considerable reduction in search space may be achieved by examining the satellite orbits. Figure 4-10 shows a snapshot of their results using the same data as in Fig. 4-9. Note that the definition of a GEO orbit will influence the results—this figure includes all objects passing within ± 300 km.

Obtaining observations for GEO satellites is difficult for many reasons. Optical sensors provide angles-only observations that are often limited by poor weather.

With sparse data and no range, the orbit determination becomes significantly more difficult. Deep space radars are limited to just a couple of sites. With limited sites, sensor networks can have coverage gaps where satellites can go undetected. New objects

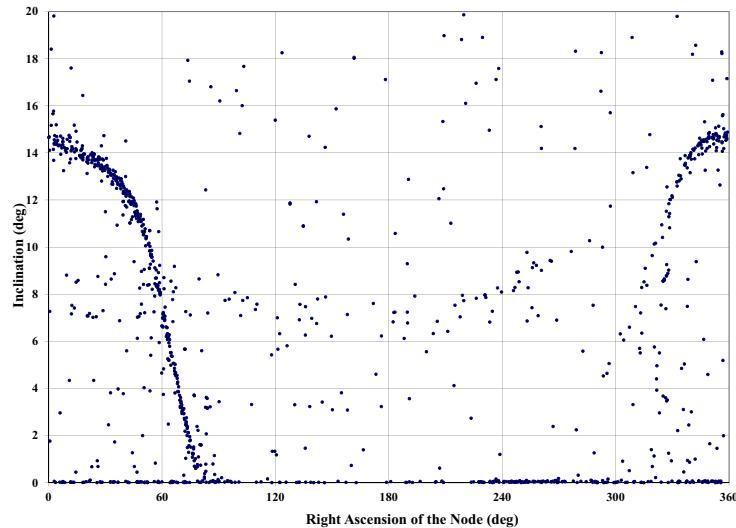


Figure 4-10. Geosynchronous Object Location. Inclination and right ascension values of all objects passing within ± 300 km of geosynchronous values. Note the “pinch point” in right ascension—a dynamic location.

are also missed because surveillance is not generally possible with limited resources. The long range from the Earth makes small objects in GEO difficult to observe (say 10 cm or less). GEO orbits also have unique observability problems where the sensor has limited or no relative motion with respect to the satellite. Satellites are often co-located in popular locations for downlink or coverage. The correlation problem with GEO observations is compounded by satellites that maneuver often (sometimes daily) to keep in a certain “box” (near continual East-West, North-South maneuvering). Lost satellites occur often, primarily because of the inability to process through unknown maneuvers. Third body perturbations affect the orbits with long-term inclination changes—both for debris and operational satellites. Solar Radiation pressure effects are difficult to model and prediction is not very accurate, similar to atmospheric drag for LEO satellites. The combination of these effects makes orbit determination of GEO objects a very difficult problem.

Tasking the sensors to obtain observations can be very difficult. First, there are operational considerations, such as the priority of a particular satellite. In cases where the entire catalog is required, satellites may be tracked to try and maintain an overall catalog accuracy level, and not to lose satellites. For commercial systems, satellites may be tracked equally, assuming no anomalies are detected. Anomalies will usually result in increased tasking and additional observations to try and determine the source of the problem. The military has different priorities based on perceived missions of satellites.

Next, the actual prioritization method must be known. Authors have discussed covariance as a means of determining when a satellite needs to be tracked (e.g., Vallado and Alfano 1999, Hill et al. 2010). With this approach, the covariance is propagated into the

future to determine when a particular satellite will exceed a certain threshold without any additional observations. As Hill points out, this can be a computationally involved exercise. The alternative of an ad hoc tabular listing of priorities and numbers of observations may seem more appealing.

Finally, the bandwidth, communication, and processing aspects are input. With modern communication systems, it's feasible to have gigantic quantities of observations produced and transmitted to a location for processing. Indeed, modern computers can handle remarkable quantities of information. However, there can be limits. Consider a radar producing 100's to 1000's of observations per second. Systems having difficulty processing this much data simply let observations "fall on the floor". Producing this much information on an entire catalog would surely impose computational constraints!

Whatever the scheme for tasking is, observations are collected. The next task is associating the observations, also known as ***observation correlation***, or ***observation association***, and this is a very difficult process. Suppose we have some optical measurements from a CCD. The streaks of the satellites (depending on the type of tracking used by the telescope), can be processed to yield start and end points. Each streak will produce 2 observations of right-ascension declination pairs. This is repeated many times, and each image can possibly have multiple satellites contained in it. Specialized software usually accomplishes this process. However, the result is a set of observations that are not yet correlated with any known object. We often have a good estimate of what satellite we are tracking, but satellites can become confused with nearby satellites to the sensor line of sight, they can maneuver out of our field-of-view, there can be a new launch, and there can be bad data. Each of these possibilities introduces uncertainty into the process. Some techniques seek to solve certain orbital elements (semimajor axis, eccentricity and inclination) and compare the results to an established catalog. However, this results in only gross approximations. Covariance analyses may have promise in this decision process. Fujimoto and Scheeres (2010) have proposed an intriguing idea using probability distributions in Poincare space.

Implementing Observation Transformations

Right ascension and declination observations change over time due to the motion of the equinox. To be precise, coordinates should use reduction techniques (Sec. 3.7), but for this example, we will not introduce this additional processing. An example shows conversions between the various observation types from surveillance and tracking systems.

▼ Example 4-1: Finding Observation Values.

GIVEN: Neptune on May 14, 1994 and a site at the U.S. Air Force Academy with $\phi_{eqd} = 39.007^\circ$, and $\lambda = -104.883^\circ$, $h_{ellp} = 2194.56$ m, and an epoch time of 13:11:20.59856 UTC.

FIND: \dot{r} , $\dot{\vartheta}$, α , δ , α_r , δ_r , β , el , $\phi_{ecliptic}$, $\lambda_{ecliptic}$ and rates

The *Astronomical Almanac* provides data for our analysis. Rates of geocentric right ascension and declination aren't usually given, but we can derive them from data on the following day. On May 14, 1994,

$$r = 29.664\,361 \text{ AU} = 4,437,725,220.51 \text{ km}$$

$$\alpha = 19^h 39^m 57.395^s = 294.989\,145\,8^\circ$$

$$\delta = -20^\circ 49' 24.58'' = -20.823\,494\,4^\circ$$

Because no rate values are published, we can approximate them using published data from May 14 to May 15, 1994. This gives us

$$\dot{r} = \frac{(149,597,870)(29,649\,616 - 29,664\,361)}{86,400} = -25.530\,330\,94 \text{ km/s}$$

$$\dot{\alpha} = \frac{15(54.856 - 57.395)}{(3600)(86,400)} = -0.000\,000\,122\,44^\circ/\text{s}$$

$$\dot{\delta} = \frac{(30.16 - 24.58)}{(3600)(86,400)} = -0.000\,000\,017\,94^\circ/\text{s}$$

Eq. (4-1) and Eq. (4-2) allow us to determine position and velocity vectors:

$$\begin{aligned}\vec{r}_{ECI} &= \begin{bmatrix} r \cos(\delta) \cos(\alpha) \\ r \cos(\delta) \sin(\alpha) \\ r \sin(\delta) \end{bmatrix} = \begin{bmatrix} 1,752,246,215 \\ -3,759,563,433 \\ -1,577,568,105 \end{bmatrix} \text{ km} \\ \vec{v}_{ECI} &= \begin{bmatrix} \dot{r} \cos(\delta) \cos(\alpha) - r \sin(\delta) \cos(\alpha) \dot{\delta} - r \cos(\delta) \sin(\alpha) \dot{\alpha} \\ \dot{r} \cos(\delta) \sin(\alpha) - r \sin(\delta) \sin(\alpha) \dot{\delta} + r \cos(\delta) \cos(\alpha) \dot{\alpha} \\ \dot{r} \sin(\delta) + r \cos(\delta) \dot{\delta} \end{bmatrix} = \begin{bmatrix} -18.323 \\ 18.332 \\ 7.777 \end{bmatrix} \text{ km/s}\end{aligned}$$

The site vector is first found in Earth fixed coordinates,

$$\vec{r}_{siteITRF} = -1275.123\,419 \hat{I} - 4797.994\,704 \hat{J} + 3994.302\,210 \hat{K} \text{ km}$$

$$\vec{v}_{siteITRF} = 0.000 \text{ km/s}$$

and then in inertial coordinates (IAU-76/FK5).

$$\vec{r}_{siteECI} = 4068.213\,205 \hat{I} - 2842.429\,051 \hat{J} + 3996.349\,956 \hat{K} \text{ km}$$

$$\vec{v}_{siteECI} = 0.207\,272\,450 \hat{I} + 0.296\,810\,996 \hat{J} + 0.000\,116\,446 \hat{K} \text{ km/s}$$

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	$r = 4,437,725,220.273$	$\alpha = 294.989\,145\,8$	$\delta = -20.823\,494\,4$
Topocentric	$\rho = 4,437,722,626.456$	$\alpha_t = 294.989\,111\,5$	$\delta_t = -20.823\,562\,4$
Horizon	$\rho = 4,437,722,626.456$	$\beta = 210.825\,066\,7$	$el = 23.859\,505\,2$
Ecliptic	$r = 4,437,725,220.273$	$\lambda_{ecliptic} = 293.258\,210\,8$	$\phi_{ecliptic} = 0.620\,750\,5$

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	$\dot{r} = -25.530\,330\,94$	$\dot{\alpha} = -0.000\,000\,122\,447$	$\dot{\delta} = -0.000\,000\,017\,941$
Topocentric	$\dot{\rho} = -25.360\,813\,8$	$\dot{\alpha}_t = -0.000\,000\,126\,774$	$\dot{\delta}_t = -0.000\,000\,017\,110$

	range rate (km/s)	lateral rate (°/s)	vertical rate (°/s)
Horizon	$\dot{\rho} = -25.360\ 813\ 8$	$\dot{\beta} = 0.003\ 862\ 975\ 427$	$\dot{el} = -0.001\ 663\ 710\ 935$
Ecliptic	$\dot{r} = -25.530\ 472\ 9$	$\dot{\lambda}_{ecliptic} = -0.000\ 000\ 115\ 843$	$\dot{\phi}_{ecliptic} = 0.000\ 000\ 001\ 547$

Notice there are many similarities, so we may question the need for all the representations. We can answer this question by recalculating the entire problem assuming a satellite at a range of 12,756 km with a velocity of 6.798 614 km/s. Because the satellite orbits at a much closer range, we expect the values to differ more dramatically.

$$\dot{\vec{r}}_{ECI} = \begin{bmatrix} 5,036.736\ 529 \\ -10,806.660\ 797 \\ -4,534.633\ 784 \end{bmatrix} \text{km} \quad \dot{\vec{v}}_{ECI} = \begin{bmatrix} 2.684\ 385\ 5 \\ -5.759\ 592\ 0 \\ -2.416\ 809\ 3 \end{bmatrix} \text{km/s}$$

The results are

	range (km)	lateral (°)	vertical (°)
Geocentric	$r = 12,756.000$	$\alpha = 294.989\ 145\ 8$	$\delta = -20.823\ 494\ 4$
Topocentric	$\rho = 11,710.812$	$\alpha_t = 276.933\ 732\ 9$	$\delta_t = -46.758\ 340\ 2$
Horizon	$\rho = 11,710.812$	$\beta = 210.877\ 774\ 7$	$el = -5.940\ 953\ 5$
Ecliptic	$r = 12,756.000$	$\lambda_{ecliptic} = 293.258\ 210\ 9$	$\phi_{ecliptic} = 0.620\ 750\ 5$

	range rate (km/s)	lateral rate (°/s)	vertical rate (°/s)
Geocentric	$\dot{r} = 6.798\ 514\ 0$	$\dot{\alpha} = -0.000\ 000\ 122\ 44$	$\dot{\delta} = -0.000\ 000\ 017\ 94$
Topocentric	$\dot{\rho} = 6.084\ 282\ 6$	$\dot{\alpha}_t = 0.012\ 339\ 704\ 05$	$\dot{\delta}_t = 0.014\ 392\ 462\ 03$
Horizon	$\dot{\rho} = 6.084\ 282\ 6$	$\dot{\beta} = 0.003\ 840\ 114\ 66$	$\dot{el} = 0.014\ 958\ 477\ 59$
Ecliptic	$\dot{r} = 6.798\ 514\ 0$	$\dot{\lambda}_{ecliptic} = -0.000\ 000\ 115\ 84$	$\dot{\phi}_{ecliptic} = 0.000\ 000\ 001\ 55$



Problems

1. 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 the values are $\beta = 55.6^\circ$ and $el = 52.3^\circ$, what are the topocentric α and δ ? Do you need to make any assumptions?
2. Your company has launched a new satellite from Cape Kennedy into a circular, 45° , inclined orbit at about 500 km altitude. Which sensor sites (use Table 4-2) 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?
3. 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 process the data.
4. 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?
5. 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?

CHAPTER 5 CELESTIAL PHENOMENA

- 5.1 Solar Phenomena
- 5.2 Lunar Phenomena
- 5.3 Celestial Applications

5.1 Solar Phenomena

We require position vectors for the Sun to analyze perturbation forces on satellites. Sunrise or sunset conditions are needed when designing a mission for sensor viewing. Using equations used in the reduction of coordinates, we can produce methods to determine the local time of sunrise and sunset. This section details useful routines to accomplish these tasks.

5.1.1 Application: Sun Position Vector

We often want to find the position vector from the Earth to the Sun for uses such as determining solar-panel illumination and remote sensing. Very precise ephemerides of the Sun with respect to the Earth are available through the Jet Propulsion Laboratory (see Appendix D); however, it's often convenient to place a mathematical algorithm inside the particular program using a less precise technique. Some algorithms require us to consider which equator and equinox are referenced (e.g., true-of-date, mean-of-date). Meeus (1991:154-164) shows an analytical method which yields modest accuracy. He discusses how to select and calculate these vectors. We'll discuss a simpler technique from the *Astronomical Almanac* (1992:C24). This latter technique results in a MOD (mean-equator of date) vector with an accuracy of 0.01° . It's valid from 1950 to 2050 because of the truncation of the expansions. Figure 5-1 shows the geometry necessary to visualize the problem.

The solution rests in part on expressions using the J2000.0 epoch. First, find the number of Julian centuries, T_{UT1} , from the epoch, using Eq. (3-42). Find the mean longitude of the Sun, λ_{M_\odot} , in a MOD frame from the Astronomical Almanac (2006:C24).

$$\lambda_{M_\odot} = 280.460^\circ + 36,000.771 T_{UT1}$$

Find the mean anomaly (see Sec. 2.2) for the Sun, M_\odot , from Eq. (3-58). We show T_{TDB} to be precise, but it's acceptable to use T_{UT1} because this is only a low-precision formula.

$$M_\odot = 357.529\ 109\ 2^\circ + 35,999.050\ 34 T_{TDB}$$

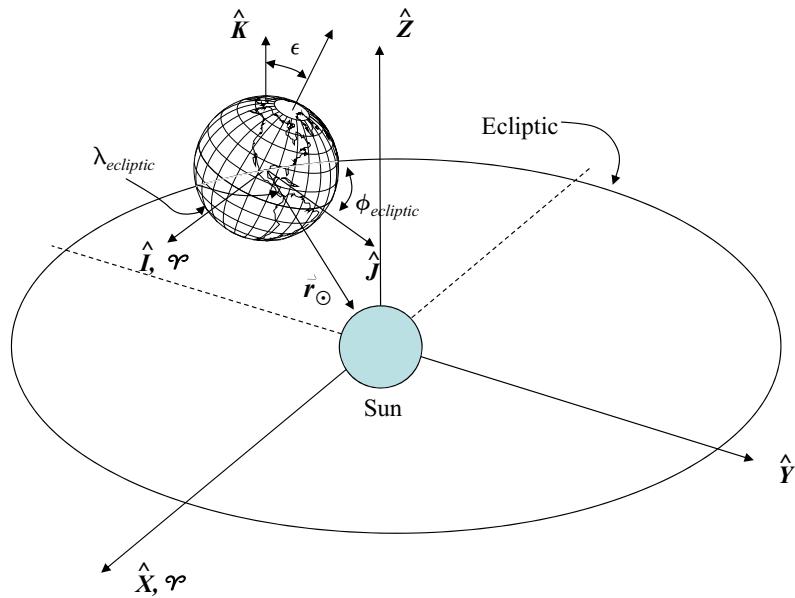


Figure 5-1. Geometry for the Sun Position Vector. The solution of the position vector to the Sun rests on determining the ecliptic longitude (ecliptic latitude is 0°) and the range. The X and I axes are parallel.

Reduce both λ_{M_\odot} and M_\odot to the range of 0° to 360° . Approximate the ecliptic longitude ($\lambda_{ecliptic} \approx \nu_\odot$) by applying the equation of center [see Eq. (2-52)]:

$$\nu_\odot = M_\odot + 2e \sin(M_\odot) + \frac{5e^2}{4} \sin(2M_\odot) + \dots$$

where we use the mean 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 the additional terms are so small. 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 often set to 0.0° .

$$\begin{aligned}\lambda_{ecliptic} &= \lambda_{M_\odot} + 1.914\ 666\ 471^\circ \sin(M_\odot) + 0.019\ 994\ 643 \sin(2M_\odot) \\ \phi_{ecliptic} &= 0^\circ\end{aligned}$$

These expressions have several applications. The trigonometric terms in the ecliptic longitude expression come from the equation of time, introduced in Chap. 3 [Eq. (3-39)].

We can approximate the obliquity of the ecliptic using only the first two terms in Eq. (3-78) 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, or Brouwer and Clemence, 1961:76):

$$\begin{aligned} r = a_e & \left[1 + \frac{e_e^2}{2} + \left\{ -e_e + \frac{3e_e^3}{8} - \frac{5e_e^5}{192} + \frac{7e_e^7}{9216} + \dots \right\} \cos(M_\odot) \right. \\ & + \left\{ -\frac{e_e^2}{2} + \frac{e_e^4}{3} - \frac{e_e^6}{16} + \dots \right\} \cos(2M_\odot) + \left\{ -\frac{3e_e^3}{8} + \frac{45e_e^5}{128} - \frac{567e_e^7}{5120} + \dots \right\} \cos(3M_\odot) \\ & + \left\{ -\frac{e_e^4}{3} + \frac{2e_e^6}{5} + \dots \right\} \cos(4M_\odot) + \left\{ -\frac{125e_e^5}{384} + \frac{4375e_e^7}{9216} + \dots \right\} \cos(5M_\odot) \\ & \left. + \left\{ -\frac{27e_e^6}{80} + \dots \right\} \cos(6M_\odot) + \left\{ -\frac{16087e_e^7}{46,080} + \dots \right\} \cos(7M_\odot) \right] \end{aligned}$$

Find the position magnitude using values of the Earth from Table D-5 ($a_e = 1.000\ 001\ 001\ 78$ AU and $e_e = 0.016\ 708\ 617$):

$$r_\odot = 1.000\ 140\ 612 - 0.016\ 708\ 617 \cos(M_\odot) - 0.000\ 139\ 589 \cos(2M_\odot)$$

If we use Eq. (4-15) and remember that the ecliptic latitude of the Sun is near-zero, the position vector in geocentric equatorial coordinates (MOD) is

$$\hat{\vec{r}}_\odot = r_\odot \cos(\lambda_{ecliptic}) \hat{I} + r_\odot \cos(\epsilon) \sin(\lambda_{ecliptic}) \hat{J} + r_\odot \sin(\epsilon) \sin(\lambda_{ecliptic}) \hat{K}$$

We find the right ascension and declination using Eq. (4-20) and other known values as necessary.

ALGORITHM 29: $SUN(JD_{UT1} \Rightarrow \hat{\vec{r}}_\odot)$

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

$$\lambda_{M_\odot} = 280.460^\circ + 36,000.771 T_{UT1}$$

LET $T_{TDB} \cong T_{UT1}$

$$\begin{aligned}
 M_{\odot} &= 357.529\ 109\ 2^{\circ} + 35,999.050\ 34 T_{TDB} \\
 \lambda_{ecliptic} &= \lambda_{M_{\odot}} + 1.914\ 666\ 471^{\circ} \sin(M_{\odot}) + 0.019\ 994\ 643 \sin(2M_{\odot}) \\
 r_{\odot} &= 1.000\ 140\ 612 - 0.016\ 708\ 617 \cos(M_{\odot}) - 0.000\ 139\ 589 \cos(2M_{\odot}) \\
 \epsilon &= 23.439\ 291^{\circ} - 0.013\ 004\ 2 T_{TDB}
 \end{aligned}$$

$$\vec{r}_{\odot} = \begin{bmatrix} r_{\odot} \cos(\lambda_{ecliptic}) \\ r_{\odot} \cos(\epsilon) \sin(\lambda_{ecliptic}) \\ r_{\odot} \sin(\epsilon) \sin(\lambda_{ecliptic}) \end{bmatrix} \text{AU}$$

▼ **Example 5-1: Finding the Sun Position Vector.**

GIVEN: April 2, 2006, 00:00 UTC
 FIND: Geocentric position vector of the Sun (MOD)

Begin by finding the Julian date using Algorithm 14, $JD_{UTC} = 2,453,827.5$. Letting $JD_{UTC} \approx JD_{UT1}$, the Julian centuries are

$$T_{UT1} = \frac{JD_{UT1} - 2,451,545.0}{36,525} = 0.062\ 491\ 444$$

Let $T_{TDB} \equiv T_{UT1}$. The mean longitude of the Sun and other parameters are

$$\lambda_{M_{\odot}} = 280.460^{\circ} + 36,000.771 T_{TDB} = 10.200\ 110^{\circ}$$

$$M_{\odot} = 357.529\ 109\ 2^{\circ} + 35,999.050\ 34 T_{TDB} = 87.160\ 369^{\circ}$$

$$\lambda_{ecliptic} = \lambda_{M_{\odot}} + 1.914\ 666\ 471^{\circ} \sin(M_{\odot}) + 0.019\ 994\ 643 \sin(2M_{\odot}) = 12.114\ 404^{\circ}$$

$$\phi_{ecliptic} = 0^{\circ}$$

As a check, the *Astronomical Almanac* lists ecliptic latitude as $0.30''$ or $0.000\ 083\ 3^{\circ}$ and longitude as $12^{\circ}07' 8.65''$ or $12.119\ 069\ 44^{\circ}$. The obliquity of the ecliptic is

$$\epsilon = 23.439\ 291^{\circ} - 0.013\ 004\ 2 T_{TDB} = 23.438\ 478^{\circ}$$

The magnitude of the distance to the Sun is

$$r_{\odot} = 1.000\ 140\ 612 - 0.016\ 708\ 617 \cos(M_{\odot}) - 0.000\ 139\ 589 \cos(2M_{\odot}) = 0.999\ 451\ 8 \text{ AU}$$

The vector referred to the mean equator of date is

$$\vec{r}_{\odot} = \begin{bmatrix} r_{\odot} \cos(\lambda_{ecliptic}) \\ r_{\odot} \cos(\epsilon) \sin(\lambda_{ecliptic}) \\ r_{\odot} \sin(\epsilon) \sin(\lambda_{ecliptic}) \end{bmatrix} = \begin{bmatrix} 0.977\ 194\ 5 \\ 0.192\ 442\ 4 \\ 0.083\ 430\ 8 \end{bmatrix} \text{AU} = \begin{bmatrix} 146,186,212 \\ 28,788,976 \\ 12,481,064 \end{bmatrix} \text{km}$$

The *Astronomical Almanac* lists the inertial (ICRS) vector as

$$\vec{r}_\odot = \begin{bmatrix} 0.977\,678\,2 \\ 0.191\,152\,1 \\ 0.082\,871\,7 \end{bmatrix} \text{AU} = \begin{bmatrix} 146,259,922 \\ 28,595,947 \\ 12,397,430 \end{bmatrix} \text{km}$$

A direct comparison is possible only after converting the MOD vector to the ICRS (GCRF). However, using an approximate relation ($TTT \approx T_{UT1}$) and accounting only for precession, the difference is about 28873 km, or 0.0002 AU.

Find the right ascension and declination from Eq. (4-20). The formula is simplified because $\phi_{ecliptic} = 0^\circ$. Computer implementations should first find δ_\odot , and then use ATAN2 to find α_\odot with

$$\begin{aligned} \sin(\alpha_\odot) &= \frac{\cos(\epsilon)\sin(\lambda_{ecliptic})}{\cos(\delta_\odot)} & \cos(\alpha_\odot) &= \frac{\cos(\lambda_{ecliptic})}{\cos(\delta_\odot)} \\ \tan(\alpha_\odot) &= \cos(\epsilon)\tan(\lambda_{ecliptic}) & \alpha_\odot &= 11.140\,899^\circ \\ \sin(\delta_\odot) &= \sin(\epsilon)\sin(\lambda_{ecliptic}) & \delta_\odot &= 4.788\,425^\circ \end{aligned}$$

The *Astronomical Almanac* lists the apparent [true of date] values:

▲ $\alpha_\odot = 0^h\,44^m\,33.42^s = 11.139\,250^\circ$ and $\delta_\odot = 4^\circ\,47'\,18.3'' = 4.788\,417^\circ$

5.1.2 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 position about the Sun. *Sunrise* and *sunset* are the times when the apparent upper limb of the Sun is seen on the horizon by 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° latitude. Above 65° , some 24^h intervals have no rising or setting phenomenon. I've made some changes to the equations to maintain notation and to improve accuracy. A key parameter in the discussion is the angle between the site and the Sun, ζ , shown in Fig. 5-2.

To start, we have various definitions for what constitutes solar illumination. The simplest method defines 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. We designate this as a sunrise or sunset. Illustrated in Fig. 5-2 is the fact that the Sun's light actually "extends" around the Earth a little, due to the Sun's immense size (the Sun should not be considered as a point source of light). 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 arcminutes 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 landmasses 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 5-2 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

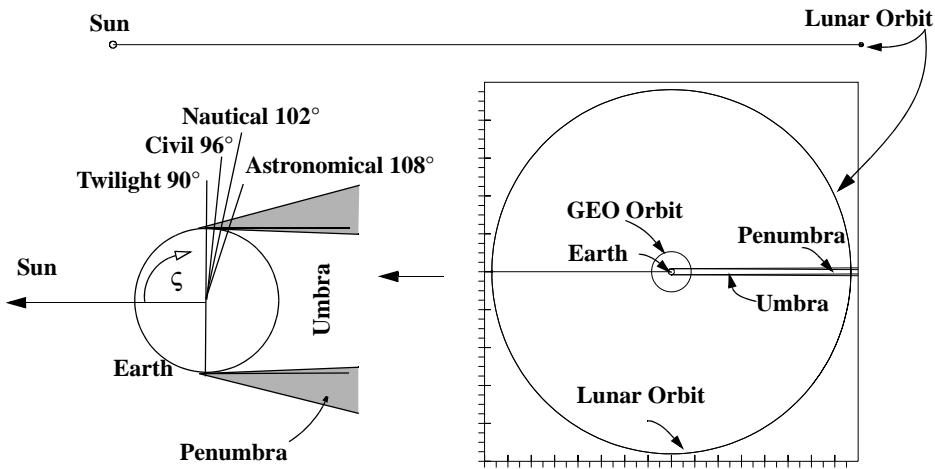


Figure 5-2. Geometry for Solar Illumination. I've shown various scales to highlight the immense distances between the Sun and Earth. Sunrise, sunset, and twilight times depend on the angle between the site and the Sun, ζ , which defines the *position* and *timing* of an event. The various angles also relate to the types of illumination.

quite small. Analysis of the time spent in these regions is important when evaluating the availability of sunlight for satellite power.

The procedure to find the sunrise, sunset, and twilight times involves several concepts of time from Chap. 3. In summary, we must process the algorithm separately for rising or setting phenomenon—*independent* of the site of interest. Once calculated, the parameters yield values of sidereal time, which we may subtract from and add to the site longitude to determine the UT for the phenomenon.*

The procedure begins identically to the first few steps in Algorithm 29 (*SUN*) except for the JD. For this case, we must calculate the JD on the date of interest for either a rise or set phenomenon. Theoretically, 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:

$$JD_{\text{sunrise}} = JD_{0^{\text{h}}} + 6^{\text{h}}/24^{\text{h}}/\text{day} - \lambda/360^{\circ}/\text{day}$$

$$JD_{\text{sunset}} = JD_{0^{\text{h}}} + 18^{\text{h}}/24^{\text{h}}/\text{day} - \lambda/360^{\circ}/\text{day}$$

Once we find the mean longitude, mean anomaly, and ecliptic longitude from Algorithm 29, we determine the right ascension and declination using Eq. (4-20). 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, or use *ATAN2*.

* 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.

$$\begin{aligned} \sin(\delta_{\odot}) &= \sin(\epsilon) \sin(\lambda_{ecliptic}) \\ \tan(\alpha_{\odot}) &= \cos(\epsilon) \tan(\lambda_{ecliptic}) \quad \left\{ \begin{array}{l} \sin(\alpha_{\odot}) = \frac{\cos(\epsilon) \sin(\lambda_{ecliptic})}{\cos(\delta_{\odot})} \\ \cos(\alpha_{\odot}) = \frac{\cos(\lambda_{ecliptic})}{\cos(\delta_{\odot})} \end{array} \right. \end{aligned}$$

Now find the local hour angle of the site. Algorithm 28 (*AZELTORADEC*) provides the relevant equation. Notice you must recalculate the resulting equation because the argument ς is necessary to determine the type of phenomenon. Using $\sin(90^\circ - \varsigma) = \cos(\varsigma)$, find this result:

$$\cos(LHA_{\odot}) = \frac{\cos(\varsigma) - \sin(\delta_{\odot}) \sin(\phi_{gc})}{\cos(\delta_{\odot}) \cos(\phi_{gc})}$$

As always, the cosine provides two solutions to the inverse trigonometric operation. Don't use the two values from one cosine operation for the rise and set values of LHA. You must recalculate the parameters for the Sun using the appropriate JD value. The primary answer is for the sunset calculation loop, and the secondary answer ($360^\circ - LHA$) is for the sunrise loop. If you find GMST [Eq. (3-45)] at 6 hours and 18 hours on the day of interest, Eq. (3-44) allows you to determine the longitudinal difference. This represents the UT of the phenomena. The universal time is

$$UT = LHA_{\odot} + \alpha_{\odot} - GMST$$

Watch your 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 30: SUNRISESET ($JD_{UT1}, \phi_{gc}, \lambda, \varsigma \Rightarrow UT_{sunrise}, UT_{sunset}$)

Do separate calculations for sunrise and sunset using

$$JD_{sunrise} = JD_{0h} + 6/24 - \lambda/360, \quad JD_{sunset} = JD_{0h} + 18/24 - \lambda/360$$

$$T_{UT1} = \frac{JD_{UT1} - 2,451,545.0}{36,525} \quad \text{Let } T_{TDB} \equiv T_{UT1}$$

$$\lambda_{M_{\odot}} = 280.460\ 618\ 4^\circ + 36,000.770\ 053\ 61 T_{UT1}$$

$$M_{\odot} = 357.529\ 109\ 2^\circ + 35,999.050\ 34 T_{TDB}$$

$$\lambda_{ecliptic} = \lambda_{M_{\odot}} + 1.914\ 666\ 471^\circ \sin(M_{\odot}) + 0.019\ 994\ 643 \sin(2M_{\odot})$$

$$\epsilon = 23.439\ 291^\circ - 0.013\ 004\ 2 T_{TDB}$$

$$\tan(\alpha_{\odot}) = \cos(\epsilon) \tan(\lambda_{ecliptic})$$

$$\sin(\delta_{\odot}) = \sin(\epsilon) \sin(\lambda_{ecliptic})$$

$$\cos(LHA_{sunset}) = \frac{\cos(\zeta) - \sin(\delta_{\odot}) \sin(\phi_{gc})}{\cos(\delta_{\odot}) \cos(\phi_{gc})}$$

IF a sunrise calculation

$$LHA_{sunrise} = 360^{\circ} - LHA_{sunset}$$

$$LSTIME (JD_{UT1}, \lambda \Rightarrow LST, GMST)$$

$$UT = LHA_{\odot} + \alpha_{\odot} - GMST$$

Consider the following example.

▼ **Example 5-2. 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. Be sure to use the JD at 0^h!

$$JD_{sunrise} = JD_{0^h} + 6/24 - \lambda/360 = 2,450,165.75$$

$$JD_{sunset} = JD_{0^h} + 18/24 - \lambda/360 = 2,450,166.25$$

Next, find several intermediate values. Consider the sunrise case, for which

$$T_{UT1} = \frac{JD_{UT1} - 2,451,545.0}{36,525} = \frac{2,450,165.75 - 2,451,545.0}{36,525} = -0.037762$$

$$\lambda_{M_{\odot}} = 280.460\ 618\ 4^{\circ} + 36.000.770\ 053\ 61 T_{UT1} = 1.006\ 488^{\circ}$$

Let $T_{TDB} \cong T_{UT1}$ and be sure to use $\zeta = 90^{\circ}\ 50'$

$$M_{\odot} = 357.529\ 109\ 2^{\circ} + 35.999.050\ 34 T_{TDB} = 78.139^{\circ}$$

$$\lambda_{ecliptic} = \lambda_{M_{\odot}} + 1.914\ 666\ 471^{\circ} \sin(M_{\odot}) + 0.019\ 994\ 643 \sin(2M_{\odot}) = 2.888\ 32^{\circ}$$

$$\epsilon = 23.439\ 291^{\circ} - 0.013\ 004\ 2 T_{TDB} = 23.439\ 782^{\circ}$$

$$\tan(\alpha_{\odot}) = \cos(\epsilon) \tan(\lambda_{ecliptic}), \alpha_{\odot} = 2.650\ 32^{\circ}$$

$$\sin(\delta_{\odot}) = \sin(\epsilon) \sin(\lambda_{ecliptic}), \delta_{\odot} = 1.148\ 52^{\circ}$$

$$\cos(LHA_{sunset}) = \frac{\cos(\zeta) - \sin(\delta_{\odot}) \sin(\phi_{gc})}{\cos(\delta_{\odot}) \cos(\phi_{gc})}, LHA_{sunset} = 92.052\ 31^{\circ}$$

Because you're considering sunrise, adjust the value of the LHA:

$$LHA_{sunrise} = 360^{\circ} - LHA_{sunset} = 267.947\ 69^{\circ}$$

Use Eq. (3-45) to find the GMST at 6^h on March 23, 1996. $GMST = -178.994^{\circ}$

Find $UT_{sunrise}$ as

$$UT_{sunrise} = LHA_{\odot} + \alpha_{\odot} - GMST = 89.59152^\circ \Rightarrow 5^h 58^m 21.9659^s$$

Similar operations for sunset [remember to use JD_{sunset} ($T_{UT1} = -0.037748$)] lead to

$$\lambda_{M_{\odot}} = 1.499312^\circ, M_{\odot} = 78.631^\circ, \lambda_{ecliptic} = 3.384^\circ, \epsilon = 23.439782^\circ$$

$$\alpha_{\odot} = 3.105^\circ, \delta_{\odot} = 1.346^\circ, LHA_{sunset} = 92.21787^\circ, GMST_{\odot} = -178.501^\circ \text{ (at } 18^h \text{ UT)}$$

$$UT_{sunset} = LHA_{\odot} + \alpha_{\odot} - GMST = 273.824006^\circ \Rightarrow 18^h 15^m 17.7614^s$$



These values are identical to those found in the Almanac (1996).

We can also examine the behavior of each twilight condition over an entire year. Figure 5-3 shows the sunrise and sunset information for a site at 40° N, 105° W, and 3000 m altitude. The Moon rising and setting phenomena will be discussed in Sec. 5.2.2.

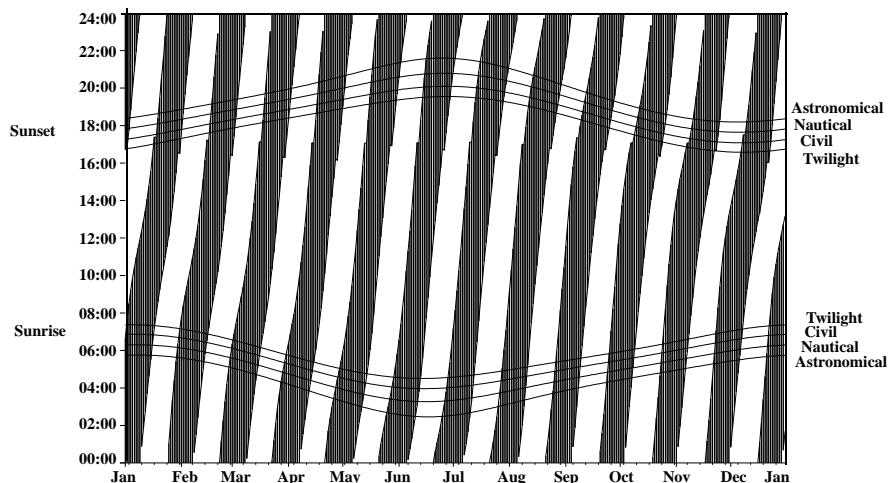


Figure 5-3. Solar and Lunar Rising and Setting Times. The local times of solar and lunar phenomena over a year (1999) show the annual variation caused by the Earth's orbit about the Sun. The dark regions show when the Moon is above the local horizon.

5.2 Lunar Phenomena

Knowledge of the Moon's location and illumination is required in many applications. For instance in astronomy, Lunar illumination is used to determine optimum observation times. Optical tracking devices often include software and hardware that prevents direct observation of the Moon. Faint objects are difficult to observe during a full Moon. We'll investigate basic routines to compute these properties.

5.2.1 Application: Moon Position Vector

We often need to know the position vector from the Earth to the Moon. The ephemerides of the Jet Propulsion Laboratory (see Appendix D) are 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, $\lambda_{\text{ecliptic}}$, 0.2° in ecliptic latitude, ϕ_{ecliptic} , and 1275 km in distance. The vectors reference the IAU-76/FK5 (mean-equator, mean equinox) frame. Consult Meeus (1991:307–314) for more accurate algorithms. If you use higher-precision theories, the complexity of the Moon’s motion becomes evident. Figure 5-4 is useful to visualize the transformations.

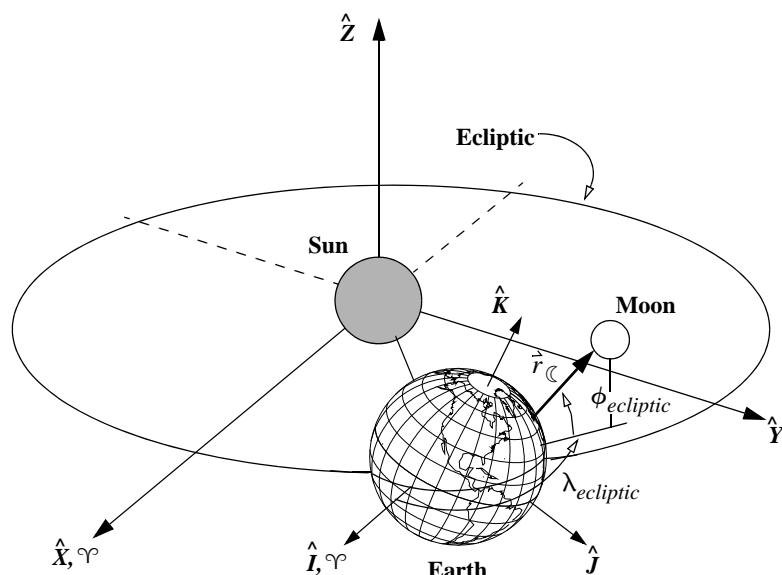


Figure 5-4. Geometry for the Moon Position Vector. Notice the ecliptic latitude is not zero for the Moon. The ecliptic coordinates determine the position vector. This figure is extremely exaggerated. X and I are parallel.

The Moon’s actual 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 theories to determine the position of the Moon—in 1900! The 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 complex, I’ll present the shortened equations and briefly describe their meanings. As with the Sun, we find the ecliptic latitude, longitude, and then include parallax using the following equations (Green, 1988:174):

$$\begin{aligned}\lambda_{ecliptic} = & \lambda_{\odot} + 6.29 \sin(M_{\odot}) - 1.27 \sin(M_{\odot} - 2D_{\odot}) + 0.66 \sin(2D_{\odot}) \\ & + 0.21 \sin(2M_{\odot}) - 0.19 \sin(M_{\odot}) - 0.11 \sin(2u_{M_{\odot}})\end{aligned}$$

$$\begin{aligned}\phi_{ecliptic} = & 5.13^\circ \sin(u_{M_{\odot}}) + 0.28 \sin(M_{\odot} + u_{M_{\odot}}) \\ & - 0.28 \sin(u_{M_{\odot}} - M_{\odot}) - 0.17 \sin(u_{M_{\odot}} - 2D_{\odot})\end{aligned}$$

$$\begin{aligned}\varphi = & 0.9508^\circ + 0.0518 \cos(M_{\odot}) + 0.0095 \cos(M_{\odot} - 2D_{\odot}) \\ & + 0.0078 \cos(2D_{\odot}) + 0.0028 \cos(2M_{\odot})\end{aligned}$$

Find expressions for the mean anomalies, M_{\odot} , the mean argument of latitude, $u_{M_{\odot}}$, and elongation, D_{\odot} , from Eq. (3-58). Meeus (1991:132) lists the Moon's longitude as

$$\lambda_{\odot} = 218.32^\circ + 481,267.883 T_{TDB}$$

Several other terms are also needed. $(2u_{M_{\odot}})$ 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 (the periodic variation is called *evection*). $M_{\odot} - 2D_{\odot}$ 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 (J2000.0) using Eq. (3-42). We use TDB exclusively for this algorithm. Eq. (3-58) 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. (5-10). Then, use Eq. (4-15) 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 nonzero component of ecliptic latitude is present. The *Astronomical Almanac* replaces the sine and cosine values of the obliquity of the ecliptic with constants rather than trigonometric terms. You may prefer these values in some situations, but you'll get higher accuracy from the trigonometric expressions. Determine the obliquity of the ecliptic from Eq. (3-68) and then find

$$\vec{r}_{\odot} = r_{\odot} \begin{bmatrix} \cos(\phi_{ecliptic}) \cos(\lambda_{ecliptic}) \\ \cos(\epsilon) \cos(\phi_{ecliptic}) \sin(\lambda_{ecliptic}) - \sin(\epsilon) \sin(\phi_{ecliptic}) \\ \sin(\epsilon) \cos(\phi_{ecliptic}) \sin(\lambda_{ecliptic}) + \cos(\epsilon) \sin(\phi_{ecliptic}) \end{bmatrix}$$

As with the algorithm for the Sun position vector, we can determine the right ascension and declination if necessary using Eq. (4-20).

ALGORITHM 31: Moon ($JD_{TDB} \Rightarrow \vec{r}_{\mathbb{C}}$)

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

$$\begin{aligned}\lambda_{ecliptic} &= 218.32^\circ + 481,267.8813T_{TDB} + 6.29 \sin(134.9 + 477,198.85T_{TDB}) \\ &\quad - 1.27 \sin(259.2 - 413,335.38T_{TDB}) + 0.66 \sin(235.7 + 890,534.23T_{TDB}) \\ &\quad + 0.21 \sin(269.9 + 954,397.70T_{TDB}) - 0.19 \sin(357.5 + 35,999.05T_{TDB}) \\ &\quad - 0.11 \sin(186.6 + 966,404.05T_{TDB})\end{aligned}$$

$$\begin{aligned}\phi_{ecliptic} &= 5.13^\circ \sin(93.3 + 483,202.03T_{TDB}) + 0.28 \sin(228.2 + 960,400.87T_{TDB}) \\ &\quad - 0.28 \sin(318.3 + 6003.18T_{TDB}) - 0.17 \sin(217.6 - 407,332.20T_{TDB})\end{aligned}$$

$$\begin{aligned}\varphi &= 0.9508^\circ + 0.0518 \cos(134.9 + 477,198.85T_{TDB}) \\ &\quad + 0.0095 \cos(259.2 - 413,335.38T_{TDB}) + 0.0078 \cos(235.7 + 890,534.23T_{TDB}) \\ &\quad + 0.0028 \cos(269.9 + 954,397.70T_{TDB})\end{aligned}$$

$$\epsilon = 23.439^\circ 291^\circ - 0.013 004 2T_{TDB} - 1.64 \times 10^{-7} T_{TDB}^2 + 5.04 \times 10^{-7} T_{TDB}^3$$

$$r_{\mathbb{C}} = \frac{1}{\sin(\varphi)}$$

$$\vec{r}_{\mathbb{C}} = r_{\mathbb{C}} \begin{bmatrix} \cos(\phi_{ecliptic}) \cos(\lambda_{ecliptic}) \\ \cos(\epsilon) \cos(\phi_{ecliptic}) \sin(\lambda_{ecliptic}) - \sin(\epsilon) \sin(\phi_{ecliptic}) \\ \sin(\epsilon) \cos(\phi_{ecliptic}) \sin(\lambda_{ecliptic}) + \cos(\epsilon) \sin(\phi_{ecliptic}) \end{bmatrix}$$

An example shows the process.

▼ **Example 5-3. Finding the Moon Position Vector.**

GIVEN: April 28, 1994, 0:00 UTC

FIND: Geocentric position vector of the Moon

Begin by letting $UTC \approx UT1$. For precision, we would rigorously find TDB using Algorithm 16, but for this example, let $TDB \equiv UT1$. Next, determine the Julian date ($JD = 2,449,470.5$). The Julian centuries are

$$T_{TDB} = \frac{2,449,470.5 - 2,451,545.0}{36,525.0} = -0.056 796 714$$

$$\begin{aligned}
\lambda_{ecliptic} &= 218.32^\circ + 481,267.8813T_{TDB} + 6.29 \sin(134.9 + 477,198.85T_{TDB}) \\
&\quad - 1.27 \sin(259.2 - 413,335.38T_{TDB}) + 0.66 \sin(235.7 + 890,534.23T_{TDB}) \\
&\quad + 0.21 \sin(269.9 + 954,397.70T_{TDB}) - 0.19 \sin(357.5 + 35,999.05T_{TDB}) \\
&\quad - 0.11 \sin(186.6 + 966,404.05T_{TDB}) \\
&= -111.762\ 880\ 1^\circ = 248.237\ 119\ 9^\circ
\end{aligned}$$

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 $\varphi = 1.009\ 936\ 11^\circ$. The obliquity of the ecliptic is

$$\epsilon = 23.439\ 291^\circ - 0.013\ 004\ 2T_{TDB} = 23.440\ 029\ 6^\circ$$

Next, find the magnitude of the position vector:

$$r_{\mathbb{C}} = \frac{1}{\sin(\varphi)} = 362144.6075 \text{ km}$$

Now determine the position vector:

$$\begin{aligned}
\hat{\vec{r}}_{\mathbb{C}} &= r_{\mathbb{C}} (\cos(\phi_{ecliptic}) \cos(\lambda_{ecliptic})) \hat{I} \\
&\quad + r_{\mathbb{C}} (\cos(\epsilon) \cos(\phi_{ecliptic}) \sin(\lambda_{ecliptic}) - \sin(\epsilon) \sin(\phi_{ecliptic})) \hat{J} \\
&\quad + r_{\mathbb{C}} (\sin(\epsilon) \cos(\phi_{ecliptic}) \sin(\lambda_{ecliptic}) + \cos(\epsilon) \sin(\phi_{ecliptic})) \hat{K} \\
\vec{r}_{\mathbb{C}} &= \begin{bmatrix} -134,240.626 \\ -311,571.590 \\ -126,693.785 \end{bmatrix} \text{ km}
\end{aligned}$$

If right ascension and declination are required, use Eq. (4-20). Computer programs should use ATAN2 with Eq. (4-20) to first find α .

$$\sin(\delta) = \sin(\phi_{ecliptic}) \cos(\epsilon) + \cos(\phi_{ecliptic}) \sin(\epsilon) \sin(\lambda_{ecliptic})$$

$$\delta = -20.477\ 702^\circ$$

$$\sin(\alpha) = \frac{-\sin(\phi_{ecliptic}) \sin(\epsilon) + \cos(\phi_{ecliptic}) \cos(\epsilon) \sin(\lambda_{ecliptic})}{\cos(\delta)}$$

$$\alpha = 113.308\ 897^\circ \text{ or } 246.691\ 103^\circ$$

A quadrant check is needed to determine the correct right ascension. The *Astronomical Almanac* again lists the apparent [true of date] values:

▲ $\alpha = 16^{\text{h}} 26^{\text{m}} 34.60^{\text{s}} = 246.644\ 165\ 91^\circ$ and $\delta = -20^\circ 22' 27.00'' = -20.374\ 166\ 67^\circ$

5.2.2 Application: Moon Rise and Set Times

The times at which moonrise and moonset occur will vary during the year based on the Earth's rotation and its place in orbit around the Sun. *Moonrise* and *moonset* are the times when the upper limb of the Moon is on the horizon for an observer on the Earth. We find this using the true altitude of the Moon, $h_{\mathbb{C}}$, which uses the horizontal refraction, the local parallax at the site, and the semi-diameter of the Moon.

$$h_{\mathbb{C}} = \sin\left(-\frac{34}{60} + 0.7275 \varphi - 0.0353 \sqrt{h_{ellp}}\right) \approx 0.00233$$

The *Almanac for Computers* (1990:B9–B11) gives a formula to find these times, although I've made some changes to maintain notational consistency and used Hiebert's (2011) suggestion for a better way to iterate. The procedure to find the moonrise/moonset times is similar to the sunrise/sunset algorithm. It's accurate to about ± 3 minutes within about 60° latitude.

The algorithm executes separately for both a rising or setting phenomenon. Once calculated, the lunar parameters yield values of sidereal time, which we may subtract from, and add to, the site longitude to determine the UT for the phenomenon. I don't specify a particular UT in this section because the method is only approximate. In the calculation of LHA_n , it's possible to obtain answers that are greater than 1.0, or less than -1.0 . These values indicate there is no phenomena (rising or setting) for the location on that day ($|\cos^{-1}(x_n)| > 1$).

The formal procedure finds many of the initial parameters from Algorithm 31 (*Moon*) including the direction cosines (l, m, n) of the Moon.

ALGORITHM 32: MOONRISESET ($JD_{UT1}, \phi_{gc}, \lambda \Rightarrow UT_{moonrise}, UT_{moonset}$)

Do separate calculations for moonrise: moonset using the same starting JD

$$JD_{temp} = JD_{0h}$$

WHILE ($\Delta UT > 30^s$

$$n = n + 1$$

$$T_{UT1} = \frac{JD_{Temp} - 2,451,545.0}{36,525} \quad \text{Let } T_{TDB} \equiv T_{UT1}$$

$$\begin{aligned} \lambda_{ecliptic} = & 218.32^\circ + 481,267.8813T_{TDB} + 6.29 \sin(134.9 + 477,198.85T_{TDB}) \\ & - 1.27 \sin(259.2 - 413,335.38T_{TDB}) + 0.66 \sin(235.7 + 890,534.23T_{TDB}) \\ & + 0.21 \sin(269.9 + 954,397.70T_{TDB}) - 0.19 \sin(357.5 + 35,999.05T_{TDB}) \\ & - 0.11 \sin(186.6 + 966,404.05T_{TDB}) \end{aligned}$$

$$\begin{aligned} \phi_{ecliptic} = & 5.13^\circ \sin(93.3 + 483,202.03T_{TDB}) + 0.28 \sin(228.2 + 960,400.87T_{TDB}) \\ & - 0.28 \sin(318.3 + 6003.18T_{TDB}) - 0.17 \sin(217.6 - 407,332.20T_{TDB}) \end{aligned}$$

$$\epsilon = 23.439\ 291^\circ - 0.013\ 004\ 2 T_{TDB}$$

$$\begin{aligned}l &= \cos(\phi_{ecliptic})\cos(\lambda_{ecliptic}) \\m &= \cos(\epsilon)\cos(\phi_{ecliptic})\sin(\lambda_{ecliptic}) - \sin(\epsilon)\sin(\phi_{ecliptic}) \\n &= \sin(\epsilon)\cos(\phi_{ecliptic})\sin(\lambda_{ecliptic}) + \cos(\epsilon)\sin(\phi_{ecliptic})\end{aligned}$$

$$\alpha_{\mathbb{C}} = \text{ATAN}2(m, l)$$

$$\sin(\delta_{\mathbb{C}}) = n$$

LSTIME (JD_{Temp}, λ ⇒ LST, GMST)

$$GHA_{\mathbb{C}_n} = GMST - \alpha_{\mathbb{C}}$$

$$LHA = GHA_{\mathbb{C}_n} + \lambda$$

$$\left[\begin{array}{l} \text{IF first time} \\ \quad \Delta GHA = 347.81^\circ \\ \text{ELSE} \\ \quad \Delta GHA = \frac{GHA_{\mathbb{C}_n} - GHA_{\mathbb{C}}}{\Delta UT} \end{array} \right]$$

$$\text{IF } \Delta GHA < 0 \text{ THEN } \Delta GHA = \Delta GHA + \frac{360^\circ}{\Delta UT}$$

$$x_n = \frac{0.00233 - \sin(\delta_{\mathbb{C}})\sin(\phi_{gc})}{\cos(\delta_{\mathbb{C}})\cos(\phi_{gc})}$$

IF $x_n > 1$, next day phenomena

$$\text{IF Moonrise calculation THEN } LHA_n = 360^\circ - \cos^{-1}(x_n)$$

$$\text{ELSE } LHA_n = \cos^{-1}(x_n)$$

$$\Delta UT = \frac{LHA_n - LHA}{\Delta GHA}$$

$$\left[\begin{array}{l} \text{IF } \Delta UT < -0.5 \text{ {event is on the day before}} \\ \quad \Delta UT = \Delta UT + \frac{360^\circ}{\Delta GHA} \\ \text{ELSE IF } \Delta UT > 0.5 \text{ {event is on the day after}} \\ \quad \Delta UT = \Delta UT - \frac{360^\circ}{\Delta GHA} \end{array} \right]$$

$$\Delta t = \Delta t + \Delta UT$$

$$JD_{temp} = JD_{temp} + \Delta UT$$

$$GHA_{\text{C}} = GHA_{\text{C}_n}$$

END WHILE LOOP

$$UT_{\text{moonrise}}, UT_{\text{moonset}} = \Delta t(24)$$

Figure 5-3 shows the moonrise and set times over one year. Notice the somewhat irregular nature of the motion. Consider the following example.

▼ Example 5-4. Calculating Moonrise-Moonset Times.

GIVEN: Site at 40°N, 0°E on August 21, 1998

FIND: Moonrise and Moonset times

Initial parameters are set to 0.0.

LOOP

$$JD_{Temp} = JD_{0^{\text{h}}} = 2,451,046.5$$

The calculations for Moonrise and Moonset each require varying numbers of iterations. In this example, there are 3 iterations for the calculations. The first iteration contains several values that are the same (denoted “*”) because the initial JD is the same. Begin by finding several intermediate values.

$$\begin{aligned} T_{TDB} &= \frac{JD_{Temp} - 2,451,545.0}{36,525} = \frac{2,451,046.5 - 2,451,545.0}{36,525} \\ &= -0.013\ 65, -0.013\ 64, -0.013\ 64: *-0.013\ 65, -0.013\ 63, -0.013\ 63 \end{aligned}$$

Let $T_{TDB} \cong T_{UT1}$

$$\begin{aligned} \lambda_{ecliptic} &= 218.32^\circ + 481,267.8813T_{TDB} + 6.29 \sin(134.9 + 477,198.85T_{TDB}) \\ &\quad - 1.27 \sin(259.2 - 413,335.38T_{TDB}) + 0.66 \sin(235.7 + 890,534.23T_{TDB}) \\ &\quad + 0.21 \sin(269.9 + 954,397.70T_{TDB}) - 0.19 \sin(357.5 + 35,999.05T_{TDB}) \\ &\quad - 0.11 \sin(186.6 + 966,404.05T_{TDB}) \\ &= -225.053\ 53^\circ, -222.666\ 07, -222.650\ 74: *-225.053\ 53, -215.612\ 38, -215.250\ 25 \end{aligned}$$

$$\begin{aligned} \phi_{ecliptic} &= 5.13^\circ \sin(93.3 + 483,202.03T_{TDB}) + 0.28 \sin(228.2 + 960,400.87T_{TDB}) \\ &\quad - 0.28 \sin(318.3 + 6003.18T_{TDB}) - 0.17 \sin(217.6 - 407,332.20T_{TDB}) \\ &= -1.423\ 04^\circ, -1.206\ 23, -1.204\ 84: *-1.423\ 04, -0.556\ 69, -0.523\ 10 \end{aligned}$$

The obliquity of the ecliptic changes too little to notice—it's essentially the same for each iteration.

$$\epsilon = 23.439\ 291^\circ - 0.013\ 004\ 2 T_{TDB} = 23.439\ 468^\circ$$

The direction cosines permit us to find the right ascension and declination.

$$\begin{aligned} l &= \cos(\phi_{ecliptic}) \cos(\lambda_{ecliptic}) \\ &= -0.706\ 228, -0.735\ 153, -0.735\ 335: *-0.706\ 228, -0.812\ 937, -0.816\ 601 \\ m &= \cos(\epsilon) \cos(\phi_{ecliptic}) \sin(\lambda_{ecliptic}) - \sin(\epsilon) \sin(\phi_{ecliptic}) \\ &= 0.659\ 041, 0.630\ 035, 0.629\ 845: *0.659\ 041, 0.538\ 087, 0.533\ 133 \end{aligned}$$

$$\begin{aligned}
n &= \sin(\epsilon) \cos(\phi_{ecliptic}) \sin(\lambda_{ecliptic}) + \cos(\epsilon) \sin(\phi_{ecliptic}) \\
&= 0.258\,664, 0.250\,211, 0.250\,156: *0.258\,664, 0.222\,702, 0.221\,192 \\
\alpha_{\zeta} &= \text{ATAN2}(m, l) = 136.9795^\circ, 139.4030, 139.419: *136.9795, 146.4992, 146.8609 \\
\sin(\delta_{\zeta}) &= n \quad \delta_{\zeta} = 14.9908^\circ, 14.4900, 14.4867: *14.9908, 12.8678, 12.7791
\end{aligned}$$

Next, we find the sidereal time and hour angles. You must initialize values the first time through the calculations.

LSTIME (JD_{Temp}, λ ⇒ LST, GMST)

$$GMST = 329.1154^\circ, 36.6446, 37.0789: *329.1154, 237.3806, 247.7356$$

$$GHA_{\zeta_n} = GMST - \alpha_{\zeta} = 192.136^\circ, -102.759, -102.340: 102.800, 90.881, 100.875$$

$$LHA = GHA_{\zeta_n} + \lambda$$

IF first time through the calculations, initialize

$$\Delta GHA = 347.81^\circ$$

ELSE

$$\Delta GHA = \frac{GHA_{\zeta_n} - GHA_{\zeta}}{\Delta UT}$$

IF $\Delta GHA < 0$

$$\Delta GHA = \Delta GHA + \frac{360^\circ}{\Delta UT}$$

$$\Delta GHA = 347.81^\circ, 348.030, 348.080: *347.81, 348.176, 348.376$$

$$x_n = \frac{0.00233 - \sin(\delta_{\zeta}) \sin(\phi_{gc})}{\cos(\delta_{\zeta}) \cos(\phi_{gc})} = -0.2115, -0.2137, -0.2137: *-0.2215, -0.1886, -0.1872$$

IF $x_n > 1$, next day

IF Moonrise calculation set $LHA_n = 360^\circ - LHA_n$. This is the point where the rise and set calculations will differ.

$$LHA_n = 257.200^\circ, 257.660, 257.663: 102.800, 100.868, 100.789$$

Now we find the times and determine if the iteration is successful.

$$\Delta UT = \frac{LHA_n - LHA}{\Delta GHA}$$

IF $\Delta UT < -0.5$ THEN

$$\Delta UT = \Delta UT + \frac{360^\circ}{\Delta GHA}$$

ELSE IF $\Delta UT > 0.5$ THEN

$$\Delta UT = \Delta UT - \frac{360^\circ}{\Delta GHA}$$

$$\Delta UT = 4.4896^h, 0.0289, 0.0002: 17.8355, 0.6884, -0.0059$$

$$\Delta t = \Delta t + \Delta UT = 4.4896^h, 4.5185, 4.5187: 17.8355, 18.5240, 18.5181$$

$$JD_{temp} = JD_{temp} + \Delta t = 2,451,046.6871, 2,451,046.6883, 2,451,046.6883:$$

$$2,451,047.2432, 2,451,047.2718, 2,451,047.2716$$

```

Set  $GHA_{\mathbb{C}} = GHA_{\mathbb{C}_n}$ 
LOOP END
 $UT_{Moonrise} = \Delta t = 4.518\ 728^h = 4:31$ 
 $UT_{Moonset} = \Delta t = 18.518\ 078^h = 18:31$ 

```

These values are close (4:32 and 18:31) to those found in the Almanac (1998). Note that when developing a general computer routine, there are many instances of the phenomena occurring on the previous, following day, or not occurring at all.

5.2.3 Phases of the Moon

The percentage of the Moon's surface that reflects light to the Earth changes constantly. There are four named ***phases of the Moon*** (new, full, last, and first quarter phases). We define the phase of the Moon using the ecliptic longitudes of the Sun and Moon.

$$phase_{\mathbb{C}} = \lambda_{ecliptic_{\odot}} - \lambda_{ecliptic_{\mathbb{C}}}$$

The corresponding phase values are 0° - new, 90° - first quarter, 180° - full, and 270° - last quarter. The phase is *close* to the angle between the Sun and the Moon. You can approximate the percentage of the Moon's surface that's illuminated as

$$\% \text{ disk} = \frac{100\%}{2}(1 - \cos(phase_{\mathbb{C}}))$$

Although the Earth does occasionally block Solar light to the Moon (a lunar eclipse), the lunar phases are due to the relative position of the Moon with respect to an observer on the Earth. Figure 5-5 shows the geometry.

The phases of the Moon also have additional names. ***Waxing*** and ***waning*** Moons describe if the apparent Moon illumination is increasing or decreasing, respectively. ***Gibbous*** refers to times when more than 50% of the Moon appears illuminated. Each season has 3 full Moons. Occasionally, as the saying goes, a ***blue Moon*** occurs whenever there are 4 full Moons in a season—the third full Moon is the blue Moon.*

5.3 Celestial Applications

For precise interplanetary and even some satellite applications, we require the location of the planets. To determine optimum viewing times for optical observations, we need the illumination, and the visibility of these objects from the Earth. This section explores some basic concepts of ephemerides and illumination.

* This occurrence happens infrequently, and was actually incorrectly interpreted for many years as the second full Moon in a season with four full Moons. The confusion resulted from a 1946 *Sky & Telescope* article which incorrectly interpreted a discussion in the 1937 *Maine Farmers Almanac*.

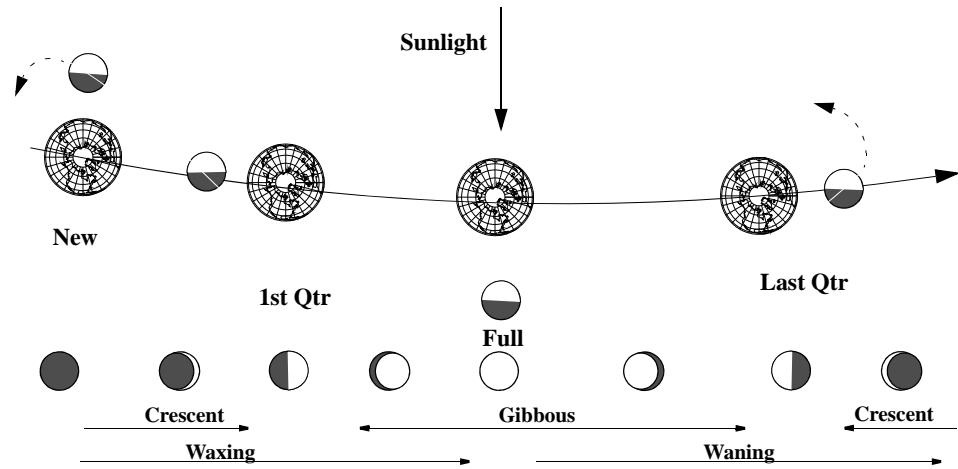


Figure 5-5. Geometry for the Motion of the Moon. This exaggerated figure shows the different phases of the Moon and its appearance from Earth. Over time, the Moon's position changes causing each phenomena. The nomenclature is combined to yield terms as waxing crescent, or waning gibbous to describe intermediate positions.

5.3.1 Application: Planetary Ephemerides

Determining a planet's location combines several topics discussed so far. First, we have to form a position and velocity vector at a given epoch in a given frame (for instance the mean equator and mean equinox of IAU-76/FK5). Often, we do so using polynomials in terms of the classical orbital elements and the elapsed Julian centuries [Eq. (3-42)]. Some programs may also specify mean or osculating elements (I'll discuss these two types of elements further in Sec. 9.2). Finally, we must select an output coordinate system. Because the polynomial expansions are usually given for the heliocentric coordinate system, we must carefully convert to other systems. Table D-3 and Table D-4 (in Appendix D) show orbital element data for the planets in the mean equator and equinox of IAU-76/FK5 and the mean equinox of date frames. 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

$$a = 5.202\,603\,191 + 1.913 \times 10^{-7} T_{TDB}$$

$$\begin{aligned}
e &= 0.048\,494\,85 + 0.000\,163\,244 T_{TDB} - 4.719 \times 10^{-7} T_{TDB}^2 - 1.97 \times 10^{-9} T_{TDB}^3 \\
i &= 1.303\,270^\circ - 0.005\,496\,6 T_{TDB} + 4.65 \times 10^{-6} T_{TDB}^2 - 4 \times 10^{-9} T_{TDB}^3 \\
\Omega &= 100.464\,441^\circ + 1.020\,955\,0 T_{TDB} + 0.000\,401\,17 T_{TDB}^2 + 5.69 \times 10^{-7} T_{TDB}^3 \\
\tilde{\omega} &= 14.331\,309^\circ + 1.612\,666\,8 T_{TDB} + 0.001\,031\,27 T_{TDB}^2 - 4.569 \times 10^{-6} T_{TDB}^3 \\
\lambda_M &= 34.351\,484^\circ + 3036.302\,788\,9 T_{TDB} + 0.000\,223\,74 T_{TDB}^2 + 2.5 \times 10^{-8} T_{TDB}^3
\end{aligned}$$

The orbital elements use alternate elements that refer to the *longitude of perihelion* and the *mean longitude*; thus, some calculations are necessary before finding 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 ω_{true} , and the overall accuracy isn't affected.

$$\begin{aligned}
\omega &= \tilde{\omega} - \Omega \\
M &= \lambda_M - \tilde{\omega}
\end{aligned}$$

Once we determine the classical orbital elements, Algorithm 10 (*COE2RV*) 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-5). To determine 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,

$$\hat{r}_{XYZ(FK5)} = \text{ROT1}[-\epsilon] \hat{r}_{XYZ}$$

We may also need to do some vector translations if we want the vectors relative to the Earth. In addition, the vectors are heliocentric, but reference the mean equator and equinox of IAU-76/FK5. Some applications require the reduction formulas of Algorithm 24 to convert the position and velocity vectors to an Earth-fixed frame (i.e., ITRF). The important point is that the vectors must be in the proper coordinate system for maximum accuracy.

Meeus (1991:205–208) discusses the method of Bretagnon and the “*Variations séculaires des orbites planétaires (VSOP)*” (VSOP-82 and VSOP-87) for determining more precise (osculating or true time-varying) locations of the planets. This solution rests on solving long polynomial equations to determine the equinoctial orbital elements. Consult either Bretagnon (1982), Bretagnon and Francou (1988), or Meeus (1991) for more information on planetary ephemerides.

ALGORITHM 33: *PLANETRV* (*Planet, JD* \Rightarrow $\hat{r}_{XYZ(FK5)}, \hat{v}_{XYZ(FK5)}$)

JULIANTOGREGORIANDATE (*JD* \Rightarrow *Yr, Mo, Day, {H, M, S} UTC*)

$$\begin{aligned} \text{CONVTIME } & (yr, mo, day, UTC, \Delta UTI, \Delta AT \Rightarrow \\ & UTI, TAI, TT, TDB, T_{UTI}, T_{TT}, T_{TDB}) \\ & T_{TT} = \frac{JD_{TT} - 2,451,545.0}{36,525} \end{aligned}$$

$$T_{TDB} \cong T_{TT}$$

Find elements $a, e, i, \Omega, \tilde{\omega}, \lambda_M$ from Table D-3 or Table D-4:

$$\begin{aligned} M &= \lambda_M - \tilde{\omega} \\ \omega &= \tilde{\omega} - \Omega \end{aligned}$$

$$\text{KEPEQTN}E (M, e \Rightarrow \nu)$$

$$p = a(1 - e^2)$$

$$\text{COE2RV } (p, e, i, \Omega, \omega, \nu(u, \lambda_{true}, \tilde{\omega}_{true}) \Rightarrow \vec{r}_{XYZ}, \vec{v}_{XYZ})$$

$$\vec{r}_{XYZ(FK5)}, \vec{v}_{XYZ(FK5)} = \text{ROT}(-\epsilon) \vec{r}_{XYZ}, \vec{v}_{XYZ}$$

This algorithm is particularly complicated, especially if it includes reduction (from Sec. 3.7), so we'll run through an example to show how it works.

▼ Example 5-5. Calculating a Planetary Location.

GIVEN: Table D-3 and Table D-4

FIND: Position and velocity vectors of Jupiter on May 20, 1994 at 20:00 UTC, with respect to the Sun, using the mean equator and equinox of IAU-76/FK5.

Begin by letting $TDB \approx UTC \approx UTI$. Then determine the Julian date. If you use Algorithm 14,

$$JD = 2,449,493.333$$

Use Eq. (3-42) to find the Julian centuries of TDB:

$$T_{TDB} = -0.056\,171\,58$$

From Table D-3 find values for the orbital elements of Jupiter. These values find the heliocentric elements for the mean equator, mean equinox of IAU-76/FK5. These equations differ from the mean of date values listed on page 295.

$$\begin{aligned} a &= 5.202\,603\,191 + 1.913 \times 10^{-7} T_{TDB} = 5.202\,603 \text{ AU} \\ e &= 0.048\,494\,85 + 0.000\,163\,244 T_{TDB} - 4.719 \times 10^{-7} T_{TDB}^2 - 1.97 \times 10^{-9} T_{TDB}^3 \\ &= 0.048\,486 \\ i &= 1.303\,270^\circ - 0.001\,987\,2 T_{TDB} + 3.318 \times 10^{-5} T_{TDB}^2 + 9.2 \times 10^{-8} T_{TDB}^3 \\ &= 1.303\,382^\circ \end{aligned}$$

$$\begin{aligned}
\Omega &= 100.464\ 441^\circ + 0.176\ 682\ 8T_{TDB} + 0.000\ 903\ 87T_{TDB}^2 - 7.032 \times 10^{-6} T_{TDB}^3 \\
&= 100.454\ 519^\circ \\
\tilde{\omega} &= 14.331\ 309^\circ + 0.215\ 552\ 5T_{TDB} + 0.000\ 722\ 52T_{TDB}^2 - 4.59 \times 10^{-6} T_{TDB}^3 \\
&= 14.319\ 203^\circ \\
\lambda_M &= 34.351\ 484^\circ + 3034.905\ 674\ 6T_{TDB} - 0.000\ 085\ 01T_{TDB}^2 + 4.0 \times 10^{-9} T_{TDB}^3 \\
&= -136.123\ 94^\circ
\end{aligned}$$

Convert to classical elements (it's not necessary to make the angles positive):

$$\begin{aligned}
M &= \lambda_M - \tilde{\omega} = -150.443\ 142^\circ \\
\omega &= \tilde{\omega} - \Omega = -86.135\ 316^\circ
\end{aligned}$$

Kepler's equation transforms elements to yield true anomaly:

$$\mathbf{KEPEQTNE}(M, e \Rightarrow \nu), \nu = 206.954\ 53^\circ$$

$$p = a(1 - e^2) = 5.190\ 372 \text{ AU}$$

Use Algorithm 10 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 solar days):

$$\begin{aligned}
\mathbf{COE2RV}(p, e, i, \Omega, \omega, \nu(u, \lambda_{true}, \tilde{\omega}_{true}) \Rightarrow \hat{r}_{XYZ} \hat{v}_{XYZ}) \\
\hat{r}_{XYZ} = -4.075\ 932 \hat{X} - 3.578\ 306 \hat{Y} + 0.105\ 970 \hat{Z} \text{ AU} \\
\hat{v}_{XYZ} = 0.004\ 889 \hat{X} - 0.005\ 320 \hat{Y} - 0.000\ 087 \hat{Z} \text{ AU/TU}
\end{aligned}$$

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 changes the reference to the mean equator and mean equinox of IAU-76/FK5. Use Eq. (3-68) to find ϵ .

$$\epsilon = 23.440\ 021^\circ$$

$$\begin{aligned}
\hat{r}_{XYZ(J2000)}, \hat{v}_{XYZ(J2000)} &= \mathbf{ROT1}(-\epsilon) \hat{r}_{XYZ} \hat{v}_{XYZ} \\
\hat{r}_{XYZ(J2000)} &= -4.075\ 932 \hat{I} - 3.325\ 167 \hat{J} - 1.326\ 185 \hat{K} \text{ AU} \\
&= -609,750,815 \hat{I} - 497,437,968 \hat{J} - 198,394,488 \hat{K} \text{ km} \\
\hat{v}_{XYZ(J2000)} &= 0.004\ 889 \hat{I} - 0.004\ 846 \hat{J} - 0.002\ 197 \hat{K} \text{ AU/TU} \\
&= 0.145\ 626 \hat{I} - 0.144\ 349 \hat{J} - 0.065\ 424 \hat{K} \text{ km/s}
\end{aligned}$$

Remember this is a heliocentric vector from the Sun's center to Jupiter in the IAU-76/FK5 frame (mean equator and equinox). If you need the vector from the Earth to Jupiter, you must subtract the vector from the Sun to the Earth. You can compare the answers with those from DE-245 for the mean equator and mean equinox of IAU-76/FK5 (see Appendix D for information on obtaining these vectors).

$$\hat{r}_{XYZ(J2000)} = -4.068\ 697 \hat{I} - 3.335\ 668 \hat{J} - 1.330\ 683 \hat{K} \text{ AU}$$

$$\blacktriangle \quad \hat{v}_{XYZ(J2000)} = 0.004\ 897 \hat{I} - 0.004\ 840 \hat{J} - 0.002\ 194 \hat{K} \text{ AU/TU}$$

5.3.2 Eclipses

Generally we think of eclipses with the Sun and Moon. However, it's actually a common stellar phenomenon. We'll introduce the basic concept and focus on solar and lunar applications. Although lunar eclipses aren't of utmost interest in studying Earth-orbiting satellites, the concept of eclipses is useful in complex orbit determination systems, and for Lunar observation. Satellites routinely enter eclipse periods during their orbits, and this can cause a great fluctuation in the 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 5-6 shows the basic eclipse geometry.

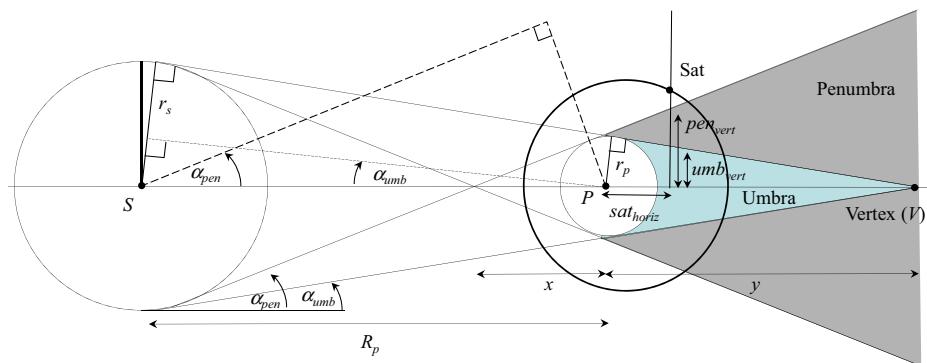


Figure 5-6. Eclipse Geometry. The general geometry for eclipses involves a primary body (Earth) and a secondary body (Sun). Eclipses occur within the umbra and the penumbra regions. Notice the scales are greatly exaggerated; the actual angular departure of the penumbra from the umbra region is actually rather small (See also Fig. 5-2). For the Sun-Earth system, $PV \approx 1.384 \times 10^6$ km, or about four times the distance to the Moon. The angles are useful for quick calculations. The dashed line represents an arbitrary orbit. The vertical and horizontal components are used in determining instantaneous shadow positions. The dashed lines and x and y parameters are useful for developing the equations.

We're interested in two main regions. The **umbra** is totally eclipsed by the primary body (Earth); the **penumbra** is only partially obscured by the primary. For the Earth, you should know the actual angles involved in these calculations because the figures are often exaggerated. The angles for the Sun are found using right triangles as follows (for the Moon, $\alpha_{umb} = 0.265\ 896^\circ$ and $\alpha_{pen} = 0.267\ 227^\circ$):

$$\begin{aligned}\text{SIN}(\alpha_{umb}) &= \frac{r_s - r_p}{R_p} = \frac{696000 - 6378}{149599870} = 0.264\ 121\ 687^\circ \\ \text{SIN}(\alpha_{pen}) &= \frac{r_s + r_p}{R_p} = \frac{696000 + 6378}{149599870} = 0.269\ 007\ 205^\circ\end{aligned}$$

Using this information, a geosynchronous satellite ($r \sim 42163$ km) traverses about 13098 km in the penumbral regions. At its orbital velocity, the satellite spends about 71

minutes in these regions. This doesn't include the time in the umbral region, which is about 12412 km (about 67 minutes).

The similar triangles defined in Fig. 5-6 shows the geometry of the distance from the planet to the vertex. Thus,

$$\frac{PV}{r_p} = \frac{SV}{r_s} \quad PV = \frac{SVr_p}{r_s}$$

We also know the distance between the secondary (Sun) and primary (Earth) bodies:

$$SV = PV + R_p$$

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,

$$PV = \frac{R_p r_p}{r_s - r_p} \quad (5-1)$$

Using values for the Earth and the Sun, we find PV to be 1.384×10^6 km. 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's shadow sometimes completely obscures the Moon, causing a *lunar eclipse*. These eclipses vary based on the specific geometry. The same phenomenon affects satellites. Figure 5-7 shows the types of Lunar eclipses.

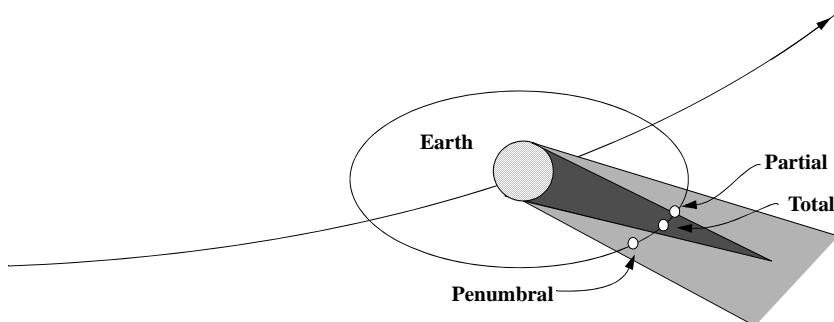


Figure 5-7. Lunar Eclipses. There are three types of lunar eclipses. The figure is greatly exaggerated to illustrate each region. For a *total eclipse*, the Moon is completely within the umbral region of the Earth's shadow. A *partial eclipse* has the Moon partially covered by the penumbra and umbra regions. *Penumbral eclipses* occur when 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. (5-1) 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's surface with its shadow, it causes a *solar eclipse*. The closeness in values is the reason solar eclipses are less common. There are certain geometries of the Moon and the Earth that permit the shadow of the Moon to lie on the Earth's surface. The Moon can also shadow a satellite, but this is much less common than lunar eclipses. Solar eclipses take three forms, shown in Fig 5-8.

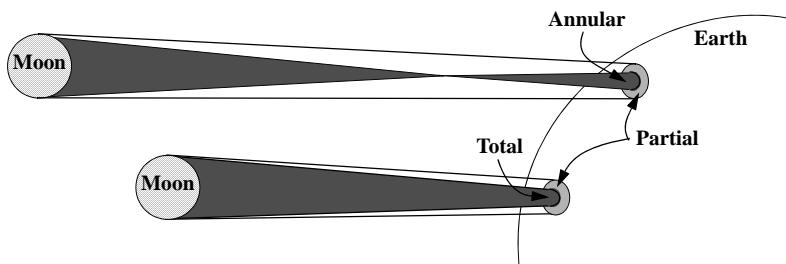


Figure 5-8. Solar Eclipses. Three types of solar eclipses result from the Moon blocking the light that travels to the Earth. The figure is greatly exaggerated to illustrate each type. For a *total eclipse*, the Moon's umbral shadow falls directly on the Earth. An *annular eclipse* occurs when the Moon's umbral cone doesn't fall directly on the Earth, but the axis of the cone intersects the Earth. A *partial eclipse* happens when only the Moon's penumbra region is involved in the shadowing.

Geometrical Shadow Analysis

We can develop a simple geometrical method (Figure 5-6) to determine if the satellite is in the umbral or penumbral regions. We use the angles developed earlier and simply examine the satellite's vertical and horizontal distances from the Earth-Sun line with the penumbra and umbra distances. Let y be the Earth-vertex distance and x be the Earth primary/secondary point, as shown in Fig. 5-6.

ALGORITHM 34: SHADOW ($\vec{r}_\odot, \vec{r} \Rightarrow$ SHADOW)

$$\text{SHADOW} = \text{'NONE'}, \alpha_{umb} = 0.264\ 121\ 687^\circ, \alpha_{pen} = 0.269\ 007\ 205^\circ$$

IF $\vec{r}_\odot \cdot \vec{r} < 0$

ANGLE $(-\vec{r}_\odot, \vec{r}, \zeta)$

$$\begin{aligned}
 SAT_{horiz} &= |\dot{\mathbf{r}}| \cos(\zeta) \\
 SAT_{vert} &= |\dot{\mathbf{r}}| \sin(\zeta) \\
 x &= \frac{r_p}{\sin(\alpha_{pen})} \\
 PEN_{vert} &= \tan(\alpha_{pen})(x + SAT_{horiz}) \\
 \text{IF } SAT_{vert} \leq PEN_{vert} \\
 &\quad \text{SHADOW} = \text{'PENUMBRA'} \\
 &\quad y = \frac{r_p}{\sin(\alpha_{umb})} \\
 &\quad UMB_{vert} = \tan(\alpha_{umb})(y - SAT_{horiz}) \\
 &\quad \text{IF } SAT_{vert} \leq UMB_{vert} \\
 &\quad \quad \text{SHADOW} = \text{'UMBRA'}
 \end{aligned}$$

Traditional Shadow Analysis

There is a more “traditional” approach that uses the classical orbital elements to find the entry and exit times of the satellite. In fact, most satellites (including geosynchronous) experience periodic eclipses. Of course when the satellite is eclipsed, it’s not exposed to solar-radiation pressure—affecting the satellite dynamics. When we need high 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), Long et al. (1978:3-28–3-29), and Neta and Vallado (1998) show the required formulas.

Begin by examining the geometry of the problem. Figure 5-9 shows the geometry for the Earth shadow. 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 product of the Sun vector and the satellite vector determines the angular separation at any time:

$$\cos(\zeta) = \frac{\dot{\mathbf{r}}_\odot \cdot \dot{\mathbf{r}}_{sat}}{r_\odot r_{sat}} \quad (5-2)$$

Ultimately, we need to find the time at which the satellite enters and exits the Earth’s shadow. From Chap. 2, this solution involves obtaining the true anomaly at each time. Unfortunately, Eq. (5-2) 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 ν). But recall the development of Eq. (2-58) for the f and g functions:

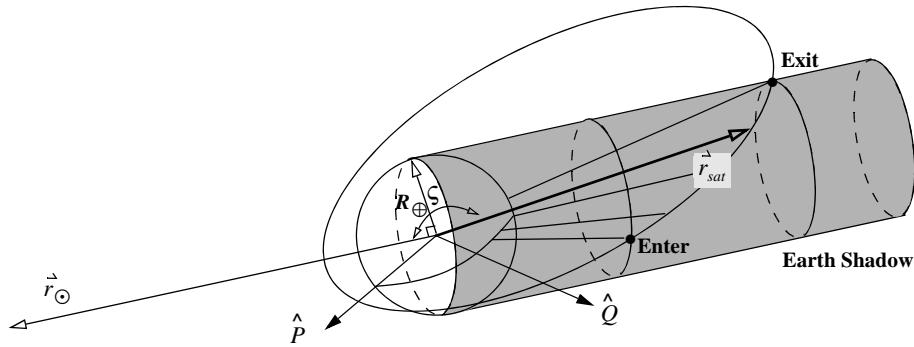


Figure 5-9. Entry-Exit Geometry for Solar Radiation. The geometry of the entry and exit points for a satellite depends on the position of the Sun and the satellite. Here we assume that the shadow is a cylindrical cone beginning at a point on the Earth that's perpendicular to the Sun, $\zeta = 90^\circ$.

$$\vec{r}_{sat} = x_w \hat{P} + y_w \hat{Q} = r_{sat} \cos(\nu) \hat{P} + r_{sat} \sin(\nu) \hat{Q}$$

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. (5-2) to include the true anomaly:

$$\cos(\zeta) = \frac{r_{sat} \cos(\nu) \hat{r}_\odot \cdot \hat{P} + r_{sat} \sin(\nu) \hat{r}_\odot \cdot \hat{Q}}{r_\odot r_{sat}}$$

Now let's define temporary parameters to simplify the 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. (5-2) as

$$(\cos(\zeta) = \beta_1 \cos(\nu) + \beta_2 \sin(\nu)) \quad (5-3)$$

Eq. (5-2) and Eq. (5-3) represent two approaches to determining the angular separation between the Sun and the satellite. If we square Eq. (5-3), insert into $r_{sat}^2 \sin^2(\zeta) < R_\oplus^2$, and use the trajectory equation [Eq. (1-24)] for r_{sat} we find the *shadow function*,

$$S = R_\oplus^2 (1 + e \cos(\nu))^2 + p^2 \{ \beta_1 \cos(\nu) + \beta_2 \sin(\nu) \}^2 - p^2 \quad (5-4)$$

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-degree equation in the cosine of the true anomaly (Escobal [1965] 1985:158). Using $\alpha = R_{\oplus}/p$

$$\begin{aligned}\alpha_1 &= \alpha^4 e^4 - 2\alpha^2(\beta_2^2 - \beta_1^2)e^2 + (\beta_1^2 + \beta_2^2)^2 \\ \alpha_2 &= \alpha^4 4e^3 - 4\alpha^2(\beta_2^2 - \beta_1^2)e \\ \alpha_3 &= \alpha^4 6e^2 - 2\alpha^2(\beta_2^2 - \beta_1^2) - 2\alpha^2(1 - \beta_2^2)e^2 + 2(\beta_2^2 - \beta_1^2)(1 - \beta_2^2) - 4\beta_2^2\beta_1^2 \\ \alpha_4 &= \alpha^4 4e - 4\alpha^2(1 - \beta_2^2)e \\ \alpha_5 &= \alpha^4 - 2\alpha^2(1 - \beta_2^2) + (1 - \beta_2^2)^2\end{aligned}\quad (5-5)$$

$$S = \alpha_1 \cos^4(\nu) + \alpha_2 \cos^3(\nu) + \alpha_3 \cos^2(\nu) + \alpha_4 \cos(\nu) + \alpha_5$$

Solution of the quartic equation 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. (5-3) must be less than zero. We can determine if it's an entry or an exit using Eq. (5-4). For example, if the shadow function, S , is zero, it's an entry or an exit (S going positive, S going negative, respectively). 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. Because the input to this process is a time dependent vector, the answer changes continually. Thus, the method is limited to an approximate solution for the true anomaly at entry and exit of the shadow. This initial guess is often refined with iterative techniques.

A drawback of this technique is that as the satellite position vector changes, the answers from the quartic also change. This requires additional analysis to determine the proper entry or exit condition for your application.

The cylindrical calculations are often adequate, but it's necessary to have a technique that allows us to find the precise penumbral and umbral regions. Umbra-penumbra corrections require us to change the angular-separation angle, ζ . Now we use Newton's method and iterate 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.

We can solve the case for an ellipsoidal Earth in various ways. First, we can determine the sublatitude point at each time using Algorithm 12, 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. (5-4) with this value.

Numerical Shadow Analysis

Eq. (5-4) can also be solved numerically. If we let $x = \cos(\nu)$ and set $S = 0$ for our solution, Eq. (5-4) becomes

$$Ax^2 + Bx + Cx\sqrt{1-x^2} + D = 0$$

We could use a Newton-Raphson technique or Halley's method (Neta and Vallado 1998) to solve for x . Halley's method is

$$x_{n+1} = x_n + \delta_n = x_n - \frac{f(x_n)}{f'(x_n) - \frac{f(x_n)f''(x_n)}{2f'(x_n)}}$$

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 the one root closest to the initial condition.

Beta Angle

The **beta angle**, β_{Sun} , is the angle between the satellite's orbital plane and the Sun. It's useful in determining eclipse times. Ross (1999) gives the formula.

$$\begin{aligned} \sin(\beta_{Sun}) &= \sin(\epsilon)\cos(i)\sin(\lambda_{ecliptic}) - \cos(\epsilon)\sin(i)\cos(\Omega)\sin(\lambda_{ecliptic}) \\ &\quad + \sin(i)\sin(\Omega)\cos(\lambda_{ecliptic}) \end{aligned}$$

Notice that the angular separation of the Earth from the vernal equinox ($\lambda_{ecliptic}$) is required. Another expression uses right ascension and declination values for the Sun.

$$\sin(\beta_{Sun}) = \cos(\delta_{\odot})\sin(i)\sin(\Omega - \alpha_{\odot}) + \sin(\delta_{\odot})\cos(i)$$

Eclipse times occur when $|\sin(\beta_{Sun})| < R_{\oplus}/r$. The duration of an eclipse for a circular orbit, $\Delta\tau_{eclipse-circ}$, is found using the satellite position magnitude, r , and the satellite period, β .

$$\Delta\tau_{eclipse-circ} = \cos^{-1}\left(\frac{\sqrt{1 - \left(\frac{R_{\oplus}}{r}\right)^2}}{\cos(\beta_{Sun})}\right)\frac{\beta}{\pi}$$

5.3.3 Application: Sight and Light

The primary problem to solve for **SIGHT** and **LIGHT** is whether or not a *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 satel-

lite 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 you know two arbitrary vectors (\vec{a} and \vec{b}). Using the dot product, the cosine of the angle between the two vectors is:

$$\cos(\theta) = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| |\vec{b}|}$$

If you now fix the maximum perpendicular distance between the central point and the line-of-sight vector to R_{\oplus} , you can determine the two half-angles shown in Fig. 5-10. The two angles for the constrained problem are

$$\cos(\theta_1) = \frac{R_{\oplus}}{|\vec{a}|} \quad \cos(\theta_2) = \frac{R_{\oplus}}{|\vec{b}|}$$

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 divisions, square roots, and trigonometric functions.

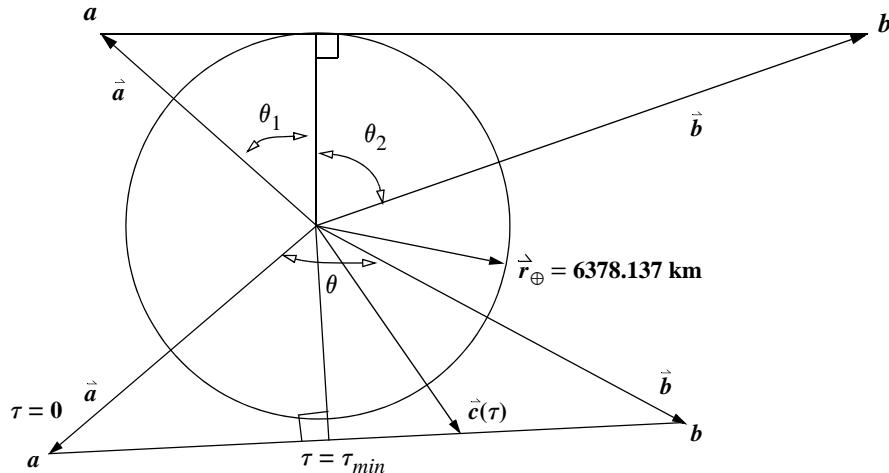


Figure 5-10. Line-of-Sight Geometry. LOS exists between the two positions if the sum of the two angles ($\theta_1 + \theta_2$) is larger than the calculated value for θ . We can also use a parametric approach in which we find the location for the minimum distance to the Earth. Each vector can be represented as a function of τ .

An alternate method by Alfano (1991) avoids the trigonometric operations but requires a little more development. Figure 5-10 shows the situation. Start with a parametric representation of a line between the two position vectors:

$$\hat{c}(\tau) = \hat{a} + (\hat{b} - \hat{a})\tau \quad (5-6)$$

where τ varies from 0 (\hat{a}) to 1.0 (\hat{b}). We want to determine the value of τ which minimizes the distance to the central body. Because we're really interested only in the magnitude, one could square $\hat{c}(\tau)$ 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 $\hat{a} = a_1\hat{I} + a_2\hat{J} + a_3\hat{K}$ and $\hat{b} = b_1\hat{I} + b_2\hat{J} + b_3\hat{K}$).

$$|\hat{c}(\tau)|^2 = (a_1 + (b_1 - a_1)\tau)^2 + (a_2 + (b_2 - a_2)\tau)^2 + (a_3 + (b_3 - a_3)\tau)^2$$

Take the derivative with respect to τ and set it equal to zero to minimize c as a function of τ . Remember to write each component separately to do the required differentiation.

$$\begin{aligned} \frac{d|\hat{c}(\tau)|^2}{d\tau} = & 2[(a_1 + (b_1 - a_1)\tau)(b_1 - a_1) + (a_2 + (b_2 - a_2)\tau)(b_2 - a_2) \\ & + (a_3 + (b_3 - a_3)\tau)(b_3 - a_3)] \end{aligned}$$

Equate this expression to zero and solve for the value of τ that minimizes c , 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:

$$\tau_{min} = \frac{|\hat{a}|^2 - \hat{a} \cdot \hat{b}}{|\hat{a}|^2 + |\hat{b}|^2 - 2 \hat{a} \cdot \hat{b}} = \frac{\hat{a} \cdot (\hat{a} - \hat{b})}{|\hat{a} - \hat{b}|^2} \quad (5-7)$$

Squaring the norm of Eq. (5-6) yields

$$|\hat{c}(\tau_{min})|^2 = |\hat{a}|^2 + |\hat{b} - \hat{a}|^2 \tau_{min}^2 + 2(\hat{b} - \hat{a}) \cdot \hat{a} \tau_{min}$$

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} from Eq. (5-7) for τ_{min}^2 :

$$|\hat{c}(\tau_{min})|^2 = |\hat{a}|^2(1 - 2\tau_{min}) + 2(\hat{a} \cdot \hat{b})\tau_{min} + \frac{(|\hat{a}|^2 - \hat{a} \cdot \hat{b})^2}{|\hat{a}|^2 + |\hat{b}|^2 - 2 \hat{a} \cdot \hat{b}}$$

Notice that τ_{min} is included 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

$$|\hat{c}(\tau_{min})|^2 = (1 - \tau_{min})|\hat{a}|^2 + (\hat{a} \cdot \hat{b})\tau_{min} \quad (5-8)$$

To use this procedure, first find τ_{min} . If $\tau_{min} < 0.0$ or $\tau_{min} > 1.0$, there is line-of-sight between the two vectors. From Fig. 5-10, 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. (5-8) to determine the square of the magnitude at the minimum location. If this squared magnitude is larger than (or equal) to R_{\oplus}^2 , 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 polar radius is almost 20 km larger than spherical. The solution is solved by scaling the K component of the position vectors using Eq. (1-5), and using the **SIGHT** algorithm exactly as before.*

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 the *Sun* rather than another satellite. It also assumes the light from the Sun acts as a point source. Be careful of the units involved—some routines give the Sun position vector in AUs, whereas the satellite is in km. The process works well for most cases. If you need extreme precision—including aberration, distortion, etc.—a more detailed process must be developed.

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 35: *SIGHT* (\hat{r}_1, \hat{r}_2 in IJK \Rightarrow LOS)

$$\tau_{min} = \frac{|\hat{r}_1|^2 - \hat{r}_1 \cdot \hat{r}_2}{|\hat{r}_1|^2 + |\hat{r}_2|^2 - 2 \hat{r}_1 \cdot \hat{r}_2} \quad \text{LOS = FALSE}$$

IF ($\tau_{min} < 0$) or ($\tau_{min} > 1$)
 LOS = TRUE
 ELSE
 IF $|\hat{c}(\tau_{min})|^2 = (1 - \tau_{min})|\hat{r}_1|^2 + (\hat{r}_1 \cdot \hat{r}_2)\tau_{min} \geq 1.0 R_{\oplus}^2$
 LOS = TRUE

* For computer software, you may want to do the scaling transformation within a separate variable because this will help remind the programmer not to pass the scaled component back to the main program. You should also check that both vectors are above the surface of the Earth. Situations where one vector is beneath the Earth's surface and the other vector is above that point can give incorrect results.

Let's work an example to test the ***SIGHT*** and ***LIGHT*** routines.

▼ **Example 5-6. Solving for *SIGHT* and *LIGHT* Conditions.**

GIVEN: $\dot{\vec{r}}_1 = -4464.696\hat{I} - 5102.509\hat{K}$ km, $\dot{\vec{r}}_2 = 5740.323\hat{J} + 3189.068\hat{K}$ km on February 15, 1995, 12:00 UTC.

FIND: LOS and whether satellite 1 is in the sunlight.

First, find the minimum parametric value:

$$\tau_{min} = \frac{|\dot{\vec{r}}_1|^2 - \dot{\vec{r}}_1 \cdot \dot{\vec{r}}_2}{|\dot{\vec{r}}_1|^2 + |\dot{\vec{r}}_2|^2 - 2 \dot{\vec{r}}_1 \cdot \dot{\vec{r}}_2} = 0.508\,235\,3$$

Now find the value of the function using this minimum value:

$$|\dot{\vec{c}}(\tau_{min})|^2 = \frac{(1 - \tau_{min})|\dot{\vec{r}}_1|^2 + (\dot{\vec{r}}_1 \cdot \dot{\vec{r}}_2)\tau_{min}}{R_\oplus^2} = 0.032\,211\,8$$

Because the value is less than 1.0, we have no LOS. Watch the units.

Next, determine the position vector for the Sun. Using Algorithm 29, find the Sun's position vector. For February 15, 1995, 0^h UTC, $JD = 2,449,763.5$. With the approximation that $JD_{UTC} \approx JD_{UT1}$,

$$\dot{\vec{r}}_\odot = 0.817\,078\,3\hat{I} - 0.509\,036\,0\hat{J} - 0.220\,700\,8\hat{K}$$
 AU

Be sure to convert to consistent (km in this case) units:

$$\dot{\vec{r}}_\odot = 122,233,179\hat{I} - 76,150,708\hat{J} - 33,016,374\hat{K}$$
 km

Apply the ***SIGHT*** algorithm to produce the minimum parametric value:

$$\tau_{min} = \frac{|\dot{\vec{r}}_1|^2 - \dot{\vec{r}}_1 \cdot \dot{\vec{r}}_2}{|\dot{\vec{r}}_1|^2 + |\dot{\vec{r}}_2|^2 - 2 \dot{\vec{r}}_1 \cdot \dot{\vec{r}}_2} = -2.3291 \times 10^{-5}$$

▲ Because the value is less than 0.0, we do have LOS and satellite 1 is illuminated.

5.3.4 Ground Illumination

Nighttime optical measurements of satellites illuminated by the Sun are an important source of observations for orbit determination. To facilitate planning for these measurements, it's useful to determine the ambient light that will fall on the sensor. Too much incoming light can inhibit a sensor from discriminating the object of interest. In general, we can use logarithmic approximations to find the ***average illumination*** (M) on a plate perpendicular to the local zenith. Seidelmann (1992:493) shows the equations for both the Sun and Moon. I've combined the relations here. The correction terms for phase (L_2) and parallax (L_3) are only applied for the Moon.

$$x = \frac{el}{90^\circ}$$

$$\begin{aligned}
 L_1 &= l_0 + l_1 x + l_2 x^2 + l_3 x^3 + L_2 + L_3 \\
 L_2 &= -8.68^\circ \times 10^{-3} f - 2.2 \times 10^{-9} f^4 \\
 L_3 &= 2 \text{LOG}_{10} \left(\frac{\varphi}{0.951^\circ} \right) \\
 f &= 180^\circ - E_\odot \\
 E &= \text{COS}^{-1} (\text{SIN}(\delta_\odot) \text{SIN}(\delta_\odot) + \text{COS}(\delta_\odot) \text{COS}(\delta_\odot) \text{COS}(\alpha_\odot - \alpha_\odot)) \\
 \text{LOG}_{10} M &= L_1 + L_2 + L_3
 \end{aligned} \tag{5-9}$$

There are several factors that affect the illumination on the Earth's surface. First, the light coming from the source can change. For the Sun, the emitted light is essentially constant. We saw in Sec 5.2.3 that the percentage of the Moon's surface that's illuminated changes over time. The phase of the Moon, or any other celestial body, is the primary factor determining ground illumination. The elevation is the next largest factor in illumination (L_1). At low elevations, the effective illumination is less than when the object is directly overhead because of the longer path through the atmosphere. Finally, parallax causes a very small change in the light reaching various locations on the Earth. We can describe these effects with the mathematical equations in Eq. (5-9).

These relations require input parameters that are based on the elevation of the Sun or Moon. Table 5-1 shows the coefficients to find the illumination values.

Figure 5-11 shows a plot of ground illumination for the Sun and Moon. Notice how quickly the illumination rises for the Sun. Also notice the large variability in the Moon's illumination, depending on phase and elevation. Finally, the Sun's twilight conditions roughly correspond to an order of magnitude decrease in the illumination ($el = -6, -12,$ and -18°). For each 6° change in elevation, we encounter a new twilight condition.

TABLE 5-1. Index Values for the Sun and Moon. The indices required for ground illumination are shown. Values for the Moon are primarily for positive elevations because the light isn't sufficient to extend to negative elevations.

Elevation Range ($^\circ$)		l_0	l_1	l_2	l_3
Sun	20 to 90	3.74	3.97	-4.07	1.47
	5 to 20	3.05	13.28	-45.98	64.33
	-0.8 to 5	2.88	22.26	-207.64	1034.30
	-5 to -0.8	2.88	21.81	-258.11	-858.36
	-12 to -5	2.70	12.17	-431.69	-1899.83
	-18 to -12	13.84	262.72	1447.42	2797.93
Moon	20 to 90	-1.95	4.06	-4.24	1.56
	5 to 20	-2.58	12.58	-42.58	59.06
	-0.8 to 5	-2.79	24.27	-252.95	1321.29

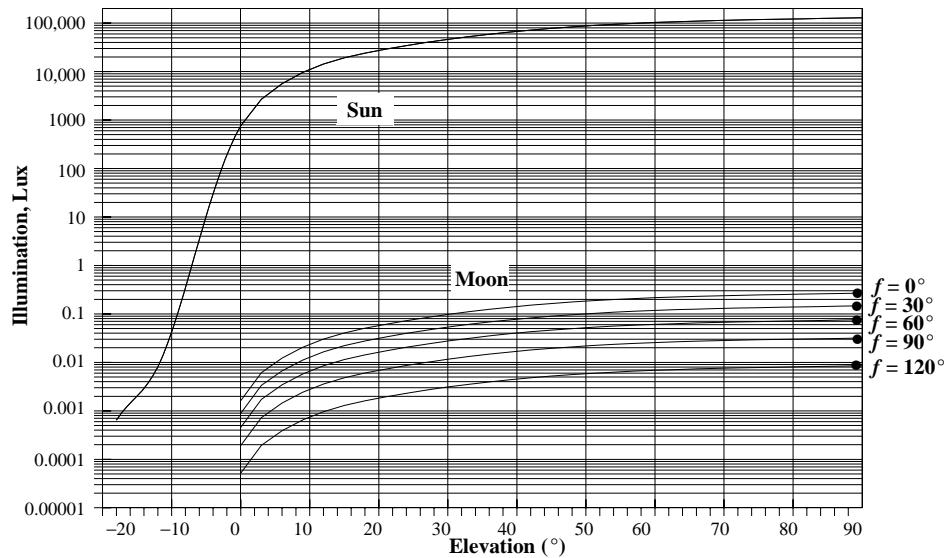


Figure 5-11. Ground Illumination. I've shown illumination (lux^{*}) values for the Sun and Moon. The Sun is very bright and is rarely obscured so that distinctions do not occur. The Moon's phases (f angles) contribute to the differences in Lunar illumination.

*Luminous emittance (or Lux = lumen / m²) is a measure of the visible light emitted

5.3.5 Miscellaneous Phenomena

So far, we've modeled observation relationships using simple geometry. Such models are convenient but too simplistic for many practical applications. This section introduces several observational effects that we can apply as additive corrections to the basic geometric relationships. Many of these corrections apply to stellar observations, and some are applicable for satellite observations. See Sec 4.2.1 for additional factors affecting the observations. Consult Seidelmann (1992:21–24, 121–139, and 721, 730) for more information.

Aberration

Aberration is the apparent angular displacement of a body 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 the correction for **light time**, because the path is no longer an ideal straight line between the points. Corrections for light time are used by all 2-way ranging systems, including SLR measurements. 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 corrections for both light time and stellar aberration.

Parallax

Parallax is another concept used in the study of Earth satellites. As the clocks and measurements of the Observatories became more accurate, observations at different times began to disagree. Frederick Bessel (1748–1846) stated one reason for the differences: parallax. **Parallax** is the angular difference in the apparent direction of an object when viewed from two different locations. Parallax wasn'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 $0.002\ 44^\circ$. The measurement 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 by classical observations.*

There are three main types of parallax. **Geocentric parallax**, φ_{geo} , also called **diurnal parallax**, is the apparent angular difference between observations from the center of the Earth and those from an Earth-fixed site (topocentric observation). When the object lies in the site's horizontal plane, we use the term **horizontal parallax**, φ_{horiz} . **Heliocentric parallax**, φ_{helio} , also called **annual parallax**, is the apparent angular difference between observations taken from the centers of the Sun and the Earth. For instance, we find φ_{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. 5-12):

$$\text{SIN}(\varphi) \approx \frac{a}{b}$$

Because the distance to the object, r , is usually much larger than the central axis, a , we use

* The discovery of parallax could have proved quite helpful to Copernicus and Galileo because it proves the Earth revolves around the Sun. Observers weren't able to accurately measure stellar 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 some modern machinery. Typical example values are $0.002\ 44^\circ$ ($8.8''$) for the Sun, 0.9507° ($57'$) for the Moon and $0.000\ 28^\circ$ ($1''$) for Alpha Centauri, so Copernicus shouldn't feel too bad!

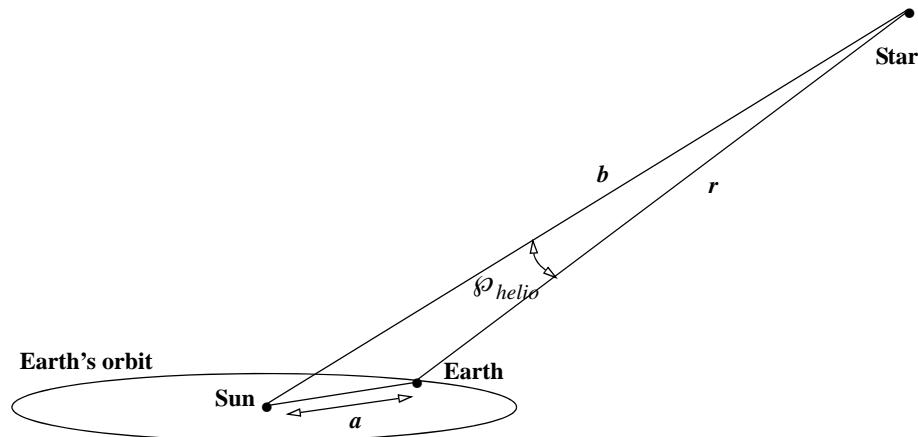


Figure 5-12. Parallax Geometry. Notice the greatly exaggerated scale to show the parallax. Depending on the accuracy, even from different points in the Earth's orbit, stars can appear to be at virtually the same location. The heliocentric parallax angle shown here illustrates the apparent difference for observations from the Sun and Earth.

$$\sin(\phi) \equiv \frac{a}{r} \quad (5-10)$$

For the Earth, the central axis, a , is the distance between the viewing location and the center of the Sun, and r is the distance to the star being observed. Of course, we can't take observations from the Earth's center or the Sun. However, the difference between two observations of a star taken 12 hours apart is twice the geocentric parallax. Similarly, the difference between observations taken six months apart is twice the annual parallax.

An example will show parallax calculations.

▼ Example 5-7. Determining Parallax.

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

FIND: ϕ_{helio} and ϕ_{geo}

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

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

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

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

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

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

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

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

Examples of parallax in astrodynamics appear in calculating the position of the Moon and planets, and in determining angular accuracies for sensor systems.

Relativistic Corrections, Deflection of Light

According to General Relativity, light travels a curved path in space as it travels near massive objects. This is a factor in position measurements involving distant galaxies and quasars because their light often passes near galaxies and massive objects which deflect it (we used these types of observations to realize the FK5 system). For satellite applications, angular deflection usually isn't important. However, the relativistic delay due to the Earth's gravity field is approximately 1 cm for ***Satellite Laser Ranging (SLR)*** and is therefore taken into account. The range delay is also used for interplanetary missions.

Proper Motion and Radial Velocity

Proper motion is defined as “total angular displacement in one year on the celestial sphere referred to a fixed equator and equinox” (Green, 1988:260). Although stars are distant from the Earth, they sometimes exhibit some relative motion. 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 proper motion with parallax corrections to accurately transform standard reference frames when using certain stars. Proper motion 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 (i.e., ICRF). Unfortunately, their extreme distances and extended structures make observing them a challenge.

The relative velocity of an object with respect to an observer is similar to proper motion, but the reference to an observer means that differences like geocentric and topocentric become important. It consists of two parts; one in the radial direction ***radial velocity***, and one in the transverse direction. We can't directly measure the transverse velocity without knowledge of the object's distance. We measure the radial velocity from Doppler measurements of the red shift in the object's light.

Conjunctions and Oppositions

When celestial objects have the same right ascension, they are said to be in conjunction or opposition. The Sun provides the direction for these two terms. For outer planets that lie 180° away from the Sun as viewed from the Earth, the phenomenon is called an ***opposition***, while all other occurrences are termed conjunctions. Outer planets have conjunctions when they are “behind” the Sun as viewed from the Earth. ***Inferior con-***

junctions occur when an inferior planet (Mercury or Venus are the inferior planets) is between the Earth and Sun. A **superior conjunction** occurs when an inferior planet is on the other side of the Sun.

Occultations and Transits

These two phenomena are closely related to eclipses. **Occultations** occur when a larger celestial object passes in front of a smaller object. **Transits** occur when a smaller object passes in front of a larger one. For example, some scientists thought that transits of inferior planets passing in front of the Sun were causing the occasional dark observations on the Sun. It was Galileo that correctly identified that these were actually sunspots. Consult Seidelmann (1992:494-498) for additional information on transits and occultations.

Semi-diameter

The semi-diameter is half the angular extent of an object. It's often used in calculations to determine if the field-of-view of a sensor will come near an object (Sun, Moon, Earth). Don't confuse this with parallax. The semi-diameter is measured from a single location. Consider Fig. 5-13.

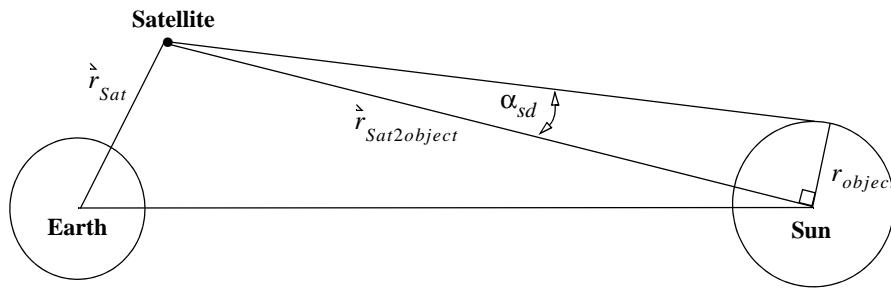


Figure 5-13. Semi-diameter. The semi-diameter is the angular extent of an object viewed from a single location. Notice that r_{object} is an approximate value. This figure is greatly exaggerated.

We can develop a simple expression for the semi-diameter

$$\tan(\alpha_{sd}) = \frac{r_{object}}{|\vec{r}_{Sat2object}|} \quad (5-11)$$

Typical values for the Sun are 0.2666° or about $16'$, and for the Moon, 0.259° , or about $15' 54''$.

Problems

1. On April 15, 2006, 10:00 UTC, is the Moon illuminated by the Sun? Explain your answer.
2. What is the UT of sunrise/sunset for a site at 40° N on June 15, 2006? 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.
3. What is the UT of moonrise/moonset for a site at 40° N on August 21, ?

CHAPTER 6 ORBITAL MANEUVERING

- 6.1 Historical Background
- 6.2 Introduction
- 6.3 Coplanar Maneuvers
- 6.4 Noncoplanar Transfers
- 6.5 Combined Maneuvers
- 6.6 Circular Rendezvous
- 6.7 Continuous-Thrust Transfers
- 6.8 Relative Motion

6.1 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-upsmanship” 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 used 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 man 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.

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 that attempts to use space for operations requires effective maneuvering to ensure a successful mission.

6.2 Introduction

Two-body analysis of orbital changes usually involves solving the Lambert problem, which examines transfer between position vectors (Sec. 7.6). 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 right ascension of the ascending node, Ω . We assume all instantaneous 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 changes in velocity we must apply are lettered as a , b , c , ... to match the positions in the orbit. Figure 6-1 shows these notations. An example of the nomenclature results in the following definitions for the change in velocity at the second maneuver.

$$\Delta v_b = v_{final} - v_{trans_b}$$

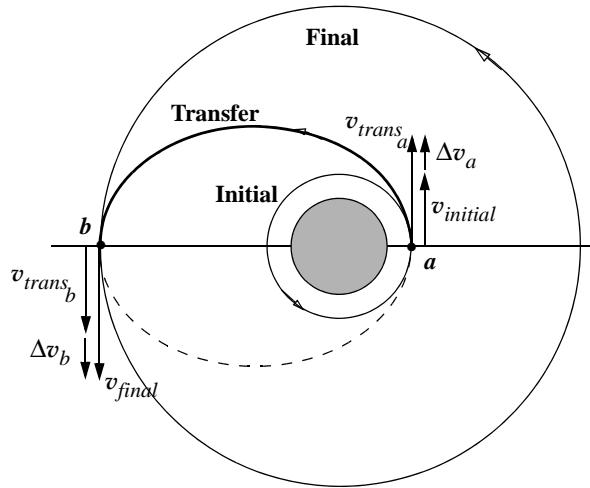


Figure 6-1. Orbit Notation. The letters a , b , ... represent changes in velocities and locations of the firings. When we use multiple transfer orbits, numbers signify which transfer we're considering.

* 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. 9. 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.

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.

6.3 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 tangent 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 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. 6-2. 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:

$$\Delta v_a = v_{final} - v_{initial} \quad (6-1)$$

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. 6-2 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. Mission planning and operations usually determine the specific direction.

We can find v_{final} and $v_{initial}$ from the velocity equations [Eq. (1-30), Eq. (1-31), Eq. (1-32) and Eq. (1-33)]:

$$\begin{aligned} v &= \sqrt{2\left(\frac{\mu}{r} + \xi\right)} & v &= \sqrt{\frac{2\mu}{r} - \frac{\mu}{a}} \\ v &= \sqrt{\frac{\mu}{r}\left(2 - \frac{1 - e^2}{1 + e \cos(\nu)}\right)} & v_{circle} &= \sqrt{\frac{\mu}{r}} \end{aligned}$$

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.

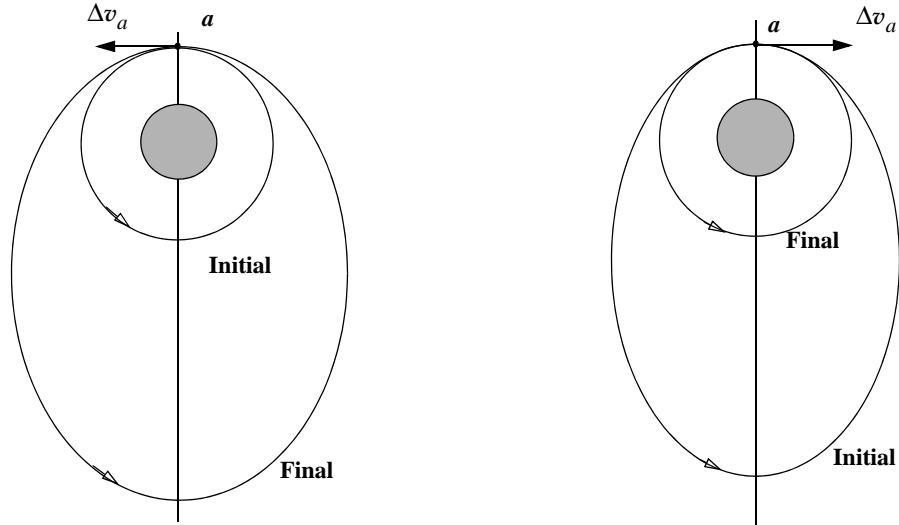


Figure 6-2. Tangential Orbit Transfer ($\phi_{fpa} = 0^\circ$). Tangential burns are defined to occur with a zero flight-path angle. Velocity increases enlarge the semimajor axis (left), whereas decreases cause the orbit to come closer to the Earth (right).

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. 6-3), although the orbits are still coplanar. From Fig. 6-3, 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. (2-94). 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. (1-30), Eq. (1-31), Eq. (1-32), or Eq. (1-33). Make sure you use the proper orbital elements for each case.

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-21). For this case, you know the final velocity from parameters of the desired final orbit, and the magnitude and vector relations become

$$\begin{aligned}\Delta v_b &= \sqrt{v_{transb}^2 + v_{final}^2 - 2v_{transb}v_{final}\cos(\phi_{transb} - \phi_{final})} \\ \hat{\Delta v}_b &= \hat{v}_{final} - \hat{v}_{transb}\end{aligned}\quad (6-2)$$

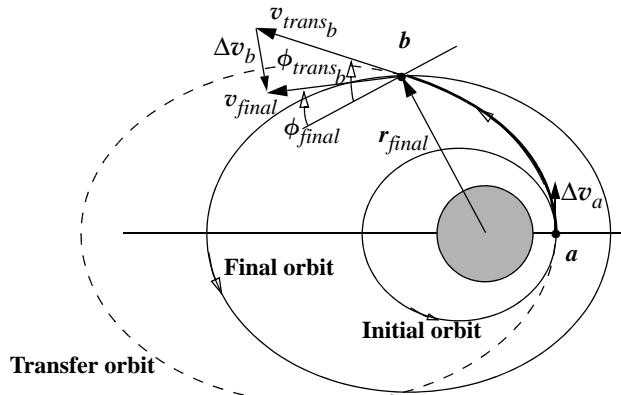


Figure 6-3. Nontangential Orbit Transfer ($\phi_{fpa} \neq 0$). The nontangential transfer, point b , has nonparallel velocity vectors because both orbits have flight-path angles greater than zero.

If the two orbits are tangent, you should be able to reduce this problem to the tangential case of Eq. (6-1). 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. 2.2.6, 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.

6.3.1 Hohmann and Bi-elliptic Transfers

Walter Hohmann (1880–1945) proposed a theory which suggested the minimum change in velocity transfer could be achieved between orbits by using two tangential burns (Hohmann, 1925).^{*} 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 6-4 shows Hohmann transfers from circular and elliptical initial orbits.

* Although Hohmann's original work considered only transfer between circular orbits, other authors have explored transfers between coaxially aligned elliptical orbits and concluded the transfer change in velocity was lowest using two tangential burns (Lawden, 1952, and Thompson, 1986:70).

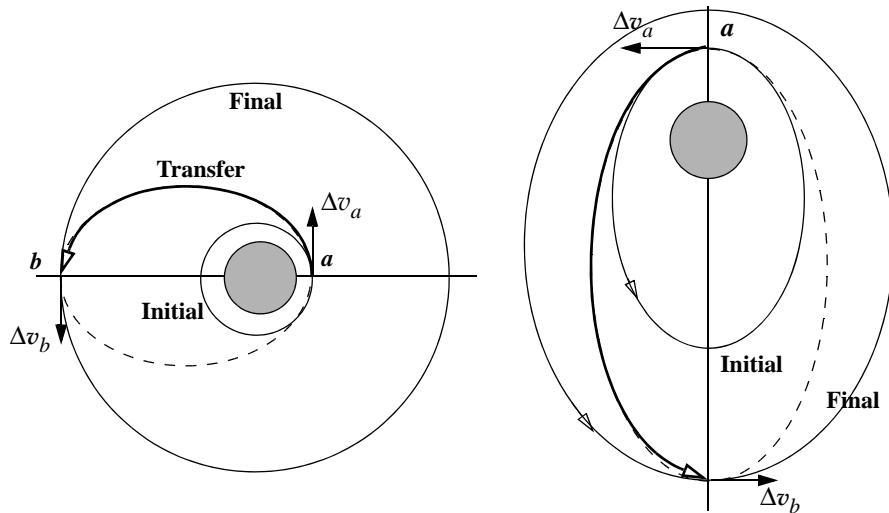


Figure 6-4. Hohmann Transfer. The two-impulse minimum change in velocity maneuver is the Hohmann transfer. Elliptical transfers must start at either apogee or perigee for this maneuver to be a minimum.

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:

$$a_{trans} = \frac{r_{initial} + r_{final}}{2} \quad (6-3)$$

$$\tau_{trans} = \frac{\theta_{trans}}{2} = \pi \sqrt{\frac{a_{trans}^3}{\mu}}$$

A variant of the Hohmann transfer is a method which actually performs two Hohmann transfers in series. Figure 6-5 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 . 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. 6.3.2 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. 6.3.2 shows that in some cases, the bi-elliptic transfer can reduce the total Δv necessary for the transfer.

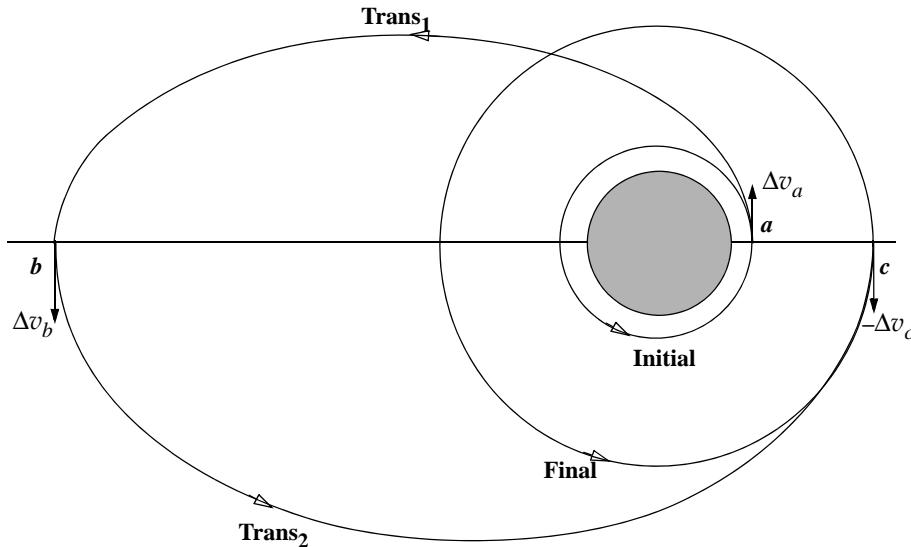


Figure 6-5. Bi-elliptic Transfer. The bi-elliptic transfer employs two Hohmann transfers in succession. This allows us to combine the middle two burns. Using three burns adds complexity but can save Δv in some instances.

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 relevant. 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.
2. Use velocity formulas for elliptical orbits in the formulas. The initial and final orbits will now have varying velocities.

$$v = \sqrt{\frac{2\mu}{r} - \frac{\mu}{a}}$$

3. Specify the semimajor axes of the initial and final orbits from the trajectory equation (+ apogee, - perigee).

$$a = r \left(\frac{1 + e \cos(\nu)}{1 - e^2} \right) = \frac{r}{1 \pm e}$$

4. Draw a diagram! You must 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 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 36: Hohmann Transfer

$$(r_{initial}, r_{final} \Rightarrow a_{trans}, \tau_{trans}, \Delta v_a, \Delta v_b)$$

$$a_{trans} = \frac{r_{initial} + r_{final}}{2}$$

$$\begin{aligned} v_{initial} &= \sqrt{\frac{\mu}{r_{initial}}} & v_{trans_a} &= \sqrt{\frac{2\mu}{r_{initial}} - \frac{\mu}{a_{trans}}} \\ v_{final} &= \sqrt{\frac{\mu}{r_{final}}} & v_{trans_b} &= \sqrt{\frac{2\mu}{r_{final}} - \frac{\mu}{a_{trans}}} \end{aligned}$$

$$\Delta v_a = v_{trans_a} - v_{initial}$$

$$\Delta v_b = v_{final} - v_{trans_b}$$

$$\Delta v = |\Delta v_a| + |\Delta v_b|$$

$$\tau_{trans} = \pi \sqrt{\frac{a_{trans}^3}{\mu}}$$

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 very useful 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 37: Bi-elliptic Transfer ($r_{initial}, r_b, r_{final} \Rightarrow a_{trans1}, a_{trans2}, \tau_{trans1}, \tau_{trans2}, \Delta v_a, \Delta v_b, \Delta v_c$)

$$\begin{aligned}
 a_{trans1} &= \frac{r_{initial} + r_b}{2} & a_{trans2} &= \frac{r_b + r_{final}}{2} \\
 v_{initial} &= \sqrt{\frac{\mu}{r_{initial}}} & v_{trans1a} &= \sqrt{\frac{2\mu}{r_{initial}} - \frac{\mu}{a_{trans1}}} \\
 v_{trans1b} &= \sqrt{\frac{2\mu}{r_b} - \frac{\mu}{a_{trans1}}} & v_{trans2b} &= \sqrt{\frac{2\mu}{r_b} - \frac{\mu}{a_{trans2}}} \\
 v_{trans2c} &= \sqrt{\frac{2\mu}{r_{final}} - \frac{\mu}{a_{trans2}}} & v_{final} &= \sqrt{\frac{\mu}{r_{final}}} \\
 \Delta v_a &= v_{trans1a} - v_{initial} \\
 \Delta v_b &= v_{trans2b} - v_{trans1b} \\
 \Delta v_c &= v_{final} - v_{trans2c} \\
 \Delta v &= |\Delta v_a| + |\Delta v_b| + |\Delta v_c| \\
 \tau_{trans} &= \pi \sqrt{\frac{a_{trans1}^3}{\mu}} + \pi \sqrt{\frac{a_{trans2}^3}{\mu}}
 \end{aligned}$$

An example transfer is useful to compare each method.

▼ Example 6-1: 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:

$$r_{initial} = \frac{191,344\,11 + 6378.137}{6378.137} = 1.03 \text{ ER} = 6569.4781 \text{ km}$$

$$r_{final} = \frac{35,781,348\,57 + 6378.137}{6378.137} = 6.61 \text{ ER} = 42,159.4856 \text{ km}$$

Next, find the velocities at each position in the transfer. From Eq. (1-33) for a satellite's circular speed,

$$v_{initial} = \sqrt{\frac{\mu}{r_{initial}}} = \sqrt{\frac{1}{1.03}} = 0.985\,329 \frac{\text{ER}}{\text{TU}} = 7.789\,390 \text{ km/s}$$

$$v_{final} = \sqrt{\frac{\mu}{r_{final}}} = \sqrt{\frac{1}{6.61}} = 0.388\,955 \frac{\text{ER}}{\text{TU}} = 3.074\,831 \text{ km/s}$$

For the transfer orbit,

$$a_{trans} = \frac{r_{initial} + r_{final}}{2} = 3.820 \text{ ER} = 24,364.4183 \text{ km}$$

From Eq. (1-31) for elliptical satellite speed, find the two transfer velocities:

$$v_{trans_a} = \sqrt{\frac{2\mu}{r_{initial} - a_{trans}}} = \sqrt{\frac{2(1)}{1.03} - \frac{1}{3.82}} = 1.296\,136 \frac{\text{ER}}{\text{TU}} = 10.246\,429 \text{ km/s}$$

$$v_{trans_b} = \sqrt{\frac{2\mu}{r_{final} - a_{trans}}} = \sqrt{\frac{2(1)}{6.61} - \frac{1}{3.82}} = 0.201\,970 \frac{\text{ER}}{\text{TU}} = 1.596\,644 \text{ km/s}$$

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:

$$\blacktriangle \quad \tau_{trans} = \pi \sqrt{\frac{a_{trans}^3}{\mu}} = \pi (3.82)^{3/2} = 23.455\,510 \text{ TU} = 5.256\,713^h = 315.403 \text{ min}$$

▼ Example 6-2. Calculating a Bi-elliptic Transfer.

GIVEN: Initial orbit alt = 191,344 11 km
 Point *b* alt = 503,873 km (79 ER)
 Final orbit alt = 376,310 km (59 ER)

FIND: Δv , time of flight (TOF)

From Example 6-1, 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. (1-33) for circular satellite speed,

$$v_{initial} = \sqrt{\frac{\mu}{r_{initial}}} = \sqrt{\frac{1}{1.03}} = 0.985\,329 \frac{\text{ER}}{\text{TU}} = 7.789\,390 \text{ km/s}$$

$$v_{final} = \sqrt{\frac{\mu}{r_{final}}} = \sqrt{\frac{1}{60.00}} = 0.129\,099 \frac{\text{ER}}{\text{TU}} = 1.020\,578 \text{ km/s}$$

For the transfer orbits,

$$a_{trans1} = \frac{r_{initial} + r_b}{2} = \frac{1.03 + 80}{2} = 40.515 \text{ ER} = 258,410.307 \text{ km}$$

$$a_{trans2} = \frac{r_b + r_{final}}{2} = \frac{80 + 60}{2} = 70.0 \text{ ER} = 446,469.637 \text{ km}$$

From Eq. (1-31) for a satellite's elliptical speed,

$$v_{trans1_a} = \sqrt{\frac{2\mu}{r_{initial} - a_{trans1}}} = \sqrt{\frac{2(1)}{1.03} - \frac{1}{40.515}} = 1.384\,581 \frac{\text{ER}}{\text{TU}} = 10.945\,624 \text{ km/s}$$

$$v_{trans2_c} = \sqrt{\frac{2\mu}{r_{final} - a_{trans2}}} = \sqrt{\frac{2(1)}{60.000} - \frac{1}{70.000}} = 0.138\,013 \frac{\text{ER}}{\text{TU}} = 1.091\,044 \text{ km/s}$$

and for the additional burn of the bi-elliptic,

$$v_{trans1_b} = \sqrt{\frac{2\mu}{r_b} - \frac{\mu}{a_{trans1}}} = \sqrt{\frac{2(1)}{80.000} - \frac{1}{40.515}} = 0.017\ 826 \frac{\text{ER}}{\text{TU}} = 0.140\ 925 \text{ km/s}$$

$$v_{trans2_b} = \sqrt{\frac{2\mu}{r_b} - \frac{\mu}{a_{trans2}}} = \sqrt{\frac{2(1)}{80.000} - \frac{1}{70.000}} = 0.103\ 510 \frac{\text{ER}}{\text{TU}} = 0.818\ 283 \text{ km/s}$$

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:

$$\tau_{trans} = \pi \sqrt{\frac{a_{trans1}^3}{\mu}} + \pi \sqrt{\frac{a_{trans2}^3}{\mu}} = \pi \sqrt{\frac{40.515^3}{1}} + \pi \sqrt{\frac{70.000^3}{1}}$$



$$\tau_{trans} = 2650.076\ 794 \text{ TU} = 593.919\ 803^{\text{h}}$$

6.3.2 Comparing Hohmann and Bi-elliptic Transfers

The Hohmann transfer is the minimum change in velocity 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. (1-31) and Eq. (1-33), respectively] and remember that $a_{initial} = r_a$ for circular orbits:

$$\Delta v_a = \sqrt{\frac{2\mu}{a_{initial}} - \frac{\mu}{a_{trans}}} - \sqrt{\frac{\mu}{a_{initial}}}$$

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):

$$R \equiv r_{final} / r_{initial} = a_{final} / a_{initial} \quad (6-4)$$

Remembering the formula for circular velocity, write

$$\frac{\Delta v_{Hohmann}}{v_{initial}} = \left(1 - \frac{1}{R}\right) \sqrt{\frac{2R}{1+R}} + \sqrt{\frac{1}{R}} - 1 \quad (6-5)$$

Figure 6-6 shows the magnitude of the total velocity change for different values of R . Find the maxima by differentiating Eq. (6-5) with respect to R and setting it equal to zero. From Escobal ([1968] 1979:60),

$$R^3 - 15R^2 - 9R - 1 = 0 \quad R = 15.581\ 76 \quad (6-6)$$

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. (6-5) 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. (6-5) but scale the result by v_{final} because $v_{initial}$ at ∞ is zero:

$$\begin{aligned} \Delta v_a &= (\sqrt{2} - 1)v_{initial} \\ \Delta v_b &= (\sqrt{2} - 1)v_{final} \end{aligned}$$

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),

$$\frac{\Delta v_\infty}{v_{initial}} = (\sqrt{2} - 1) \left(1 + \sqrt{\frac{1}{R}}\right) \quad (6-7)$$

Figure 6-6 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. (6-5) and Eq. (6-7), so

$$R^3 - (7 + 4\sqrt{2})R^2 + (3 + 4\sqrt{2})R - 1 = 0 \quad R = 11.938\ 76 \quad (6-8)$$

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, absolute values for generality):

$$\frac{\Delta v_{bi-elliptic}}{v_{initial}} = \left| \sqrt{\frac{2R^*}{1+R^*}} - 1 \right| + \left| \sqrt{\frac{2}{R^*}} \left(\sqrt{\frac{1}{1+\frac{R^*}{R}}} - \sqrt{\frac{1}{1+R^*}} \right) \right| + \left| \sqrt{\frac{1}{R}} \left(\sqrt{\frac{2R^*}{R+R^*}} - 1 \right) \right| \quad (6-9)$$

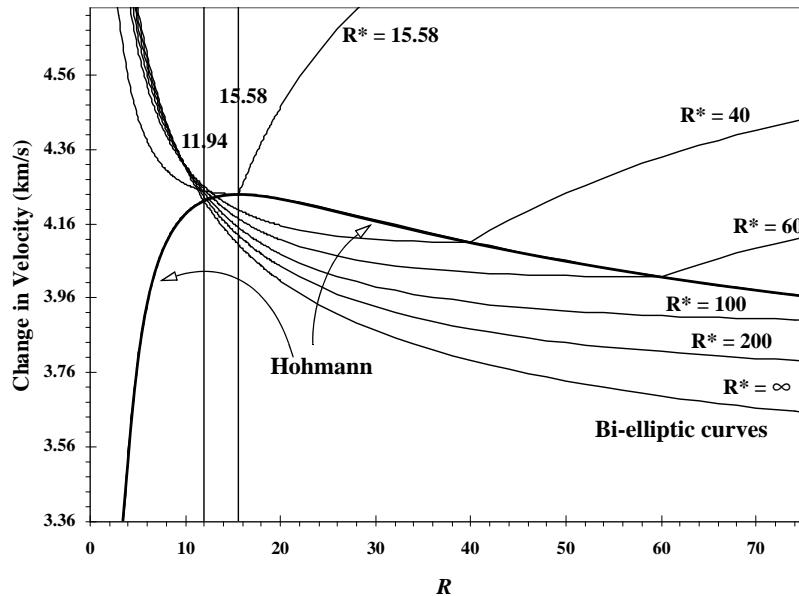


Figure 6-6. Hohmann Transfer—Comparison of Change in Velocity. The change in velocity is shown when using Hohmann and bi-elliptic orbit changes. Be aware of the different notations for the ratios (final / initial) (R for Hohmann and R^* for bi-elliptic). Also, for all cases, $v_{initial} = 1$ ER/TU = 7.905 km/s.

Figure 6-6 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.

6.3.3 Transfers Using the One-Tangent Burn

A 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 6-7 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. 6-7 because of the requirement for the ellipses to be coaxially aligned.

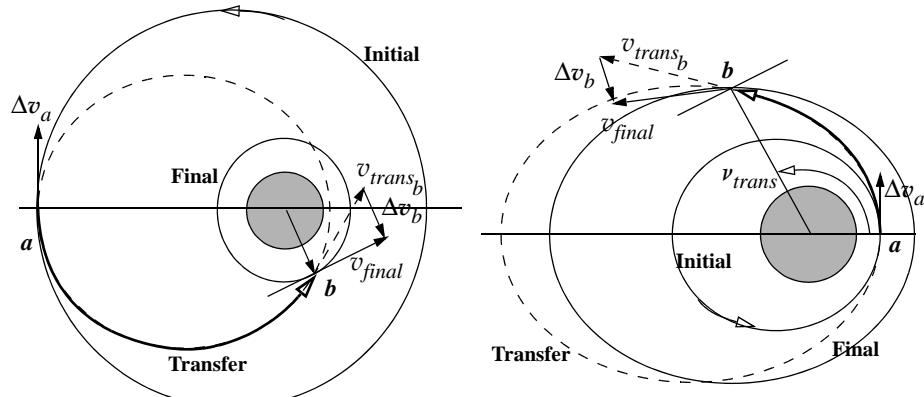


Figure 6-7. One-Tangent Burns. For both circular (left) and elliptical (right) orbit changes, the true anomalies of the transfer and final orbits are equal. The maneuver starts at apoapsis (left, $\nu > 180^\circ$) and periapsis (right, $\nu < 180^\circ$) of the transfer orbit.

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. (2-75) applies for the radius of periapsis (-) or apoapsis (+).

$$r_{initial} = a_{trans}(1 \mp e_{trans})$$

$$r_{final} = \frac{a_{trans}(1 - e_{trans}^2)}{1 + e_{trans}\cos(\nu_{trans_b})}$$

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

$$e_{trans}(r_{final} \cos(\nu_{trans_b}) \mp r_{initial}) = r_{initial} - r_{final}$$

Using the reciprocal of Eq. (6-4), we see that

$$R^{-1} = \frac{r_{initial}}{r_{final}}$$

and solving for eccentricity yields the desired result:

$$e_{trans} = \frac{R^{-1} - 1}{\cos(\nu_{trans_b}) \mp R^{-1}} \quad \begin{cases} - \text{periapsis} \\ + \text{apoapsis} \end{cases} \quad (6-10)$$

Notice that ν_{trans_b} is usually less than 180° for transfers that begin at perigee, and larger than 180° for transfers that begin at apogee. All solutions have two answers—the other occurs at $360^\circ - \nu_{trans_b}$ —the only difference is the time because Δv is the same.

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. Besides the general restrictions on ν based on the origin of the maneuver, we want to keep the eccentricity positive. By solving for the denominator of Eq. (6-10), we see that the true anomaly must obey the following restrictions.

Periapsis	Apoapsis
$IF(R^{-1} > 1), \cos(\nu_{trans_b}) > R^{-1}$	$IF(R^{-1} > 1), \cos(\nu_{trans_b}) > -R^{-1}$
$IF(R^{-1} < 1), \cos(\nu_{trans_b}) < R^{-1}$	$IF(R^{-1} < 1), \cos(\nu_{trans_b}) < -R^{-1}$

We need the semimajor axis for the transfer ellipse, but we can't calculate it as in Eq. (6-3) 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. (2-75), we see that

$$a_{trans} = \frac{r_{initial}}{1 \mp e_{trans}}$$

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. (6-2)]. Remember, you must know the flight-path angle [Eq. (2-95)].

Finally, calculate the transfer time using Kepler's equation [Eq. (2-7)] 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. (2-8). 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 38: One-Tangent Burn

$$(r_{initial}, r_{final}, v_{trans_b} \Rightarrow e_{trans}, a_{trans}, \tau_{trans}, \Delta v_a, \Delta v_b)$$

$$R^{-1} = \frac{r_{initial}}{r_{final}}$$

$$e_{trans} = \frac{R^{-1} - 1}{\cos(\nu_{trans_b}) \mp R^{-1}}$$

$$a_{trans} = \frac{r_{initial}}{1 \mp e_{trans}} \quad \begin{cases} - \text{periapsis} \\ + \text{apoapsis} \end{cases}$$

$$v_{initial} = \sqrt{\frac{\mu}{r_{initial}}} \quad v_{trans_a} = \sqrt{\frac{2\mu}{r_{initial}} - \frac{\mu}{a_{trans}}}$$

$$v_{final} = \sqrt{\frac{\mu}{r_{final}}} \quad v_{trans_b} = \sqrt{\frac{2\mu}{r_{final}} - \frac{\mu}{a_{trans}}}$$

$$\Delta v_a = v_{trans_a} - v_{initial}$$

$$\tan(\phi_{trans_b}) = \frac{e_{trans} \sin(\nu_{trans_b})}{1 + e_{trans} \cos(\nu_{trans_b})}$$

$$\Delta v_b = \sqrt{v_{trans_b}^2 + v_{final}^2 - 2v_{trans_b}v_{final} \cos(\phi_{trans_b})}$$

$$\Delta v_{otb} = |\Delta v_a| + |\Delta v_b|$$

$$\cos(E) = \frac{e_{trans} + \cos(\nu_{trans_b})}{1 + e_{trans} \cos(\nu_{trans_b})}$$

$$\tau_{trans} = \sqrt{\frac{a_{trans}^3}{\mu}} \left\{ 2k\pi + (E - e_{trans} \sin(E)) - (E_o - e_{trans} \sin(E_o)) \right\}$$

Let's look at an example of a one-tangent burn, so we can compare it to other methods.

▼ Example 6-3. Performing a One-tangent Burn.

GIVEN: Initial orbit alt = 191,344 11 km
 Final orbit alt = 35,781,348 57 km
 Transfer point $\nu_{trans_b} = 160^\circ$
 FIND: Δv , time of flight (τ_{trans})

From Example 6-1, 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:

$$R^{-1} = \frac{r_{initial}}{r_{final}} = \frac{1.03}{6.61} = 0.155\ 821$$

$$e_{trans} = \frac{R^{-1} - 1}{\cos(\nu_{trans_b}) - R^{-1}} = \frac{0.155\ 821 - 1}{\cos(160^\circ) - 0.155\ 821} = 0.770\ 578$$

$$a_{trans} = \frac{r_{initial}}{(1 - e_{trans})} = \frac{1.03}{1 - 0.770\ 578} = 4.489\ 438 \text{ ER} = 28,634.2505 \text{ km}$$

From the formula for a satellite's circular speed,

$$v_{initial} = \sqrt{\frac{\mu}{r_{initial}}} = \sqrt{\frac{1}{1.03}} = 0.985\ 355 \frac{\text{ER}}{\text{TU}} = 7.789\ 390 \text{ km/s}$$

$$v_{final} = \sqrt{\frac{\mu}{r_{final}}} = \sqrt{\frac{1}{6.61}} = 0.388\ 961 \frac{\text{ER}}{\text{TU}} = 3.074\ 831 \text{ km/s}$$

From the equation for a satellite's elliptical speed,

$$v_{trans_a} = \sqrt{\frac{2\mu}{r_{initial}} - \frac{\mu}{a_{trans}}} = \sqrt{\frac{2(1)}{1.03} - \frac{1}{4.489\ 438}} = 1.311\ 108 \frac{\text{ER}}{\text{TU}} = 10.364\ 786 \text{ km/s}$$

$$v_{trans_b} = \sqrt{\frac{2\mu}{r_{final}} - \frac{\mu}{a_{trans}}} = \sqrt{\frac{2(1)}{6.61} - \frac{1}{4.489\ 438}} = 0.282\ 536 \frac{\text{ER}}{\text{TU}} = 2.233\ 554 \text{ km/s}$$

$$\Delta v_a = v_{trans_a} - v_{initial} = 1.311\ 108 - 0.985\ 355 = 0.325\ 778 \frac{\text{ER}}{\text{TU}} = 2.575\ 396 \text{ km/s}$$

Use the flight-path angle for the nontangential transfer and, from Eq. (2-95) and Eq. (2-96),

$$\tan(\phi_{trans_b}) = \frac{e_{trans} \sin(\nu_{trans_b})}{1 + e_{trans} \cos(\nu_{trans_b})} = \frac{0.770\,578 \sin(160^\circ)}{1 + 0.770\,578 \cos(160^\circ)} \quad \phi_{trans_b} = 43.688\,825^\circ$$

Because the final orbit is circular, the angle change is ϕ_{trans_b} , so

$$\Delta v_b = \sqrt{v_{final}^2 + v_{trans_b}^2 - 2v_{final}v_{trans_b} \cos(\phi_{trans_b})} = \sqrt{(0.388\,961)^2 + (0.282\,536)^2 - 2(0.388\,961)(0.282\,536) \cos(43.688\,825^\circ)} = 0.268\,670 \frac{\text{ER}}{\text{TU}}$$

The total Δv_{otb} is simply the sum of the two burns:

$$\Delta v_{otb} = |\Delta v_a| + |\Delta v_b| = 0.325\,778 + 0.268\,670 = 0.594\,448 \frac{\text{ER}}{\text{TU}} = 4.699\,333 \frac{\text{km}}{\text{s}}$$

Calculate time of flight using Kepler's equation:

$$\tau_{trans} = \sqrt{\frac{a_{trans}^3}{\mu}} \left\{ 2k\pi + (E - e_{trans} \sin(E)) - (E_o - e_{trans} \sin(E_o)) \right\}$$

Because this transfer starts at periapsis, $E_o = 0^\circ$. The transfer doesn't pass perigee, so k must equal zero. From Eq. (2-9),

$$\cos(E) = \frac{e_{trans} + \cos(\nu_{trans_b})}{1 + e_{trans} \cos(\nu_{trans_b})} = \frac{0.770\,578 + \cos(160^\circ)}{1 + 0.770\,578 \cos(160^\circ)}$$

$$E = 127.805\,194^\circ \text{ and } E = 232.194\,810^\circ$$

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. (2-6)] results in

$$\tau_{trans} = \sqrt{4.489\,303^3} \left(\frac{127.805\,194}{57.2958} - 0.770\,578 \sin(127.805\,194^\circ) \right)$$

▲ $\tau_{trans} = 15.427\,06 \text{ TU} = 207.445 \text{ min}$

Table 6-1 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 6-1 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 a large penalty in time.*

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. (1-20)]. 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,

* 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. 6.6 will discuss rendezvous.

TABLE 6-1. 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. 6-6.

	Initial Alt (km)	Final Alt (km)	v_{transb}	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

$$\xi_{new} = \frac{(v + \Delta v)^2}{2} - \frac{\mu}{r}$$

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 periapsis 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 periapsis. 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!

6.3.4 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 many parameters about the orbit. In practice, if you’re trying to do this complex a maneuver, you must solve Lambert’s problem (Sec. 7.6). 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!

6.4 Noncoplanar Transfers

Coplanar maneuvers enable a satellite to alter three orbital elements: semimajor axis, eccentricity, and argument of perigee. To do an intercept or rendezvous, we may need to change two other elements: inclination, i , and the right ascension 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 right ascension of the ascending node. A Δv applied at *any other point* will change both inclination and the node. Once we’ve explored general noncoplanar maneuvers, we’ll look at combined (coplanar and noncoplanar) maneuvers.

6.4.1 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. 6-8.

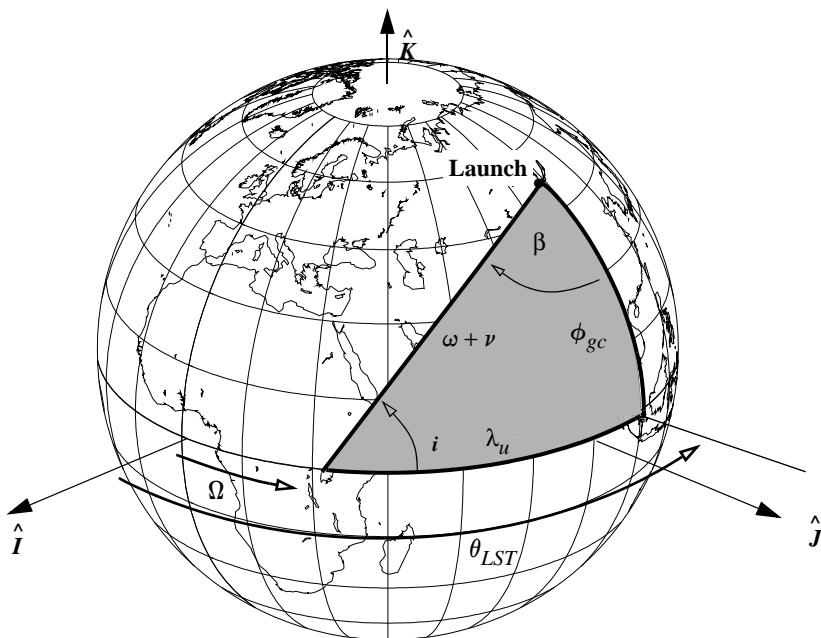


Figure 6-8. Launch-Window Geometry. Direct launches can occur only if the launch site's latitude is less than or equal to the satellite's inclination, i . Determining the launch window involves three basic steps: (1) finding the launch azimuth, β , from the site (Γ ve shown an equivalent angle in the triangle because β is measured from north), (2) determining the auxiliary angle, λ_u , and (3) finding the GMST (or LST).

By examining the right spherical triangle, we can use Eq. (C-25) to develop an expression for the inclination:

$$\cos(i) = \cos(\phi_{gc})\sin(\beta) \quad (6-11)$$

and then solve for the launch azimuth, β :

$$\sin(\beta) = \frac{\cos(i)}{\cos(\phi_{gc})} \quad (6-12)$$

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:

$$\sin(\phi_{gc}) = \sin(i)\sin(\omega + \nu) \quad (6-13)$$

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

$$\vec{v}_L = \vec{\omega}_{\oplus} \times \vec{r}_{site} \quad \text{or} \quad v_L = |\vec{\omega}_{\oplus} \times \vec{r}_{site}| = \omega_{\oplus} r_{site} \cos(\phi_{gc}) \quad (6-14)$$

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 back cover, we find the velocity for a site at the equator.

$$v_L = \omega_{\oplus} r_{site} = 7.292\ 115\ 9 \times 10^{-5} (6,378,137.0) = 465.101\ 14 \text{ m/s} \quad (6-15)$$

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. 6-9, along with their recommended firing azimuths (due to safety considerations during launches). Notice how most sites are at low latitudes. Table 6-2 lists data for the launch sites.

Launch Windows

Finally, launch windows are an integral part of 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. 6-8 to visualize these relationships. From Eq. (6-12), we can determine the launch azimuth. Remember the sine will always give two answers (the other angle is $180^\circ - \beta$). For Eq. (6-12), 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. 6-8. Thus,

$$\cos(\lambda_u) = \frac{\cos(\beta)}{\sin(i)} \quad (6-16)$$

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_{GMST} , when the angles are in the proper position. Notice that

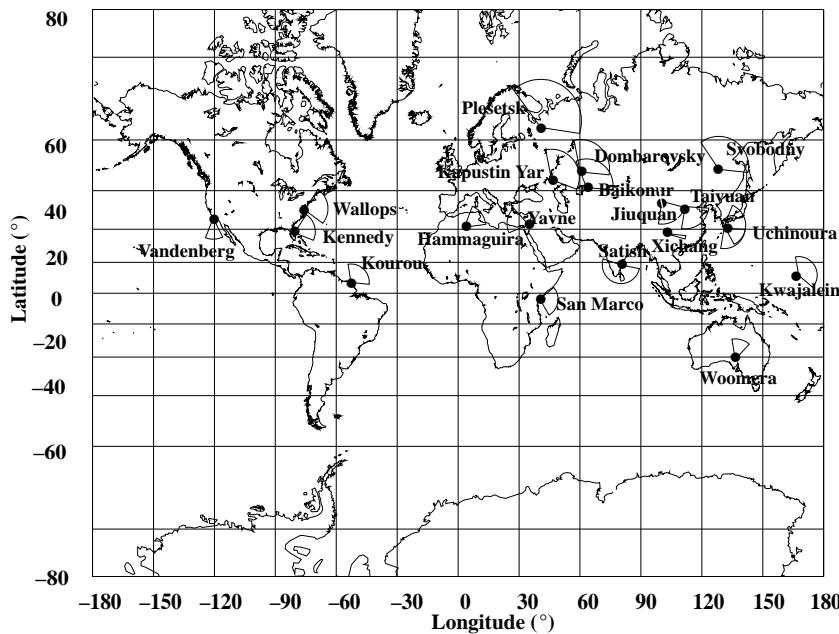


Figure 6-9. Locations for Selected Launch Sites. Direct launches are shown from various launch sites. Note the restrictions so boosters don't hit civilian areas and the relatively uniform low latitudes to take advantage of the Earth's rotation. Nominal 1000 km "fans" are shown for each site.

from Fig. 6-8, $\lambda_u = \theta_{LST} - \Omega$. For this step, we need the longitude of the launch site, λ , so we can use GMST and

$$\theta_{GMST} = \Omega + \lambda_u - \lambda$$

We know the latitude and inclination are fixed, and we usually have a tolerance for the right ascension of the ascending node ($\pm\Delta\Omega$) (See the discussions with Figures 11-13 and 11-15). This tolerance creates the launch window. Once we determine the range of values for GMST, the final step is to find the time of day (UT) for launch. Solving Eq. (3-46) for UT gives us

$$UT = \frac{\theta_{GMST} - \theta_{GMST\ 0^h}}{\omega_{\oplus}} \quad (6-17)$$

Recall that $\theta_{GMST\ 0^h}$ is the GMST at 0^h on a particular day. As we substitute particular values of $\theta_{GMST\ 0^h}$ for each day into Eq. (6-17), we get a series of potential launch times. Any variability in Ω gives us a window on each day, where the mission can succeed.

TABLE 6-2. Locations of Selected Launch Sites. Minimum and maximum azimuth values are usually determined so launches don't fly over populated areas. Some sites do not list specific azimuth restrictions. You can find the orbital inclinations using Eq. (6-11). Don't forget to evaluate the east/west azimuths for min/max inclinations. [Source: Kelso (2007), Larson and Wertz (1992:679–681) and Chiulli (1994)]

Name	Location	Lat	Lon	Min Az	Max Az	Min <i>i</i>	Max <i>i</i>
Alcantara Space Center	Alcantara, Brazil	2.283	-44.38	343	90	2.3	107.0
Baikonur Cosmodrome, aka Tyuratam Missile and Space Center	Kazakhstan	45.997	63.301	340	90	46.0	103.7
Dombarovsky Launch Site, Russia	Dombarovsky, Russia	50.803	59.512 also Yasny nearby				
Guiana Space Center	Kourou, French Guiana	5.235	-52.772	351	95	5.2	99.0
Hammaguir Space Track Range	Algeria	30.876	3.066 closed				
Jiuquan Space Center	China	40.966	100.285	101	190	42.2	97.5
Kapustin Yar Missile & Space Complex	Russia	48.648	46.017	350	107	48.6	96.6
Kodiak Launch Complex	Kodiak, Alaska	57.436	-152.338	110	220	59.6	110.2
Plesetsk Missile and Space Complex	Russia	62.922	40.661	330	90	62.9	103.2
Reagan Test Site	Kwajalein Atoll	8.720	167.732	-15	145	8.7	104.8
San Marco Launch Platform	Indian Ocean (Kenya)	-2.941	40.213	50	150	2.9	40.1
Satish Dhawan Space Center, formerly Sriharikota Launching Range	India	13.727	80.232	100	290	16.9	155.9
Svobodny Launch Complex	Russia	51.880	128.374 closed, Vostochny replacement				
Taiyuan Space Center	China	38.849	111.608	90	190	38.8	97.8
Tanegashima Space Center	Yoshinobu, Japan	30.400	130.974	90	190	30.4	98.6
Uchinoura Space Center, formerly Kagoshima Space Center	Japan	31.251	131.081	20	150	31.3	73.0
USAF Eastern Test Range	Cape Canaveral, FL	28.533	-80.575	32	111	28.5	62.3
USAF Western Test Range	Vandenburg AFB, CA	34.669	-120.61	152	202	67.3	107.9
Wallops Island	Wallops Island, VA	37.843	-75.479	90	160	37.8	74.3
Woomera, Australia	Woomera, Australia	-30.943	136.521	350	45	52.7	98.6
Xichang	China	28.246	102.027	94	105	31.7	28.5
Yavne Launch Facility	Israel	31.898	34.701				

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. (6-12). You must use this value in Eq. (6-16) for consistency.

For instance, proceed as follows to determine the correct GMST at the descending node of a direct orbit. Find the launch azimuth using Eq. (6-12). Select the second angle ($180^\circ - \beta$) because you want the descending node. Next find the auxiliary angle from Eq. (6-16). 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.

6.4.2 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. 6-10. Notice the only two common points in the old and new orbits occur at the nodes.

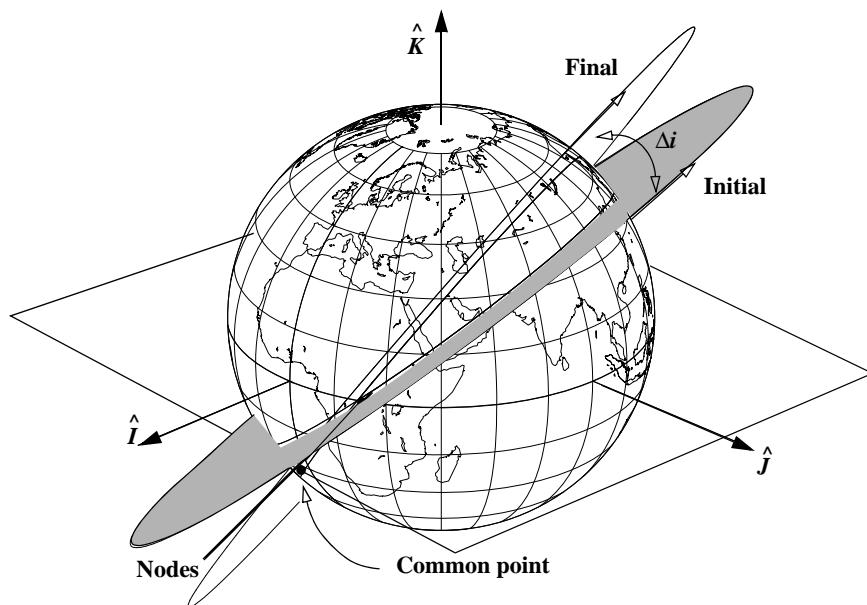


Figure 6-10. Inclination-Only Changes. An inclination-only change must occur at one of the nodal crossings because these are the only two points common to *both* orbits. The angular separation is equal to the change in inclination.

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:

$$\vec{v}_o \cdot \vec{v} = v_o v \cos(\Theta)$$

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:

$$\vec{v}_o = v_o \sin(\phi_{fp}) \hat{R} + v_o \cos(\phi_{fp}) \hat{S}$$

The second vector must use the same RSW system, but because there is an additional change in the inclination,

$$\hat{\vec{v}} = v \sin(\phi_{fpa}) \hat{R} + v \cos(\phi_{fpa}) \cos(\Delta i) \hat{S} + v \cos(\phi_{fpa}) \sin(\Delta i) \hat{W}$$

Use these velocity-vector expressions in the dot product and solve for the transfer angle:

$$\cos(\Theta) = \sin^2(\phi_{fpa}) + \cos^2(\phi_{fpa}) \cos(\Delta i) \quad (6-18)$$

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

$$\Delta v_{i \text{ only}} = 2v_{initial} \cos(\phi_{fpa}) \sin\left(\frac{\Delta i}{2}\right) \quad (6-19)$$

where we include the subscripts to remind ourselves of the initial values, even though they'll equal the final values. Analysis of Eq. (6-19) 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 $\Delta v_{i \text{ only}}$. Recognize that plane changes are expensive. For circular orbits, if $\Delta i = 60^\circ$, $\Delta v_{i \text{ only}} = v_{initial}$! 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

$$\Delta v_{i \text{ only}} = 2v_{initial} \sin\left(\frac{\Theta}{2}\right) \quad (6-20)$$

Comparing Eq. (6-19) and Eq. (6-20), 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 required, be sure to check *both* nodes and use the node with the smaller initial velocity. For circular orbits, there will be no difference. The flight-path angle is necessary for elliptical orbits.

ALGORITHM 39: 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 6-4. 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:

$$a = \frac{p}{1 - e^2} = \frac{2.8}{1 - 0.3^2} = 3.076\ 923 \text{ ER} = 19,625.037 \text{ km}$$

Then, find the position using the trajectory equation:

$$r = \frac{p}{1 + e \cos(\nu)} = \frac{2.8}{1 + 0.3 \cos(330)} = 2.222\ 562 \text{ ER} = 14,175.802 \text{ km}$$

The velocity's magnitude is

$$v = \sqrt{\frac{2\mu}{r} - \frac{\mu}{a}} = \sqrt{\frac{2(1)}{2.222\ 56} - \frac{1}{3.0769}} = 0.758\ 197 \frac{\text{ER}}{\text{TU}} = 5.993\ 824 \text{ km/s}$$

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

$$\Delta v_{i\ only} = 2v_{initial} \cos(\phi_{fpa}) \sin\left(\frac{\Delta i}{2}\right) = 2(0.758\ 196\ 9) \cos(-6.7900) \sin\left(\frac{15.0}{2}\right)$$

$$\Delta v_{i\ only} = 0.196\ 540\ 8 \text{ ER/TU} = 1.553\ 727 \text{ km/s}$$

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

$$r = \frac{p}{1 + e \cos(\nu)} = \frac{2.8}{1 + 0.3 \cos(150^\circ)} = 3.7828 \text{ ER} = 24,127.2190 \text{ km}$$

$$v = \sqrt{\frac{2\mu}{r} - \frac{\mu}{a}} = \sqrt{\frac{2(1)}{3.7828} - \frac{1}{3.0769}} = 0.451\ 341\ 2 \frac{\text{ER}}{\text{TU}} = 3.568\ 017 \text{ km/s}$$

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

$$\Delta v_i \text{ only} = 0.115\ 476 \frac{\text{ER}}{\text{TU}} = 0.912\ 883 \text{ km/s}$$

▲ Notice the reduced overall velocity. You should check both nodes for elliptical orbits.

6.4.3 Changes in the Right Ascension of the 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 right ascension of the ascending node. This is just one possible case that 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. 6-11. 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 right ascension of the ascending node (Sec. 6.4.4). 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 .

We need 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. 6-11 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 :

$$\cot(u_{\text{initial}}) \sin(\Delta\Omega) = -\sin(i_{\text{initial}}) \cot(i_{\text{final}}) + \cos(i_{\text{initial}}) \cos(\Delta\Omega)$$

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 use a cosine relation. Eq. (C-28) 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:

$$\cos(u_{\text{initial}}) \sin(\vartheta) \cos(i_{\text{initial}}) = \sin(i_{\text{final}}) \cos(\Delta\Omega) - \cos(\vartheta) \sin(i_{\text{initial}})$$

Simplifying, because the inclinations are equal, we get

$$\cos(u_{\text{initial}}) = \tan(i_{\text{initial}}) \frac{\cos(\Delta\Omega) - \cos(\vartheta)}{\sin(\vartheta)} \quad (6-21)$$

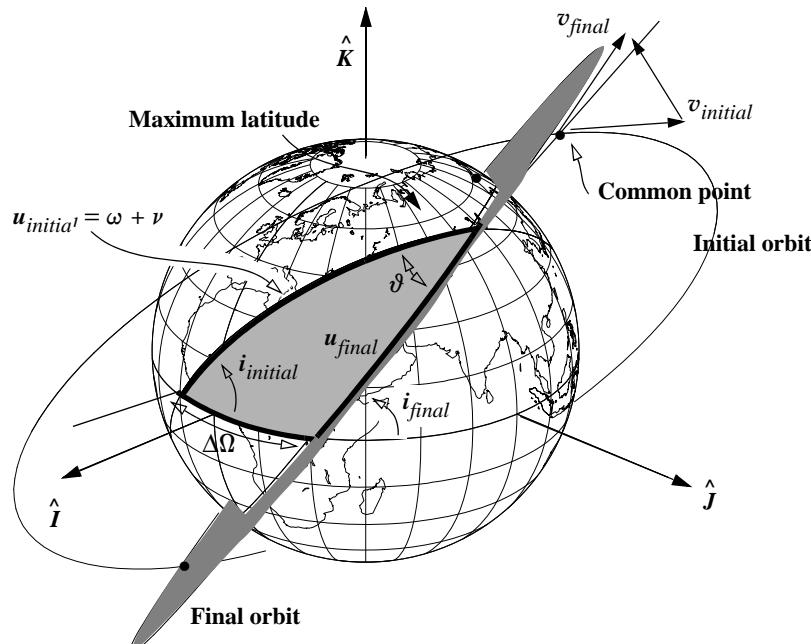


Figure 6-11. Effect of Changing the Right ascension of the Ascending Node. Circular orbits will always have two common points where the orbit paths cross each other. Because these points occur at different true-anomaly values in an elliptical orbit, we must analyze them separately.

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

$$\cos(u_{final}) = \cos(i_{initial})\sin(i_{initial})\left(\frac{1-\cos(\Delta\Omega)}{\sin(\vartheta)}\right) \quad (6-22)$$

To determine the burn angle ϑ , use Eq. (C-26). 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

$$\cos(\vartheta) = \cos^2(i_{initial}) + \sin^2(i_{initial})\cos(\Delta\Omega) \quad (6-23)$$

Although the sine law is equally valid, 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, occurring in the Southern Hemisphere in Fig. 6-11.

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. (6-19) but replacing Δi with ϑ and eliminating ϕ_{fpa} . Thus,

$$\Delta v_{\Omega \text{ only}_{cir}} = 2v_{initial} \sin\left(\frac{\vartheta}{2}\right) \quad (6-24)$$

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. 6-11, $u_{initial} \neq u_{final}$. We won't develop options to solve this problem because it's much more appropriate to solve the Lambert problem to find the actual velocity required for the maneuver. I'll show the algorithm and an example for circular orbits only.

ALGORITHM 40: Change in the 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 node for a circular orbit.

▼ **Example 6-5. Finding Velocity Change for Changes in the Ascending Node Only.**

GIVEN: $i = 55.0^\circ$, $\Delta\Omega = 45.0^\circ$, $v_{initial} = 5.892\ 311 \text{ km/s} = 5.892\ 311 / 7.905\ 365$
 $= 0.745\ 356 \text{ ER/TU}$, $e = 0.0$ (therefore $a = 11,480.649 \text{ km}$)

FIND: $\Delta v_{\Omega \text{ only}}$

First, determine the angle through which the velocity is rotated:

$$\begin{aligned} \cos(\vartheta) &= \cos^2(i) + \sin^2(i) \cos(\Delta\Omega) = \cos^2(55^\circ) + \sin^2(55^\circ) \cos(45^\circ) \\ \vartheta &= 36.537\ 66^\circ \end{aligned}$$

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. (6-21) and Eq. (6-22).

▲ $u_{initial} = 103.3647^\circ$ and $u_{final} = 76.6353^\circ$.

6.4.4 Changes to Inclination and the 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. 6-12 shows. Algorithm 41 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-26)] so you can avoid quadrant checks. Substituting in known values for the triangle, find

$$\cos(u_{initial}) = \frac{\sin(i_{final})\cos(\Delta\Omega) - \cos(\vartheta)\sin(i_{initial})}{\sin(\vartheta)\cos(i_{initial})} \quad (6-25)$$

You can't simplify this equation, as we did for changes in the right ascension of the ascending node, because the inclination change is predetermined. Find the final orbit argument of latitude from the development of Eq. (6-22), also with no simplification:

$$\cos(u_{final}) = \frac{\cos(i_{initial})\sin(i_{final}) - \sin(i_{initial})\cos(i_{final})\cos(\Delta\Omega)}{\sin(\vartheta)} \quad (6-26)$$

The separation angle is the angle between the two orbits, which you can find developed in Eq. (6-23). However, the inclinations are *not* the same, so don't simplify the equation.

The Δv will be the same as for the node change discussed previously [Eq. (6-24)]. 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. 7.6).

ALGORITHM 41: 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.

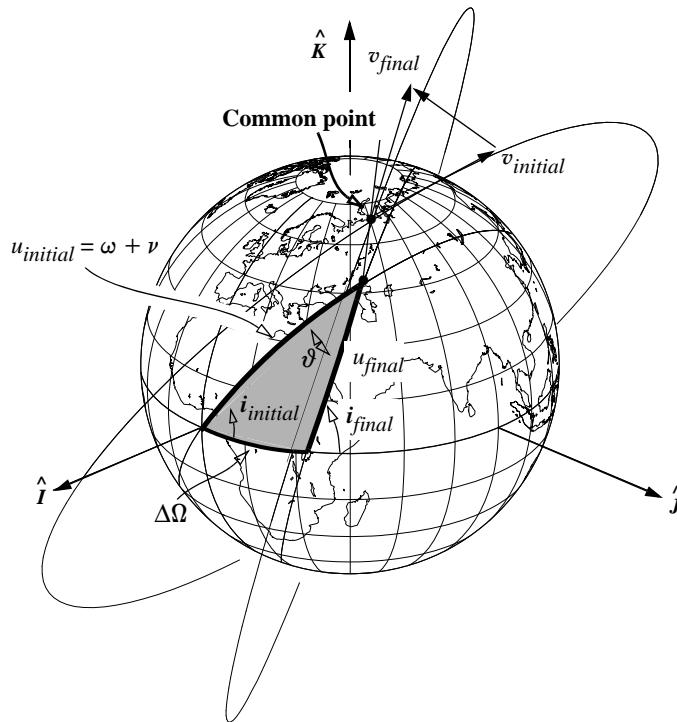


Figure 6-12. Geometry for General Noncoplanar Change. This type of change requires a general solution to find the common point of the two orbits. Note that both the inclination and ascending node are changing.

▼ **Example 6-6. 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: Δv_i and Ω

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 $u_{final} = 97.3803^\circ$.

Lovell (2006) suggests that for small changes in inclination and ascending node, we can use similar expressions. Begin with Fig. 6-12 and use Eq. (C-25).

$$\cos(\pi - (i_{initial} + \Delta i)) = -\cos(\vartheta)\cos(i_{initial}) + \sin(\vartheta)\sin(i_{initial})\cos(u_{initial})$$

Simplify this using Eq. (C-14) for the left hand side to get

$$\cos(\pi - (i_{initial} + \Delta i)) = -\{\cos(i_{initial})\cos(\Delta i) - \sin(i_{initial})\sin(\Delta i)\}$$

Next, recognize that Δi and ϑ will be small. Thus the equation may be written as

$$\begin{aligned} -\cos(i_{initial}) + \Delta i \sin(i_{initial}) &= -\cos(i_{initial}) + \vartheta \sin(i_{initial}) \cos(u_{initial}) \\ \Delta i &= \vartheta \cos(u_{initial}) \end{aligned} \quad (6-27)$$

Now, use Eq. (C-26) and find

$$\sin(\vartheta)\sin(u_{initial}) = \sin(\Delta\Omega)\sin(\pi - (i_{initial} + \Delta i))$$

Small angle approximations let us find

$$\vartheta \sin(u_{initial}) = \Delta\Omega \sin(i_{initial}) \quad (6-28)$$

Solving Eq. (6-27) for ϑ and inserting into Eq. (6-28), we can simplify to find

$$u_{initial} = \text{ATAN}\left(\frac{\Delta\Omega \sin(i_{initial})}{\Delta i}\right) \quad (6-29)$$

The central angle is found by squaring Eq. (6-27) and Eq. (6-28). Simplifying,

$$\vartheta = \sqrt{(\Delta\Omega \sin(i))^2 + \Delta i^2} \quad (6-30)$$

6.5 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.

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. 7.6. Although we can also determine ellipti-

cal 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 6-13 shows the situation, which we need to address before we try to solve a particular mission with changes in orbital inclination, size, and shape.

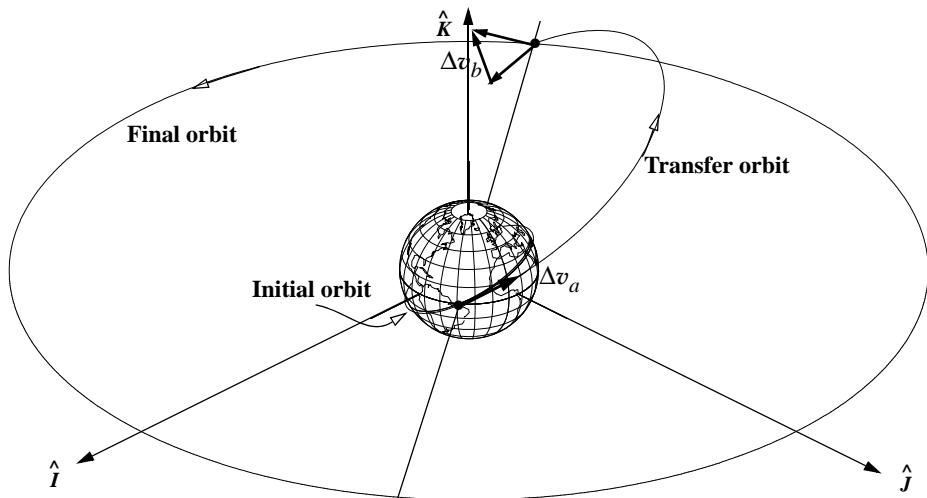


Figure 6-13. Geometry for a Combined Maneuver. A transfer from LEO to GEO is shown to the correct scale. The transfer orbit is just slightly below the initial 28.5° LEO orbit. Notice how close the LEO satellite is to the Earth when it's drawn to scale.

6.5.1 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

* Practical maneuver planning uses optimization techniques to simultaneously solve specific thruster characteristics (thrust, fixed- Δv , orientations, etc.) and mission constraints (placement and frequency of burns), given knowledge of the vehicle (ephemeris accuracy, thrust models, etc.).

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. (6-2). The major difference from Eq. (6-2) is that the angular separation will equal the change in inclination. Thus,

$$\Delta v = \sqrt{v_{trans}^2 + v_{initial}^2 - 2v_{initial}v_{trans}\cos(\Delta i)} \quad (6-31)$$

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 ($\Delta i_{initial} = s\Delta i$, $\Delta i_{final} = (1 - s)\Delta i$).

$$\begin{aligned} \Delta v = \Delta v_a + \Delta v_b &= \sqrt{v_{initial}^2 + v_{trans_a}^2 - 2v_{initial}v_{trans_a}\cos(s\Delta i)} \\ &+ \sqrt{v_{final}^2 + v_{trans_b}^2 - 2v_{final}v_{trans_b}\cos((1-s)\Delta i)} \end{aligned}$$

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 $\text{SIN}(s\Delta i)$.

$$\text{SIN}(s\Delta i) = \frac{\Delta v_a v_{final} v_{trans_b} \text{SIN}((1-s)\Delta i)}{\Delta v_b v_{initial} v_{trans_a}}$$

Because $\text{SIN}(s\Delta i)$ 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:

$$\begin{aligned} \Delta v_a^2 + \Delta v_b^2 &\approx v_{initial}^2 + v_{trans_a}^2 - 2v_{initial}v_{trans_a}\cos(s\Delta i) \\ &+ v_{final}^2 + v_{trans_b}^2 - 2v_{final}v_{trans_b}\cos((1-s)\Delta i) \end{aligned}$$

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:

$$\frac{\partial(\Delta v_a^2 + \Delta v_b^2)}{\partial s} \approx \Delta i 2v_{initial}v_{trans_a} \text{SIN}(s\Delta i) - \Delta i 2v_{final}v_{trans_b} \text{SIN}((1-s)\Delta i) = 0$$

Collecting terms and rearranging results in

$$\frac{\text{SIN}(s\Delta i)}{\text{COS}(s\Delta i)} \approx \frac{v_{trans_b}v_{final}\text{SIN}(\Delta i)}{v_{initial}v_{trans_a} + v_{trans_b}v_{final}\text{COS}(\Delta i)}$$

and finally,

$$s \approx \frac{1}{\Delta i} \text{TAN}^{-1} \left[\frac{\text{SIN}(\Delta i)}{\frac{v_{initial}v_{trans_a}}{v_{final}v_{trans_b}} + \cos(\Delta i)} \right]$$

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. (1-31) and Eq. (1-33) results in

$$\frac{v_{initial}v_{trans_a}}{v_{final}v_{trans_b}} = \frac{\sqrt{\frac{\mu}{r_{initial}}} \sqrt{\frac{2\mu}{r_{initial}} - \frac{\mu}{a_{trans}}}}{\sqrt{\frac{\mu}{r_{final}}} \sqrt{\frac{2\mu}{r_{final}} - \frac{\mu}{a_{trans}}}} = \frac{\sqrt{\frac{\mu}{r_{initial}}} \sqrt{2\mu}}{\sqrt{\frac{\mu}{r_{final}}} \sqrt{2\mu}} \frac{\frac{r_{final}}{r_{initial}(r_{initial} + r_{final})}}{\frac{r_{initial}}{r_{final}(r_{initial} + r_{final})}}$$

Using the ratio defined by Eq. (6-4), 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 42, below, but be careful with the units of s .

ALGORITHM 42: Minimum Combined Plane Change

$(i_{initial}, r_{initial}, v_{initial}, v_{trans_a}, v_{trans_b}, r_{final}, v_{final}, \Delta i \Rightarrow \Delta i_{initial}, \Delta i_{final})$

$$\left\{ \begin{array}{ll} \text{Iterate:} & \text{SIN}(s\Delta i) = \frac{\Delta v_a v_{final} v_{trans_b} \text{SIN}((1-s)\Delta i)}{\Delta v_b v_{initial} v_{trans_a}} \\ \text{or Estimate:} & R = \frac{r_{final}}{r_{initial}} \quad s \approx \frac{1}{\Delta i} \text{TAN}^{-1} \left[\frac{\text{SIN}(\Delta i)}{R^{3/2} + \cos(\Delta i)} \right] \\ \Delta i_{initial} = s\Delta i & \Delta i_{final} = (1-s)\Delta i \\ \Delta v_{initial} = \sqrt{v_{initial}^2 + v_{trans_a}^2 - 2v_{initial}v_{trans_a}\cos(\Delta i_{initial})} & \Delta v = \Delta v_{initial} + \Delta v_{final} \\ \Delta v_{final} = \sqrt{v_{final}^2 + v_{trans_b}^2 - 2v_{final}v_{trans_b}\cos(\Delta i_{final})} & \end{array} \right.$$

Table 6-3 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 6-3. 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}	Δi°	Δi_a°	Δi_b°	$\Delta v_{total} \text{ m/s}$
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

6.5.2 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, but if the fixed- Δv Δi isn't acceptable, another motor may be required. 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. (6-31) 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:

$$\cos(\Delta i_1) = \frac{v_{initial}^2 + v_{trans_a}^2 - \Delta v^2}{2v_{initial}v_{trans_a}} \quad (6-32)$$

Compare this change in inclination with the *optimum* amount of an inclination change for the first burn ($\Delta i_{initial}$ in Algorithm 42). 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. 6-14, 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.

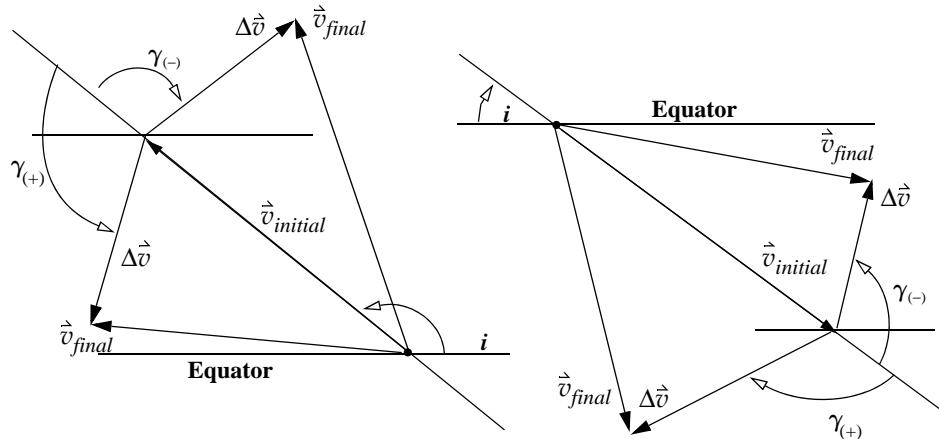


Figure 6-14. Geometry for Fixed- Δv Maneuvers. This figure includes changes to the retrograde ascending-node (left) and direct descending-node (right). I've shown two inclination changes (+ and -) for each case. Determine the proper sign for the payload angle, γ , based on the sign of the inclination 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 43: Fixed- Δv Maneuvers ($v_{initial}, v_{final}, v_{trans_a}, \Delta v \Rightarrow \gamma, \Delta i_1$)

$$\cos(\gamma) = -\frac{v_{initial}^2 + \Delta v^2 - v_{final}^2}{2v_{initial}\Delta v}$$

$$\begin{array}{ll} \Delta i - \text{decreasing} & -180^\circ \leq \gamma \leq 0^\circ \\ \Delta i + \text{increasing} & 0^\circ < \gamma < 180^\circ \end{array}$$

$$\cos(\Delta i_1) = \frac{v_{initial}^2 + v_{trans_a}^2 - \Delta v^2}{2v_{initial}v_{trans_a}}$$

Example 6-7 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. 7.6). Example 6-7 shows the complete process.

▼ Example 6-7. 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} , γ_a , γ_b

Find the radii for the orbits along with the transfer semimajor axis:

$$r_{initial} = \frac{6378.137 + 191}{6378.137} = 1.029\,946 \text{ ER} = 6569.137 \text{ km}$$

$$r_{final} = \frac{6378.137 + 35,780}{6378.137} = 6.609\ 789 \text{ ER} = 42,158.137 \text{ km}$$

$$a_{trans} = \frac{r_{initial} + r_{final}}{2} = \frac{1.029\ 94 + 6.609\ 789}{2} = 3.819\ 867 \text{ ER} = 24,363.637 \text{ km}$$

First, determine the velocities:

$$v_{initial} = \sqrt{\frac{\mu}{r_{initial}}} = \sqrt{\frac{1}{1.0299}} = 0.985\ 355\ 1 \text{ ER/TU} = 7.789\ 592 \text{ km/s}$$

$$v_{final} = \sqrt{\frac{\mu}{r_{final}}} = \sqrt{\frac{1}{6.6098}} = 0.388\ 961 \text{ ER/TU} = 3.074\ 880 \text{ km/s}$$

$$v_{trans_a} = \sqrt{\frac{2\mu}{r_{initial}} - \frac{\mu}{a_{trans}}} = \sqrt{\frac{2(1)}{1.0299} - \frac{1}{3.819}} = 1.296\ 171 \text{ ER/TU} = 10.2467 \text{ km/s}$$

$$v_{trans_b} = \sqrt{\frac{2\mu}{r_{final}} - \frac{\mu}{a_{trans}}} = \sqrt{\frac{2(1)}{6.6098} - \frac{1}{3.819}} = 0.201\ 971 \text{ ER/TU} = 1.5966 \text{ km/s}$$

Then, determine how much to change the inclination with the first burn with Algorithm 42. The orbital ratio is $R = 6.417\ 724$.

$$s \equiv \frac{1}{\Delta i} \tan^{-1} \left(\frac{\sin(\Delta i)}{R^{3/2} + \cos(\Delta i)} \right) = \frac{1}{-28.5} \tan^{-1} \left(\frac{\sin(-28.5^\circ)}{6.417\ 724^{3/2} + \cos(-28.5)} \right) = 0.055\ 962$$

$$\Delta i_1 = s\Delta i = (0.055\ 962)(-28.5^\circ) = -1.594\ 920^\circ$$

$$\Delta i_2 = (1-s)\Delta i = (1 - 0.055\ 962)(-28.5^\circ) = -26.905\ 080^\circ$$

Now, find the change in velocities:

$$\begin{aligned} \Delta v_a &= \sqrt{v_{initial}^2 + v_{trans_a}^2 - 2v_{initial}v_{trans_a} \cos(\Delta i_1)} = 0.312\ 404 \text{ ER/TU} \\ &= 2.469\ 669 \text{ km/s} \end{aligned}$$

$$\begin{aligned} \Delta v_b &= \sqrt{v_{final}^2 + v_{trans_b}^2 - 2v_{final}v_{trans_b} \cos(\Delta i_2)} = 0.227\ 974 \text{ ER/TU} \\ &= 1.802\ 216 \text{ km/s} \end{aligned}$$

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). If we assume that the booster can provide the calculated change in velocity, we can determine the orientation (payload angles) of the burns using Algorithm 43. Note that this is usually *not* the case. If the fixed- Δv doesn't match 4721 m/s, another analysis, or motor will be needed.

$$\cos(\gamma_a) = -\frac{v_{initial}^2 + \Delta v_a^2 - v_{trans_a}^2}{2v_{initial}\Delta v_a} = -\frac{0.985\ 355^2 + 0.312\ 404^2 - 1.296\ 729^2}{2(0.985\ 355)0.312\ 404}$$

$$\cos(\gamma_b) = -\frac{v_{trans_b}^2 + \Delta v_b^2 - v_{final}^2}{2v_{trans_b}\Delta v_b} = -\frac{0.201\ 971^2 + 0.227\ 974^2 - 0.388\ 961^2}{2(0.201\ 971)0.227\ 974}$$



$$\gamma_a = 6.6313^\circ, \text{ and } \gamma_b = 50.5392^\circ$$

6.6 Circular Rendezvous

After determining the Δv required to do an orbit maneuver, we 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. 7.6 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. Suppose a communications firm is 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.

6.6.1 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. 6-15. A sign convention is needed. 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 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,

$$\omega_{tgt} = \sqrt{\frac{\mu}{a_{tgt}^3}} = n \quad (6-33)$$

In the time the interceptor revolves once in the phasing orbit, the target completes one revolution *minus* the initial separation, or ϑ . The sign of ϑ is important. If we remember the sign convention for the phase angle, the formula is

$$\tau_{phase} = \Delta t = t_1 - t_0 = \frac{k_{tgt}(2\pi) + \vartheta}{\omega_{tgt}} \quad (6-34)$$

where multiple revolutions of the target satellite ($k_{tgt} = 1, 2, \dots$) are used to obtain solutions.

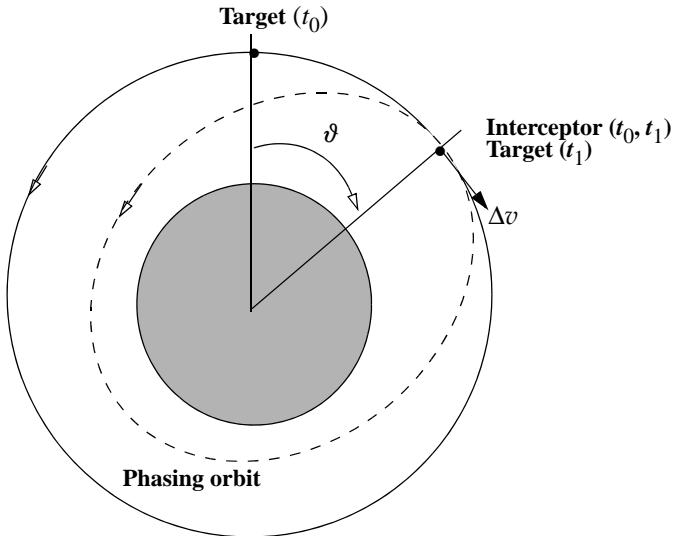


Figure 6-15. Circular Coplanar Rendezvous (Interceptor Trailing Target). Because the interceptor trails the target, the phase angle, ϑ , is negative. To rendezvous, the interceptor must move into a lower orbit with a shorter period to make up for the initial time displacement.

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. (1-26)] for the semimajor axis:

$$a_{phase} = \left(\mu \left(\frac{\tau_{phase}}{k_{int}(2\pi)} \right)^2 \right)^{1/3} \quad (6-35)$$

Notice a_{phase} must be large enough so the perigee radius is greater than the radius of the Earth. Multiple revolutions of the interceptor ($k_{int} = 1, 2, \dots$) 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 direc-

tions. 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,

$$\Delta v = 2|v_{phase} - v_{ini}| = 2 \left| \sqrt{\frac{2\mu}{a_{tgt}}} - \sqrt{\frac{\mu}{a_{phase}}} \right| \quad (6-36)$$

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. 6-16. 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. (6-34). You can find the semimajor axis of the phasing orbit using Eq. (6-35) and the change in velocity using Eq. (6-36).

If time permits, you can slightly improve Eq. (6-36). 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 .

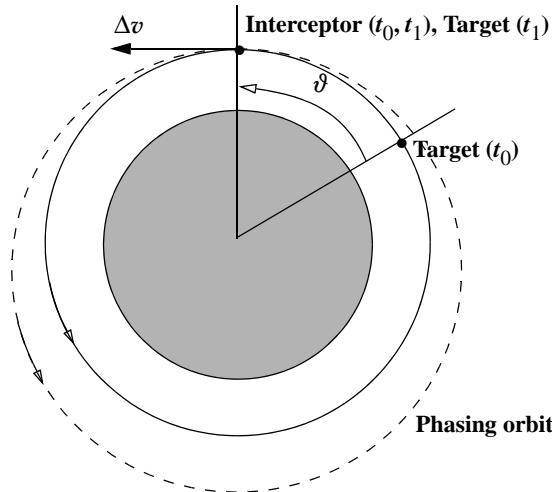


Figure 6-16. Circular Coplanar Rendezvous (Interceptor Leading Target). When the interceptor leads the target, the phase angle is positive, and the phasing orbit must be longer than the target orbit to permit the target satellite to *catch up*.

The basic idea is to adjust the phasing orbit as little as possible and make several revolutions in the phasing orbit until rendezvous occurs. The semimajor axis of the phasing orbit depends on the available time, so constraints on the phasing time drive this process. For instance, if time permits, you could attempt two phasing revolutions ($k_{int} = 2$).

In Fig. 6-16 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. 6-17. 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):

$$\tau_{trans} = \pi \sqrt{\frac{a_{trans}^3}{\mu}}$$

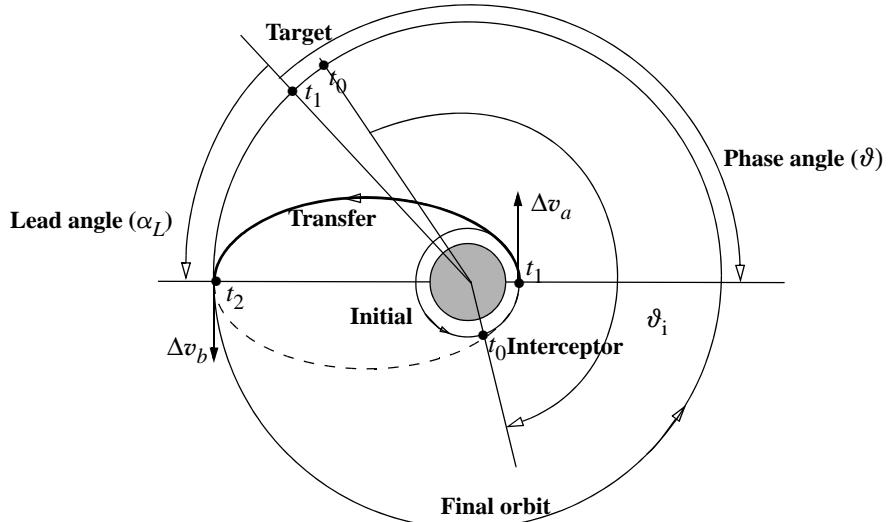


Figure 6-17. Circular Coplanar Rendezvous (Different Orbits). Whenever the satellites are in different orbits, we must calculate a lead angle (and therefore a phase angle). Also, the initial separation of the satellites (at t_0) usually differs from the required phasing for transfer, thus the symbol 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

$$\alpha_L = \omega_{tgt} \tau_{trans} \quad (6-37)$$

Then, find the required phase angle, ϑ , from the target to the interceptor. Remember that this must be positive.

$$\vartheta = \alpha_L - \pi \quad (6-38)$$

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,

$$\tau_{\text{wait}} = \frac{\vartheta - \vartheta_i + 2\pi k}{\omega_{\text{int}} - \omega_{\text{tgt}}} \quad (6-39)$$

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 $\frac{2\pi}{\omega_{\text{int}} - \omega_{\text{tgt}}}$ is the time during which the relative geometry repeats. This repeating time is 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. (6-39) 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 44 and 45 summarize these techniques, and a few examples illustrate the possibilities of changes to coplanar orbits.

ALGORITHM 44: Circular Coplanar Phasing (Same Orbits)

$$(a_{\text{tgt}}, \vartheta, k_{\text{tgt}}, k_{\text{int}} \Rightarrow \tau_{\text{phase}}, \Delta v, a_{\text{phase}})$$

$$\begin{aligned} \omega_{\text{tgt}} &= \sqrt[3]{\frac{\mu}{a_{\text{tgt}}^3}} \\ \tau_{\text{phase}} &= \frac{2\pi k_{\text{tgt}} + \vartheta}{\omega_{\text{tgt}}} \\ a_{\text{phase}} &= \left(\mu \left(\frac{\tau_{\text{phase}}}{2\pi k_{\text{int}}} \right)^2 \right)^{1/3} \end{aligned}$$

$$\Delta v = 2 \left| \sqrt{\frac{2\mu}{a_{\text{tgt}}}} - \sqrt{\frac{\mu}{a_{\text{phase}}}} \right|$$

ALGORITHM 45: Circular Coplanar Phasing (Different Orbits)

$$(\vartheta_i, a_{int}, a_{tgt}, k \Rightarrow \tau_{trans}, \Delta v, a_{trans})$$

$$\omega_{tgt} = \sqrt{\frac{\mu}{a_{tgt}^3}} \quad \omega_{int} = \sqrt{\frac{\mu}{a_{int}^3}}$$

$$a_{trans} = \frac{a_{int} + a_{tgt}}{2} \quad \tau_{trans} = \pi \sqrt{\frac{a_{trans}^3}{\mu}}$$

$$\alpha_L = \omega_{tgt} \tau_{trans} \quad \vartheta = \alpha_L - \pi$$

$$\tau_{wait} = \frac{\vartheta - \vartheta_i + 2\pi k}{\omega_{int} - \omega_{tgt}}$$

$$\Delta v = \left| \sqrt{\frac{2\mu}{a_{int}}} - \sqrt{\frac{\mu}{a_{trans}}} \right| + \left| \sqrt{\frac{2\mu}{a_{tgt}}} - \sqrt{\frac{\mu}{a_{trans}}} \right|$$

▼ Example 6-8. 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. 6-16 for the general geometry. First, convert the input radii to canonical units:

$$a_{int} = a_{tgt} = 12,756.274 / 6378.137 = 2.0 \text{ ER} = 12,756.274 \text{ km}$$

$$\omega_{tgt} = \sqrt{\frac{1}{2^3}} = 0.35355 \text{ rad/TU} = 0.020801 \text{ rad/s}$$

Let the target satellite complete one more orbit ($k_{tgt} = 1$) because the initial separation is small. Thus,

$$\tau_{phase} = \frac{2\pi k_{tgt} + \vartheta}{\omega_{int}} = \frac{2\pi - 20(\pi/180)}{0.35355} = 16.8742 \text{ TU} = 225.6950 \text{ min}$$

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.

$$a_{phase} = \left(\mu \left(\frac{\tau_{phase}}{2\pi k_{int}} \right)^2 \right)^{1/3} = 1.9252 \text{ ER} = 12,279.3332 \text{ km}$$

Next, find the initial change in velocity for the interceptor to enter the higher orbit. Note the negative sign indicates firing opposite the velocity vector.

$$\Delta v_{initial} = \left| \sqrt{\frac{2\mu}{a_{int}}} - \sqrt{\frac{\mu}{a_{phase}}} \right| = 0.693238 - 0.707107 = -0.013868 \text{ ER/TU} =$$

$$-0.109634 \text{ km/s}$$

The total Δv to rendezvous will be twice this amount, or

$$\Delta v = 2(\Delta v_{initial}) = -0.027\ 734 \text{ ER/TU} = -0.219\ 269 \text{ km/s}$$

Now evaluate the solution for the case in which the target can complete two revolutions, $k_{tgt} = 2$. The phasing time is

$$\tau_{phase} = \frac{2\pi k_{tgt} + \vartheta}{\omega_{int}} = \frac{2(2\pi) - 20(\pi/180)}{0.353\ 55} = 34.5558 \text{ TU} = 464.6661 \text{ min}$$

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.

$$a_{phase} = \left(\mu \left(\frac{\tau_{phase}}{2\pi k_{int}} \right)^2 \right)^{1/3} = 1.9628 \text{ ER} = 12,518.9393 \text{ km}$$

$$\Delta v_{initial} = -0.0067 \text{ ER/TU} = -0.0532 \text{ km/s}$$

$$\Delta v = 2(\Delta v_{initial}) = -0.0135 \text{ ER/TU} = -0.106\ 481 \text{ km/s}$$

▲ Notice the large reduction in velocity when you use two phasing orbits.

▼ Example 6-9. Velocity Change for a Rendezvous from Different Orbits.

GIVEN: $a_{int} = 12,756.274 \text{ km}$, $a_{tgt} = 42,159.48 \text{ km}$, $\vartheta_i = -20^\circ$ (target behind interceptor)

FIND: Δv

First convert the radii to canonical units:

$$a_{int} = 12,756.274 / 6378.137 = 2 \text{ ER}, a_{tgt} = 42,159.48 / 6378.137 = 6.61 \text{ ER}$$

Now find the mean motions, using the radii as the semimajor axis because the orbits are circular:

$$\omega_{tgt} = \sqrt{\frac{\mu}{a_{tgt}^3}} = \sqrt{\frac{1}{6.61^3}} = 0.058\ 843 \text{ rad/TU} = 0.003\ 46 \text{ rad/s}$$

$$\omega_{int} = \sqrt{\frac{\mu}{a_{int}^3}} = \sqrt{\frac{1}{2^3}} = 0.353\ 553 \text{ rad/TU} = 0.020\ 801 \text{ rad/s}$$

The semimajor axis of the phasing orbit is

$$a_{trans} = \frac{a_{int} + a_{tgt}}{2} = \frac{2 + 6.61}{2} = 4.305 \text{ ER} = 27,457.877 \text{ km}$$

The time required for the transfer is half the period of the Hohmann transfer, or

$$\tau_{trans} = \pi \sqrt{\frac{a_{trans}^3}{\mu}} = \pi \sqrt{\frac{4.305^3}{1}} = 28.1072 \text{ TU} = 377.9537 \text{ min}$$

The lead angle the target satellite must have for the transfer to take place is

$$\alpha_L = \omega_{tgt} \tau_{trans} = 0.058\ 843(28.1072) = 94.608\ 48^\circ$$

The phasing angle for the target satellite, when it is in the required position, is (it must be positive)

$$\vartheta = \alpha_L - \pi = 94.608\ 48^\circ - 180^\circ = -85.391\ 52^\circ \Rightarrow 94.608\ 48^\circ = 1.651\ 229 \text{ rad}$$

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,

$$\tau_{wait} = \frac{\vartheta - \vartheta_i + 2\pi k}{\omega_{int} - \omega_{tgt}} = \frac{-1.651\ 229 - 0.349\ 065\ 85 + 2\pi}{0.353\ 553 - 0.058\ 843} = 28.1072 \text{ TU} = 377.9537 \text{ min}$$

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:

$$\Delta v = \left| \sqrt{\frac{2\mu}{a_{int}} - \frac{\mu}{a_{trans}}} - \sqrt{\frac{\mu}{a_{int}}} \right| + \left| \sqrt{\frac{2\mu}{a_{tgt}} - \frac{\mu}{a_{trans}}} - \sqrt{\frac{\mu}{a_{tgt}}} \right| =$$

▲ $\left| \sqrt{\frac{2(1)}{2} - \frac{1}{4.305}} - \sqrt{\frac{1}{2}} \right| + \left| \sqrt{\frac{2(1)}{6.61} - \frac{1}{4.305}} - \sqrt{\frac{1}{6.61}} \right| = 0.292\ 928\ 9 \text{ ER/TU} = 2.315\ 71 \text{ km/s}$

6.6.2 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. 6-18.

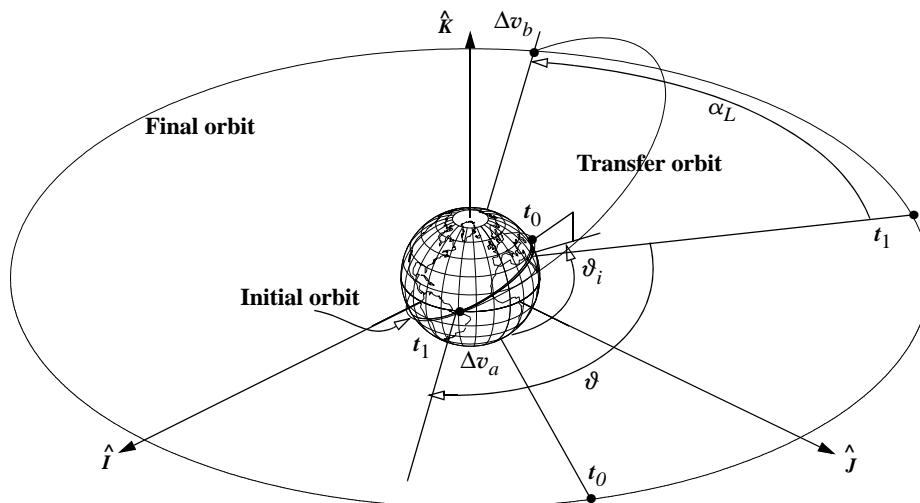


Figure 6-18. Geometry for a Circular, Noncoplanar, Orbit Transfer. This transfer combines the Hohmann transfer from LEO to GEO with the change in inclination plane. The maneuver can't start until the interceptor is at a node, t_1 , and the target, t_1 , is at the required lead angle. Because both satellites 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.

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. 6-19. 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 angle is correct to start the Hohmann-transfer orbit. But we must make sure the phasing orbit doesn't hit the Earth!

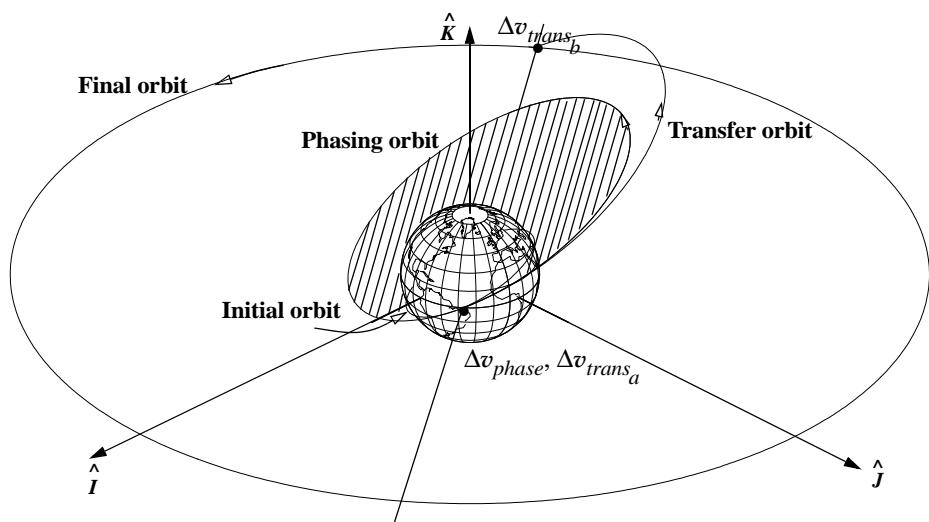


Figure 6-19. Alternate Geometry for a Circular, Noncoplanar, Orbit Transfer. An intermediate phasing orbit is useful to permit a quicker time to rendezvous. The size of the phasing orbit is calculated so the satellite returns to the node in time for the transfer into the final orbit. Both burns (Δv_{phase} and Δv_{trans_a}) occur at the same point.

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. 6-19. 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 semimajor 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. (6-37).

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:

$$\Delta t_{node} = \frac{\Delta\vartheta_{int}}{\omega_{int}} \quad (6-40)$$

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,

$$\lambda_{true1} = \lambda_{true0} + \omega_{tgt}\Delta t_{node}$$

Determine the new phase angle, ϑ_{new} , based on the new locations of both satellites at time t_1 . This often involves using Eq. (2-92) 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

$$\rho_{phase} = \frac{\alpha_{new} - \alpha_L + 2\pi k_{tgt}}{\omega_{tgt}} \quad (6-41)$$

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. (6-35).

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 46: Noncoplanar Phasing

$$(\vartheta_i, a_{int}, a_{tgt}, k_{tgt}, u_{int}, \Omega_{int}, \lambda_{true0}, \Delta i \Rightarrow$$

$$\tau_{trans}, \tau_{phase}, \Delta v_{phase}, \Delta v_{trans1}, \Delta v_{trans2}, a_{phase})$$

$$\omega_{tgt} = \sqrt{\frac{\mu}{a_{tgt}^3}} \quad \omega_{int} = \sqrt{\frac{\mu}{a_{int}^3}}$$

$$a_{trans} = \frac{a_{initial} + a_{final}}{2}$$

$$\tau_{trans} = \pi \sqrt{\frac{a_{trans}^3}{\mu}}$$

$$\alpha_L = \omega_{tgt} \tau_{trans}$$

Find $\Delta\vartheta_{int}$ to reach a node (180° or $360^\circ - u_{int}$)

$$\Delta t_{node} = \frac{\Delta\vartheta_{int}}{\omega_{int}}$$

$$\lambda_{true_{tgt1}} = \lambda_{true0} + \omega_{tgt} \Delta t_{node}$$

Find λ_{true} for interceptor at t_1 $\lambda_{true_{int1}} = \Omega + \pi$

$$\vartheta_{new} = \lambda_{true_{int1}} - \lambda_{true_{tgt1}}$$

$$\alpha_{new} = \pi + \vartheta_{new}$$

$$\begin{aligned}\beta_{phase} &= \frac{\alpha_{new} - \alpha_L + 2\pi k_{tgt}}{\omega_{tgt}} \\ a_{phase} &= \left(\mu \left(\frac{\beta_{phase}}{k_{int} 2\pi} \right)^2 \right)^{1/3} \\ \Delta v_{phase} &= \left| \sqrt{\frac{2\mu}{a_{int} - a_{phase}}} - \sqrt{\frac{\mu}{a_{int}}} \right| \\ \Delta v_{trans_1} &= \left| \sqrt{\frac{2\mu}{a_{int} - a_{trans}}} - \sqrt{\frac{2\mu}{a_{int} - a_{phase}}} \right| \\ \Delta v_{trans_2} &= \sqrt{\left(\frac{2\mu}{a_{tgt} - a_{trans}} \right) + \left(\frac{\mu}{a_{tgt}} \right)} - 2 \sqrt{\frac{2\mu}{a_{tgt} - a_{trans}}} \sqrt{\frac{\mu}{a_{tgt}}} \cos(\Delta i) \\ \tau_{total} &= 2\pi \sqrt{\frac{a_{phase}^3}{\mu}} + \tau_{trans} + \Delta t_{node}\end{aligned}$$

Because this problem is rather involved, let's look at an example to show the operation.

▼ Example 6-10. Velocity Change for a Rendezvous Using Noncoplanar Orbits.

Consider the following problem of two circular orbits, as illustrated in Fig. 6-20.

Interceptor	Target
$a = 7143.51 \text{ km} = 1.12 \text{ ER}$	$a = 42,159.4855 \text{ km} = 6.61 \text{ ER}$
$i = 28.5^\circ$	$i = 0$
$\Omega = 45^\circ$	$\lambda_{true} = 200^\circ$
$u = 15^\circ$	

First, find the desired lead angle, α_L for the final transfer, and some initial quantities:

$$\begin{aligned}\omega_{tgt} &= \frac{\sqrt{\mu}}{\sqrt{a_{tgt}^3}} = \frac{0.058\,843 \text{ rad/TU}}{0.003\,462 \text{ rad/s}} & a_{trans} &= \frac{1.12 + 6.61}{2} = \frac{3.865 \text{ ER}}{24,651.498 \text{ km}} \\ \omega_{int} &= \frac{\sqrt{\mu}}{\sqrt{a_{int}^3}} = \frac{0.843\,67 \text{ rad/TU}}{0.049\,636 \text{ rad/s}} & \tau_{trans} &= \pi \sqrt{\frac{a_{trans}^3}{\mu}} = \frac{23.871 \text{ TU}}{320.992 \text{ min}} \\ \alpha_L &= \omega_{tgt} \tau_{trans} \frac{180}{\pi} = 80.481^\circ\end{aligned}$$

To find α_i , 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

$$\Delta t_{node} = \frac{\Delta\vartheta_{int}}{\omega_{int}} = \frac{165^\circ \left(\frac{\pi}{180} \right)}{0.843\,67} = 3.4134 \text{ TU} = 45.8996 \text{ min}$$

Now, how far has the target moved in this length of time?

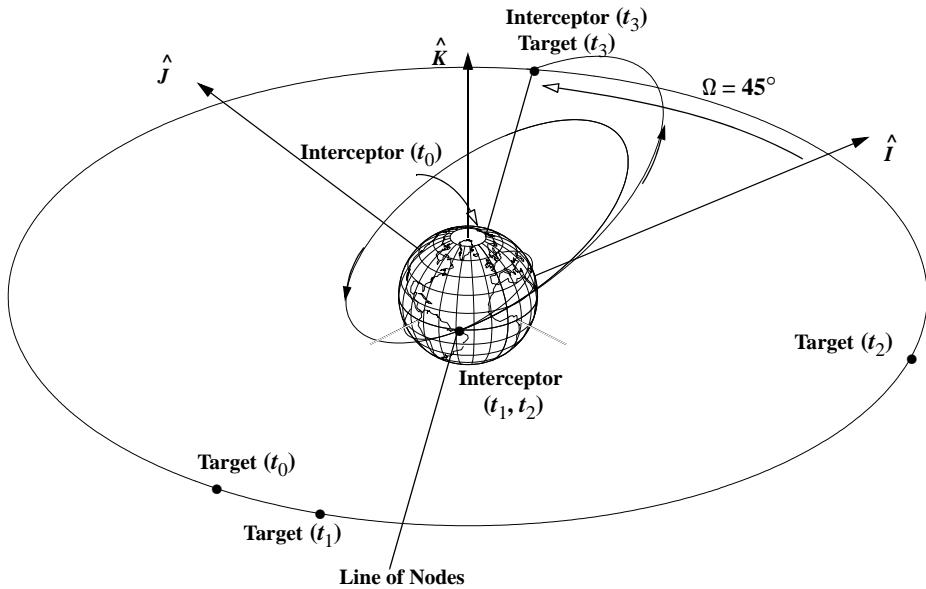


Figure 6-20. Geometry for the Noncoplanar Orbit Transfer in Example 6-10. The interceptor starts in a 28.5° inclined orbit, behind the Earth, whereas the final orbit is equatorial. The phasing orbit is designed to place the interceptor at the node at the proper time for transfer to the final orbit.

$$\omega_{tgt} \Delta t_{node} = (0.0588434)(3.4134) \left(\frac{180}{\pi} \right) = 11.5^\circ$$

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 $u = 180^\circ$, so $\lambda_{int1} = 45 + 180^\circ$. From Fig. 6-20, 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 ($k_{tgt} = 0$):

$$\rho_{phase} = \frac{\alpha_{new} - \alpha_L + 2\pi k_{tgt}}{\omega_{tgt}} = \frac{193.5^\circ - 80.48^\circ}{0.0588434} = 33.5196 \text{ TU} = 450.7328 \text{ min}$$

As a check, the target is now at $(0.05884)(33.5224)(180/\pi) = 113.019^\circ$. If the interceptor revolves once in a phasing orbit ($k_{int} = 1$),

$$a_{phase} = \left(\mu \left(\frac{\rho_{phase}}{k_{int} 2\pi} \right)^2 \right)^{1/3} = 3.0533 \text{ ER} = 19,473.304 \text{ km}$$

Check $a_{int} < a_{phase} < a_{tgt}$. Finally, calculate the velocities required for each maneuver. The initial velocities are v_{int} and v_{tgt} . Note that we do not consider the optimum inclination change (Sec. 6.5.1) in which we would split the plane change between two burns. Here we consider all the inclination change to occur on the second and final maneuver.

$$v_{int} = \sqrt{\frac{\mu}{a_{int}}} = 0.944\ 91 \text{ ER/TU} = 7.4699 \text{ km/s and}$$

$$v_{tgt} = \sqrt{\frac{\mu}{a_{tgt}}} = 0.388\ 95 \text{ ER/TU} = 3.0748 \text{ km/s}$$

Also, find v_{phase} , v_{trans1} , and v_{trans2} :

$$v_{phase} = \sqrt{\frac{2\mu}{a_{int}} - \frac{\mu}{a_{phase}}} = 1.207\ 552 \text{ ER/TU} = 9.5461 \text{ km/s}$$

$$v_{trans1} = \sqrt{\frac{2\mu}{a_{int}} - \frac{\mu}{a_{trans}}} = 1.235\ 711 \text{ ER/TU} = 9.7687 \text{ km/s}$$

$$v_{trans2} = \sqrt{\frac{2\mu}{a_{tgt}} - \frac{\mu}{a_{trans}}} = 0.209\ 379 \text{ ER/TU} = 1.6552 \text{ km/s}$$

Therefore, to find the Δv s, use

$$\Delta v_{phase} = |v_{phase} - v_{int}| = 0.262\ 641 \text{ ER/TU} = 2.0762 \text{ km/s}$$

$$\Delta v_{trans1} = |v_{trans1} - v_{phase}| = 0.028\ 159 \text{ ER/TU} = 0.222\ 608 \text{ km/s}$$

$$\Delta v_{trans2}^2 = v_{trans2}^2 + v_{tgt}^2 - 2v_{trans2}v_{tgt}\cos(\Delta i)$$

$$\Delta v_{trans2} = 0.228\ 003 \text{ ER/TU} = 1.802\ 449 \text{ km/s}$$

Finally, the total time is

$$\tau_{total} = \tau_{phase} + \tau_{trans} + \Delta\tau_{node} = 33.524 + 23.871 + 3.413 = 60.8042 \text{ TU} = 817.624 \text{ min}$$

and the total Δv is

▲ $\Delta v_{TOTAL} = \Delta v_{phase} + \Delta v_{trans1} + \Delta v_{trans2} = 0.518\ 803 \text{ ER/TU} = 4.101\ 330 \text{ km/s.}$

6.7 Continuous-Thrust Transfers

For most maneuvers discussed so far, we've assumed the use of *impulsive* thrust. Although impulsive models approximate most maneuvers well, realistic burns require finite time. In addition, we now have the technology to thrust continuously, so we need to model 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 continuous-thrust to raise or resize an orbit, and (2) exploring low-continuous-thrust maneuvers to simultaneously change orbital inclination and size (combined maneuvers in Sec. 6.5).

It's beneficial for you to be familiar with *Hamiltonian* formulations, *Lagrange multipliers*, and *optimal-control theory*, because this section goes beyond 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 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 when the constraints are satisfied. This rarely happens, but the partial derivatives give an indication of how to move from the 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).

6.7.1 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 continuous 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_{thrust}^* , 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 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*, Δv_{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.

Accumulated Velocity Change

The *accumulated velocity change*, Δv_{acc} , is the total velocity imparted by the force of thrust during the elapsed transfer time, t_f , so that

$$\Delta v_{acc} = \int_0^{t_f} \frac{a_{thrusti}}{1 + \dot{m}_{spec} t} dt \quad (6-42)$$

The *specific propellant mass flow rate*, \dot{m}_{spec} , is the actual mass flow rate divided by the initial mass; this value is negative when dealing with propellant loss. Eq. (6-42) is a variation of the *ideal rocket equation*, which is more commonly known as

$$\Delta v = g I_{sp} \ln \frac{m_{init}}{m_{init} - m_{prop}}$$

where m_{init} = the vehicle's initial mass

m_{prop} = propellant mass

g = gravity

I_{sp} = specific impulse

Eq. (6-42) models the effects of continuously expending fuel recasting the solution in terms of accumulated velocity change. The formula serves two purposes: it models the effect of fuel depletion, and it defines the relationship between time and thrust-induced velocity change through the equation that completes the integration in Eq. (6-42):

$$\Delta v_{acc} = \frac{a_{thrusti}}{\dot{m}_{spec}} \ln(1 + \dot{m}_{spec} t_f) \quad (6-43)$$

* Some common notations may be confusing, especially “ a ”. We use $a_{thrusti}$ for the initial acceleration due to thrust, a_{curr} for the current value of the semimajor axis, and a_{init}/a_{final} for the initial and final semimajor axes values.

For general solutions, we solve Eq. (6-43) to find the time of flight, t_f

$$t_f = \frac{1}{-\dot{m}_{spec}} \left(1 - \exp \left[\frac{\dot{m}_{spec} \Delta v_{acc}}{a_{thrusti}} \right] \right) \quad (6-44)$$

Recasting the minimum-time solution in terms of Δv_{acc} allows us to find the *propellant mass fraction*, p_{mf}

$$m_{prop} = m_{init} - m_{final} = -\dot{m} t_f$$

Given a_{acc} , p_{mf} , and Δv_{acc} , we can find the specific mass flow rate and transfer time.

$$\dot{m}_{spec} = \frac{a_{thrusti}}{\Delta v_{acc}} \ln(1 - p_{mf}) = \frac{\dot{m}}{m_{init}} \quad t_f = -\frac{m_{prop}}{\dot{m}} = \frac{p_{mf}}{\dot{m}_{spec}} \quad (6-45)$$

For the limiting case in which \dot{m}_{spec} and p_{mf} approach zero, the total transfer time from Eq. (6-42) is simply

$$t_f = \frac{\Delta v_{acc}}{a_{thrusti}} \quad (6-46)$$

Another useful relation specifies thrust (F) in kg m/s² (with a negative mass flow rate to indicate mass being lost)

$$F = -I_{sp} \dot{m} g \quad (6-47)$$

6.7.2 Orbit Raising

The first major continuous-thrust application is the ability to change an orbit's size, which is extremely common for any operation in space. We usually can't insert a payload directly into a geosynchronous orbit because it requires a very large booster. An appealing alternative is to launch the satellite into a low-Earth orbit and have a separate transfer booster place it into the correct final orbit. Yet another alternative is to use the same station keeping thrusters for the orbital transfer as well. Because only a single vehicle is used, this minimizes debris generation. 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. We'll first consider only changes to an orbit's size from circular orbits. Low thrust transfers from elliptical orbits require more complex analyses (primer vector theory, etc) and are beyond our current scope.

Equations of Motion

The differential equations defining 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 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

$$\begin{aligned}\dot{r} &= v_R & \dot{v}_R &= \frac{v_S^2}{r} - \frac{\mu}{r^2} + \frac{a_{thrusti} \sin(\phi_c)}{1 + \dot{m}_{spec} t} \\ \dot{v}_S &= -\frac{v_R v_S}{r} + \frac{a_{thrusti} \cos(\phi_c)}{1 + \dot{m}_{spec} t} & \dot{\theta}_p &= \frac{v_S}{r}\end{aligned}\quad (6-48)$$

where v_R = radial-velocity component (RSW system)
 v_S = transverse velocity component (RSW system)
 $a_{thrusti}$ = initial vehicle acceleration due to thrust
 ϕ_c = control angle (measured from the velocity vector to the thrust vector with positive values in the angular momentum direction)
 θ_p = angle denoting the vehicle's location referenced from the start of the maneuver

Optimal-Control Formula—Orbit Raising

Given an initial and final orbit 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 similar to the Hohmann transfer, except that for the use of nonimpulsive maneuvers. The control formula for this problem follows Bryson and Ho (1987:42–89), but I repeat 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. (6-48)]. Capture the behavior of the Lagrange multipliers, called λ *dynamics*, by

$$\begin{aligned}\dot{\lambda}_r &= -\lambda_{v_R} \left(-\frac{v_S^2}{r^2} + \frac{2\mu}{r^3} \right) - \lambda_{v_S} \left(\frac{v_R v_S}{r^2} \right) \\ \dot{\lambda}_{v_R} &= -\lambda_r + \lambda_{v_S} \left(\frac{v_S}{r} \right) \\ \dot{\lambda}_{v_S} &= -\lambda_{v_R} \left(\frac{2v_S}{r} \right) + \lambda_{v_S} \left(\frac{v_R}{r} \right)\end{aligned}$$

Because both semimajor axis and eccentricity are changing, the control parameter is operating in the pitch plane (See page 156), so we'll use ϕ_c . The partial of \mathcal{H} with respect to the control parameter (in-plane thrust control angle) must equal zero, so

$$\frac{\partial \mathcal{H}}{\partial \phi_c} = (\lambda_{v_R} \cos(\phi_c) - \lambda_{v_S} \sin(\phi_c)) \frac{a_{thrust}}{1 + \dot{m}_{spec} t} = 0$$

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 Using 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. 3.8 offer this type of scaling, but these canonical units are based on the initial satellite orbit. If r_i is the initial radial distance,

$$1 \text{ DU}^* = r_i \quad 1 \text{ TU}^* = \sqrt{\frac{r_i^3}{\mu}} \quad 1 \frac{\text{DU}^*}{\text{TU}^*} = \sqrt{\frac{\mu}{r_i}} \quad (6-49)$$

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, v_R, v_S) array for any circle-to-circle coplanar transfer are simply $\hat{r} = 1\hat{R} \text{ DU}^*$ and $\hat{v} = 1\hat{S} \text{ DU}^*/\text{TU}^*$. The final values are $\hat{r} = R\hat{R} \text{ DU}^*$ and $\hat{v} = \sqrt{1/R}\hat{S} \text{ DU}^*/\text{TU}^*$, where R is the ratio [recall Eq. (6-4)] of the final radial distance to the initial. This section considers only orbit raising ($R > 1.0$).

As mentioned before, 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*, p_{mf} , into accumulated velocity change and plot them against initial acceleration to produce Figs. 6-21 through 6-23. The ripples in the curves of Figs. 6-21 and 6-22 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.

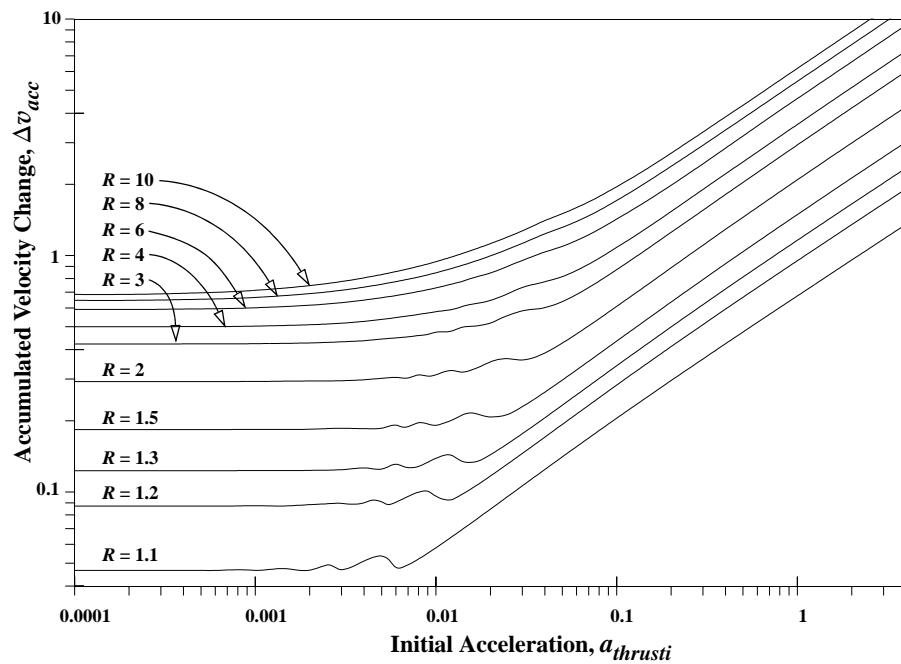


Figure 6-21. Δv_{acc} Contours for Various Orbit Ratios, R ($p_{mf} = 0.25$). Notice the various non-dimensional initial accelerations with their resulting accumulated velocity changes for different orbit ratios (final to initial). The ripples represent a transition between gravitational and thrust dominance, which induces a slight eccentricity in the orbit.

Figure 6-23 results from taking the optimal solutions for $p_{mf} = 0$. 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 eccentricity to meet the final condition of a circular orbit. The peaks in the curves

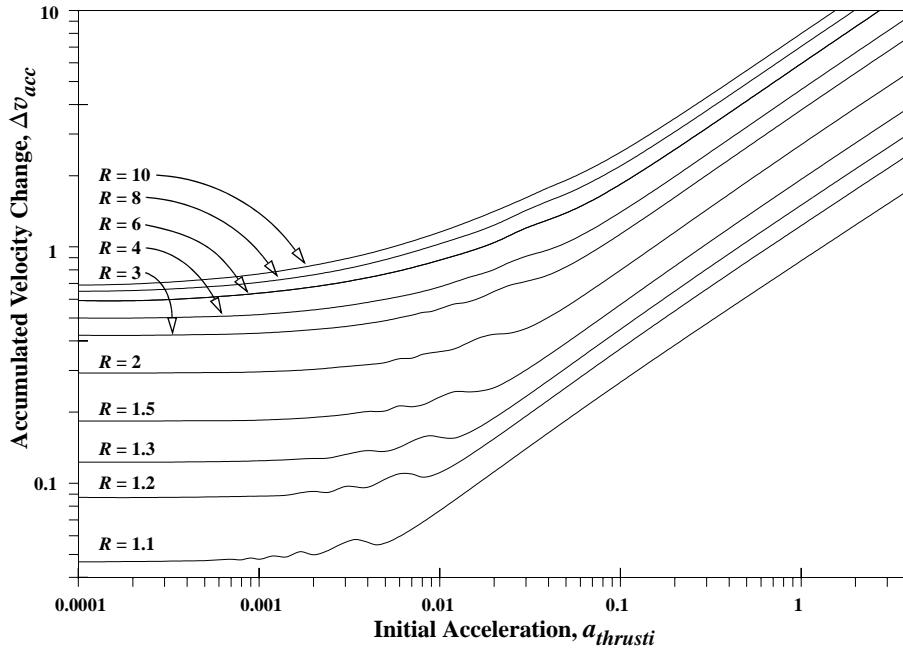


Figure 6-22. Δv_{acc} Contours for Various Orbit Ratios, R ($p_{mf} = 0.75$). The nondimensional initial accelerations versus velocity change curve is similar to Fig. 6-21, although the ripples occur at lower initial accelerations due to the increased mass loss.

reflect the added cost of this process. When the satellite orbits 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. 6-21, 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 $a_{thrusti} < 10^{-4}$ DU*/TU *2 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

$$\frac{dr}{dv} = 2 \sqrt{\frac{r^3}{\mu}}$$

Using the previously defined units and boundary conditions, we discover the accumulated velocity change for orbit-raising is simply (canonical units)

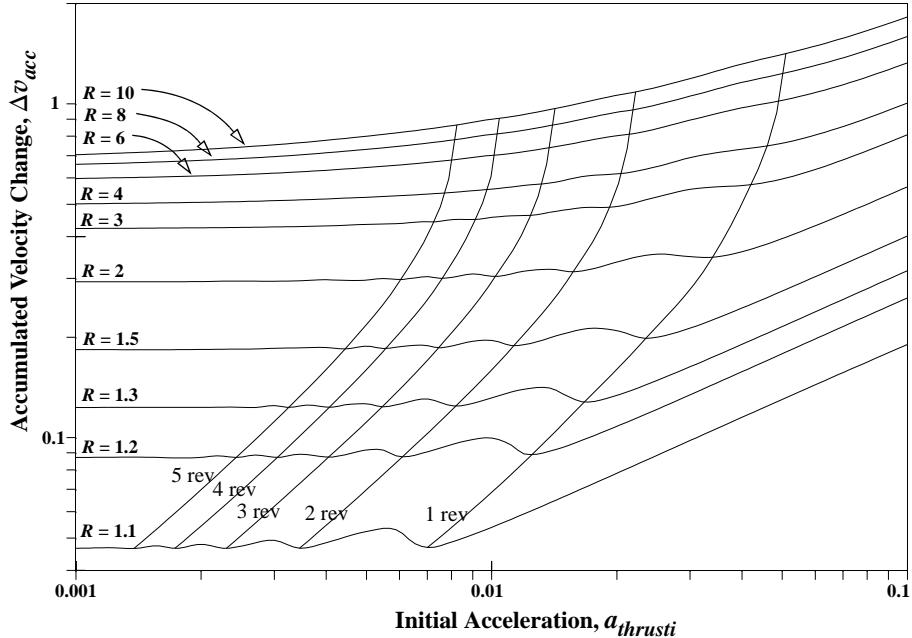


Figure 6-23. Δv_{acc} Contours for Various Orbit Ratios, R ($p_{mf} = 0$). The revolution lines are superimposed to show the penalty for not completing a revolution during the transfer. As we use more revolutions, the penalty diminishes, but for the first iteration, it can be significant. Mission planners should choose thrust levels which give polar angles that are multiples of 360° .

$$\Delta v_{acc} = 1 - \sqrt{\frac{1}{R}} \quad (6-50)$$

With this equation we can find the accumulated velocity change given the ratio of the desired orbital change.

We analytically determine the constant upward-sloping lines of Figs. 6-21 to 6-22 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 $a_{thrusti} > 4 \text{ DU}^*/\text{TU}^{*2}$ by the equation (canonical units)

$$\begin{aligned} \Delta v_{acc} &= 2\sqrt{(R-1)a_{thrusti}} \quad p_{mf} = 0 \\ \Delta v_{acc} &= -\ln(1-p_{mf}) \sqrt{\frac{(R-1)a_{thrusti}}{2-p_{mf}-2\sqrt{1-p_{mf}}}} \quad p_{mf} > 0 \end{aligned}$$

Then, find the final time from Eq. (6-45) or Eq. (6-46) and discover the switching time.

$$\begin{aligned} t_s &= 0.5t_f \quad p_{mf} = 0 \\ t_s &= \left(\frac{1 - \sqrt{1 - p_{mf}}}{p_{mf}} \right) t_f \quad p_{mf} > 0 \end{aligned}$$

Implementing Continuous-Thrust Solutions

By convention, the input parameters will be listed as an array $(r_i, r_f, \mu, a_{thrusti}, p_{mf})$. The total accumulated velocity should have two significant digits from the transfer charts; Six significant digits reduce round-off errors in all computations. See the table on the inside back cover for the physical constants.

To use the charts, begin with the input parameters. Scale these values into 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. (6-45) and Eq. (6-46) if they're needed. Let's look at an example.

▼ Example 6-11. Determining the Time Needed for a Transfer between Earth and Mars.

Consider a transfer between Earth and Mars, for which 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, thus 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, or a specific mass flow rate of $-1.29 \times 10^{-3} \text{ /day}$, or $-1.493 \ 06 \times 10^{-7} \text{ /s}$ ($0.074 \ 990 \text{ /TU}^*$). The initial acceleration, $a_{thrusti}$, is $8.331 \ 73 \times 10^{-7} \text{ km/s}^2$. We must design a trajectory to find the minimum transfer time. The problem is to find a p_{mf} that gives this result. By interpolating, the accuracy will depend on several factors including the "distance" of p_{mf} from known figures and the degree of linearity of the curve.

Begin with $p_{mf} = 0.25$. Find initial data from Table D-5 and Table D-3. The radius of the Earth from the Sun is $1.495 \ 98 \times 10^8 \text{ km}$ and that of Mars is $2.279 \ 39 \times 10^8 \text{ km}$. The gravitational parameter of the Sun is $1.327 \ 12 \times 10^{11} \text{ km}^3/\text{s}^2$ and a solar time unit is 58.132 days, or $5.023 \times 10^6 \text{ s}$. Converting to canonical units $(r_i, r_f, \mu, a_i, p_{mf})$: 1.0 DU^* , $1.523 \ 68 \text{ DU}^*$, $1.0 \text{ DU}^{*3}/\text{TU}^{*2}$, $0.1405 \text{ DU}^*/\text{TU}^{*2}$, 0.25 . For this problem, these units are identical to the "starred" canonical units because the radius of the Earth, r_i , is already unity. From Fig. 6-21, find $\Delta v_{acc} = 0.54 \text{ AU/TU}_\odot$. From Eq. (6-45),

$$\dot{m}_{spec} = \frac{a_{thrusti}}{\Delta v_{acc}} \ln(1 - p_{mf}) = \frac{0.1405}{0.540} \ln(1 - 0.25) = -0.074 \ 851 \text{ /TU}^*$$

Now let $p_{mf} = 0.75$ because we don't have a figure for $p_{mf} = 0.5$ —see Alfano and Thorne (1994) for the figure. From Fig. 6-22, find $\Delta v_{acc} = 0.66 \text{ AU/TU}_\odot$.

$$\dot{m}_{spec} = \frac{a_{thrusti}}{\Delta v_{acc}} \ln(1 - p_{mf}) = \frac{0.1405}{0.660} \ln(1 - 0.75) = -0.295 \ 126 \text{ /TU}^*$$

Now interpolate p_{mf} and the mass flow rate to match the given value ($0.074 \ 990 \text{ /TU}^*$).

$$\dot{m}_{spec} = 0.25 + (0.5)(0.000 \ 139 \ 6) / 0.453 \ 393 \ 4 = 0.250 \ 153 \ 9 \text{ /TU}^*$$

By Eq. (6-45), $t_{trans} = 0.250 \ 153 \ 9 / 0.074 \ 990 = 3.335 \ 82 \text{ TU}^* = (3.335 \ 82)(5.023 \times 10^6) = 1.675$

$583 \times 10^7 \text{ s} = 193.9332 \text{ days}$. The answer from Bryson and Ho (1975:68) is 193 days. Our answer is within two significant figures of the 193 day answer, so it is acceptable. Notice I've dropped the minus sign since it depends only on the convention for the sign of the mass flow rate.



▼ **Example 6-12. Transferring from LEO to GEO.**

This case involves a coplanar transfer from a circular, low-Earth parking orbit of about 318.9 km altitude (radius = 1.05 ER) to a final circular GEO orbit radius of 42159.5 km (6.61 ER). The vehicle's initial acceleration is $4.0 \times 10^{-3} \text{ m/s}^2$, with a propellant mass fraction of 0.25.

Initial data is $(r_i, r_f, \mu, a_{thrusti}, p_{mf})$ 1.05 ER, 6.61 ER, $1.0 \text{ ER}^3/\text{TU}^2$, $4.0 \times 10^{-3} \text{ km/s}^2$, 0.25

The "new" units are: 1.0 DU*, 6.295 24 DU*, $1.0 \text{ DU}^3/\text{TU}^2$, $4.500 79 \times 10^{-4} \text{ DU}^2/\text{TU}^2$, 0.25

From Eq. (6-50), find $\Delta v_{acc} = 0.601 44 \text{ DU}^2/\text{TU}$. Then by Eq. (6-45),

$$\begin{aligned}\dot{m}_{spec} &= \frac{a_{thrusti}}{\Delta v_{acc}} \ln(1 - p_{mf}) = \frac{4.501 \times 10^{-4}}{0.601 44} \ln(1 - 0.25) \\ &= -2.152 82 \times 10^{-4} / \text{TU}^2 = -2.48 \times 10^{-7} / \text{s} \\ t_f &= p_{mf} / \dot{m}_{spec} = 1161.267 \text{ TU}^2 = (1161.267)(868.0722) = 1.008 06 \times 10^6 \text{ s} = 11.667 \text{ days}\end{aligned}$$

▲ This case involves low thrust (flat curve portion for Fig. 6-21 to Fig. 6-23, so Δv_{acc} is constant for a given orbit ratio R regardless of p_{mf} or $a_{thrusti}$), and agrees with the numerical solution.

6.7.3 Low-Thrust, Noncoplanar Transfers

Several authors have studied low-thrust orbital transfers using minimum time. In this section we'll 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 many orbits). The *Alfano transfer* (1982) is presented—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, remember that thrust is treated as a perturbation. Chapters 8 and 9 detail the solution techniques for perturbations.

Short-Time-Scale Problem

Edelbaum (1961) first solved the optimal-control problem over the fast time scale of one orbit. For a low-thrust vehicle in a nearly circular orbit, the Lagrange planetary equations for semimajor axis [Eq. (9-12)] and inclination [Eq. (9-24)] describe the orbital change from perturbations. We'll also use *initial acceleration*, $a_{thrusti}$, and the *thrust-control angle*, ϕ_c , from the normal plane to the radius. ϕ_c is sometimes thought of as the *vehicle's yaw angle* depending on whether the thruster is fixed on the satellite or if it has a gimbal mechanism. The yaw direction results because the semimajor axis and inclination are changing. For small eccentricity, we don't consider any pitch component of the vehicle's acceleration.

The control law is exercised over the duration of the maneuver. We track this from the moment the engine switches on. Ignition occurs at the intersecting node when performing non-coplanar transfers, but may occur anywhere for orbit raising. The *control*

$\vartheta(\theta_p)$ tries to get the most inclination change over one orbit while still changing the semimajor axis by a given amount, Δa .

Assuming small changes over the orbit leads to

$$\delta \int_0^{2\pi} \left[\frac{a_{curr}^2 a_{thrusti}}{\mu} \sin(\phi_c) \cos(\theta_p) + \lambda \left(\frac{2a_{curr}^3 a_{thrusti}}{\mu} \cos(\phi_c) - \frac{\Delta a_{curr}}{2\pi} \right) \right] d\theta_p = 0$$

Optimizing yields the **control law angle** for the short-time-scale problem, ϕ_c :

$$\tan(\phi_c) = \frac{\cos(\theta_p)}{\sqrt{\frac{1}{c_v} - 1}} \quad (6-51)$$

Here, anticipating the next section, we've used the *control variable*, c_v , with an operative range from 0.0 to 1.0. Also note that c_v changes as the vehicle's semimajor axis changes. a_{curr} represents the amount the current semimajor axis value (canonical units).

$$c_v = \frac{1}{4\lambda^2 a_{curr}^2 + 1} \quad (6-52)$$

The changes in semimajor axis and inclination per revolution are (canonical units)

$$\begin{aligned} \Delta a_{rev} &= \frac{8a_{curr}^3 a_{thrusti}}{\mu} \sqrt{1 - c_v} K(c_v) \\ \Delta i_{rev} &= \frac{4a_{curr}^2 a_{thrusti}}{\mu} \left[\frac{1}{\sqrt{c_v}} E(c_v) + \left(\sqrt{c_v} - \frac{1}{\sqrt{c_v}} \right) K(c_v) \right] \end{aligned}$$

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 observations are relevant. The net change in eccentricity and node per orbit is zero with this control program; therefore, circular orbits stay nearly circular (maximum eccentricity of about 0.01). The control law angle ϕ_c varies from a pure in-track direction 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. (1-26)] (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 using fuel, which (for constant thrust) cause the vehicle's acceleration to vary as

$$a_{thrust}(t) = \frac{a_{thrusti}}{1 + \dot{m}_{spec}\Delta t}$$

This last effect introduces explicit time dependence into the control Hamiltonian, but we can eliminate it by introducing the new independent variable τ and by letting $d\tau = a_{thrust}(t)dt$. This leads to the time transformation already presented in Eq. (6-43) and Eq. (6-44), where $\Delta v_{acc} = \tau$. Physically, Δv_{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

$$\frac{\Delta a}{\Delta \tau} = \frac{4}{\pi \sqrt{\mu}} \sqrt{\frac{a_{curr}^3}{1 - c_v}} K(c_v) \quad \frac{\Delta i}{\Delta \tau} = \frac{2}{\pi \sqrt{\mu}} \left[\frac{1}{\sqrt{c_v}} E(c_v) + \left(\sqrt{c_v} - \frac{1}{\sqrt{c_v}} \right) K(c_v) \right]$$

In this form the equations of motion are free of the values of $a_{thrusti}$ and \dot{m} for a particular vehicle and are also free of the variable τ .

We can now pose the long-time-scale problem—to transfer from given initial values of a_{init} and i_{init} to final values of a_{final} and i_{final} using the minimum change in velocity, Δv_{acc} . At the same time, elapsed time and fuel consumption must be minimized. This is a standard problem: optimal control in minimum time (Bryson and Ho, 1987:76–89). The *control Hamiltonian*, \mathcal{H}_c , is

$$\mathcal{H}_c = 1 + \lambda_a \frac{4}{\pi \sqrt{\mu}} \sqrt{\frac{a_{curr}^3}{1 - c_v}} P(c_v) + \lambda_i \frac{2}{\pi \sqrt{\mu}} \sqrt{\frac{a_{curr}}{1 - c_v}} R(c_v) \quad P(c_v) = \sqrt{1 - c_v} K(c_v) \quad R(c_v) = \frac{1}{c_v} E(c_v) + \left(\sqrt{c_v} - \frac{1}{\sqrt{c_v}} \right) K(c_v)$$

where I've introduced two Lagrange multipliers (λ_a and λ_i). Get the optimal control profile by solving

$$\frac{da}{d\tau} = \frac{\partial \mathcal{H}_c}{\partial \lambda_a} \quad \frac{di}{d\tau} = \frac{\partial \mathcal{H}_c}{\partial \lambda_i}$$

$$\frac{d\lambda_a}{d\tau} = -\frac{\partial \mathcal{H}_c}{\partial a} \quad \frac{d\lambda_i}{d\tau} = -\frac{\partial \mathcal{H}_c}{\partial i}$$

$$\frac{\partial \mathcal{H}_c}{\partial c_v} = 0$$

$$\mathcal{H}_c(\tau) = 0$$

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.

Now 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 variable for the long-term problem as a function of the semimajor axis and the constant value of λ_i (canonical units).

$$c_v = \varphi_{LS}^{-1} \left[\frac{\pi}{2\lambda_i} \sqrt{\frac{1}{a_{curr}}} \right] \quad (6-53)$$

The control law φ_{LS} is defined as

$$\varphi_{LS}(c_v) = \frac{R'(c_v)P(c_v)}{P'(c_v)} - R(c_v) \quad (6-54)$$

where primes refer to differentiation with respect to c_v . This function is monotonic, so we can't invert it, but there is 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. (6-48)]. This solution is subject to the control law [Eq. (6-53)], with given boundary conditions. Thus, we've significantly reduced the order compared to that of the original problem. Now, starting from given initial conditions, we must determine two unknown constants, λ_i , and v_{acc} , while meeting the final conditions. We can easily solve this boundary-value problem with two variables by using numerical techniques and then globally mapping the results.

Now, use canonical units by setting $a_{init} = 1 \text{ DU}^*$, $i_{init} = 0^\circ$, and setting the gravitational parameter $\mu = 1 \text{ DU}^{*3}/\text{TU}^{*2}$, which—by Kepler's third law—implicitly sets the time unit. Numerically integrating the state equations of motion with the control [Eq. (6-53)] produces Fig. 6-24, in which contours of λ_i and Δv_{acc} are plotted on the a_{final} - i_{final} plane.

Curves of constant λ_i are also trajectories across the (a_{final}, i_{final}) plane. 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 Δv_{acc} contour interval is 0.1 in the dimensionless units (DU^*/TU^*). The figure permits simple interpolation of approximate values of the constants Δv_{acc} 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 solve the boundary-value problem for the slow time scale. The vehicle then executes a periodic program for yaw angle given by Eq. (6-51), with the slow-control variable, c_v , from Eq. (6-53). Whenever we meet the basic assumptions of the two time-scale methods, the fast-time variable, c_v , and the slow-time variable, Δv_{acc} , are independent of one another. As long as at least five or more revolutions elapse, the vehicle arrives in the final orbit when $\tau = \Delta v_{acc}$. The vehicle need not be able to solve the opti-

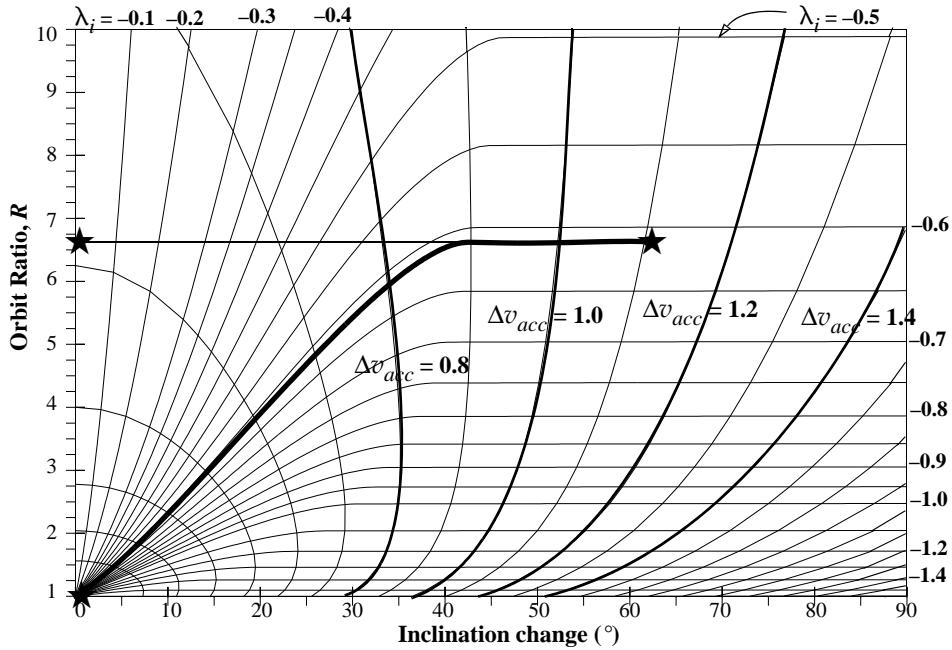


Figure 6-24. Low Thrust, Noncoplanar Transfers using the Alfano Transfer. This chart permits us to estimate changes in semimajor axis and inclination for satellites using continuous low thrust. Contours of λ_i represent trajectories and are labeled with the appropriate value. The contour interval for Δv_{acc} is 0.1. The example shows a transfer from LEO to GEO, and about a 63° inclination change. The Hohmann transfer changes the size first ($R = 6.6$), then the inclination, and then the final maneuver to end the process (each burn is shown by a \star). The low-thrust transfer performs a single continuous maneuver and is shown as the curved line.

mal-control problem. Given the values of λ_i and Δv_{acc} for the desired transfer, the vehicle only needs access to the current values of a_{curr} and c_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 Δv_{acc} to t conversion changes [Eq. (6-44)] 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 Δv_{acc} -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, $a_{thrusti}(t)$, used in the time transformation produces the same trajectory results with an optimal Δv_{acc} -time, as long as $a_{thrusti}(t)$ 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

Wiesel and Alfano (1985) provide a closed-form solution 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 $c_v = \varphi^{-1}(X)$ as a rational function of the variable $Z = X^{-2}$ in the form

$$c_v = \varphi^{-1}(Z = X^{-2}) \approx \sum_{i=0}^{10} \alpha_i Z^i \sum_{i=0}^{10} \beta_i Z^i \quad (6-55)$$

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 6-4). This is a Chebyshev fit, not a Taylor series, so don't truncate it.

TABLE 6-4. 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

An algorithm encapsulates the preceding equations and discussions. Note that transfers generally increase the semimajor axis, but you can use the same equations to decrease the orbit size by keeping the larger radius (semimajor axis) on top of the ratio (R) calculation. There are also several ways to establish the thrust of the rocket accomplishing the transfer. I have chosen I_{sp} , thrust, and initial mass. You can also use \dot{m} , and $a_{thrusti}$, or other combinations. Be aware that some solutions can take significant time to calculate due to the small time steps, and multiple revolutions. Both the short and long scale approaches are shown. The long-scale case could use Eq. (6-54), but the numerical approach from Eq. (6-55) is shown instead. Be careful of canonical equations—usually those with a_{curr} and R . Finally, the algorithm uses a simple Euler integrator. You could also use a combined position-velocity-acceleration integrator.

ALGORITHM 47: Low Thrust Transfer

$$(a_{init}, a_{final}, i_{init}, i_{final}, I_{sp}, F, m_i \Rightarrow \tau_{trans}, \Delta v)$$

$$R = a_{final} / a_{init}$$

$$\dot{m} = -\frac{F}{I_{sp}g} \quad \dot{m}_{spec} = \frac{\dot{m}}{m_i} \quad a_{thrusti} = \frac{F}{m_i}$$

$$\Delta i = i_{final} - i_{init}$$

$$\theta_p = 0.0, \Delta t = 0.0, \Delta t_s = \text{Period/steps}$$

$$\dot{\vec{r}}_o = a_{init} \Rightarrow 1.0 \text{ DU*}, \text{TU*} = \sqrt{\frac{r^3}{\mu}}$$

Find λ (need to iterate) and common node vector, \hat{n}_{common}

LOOP short scale long scale

$$c_v = \frac{1}{4\lambda^2 a_{curr}^2 + 1} \quad \text{or} \quad X = \frac{\pi}{2\lambda} \sqrt{\frac{1.0}{a_{curr}}}$$

$$c_v = \varphi^{-1}(Z = X^{-2}) \approx \sum_{i=0}^{10} \alpha_i Z^i \sum_{i=0}^{10} \beta_i Z^i$$

$$\text{TAN}(\phi_c) = \frac{\cos(\theta_p)}{\sqrt{\frac{1}{c_v} - 1}}$$

$$a_{thrust} = \frac{a_{thrusti}}{1 + \dot{m}_{spec} \Delta t}$$

$$\dot{\vec{a}}_{NTW} = a_{thrust} [\cos(\phi_c) \hat{T} + \sin(\phi_c) \hat{W}]$$

Form ($NTW \Rightarrow ECI$) conversion from target

$$\dot{\vec{a}}_{ECI} = \left[\frac{ECI}{NTW} \right] \dot{\vec{a}}_{NTW}$$

$$\dot{\vec{v}} = \dot{\vec{v}}_o + \dot{\vec{a}}_{ECI} \Delta t_s$$

PKEPLER($\dot{\vec{r}}_o, \dot{\vec{v}}_o, \Delta t \Rightarrow \dot{\vec{r}}_{curr}, \dot{\vec{v}}_{curr}$)

$$\Delta t = \Delta t + \Delta t_s$$

Find orbital elements $\left(a_{curr}, i, \cos(\theta_p) = \frac{\dot{\vec{r}}_{curr} \cdot \hat{n}_{common}}{|\dot{\vec{r}}_{curr}|} \right)$

$$\Delta t_s = 2\pi \sqrt{a_{curr}^3} \frac{\text{TU*}}{steps}$$

UNTIL Transfer complete ($a = a_{final}, i = i_{final}$)

$$\Delta v_{acc} = 1 - \sqrt{\frac{1}{R}}$$

$$t_f = \frac{1}{\dot{-m}_{spec}} \left(1 - \exp \left[\frac{\dot{m}_{spec} \Delta v_{acc}}{a_{thrusti}} \right] \right)$$

▼ **Example 6-13. Transferring from LEO to GEO.**

This case involves a coplanar transfer from a circular, low-Earth parking orbit of 200 km altitude (radius of 6578.136 km) to a final circular GEO orbit of 42,159.48 km (6.61 ER), including a 28.5° inclination change. The vehicle's specific impulse is 5000 sec from an ion engine, and the specific mass flow rate is -2×10^{-7} /sec.

The initial data for the altitudes is 1.0314 ER and 6.6105 ER. This gives us non-dimensional units of

$$1 \text{ DU}^* = 6578.136 \text{ km} = 1.0314 \text{ ER} \text{ and } 1 \text{ TU}^* = \sqrt{\frac{a^3}{\mu}} = \sqrt{\frac{1.0314^3}{1}} = 1.0474 \text{ TU} = 845.0472 \text{ sec}$$

The ratio of the orbits is $6.6105/1.0314 = 6.4095$. From Fig. 6-24, we find $v_{acc} = 0.75 \text{ DU}^*/\text{TU}^*$ and $\lambda_i = -0.54$. Converting the units on v_{acc} ,

$$v_{acc} = 0.75 \text{ DU}^*/\text{TU}^* = 0.75 (6578.136) (1/845.0472) = 5.8382 \text{ km/s}$$

The exhaust velocity is gI_{sp} , which is $0.010(5000) = 50 \text{ km/s}$ (remember that gI_{sp} is also $-a_i/\dot{m}_{spec}$ so a_i is $1 \times 10^{-5} \text{ km/s}^2$). We can now convert to find the time to complete the maneuver using Eq. (6-44).

$$\blacktriangleleft t_f = \frac{1}{\dot{-m}_{spec}} \left(1 - \exp \left[\frac{\dot{m}_{spec} \Delta v_{acc}}{a_i} \right] \right) = \frac{-1}{-2 \times 10^{-7}} \left(1 - \exp \left[\frac{-2 \times 10^{-7} (5.8382)}{1 \times 10^{-5}} \right] \right) = 551,024 \text{ sec,}$$

or 6.377 days.

6.8 Relative Motion

This section develops and analyzes equations of motion between two close-orbiting satellites. These equations are used extensively in *formation flying*. 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. 6-25, 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 completely arbitrary. 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.

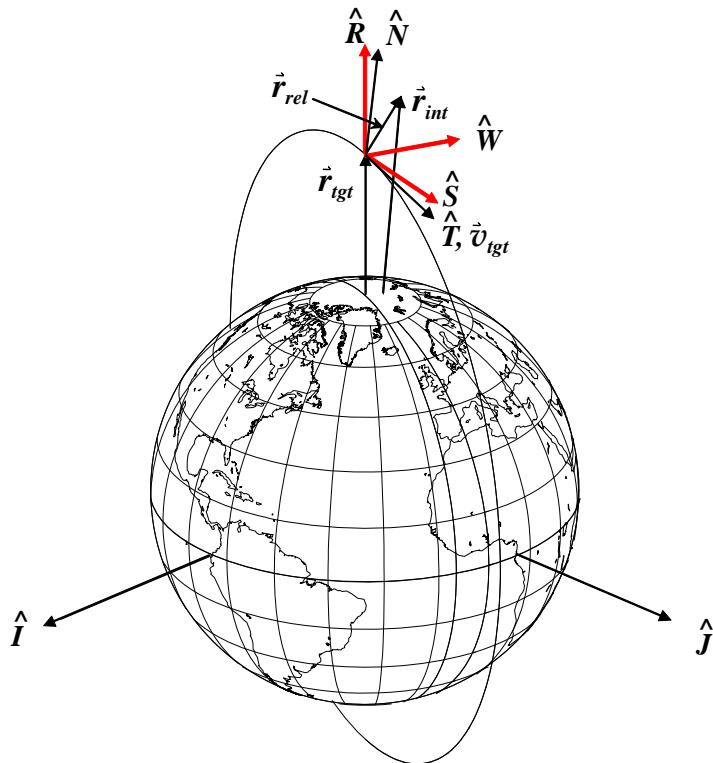


Figure 6-25. Coordinate-System Geometry for Relative Motion. The traditional Hill's coordinate system aligns with the RSW system as shown. The variables x , y and z and their rates describe the displacements between the target and interceptor vehicles in the RSW frame. Remember that the S and y axes will be aligned with the velocity vector only in circular orbits—a major assumption for Hill's solutions. The NTW frame is necessary for later transformations admitting elliptical orbits.

To derive the equations of motion, we use the RSW coordinate system introduced in Sec. 3.3. 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.*

We can define the equation of motion for the target satellite using two-body motion.

* 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.

$$\ddot{\vec{r}}_{tgt} = -\frac{\mu \vec{r}_{tgt}}{r_{tgt}^3} \quad (6-56)$$

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,

$$\ddot{\vec{r}}_{int} = -\frac{\mu \vec{r}_{int}}{r_{int}^3} + \vec{F}_{thrust} \quad (6-57)$$

From Fig. 6-25, 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. (6-56) and Eq. (6-57)—gives us

$$\ddot{\vec{r}}_{rel} = -\frac{\mu \vec{r}_{int}}{r_{int}^3} + \vec{F}_{thrust} + \frac{\mu \vec{r}_{tgt}}{r_{tgt}^3} \quad (6-58)$$

To be useful, this equation needs a little work. Our choice of coordinate systems makes a difference. I'm using Kaplan's convention. You can change it to match Clohessy-Wiltshire, or any other convention, 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:

$$\frac{\ddot{\mathbf{r}}_{int}}{r_{int}^3} = \frac{\dot{\mathbf{r}}_{tgt} + \dot{\mathbf{r}}_{rel}}{r_{tgt}^3} \left\{ \frac{1}{\left(1 + \frac{2\dot{\mathbf{r}}_{tgt} \cdot \dot{\mathbf{r}}_{rel}}{r_{tgt}^2} \right)^{3/2}} \right\}$$

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

$$\frac{\ddot{\mathbf{r}}_{int}}{r_{int}^3} = \frac{\dot{\mathbf{r}}_{tgt} + \dot{\mathbf{r}}_{rel}}{r_{tgt}^3} \left\{ 1 - \frac{3}{2} \left(\frac{2\dot{\mathbf{r}}_{tgt} \cdot \dot{\mathbf{r}}_{rel}}{r_{tgt}^2} \right) + \dots \right\}$$

Now substitute this result into Eq. (6-58):

$$\ddot{\mathbf{r}}_{rel} = -\mu \left(\frac{\dot{\mathbf{r}}_{tgt} + \dot{\mathbf{r}}_{rel}}{r_{tgt}^3} \left\{ 1 - \frac{3}{2} \left(\frac{2\dot{\mathbf{r}}_{tgt} \cdot \dot{\mathbf{r}}_{rel}}{r_{tgt}^2} \right) + \dots \right\} \right) + \vec{F}_{thrust} + \frac{\mu \dot{\mathbf{r}}_{tgt}}{r_{tgt}^3}$$

Expanding and removing terms of opposite signs and keeping only first-order terms gives you

$$\ddot{\mathbf{r}}_{rel} = -\frac{\mu}{r_{tgt}^3} \left\{ -\frac{3\dot{\mathbf{r}}_{tgt}}{2} \left(\frac{2\dot{\mathbf{r}}_{tgt} \cdot \dot{\mathbf{r}}_{rel}}{r_{tgt}^2} \right) + \dot{\mathbf{r}}_{rel} - \frac{3\dot{\mathbf{r}}_{rel}}{2} \left(\frac{2\dot{\mathbf{r}}_{tgt} \cdot \dot{\mathbf{r}}_{rel}}{r_{tgt}^2} \right) \right\} + \vec{F}_{thrust}$$

Here it's possible to reduce some terms using the first-order criterion with the product of $\dot{\mathbf{r}}_{rel}$ and the dot product. Assume this quantity is small and drop it to simplify the problem:

$$\ddot{\mathbf{r}}_{rel} = -\frac{\mu}{r_{tgt}^3} \left\{ -\frac{3\dot{\mathbf{r}}_{tgt}}{2} \left(\frac{2\dot{\mathbf{r}}_{tgt} \cdot \dot{\mathbf{r}}_{rel}}{r_{tgt}^2} \right) + \dot{\mathbf{r}}_{rel} \right\} + \vec{F}_{thrust}$$

I've rearranged this equation to show unit vectors, $\hat{\mathbf{R}}$, in the direction of the target satellite. The unit vector and relative vector dot product yield only the x -component. Thus,

$$\ddot{\mathbf{r}}_{rel} = -\frac{\mu}{r_{tgt}^3} \left\{ -3x\hat{\mathbf{R}} + \dot{\mathbf{r}}_{rel} \right\} + \vec{F}_{thrust} \quad (6-59)$$

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. (3-26), which relates inertial and relative vectors in a rotating frame. Therefore,

$$\ddot{\vec{r}}_I = \ddot{\vec{r}}_R + \vec{\omega}_R \times \dot{\vec{r}}_R + 2\vec{\omega}_R \times \dot{\vec{r}}_R + \vec{\omega}_R \times (\vec{\omega}_R \times \dot{\vec{r}}_R)$$

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 $\dot{\vec{r}}$. Solving for the relative acceleration gives us

$$\ddot{\vec{r}}_R = \ddot{\vec{r}}_I - \vec{\omega}_R \times \dot{\vec{r}}_R - 2\vec{\omega}_R \times \dot{\vec{r}}_R - \vec{\omega}_R \times (\vec{\omega}_R \times \dot{\vec{r}}_R)$$

Writing this in terms of the vectors in this section results in ^{*}

$$\begin{aligned} \ddot{\vec{r}}_{rel,R} &= \ddot{\vec{r}}_{rel,I} - \vec{\omega}_R \times \dot{\vec{r}}_{rel} - 2\vec{\omega}_R \times \dot{\vec{r}}_{rel} - \vec{\omega}_R \times (\vec{\omega}_R \times \dot{\vec{r}}_{rel}) \\ \omega &= \sqrt{\frac{\mu}{r_{tgt}^3}} \end{aligned} \quad (6-60)$$

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 $\dot{\vec{r}}_{rel}$ in the target's coordinate system:

$$\begin{aligned} \vec{\omega}_R \times \dot{\vec{r}}_{rel} &= \begin{bmatrix} \hat{R} & \hat{S} & \hat{W} \\ 0 & 0 & \dot{\omega} \\ x & y & z \end{bmatrix} = -\dot{\omega}y\hat{R} + \dot{\omega}x\hat{S} \\ \vec{\omega}_R \times \dot{\vec{r}}_{rel} &= \begin{bmatrix} \hat{R} & \hat{S} & \hat{W} \\ 0 & 0 & \omega \\ \dot{x} & \dot{y} & \dot{z} \end{bmatrix} = -\omega\dot{y}\hat{R} + \omega\dot{x}\hat{S} \\ \vec{\omega}_R \times (\vec{\omega}_R \times \dot{\vec{r}}_{rel}) &= \vec{\omega}_R \times \begin{bmatrix} \hat{R} & \hat{S} & \hat{W} \\ 0 & 0 & \omega \\ x & y & z \end{bmatrix} = \vec{\omega}_R \times (-\omega y\hat{R} + \omega x\hat{S}) \\ &= \begin{bmatrix} \hat{R} & \hat{S} & \hat{W} \\ 0 & 0 & \omega \\ -\omega y & \omega x & 0 \end{bmatrix} = -\omega^2x\hat{R} - \omega^2y\hat{S} \end{aligned}$$

* Note that this result is valid only for circular orbits. In general, $\omega = \sqrt{\frac{\mu p}{r_{tgt}^4}}$

Substituting these values and the derived value for the inertial acceleration of the interceptor [Eq. (6-59)] into the relative acceleration [Eq. (6-60)] results in the generalized equations of motion:

$$\ddot{\vec{r}}_{relR} = -\frac{\mu}{r_{tgt}^3} \{ \hat{r}_{rel} - 3x\hat{R} \} + \vec{F}_{thrust} + \dot{\omega}y\hat{R} - \dot{\omega}x\hat{S} - 2(-\omega y\hat{R} + \omega x\hat{S}) + \omega^2 x\hat{R} + \omega^2 y\hat{S}$$

If we *assume* circular motion, the constant in the first term is simply ω^2 and $\dot{\omega} = 0$. Expanding the terms gives us

$$\ddot{\vec{r}}_{relR} = -\omega^2 \{ x\hat{R} + y\hat{S} + z\hat{W} - 3x\hat{R} \} + \vec{F}_{thrust} + 2\omega y\hat{R} - 2\omega x\hat{S} + \omega^2 x\hat{R} + \omega^2 y\hat{S}$$

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*:

$$\left. \begin{aligned} \ddot{x} - 2\omega\dot{y} - 3\omega^2x &= f_x \\ \ddot{y} + 2\omega\dot{x} &= f_y \\ \ddot{z} + \omega^2z &= f_z \end{aligned} \right\} \text{near-circular orbits} \quad (6-61)$$

Note the types of acceleration in these equations. In the f_x equation—from left to right—they are total, Coriolis, and centripetal acceleration.

6.8.1 Position Solutions for Nearly Circular Orbits

To solve Hill's equations, 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. (6-61), first take the derivative of the x equation, without any external forces and with ω 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

$$\mathcal{L}[\ddot{x} + \omega^2 \dot{x}] = \{s^3 X(s) - s^2 x_o - s \dot{x}_o - \ddot{x}_o\} + \omega^2 \{s X(s) - x_o\} = 0$$

and expand to get

$$s^3 X(s) - s^2 x_o - s \dot{x}_o - \ddot{x}_o + s X(s) \omega^2 - x_o \omega^2 = 0$$

Group terms to reduce this expression to

$$s X(s)(s^2 + \omega^2) = (s^2 + \omega^2)x_o + s \dot{x}_o + \ddot{x}_o$$

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. (6-61) (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. (6-61)] 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:

$$\dot{x}(t) = \frac{\dot{x}_o}{\omega} \cos(\omega t) + \left(3x_o + \frac{2\dot{y}_o}{\omega} \right) \omega \sin(\omega t)$$

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

$$\begin{aligned} s^2 Z(s) - sz_o - \dot{z}_o + \omega^2 Z(s) &= 0 \\ Z(s)(s^2 + \omega^2) &= sz_o + \dot{z}_o \\ Z(s) &= \frac{sz_o}{(s^2 + \omega^2)} + \frac{\dot{z}_o}{(s^2 + \omega^2)} \end{aligned}$$

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, we combine the position and velocity equations into a computational algorithm.

ALGORITHM 48: Hill's Equations

$(x_o, y_o, z_o, \dot{x}_o, \dot{y}_o, \dot{z}_o \text{ of the interceptor}, \omega_{tgt}, \Delta t \Rightarrow$

$x, y, z, \dot{x}, \dot{y}, \dot{z} \text{ of the interceptor})$

$$\begin{aligned}
 x(t) &= \frac{\dot{x}_o}{\omega} \sin(\omega t) - \left(3x_o + \frac{2\dot{y}_o}{\omega} \right) \cos(\omega t) + \left(4x_o + \frac{2\dot{y}_o}{\omega} \right) \\
 y(t) &= \left(6x_o + \frac{4\dot{y}_o}{\omega} \right) \sin(\omega t) + \frac{2\dot{x}_o}{\omega} \cos(\omega t) - (6\omega x_o + 3\dot{y}_o)t + \left(y_o - \frac{2\dot{x}_o}{\omega} \right) \\
 z(t) &= z_o \cos(\omega t) + \frac{\dot{z}_o}{\omega} \sin(\omega t) \\
 \dot{x}(t) &= \dot{x}_o \cos(\omega t) + (3\omega x_o + 2\dot{y}_o) \sin(\omega t) \\
 \dot{y}(t) &= (6\omega x_o + 4\dot{y}_o) \cos(\omega t) - 2\dot{x}_o \sin(\omega t) - (6\omega x_o + 3\dot{y}_o) \\
 \dot{z}(t) &= -z_o \omega \sin(\omega t) + \dot{z}_o \cos(\omega t)
 \end{aligned}$$

An example shows Hill's equations in operation:

▼ Example 6-14. 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 down, 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

$$\dot{x}_o = -0.1 \quad \dot{y}_o = -0.04 \quad \dot{z}_o = -0.02 \quad \text{m/s}$$

Next, find the angular rate of the target (Shuttle). In a 590 km altitude orbit,

$$\omega = \sqrt{\frac{\mu}{r_{tgt}^3}} = \sqrt{\frac{398,600.5}{6968.1363^3}} = 0.001\ 085\ 4 \text{ rad/s}$$

The initial position of 0.0 simplifies the equations in Algorithm 48 to

$$\begin{aligned}
 x(t) &= \frac{\dot{x}_o}{\omega} \sin(\omega t) - \frac{2\dot{y}_o}{\omega} \cos(\omega t) + \frac{2\dot{y}_o}{\omega} \\
 y(t) &= \frac{4\dot{y}_o}{\omega} \sin(\omega t) + \frac{2\dot{x}_o}{\omega} \cos(\omega t) - 3\dot{y}_o t - \frac{2\dot{x}_o}{\omega} \\
 z(t) &= \frac{\dot{z}_o}{\omega} \sin(\omega t)
 \end{aligned}$$

$$\begin{aligned}\dot{x}(t) &= \dot{x}_o \cos(\omega t) + 2\dot{y}_o \sin(\omega t) \\ \dot{y}(t) &= 4\dot{y}_o \cos(\omega t) - 2\dot{x}_o \sin(\omega t) - 3\dot{y}_o \\ \dot{z}(t) &= \dot{z}_o \cos(\omega t)\end{aligned}$$

Substituting the time of 5 minutes (300^s) yields

$$\begin{array}{ll}x(5) = -33.345 \text{ m} & \dot{x}(5) = -0.12 \text{ m/s} \\ y(5) = -1.473 \text{ m} & \dot{y}(5) = 0.03 \text{ m/s} \\ z(5) = -5.894 \text{ m} & \dot{z}(5) = -0.02 \text{ m/s}\end{array}$$

and, for the case of 20 minutes (1200^s),

$$\begin{array}{ll}x(20) = -143.000 \text{ m} & \dot{x}(20) = -0.10 \text{ m/s} \\ y(20) = 137.279 \text{ m} & \dot{y}(20) = 0.27 \text{ m/s} \\ z(20) = -17.766 \text{ m} & \dot{z}(20) = -0.01 \text{ m/s}\end{array}$$

▲ Notice how both the position and velocity change dramatically.

6.8.2 Trend Analysis

We can use the equations in Algorithm 48 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. 6.8.3. By examining the x or y positions in Algorithm 48, 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

$$x(t) = C \left\{ \frac{\dot{x}_o}{C} \sin(\omega t) - \frac{3x_o + \frac{2\dot{y}_o}{\omega}}{C} \cos(\omega t) \right\} + 4x_o + \frac{2\dot{y}_o}{\omega}$$

Define the two trigonometric terms:

$$\sin(\psi_o) \equiv -\frac{3x_o + \frac{2\dot{y}_o}{\omega}}{C} \quad \cos(\psi_o) \equiv \frac{\dot{x}_o}{C}$$

Substitute to get

$$x(t) = C\{\cos(\psi_o)\sin(\omega t) + \sin(\psi_o)\cos(\omega t)\} + 4x_o + \frac{2\dot{y}_o}{\omega}$$

Recognizing the trigonometric summation identity [Eq. (C-15)] and performing similar operations on $y(t)$ result in

$$\begin{aligned} x &= x_c + C\sin(\omega t + \psi_o) \\ y &= y_{c_o} + \dot{y}_c t + 2C\cos(\omega t + \psi_o) \end{aligned} \tag{6-62}$$

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 48 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:

$$\begin{aligned} D &\equiv \sqrt{\left(\frac{\dot{z}_o}{\omega}\right)^2 + z_o^2} \\ \sin(\phi_o) &\equiv \frac{\dot{z}_o}{D} \quad \cos(\phi_o) \equiv \frac{z_o}{D} \end{aligned}$$

Now, rewrite the general form for the z equation using Eq. (C-15):

$$z = D \cos(\omega t - \phi_o) \quad (6-63)$$

Find the center z_c by substituting $t = 0$ into Algorithm 48, or $z_c = z_o$. Summarize the equations above to get

$$\begin{aligned} y_{c_o} &= y_o - \frac{2\dot{x}_o}{\omega} & C &= \sqrt{\left(3x_o + \frac{2\dot{y}_o}{\omega}\right)^2 + \left(\frac{\dot{x}_o}{\omega}\right)^2} \\ \dot{y}_c &= -6\omega x_o - 3\dot{y}_o & D &= \sqrt{\left(\frac{\dot{z}_o}{\omega}\right)^2 + z_o^2} \\ x_c &= -\frac{2\dot{y}_c}{3\omega} & z_c &= z_o \end{aligned} \quad (6-64)$$

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. (6-62)] in a moving frame, we can separate terms as follows:

$$\begin{aligned} x_m &= x - x_c = C \sin(\omega t + \psi_o) \\ y_m &= y - y_{c_o} = 2C \cos(\omega t + \psi_o) \end{aligned}$$

Thus, we have equations of motion that represent an ellipse in the x_m - y_m plane.

$$\frac{x_m^2}{C^2} + \frac{y_m^2}{4C^2} = 1$$

Similar analysis with Eq. (6-63) indicates a circle in the y_m - z plane. The preceding results let us to draw an elliptical cylinder of the interceptor's instantaneous motion (Fig. 6-26).

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. (6-64) 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 ($r_{tgt} = 6378.137 + 222$ km = 6600.137 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:

$$\omega = \sqrt{\frac{\mu}{r_{tgt}^3}} = \sqrt{\frac{398,600.5}{6600.137^3}} = 0.001\ 177\ 44 \text{ rad/s}$$

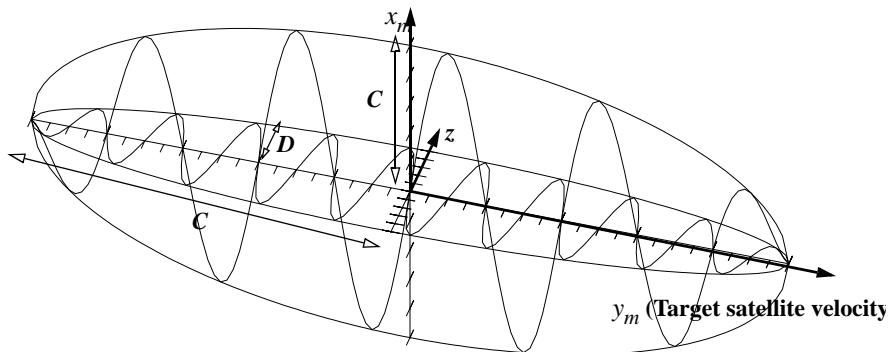


Figure 6-26. Interceptor's Instantaneous Relative Motion about the Target. We can define an elliptical region in which the interceptor “orbits” about the target satellite. (This is similar to a prolate spheroid, discussed in Sec. 3.2.2.) Although it’s not actually an orbit, we can approximate the motion as an ellipse for short time spans. Over time, this instantaneous figure stretches (due to the y_c drift term) into an elliptic cylinder. This is the actual surface for the satellites motion. 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.

We can examine the x and y directions independently of z because their motion doesn’t couple with z . The analysis in Fig. 6-27 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.

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 6-27 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

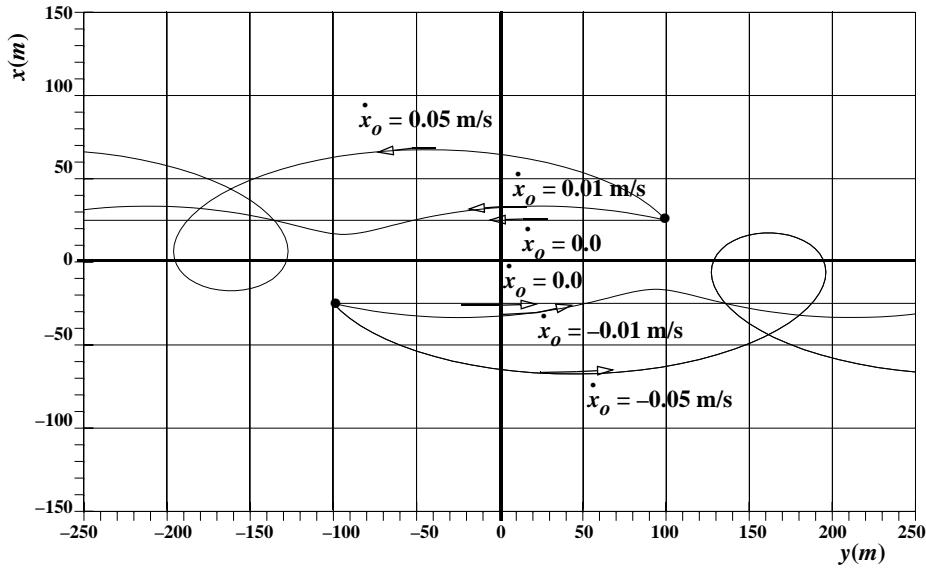


Figure 6-27. Interceptor Motion for x_o and y_o Displacements, \dot{x}_o and \dot{y}_o Variations. By varying the interceptor's initial conditions, several interesting motions are possible, including straight lines — $\dot{y}_o = 0.04415 \text{ m/s}$ for each x_o negative case, and $\dot{y}_o = -0.04415 \text{ m/s}$ for each x_o positive case. Note, these straight lines are not physically possible for an extended period of time.

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. 6-27, you would expect loops due to the eccentricity and a backward or forward motion depending on the initial x displacement. Figure 6-28 shows various initial x displacements, positive and negative, between 5 and 500 m.

You can also deduce the results of Fig. 6-28 using the formulas given in Eq. (6-64). 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. To find these combinations, you must set the C expression equal to zero and allow x_o and \dot{y}_o , respectively, to vary. Thus,

$$\begin{aligned} \dot{x}_o &= \dot{y}_o = x_o = 0 \\ \text{or} \\ \dot{x}_o &= 0 \quad \dot{y}_o = -\frac{3}{2}x_o\omega \\ \text{or} \\ \dot{x}_o &= 0 \quad x_o = \frac{-2\dot{y}_o}{3\omega} \end{aligned} \tag{6-65}$$

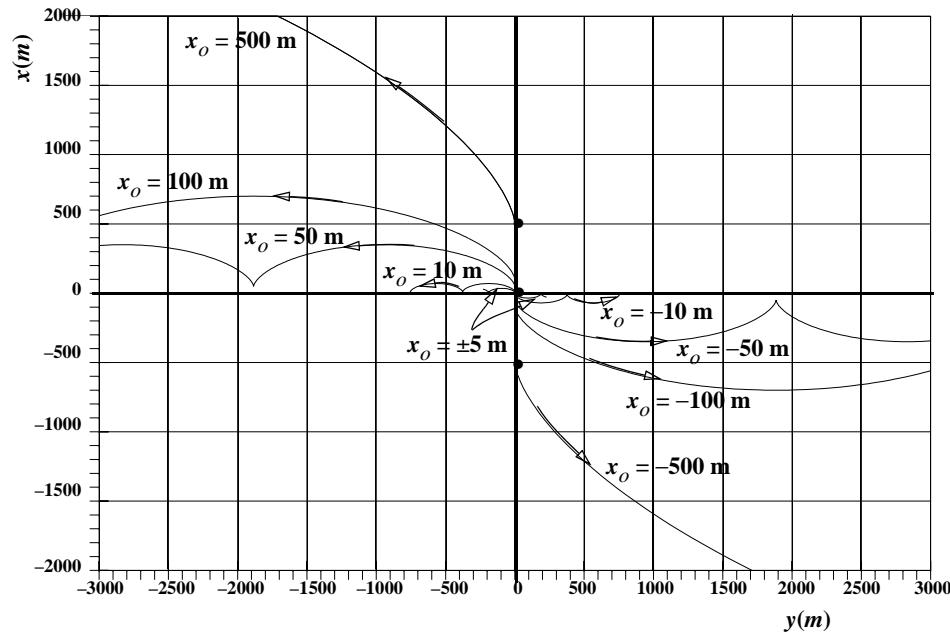


Figure 6-28. Motion of the Interceptor for Various x_o Displacements. Notice how all displacements above the satellite move to the left (lower velocity causes a lagging effect), whereas those below move to the right (higher velocity causes a lead). In addition, as the satellite's initial location gets farther from the target, the motion away from the target increases.

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. 6-27.

As x_o grows from 5 to 500 m from the target in Fig. 6-28 (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. 6-29). 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.

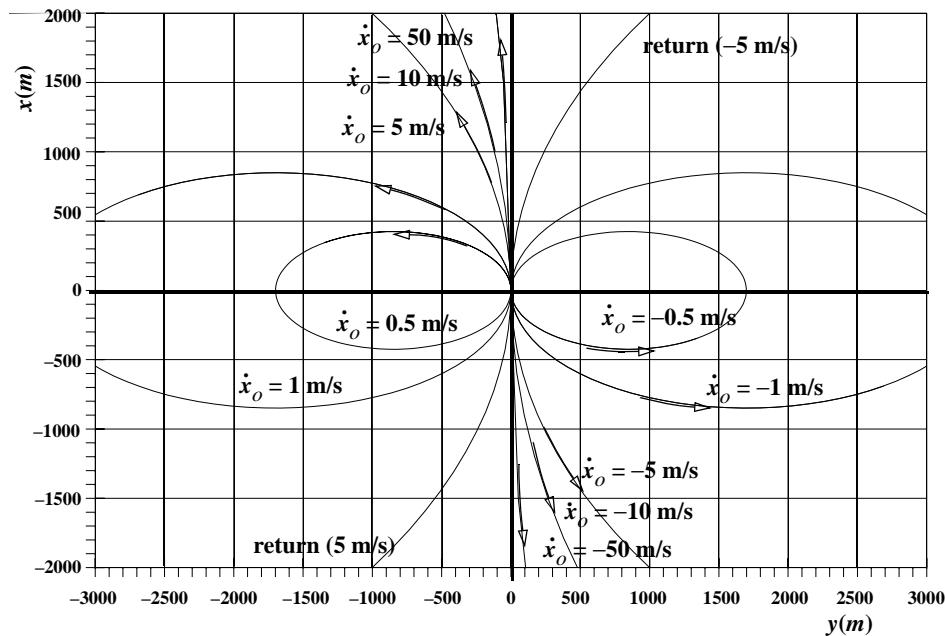


Figure 6-29. Motion of the Interceptor for \dot{x}_o Variations. All motions which result from variations in the x velocity become periodic motions about the target. Notice each path *will* return to the target in one period.

Nonzero values of \dot{y}_o will also produce an *elliptical motion* (Fig. 6-30). Both the mathematics from Eq. (6-64) 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. 6-29 and 6-30, except they are displaced to the initial y position. Note that certain values of x velocity,

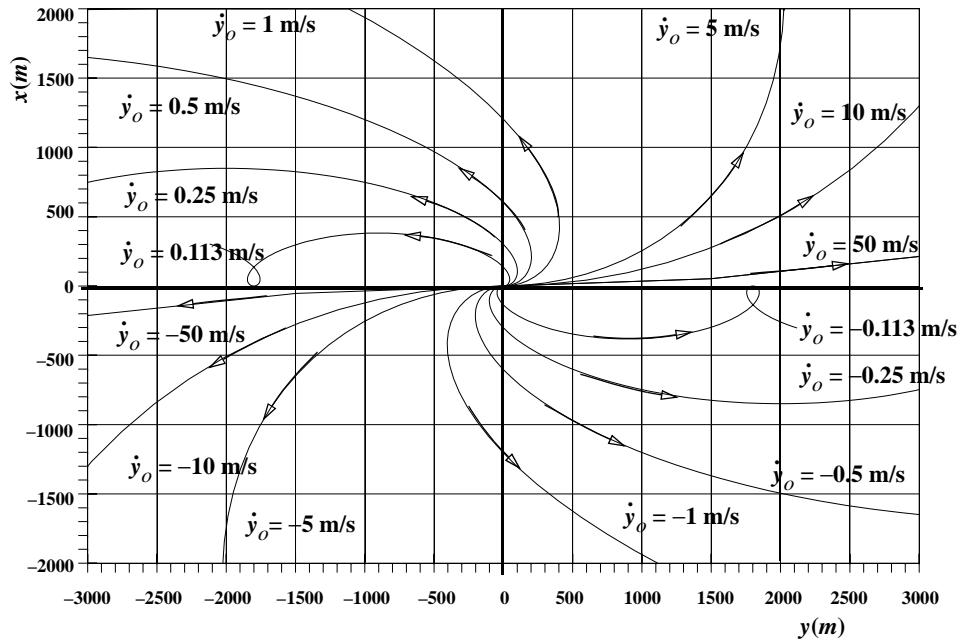


Figure 6-30. Motion of the Interceptor for \dot{y}_o Variations. All positive y velocity changes cause the interceptor to travel left, whereas negative velocities cause motion to the right. Although $\dot{y}_o = 50 \text{ m/s}$ appears to travel to the right, it will curve back around to the left over time.

combined with y displacements, determine the initial center of the *ellipse*. From Eq. (6-64), it's apparent that \dot{y}_o and \dot{x}_o are the only two variables that affect this center. If

$$\dot{y}_o = \frac{2\dot{x}_o}{\omega}$$

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. 6-30 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 6-31 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. (6-64), 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 .

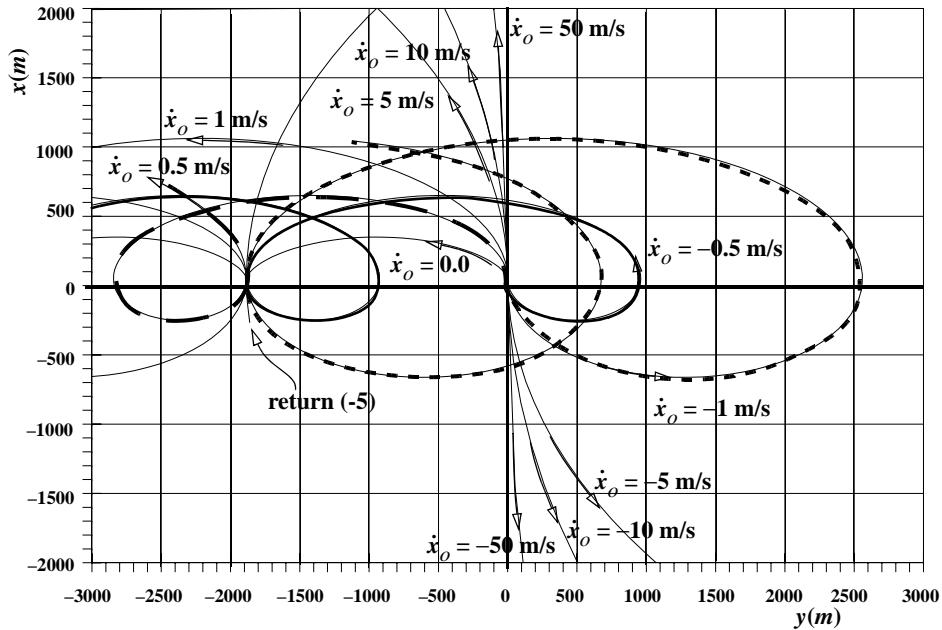


Figure 6-31. Motion of the Interceptor for x_o Displacements and \dot{x}_o Variations. This figure represents the most complex motion of the individual changes. The graph is based on a constant offset of 50 m for x_o , with velocity changes in the x direction as indicated. Notice how many of the curves converge at the same point (near the return). Eventually, all paths will end up drifting behind (to the left).

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. 6-32 uses a 50 m displacement for x . In this case,

$$\dot{y}_o = -2\omega x_o = -2(0.00117744)(50) = -0.117744 \text{ m/s}$$

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.

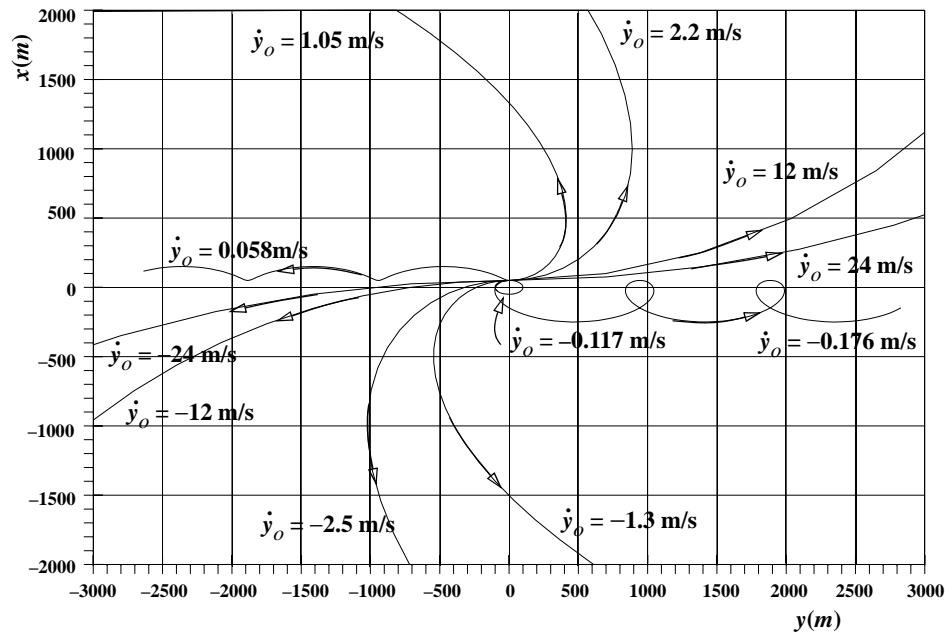


Figure 6-32. Motion of the Interceptor for x_o Displacements and \dot{y}_o Variations. Each of the variations above comes from a 50 m x_o displacement coupled with varying values of y velocity as shown. Notice the special case of an “orbit” around the target satellite at -0.117 m/s.

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. (6-64) 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 6-33 shows this combination. The x-y motion is the same as the loops in Fig. 6-27, 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 plots.
4. Adjusting the value of \dot{y}_o to $\dot{y}_o = -2\omega x_o$ will stop any drift in the *ellipse*.

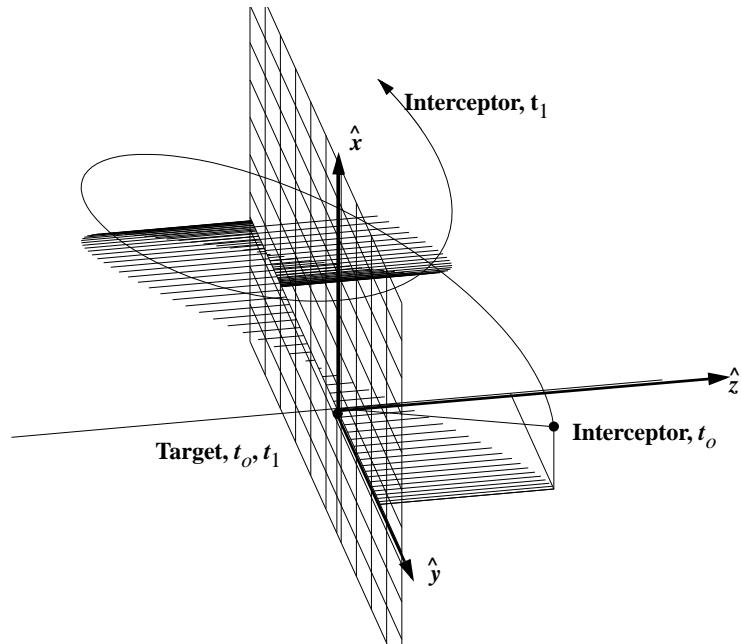


Figure 6-33. Relative Motion of the Interceptor with a z Variation. Adding a z variation to any of the other motions described simply adds a three-dimensional component. In this example, the interceptor goes above and then below the target satellite.

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 48 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

$$\begin{aligned} \frac{\dot{x}_o}{\omega} \sin(\omega t) - \left(3x_o + \frac{2\dot{y}_o}{\omega} \right) \cos(\omega t) + \left(4x_o + \frac{2\dot{y}_o}{\omega} \right) = 0 \\ \left(6x_o + \frac{4\dot{y}_o}{\omega} \right) \sin(\omega t) + \frac{2\dot{x}_o}{\omega} \cos(\omega t) - (6\omega x_o + 3\dot{y}_o)t + \left(y_o - \frac{2\dot{x}_o}{\omega} \right) = 0 \\ z_o \cos(\omega t) + \frac{\dot{z}_o}{\omega} \sin(\omega t) = 0 \end{aligned}$$

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

$$\dot{y}_o = \frac{-2\omega x_o (1 - \cos(\omega t))(4 - 3\cos(\omega t)) - (y_o \omega \sin(\omega t) + 6x_o \omega \sin^2(\omega t) - 6\omega^2 x_o t \sin(\omega t))}{4(1 - \cos(\omega t))^2 + (4 \sin(\omega t) - 3\omega t)(\sin(\omega t))}$$

Factoring out an $\omega \sin(\omega t)$ in the numerator, and then a $6x_o$, results in

$$\dot{y}_o = \frac{-2\omega x_o(1 - \cos(\omega t))(4 - 3\cos(\omega t)) - (\omega \sin(\omega t)\{y_o + 6x_o(\sin(\omega t) - \omega t)\})}{4(1 - \cos(\omega t))^2 + (4\sin(\omega t) - 3\omega t)(\sin(\omega t))}$$

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

$$\begin{aligned} \dot{y}_o &= \frac{(6x_o(\omega t - \sin(\omega t)) - y_o)\omega \sin(\omega t) - 2\omega x_o(4 - 3\cos(\omega t))(1 - \cos(\omega t))}{(4\sin(\omega t) - 3\omega t)\sin(\omega t) + 4(1 - \cos(\omega t))^2} \\ \dot{x}_o &= -\frac{\omega x_o(4 - 3\cos(\omega t)) + 2(1 - \cos(\omega t))\dot{y}_o}{\sin(\omega t)} \\ \dot{z}_o &= -z_o \omega \cot(\omega t) \end{aligned} \quad (6-66)$$

Let's look at an example of this technique.

▼ Example 6-15. Solving for Velocity in Hill's Equation.

GIVEN: A satellite in circular orbit at 590 km altitude. Suppose we continue Example 6-14; after 10 minutes, we decide to retrieve the Hubble Space Telescope.

FIND: Δv to rendezvous with the Shuttle in 5 and 15 minutes.

Examining Eq. (6-66) reveals quantities similar to those in Example 6-14:

$$\omega = 0.001\ 085\ 4\ \text{rad/s}$$

First, we find the position of the Hubble Space Telescope after 10 minutes (600^s). Using Algorithm 48,

$$x(600) = -70.933\ \text{m}, y(600) = 20.357\ \text{m}, \text{ and } z(600) = -11.170\ \text{m}$$

Now, solve for the required velocity to rendezvous [Eq. (6-66)] in 5 and 15 minutes (300^s and 900^s), but be sure to calculate \dot{y}_o first because you need it to calculate \dot{x}_o .

At $t = 5$ minutes (300 seconds), At $t = 15$ minutes (900 seconds),

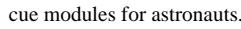
$$\dot{y}_o = 0.0135 \text{ m/s} \quad \dot{y}_o = 0.0753 \text{ m/s}$$

$$\dot{x}_o = 0.2742 \text{ m/s} \quad \dot{x}_o = 0.1356 \text{ m/s}$$

$$\dot{z}_o = 0.0359 \text{ m/s} \quad \dot{z}_o = 0.0082 \text{ m/s}$$

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 $\dot{\omega}$ 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 sidereal 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.

6.8.3 Transformation to Inertial Coordinates

The preceding two-dimensional linearized analysis suggests a question: how accurate are these results compared to simply propagating each individual satellite state with complete force models? The answer requires two steps. First we need a rigorous method to transform between Cartesian (ECI) and Hill's (curvilinear) space. Then we can examine sample cases and determine the effect of the Hill's linearization. The transformation

presented here is based on Vallado and Alfano (2011).^{*} This new approach has the advantage of working for circular and elliptical orbits, has corrections to the velocity transformation, and references the actual satellite orbit to better specify the separation distance. The approach is completely general because it lets us use *any* force model. It can evaluate any size gravity field, atmospheric drag, solar radiation pressure, etc.—a significant increase in applicability from circular only, two-body motion!

Several authors have extended Hill's equations into perturbed motion (Schweighart and Sedwick, 2002, Gim and Alfriend, 2003). This is most often accomplished by adding the perturbing forces to the right hand side of Eq. (6-56) and Eq. (6-57), however, this is much more complex than using the transformations described in this section.

For all orbits, the linear xyz distances can be misleading in the portrayal of the actual relative differences. The x component is unambiguous, being measured radially out from the central body. However, the y and z components contain differences due to the curvilinear orbital motion. Consider a satellite with only y and z displacements from the target satellite. Figure 6-34 shows that the y displacement actually has an x component as well, due to the curvature of the target satellite's orbit (e.g., a y displacement of 100 m is really slightly less than 100 m, and it includes a small negative x component). The same is true for z and for velocity components.

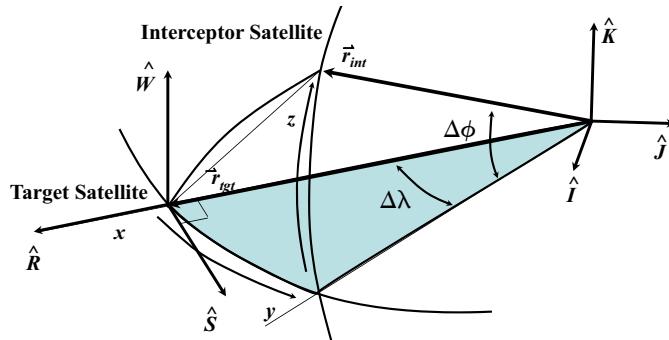


Figure 6-34. Relative Position of Interceptor and Target. This diagram is extremely exaggerated to illustrate the inconsistency in assuming the Hill's y and z component displacements transfer directly to cartesian position and velocity vectors.

The question we must address is how to convert between Cartesian (ECI) and Hill's—between the true vector positions, and the linearized, or relative Hill's-equation positions. We begin with the **Equidistant Cylindrical** (EQC) system (Weisstein, 2011).[†]

* In my first edition (1997), a transformation was used to examine the accuracy of Hill's equations using complete force models. It properly converted the position vectors, but had some velocity vector errors. Melton (2000) analyzed accuracy results for modestly eccentric orbits using two-body motion and relative position vectors, and Gim and Alfriend (2003) examined the accuracy for eccentric perturbed motion. Hill, Alfriend, and Sabol, (2008) examined the transformation while investigating covariance propagation issues. While mentioning an ECI-curvilinear transformation, the latter papers did not detail the transformation for position and velocity vectors.

Elliptical orbits require a more complex formulation because the arc length (the y distance shown in Fig. 6-34) is actually a function of the true anomaly. Although the EQC system appears to solve the problem, the distance along an elliptical orbit between two true anomaly values introduces an arc length dependency that ultimately requires an elliptic integral (as suggested by Hill, Alfriend, and Sabol, 2008). Therefore, we use a modified EQC (EQCM) notation for the coordinate system, in which we bend the frame about the target orbital path (to be consistent with the target orbital track).

To derive the transformations, examine Fig. 6-35. Note the use of points (1, 2, and 3) to accomplish the transformation and recognize that the interceptor may be in any orbit, just like the target satellite. All initial points and distances relate to the target satellite and its orbit.

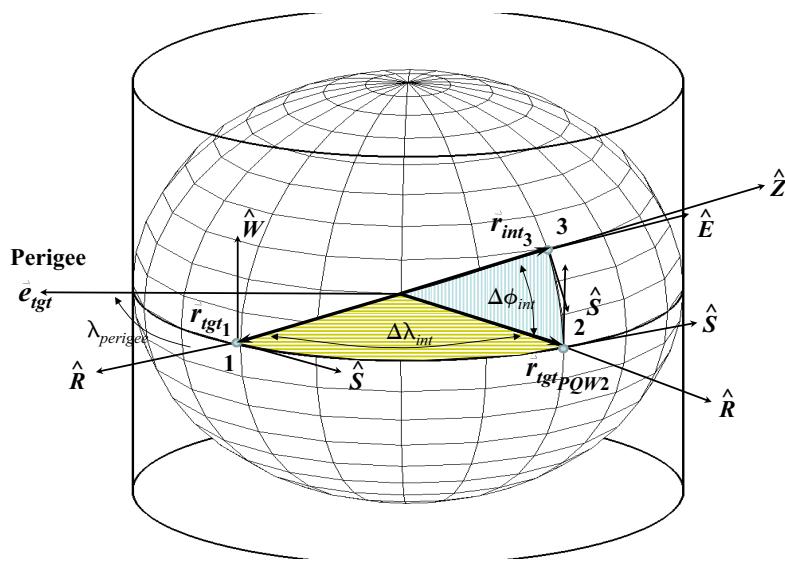


Figure 6-35. Velocity Transformation. Transforming the velocity vectors requires angles and rates of change as depicted here. Note that although these coordinates differ in curvilinear space, we are only interested in the magnitude and rate differences.

The transformation process has several steps and we rely on the 3 points shown in Fig. 6-35, naming each vector as target (tgt) or interceptor (int), the coordinate system (ECI, RSW, PQW, SEZ), and the point (1-3). We start in Fig. 6-35 at point 1. The target satellite state is used to define the RSW coordinates and the target and interceptor satellites are transformed to these RSW frame. Next, we setup latitude and longitude-like transformations to move the target RSW vector to the interceptor longitude location at

† The EQC system has equal distances for polar and equatorial mappings. Using a central meridian (λ_o) and latitude (ϕ), the following equations apply. $x = \lambda - \lambda_o$, with the reverse equation $\lambda = \lambda_o + x \sec(\phi)$, and $y = \phi$.

point 2 in Fig. 6-35. Then we rotate the target RSW vector to a RSW frame centered at the interceptor position to find the solution. Finding the arc length from the true anomaly (and reverse) can be challenging. Finally, we “unwrap” the final position and determine the components.

Figures 6-34 and 6-35 are helpful in determining the rates of change for the velocity expressions. The transformation is shown in Algorithm 49.

ALGORITHM 49: HilleQCM to ECI

$$([\dot{\vec{r}}_{tgt}, \dot{\vec{v}}_{tgt}]_{ECI}, x, y, z, \dot{x}, \dot{y}, \dot{z} \Rightarrow [\dot{\vec{r}}_{int}, \dot{\vec{v}}_{int}]_{ECI})$$

$$[\dot{\vec{r}}_{tgt}, \dot{\vec{v}}_{tgt}]_{ECI} \Rightarrow [\hat{R}\hat{S}\hat{W}]_1$$

$$[\dot{\vec{r}}_{tgt}, \dot{\vec{v}}_{tgt}]_{RSW1} = [\hat{R}\hat{S}\hat{W}]_1 [\dot{\vec{r}}_{tgt}, \dot{\vec{v}}_{tgt}]_{ECI}$$

$$\dot{\vec{r}}_{tgtRSW1}, \dot{\vec{v}}_{tgtRSW1} \Rightarrow \dot{\vec{e}}_{tgt}, a_{tgt}$$

$$\lambda_{perigee} = \text{TAN}^{-1}\left(\frac{\dot{e}_{tgt}(S)}{e_{tgt}(R)}\right)$$

$$\nu_1 = -\lambda_{perigee}$$

Find the ending true anomaly (ν_2) via numerical approaches or an elliptic integral.

$$\dot{\vec{r}}_{tgtPQW2} = \begin{bmatrix} \frac{p_{tgt}\cos(\nu_2)}{1+e\cos(\nu_2)} \\ \frac{p_{tgt}\sin(\nu_2)}{1+e\cos(\nu_2)} \\ 0 \end{bmatrix} \quad \dot{\vec{v}}_{tgtPQW2} = \begin{bmatrix} -\sqrt{\frac{\mu}{p_{tgt}}}\sin(\nu_2) \\ \sqrt{\frac{\mu}{p_{tgt}}}(e_{tgt} + \cos(\nu_2)) \\ 0 \end{bmatrix}$$

$$[\dot{\vec{r}}_{tgt}, \dot{\vec{v}}_{tgt}]_{PQW2} \Rightarrow [\hat{R}\hat{S}\hat{W}]_2$$

$$[\dot{\vec{r}}_{tgt}, \dot{\vec{v}}_{tgt}]_{RSW2} = [\hat{R}\hat{S}\hat{W}]_2 [\dot{\vec{r}}_{tgt}, \dot{\vec{v}}_{tgt}]_{PQW2}$$

$$\Delta\phi_{int} = \text{SIN}^{-1}\left(\frac{z}{|\dot{\vec{r}}_{tgtRSW2}|}\right)$$

$$\Delta\lambda = \nu_2 - \nu_1$$

$$[\hat{r}_{int}]_{RSW1} = \begin{bmatrix} \text{SIN}(\Delta\phi_{int}) \\ \text{COS}(\Delta\phi_{int})\text{SIN}(\Delta\lambda) \\ \text{COS}(\Delta\phi_{int})\text{COS}(\Delta\lambda) \end{bmatrix}$$

$$[\dot{\vec{r}}_{int}]_{SEZ2} = \begin{bmatrix} \sin(\Delta\phi_{int})\cos(\Delta\lambda) & \sin(\Delta\phi_{int})\sin(\Delta\lambda) & -\cos(\Delta\phi_{int}) \\ -\sin(\Delta\lambda) & \cos(\Delta\lambda) & 0 \\ \cos(\Delta\phi_{int})\cos(\Delta\lambda) & \cos(\Delta\phi_{int})\sin(\Delta\lambda) & \sin(\Delta\phi_{int}) \end{bmatrix} [\dot{\vec{r}}_{int}]_{RSW1}$$

$$[\dot{\vec{r}}_{int}]_{SEZ2}(Z) = x + \dot{\vec{r}}_{tgtRSW2}(R)$$

$$r_{scale} = \frac{[\dot{\vec{r}}_{int}]_{SEZ2}(Z)}{[\dot{\vec{r}}_{int}]_{RSW1}(Z)}$$

$$\dot{\vec{r}}_{intRSW1} = r_{scale}[\dot{\vec{r}}_{int}]_{RSW1}$$

$$\omega_1 = \frac{\dot{\vec{v}}_{tgtRSW1}(R)}{|\dot{\vec{r}}_{tgtRSW1}|} \quad \dot{\lambda} = \frac{\dot{y} + \omega_1 |\dot{\vec{r}}_{tgtRSW1}|}{p_{tgt}} \quad \dot{\phi} = \frac{\dot{z}}{p_{tgt}}$$

$$\dot{\vec{v}}_{intSEZ} = \begin{bmatrix} -\dot{\phi}r_{scale} \\ \dot{\lambda}r_{scale}\cos(\Delta\phi_{int}) \\ \dot{z} + \dot{\vec{v}}_{tgtRSW2}(W) \end{bmatrix}$$

$$\dot{\vec{v}}_{intRSW1} = \begin{bmatrix} \sin(\Delta\phi_{int})\cos(\Delta\lambda) & -\sin(\Delta\lambda) & \cos(\Delta\phi_{int})\cos(\Delta\lambda) \\ \sin(\Delta\phi_{int})\sin(\Delta\lambda) & \cos(\Delta\lambda) & \cos(\Delta\phi_{int})\sin(\Delta\lambda) \\ -\cos(\Delta\phi_{int}) & 0 & \sin(\Delta\phi_{int}) \end{bmatrix} \dot{\vec{v}}_{intSEZ}$$

$$[\dot{\vec{r}}_{intEQCM}, \dot{\vec{v}}_{intEQCM}] = [\hat{R}\hat{S}\hat{W}]_1^T [\dot{\vec{r}}_{intRSW1}, \dot{\vec{v}}_{intRSW1}]$$

The reverse process is similar, but it has several subtleties and is thus presented separately. The curved arc distance is recovered using numerical techniques.

ALGORITHM 50: ECI to HillEQCM

$$([\dot{\vec{r}}_{tgt}, \dot{\vec{v}}_{tgt}]_{ECI}, [\dot{\vec{r}}_{int}, \dot{\vec{v}}_{int}]_{ECI}) \Rightarrow (x, y, z, \dot{x}, \dot{y}, \dot{z})$$

$$[\dot{\vec{r}}_{tgt}, \dot{\vec{v}}_{tgt}]_{ECI} \Rightarrow [\hat{R}\hat{S}\hat{W}]_1$$

$$[\dot{\vec{r}}_{tgt}, \dot{\vec{v}}_{tgt}]_{RSW1} = [\hat{R}\hat{S}\hat{W}]_1 [\dot{\vec{r}}_{tgt}, \dot{\vec{v}}_{tgt}]_{ECI}$$

$$[\dot{\vec{r}}_{int}, \dot{\vec{v}}_{int}]_{RSW1} = [\hat{R}\hat{S}\hat{W}]_1 [\dot{\vec{r}}_{int}, \dot{\vec{v}}_{int}]_{ECI}$$

$$\Delta\phi_{int} = \sin^{-1}\left(\frac{z}{|\dot{\vec{r}}_{tgtRSW2}|}\right) \quad \Delta\lambda = \tan^{-1}\left(\frac{\dot{\vec{r}}_{intRSW1}(S)}{\dot{\vec{r}}_{intRSW1}(R)}\right)$$

$$\dot{\vec{r}}_{tgtRSW1}, \dot{\vec{v}}_{tgtRSW1} \Rightarrow \dot{\vec{e}}_{tgt}, a_{tgt}$$

$$\lambda_{perigee} = \tan^{-1}\left(\frac{\hat{\vec{e}}_{tgt}(S)}{\hat{\vec{e}}_{tgt}(R)}\right)$$

$$\nu_1 = -\lambda_{perigee} \quad \nu_2 = \Delta\lambda - \lambda_{perigee}$$

$$\dot{\vec{r}}_{tgtPQW2} = \begin{bmatrix} p_{tgt} \cos(\nu_2) \\ \frac{p_{tgt} \cos(\nu_2)}{1 + e \cos(\nu_2)} \\ \frac{p_{tgt} \sin(\nu_2)}{1 + e \cos(\nu_2)} \\ 0 \end{bmatrix} \quad \dot{\vec{v}}_{tgtPQW2} = \begin{bmatrix} -\sqrt{\frac{\mu}{p_{tgt}}} \sin(\nu_2) \\ \sqrt{\frac{\mu}{p_{tgt}}} (e_{tgt} + \cos(\nu_2)) \\ 0 \end{bmatrix}$$

$$[\dot{\vec{r}}_{tgt}, \dot{\vec{v}}_{tgt}]_{PQW2} \Rightarrow [\hat{R} \hat{S} \hat{W}]_2$$

$$[\dot{\vec{r}}_{tgt}, \dot{\vec{v}}_{tgt}]_{RSW2} = [\hat{R} \hat{S} \hat{W}]_2 [\dot{\vec{r}}_{tgt}, \dot{\vec{v}}_{tgt}]_{PQW2}$$

$$[\dot{\vec{r}}_{int}, \dot{\vec{v}}_{int}]_{SEZ} = \begin{bmatrix} \sin(\Delta\phi_{int}) \cos(\Delta\lambda) & \sin(\Delta\phi_{int}) \sin(\Delta\lambda) & -\cos(\Delta\phi_{int}) \\ -\sin(\Delta\lambda) & \cos(\Delta\lambda) & 0 \\ \cos(\Delta\phi_{int}) \cos(\Delta\lambda) & \cos(\Delta\phi_{int}) \sin(\Delta\lambda) & \sin(\Delta\phi_{int}) \end{bmatrix} [\dot{\vec{r}}_{int}, \dot{\vec{v}}_{int}]_{RSW1}$$

$$\dot{\vec{r}}_{intEQCM} = \begin{bmatrix} \dot{\vec{r}}_{intSEZ2}(Z) - \dot{\vec{r}}_{tgtRSW2}(R) \\ arc1 \\ \Delta\phi_{int} r_{tgt2} \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

$$\omega_1 = \frac{\dot{\vec{v}}_{tgtRSW1}(S)}{|\dot{\vec{r}}_{tgtRSW1}|} \quad \dot{\lambda} = \frac{\dot{\vec{v}}_{intSEZ2}(E)}{|\dot{\vec{r}}_{int}| \cos(\Delta\phi_{int})} \quad \dot{\phi} = \frac{-\dot{\vec{v}}_{intSEZ2}(S)}{|\dot{\vec{r}}_{int}|}$$

$$\dot{\vec{v}}_{intEQCM} = \begin{bmatrix} \dot{\vec{v}}_{intSEZ2}(Z) - \dot{\vec{v}}_{tgtRSW2}(R) \\ \dot{\lambda} r_{tgt2} - \omega_1 |\dot{\vec{r}}_{tgtRSW1}| \\ \dot{\phi} r_{tgt2} \end{bmatrix} = \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{bmatrix}$$

Once these algorithms are programmed, they should operate consistently in both directions. Tests showed the routines are accurate to about 1.0×10^{-6} mm in position and 1.0×10^{-6} mm/s in velocity. There are two general areas to explore. First, the error in the original Hill's formulation for circular orbits, and how much error is present in elliptical orbits. Second, the error in the original Hill's formulation resulting from the use of two-body motion.

The error from the circular orbit restriction is evident when looking at initial displacements, and what they would be in ECI space. Figure 6-36 shows the positional differences between a Hill's propagation, and a two-body numerical propagation of the two LEO satellites ($a = 6878$ km, $e = 0.00$) in close proximity (each position starts 100 m away from the target, and each velocity differs by 0.01 m/s). For circular orbits using two-body propagation, the Hill's propagation should closely match the two satellite independent propagations (also using two-body techniques) which are converted to a Hill's equivalent frame.

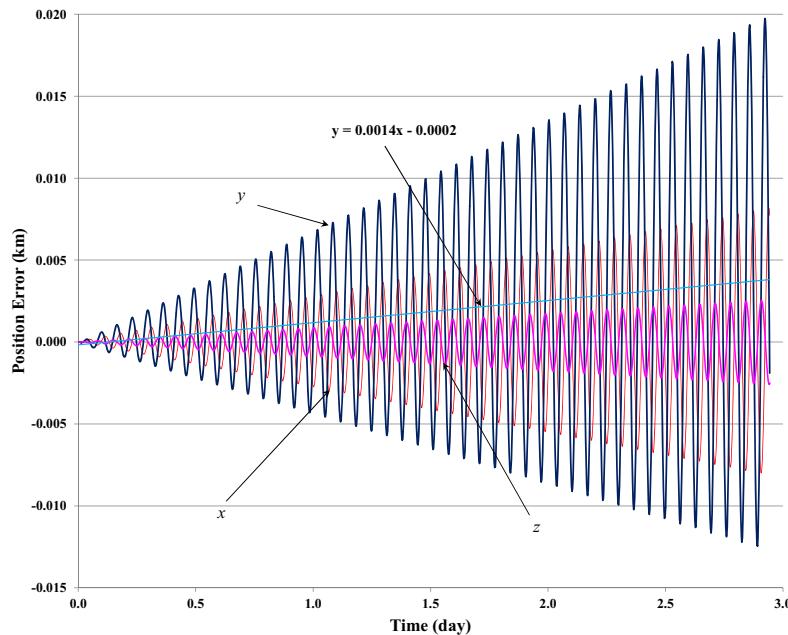


Figure 6-36. Comparing Hill's and Keplerian Motion—Circular. This figure shows the differences (averaging about 1.4 m /day) of simulations using Hill's and Keplerian propagations. The straight line is a least squares fit to the data in km. The period of the target is about 95 minutes.

Figure 6-37 shows data for a LEO slightly eccentric orbit ($a = 8164$ km, $e = 0.152$). Notice the significant increase in uncertainty from the Hill's solutions (about 9 km / day). This makes sense because Hill's assumes circular motion, and the ωt term experiences the largest variation from a “mean” value in an eccentric orbit. Selecting the alti-

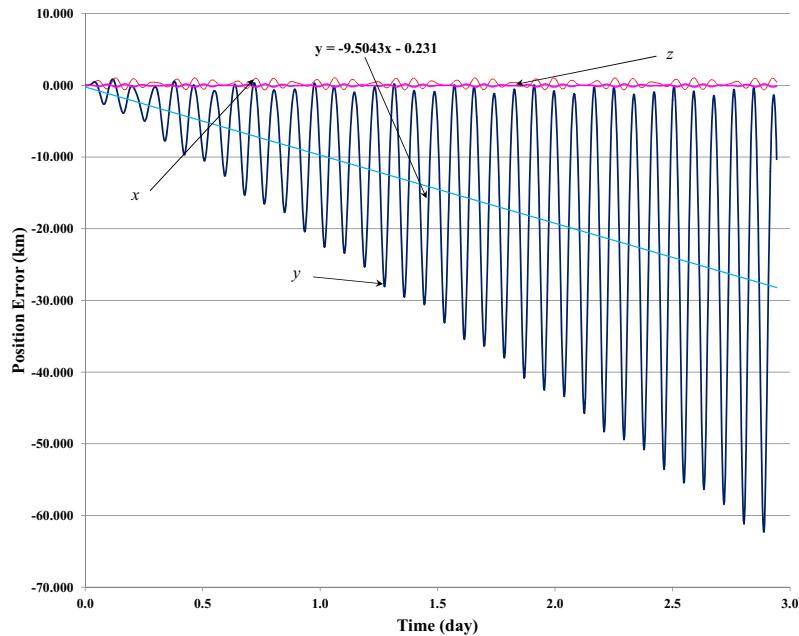


Figure 6-37. Comparing Hill's and Numerically Integrated Motion—LEO Eccentric. This figure shows the component differences resulting from numerically propagating (70×70 gravity, drag, third-body, solar radiation pressure) the target and interceptor satellites. Although the eccentricity is only 0.15, the error from Hill's is actually larger than the HEO case.

tude for which the ωt term is found can improve the results, but each revolution will have maximum departures from this central value. The effect is exacerbated at lower altitudes and thus, the worst place to use Hill's equations for eccentric orbits is in LEO. The algorithms in this section are recommended for relative motion on eccentric LEO orbits to avoid these difficulties.

When adding force models, most studies focus on adding J_2 or another single perturbation force. We chose to incorporate complete force models. In Fig. 6-38, we used a 70×70 gravitational model, NRL-MSIS00 drag, tides, solar radiation pressure, and third-body forces for a HEO satellite ($a = 25533$ km, $e = 0.73$). The additional uncertainty for an eccentric orbit is not unexpected because Hill's assumes circular orbits. Notice that the error growth is actually less than the LEO case—due to the ωt effect mentioned previously.

Additional tests showed that using a two-body or a full force model numerical propagation did not usually make much, if any difference, in comparison to Hill's. The exception was for LEO orbits where perturbations cause larger variations. For the case in Fig. 6-38, the trend line grew at 4.168 km / day (compared to 4.188) for a complete force model propagation of each satellite. This is because Hill's considers only the relative positions, and is remarkably accurate for that. Unfortunately, if you need to interface

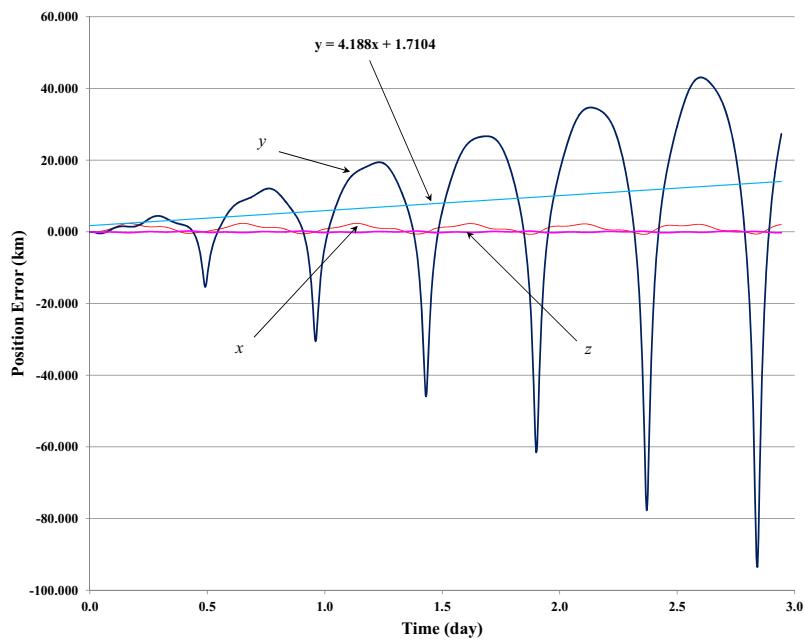


Figure 6-38. Comparing Hill's and Numerically Integrated Motion—Highly Elliptical. This figure shows the component differences (RSW) resulting from numerically propagating (70×70 gravity, tides, drag, third-body, solar radiation pressure) the target and interceptor satellites and converting to an equivalent Hill's representation, compared to a Hill's propagation. The straight line now shows about 4.18 km /day.

with another organization and need to make the transformation to inertial (ECI for instance) space, Hill's will introduce significant error for eccentric orbits.

The assumptions used in the Hill's formulation clearly introduce error into the solution. However, they are still widely used for initial planning of eccentric orbits. Additional realism may be inserted by using the perturbed formulations, but developing these equations, even for just zonal effects, can be challenging. Because the Hill's formulation works well on all orbit types for the *relative* positions, they are often sufficient for an application. However, LEO eccentric orbits require the equations from this section to properly account for the true motion in inertial space.

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 110 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 6-11 [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. 7. 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: $a = 6563.3698$ 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 ($a = 6727.7061$ km, $e = 0.000\ 744\ 4$, $i = 51.6598^\circ$, $\Omega = 159.7517^\circ$, $\omega = 117.6835^\circ$, $M = 242.4909^\circ$ on the epoch of December 24, 1998, at 3:40:53.6148 UT). 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 $0\text{ km} \times 80\text{ km}$ 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° 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* 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 6.8.3 shows how we can transform results to position and velocity vectors. Analyze the results in Fig. 6-28. How accurate are they? Hint: use a propagation technique from Chap. 8 or Chap. 9 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 6-2.
10. Can a combined bielliptic transfer with the inclination change at the farthest apogee be more efficient than a Hohmann transfer from LEO to GEO? Consider the entire range of total inclination changes from 0 to 180° .

422

ORBITAL MANEUVERING

6.8

CHAPTER 7

INITIAL ORBIT DETERMINATION

- 7.1 Historical Background
- 7.2 Observations of Range, Azimuth, and Elevation
- 7.3 Angles-only Observations
- 7.4 Mixed Observations
- 7.5 Three Position Vectors and Time
- 7.6 Two Position Vectors and Time—Lambert’s Problem
- 7.7 Application: Targeting Problem

7.1 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 geometri-cian 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 solution 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 fundamental 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. (9-12)].

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. 7.3.1, 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 Xavor von Zach (1754–1832) was the editor of *Monatliche 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 *Monatliche 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 astrodynamists, 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 a 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. 10), we couldn’t process data without the individual vectors determined through these techniques.

7.2 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. 7.1 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 that process angular information (see Sec. 7.3), range information allows us to analyze data faster, more simply, and more accurately. In Sec. 10.4.2 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 for general applications.

7.2.1 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, so 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. Some 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), combine the vectors in a common coordinate frame (ECEF), and then transform the position and velocity vectors to a frame for the particular application (ECI). Highly accurate simulations may also require deflection of the vertical calculations. The horizon will be perpendicular to a vertical line along geodetic or astronomical latitude. Figure 7-1 shows the geometry.

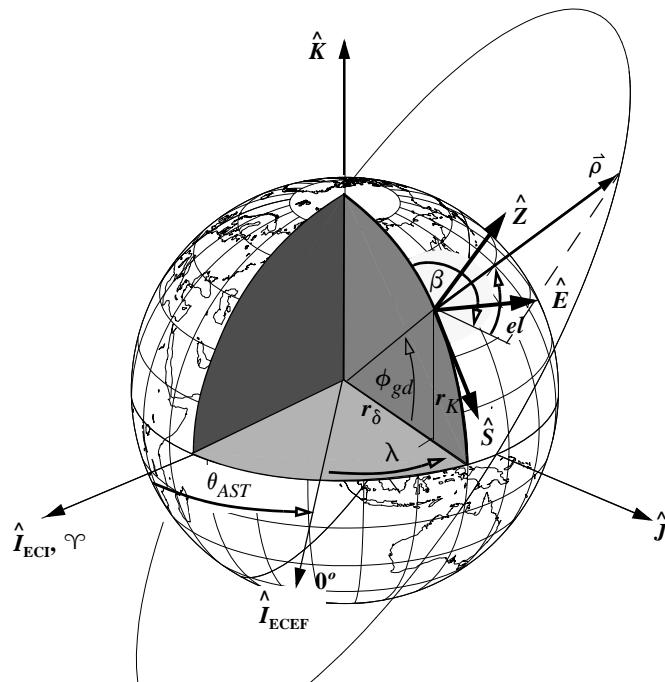


Figure 7-1. Geometry of Site Coordinates. The site's position depends on latitude and longitude, and the r_δ and r_K components shown above. Geodetic latitude is common, but other conventions exist. We must also know θ_{AST} for additional processing.

Distinguishing coordinate systems is required for accurate results. The ITRF system is a common choice because it's consistent with the body-fixed nature of the observations. If we require transformation to another system, the process can involve accounting for polar motion, sidereal time, nutation, and precession. See Sec. 3.7.

The first step is to find the site's position vector. Recalling the discussion in Sec. 3.2.2 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. (3-7). Using Fig. 7-1, we construct the position vector directly in the Geocentric Equatorial System. The site's position vector in the ECEF frame is*

$$\begin{aligned}\hat{\mathbf{r}}_{siteECEF} &= \begin{bmatrix} r_\delta \cos(\lambda) \\ r_\delta \sin(\lambda) \\ r_K \end{bmatrix} \\ &= \begin{bmatrix} (C_\oplus + h_{ellp}) \cos(\phi_{gd}) \cos(\lambda) \\ (C_\oplus + h_{ellp}) \cos(\phi_{gd}) \sin(\lambda) \\ (S_\oplus + h_{ellp}) \sin(\phi_{gd}) \end{bmatrix}\end{aligned}\quad (7-1)$$

Once we know the site's location, we find the satellite's position and velocity vectors in the topocentric coordinate system, SEZ. In Fig. 7-2, I have enlarged a section of Fig. 7-1 to show the geometry. Don't forget the deflection of the vertical if applicable.

Defining *elevation*, *el*, and *azimuth*, β , is important. As discussed in Sec. 3.3, 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. 7-2, we can reconfirm Eq. (4-4) and Eq. (4-5). Notice we now have the site's position vector in the ECEF coordinate system, whereas the satellite's vectors are in the topocentric SEZ system. Because we want to determine the satellite's position and velocity vectors in the ECEF coordinate system, we'll transform from SEZ to ECEF, using one of the methods discussed in Sec. 3.4. From Eq. (3-28)

$$\begin{aligned}\hat{\rho}_{ECEF} &= [\text{ROT3}(-\lambda)][\text{ROT2}(-(90^\circ - \phi_{gd}))]\hat{\rho}_{SEZ} \\ \dot{\hat{\rho}}_{ECEF} &= [\text{ROT3}(-\lambda)][\text{ROT2}(-(90^\circ - \phi_{gd}))]\dot{\hat{\rho}}_{SEZ}\end{aligned}$$

We must reexamine the physical situation to complete the process. Figure 7-3 shows the geometry of the site and satellite. The geocentric position vector is

$$\hat{\mathbf{r}}_{ECEF} = \hat{\rho}_{ECEF} + \hat{\mathbf{r}}_{siteECEF} \quad (7-2)$$

The velocity, relative to the rotating frame, is simply the slant range velocity (see also Eq. (4-7)).

$$\hat{\mathbf{v}}_{ECEF} = \dot{\hat{\rho}}_{ECEF} \quad (7-3)$$

* For applications that don't distinguish between ECI and ECEF, we use the Local Sidereal Time, LST or θ_{LST} , in place of the site longitude. LST approximately locates the site at the time of each observation. Remember the time is usually given as UTC. To obtain LST, we must first convert UTC to UT1 using the EOP data. See Algorithm 16 for additional information.

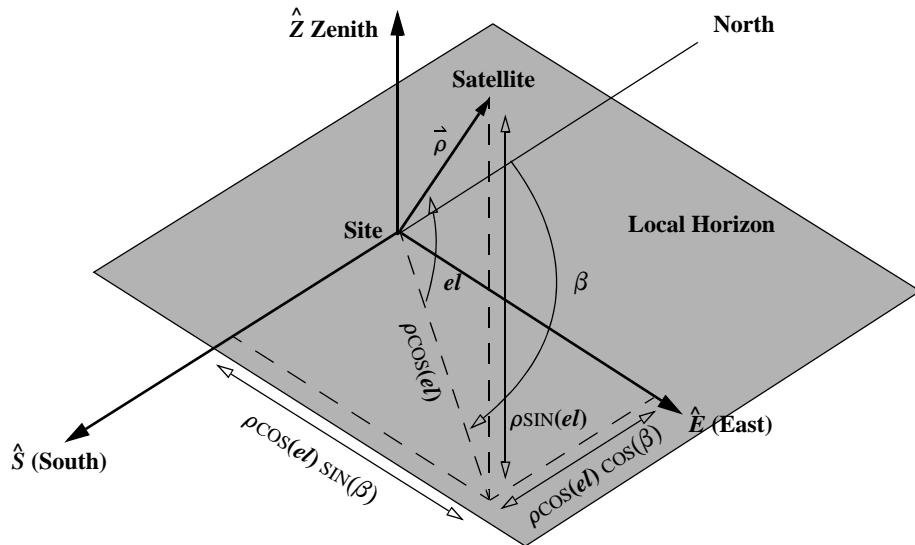


Figure 7-2. Geometry of the Sensor Site. This figure shows the site geometry from a different perspective. Elevation is measured positive up from the local horizon. Notice the positive definition because we measure azimuth from north.

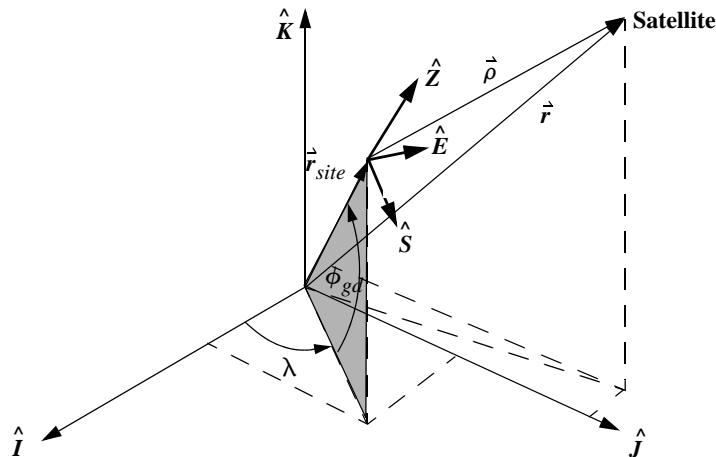


Figure 7-3. Site-to-Satellite Geometry. To complete *SITE-TRACK*, we use the slant-range vector and the site vector and rotate through geodetic latitude, ϕ_{gd} , and longitude, λ .

If the less accurate approximation is used, we must add the site's relative motion (Sec. 3.4.2) because the (FK5) reduction is no longer used. Thus, the velocity for this case is

$$\dot{\vec{v}}_{ECI} \equiv \dot{\vec{r}}_{ECI} + \vec{\omega}_{\oplus} \times \vec{r}_{siteECI}$$

Implementing the SITE-TRACK Algorithm

Summarizing, we observe a subtlety in the formulation—we need geodetic latitude. Remember the discussion on geodesy from Chap. 3 (Sec. 3.2.3). 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 scaling. 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 valid.

To speed computer 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 ECEF 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 ideas will yield a robust and fast algorithm.

ALGORITHM 51: SITE-TRACK

$(\phi_{gd}, \lambda, h_{ellp}, \rho, \beta, el, \dot{\rho}, \dot{\beta}, \dot{el}, yr, mo, day, UTC, \Delta UT1, \Delta AT, x_p, y_p \Rightarrow$

$$\vec{r}_{ECI}, \dot{\vec{v}}_{ECI})$$

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

$$r_{\delta} = (C_{\oplus} + h_{ellp}) \cos(\phi_{gd})$$

$$r_K = (S_{\oplus} + h_{ellp}) \sin(\phi_{gd})$$

$$\vec{r}_{siteECEF} = \begin{bmatrix} r_{\delta} \cos(\lambda) \\ r_{\delta} \sin(\lambda) \\ r_K \end{bmatrix}$$

$$\vec{\rho}_{SEZ} = \begin{bmatrix} -\rho \cos(el) \cos(\beta) \\ \rho \cos(el) \sin(\beta) \\ \rho \sin(el) \end{bmatrix}$$

$$\begin{aligned}\dot{\rho}_{SEZ} &= \begin{bmatrix} -\dot{\rho}\cos(el)\cos(\beta) + \rho\sin(el)\cos(\beta)\dot{el} + \rho\cos(el)\sin(\beta)\dot{\beta} \\ \dot{\rho}\cos(el)\sin(\beta) - \rho\sin(el)\sin(\beta)\dot{el} + \rho\cos(el)\cos(\beta)\dot{\beta} \\ \dot{\rho}\sin(el) + \rho\cos(el)\dot{el} \end{bmatrix} \\ \dot{\rho}_{ECEF} &= [\text{ROT3}(-\lambda)][\text{ROT2}(-(90^\circ - \phi_{gd}))]\dot{\rho}_{SEZ} = \left[\frac{ECEF}{SEZ}\right]\dot{\rho}_{SEZ} \\ \dot{\rho}_{ECEF} &= [\text{ROT3}(-\lambda)][\text{ROT2}(-(90^\circ - \phi_{gd}))]\dot{\rho}_{SEZ} = \left[\frac{ECEF}{SEZ}\right]\dot{\rho}_{SEZ} \\ \left[\frac{ECEF}{SEZ}\right] &= \begin{bmatrix} \sin(\phi_{gd})\cos(\lambda) & -\sin(\lambda) & \cos(\phi_{gd})\cos(\lambda) \\ \sin(\phi_{gd})\sin(\lambda) & \cos(\lambda) & \cos(\phi_{gd})\sin(\lambda) \\ -\cos(\phi_{gd}) & 0 & \sin(\phi_{gd}) \end{bmatrix} \\ \dot{r}_{ECEF} &= \dot{\rho}_{ECEF} + \dot{r}_{siteECEF} \\ \dot{v}_{ECEF} &= \dot{\rho}_{ECEF}\end{aligned}$$

$$FK5(\dot{r}_{ECEF}, \dot{v}_{ECEF}, yr, mo, day, UTC, \Delta UT1, \Delta AT, x_p, y_p \Rightarrow \dot{r}_{ECI}, \dot{v}_{ECI})$$


Example 7-1: Using Algorithm 51 to Determine the SITE-TRACK.

GIVEN: $\phi_{gd} = 39.007^\circ$, $\lambda = 104.883^\circ W$, alt = 2187 m

with an observation of $\rho = 604.68$ km, $\beta = 205.6^\circ$, $el = 30.7^\circ$, and rate values of $\dot{\rho} = 2.08$ km/s, $\dot{\beta} = 0.15^\circ/s$, and $\dot{el} = 0.17^\circ/s$ on May 20, 1995 at 03:17:02.00 UT

Let $\Delta UT1 = 0.0$ s and $\Delta AT = 29$ sec. Also, let $x_p = y_p = 0.0$.

FIND: \dot{r}_{site} , \dot{r} , \dot{v}

The first task is to determine the site's position vector. The auxiliary terms are

$$C_\oplus = \frac{R_\oplus}{\sqrt{1 - e_\oplus^2 \sin^2(\phi_{gd})}} = 6386.6116 \text{ km}$$

$$S_\oplus = C_\oplus(1 - e_\oplus^2) = 6343.8545 \text{ km}$$

The vertical and horizontal components are

$$r_\delta = (C_\oplus + h_{ellp}) \cos(\phi_{gd}) = 4964.5377 \text{ km}$$

$r_K = (S_\oplus + h_{ellp}) \sin(\phi_{gd}) = 3994.2955 \text{ km}$ The site vector in the ECEF coordinate system is

$$\dot{r}_{site} = \begin{bmatrix} r_\delta \cos(\lambda) \\ r_\delta \sin(\lambda) \\ r_K \end{bmatrix} = \begin{bmatrix} -1275.1219 \\ -4797.9890 \\ 3994.2975 \end{bmatrix} \text{ km}$$

The slant-range position and velocity vectors in the SEZ coordinate system are

$$\begin{aligned}\hat{\rho}_{SEZ} &= \begin{bmatrix} -\rho \cos(eI) \cos(\beta) \\ \rho \cos(eI) \sin(\beta) \\ \rho \sin(eI) \end{bmatrix} = \begin{bmatrix} 468.8947 \\ -224.6567 \\ 308.7151 \end{bmatrix} \text{ km} \\ \dot{\hat{\rho}}_{SEZ} &= \begin{bmatrix} -\dot{\rho} \cos(eI) \cos(\beta) + \rho \sin(eI) \cos(\beta) eI + \rho \cos(eI) \sin(\beta) \dot{\beta} \\ \dot{\rho} \cos(eI) \sin(\beta) - \rho \sin(eI) \sin(\beta) eI + \rho \cos(eI) \cos(\beta) \dot{\beta} \\ \dot{\rho} \sin(eI) + \rho \cos(eI) eI \end{bmatrix} = \begin{bmatrix} 0.198714 \\ -1.604565 \\ 2.604609 \end{bmatrix} \text{ km/s}\end{aligned}$$

The transformation matrix from SEZ to ECEF is determined as a complete matrix.

$$\begin{aligned}[\frac{ECEF}{SEZ}] &= \begin{bmatrix} \sin(\phi_{gd}) \cos(\lambda) & -\sin(\lambda) & \cos(\phi_{gd}) \cos(\lambda) \\ \sin(\phi_{gd}) \sin(\lambda) & \cos(\lambda) & \cos(\phi_{gd}) \sin(\lambda) \\ -\cos(\phi_{gd}) & 0 & \sin(\phi_{gd}) \end{bmatrix} \\ [\frac{ECEF}{SEZ}] &= \begin{bmatrix} -0.161663 & 0.966452 & -0.199587 \\ -0.608300 & -0.256846 & -0.751000 \\ -0.777069 & 0 & 0.629415 \end{bmatrix}\end{aligned}$$

Doing the matrix operations gives the slant range position and velocity vectors in the ECEF frame.

$$\begin{aligned}\hat{\rho}_{ECEF} &= [\frac{ECEF}{SEZ}] \hat{\rho}_{SEZ} = \begin{bmatrix} -354.538409 \\ -459.371529 \\ -170.053572 \end{bmatrix} \text{ km and} \\ \dot{\hat{\rho}}_{ECEF} &= [\frac{ECEF}{SEZ}] \dot{\hat{\rho}}_{SEZ} = \begin{bmatrix} -2.102707 \\ -1.664813 \\ 1.484966 \end{bmatrix} \text{ km/s}\end{aligned}$$

Next, determine the position and velocity vectors in the ECEF frame:

$$\hat{r} = \hat{\rho}_{ECEF} + \hat{r}_{siteECEF} = \begin{bmatrix} -1629.66032 \\ -5257.36056 \\ 3824.24388 \end{bmatrix} \text{ km and } \hat{v} = \dot{\hat{\rho}}_{ECEF}$$

Finally, we determine the ECI vectors. Using Algorithm 24,

$$\begin{aligned}\hat{r} &= \begin{bmatrix} -5505.504883 \\ 56.449170 \\ 3821.871726 \end{bmatrix} \text{ km} \\ \hat{v} &= \begin{bmatrix} -2.201803 \\ 1.136781 \\ 1.484060 \end{bmatrix} \text{ km/s}\end{aligned}$$

If we approximate the results with LST and have no rigorous transformation of coordinates:

$$\blacktriangle \quad \hat{r} = \begin{bmatrix} -5503.79562 \\ 62.28191 \\ 3824.24480 \end{bmatrix} \text{ km} \quad \dot{\hat{v}} = \dot{\hat{\rho}}_{IJK} + \hat{\omega} \oplus \times \hat{r}_{siteIJK} = \begin{bmatrix} -2.199987 \\ 1.139111 \\ 1.484966 \end{bmatrix} \text{ km/s}$$

7.3 Angles-only Observations

Modifications of the *SITE-TRACK* procedure we've discussed do much of the processing for today's systems, but some measurements need different techniques. Angles-only methods work for angular data; yet, they're fundamentally limited because they lack range information. Iteration or more processing overcomes this problem, but not without additional complexity and computation. Angles-only techniques are used for optical sensors, such as the early Baker-Nunn camera or the *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 satellite magnitude and lighting determine if it is visible to the sensor.

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 7-4 shows such an example case for which an optical site provides angular data from at least three line-of-sight vectors. We may tend to view the unit vectors as position vectors. They're not.

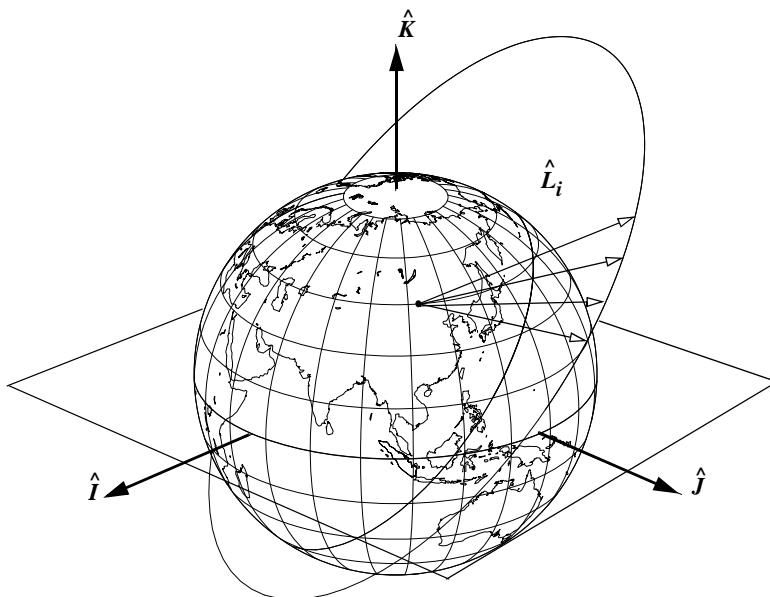


Figure 7-4. Geometry of Angles-only Observations. The underlying principle for the angles-only technique is the use of topocentric angles to form line-of-sight vectors. For objects that are very distant, the distinction for topocentric values diminishes.

We'll explore several techniques in this section: Laplace's, Gauss's, Double- r iteration, and Gooding. You should know all these methods because each solves one aspect of the problem. The first two are generally best for interplanetary operations, while the last two are better for Earth orbiting satellites. Laplace's 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 satellites. Gauss's 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). 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. Gooding (1993 and 1997) presents another robust approach that enjoys use in commercial operations. Finally, Der (2012) has some interesting approaches to finding the real roots that may warrant further investigation.

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 $[(\alpha_{t1}, \delta_{t1}), (\alpha_{t2}, \delta_{t2}), (\alpha_{t3}, \delta_{t3})]$ 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 4-1).^{*} Finally, it's crucial to use a common reference frame for all processing. Typically, the observations are obtained in an Earth-fixed frame (ITRF). The calculations often proceed in a common inertial frame—like the GCRF, but may take place in older frames like J2000 (IAU-76/FK5). The first step is to form line-of-sight unit vectors, \hat{L} , to the satellite at each observation time. Use Eq. (4-17):

$$\hat{L}_i = \begin{bmatrix} \cos(\delta_{ti}) \cos(\alpha_{ti}) \\ \cos(\delta_{ti}) \sin(\alpha_{ti}) \\ \sin(\delta_{ti}) \end{bmatrix} \quad i = 1 \dots 3 \quad (7-4)$$

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

$$\hat{r} = \rho \hat{L} + \hat{r}_{site} \quad (7-5)$$

This relation formalizes our use of the *topocentric* values for right ascension and declination.

* We can also use the topocentric angles β and el , but we must first convert them using Algorithm 28, or another appropriate technique.

We use Algorithm 51 to obtain the site vectors. As with the observations, we must form these vectors in the proper reference frame. If we dot Eq. (7-5) with itself, we get an expression for the satellite's distance:

$$r = \sqrt{\rho^2 + 2\rho\hat{L} \cdot \dot{\vec{r}}_{site} + \dot{\vec{r}}_{site}^2} \quad (7-6)$$

7.3.1 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.

You first do the preliminary calculations of Eq. (7-4) and Eq. (7-5) 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. (7-5) twice produces the velocity and acceleration for each observation:

$$\begin{aligned} \dot{\vec{r}} &= \dot{\rho}\hat{L} + \rho\dot{\hat{L}} + \dot{\vec{r}}_{site} \\ \ddot{\vec{r}} &= \ddot{\rho}\hat{L} + 2\dot{\rho}\dot{\hat{L}} + \ddot{\rho}\hat{L} + \ddot{\vec{r}}_{site} \end{aligned} \quad (7-7)$$

The process assumes two-body motion [Eq. (1-14)], so we may substitute this relation for the satellite's acceleration above and combine with Eq. (7-5) to get

$$-\frac{\mu}{r^3}(\rho\hat{L} + \dot{\vec{r}}_{site}) = \ddot{\rho}\hat{L} + 2\dot{\rho}\dot{\hat{L}} + \ddot{\rho}\hat{L} + \ddot{\vec{r}}_{site}$$

Rearranging gives us

$$\ddot{\rho}\hat{L} + 2\dot{\rho}\dot{\hat{L}} + \rho\left(\ddot{\hat{L}} + \frac{\mu}{r^3}\hat{L}\right) = -\ddot{\vec{r}}_{site} - \frac{\mu}{r^3}\dot{\vec{r}}_{site} \quad (7-8)$$

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),

$$\dot{\vec{r}}(t) = \sum_{i=1}^n \vec{r}_i \prod_{k \neq i} \frac{t - t_k}{t_i - t_k} \quad (7-9)$$

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. 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. 7.5.2) approach to determine the average site velocity and acceleration vectors. Again, use Lagrange's interpolation formula [Eq. (7-9)], with a shorthand notation for the change in the times (Escobal [1965] 1985:268).

$$\tau_i = t_i - t_2$$

$$\dot{\vec{r}}_{site_2} = -\frac{\tau_3}{\tau_1(\tau_1-\tau_3)} \dot{\vec{r}}_{site_1} - \frac{\tau_3+\tau_1}{\tau_1\tau_3} \dot{\vec{r}}_{site_2} - \frac{\tau_1}{\tau_3(\tau_3-\tau_1)} \dot{\vec{r}}_{site_3}$$

Find the acceleration equation by the interpolation formula:

$$\ddot{\vec{r}}_{site_2} = \frac{2}{\tau_1(\tau_1-\tau_3)} \dot{\vec{r}}_{site_1} + \frac{2}{\tau_1\tau_3} \dot{\vec{r}}_{site_2} + \frac{2}{\tau_3(\tau_3-\tau_1)} \dot{\vec{r}}_{site_3}$$

Referring to the acceleration [Eq. (7-8)], 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. (7-8) become

$$\begin{bmatrix} \hat{L} + 2\dot{\hat{L}} + \frac{\mu}{r^3}\hat{L} \\ \ddot{\rho} \\ \dot{\rho} \\ \rho \end{bmatrix} = -\begin{bmatrix} \ddot{\hat{r}}_{site} + \frac{\mu}{r^3}\hat{r}_{site} \end{bmatrix}$$

They can be solved using determinants and Cramer's rule (see Sec. C.3). Find the *determinant*, D , for the system of equations in slant range:

$$D = \left| \begin{array}{c} \hat{L} + 2\dot{\hat{L}} + \frac{\mu}{r^3}\hat{L} \\ \ddot{\rho} \\ \dot{\rho} \\ \rho \end{array} \right| \quad (7-10)$$

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.3) yields

$$D = 2 \left| \begin{array}{c} \hat{L} + \dot{\hat{L}} + \frac{\ddot{\hat{r}}_{site} + \frac{\mu}{r^3}\hat{r}_{site}}{2} \end{array} \right|$$

Now apply Cramer's rule using Eq. (7-8). Because the third column represents the slant-range component, replace that column with the right-hand side of Eq. (7-8) (factor out the -1 from that column).

$$D\rho = -2 \left| \begin{array}{c} \hat{L} + \dot{\hat{L}} + \frac{\ddot{\hat{r}}_{site} + \frac{\mu}{r^3}\hat{r}_{site}}{2} \end{array} \right|$$

Then, split this determinant as follows:

$$D\rho = -2 \left| \begin{array}{c} \hat{L} + \dot{\hat{L}} + \frac{\ddot{\hat{r}}_{site}}{2} \end{array} \right| - 2 \frac{\mu}{r^3} \left| \begin{array}{c} \hat{L} + \dot{\hat{L}} + \hat{r}_{site} \end{array} \right|$$

Define the first determinant as D_1 and the second determinant as D_2 . Solve for the range, ρ , to get

$$\rho = \frac{-2D_1 - 2\mu D_2}{D} \quad (7-11)$$

Notice the determinant could be zero, which causes numerical difficulties.*

* Escobal ([1965] 1985:265–266) points out that a zero determinant 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 make an initial guess for the middle radius, 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. (7-11) into Eq. (7-6) results in

$$r_2^2 = \left(\frac{-2D_1}{D} - \frac{2\mu D_2}{r_2^3 D} \right)^2 + 2 \left(\frac{-2D_1}{D} - \frac{2\mu D_2}{r_2^3 D} \right) \hat{L}_2 \cdot \dot{\hat{r}}_{site_2} + r_{site_2}^2$$

Expanding this equation and letting C equal the dot product will produce an eighth-degree polynomial:

$$\begin{aligned} C &= \hat{L}_2 \cdot \dot{\hat{r}}_{site_2} \\ r_2^8 + \left(\frac{4CD_1}{D} - \frac{4D_1^2}{D^2} - r_{site_2}^2 \right) r_2^6 + \mu \left(\frac{4CD_2}{D} - \frac{8D_1 D_2}{D^2} \right) r_2^3 - \frac{4\mu^2 D_2^2}{D^2} &= 0 \end{aligned} \quad (7-12)$$

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. (7-11) for the slant range at the middle time.

Now, repeat this process for the velocity. Use Cramer's rule again starting from Eq. (7-8) and remember that the velocity component was the middle term in the determinant expression, so

$$D\dot{\rho} = - \left| \begin{array}{c|ccc} \hat{L} & \ddot{\hat{r}}_{site} & \frac{\mu}{r^3} \dot{\hat{r}}_{site} & \ddot{\hat{L}} \end{array} \right|$$

Split this determinant:

$$D\dot{\rho} = - \left| \begin{array}{c|cc} \hat{L} & \ddot{\hat{r}}_{site} & \ddot{\hat{L}} \\ \hline & -\frac{\mu}{r^3} & \end{array} \right| \left| \begin{array}{c} \hat{L} \\ \dot{\hat{r}}_{site} \\ \ddot{\hat{L}} \end{array} \right|$$

Define the first velocity determinant as D_3 and the second determinant as D_4 . Solving for the range rate, $\dot{\rho}$, results in

$$\dot{\rho} = - \frac{D_3}{D} - \frac{\mu D_4}{r^3 D} \quad (7-13)$$

So you can use the velocity determinants and solve for the middle velocity vector in Eq. (7-7):

$$\dot{\hat{v}}_2 = \dot{\rho} \hat{L}_2 + \rho \dot{\hat{L}}_2 + \dot{\hat{r}}_{site_2}$$

Notice this method doesn't require iteration if we find the correct root the first time in Eq. (7-12). 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 rou-

tine satellite observations using only sparse data (3–5 points/pass). The method becomes more efficient when combined with orbit determination techniques discussed in Chap. 10. But it's historically important, and it remains valid for interplanetary observations.

7.3.2 Gauss's Technique

Gauss's method using 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, Gauss's 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 Gauss's 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,

$$c_1 \hat{\vec{r}}_1 + c_2 \hat{\vec{r}}_2 + c_3 \hat{\vec{r}}_3 = \hat{\vec{0}} \quad (7-14)$$

We can get expressions for the coefficients by taking the cross product of the first and third position vectors with Eq. (7-14). Assuming c_2 is not zero, we have

$$\hat{\vec{r}}_1 \times \hat{\vec{r}}_3 (c_1) = \hat{\vec{r}}_2 \times \hat{\vec{r}}_3 (-c_2) \quad \hat{\vec{r}}_1 \times \hat{\vec{r}}_3 (c_3) = \hat{\vec{r}}_1 \times \hat{\vec{r}}_2 (-c_2)$$

But from Chap. 2's discussion of the “Classical Formula” on page 82, we know we can apply the f and g functions to form the position vectors, using the middle position and velocity vectors. Thus,

$$\hat{\vec{r}}_i = f_i \hat{\vec{r}}_2 + g_i \hat{\vec{v}}_2, \quad i = 1, 3$$

Eq. (7-14) 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:

$$c_1 = \frac{\hat{\vec{r}}_2 \times (f_3 \hat{\vec{r}}_2 + g_3 \hat{\vec{v}}_2)}{(f_1 \hat{\vec{r}}_2 + g_1 \hat{\vec{v}}_2) \times (f_3 \hat{\vec{r}}_2 + g_3 \hat{\vec{v}}_2)} = \frac{g_3}{f_1 g_3 - f_3 g_1}$$

$$c_3 = \frac{(f_1 \hat{\vec{r}}_2 + g_1 \hat{\vec{v}}_2) \times \hat{\vec{r}}_2}{(f_1 \hat{\vec{r}}_2 + g_1 \hat{\vec{v}}_2) \times (f_3 \hat{\vec{r}}_2 + g_3 \hat{\vec{v}}_2)} = \frac{-g_1}{f_1 g_3 - f_3 g_1}$$

If we knew the position and velocity vectors, the f and g functions could be solved to get a final solution [Eq. (2-64)]. Because the velocity is still unknown, we can use the *series* form of the f and g functions (see Chap. 2, “Series Forms,” page 86). 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:

$$c_1 = \frac{g_3}{f_1 g_3 - f_3 g_1} \approx \frac{\tau_3}{\tau_3 - \tau_1} + \frac{u \tau_3 ((\tau_3 - \tau_1)^2 - \tau_3^2)}{6(\tau_3 - \tau_1)} = a_1 + a_{1u} u$$

$$c_3 = \frac{-g_1}{f_1 g_3 - f_3 g_1} \approx -\frac{\tau_1}{\tau_3 - \tau_1} - \frac{u \tau_1 ((\tau_3 - \tau_1)^2 - \tau_1^2)}{6(\tau_3 - \tau_1)} = a_3 + a_{3u} u$$

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. (7-5) we know the relation for each observation, including the site vector. Substituting into Eq. (7-14) 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. (7-5)] and write the next relation in matrix form. This permits matrix solution of simultaneous equations and shows that we *must* solve them simultaneously.

$$\underbrace{\begin{bmatrix} \hat{L}_1 & \hat{L}_2 & \hat{L}_3 \end{bmatrix}}_{\mathbf{L}} \begin{bmatrix} c_1 \hat{\rho}_1 \\ c_2 \hat{\rho}_2 \\ c_3 \hat{\rho}_3 \end{bmatrix} = \begin{bmatrix} \hat{r}_{site_1} & \hat{r}_{site_2} & \hat{r}_{site_3} \end{bmatrix} \begin{bmatrix} -c_1 \\ -c_2 \\ -c_3 \end{bmatrix}$$

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.3 for additional information. With unit column vectors written as $\hat{L}_i = L_{xi}\hat{I} + L_{yi}\hat{J} + L_{zi}\hat{K}$, the solution is

$$\mathbf{L}^{-1} = \frac{\begin{bmatrix} L_{y2}L_{z3} - L_{y3}L_{z2} & -L_{y1}L_{z3} + L_{y3}L_{z1} & L_{y1}L_{z2} - L_{y2}L_{z1} \\ -L_{x2}L_{z3} + L_{x3}L_{z2} & L_{x1}L_{z3} - L_{x3}L_{z1} & -L_{x1}L_{z2} + L_{x2}L_{z1} \\ L_{x2}L_{y3} - L_{x3}L_{y2} & -L_{x1}L_{y3} + L_{x3}L_{y1} & L_{x1}L_{y2} - L_{x2}L_{y1} \end{bmatrix}}{|L|}$$

$$\begin{bmatrix} c_1 \rho_1 \\ c_2 \rho_2 \\ c_3 \rho_3 \end{bmatrix} = \mathbf{L}^{-1} \begin{bmatrix} \hat{\mathbf{r}}_{site_1} + \hat{\mathbf{r}}_{site_2} + \hat{\mathbf{r}}_{site_3} \end{bmatrix} \begin{bmatrix} -c_1 \\ -c_2 \\ -c_3 \end{bmatrix} \quad (7-15)$$

At this point, we can examine the slant-range equation from Eq. (7-15) 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 $\mathbf{L}^{-1}\mathbf{r}_{site}$ 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. (7-6), gives you an eighth-degree polynomial in r_2 . Although the form is similar to Eq. (7-12) 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. (7-12) apply here, too, so

$$\begin{aligned} C &= \hat{\mathbf{L}}_2 \cdot \hat{\mathbf{r}}_{site_2} \\ r_2^8 - (d_1^2 + 2Cd_1 + r_{site_2}^2)r_2^6 - 2\mu(Cd_2 + d_1d_2)r_2^3 - \mu^2d_2^2 &= 0 \end{aligned} \quad (7-16)$$

Once we pick an appropriate root, we can update our guess for the f and g series coefficient, u . After calculating 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 54), **HERRICK-GIBBS** (Algorithm 55), or **LAMBERT-UNIVERSAL** (Algorithm 58), 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 Gauss's, 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-degree 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. 10. Notice that the site vectors are all in the ECI frame for this algorithm. The technique seems to perform better with closely spaced observations.

ALGORITHM 52: ANGLES-ONLY GAUSS

$$(\hat{L}_1, \hat{L}_2, \hat{L}_3, JD_1, JD_2, JD_3, \hat{\vec{r}}_{site_1}, \hat{\vec{r}}_{site_2}, \hat{\vec{r}}_{site_3} \Rightarrow \hat{\vec{r}}_2, \hat{\vec{v}}_2)$$

$$\tau_1 = JD_1 - JD_2 \text{ and } \tau_3 = JD_3 - JD_2$$

$$a_1 = \frac{\tau_3}{\tau_3 - \tau_1} \quad a_{1u} = \frac{\tau_3((\tau_3 - \tau_1)^2 - \tau_3^2)}{6(\tau_3 - \tau_1)}$$

$$a_3 = -\frac{\tau_1}{\tau_3 - \tau_1} \quad a_{3u} = -\frac{\tau_1((\tau_3 - \tau_1)^2 - \tau_1^2)}{6(\tau_3 - \tau_1)}$$

$$\mathbf{L}^{-1} = \frac{\begin{bmatrix} L_{y2}L_{z3} - L_{y3}L_{z2} & -L_{y1}L_{z3} + L_{y3}L_{z1} & L_{y1}L_{z2} - L_{y2}L_{z1} \\ -L_{x2}L_{z3} + L_{x3}L_{z2} & L_{x1}L_{z3} - L_{x3}L_{z1} & -L_{x1}L_{z2} + L_{x2}L_{z1} \\ L_{x2}L_{y3} - L_{x3}L_{y2} & -L_{x1}L_{y3} + L_{x3}L_{y1} & L_{x1}L_{y2} - L_{x2}L_{y1} \end{bmatrix}}{|L|}$$

$$[\mathbf{M}] = [\mathbf{L}^{-1}][\mathbf{r}_{site}]$$

$$d_1 = M_{21}a_1 - M_{22} + M_{23}a_3$$

$$d_2 = M_{21}a_{1u} + M_{23}a_{3u}$$

$$C = \hat{L}_2 \cdot \hat{\vec{r}}_{site_2}$$

$$r_2^8 - (d_1^2 + 2Cd_1 + r_{site_2}^2)r_2^6 - 2\mu(Cd_2 + d_1d_2)r_2^3 - \mu^2d_2^2 = 0$$

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

$$\hat{\vec{r}}_i = \rho_i \hat{\vec{L}}_i + \hat{\vec{r}}_{site_i} \quad i = 1 \dots 3$$

Use **HERRICK-GIBBS** (small $\Delta\nu$) or **GIBBS** (large $\Delta\nu$) for $\hat{\vec{v}}_2$

Find semiparameter p using **RV2COE**

$$f_i = 1 - \left(\frac{r_i}{p}\right)(1 - \cos(\Delta\nu_i)) \quad i = 1, 3$$

$$g_i = \frac{r_i r_2 \sin(\Delta\nu_i)}{\sqrt{\mu p}} \quad i = 1, 3$$

$$c_1 = \frac{g_3}{f_1 g_3 - f_3 g_1} \quad c_3 = \frac{-g_1}{f_1 g_3 - f_3 g_1}$$

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

$$\hat{\vec{r}}_i = \rho_i \hat{\vec{L}}_i + \hat{\vec{r}}_{site_i} \quad i = 1 \dots 3$$

7.3.3 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 more efficient for observations which are far apart—something that Gauss's technique doesn't do well. 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 subsequent iterations use

the second portion to determine intermediate guesses, so it's important to have a modular routine. 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. (7-5). Consult Escobal ([1965] 1985:281) for hyperbolic relations. t_m describes the orbital motion, (+1) for direct and (-1) for retrograde orbits. All site vectors are ECI. The observations need not be evenly spaced, but must be topocentric for satellite orbits. The initial guess is important and the change on each iteration should be limited (0.005 is suggested). The technique seems to obtain better answers over a wide range of observational spacing, but some cases can cause it to fail.

ALGORITHM 53: ANGLES-DOUBLE-R

$$(\hat{L}_1, \hat{L}_2, \hat{L}_3, JD_1, JD_2, JD_3, \hat{r}_{site_1}, \hat{r}_{site_2}, \hat{r}_{site_3} \Rightarrow \hat{r}_2, \hat{v}_2)$$

$$\tau_1 = JD_1 - JD_2$$

$$\tau_3 = JD_3 - JD_2$$

Guess r_1 or let $r_1 = 12756.274$ km

Guess r_2 or let $r_2 = 12820.055$ 37km

$$c_i = 2\hat{L}_i \cdot \hat{r}_{site_i} \quad i = 1, 2$$

LOOP

$$\rho_i = \frac{-c_i + \sqrt{c_i^2 - 4(r_{site_i}^2 - r_i^2)}}{2} \quad i = 1, 2$$

$$\hat{r}_i = \rho_i \hat{L}_i + \hat{r}_{site_i} \quad i = 1, 2$$

$$\hat{W} = \frac{\hat{r}_1 \times \hat{r}_2}{|\hat{r}_1| |\hat{r}_2|}$$

$$\rho_3 = \frac{-\hat{r}_{site_3} \cdot \hat{W}}{\hat{L}_3 \cdot \hat{W}}$$

$$\hat{r}_3 = \rho_3 \hat{L}_3 + \hat{r}_{site_3}$$

FOR $j = 2, 3$ and $k = 1, 2$,

$$\cos(\Delta\nu_{jk}) = \frac{\hat{r}_j \cdot \hat{r}_k}{r_j r_k}$$

$$\sin(\Delta\nu_{jk}) = t_m \sqrt{1 - \cos^2(\Delta\nu_{jk})}$$

$$\begin{cases}
 \text{IF } \Delta\nu_{31} > 180^\circ \\
 c_1 = \frac{r_2 \sin(\Delta\nu_{32})}{r_1 \sin(\Delta\nu_{31})} \quad c_3 = \frac{r_2 \sin(\Delta\nu_{21})}{r_3 \sin(\Delta\nu_{31})} \\
 p = \frac{c_1 r_1 + c_3 r_3 - r_2}{c_1 + c_3 - 1} \\
 \text{ELSE} \\
 c_1 = \frac{r_1 \sin(\Delta\nu_{31})}{r_2 \sin(\Delta\nu_{32})} \quad c_3 = \frac{r_1 \sin(\Delta\nu_{21})}{r_3 \sin(\Delta\nu_{32})} \\
 p = \frac{c_3 r_3 - c_1 r_2 + r_1}{-c_1 + c_3 + 1} \\
 e \cos(\nu_i) = \frac{p}{r_i} - 1, i = 1, 2, 3 \\
 \text{IF } \Delta\nu_{21} \neq 180^\circ \\
 e \sin(\nu_2) = \frac{-\cos(\Delta\nu_{21})e \cos(\nu_2) + e \cos(\nu_1)}{\sin(\Delta\nu_{21})} \\
 \text{ELSE} \\
 e \sin(\nu_2) = \frac{\cos(\Delta\nu_{32})e \cos(\nu_2) - e \cos(\nu_3)}{\sin(\Delta\nu_{31})} \\
 e = \sqrt{(e \cos(\nu_2))^2 + (e \sin(\nu_2))^2} \\
 a = \frac{p}{1 - e^2} \quad n = \sqrt{\frac{\mu}{a^3}} \\
 S = \frac{r_2}{p} \sqrt{1 - e^2} e \sin(\nu_2) \quad C = \frac{r_2}{p} \{e^2 + e \cos(\nu_2)\} \\
 \sin(\Delta E_{32}) = \frac{r_3}{\sqrt{ap}} \sin(\Delta\nu_{32}) - \frac{r_3}{p} (1 - \cos(\Delta\nu_{32})) S \\
 \cos(\Delta E_{32}) = 1 - \frac{r_2 r_3}{ap} (1 - \cos(\Delta\nu_{32})) \\
 \sin(\Delta E_{21}) = \frac{r_1}{\sqrt{ap}} \sin(\Delta\nu_{21}) + \frac{r_1}{p} (1 - \cos(\Delta\nu_{21})) S \\
 \cos(\Delta E_{21}) = 1 - \frac{r_2 r_1}{ap} (1 - \cos(\Delta\nu_{21}))
 \end{cases}$$

$$\Delta M_{32} = \Delta E_{32} + 2S \sin^2\left(\frac{\Delta E_{32}}{2}\right) - C \sin(\Delta E_{32})$$

$$\Delta M_{12} = -\Delta E_{21} + 2S \sin^2\left(\frac{\Delta E_{21}}{2}\right) + C \sin(\Delta E_{21})$$

$$F_1 = \tau_1 - \frac{\Delta M_{12}}{n} \quad F_2 = \tau_3 - \frac{\Delta M_{32}}{n}$$

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 $\Delta r_1 \approx 0.005 r_1$

$$\frac{\partial F_1}{\partial r_1} = \frac{F_1(r_1 + \Delta r_1, r_2) - F_1(r_1, r_2)}{\Delta r_1}$$

$$\frac{\partial F_2}{\partial r_1} = \frac{F_2(r_1 + \Delta r_1, r_2) - F_2(r_1, r_2)}{\Delta r_1}$$

Repeat all calculations for $F_2(r_1, r_2 + \Delta r_2)$. Let $\Delta r_2 \approx 0.005 r_2$

$$\frac{\partial F_1}{\partial r_2} = \frac{F_1(r_1, r_2 + \Delta r_2) - F_1(r_1, r_2)}{\Delta r_2}$$

$$\frac{\partial F_2}{\partial r_2} = \frac{F_2(r_1, r_2 + \Delta r_2) - F_2(r_1, r_2)}{\Delta r_2}$$

$$\Delta = \frac{\partial F_1}{\partial r_1} \left(\frac{\partial F_2}{\partial r_2} \right) - \frac{\partial F_2}{\partial r_1} \left(\frac{\partial F_1}{\partial r_2} \right)$$

$$\Delta_1 = \frac{\partial F_2}{\partial r_2} F_1 - \frac{\partial F_1}{\partial r_2} F_2$$

$$\Delta_2 = \frac{\partial F_1}{\partial r_1} F_2 - \frac{\partial F_2}{\partial r_1} F_1$$

$$\Delta r_1 = -\frac{\Delta_1}{\Delta} \quad \Delta r_2 = -\frac{\Delta_2}{\Delta}$$

UNTIL converged (if not, set $r_1 = r_1 + \Delta r_1$, and $r_2 = r_2 + \Delta r_2$)

$$f = 1 - \frac{a}{r_2} (1 - \cos(\Delta E_{32}))$$

$$g = \tau_3 - \sqrt{\frac{a^3}{\mu}} (\Delta E_{32} - \sin(\Delta E_{32}))$$

$$\hat{\vec{v}}_2 = \frac{\hat{\vec{r}}_3 - f\hat{\vec{r}}_2}{g}$$

▼ **Example 7-2. Determining an Initial Orbit 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, 2012. Use observations number 3, 5, and 6. $\Delta AT = 35\text{s}$, $\Delta UT1 = -0.609\ 641\ \text{s}$, $x_p = 0.137\ 495''$, $y_p = 0.342\ 416''$. Assume LOD = 0.0 sec. Note the use of degrees for the observations. α_t is usually given in HMS and δ_t in DMS.

Obs. #	h min s	α_t °	δ_t °
1	11: 32: 28.00	333.028738	-2.022317
2	11: 36: 28.00	345.235515	7.648921
3	11: 40: 28.00	0.939913	18.667717
4	11: 44: 28.00	21.235600	29.086871
5	11: 48: 28.00	45.025748	35.664741
6	11: 52: 28.00	67.886655	36.996583
7	11: 56: 28.00	86.208078	34.719667
8	12: 00: 28.00	99.845522	30.928387
9	12: 04: 28.00	110.078585	26.767438
10	12: 08: 28.00	118.058822	22.680214
11	12: 12: 28.00	124.552101	18.801899
12	12: 16: 28.00	130.038810	15.154469
13	12: 20: 28.00	134.823380	11.722965
14	12: 24: 28.00	139.104912	8.483073
15	12: 26: 28.00	141.101028	6.927305

FIND: Position and velocity vectors for the observation using Gauss's method.

First, find the Julian dates, and the change in times between the observations:

$$\tau_1 = JD_1 - JD_2 = 2,456,159.986\ 435 - 2,456,159.991\ 991 \Rightarrow -8.0\ \text{min}$$

$$\tau_3 = JD_3 - JD_2 = 2,456,159.994\ 769 - 2,456,159.991\ 991 \Rightarrow 4.0\ \text{min}$$

Next, find the line-of-sight unit vector. Remember that each observation produces one unit vector.

$$\hat{L}_i = \begin{bmatrix} \cos(\delta_i) \cos(\alpha_i) \\ \cos(\delta_i) \sin(\alpha_i) \\ \sin(\delta_i) \end{bmatrix}, \quad \mathbf{L} = \begin{bmatrix} 0.9473 & 0.5742 & 0.3007 \\ 0.0155 & 0.5747 & 0.7399 \\ 0.3201 & 0.5830 & 0.6018 \end{bmatrix}$$

Find the matrix of site vectors using Algorithm 51. The ECEF vector is

$$\mathbf{r}_{site_{ECEF}} = \begin{bmatrix} -1673.9286 \\ -4599.0809 \\ 4079.2711 \end{bmatrix} \text{ km}$$

The observation times will yield different ECI site vectors.

$$[\mathbf{r}_{site_{ECI}}] = \begin{bmatrix} 4054.881 & 3956.224 & 3905.073 \\ 2748.195 & 2888.232 & 2956.935 \\ 4074.237 & 4074.364 & 4074.430 \end{bmatrix} \text{ km}$$

The change in times lets you estimate parameters that lead to finding the middle range magnitude.

$$a_1 = \frac{\tau_3}{\tau_3 - \tau_1} \quad a_{1u} = \frac{\tau_3((\tau_3 - \tau_1)^2 - \tau_1^2)}{6(\tau_3 - \tau_1)}, \quad a_1 = 0.333\ 333 \text{ and } a_{1u} = 25,599.9987 \text{ sec}^2$$

$$a_3 = -\frac{\tau_1}{\tau_3 - \tau_1} \quad a_{3u} = -\frac{\tau_1((\tau_3 - \tau_1)^2 - \tau_1^2)}{6(\tau_3 - \tau_1)}, \quad a_3 = 0.666\ 667 \text{ and } a_{3u} = 31,999.9992 \text{ sec}^2$$

You also need the inverse of the unit vector:

$$\mathbf{L}^{-1} = \frac{1}{|L|} \begin{bmatrix} L_{y2}L_{z3} - L_{y3}L_{z2} & -L_{y1}L_{z3} + L_{y3}L_{z1} & L_{y1}L_{z2} - L_{y2}L_{z1} \\ -L_{x2}L_{z3} + L_{x3}L_{z2} & L_{x1}L_{z3} - L_{x3}L_{z1} & -L_{x1}L_{z2} + L_{x2}L_{z1} \\ L_{x2}L_{y3} - L_{x3}L_{y2} & -L_{x1}L_{y3} + L_{x3}L_{y1} & L_{x1}L_{y2} - L_{x2}L_{y1} \end{bmatrix}$$

$$\mathbf{L}^{-1} = \begin{bmatrix} 28.593 & 56.908 & -84.259 \\ -76.035 & -158.366 & 232.712 \\ 58.460 & 123.169 & -178.992 \end{bmatrix}$$

Using matrix operations, form intermediate relations ($[\mathbf{M}] = [\mathbf{L}^{-1}][\mathbf{r}_{site}]$), and ultimately determine parameters for use in the eighth-degree equation:

$$d_1 = M_{21}a_1 - M_{22} + M_{23}a_3 = 231.7233 \text{ km}$$

$$d_2 = M_{21}a_{1u} + M_{23}a_{3u} = 1.1093 \times 10^{10} \text{ km sec}^2$$

The eighth-degree equation is

$$C = \hat{\mathbf{L}}_2 \cdot \hat{\mathbf{r}}_{site_2} = 6307.2743 \text{ km}$$

$$r_2^8 - (d_1^2 + 2Cd_1 + r_{site_2}^2)r_2^6 - 2\mu(Cd_2 + d_1d_2)r_2^3 - \mu^2d_2^2 = 0$$

$$r_2^8 - 4.3571 \times 10^7 r_2^6 - 5.7824 \times 10^{19} r_2^3 - 1.9550 \times 10^{31} = 0$$

The real roots for this equation are 10,451.1692 and -6,911.6045 km. Use the first root, r_2 , and calculate

$$u = \frac{\mu}{r_2^3} = 3.4917 \times 10^{-7} / \text{sec}^2$$

This value allows you to find an initial estimate of the slant-range values, which ends Gauss's part of the solution. First find the three c_i coefficients.

$$\begin{bmatrix} -c_1 \\ -c_2 \\ -c_3 \end{bmatrix} = \begin{bmatrix} -a_1 - a_{1u}u \\ 1 \\ -a_3 - a_{3u}u \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 0.342\ 884 \\ -1.0 \\ 0.678\ 177 \end{bmatrix}$$

Determine the initial guess of the slant ranges. Be sure to multiply the matrices out first.

$$\begin{bmatrix} c_1 \rho_1 \\ c_2 \rho_2 \\ c_3 \rho_3 \end{bmatrix} = [\mathbf{M}] \begin{bmatrix} -c_1 \\ -c_2 \\ -c_3 \end{bmatrix} \text{ solving: } \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \end{bmatrix} = \begin{bmatrix} 4169.739 \\ 4104.959 \\ 4546.683 \end{bmatrix} \text{ km}$$

Formally, Gauss's method ends at this point. We can form the initial estimate of the ranges to find the position vectors. ($\hat{\vec{r}}_i = \rho_i \hat{\vec{L}}_i + \vec{r}_{site_i} \quad i = 1 \dots 3$)

$$\begin{bmatrix} \hat{\vec{r}}_1^T \\ \hat{\vec{r}}_2^T \\ \hat{\vec{r}}_3^T \end{bmatrix} = \begin{bmatrix} 8004.7213 & 2812.9960 & 5408.8835 \\ 6313.3958 & 5247.5237 & 6467.7250 \\ 5272.0417 & 6321.1254 & 6810.4754 \end{bmatrix} \text{ km}$$

Because the vectors are “far” apart (17° and 8° , use **GIBBS** to find $\hat{\vec{v}}_2$. The result is

$$\hat{\vec{v}}_2 = -4.185\,518 \hat{I} \ 4.788\,532 \hat{J} \ 1.721\,732 \hat{K} \text{ km/s}$$

This example was designed from a known orbit with an epoch at 11:48:28.000. The epoch values are

$$\hat{\vec{r}}_2 = 6356.486\,034 \hat{I} + 5290.532\,2578 \hat{J} + 6511.396\,979 \hat{K} \text{ km}$$

$$\hat{\vec{v}}_2 = -4.172\,948 \hat{I} + 4.776\,550 \hat{J} + 1.720\,271 \hat{K} \text{ km/s}$$

You may be tempted to think that this result is not very good, however the orbital elements appear more reasonable. Because the observations were relatively close together, the Gaussian solution looks reasonable. Choosing points that are farther apart increases the uncertainty with the Gaussian solution. Comparing the known orbital elements with Gaussian, and Double- r iteration results,

$a = 12,246.023$ km vs.	$12,641.649$ km vs.	$12,245.8297$ km
$e = 0.200\,000$ vs.	$0.212\,951$ vs.	0.2000
$i = 40.0000^\circ$ vs.	39.9642° vs.	40.0000°
$\Omega = 330.0000^\circ$ vs.	330.0024° vs.	330.0000°

A few comments are necessary about this example, and angles-only techniques in general. The overall method may not seem to work for three perfect measurements as one might expect after using Gibbs, Lambert, and other techniques. Sometimes, the results look “reasonable”, but many times, the user is left with a diverging solution which does not provide an exact answer. This is perhaps why operational programs devote so much effort to refining the answer, as opposed to obtaining the first estimate. Some approaches are discussed in Long et al. (1989:9-9 to 9-44). For this example, the Double- r method is superior (converging in 6 iterations), but in other cases the roles are reversed. Note that hyperbolic orbits are often encountered during the iterations. Some interest has centered around the Gooding approach (1993, 1997). This technique may avoid some of the vagaries of the techniques in general, but there are still cases where multiple solutions arise.



7.3.4 Gooding's Method

Gooding (1997) proposed an approach for angles-only orbit determination, somewhat following the Double- r iteration, but using his Lambert solution (Gooding 1988, 1990). The solution uses iteration and can handle multiple revolutions. A challenge is setting all the parameters correctly, and ascertaining which solution (when multiple solutions arise) is correct. He provides FORTRAN code in the paper. I do not list the algorithm nor code as they are long, and contained in his papers. AGI's Orbit Determination Toolkit (ODTK) uses this approach extensively for angles-only IOD processing.

7.4 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.

We develop techniques for determining initial orbits based on the data available for processing. For instance, *Satellite Laser Ranging* (SLR) data is very popular as highly accurate observations for orbit determination. Unfortunately, there is no *direct* technique to process the resulting range-only information without already knowing the orbit. Thus, the range-only technique we'll introduce is a numerical method to process the SLR observations.

The overall topic of mixed observations is closely related to the *differential correction* of orbits (Chap. 10). 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.

7.4.1 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 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-degree 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. (7-5) and Eq. (7-6). Recognize that you don't need unit vectors. Solving for the slant range results in

$$\begin{aligned}\hat{\rho}_i &= \hat{r} - \hat{r}_{site_i}, \\ \rho_i^2 &= r^2 - 2\hat{r} \cdot \hat{r}_{site_i} + r_{site_i}^2, \quad i = 1 \dots 3\end{aligned}\tag{7-17}$$

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. (7-17)b with $i = 1$ from Eq. (7-17)b with $i = 2, 3$, and leaving the dot product on one side, you get

$$\begin{aligned}-\frac{1}{2}[\rho_3^2 - \rho_1^2 - (r_{site3}^2 - r_{site1}^2)] &= \dot{\vec{r}} \cdot (\dot{\vec{r}}_{site3} - \dot{\vec{r}}_{site1}) \\-\frac{1}{2}[\rho_2^2 - \rho_1^2 - (r_{site2}^2 - r_{site1}^2)] &= \dot{\vec{r}} \cdot (\dot{\vec{r}}_{site2} - \dot{\vec{r}}_{site1})\end{aligned}$$

Now, replace the left-hand side with temporary variables (d_3, d_2). Then, separate the dot-product terms:

$$\begin{aligned}d_3 &= (r_{site3I} - r_{site1I})r_I + (r_{site3J} - r_{site1J})r_J + (r_{site3K} - r_{site1K})r_K \\d_2 &= (r_{site2I} - r_{site1I})r_I + (r_{site2J} - r_{site1J})r_J + (r_{site2K} - r_{site1K})r_K\end{aligned}\quad (7-18)$$

Because the ultimate goal is to determine the position vector, isolate the J and K terms by multiplying Eq. (7-18) 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:

$$\begin{aligned}D &= (r_{site3K} - r_{site1K})(r_{site2J} - r_{site1J}) - (r_{site2K} - r_{site1K})(r_{site3J} - r_{site1J}) \\a_1 &= \frac{(r_{site2I} - r_{site1I})(r_{site3J} - r_{site1J}) - (r_{site3I} - r_{site1I})(r_{site2J} - r_{site1J})}{D} \\a_2 &= \frac{d_3(r_{site2J} - r_{site1J}) - d_2(r_{site3J} - r_{site1J})}{D} \\a_3 &= \frac{(r_{site2I} - r_{site1I})(r_{site3K} - r_{site1K}) - (r_{site3I} - r_{site1I})(r_{site2K} - r_{site1K})}{-D} \\a_4 &= \frac{d_3(r_{site2K} - r_{site1K}) - d_2(r_{site3K} - r_{site1K})}{-D}\end{aligned}$$

The result is

$$\begin{aligned}r_J &= a_3r_I + a_4 \\r_K &= a_1r_I + a_2\end{aligned}\quad (7-19)$$

We know that $r^2 = r_I^2 + r_J^2 + r_K^2$. If you use this expression with Eq. (7-19) and substitute it into Eq. (7-17), a quadratic equation in r_I results:

$$\begin{aligned}&\left\{a_1^2 + a_3^2 + 1\right\}r_I^2 + \left\{2(a_1a_2 + a_3a_4 - r_{site1I} - a_3r_{site1J} - a_1r_{site1K})\right\}r_I \\&+ \left\{a_2^2 + a_4^2 - 2a_4r_{site1J} - 2a_2r_{site1K} + r_{site1}^2 - \rho_1^2\right\} = 0\end{aligned}\quad (7-20)$$

The quadratic results in two roots. You can reject the root that yields an angle θ greater than 90° because it implies the satellite is not in the field-of-view, where,

$$\cos(\theta) = \frac{\dot{\vec{r}}_{site_1} \cdot \dot{\vec{\rho}}_1}{\vec{r}_{site_1} \cdot \vec{\rho}_1}$$

Once r_I is known, you can use Eq. (7-19) to find r_J and r_K , which determine the position vector. Consult Escobal ([1965] 1985:311–312) for the derivation of range-rate information.

$$E_i = \rho_i \dot{\rho}_i - \dot{\vec{r}}_{site_i} \cdot \dot{\vec{\rho}}_i, \quad i = 1, 2, 3$$

$$\mathbf{M} = \begin{bmatrix} \dot{\vec{\rho}}_1 \\ \dot{\vec{\rho}}_2 \\ \dot{\vec{\rho}}_3 \end{bmatrix} \quad \vec{v} = \mathbf{M}^{-1} \begin{bmatrix} E_1 \\ E_2 \\ E_3 \end{bmatrix}$$

7.4.2 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 (b). We now have the state as a function

of the observations, $\mathbf{X} = f(\mathbf{b})$. We can also determine the observations as a function of the state $\mathbf{b} = g(\mathbf{X})$.

Here, \mathbf{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. (2-51)]. In orbit determination, the Newton-Raphson iteration is sometimes called the method of *differential correction* (which we'll explore in Chap. 10), 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 \mathbf{X} such that $g(\mathbf{X}) = \mathbf{b}$. Let \mathbf{X}_1 be an initial guess and put $\mathbf{b}_1 = g(\mathbf{X}_1)$. Put $\mathbf{b}_i = \mathbf{b}_{i-1}$ and for each real number λ put

$$\mathbf{b}_\lambda = (1 - \lambda)\mathbf{b}_{i-1} + \lambda\mathbf{b}_i$$

We'll seek a continuous family (\mathbf{X}_λ) of solutions to $g(\mathbf{X}_\lambda) = \mathbf{b}_\lambda$. 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}_i) yields

$$\frac{\partial g}{\partial \mathbf{X}} \frac{d\mathbf{X}}{ds} = (\mathbf{b}_i - \mathbf{b}_{i-1}) \frac{d\lambda}{ds}$$

or in matrix form,

$$\left[\frac{\partial g}{\partial \mathbf{X}} (\mathbf{b}_{i-1} - \mathbf{b}_i) \right] \begin{bmatrix} \frac{d\mathbf{X}}{ds} \\ \frac{d\lambda}{ds} \end{bmatrix} = 0$$

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.

7.5 Three Position Vectors and Time

The previous methods provide data which we often have to process further to determine the initial orbit. In the *SITE-TRACK* discussion, I mentioned that some pieces of observational data might 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 that let us generate position and velocity information from position-only data. The simplest solutions use three position vectors [and times (UTC)] 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.

7.5.1 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 Gauss 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 needed 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 erroneous results.

Finally, we require the vectors to be coplanar, but we can relax this somewhat in practice. Because actual observations aren't perfect, we may obtain vectors which deviate by a degree or two from each other (in inclination). You should define the tolerance for the coplanar requirement before solving this problem. To describe the procedure, assume the vectors are coplanar. Figure 7-5 depicts the geometry for the Gibbs problem.

Notice the spacing between the vectors. As mentioned earlier, this procedure is basically 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. (7-14). Next, we need a method to isolate the coefficients. Start by crossing Eq. (7-14) with \vec{r}_1, \vec{r}_2 , and \vec{r}_3 respectively, recalling that $\vec{r}_i \times \vec{r}_i = 0$:

$$\begin{aligned} c_2 \vec{r}_1 \times \vec{r}_2 &= c_3 \vec{r}_3 \times \vec{r}_1 \\ c_1 \vec{r}_1 \times \vec{r}_2 &= c_3 \vec{r}_2 \times \vec{r}_3 \\ c_1 \vec{r}_3 \times \vec{r}_1 &= c_2 \vec{r}_2 \times \vec{r}_3 \end{aligned} \quad (7-21)$$

Dot Eq. (7-14) with the eccentricity vector:

$$\hat{\vec{e}} \cdot c_1 \vec{r}_1 + \hat{\vec{e}} \cdot c_2 \vec{r}_2 + \hat{\vec{e}} \cdot c_3 \vec{r}_3 = 0 \quad (7-22)$$

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:

$$\hat{\vec{e}} \cdot \vec{r}_i = p - r_i \quad (7-23)$$

Substituting this expression into Eq. (7-22) 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. (7-21) into the last equation to generate an equation with only the constant c_2 :

* This isn't surprising, because Gibbs is considered a pioneer of vector analysis.

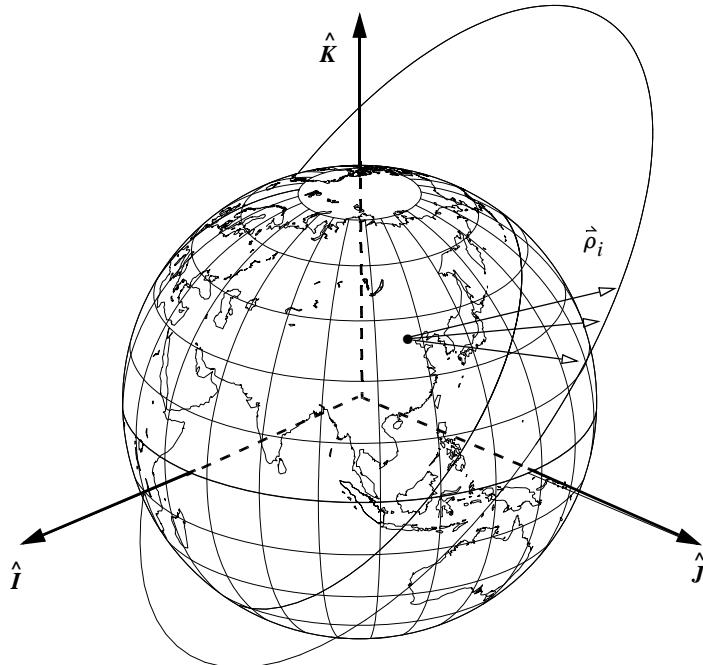


Figure 7-5. Orbit Geometry for the Gibbs Method. To determine the initial orbit using the Gibbs method, we must have at least three sequential observations.

$$c_2 \vec{r}_2 \times \vec{r}_3 (p - r_1) + c_2 \vec{r}_3 \times \vec{r}_1 (p - r_2) + c_2 \vec{r}_1 \times \vec{r}_2 (p - r_3) = \vec{0}$$

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} :

$$\begin{aligned} \vec{D} &= \vec{r}_1 \times \vec{r}_2 + \vec{r}_2 \times \vec{r}_3 + \vec{r}_3 \times \vec{r}_1 \\ \vec{N} &= r_1(\vec{r}_2 \times \vec{r}_3) + r_2(\vec{r}_3 \times \vec{r}_1) + r_3(\vec{r}_1 \times \vec{r}_2) = p\vec{D} \end{aligned} \quad (7-24)$$

\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 $\vec{N} \cdot \vec{D} = ND$. 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.*

Next, recall the perifocal coordinate system. The \hat{P} vector pointed towards periapsis, 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

$$\hat{P} = \frac{\hat{e}}{|\hat{e}|} \quad \hat{W} = \frac{\vec{N}}{|\vec{N}|}$$

The perifocal coordinate system is orthogonal, so you can define the third vector as

$$\hat{Q} = \hat{W} \times \hat{P} = \frac{\vec{N} \times \hat{e}}{|\vec{N}| |\hat{e}|}$$

Substitute the definition of \vec{N} from Eq. (7-24) and find

$$Ne\hat{Q} = \vec{N} \times \hat{e} = r_1(\hat{r}_2 \times \hat{r}_3) \times \hat{e} + r_2(\hat{r}_3 \times \hat{r}_1) \times \hat{e} + r_3(\hat{r}_1 \times \hat{r}_2) \times \hat{e}$$

Now use the general relationship for triple cross products of vectors [Eq. (C-5)], and rewrite as

$$\begin{aligned} Ne\hat{Q} &= r_1(\hat{r}_2 \cdot \hat{e})\hat{r}_3 - r_1(\hat{r}_3 \cdot \hat{e})\hat{r}_2 + r_2(\hat{r}_3 \cdot \hat{e})\hat{r}_1 - r_2(\hat{r}_1 \cdot \hat{e})\hat{r}_3 \\ &\quad + r_3(\hat{r}_1 \cdot \hat{e})\hat{r}_2 - r_3(\hat{r}_2 \cdot \hat{e})\hat{r}_1 \end{aligned}$$

Again using the relationship in Eq. (7-23) 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. (7-24)] 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 \vec{D} , \vec{N} , and \vec{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. 7-6.

From Eq. (1-23), recall the velocity vector at the middle time satisfies the expression

$$\dot{\vec{r}}_2 \times \vec{h} = \mu \left(\frac{\vec{r}_2}{r_2} + \hat{e} \right)$$

Cross the angular momentum vector into the equation above to obtain

* 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 precision of your observations.

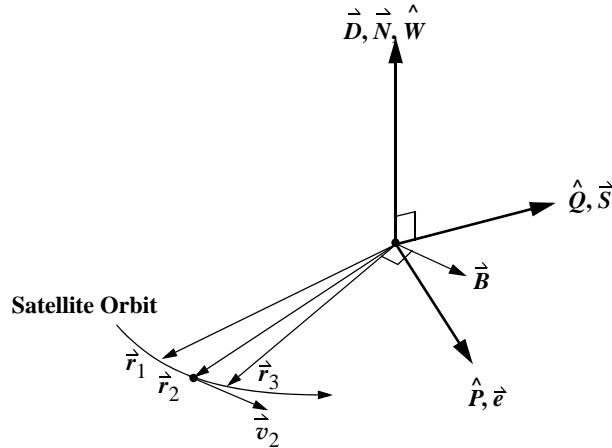


Figure 7-6. Geometry of Vectors for the Gibbs Method. It's useful to visualize the vectors which we use to develop the Gibbs solution. The vectors are not all equal. For instance, P and the eccentricity vector aren't equal, but they're in the same direction. The B vector is normal to \vec{r}_2 and \hat{D} .

$$\hat{h} \times (\dot{\hat{r}}_2 \times \hat{h}) = \mu \left(\frac{\hat{h} \times \hat{r}_2}{r_2} + \hat{h} \times \hat{e} \right)$$

Using the vector identity in Eq. (C-5), simplify the left side of the equation:

$$h^2 \dot{\hat{v}}_2 = \mu \left(\frac{\hat{h} \times \hat{r}_2}{r_2} + \hat{h} \times \hat{e} \right)$$

You can write $\hat{h} = h\hat{W}$ and $\hat{e} = e\hat{P}$, so

$$\dot{\hat{v}}_2 = \frac{\mu}{h} \left(\frac{\hat{W} \times \hat{r}_2}{r_2} + e\hat{W} \times \hat{P} \right) = \frac{\mu}{h} \left(\frac{\hat{W} \times \hat{r}_2}{r_2} + e\hat{Q} \right)$$

Using $N = pD$ and Eq. (1-19) 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

$$\dot{\hat{v}}_2 = \frac{1}{r_2} \sqrt{\frac{\mu}{ND}} \hat{D} \times \hat{r}_2 + \sqrt{\frac{\mu}{ND}} \hat{S}$$

Notice we've introduced notation for the middle velocity vector because this is our end goal. Define the following quantities to streamline the calculations:

$$\vec{B} \equiv \vec{D} \times \vec{r}_2$$

$$L_g \equiv \sqrt{\frac{\mu}{ND}}$$

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

$$\vec{v}_2 = \frac{L_g}{r_2} \vec{B} + L_g \vec{S}$$

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, find the two separation angles (α_{12} , α_{23} between vectors 1 and 2 and vectors 2 and 3):

$$\cos(\alpha_{12}) = \frac{\vec{r}_1 \cdot \vec{r}_2}{|\vec{r}_1||\vec{r}_2|} \quad \cos(\alpha_{23}) = \frac{\vec{r}_2 \cdot \vec{r}_3}{|\vec{r}_2||\vec{r}_3|} \quad (7-25)$$

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 \vec{Z}_{23} be the normal vector from the second and third position vectors.

$$\vec{Z}_{23} = \vec{r}_2 \times \vec{r}_3$$

The vectors are coplanar if \vec{Z}_{23} is perpendicular to \vec{r}_1 , so use a dot product:

$$\hat{\vec{Z}}_{23} \cdot \hat{\vec{r}}_1 = 0 \quad (7-26)$$

Your particular tolerance for “zero” determines the accuracy you wish to place on this coplanar requirement. The actual angle is

$$\alpha_{cop} = 90^\circ - \cos^{-1} \left(\frac{\hat{\vec{Z}}_{23} \cdot \hat{\vec{r}}_1}{|\hat{\vec{Z}}_{23}| |\hat{\vec{r}}_1|} \right) \quad (7-27)$$

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 54: GIBBS ($\hat{\vec{r}}_1 \hat{\vec{r}}_2 \hat{\vec{r}}_3$ in IJK $\Rightarrow \hat{\vec{v}}_2$)

$$\hat{\vec{Z}}_{12} = \hat{\vec{r}}_1 \times \hat{\vec{r}}_2 \quad \hat{\vec{Z}}_{23} = \hat{\vec{r}}_2 \times \hat{\vec{r}}_3 \quad \hat{\vec{Z}}_{31} = \hat{\vec{r}}_3 \times \hat{\vec{r}}_1$$

$$\alpha_{cop} = \sin^{-1} \left(\frac{\hat{\vec{Z}}_{23} \cdot \hat{\vec{r}}_1}{|\hat{\vec{Z}}_{23}| |\hat{\vec{r}}_1|} \right)$$

$$\cos(\alpha_{12}) = \frac{\hat{\vec{r}}_1 \cdot \hat{\vec{r}}_2}{|\hat{\vec{r}}_1| |\hat{\vec{r}}_2|} \quad \cos(\alpha_{23}) = \frac{\hat{\vec{r}}_2 \cdot \hat{\vec{r}}_3}{|\hat{\vec{r}}_2| |\hat{\vec{r}}_3|}$$

$$\hat{\vec{N}} = r_1 \hat{\vec{Z}}_{23} + r_2 \hat{\vec{Z}}_{31} + r_3 \hat{\vec{Z}}_{12}$$

$$\hat{\vec{D}} = \hat{\vec{Z}}_{12} + \hat{\vec{Z}}_{23} + \hat{\vec{Z}}_{31}$$

$$\hat{\vec{S}} = (r_2 - r_3) \hat{\vec{r}}_1 + (r_3 - r_1) \hat{\vec{r}}_2 + (r_1 - r_2) \hat{\vec{r}}_3$$

$$\hat{\vec{B}} = \hat{\vec{D}} \times \hat{\vec{r}}_2$$

$$L_g = \sqrt{\frac{\mu}{ND}}$$

$$\hat{\vec{v}}_2 = \frac{L_g \hat{\vec{B}}}{r_2} + L_g \hat{\vec{S}}$$

▼ **Example 7-3. Using Three Position Vectors for GIBBS Orbit Determination.**

GIVEN: $\hat{r}_1 = 6378.137\hat{K}$ km, $\hat{r}_2 = -4464.696\hat{J} - 5102.509\hat{K}$ km, and

$$\hat{r}_3 = 5740.323\hat{J} + 3189.068\hat{K}$$
 km

FIND: \hat{v}_{IJK}

Always test to be sure the input vectors are coplanar. In this case,

$$\hat{Z}_{23} = \hat{r}_2 \times \hat{r}_3 = 15,051,830.62\hat{I}$$
 km²

$$\alpha_{cop} = \text{SIN}^{-1}\left(\frac{\hat{Z}_{23} \cdot \hat{r}_1}{|\hat{Z}_{23}| |\hat{r}_1|}\right) = 0^\circ$$

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:

$$\hat{N} = \hat{r}_1(\hat{Z}_{23}) + \hat{r}_2(\hat{Z}_{31}) + \hat{r}_3(\hat{Z}_{12}) = 5.3123 \times 10^{11}\hat{I}$$
 km³

$$\hat{D} = \hat{Z}_{12} + \hat{Z}_{23} + \hat{Z}_{31} = 8.0141 \times 10^8\hat{I}$$
 km²

$$\hat{S} = (\hat{r}_2 - \hat{r}_3)\hat{r}_1 + (\hat{r}_3 - \hat{r}_1)\hat{r}_2 + (\hat{r}_1 - \hat{r}_2)\hat{r}_3 = -3.1490 \times 10^6\hat{J} - 8.8301 \times 10^5\hat{K}$$
 km²

$$\hat{B} = \hat{D} \times \hat{r}_2 = 4.0892 \times 10^{11}\hat{J} - 3.5780 \times 10^{11}\hat{K}$$
 km³

Use one more constant to get the final solution:

$$L_g = \sqrt{\frac{\mu}{ND}} = 9.6761 \times 10^8$$
 1/km s

so

$$\hat{v}_2 = \frac{L_g \hat{B}}{r_2} + L_g \hat{S} = 5.531\ 148\hat{J} - 5.191\ 806\hat{K}$$
 km/s



7.5.2 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 that 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** 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). **HERRICK-GIBBS** 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, 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 7-7 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.

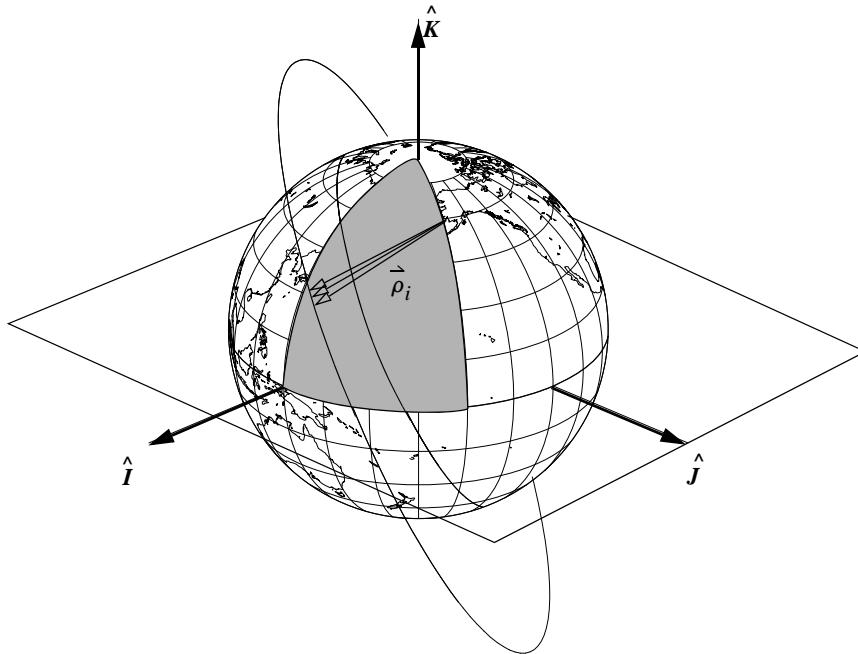


Figure 7-7. Orbit Geometry for the Herrick-Gibbs Method. Orbit determination using the Herrick-Gibbs method works well with closely spaced observations.

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. (7-27) to determine the coplanar property. The separation between the vectors is also important, but now we want a *small* separation. Eq. (7-25) 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

$$\dot{\vec{r}}(t) = \dot{\vec{r}}_2 + \ddot{\vec{r}}_2(t - t_2) + \frac{\ddot{\vec{r}}_2(t - t_2)^2}{2!} + \frac{\ddot{\vec{r}}_2(t - t_2)^3}{3!} + \frac{\ddot{\vec{r}}_2(t - t_2)^4}{4!} + \dots$$

Now, use this form for the two position vectors, $\dot{\vec{r}}_1$, $\dot{\vec{r}}_3$, and simplify the notation for the time differences as $\Delta t_{ij} = t_i - t_j$:

$$\begin{aligned}\dot{\vec{r}}_1 &= \dot{\vec{r}}_2 + \ddot{\vec{r}}_2 \Delta t_{12} + \frac{\ddot{\vec{r}}_2 \Delta t_{12}^2}{2!} + \frac{\ddot{\vec{r}}_2 \Delta t_{12}^3}{3!} + \frac{\overset{iv}{\vec{r}}_2 \Delta t_{12}^4}{4!} + \dots \\ \dot{\vec{r}}_3 &= \dot{\vec{r}}_2 + \ddot{\vec{r}}_2 \Delta t_{32} + \frac{\ddot{\vec{r}}_2 \Delta t_{32}^2}{2!} + \frac{\ddot{\vec{r}}_2 \Delta t_{32}^3}{3!} + \frac{\overset{iv}{\vec{r}}_2 \Delta t_{32}^4}{4!} + \dots\end{aligned}\tag{7-28}$$

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

$$\begin{aligned}-\dot{\vec{r}}_1 \Delta t_{32}^2 + \dot{\vec{r}}_3 \Delta t_{12}^2 &= \dot{\vec{r}}_2 (-\Delta t_{32}^2 + \Delta t_{12}^2) + \ddot{\vec{r}}_2 (-\Delta t_{32}^2 \Delta t_{12} + \Delta t_{12}^2 \Delta t_{32}) \\ &\quad + \frac{\ddot{\vec{r}}_2}{6} (-\Delta t_{32}^2 \Delta t_{12}^3 + \Delta t_{12}^2 \Delta t_{32}^3) \\ &\quad + \frac{\overset{iv}{\vec{r}}_2}{24} (-\Delta t_{32}^2 \Delta t_{12}^4 + \Delta t_{12}^2 \Delta t_{32}^4)\end{aligned}$$

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, $\dot{\vec{r}}_2$, switched to account for its being on the other side of the equation. Thus,

$$\begin{aligned}\dot{\vec{r}}_2 (\Delta t_{12} \Delta t_{32} \Delta t_{31}) &= \dot{\vec{r}}_1 \Delta t_{32}^2 + \dot{\vec{r}}_2 (-\Delta t_{32}^2 + \Delta t_{12}^2) - \dot{\vec{r}}_3 \Delta t_{12}^2 \\ &\quad + \frac{\ddot{\vec{r}}_2}{6} (\Delta t_{12}^2 \Delta t_{32}^2 \Delta t_{31}) + \frac{\overset{iv}{\vec{r}}_2}{24} (\Delta t_{12}^2 \Delta t_{32}^2 \Delta t_{31} \{ \Delta t_{32} + \Delta t_{12} \})\end{aligned}\tag{7-29}$$

The problem now is to determine expressions for the third and fourth derivatives because everything else in Eq. (7-29) is known. Using Eq. (7-28), differentiate each equation twice, which provides the starting point to remove the higher derivatives:

$$\dot{\vec{r}}_1 = \ddot{\vec{r}}_2 + \ddot{\vec{r}}_2 \Delta t_{12} + \frac{\overset{iv}{\vec{r}}_2}{2} \Delta t_{12}^2 + \dots$$

$$\ddot{\vec{r}}_3 = \ddot{\vec{r}}_2 + \ddot{\vec{r}}_2 \Delta t_{32} + \frac{i_v}{2} \dot{\vec{r}}_2 \Delta t_{32}^2 + \dots$$

Eliminate the third derivative term by forming $-\Delta t_{32} \ddot{\vec{r}}_1 + \Delta t_{12} \ddot{\vec{r}}_3$

$$-\ddot{\vec{r}}_1 \Delta t_{32} + \ddot{\vec{r}}_3 \Delta t_{12} = \ddot{\vec{r}}_2 (-\Delta t_{32} + \Delta t_{12}) + \frac{i_v}{2} (\Delta t_{32} \Delta t_{12}^2 + \Delta t_{12} \Delta t_{32}^2)$$

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

$$-\ddot{\vec{r}}_1 \Delta t_{32} + \ddot{\vec{r}}_3 \Delta t_{12} = \ddot{\vec{r}}_2 (-\Delta t_{31}) + \frac{i_v}{2} (\Delta t_{12} \Delta t_{32} \Delta t_{31})$$

which you can solve to isolate the fourth derivative:

$$\frac{i_v}{\Delta t_{12} \Delta t_{32} \Delta t_{31}} (-\ddot{\vec{r}}_1 \Delta t_{32} + \ddot{\vec{r}}_2 \Delta t_{31} + \ddot{\vec{r}}_3 \Delta t_{12}) \quad (7-30)$$

Similarly, isolate the third derivative by forming

$$-\Delta t_{32}^2 \ddot{\vec{r}}_1 + \Delta t_{12}^2 \ddot{\vec{r}}_3$$

and

$$-\ddot{\vec{r}}_1 \Delta t_{32}^2 + \ddot{\vec{r}}_3 \Delta t_{12}^2 = \ddot{\vec{r}}_2 (-\Delta t_{32}^2 + \Delta t_{12}^2) + \ddot{\vec{r}}_2 (-\Delta t_{32}^2 \Delta t_{12} + \Delta t_{12}^2 \Delta t_{32}) \quad (7-31)$$

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. (7-31) 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

$$\ddot{\vec{r}}_2 = \frac{1}{\Delta t_{12}\Delta t_{32}\Delta t_{13}}(-\ddot{\vec{r}}_1\Delta t_{32}^2 + \ddot{\vec{r}}_2(\Delta t_{31}[\Delta t_{32} + \Delta t_{12}]) + \ddot{\vec{r}}_3\Delta t_{12}^2) \quad (7-32)$$

Now, substitute these two expressions for the derivatives into Eq. (7-29), divide through by $(\Delta t_{12}\Delta t_{32}\Delta t_{31})$, and simplify to get

$$\begin{aligned} \dot{\vec{r}}_2 &= \frac{1}{\Delta t_{12}\Delta t_{32}\Delta t_{31}}[\dot{\vec{r}}_1\Delta t_{32}^2 + \dot{\vec{r}}_2(-\Delta t_{32}^2 + \Delta t_{12}^2) - \dot{\vec{r}}_3\Delta t_{12}^2] \\ &\quad - \frac{1}{6\Delta t_{31}}[-\ddot{\vec{r}}_1\Delta t_{32}^2 + \ddot{\vec{r}}_2\Delta t_{31}(\Delta t_{32} + \Delta t_{12}) + \ddot{\vec{r}}_3\Delta t_{12}^2] \\ &\quad + \frac{(\Delta t_{32} + \Delta t_{12})}{12\Delta t_{31}}[-\ddot{\vec{r}}_1\Delta t_{32} + \ddot{\vec{r}}_2\Delta t_{31} + \ddot{\vec{r}}_3\Delta t_{12}] \\ &= \frac{\dot{\vec{r}}_1\Delta t_{32}}{\Delta t_{12}\Delta t_{31}} - \frac{\dot{\vec{r}}_2(\Delta t_{32} + \Delta t_{12})}{\Delta t_{12}\Delta t_{32}} - \frac{\dot{\vec{r}}_3\Delta t_{12}}{\Delta t_{32}\Delta t_{31}} + \ddot{\vec{r}}_1\left(\frac{\Delta t_{32}^2}{6\Delta t_{31}} - \frac{\Delta t_{32}(\Delta t_{32} + \Delta t_{12})}{12\Delta t_{31}}\right) \\ &\quad + \ddot{\vec{r}}_2\left(\frac{-\Delta t_{31}(\Delta t_{32} + \Delta t_{12})}{6\Delta t_{31}} + \frac{\Delta t_{31}(\Delta t_{32} + \Delta t_{12})}{12\Delta t_{31}}\right) \\ &\quad + \ddot{\vec{r}}_3\left(\frac{-\Delta t_{12}^2}{6\Delta t_{31}} + \frac{\Delta t_{12}(\Delta t_{32} + \Delta t_{12})}{12\Delta t_{31}}\right) \\ &= \frac{\dot{\vec{r}}_1\Delta t_{32}}{\Delta t_{12}\Delta t_{31}} - \frac{\dot{\vec{r}}_2(\Delta t_{32} + \Delta t_{12})}{\Delta t_{12}\Delta t_{32}} - \frac{\dot{\vec{r}}_3\Delta t_{12}}{\Delta t_{32}\Delta t_{31}} + \ddot{\vec{r}}_1\left(\frac{\Delta t_{32}(2\Delta t_{32} - \Delta t_{32} - \Delta t_{12})}{12\Delta t_{31}}\right) \\ &\quad + \ddot{\vec{r}}_2\left(\frac{-\Delta t_{31}(\Delta t_{32} + \Delta t_{12})}{12\Delta t_{31}}\right) + \ddot{\vec{r}}_3\left(\Delta t_{12}\frac{-2\Delta t_{12} + \Delta t_{32} + \Delta t_{12}}{12\Delta t_{31}}\right) \end{aligned}$$

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. (7-28).*

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 55, below, I've shown the Julian date 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 (*MJD*) to resolve this difficulty.

ALGORITHM 55: HERRICK-GIBBS ($\hat{\vec{r}}_1, \hat{\vec{r}}_2, \hat{\vec{r}}_3, JD_1, JD_2, JD_3 \Rightarrow \hat{\vec{v}}_2$)

$$\begin{aligned} \Delta t_{31} &= JD_3 - JD_1 \\ \Delta t_{32} &= JD_3 - JD_2 \\ \Delta t_{21} &= JD_2 - JD_1 \\ \hat{\vec{Z}}_{23} &= \hat{\vec{r}}_2 \times \hat{\vec{r}}_3 \quad \alpha_{cop} = 90^\circ - \cos^{-1} \left(\frac{\hat{\vec{Z}}_{23} \cdot \hat{\vec{r}}_1}{|\hat{\vec{Z}}_{23}| |\hat{\vec{r}}_1|} \right) \\ \cos(\alpha_{12}) &= \frac{\hat{\vec{r}}_1 \cdot \hat{\vec{r}}_2}{|\hat{\vec{r}}_1| |\hat{\vec{r}}_2|} \quad \cos(\alpha_{23}) = \frac{\hat{\vec{r}}_2 \cdot \hat{\vec{r}}_3}{|\hat{\vec{r}}_2| |\hat{\vec{r}}_3|} \\ \hat{\vec{v}}_2 &= -\Delta t_{32} \left(\frac{1}{\Delta t_{21} \Delta t_{31}} + \frac{\mu}{12 r_1^3} \right) \hat{\vec{r}}_1 + (\Delta t_{32} - \Delta t_{21}) \left(\frac{1}{\Delta t_{21} \Delta t_{32}} + \frac{\mu}{12 r_2^3} \right) \hat{\vec{r}}_2 \\ &\quad + \Delta t_{21} \left(\frac{1}{\Delta t_{32} \Delta t_{31}} + \frac{\mu}{12 r_3^3} \right) \hat{\vec{r}}_3 \end{aligned}$$

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 approxi-

* 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 the differential correction processing, which I'll talk about in Chap. 10.

mate cross-over point between 1° and 5° . Below 1° , **HERRICK-GIBBS** is superior; above 5° , **GIBBS** is superior.

▼ **Example 7-4. Using Three Position Vectors and Time for Herrick-Gibbs Orbit Determination.**

$$\begin{aligned} \text{GIVEN: } \hat{\vec{r}}_1 &= 3419.855\ 64\hat{I} + 6019.826\ 02\hat{J} + 2784.600\ 22\hat{K} \text{ km} & t_1 &= 0.0 \\ \hat{\vec{r}}_2 &= 2935.911\ 95\hat{I} + 6326.183\ 24\hat{J} + 2660.595\ 84\hat{K} \text{ km} & t_2 &= 1 \text{ min } 16.48 \text{ s} \\ \hat{\vec{r}}_3 &= 2434.952\ 02\hat{I} + 6597.386\ 74\hat{J} + 2521.523\ 11\hat{K} \text{ km} & t_3 &= 2 \text{ min } 33.04 \text{ s} \end{aligned}$$

$$\text{FIND: } \hat{\vec{v}}_2$$

Find the delta times:

$$\Delta t_{21} = 76.48 \text{ s}$$

$$\Delta t_{31} = 153.04 \text{ s}$$

$$\Delta t_{32} = 76.56 \text{ s}$$

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.

$$\hat{\vec{Z}}_{23} = \hat{\vec{r}}_2 \times \hat{\vec{r}}_3 = -1.6014 \times 10^6 \hat{I} - 9.2455 \times 10^5 \hat{J} + 3.9654 \times 10^6 \hat{K} \text{ km}^2$$

$$\alpha_{cop} = 90^\circ - \cos^{-1} \left(\frac{\hat{\vec{Z}}_{23} \cdot \hat{\vec{r}}_1}{\|\hat{\vec{Z}}_{23}\| \|\hat{\vec{r}}_1\|} \right) = 0.0^\circ$$

$$\cos(\alpha_{12}) = \frac{\hat{\vec{r}}_1 \cdot \hat{\vec{r}}_2}{\|\hat{\vec{r}}_1\| \|\hat{\vec{r}}_2\|} \Rightarrow 4.5^\circ \quad \cos(\alpha_{23}) = \frac{\hat{\vec{r}}_2 \cdot \hat{\vec{r}}_3}{\|\hat{\vec{r}}_2\| \|\hat{\vec{r}}_3\|} \Rightarrow 4.5^\circ$$

Now, find the middle velocity vector directly:

$$\hat{\vec{v}}_2 = -6.441\ 645\hat{I} + 3.777\ 634\ 3\hat{J} - 1.720\ 587\hat{K} \text{ km/s}$$



7.6 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. 2.3, 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. 6. 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. 7.7, 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 origi-

nal geometric formulation, which provides equations to determine the minimum-energy orbit, and the original Gauss formulation, which gives geometrical insight into the problem. Lambert's original method ultimately reduces to two separate cases to provide a complete answer: whether or not the orbital segment includes the attracting focus, and if it includes the vacant focus (Danby, 1992:182). The universal-variable solution and algorithms by Battin (1987) are elegant and robust algorithms for general purposes. Gooding's algorithm (1988 and 1990), and Thorne's solution (Thorne, 2004) 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.

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. 7-8 illustrates.

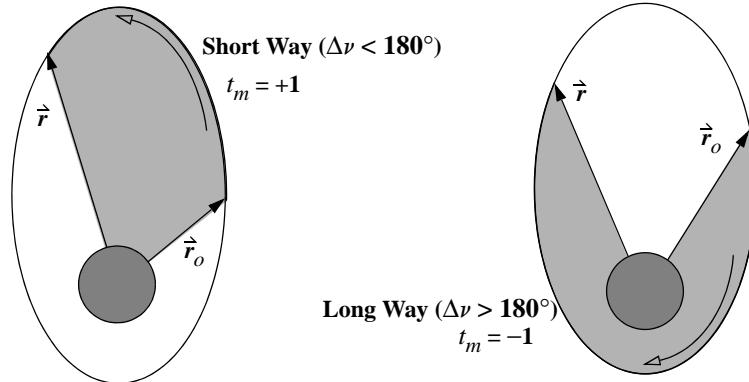


Figure 7-8. Transfer Methods, t_m , for the Lambert Problem. Traveling between the two specified points can take the long way or the short way. For the long way, the change in true anomaly exceeds 180° .

We can see right away from Fig. 7-8 that two paths will satisfy the requirement to find an orbit between the two given position vectors in a given time. 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 must assume a direction:

$$\cos(\Delta\nu) = \frac{\hat{r}_o \cdot \hat{r}}{r_o r} \quad (7-33)$$

$$\sin(\Delta\nu) = t_m \sqrt{1 - \cos^2(\Delta\nu)}$$

In these relations, recognize that t_m is (+1) for short-way transfers and (-1) for long-way transfers.

It's also useful to understand the general solutions that are possible for the Lambert problem. Figure 7-9 shows the possible results. Notice that the units are normalized with canonical units. The solution space includes all orbit types, and the elliptical transfers may also include multi-revolution cases.

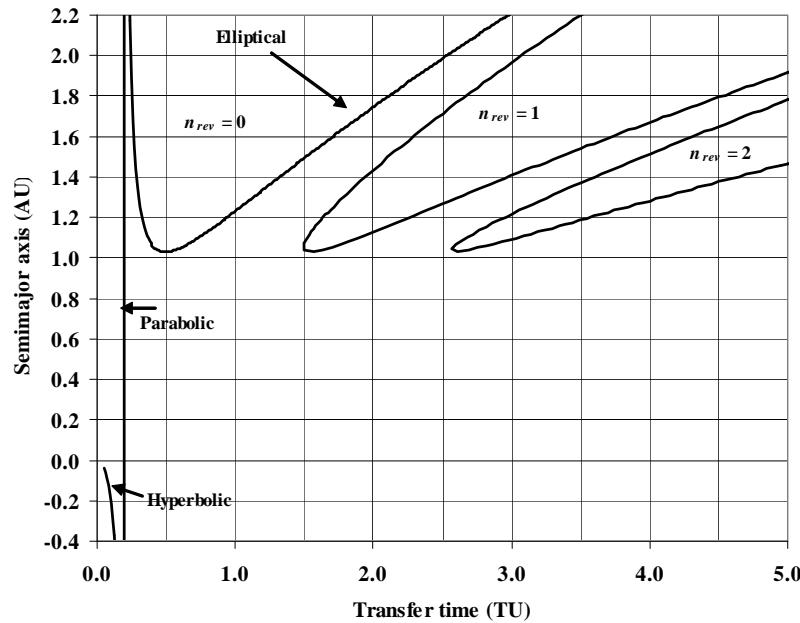


Figure 7-9. General solution options for the Lambert problem. This figure shows the various orbits that are possible for transfers considering the Lambert problem. The units are normalized, and the initial separation between the spacecraft is 75° . The interceptor and target positions are 1.0 and 1.524 AU respectively. Long and short way trajectories are possible with the elliptical transfers. Notice the corresponding increase in minimum transfer time as the number of revolutions increases. (Source Thorne, 2007 using his series solution discussed later)

7.6.1 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. Battin (1987:276) states Lambert's basic premise 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. 7-10. 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 !

$$s = \frac{r_o + r + c}{2} \quad (7-34)$$

From Eq. (1-1), 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. 7-10 with F , F' , and r ,

$$2a = r + (2a - r)$$

Knowing the orbit's semimajor axis, Figure 7-10 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. 7-10, I've shown the foci for one semimajor axis, along with a corresponding orbit. In Fig. 7-11 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. 7-11.

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 $r - r_o$.

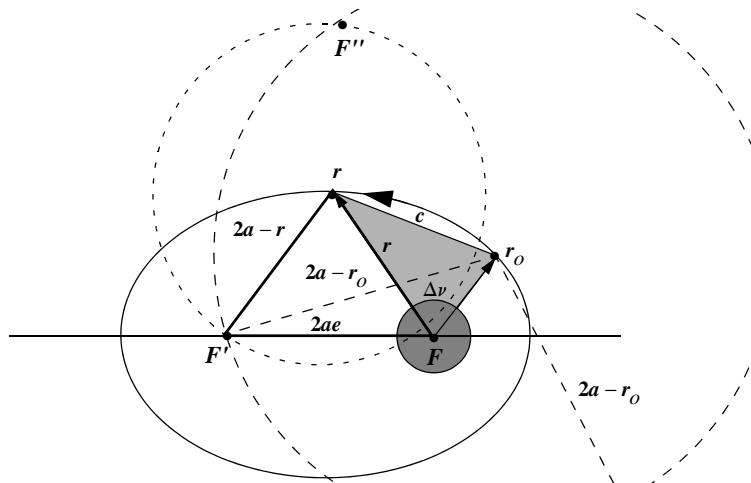


Figure 7-10. Geometry for the Lambert Problem (I). This figure shows how we locate the secondary focus—the intersection of the dashed circles. The chord length, c , is the shortest distance between the two position vectors. The sum of the distances from the foci to any point, r or r_o , is equal to twice the semimajor axis.

4. The secondary focus whose distance to the primary is shorter has an orbit with a smaller eccentricity. In Fig. 7-10, 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. 7-11 where the darker circles just touch.

First, 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. (7-33)] is greater than 180° . The quadrant of α_e requires the time of flight for a minimum-energy transfer, Eq. (7-39). Let $\alpha_e = 2\pi - \alpha_e$ if the desired time of flight is greater than Δt_{min} .

$$\begin{aligned} \text{SIN}\left(\frac{\alpha_e}{2}\right) &= \sqrt{\frac{r_o + r + c}{4a}} = \sqrt{\frac{s}{2a}} & \text{SINH}\left(\frac{\alpha_h}{2}\right) &= \sqrt{\frac{s}{-2a}} = \sqrt{\frac{r_o + r + c}{-4a}} \\ \text{SIN}\left(\frac{\beta_e}{2}\right) &= \sqrt{\frac{r_o + r - c}{4a}} = \sqrt{\frac{s - c}{2a}} & \text{SINH}\left(\frac{\beta_h}{2}\right) &= \sqrt{\frac{s - c}{-2a}} = \sqrt{\frac{r_o + r - c}{-4a}} \end{aligned} \quad (7-35)$$

Kaplan (1976:319–324) shows that the general time of flight, sometimes called **Lagrange time equation**, is

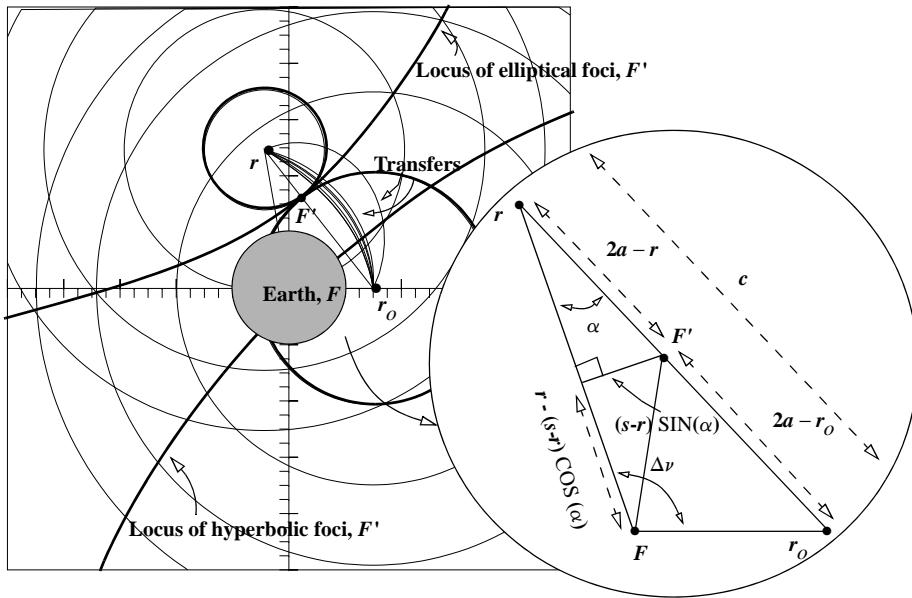


Figure 7-11. Geometry for the Lambert Problem (II). Solving Lambert's problem relies on many geometrical quantities. Be sure to allow for the Earth when viewing representations like this. I've shown transfers between a satellite at 9567.2 km and 15,307.5 km from the Earth's center. We can use the inset 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' .

$$\begin{aligned}\Delta t &= \sqrt{\frac{a^3}{\mu}} [2n_{rev}\pi + \alpha_e - \sin(\alpha_e) \mp (\beta_e - \sin(\beta_e))] \\ &= \sqrt{\frac{-a^3}{\mu}} [\sinh(\alpha_h) - \alpha_h \mp (\sinh(\beta_h) - \beta_h)]\end{aligned}\quad (7-36)$$

The minus/plus sign indicates solutions for short and long way trajectories, respectively. Notice that I've included the number of revolutions ($n_{rev} = 0, 1, 2, \dots$) completed for multiple revolution solutions. Ochoa and Prussing (1992) provide additional information on multiple revolution solutions.

Now, let's explore the minimum energy cases. They represent 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:

$$a_{min} = \frac{s}{2} = \frac{r_o + r + c}{4} \quad (7-37)$$

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. 7-11 permit

$$\begin{aligned} 4a_{min}^2 e_{min}^2 &= \{(s - r) \sin(\alpha)\}^2 + \{r - (s - r) \cos(\alpha)\}^2 \\ &= (s - r)^2 \{1 - \cos^2(\alpha)\} + r^2 + (s - r)^2 \cos^2(\alpha) - 2r(s - r) \cos(\alpha) \quad (7-38) \\ &= s^2 - 2r(s - r)(1 + \cos(\alpha)) \end{aligned}$$

It’s useful to determine an expression for the central angle with the cosine law:

$$r_o^2 = r^2 + c^2 - 2rc \cos(\alpha)$$

$$\cos(\alpha) = \frac{-r_o^2 + r^2 + c^2}{2rc}$$

Now consider the expression,

$$2s(s - r_o) - rc$$

and expand it using Eq. (7-34). 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. (7-38)] 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. (1-10), develop $4a^2 e^2 = 4a(a - p)$. Remembering Eq. (7-37), 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. (2-79), it's easy to determine the eccentricity of this minimum-energy transfer. Kaplan (1976:313) shows that the minimum value for the eccentricity is

$$e_{min} = \sqrt{1 - \frac{2p_{min}}{s}}$$

We observe several things before finding the minimum time. If we solve Eq. (7-37) and Eq. (7-36) using the minimum semimajor axis, we can find corresponding time of flight. This will not be the absolute minimum time of flight, nor will it be the minimum for each revolution case. The absolute minimum will come from a parabolic solution, given by Kaplan (1976:324). Notice that since hyperbolic transfers don't [realistically] repeat, they are not considered here.

$$t_{min(absolute)} = \frac{1}{3}\sqrt{\frac{2}{\mu}}\{s^{3/2} - (s - c)^{3/2}\} \quad (7-39)$$

To find the minimum time for each multiple revolution case, you must iterate on the semimajor axis. Ochoa and Prussing (1992) provide additional details on the formulation, including how to setup the Newton iteration to find the semimajor axis that will yield the minimum transfer time. The approach uses the minimum semimajor axis from Eq. (7-37).

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. (7-36) 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. (2-64), Eq. (2-65), and Eq. (2-67) to find the f and g functions, and then the position and velocity vectors. Solving Eq. (2-58) for the initial velocity vector and using Eq. (2-64) produces

$$\hat{\vec{v}}_o = \frac{1}{g}\{\hat{\vec{r}} - f\hat{\vec{r}}_o\} = \frac{\sqrt{\mu p}}{r_o r \sin(\Delta\nu)} \left\{ \hat{\vec{r}} - \left[1 - \frac{r}{p} \{1 - \cos(\Delta\nu)\} \right] \hat{\vec{r}}_o \right\}$$

Algorithm 56 summarizes the results of the minimum-energy transfers. Note that there are no multiple revolution cases and that because α_e is equal to π , some simplification is possible.

ALGORITHM 56: Lambert’s Problem—Minimum Energy

$$\begin{aligned}
 & (\dot{\vec{r}}_o, \vec{r} \Rightarrow a_{min}, e_{min}, t_{min}, \dot{\vec{v}}_o) \\
 & \cos(\Delta\nu) = \frac{\dot{\vec{r}}_o \cdot \vec{r}}{r_o r} \\
 & c = \sqrt{r_o^2 + r^2 - 2r_o r \cos(\Delta\nu)} \\
 & s = \frac{r_o + r + c}{2} \\
 & a_{min} = \frac{s}{2} \quad p_{min} = \frac{r_o r}{c} (1 - \cos(\Delta\nu)) \\
 & e_{min} = \sqrt{1 - \frac{2p_{min}}{s}} \\
 & \alpha_e = \pi \quad \sin\left(\frac{\beta_e}{2}\right) = \sqrt{\frac{s-c}{s}} \\
 & t_{min(amin)} = \sqrt{\frac{a_{min}^3}{\mu}} [\alpha_e \mp (\beta_e - \sin(\beta_e))] \\
 & t_{min(absolute)} = \frac{1}{3} \sqrt{\frac{2}{\mu}} \{s^{3/2} - (s-c)^{3/2}\} \\
 & \dot{\vec{v}}_o = \frac{\sqrt{\mu p_{min}}}{r_o r \sin(\Delta\nu)} \left\{ \dot{\vec{r}} - \left[1 - \frac{r}{p_{min}} \{1 - \cos(\Delta\nu)\} \right] \dot{\vec{r}}_o \right\}
 \end{aligned}$$

7.6.2 Lambert—Gauss’s 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 formula 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

Gauss's technique I've presented for dealing with angles-only data, even though the two are related. Indeed, Herrick (1971:247) remarks that

The process [Gauss's 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 7-12 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:

$$y = \frac{\sqrt{\mu p} \Delta t}{r_o r \sin(\Delta\nu)} \quad (7-40)$$

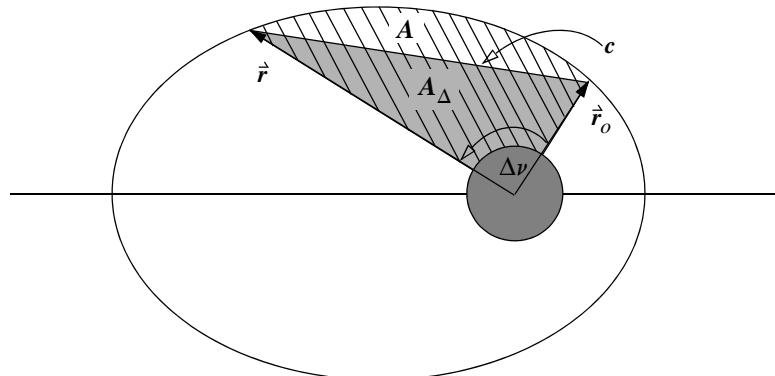


Figure 7-12. Area Swept Out by a Satellite During Transfer. Gauss's method relies on the area swept out by the satellite during its orbit. Two regions are useful—the triangular shaded region, A_{Δ} , and the total area, A , swept out by the satellite.

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. (2-109) 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. (7-40)] and square the result:

$$y^2 = \frac{\mu \Delta t^2 \sec^2\left(\frac{\Delta\nu}{2}\right)}{2r_o r \left(r_o + r - 2\sqrt{r_o r} \cos\left(\frac{\Delta\nu}{2}\right) \cos\left(\frac{\Delta E}{2}\right) \right)}$$

This is usually known as Gauss's first equation. Using several temporary variables simplifies the notation, but their formation isn't obvious:

$$\ell \equiv \frac{r_o + r}{4\sqrt{r_o r} \cos\left(\frac{\Delta\nu}{2}\right)} - \frac{1}{2} \quad m \equiv \frac{\mu \Delta t^2}{\left\{ 2\sqrt{r_o r} \cos\left(\frac{\Delta\nu}{2}\right) \right\}^3} \quad x_1 \equiv \sin^2\left(\frac{\Delta E}{4}\right)$$

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):

$$y^2 = \frac{m}{\ell + x_1} \tag{7-41}$$

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*:

$$y^3 - y^2 = mx_2 = y^2(y-1) \quad (7-42)$$

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. (7-41) for y^2 into Eq. (7-42) and solve for the sector ratio,

$$y = (\ell + x_1)x_2 + 1$$

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,

$$\begin{aligned} y^3 - y^2 - hy - h &= 0 \\ h &= \frac{m}{\frac{5}{6} + 1 + \xi} \end{aligned}$$

We use Eq. (2-108) and Eq. (2-109) and appropriate end conditions to yield an expression for the semimajor axis if we need it.

$$a = \frac{r_o + r - 2\sqrt{r_o r} \cos\left(\frac{\Delta\nu}{2}\right) \cos\left(\frac{\Delta E}{2}\right)}{2 \sin^2\left(\frac{\Delta E}{2}\right)} \quad p = \frac{y^2 r_o^2 r^2 \sin^2(\Delta\nu)}{\mu \Delta t^2}$$

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 57 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 ℓ 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 57: Lambert—Gauss's Solution ($\hat{\vec{r}}_o, \hat{\vec{r}}, \Delta t, t_m \Rightarrow \hat{\vec{v}}_o, \hat{\vec{v}}$)

$$\begin{aligned} \cos(\Delta\nu) &= \frac{\hat{\vec{r}}_o \cdot \hat{\vec{r}}}{r_o r} \\ \sin(\Delta\nu) &= t_m \sqrt{1 - \cos^2(\Delta\nu)} \end{aligned}$$

$$\ell \equiv \frac{r_o + r}{4\sqrt{r_o r} \cos\left(\frac{\Delta\nu}{2}\right)} - \frac{1}{2} \quad m \equiv \frac{\mu \Delta t^2}{\left\{2\sqrt{r_o r} \cos\left(\frac{\Delta\nu}{2}\right)\right\}^3}$$

Guess $y = 1$

LOOP

$$x_1 = \frac{m}{y^2} - \ell$$

$$x_2 = \frac{4}{3} \left(1 + \frac{6x_1}{5} + \frac{6(8)x_1^2}{5(7)} + \frac{6(8)(10)x_1^3}{5(7)(9)} + \dots \right)$$

$$y = 1 + x_2(\ell + x_1)$$

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,

$$\cosh\left(\frac{\Delta H}{2}\right) = 1 - 2x_1$$

$$p = \frac{r_o r (1 - \cos(\Delta\nu))}{r_o + r - 2\sqrt{r_o r} \cos\left(\frac{\Delta\nu}{2}\right) \cosh\left(\frac{\Delta H}{2}\right)}$$

$$f = 1 - \frac{r}{p}(1 - \cos(\Delta\nu)) \quad g = \frac{rr_o \sin(\Delta\nu)}{\sqrt{\mu p}}$$

$$\dot{f} = \sqrt{\frac{1}{p}} \tan\left(\frac{\Delta\nu}{2}\right) \left(\frac{1 - \cos(\Delta\nu)}{p} - \frac{1}{r} - \frac{1}{r_o} \right)$$

$$\dot{g} = 1 - \frac{r_o}{p}(1 - \cos(\Delta\nu))$$

$$\dot{\hat{v}}_o = \frac{\dot{\hat{r}} - \hat{r}_o}{g} \quad \dot{\hat{v}} = \frac{\dot{g} \hat{r} - \dot{\hat{r}}_o}{g}$$

7.6.3 Lambert—Thorne's Solution

Thorne has developed a complete power series solution to the Lambert problem (Thorne 2004). He has graciously provided this section for the present text. It's equally valid for any value of the transfer time, whether the orbital path is hyperbolic or elliptical. As shown earlier, the transfer time is equal to a complicated function of the unknown semimajor axis of the observed orbit, and is case dependent. We can solve these equations using series reversion and inversion, if necessary, to improve convergence properties. The complete mathematical technique is too lengthy for this book, but the resulting three power series are the only known direct solutions of the Lambert problem as an explicit function of the transfer time. The three series solutions are needed to cover all of the various cases of hyperbolic and elliptical transfer orbits. Thorne's solution has the advantage of not requiring any initial guess, unlike the traditional root-finding methods, and it works even with colinear position vectors. Finally, because Thorne's solution contains the time of flight explicitly, we can take time derivatives of the series equations to minimize the delta- v required for any number of sequential orbital maneuvers.

If the transfer angle is less or more than 180 degrees, there is a corresponding sign change (\mp) in the Lagrange time equations, as shown in Eq. (7-35) and Eq. (7-36).

In equation Eq. (7-35), c is the chord which is a line segment connecting the two observed position vectors (Fig. 7-11), and s is from Eq. (7-34). These quantities are known from the given observations and are not dependent on the transfer time, Δt . To determine which case of the Lagrange time equations to use for a particular orbital transfer problem, you must account for the given values of Δv and Δt , which will then narrow the possibilities down to an arc of a single conic section. As stated previously, Δv may be used to establish the choice of sign in Eq. (7-36). Once the sign has been determined, corresponding to a short-way or long-way transfer, the given transfer time will then completely determine the type of conic section by comparison to the known parabolic and minimum-energy transfer times. If Δt is less than Δt_p , the arc is hyperbolic. If Δt lies between Δt_p and Δt_{min} , the path is a type A, short-way elliptic arc. Finally, if Δt is greater than Δt_{min} , the path is a type B, long-way elliptic arc. The expressions for Δt_p and Δt_{min} are as follows:

$$\Delta t_p = \frac{1}{3} \sqrt{\frac{2s^3}{\mu}} \left\{ 1 \mp \left(\frac{s-c}{s} \right)^{\frac{3}{2}} \right\} \quad (7-43)$$

$$\Delta t_{min} = \sqrt{\frac{s^3}{8\mu}} \left\{ \pi \mp \left(2 \sin^{-1} \sqrt{\frac{s-c}{c}} - \sin \left(2 \sin^{-1} \sqrt{\frac{s-c}{c}} \right) \right)^{\frac{3}{2}} \right\}$$

Lagrange's time equations [Eq. (7-36)] give a set of closed-form relationships between the observed transfer time and the semimajor axis (a) of the conic arc. However, the functional dependence is the opposite of what would be desired, because Δt comes from observations or mission plans, but a is the unknown quantity to be found. If values of a are plotted as a function of Δt as shown in Fig. 7-9, it is evident that semima-

major axis is a single-valued function of time-of-flight. The plot may be divided into three regions, representing the hyperbolic, elliptical short-way and elliptical long-way cases. The hyperbolic region is characterized by negative values of semimajor axis. The parabolic transfer time Δt_p forms the boundary between the hyperbolic and short-way elliptic cases. The minimum energy transfer time Δt_{min} forms the boundary between the second and third regions, which correspond to the short-way and long-way elliptic cases. In the third region, there are two separate boundaries (B_{min} , B_{long}) to examine. The first boundary is the minimum-energy transfer time, and the second boundary is the region where the transfer time approaches infinity.

Physically, as the semimajor axis of a long-way elliptic transfer becomes very large, the transfer time will approach the value of the period of the complete closed orbit since the complementary time on the short-way arc becomes very small. So, the plot of the semimajor axis in the third region of Fig. 7-9 will approach a function that is proportional to $\Delta t^{2/3}$, as may be seen from solving the orbital period for the semimajor axis [Eq. (1-29)].

Many excellent root-finding techniques exist to solve the Lagrange time equations and some of them are described elsewhere in this text. However, because a is the unknown quantity to find, it is quite useful to find a direct solution which is explicit in the transfer time. To begin the analytical development of the series solution of Lambert's Theorem, you must first convert the three cases of the Lagrange time equations [Eq. (7-36)] into series form.

Hyperbolic and Short-Way Elliptical Series Expression

For both the hyperbolic and short-way elliptical cases, use the same series expression.

$$\frac{\Delta t}{\Delta t_p} - 1 = \sum_{i=1}^{\infty} \frac{\left(1 \mp \left(\frac{s-c}{s}\right)^{\frac{3}{2}}\right)^{i+\frac{3}{2}}}{\left(1 \mp \left(\frac{s-c}{s}\right)^{\frac{3}{2}}\right)^i} \frac{\left(\frac{1}{2}\right)_i \left(\frac{3}{2}\right)_i \left(\frac{s}{2a}\right)^i}{\left(\frac{5}{2}\right)_i i!} = \sum_{i=1}^{\infty} A_i \left(\frac{s}{2a}\right)^i \quad (7-44)$$

where the parabolic transfer time is given by Euler's equation

$$\Delta t_p = \frac{\sqrt{2}}{3} \sqrt{\frac{s^{\frac{3}{2}}}{\mu}} \left\{ 1 \mp \left(\frac{s-c}{s}\right)^{\frac{3}{2}} \right\}$$

The terms in equation Eq. (7-44) with subscripts i are Pochhammer symbols, which were chosen to simplify the expression. Thorne and Bain (1995) show that the time equation for the hyperbolic case differs only in the sign of the argument, $-s/2a$, and the series coefficients are identical. Eq. (7-44) is expressed more generally as follows:

$$T = \left(\frac{\Delta t}{\Delta t_p} - 1 \right) = \sum_{n=1}^{\infty} A_n \left(\frac{s}{2} \right)^n a^{-n} \quad (7-45)$$

The series coefficients A_n in Eq. (7-45), as well as other series coefficients throughout this section, are labeled underneath as appropriate to indicate which case of the Lagrange time equations they represent. Eq. (7-45) is the first of three series expansions needed to represent the Lambert problem.

The remaining case of the Lagrange time equations, given by Eq. (7-36), addresses the long-way elliptical transfers. There are two series expressions required to represent boundary conditions when the transfer time is either near the minimum energy time or as the transfer time approaches infinity, and they will be derived next.

Long-Way Minimum Energy Elliptical Series Expression (B_{me})

For a minimum energy orbital transfer, the semimajor axis of the conic section is given by Eq. (7-37). This is the smallest possible value of a for an elliptic orbit, which may be seen in Fig. 7-9. We can expand Eq. (7-36) about the minimum energy value for the semimajor axis by substituting $a = s/2 + x^2$ and using x as the new independent variable, because the sum of $s/2 + x^2$ will always be greater than or equal to the minimum value of semimajor axis. Using this substitution, the Lagrange time equation for a long-way, elliptical transfer in the neighborhood of the minimum energy time is as follows:

$$\begin{aligned} \Delta t = & \left(\frac{\left(\frac{s}{2} + x^2 \right)^{3/2}}{\mu} \right) \left(2\pi - \left[2 \sin^{-1} \left(\sqrt{\frac{s}{s+2x^2}} \right) - \sin \left(2 \sin^{-1} \left(\sqrt{\frac{s}{s+2x^2}} \right) \right) \right] \right. \\ & \left. \mp \left[2 \sin^{-1} \left(\sqrt{\frac{s-c}{s+2x^2}} \right) - \sin \left(2 \sin^{-1} \left(\sqrt{\frac{s-c}{s+2x^2}} \right) \right) \right] \right) \end{aligned} \quad (7-46)$$

where the series expressions for the various terms in Eq. (7-46) are given by:

$$\begin{aligned} \left(\frac{s}{2} + x^2 \right)^{3/2} &= \sum_{n=0}^{\infty} \frac{\left(\frac{3}{2}\right)\left(\frac{1}{2}\right)\dots\left(\frac{3}{2}-n+1\right)}{n!} \left(\frac{s}{2} \right)^{\left(\frac{3}{2}-n\right)} x^{2n} \\ \alpha &= 2 \sin^{-1} \left(\sqrt{\frac{s}{s+2x^2}} \right) = \pi - 2\sqrt{2} \sum_{n=0}^{\infty} \frac{(-2)^n}{2n+1} s^{\left(-n-\frac{1}{2}\right)} x^{(2n+1)} \\ \sin(\alpha) &= \sin \left(2 \sin^{-1} \left(\sqrt{\frac{s}{s+2x^2}} \right) \right) = -\sqrt{2} \sum_{n=0}^{\infty} (-2)^{(n+1)} s^{\left(-n-\frac{1}{2}\right)} x^{(2n+1)} \end{aligned}$$

The series expressions for the β terms involve double summations:

$$\beta = 2 \sin^{-1} \left(\sqrt{1 - \frac{c}{s}} + 2 \sqrt{c(s-c)} \right) \sum_{n=1}^{\infty} \left[\sum_{m=1}^n S\beta C_{n,m} c^{(n-m)} s^{(m-1)} \right] \left(\frac{-1}{cs} \right)^{(n-1)} \frac{x^{2n}}{(n-1)!}$$

where the βC coefficients are given by:

$$\beta C = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots \\ 2 & 1 & 0 & 0 & \dots \\ 8 & 4 & 3 & 0 & \dots \\ 48 & 24 & 18 & 15 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix} \quad (7-47)$$

The matrix in Eq. (7-47) is formed by multiplying row $(n-1)$ times $(2n-2)$ and the last non-zero element in each row is given by $(2n-3)$ times the previous last non-zero element, and this process may be continued to any order necessary.

$$\sin(\beta) = 2 \sqrt{\frac{c(s-c)}{s^2}} \left\{ 1 + \sum_{n=1}^{\infty} \left[\sum_{m=1}^n S\beta C_{n,m} c^{(n-m)} s^{(m-1)} \right] \left(\frac{-1}{cs} \right)^{(n-1)} \frac{x^{2n}}{(n-1)!} \right\}$$

where the $S\beta C$ coefficients are given by:

$$S\beta C = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots \\ 2 & -1 & 0 & 0 & \dots \\ 8 & -4 & -1 & 0 & \dots \\ 48 & -24 & -6 & -3 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix}$$

This matrix is formed similar to the matrix in Eq. (7-47). To get the final series form of Eq. (7-46), a series multiplication is required after combining the α and β terms in Eq. (7-46). Given two general series (A and B), the coefficients of the product $AB = C$ is

$$A = \sum_{n=0}^{\infty} A_n x^n \quad B = \sum_{n=0}^{\infty} B_n x^n$$

$$C_n = AB = \sum_{k=0}^n A_k B_{n-k}$$

Thus, the A coefficients may be determined for Eq. (7-48)

$$\Delta t - \Delta t_{min} = \sum_{n=1}^{\infty} A_n x^n \quad (7-48)$$

where Δt_{min} is the known minimum-energy transfer time for the given problem geometry, when $a = s/2$. Eq. (7-48) is the second of the three required series expansions for the Lambert problem.

Long-Way Infinity Elliptical Series Expression

For the case where the transfer time approaches infinity, the series expansion of Eq. (7-36) is:

$$\Delta t = \frac{2\pi n_{rev}}{\mu} a^{\frac{3}{2}} - \sum_{n=1}^{\infty} \left[1 \mp \left(\frac{s-c}{s} \right)^{\left(n + \frac{3}{2} \right)} \right] + \frac{\left(\frac{1}{2} \right) n \left(\frac{3}{2} \right) n}{\left(\frac{5}{2} \right) n} \left(\frac{s}{2a} \right)^n$$

In this case, the first term of the series has a power of $(3/2)$, and the remaining powers are the negative integers. The integer number of complete revolutions (n_{rev}) is included in the first series coefficient (A_0), and it may be chosen as part of the input data. Thus, you may form a single power series to combine all terms as follows, where the coefficients of the remaining half-powers below $3/2$ are known to be zero:

$$\Delta t = \sum_{n=0}^{\infty} A_n a^{\left(\frac{3-n}{2} \right)} \quad (7-49)$$

Eq. (7-49) gives the final series expansion to cover all cases of the Lambert problem. Next, all three series expansions are solved for the semimajor axis.

Hyperbolic and Short-Way Elliptical Series Solution

Thorne and Bain (1995) show that the series solution for both the hyperbolic and short-way elliptical classes of transfer arcs is given by:

$$a = \frac{s}{2} \sum_{n=1}^{\infty} B_n \left(\frac{\Delta t}{\Delta t_p} - 1 \right)^{n-2} \quad (7-50)$$

The B_n coefficients in Eq. (7-50) may be determined through the construction of a matrix equation, where the matrix Q transforms the original series coefficients A_n from Eq. (7-45) to produce the required simultaneous series reversion and inversion. Once the Q matrix has been constructed, the unknown coefficients B are generated by a simple matrix multiplication:

$$\vec{B} = \vec{Q}\vec{A} \quad (7-51)$$

where $B_1 = A_1$, $\vec{A} = [A_2, A_3, \dots, A_n]$ and $\vec{B} = [B_2, B_3, \dots, B_n]$ in Eq. (7-51). The matrix \vec{Q} is defined as

$$\vec{Q} = \begin{bmatrix} A_1^{-1} & 0 & 0 & 0 & 0 & \dots \\ -\frac{A_2}{A_1^3} & A_1^{-2} & 0 & 0 & 0 & \dots \\ \frac{2A_2^2 - A_1 A_3}{A_1^5} & -\frac{2A_2}{A_1^4} & A_1^{-3} & 0 & 0 & \dots \\ -\frac{5A_2^3 + 5A_1 A_2 A_3 - A_1^2 A_4}{A_1^7} & \frac{5A_2^2 - 2A_1 A_3}{A_1^6} & -\frac{3A_2}{A_1^5} & A_1^{-4} & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots & A_1^{-i} \end{bmatrix}$$

Although the \vec{Q} matrix was originally derived to accomplish the simultaneous reversion and inversion of a general series, it is interesting to notice that the first column of \vec{Q} is composed of the expressions needed for a series reversion only. Thus, \vec{Q} may also be used in the solution of Eq. (7-48). The recursion relationships needed to produce the elements Q_{ij} of the \vec{Q} matrix are as follows:

$$\begin{aligned} Q(1, 1) &= A_1^{-1} \\ Q(i, j) &= \sum_{k=1}^{i-1} Q(i-k, j-1)(Q(k, 1)) \quad i = 2, 3, 4, \dots \quad 1 < j \leq i \\ Q(i, 1) &= \sum_{k=1}^{i-1} \left(\frac{-1}{A_1}\right) Q(i, k+1) A(k+1) \end{aligned} \quad (7-52)$$

Eq. (7-52) gives the upper left element of \vec{Q} , and the following two series expressions complete the \vec{Q} matrix to the desired order. Finally, the series coefficients to solve the hyperbolic and short-way elliptical cases are given by Eq. (7-53).

$$B_n = \sum_{m=1}^{n-1} Q(n-1, m) A(m+1) \quad (7-53)$$

Long-Way Minimum Energy Elliptical Series Solution

To solve Eq. (7-48), you may perform a straightforward series reversion to get the new series coefficients using only the first column of elements taken from the \vec{Q} matrix in Eq. (7-52) once it has been formed. The series solution of Eq. (7-48) for long-way transfers near the minimum-energy time is given by:

$$a = \frac{s}{2} + x^2 \quad x = \sum_{n=1}^{\infty} B_n (\Delta t - \Delta t_{min})^n \quad (7-54)$$

Long-Way Infinity Elliptical Series Solution

Finally, for the long-way elliptical case with transfer times that are much larger than the minimum-energy time, the series representation of Eq. (7-49) must first be put into the proper form for a series reversion, then reversed, and finally solved for the semimajor axis. These steps are as follows. Isolate $a^{-1/2}$ from Eq. (7-49) as the independent variable:

$$\Delta t = \sum_{n=0}^{\infty} A_n \left(a^{-\frac{1}{2}} \right)^{n-3}$$

Factor out $(a^{-1/2})^{-3}$ so that the leading series term is a constant.

$$\Delta t = \left(a^{-\frac{1}{2}} \right)^{-3} \sum_{n=0}^{\infty} A_n \left(a^{-\frac{1}{2}} \right)^n$$

Raise the series to the $(-1/3)$ power so that the $a^{-1/2}$ term may later be brought back in to the series:

$$\Delta t^{-1/3} = \left(a^{-\frac{1}{2}} \right) \left[\sum_{n=0}^{\infty} A_n \left(a^{-\frac{1}{2}} \right)^n \right]^{-1/3}$$

Using a general technique which will be described beginning with Eq. (7-56) below, you can transform the A_n coefficients to the new A'_n and retain $a^{-1/2}$ as the independent variable.

$$\Delta t^{-1/3} = \left(a^{-\frac{1}{2}} \right) \sum_{n=0}^{\infty} A'_n \left(a^{-\frac{1}{2}} \right)^n$$

Return $a^{-1/2}$ to the series, thus putting it into the proper form for reversion by starting with a linear term.

$$\Delta t^{-1/3} = \sum_{n=0}^{\infty} A'_n \left(a^{-\frac{1}{2}} \right)^{n+1}$$

Perform the series reversion, so that $t^{-1/3}$ is now the independent variable with B'_n as the reversion coefficients.

$$a^{-1/2} = \sum_{n=0}^{\infty} B'_n \left(\Delta t^{-\frac{1}{3}} \right)^{n+1}$$

Factor out $t^{-1/3}$ to prepare for squaring the series.

$$a^{-1/2} = \Delta t^{-\frac{1}{3}} \sum_{n=0}^{\infty} B'_n \left(\Delta t^{-\frac{1}{3}} \right)^n$$

Square the series to solve for a .

$$a = \left(\Delta t^{-\frac{1}{3}} \right)^{-2} \left[\sum_{n=0}^{\infty} B'_n \left(\Delta t^{-\frac{1}{3}} \right)^n \right]^{-2}$$

Transform the B'_n coefficients to the new B_n based on squaring the series.

$$a = \Delta t^{\frac{2}{3}} \sum_{n=0}^{\infty} B'_n \left(t^{-\frac{1}{3}} \right)^n$$

Simplify by returning $t^{2/3}$ to the series, and thus we have the form of the solution for the elliptic case as the transfer time approaches infinity, which, along with Eq. (7-50) and Eq. (7-54), completes the set of series solutions for Lambert's Theorem.

$$a = \sum_{n=0}^{\infty} B'_n \Delta t^{\frac{2-n}{3}} \quad (7-55)$$

Two of the above steps involve raising a power series to the negative exponents $-1/3$ and -2 . Because no generalized method for raising a power series to an arbitrary exponent was found in the literature, this represents an original recursive technique to accomplish this task. The series transformation takes the following form.

$$\sum_{n=0}^{\infty} A'_n x^n = \left[\sum_{n=0}^{\infty} A_n x^n \right]^m \quad (7-56)$$

To accomplish this transformation to an arbitrary number of terms, it is useful to develop recursive relationships among the coefficients. If the coefficients of a series

which has been raised to the power m are arranged such as those shown below, it is possible to recognize patterns in the individual terms.

$$\begin{aligned} A'_0 &= A_0^m \\ A'_1 &= mA_0^{m-1}A_1 \\ A'_2 &= \frac{m}{2}(m-1)A_0^{m-2}A_1^2 + mA_0^{m-1}A_2 \\ A'_3 &= \frac{m}{6}(m-1)(m-2)A_0^{m-3}A_1^3 + m(m-1)A_0^{m-2}A_1A_2 + mA_0^{m-1}A_3 \end{aligned} \quad (7-57)$$

Assigning subscripts i and j to the individual terms in each coefficient shown on right hand side of Eq. (7-57), you may then form a matrix \mathbf{P} to examine the relationships between the elements.

$$\mathbf{P} = \begin{bmatrix} A_0 & 0 & 0 \\ mA_0^{m-1}A_1 & 0 & 0 \\ \frac{m}{2}(m-1)A_0^{m-2}A_1^2 & mA_0^{m-1}A_2 & 0 \\ \frac{m}{6}(m-1)(m-2)A_0^{m-3}A_1^3 & m(m-1)A_0^{m-2}A_1A_2 & mA_0^{m-1}A_3 \end{bmatrix}$$

The new series coefficients can be generated to any desired order by using the following recursive relationships to construct the elements of the \mathbf{P} matrix and to combine the results.

$$\begin{aligned} P_{(1,1)} &= A_0^m \\ P_{(i,j)} &= \left[\frac{(m-i+1)...m}{A_0(i-j)!} \right] [P_{(i-j,1)}, P_{(i-j+1,2)}, \dots, P_{(i-1,j)}] \times [A_j, A_{j-1}, \dots, A_1] \\ A'_n &= \sum_{k=0}^n P_{(n+1,k+1)} \end{aligned}$$

To summarize this analysis, the complete series solution of all cases of the Lambert problem is given by Eq. (7-50), Eq. (7-54) and Eq. (7-55). With these three solutions, it is possible to express the Δv required to maneuver a satellite from one orbit to another as an explicit function of the transfer time. The velocity at any point on an orbit is a function of the semimajor axis a , and the three series solutions for a may be truncated to any desired order and substituted, as appropriate, into the Δv equations. The resulting overall function for Δv becomes explicit in the transfer time, which allows it to be minimized using ordinary calculus techniques. Explicit differentiation would not be possi-

ble using the traditional root-finding methods, but the three series solutions presented here allow for any number of orbital maneuvers to be optimized simultaneously.

7.6.4 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. 2.3 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 useful 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. 2.3. The universal-variable solution contained f and g expressions which permitted us to determine the future position and velocity vectors [Eq. (2-68)]:

$$\begin{aligned} f &= 1 - \frac{\chi^2}{r_o} c_2 & g &= t - t_o - \frac{\chi^3}{\sqrt{\mu}} c_3 \\ \dot{g} &= 1 - \frac{\chi^2}{r} c_2 & \dot{f} &= \frac{\sqrt{\mu}}{rr_o} \chi (\psi c_3 - 1) \end{aligned}$$

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. (2-64)]:

$$\begin{aligned} f &= 1 - \frac{r}{p} (1 - \cos(\Delta\nu)) & g &= \frac{rr_o \sin(\Delta\nu)}{\sqrt{\mu p}} \\ \dot{f} &= \frac{\sqrt{\mu}}{p} \left(\frac{1 - \cos(\Delta\nu)}{\sin(\Delta\nu)} \right) \left(\frac{1 - \cos(\Delta\nu)}{p} - \frac{1}{r} - \frac{1}{r_o} \right) & \dot{g} &= 1 - \frac{r_o}{p} (1 - \cos(\Delta\nu)) \end{aligned}$$

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,

$$\chi = \sqrt{\frac{rr_o(1 - \cos(\Delta\nu))}{pc_2}} \quad (7-58)$$

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:

$$\frac{rr_o(1 - \cos(\Delta\nu))}{p} = r_o + r + \frac{\sqrt{rr_o}\sin(\Delta\nu)}{\sqrt{1 - \cos(\Delta\nu)}} \frac{(\psi c_3 - 1)}{\sqrt{c_2}} \quad (7-59)$$

Now, define two new variables:

$$y \equiv \frac{rr_o(1 - \cos(\Delta\nu))}{p} \quad A \equiv \frac{\sqrt{rr_o}\sin(\Delta\nu)}{\sqrt{1 - \cos(\Delta\nu)}} \quad (7-60)$$

and write Eq. (7-59) as

$$y = r_o + r + \frac{A(\psi c_3 - 1)}{\sqrt{c_2}}$$

From Eq. (7-58) and the definition in Eq. (7-60), the universal variable becomes:

$$\chi = \sqrt{\frac{y}{c_2}}$$

Now you can solve the g function equation for time,

$$t - t_o = \frac{\chi^3}{\sqrt{\mu}}c_3 + \frac{rr_o\sin(\Delta\nu)}{\sqrt{\mu p}}$$

and, when you include the new variables,

$$t - t_o = \frac{\chi^3 c_3}{\sqrt{\mu}} + \frac{A \sqrt{y}}{\sqrt{\mu}}$$

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. (2-64) to a form using the newly defined variables:

$$f = 1 - \frac{y}{r_o} \quad \dot{g} = 1 - \frac{y}{r} \quad g = A \sqrt{\frac{y}{\mu}}$$

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,

$$\dot{f} = \frac{f\dot{g} - 1}{g} = \frac{\sqrt{\mu y}(-r - r_o + y)}{r_o r A}$$

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:

$$\hat{r} = f\hat{r}_o + g\hat{v}_o$$

If we solve this expression for the initial velocity vector, we’ll get half the solution:

$$\hat{v}_o = \frac{\hat{r} - f\hat{r}_o}{g}$$

Now, let’s determine the final velocity vector using the velocity-update equation if we’ve calculated the rate of f .

$$\hat{v} = \dot{f}\hat{r}_o + \dot{g}\hat{v}_o$$

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

$$\hat{v} = \frac{f\dot{g} - 1}{g}\hat{r}_o + \dot{g}\frac{\hat{r} - f\hat{r}_o}{g} = \frac{\dot{g}\hat{r} - \dot{g}\hat{r}_o}{g}$$

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. 7-13, 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 computer 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.

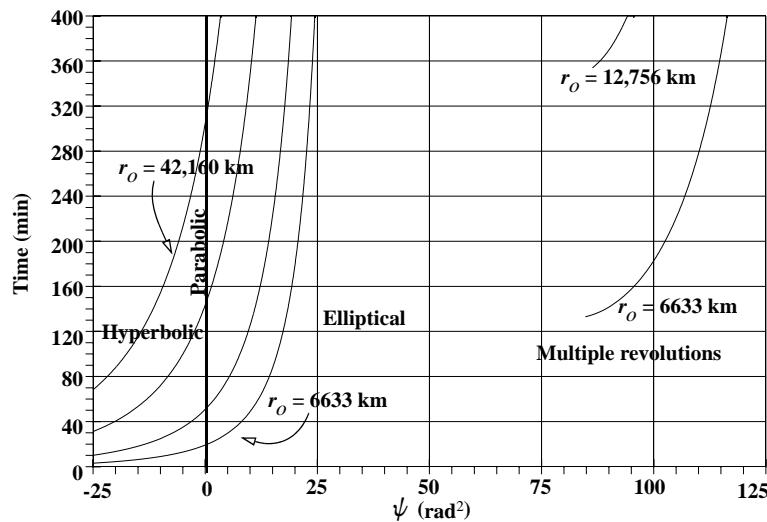


Figure 7-13. Plot of Time versus ψ . Several different transfers are shown with their equivalent values of ψ . Hyperbolic transfers have negative values of ψ , and elliptical orbits have positive values of ψ . Parabolic orbits occur when $\psi = 0$. As the initial orbit position, r_o , gets farther from 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 58: Lambert—Universal Variables ($\hat{r}_o, \hat{r}, \Delta t, t_m \Rightarrow \hat{v}_o, \hat{v}$)

$$\cos(\Delta\nu) = \frac{\hat{r}_o \cdot \hat{r}}{r_o r}$$

$$\sin(\Delta\nu) = t_m \sqrt{1 - \cos^2(\Delta\nu)}$$

$$A = t_m \sqrt{rr_o(1 + \cos(\Delta\nu))}$$

If $A = 0.0$, we can't calculate the orbit.

$$\psi_n = 0.0, \text{ therefore } c_2 = \frac{1}{2} \text{ and } c_3 = \frac{1}{6}$$

$\psi_{up} = 4\pi^2$ 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 $y > 0.0$

$$\chi_n = \sqrt{\frac{y_n}{c_2}}$$

$$\Delta t_n = \frac{\chi_n^3 c_3 + A \sqrt{y_n}}{\sqrt{\mu}}$$

IF $\Delta t_n \leq \Delta t$
 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 $|\Delta t_n - \Delta t| < 1 \times 10^{-6}$

$$f = 1 - \frac{y_n}{r_o} \quad \dot{g} = 1 - \frac{y_n}{r} \quad g = A \sqrt{\frac{y_n}{\mu}}$$

$$\vec{v}_o = \frac{\vec{r} - f \vec{r}_o}{g} \quad \vec{v} = \frac{\dot{g} \vec{r} - \vec{r}_o}{g}$$

7.6.5 Lambert Solution—Battin Method

Battin (1987:325–342) presents 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 Gauss formulations. He extensively uses con-

tinued 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 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 also 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 59 is *not* the same as the sector-to-triangle ratio used in Gauss's development. Relations from Gauss's original 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. (2-65) or Eq. (2-67), 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). Shen and Tsiotras (2003) discuss the iterative modifications necessary to solve multi-revolution problems with Battin's method.

ALGORITHM 59: Lambert—Battin Method ($\vec{r}_o, \vec{r}, \Delta t, t_m \Rightarrow \vec{v}_o, \vec{v}$)

$$\cos(\Delta\nu) = \frac{\vec{r}_o \cdot \vec{r}}{r_o r}$$

$$\sin(\Delta\nu) = t_m \sqrt{1 - \cos^2(\Delta\nu)}$$

$$c = \sqrt{r_o^2 + r^2 - 2r_o r \cos(\Delta\nu)}$$

$$s = \frac{r_o + r + c}{2}$$

$$\epsilon = \frac{r - r_o}{r_o}$$

$$\tan^2(2w) = \frac{\frac{\epsilon^2}{4}}{\sqrt{\frac{r}{r_o} + \frac{r}{r_o} \left(2 + \sqrt{\frac{r}{r_o}}\right)}}$$

$$r_{op} = \sqrt{r_o r} \left\{ \cos^2\left(\frac{\Delta\nu}{4}\right) + \tan^2(2w) \right\}$$

$$\ell = \frac{\sin^2\left(\frac{\Delta\nu}{4}\right) + \tan^2(2w)}{\sin^2\left(\frac{\Delta\nu}{4}\right) + \tan^2(2w) + \cos\left(\frac{\Delta\nu}{2}\right)} \quad 0^\circ < \Delta\nu < 180^\circ$$

$$\ell = \frac{\cos^2\left(\frac{\Delta\nu}{4}\right) + \tan^2(2w) - \cos\left(\frac{\Delta\nu}{2}\right)}{\cos^2\left(\frac{\Delta\nu}{4}\right) + \tan^2(2w)} \quad 180^\circ < \Delta\nu < 360^\circ$$

$$m = \frac{\mu\Delta t^2}{8r_{op}^3}$$

Let $x = \ell$ if the orbit is elliptical, 0.0 otherwise

LOOP

$$\eta = \frac{x}{(\sqrt{1+x}+1)^2}$$

$$\xi(x) = \frac{8(\sqrt{1+x}+1)}{3 + \frac{1}{9\eta}}$$

$$c_\eta = \frac{n^2}{(2n)^2 - 1}$$

$$n = 4, 5 \dots$$

$$5 + \eta + \frac{16}{1 + \frac{63\eta}{1 + \frac{c_\eta\eta}{1 + \dots}}}$$

$$h_1 = \frac{(\ell+x)^2(1+3x+\xi(x))}{(1+2x+\ell)(4x+\xi(x)(3+x))}$$

$$h_2 = \frac{m(x-\ell+\xi(x))}{(1+2x+\ell)(4x+\xi(x)(3+x))}$$

Solve Cubic ($y^3 - y^2 - h_1 y^2 - h_2 = 0$) or use:

$$B = \frac{27h_2}{4(1+h_1)^3} \quad U = \frac{B}{2(\sqrt{1+B}+1)}$$

$$\begin{aligned}
& \text{IF } n = (1, 3, 5, \dots) \\
K(U) &= \frac{\frac{1}{3}}{1 + \frac{\frac{4}{27}U}{1 + \frac{\frac{8}{27}U}{1 + \frac{c_U U}{1 + \dots}}}} \quad c_U = \frac{2(3n+1)(6n-1)}{9(4n-1)(4n+1)} \\
&\text{IF } n = (0, 2, 4, \dots) \\
&\quad c_U = \frac{2(3n+2)(6n+1)}{9(4n+1)(4n+3)} \\
y &= \frac{1+h_1}{3} \left(2 + \frac{\sqrt{1+B}}{1+2U(K(U))^2} \right) \\
x &= \sqrt{\left(\frac{1-\ell}{2}\right)^2 + \frac{m}{y^2}} - \frac{1+\ell}{2} \\
&\text{UNTIL } x \text{ stops changing} \\
a &= \frac{\mu \Delta t^2}{16 r_{op}^2 x y^2} \\
&\text{IF } a > 0.0 \\
&\quad \text{SIN}\left(\frac{\beta_e}{2}\right) = \sqrt{\frac{s-c}{2a}} \\
&\quad \text{IF } \Delta\nu > \pi \\
&\quad \beta_e = -\beta_e \\
&\quad a_{min} = \frac{s}{2} \\
&\quad t_{min} = \sqrt{\frac{a_{min}^3}{\mu}} \{ \pi - \beta_e + \text{SIN}(\beta_e) \} \\
&\quad \text{SIN}\left(\frac{\alpha_e}{2}\right) = \sqrt{\frac{s}{2a}} \\
&\quad \text{IF } \Delta t > t_{min} \\
&\quad \alpha_e = 2\pi - \alpha_e \\
&\quad \Delta E = \alpha_e - \beta_e \\
f &= 1 - \frac{a}{r_o} (1 - \cos(\Delta E)) \quad g = \Delta t - \sqrt{\frac{a^3}{\mu}} (\Delta E - \text{SIN}(\Delta E)) \\
\dot{g} &= 1 - \frac{a}{r} (1 - \cos(\Delta E))
\end{aligned}$$

$$\begin{cases}
 \text{IF } a < 0.0 \\
 \sinh\left(\frac{\alpha_h}{2}\right) = \sqrt{\frac{s}{-2a}} \\
 \sinh\left(\frac{\beta_h}{2}\right) = \sqrt{\frac{s-c}{-2a}} \\
 \Delta H = \alpha_h - \beta_h \\
 f = 1 - \frac{a}{r_o}(1 - \cosh(\Delta H)) \quad g = \Delta t - \sqrt{\frac{-a^3}{\mu}}(\sinh(\Delta H) - \Delta H) \\
 \dot{g} = 1 - \frac{a}{r}(1 - \cosh(\Delta H)) \\
 \dot{\vec{v}_o} = \frac{\dot{\vec{r}} - f\vec{r}_o}{g} \quad \dot{\vec{v}} = \frac{\dot{g}\dot{\vec{r}} - \dot{\vec{r}}_o}{g}
 \end{cases}$$

▼ **Example 7-5. Solving Lambert's Problem.**

GIVEN: $\vec{r}_o = 15,945.34 \hat{I}$ km $\Delta t = 76.0$ min

$\vec{r} = 12,214.83899 \hat{I} + 10,249.46731 \hat{J}$ km t_m = short way

FIND: $\dot{\vec{v}}_o, \dot{\vec{v}}$ using each technique

Find several initial parameters for solution:

$$\psi_n = 0.0, \text{ therefore } c_2 = \frac{1}{2} \text{ and } c_3 = \frac{1}{6}$$

$$\psi_{up} = 4\pi^2 \text{ and } \psi_{low} = -4\pi$$

$$\begin{aligned}
 \cos(\Delta\nu) &= \frac{\dot{\vec{r}}_o \cdot \vec{r}}{r_o r} \\
 \sin(\Delta\nu) &= t_m \sqrt{1 - \cos^2(\Delta\nu)}
 \end{aligned}
 \left. \right\} \Delta\nu = 40^\circ$$

Now check to see if the orbit is possible. A must not be zero.

$$A = t_m \sqrt{r r_o (1 + \cos(\Delta\nu))} = 21,188.17 \text{ km}$$

Table 7-1 shows the results of the iteration loop.

The final answers are

$$\dot{\vec{v}}_o = 2.058913 \hat{I} + 2.915965 \hat{J} \text{ km/s}$$

$$\dot{\vec{v}} = -3.451565 \hat{I} + 0.910315 \hat{J} \text{ km/s}$$

Using the minimum-energy solution, we find that

$$\dot{\vec{v}}_o = 2.047409 \hat{I} + 2.924003 \hat{J} \text{ km/s}$$

TABLE 7-1. 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 (ER)	x_n (VER)	Δt_n (TU)	Δt_n (min)	ψ_{n+1}
1	7.845 86	11.059 22	92.696 71	1246.48	19.739 21
2	5.000 00	4.967 29	19.847 16	266.88	9.869 60
3	2.913 81	2.992 49	9.156 43	123.13	4.934 80
10	1.738 69	2.081 70	5.701 62	75.96	2.583 06
17	1.719 50	2.066 89	5.651 77	75.99	2.546 62
25	1.719 54	2.066 93	5.651 88	76.00	2.546 70

$$\hat{\vec{v}} = -3.447\,919 \hat{\vec{I}} + 0.923\,867 \hat{\vec{J}} \text{ km/s}$$

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.

Gauss's original technique finds

$$\hat{\vec{v}}_o = 2.058\,925 \hat{\vec{I}} + 2.915\,956 \hat{\vec{J}} \text{ km/s}$$

$$\hat{\vec{v}} = -3.451\,569 \hat{\vec{I}} + 0.910\,301 \hat{\vec{J}} \text{ km/s}$$

Battin's method requires much fewer iterations (5) and arrives at the same answer:

$$\hat{\vec{v}}_o = 2.058\,913 \hat{\vec{I}} + 2.915\,965 \hat{\vec{J}} \text{ km/s}$$

$$\hat{\vec{v}} = -3.451\,565 \hat{\vec{I}} + 0.910\,315 \hat{\vec{J}} \text{ km/s}$$



7.7 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. 6 for the complex problems of orbital maneuvering. We call this problem intercept and rendezvous or, more generally, the *targeting problem*.

We can apply our solution of Lambert's problem to the 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 6 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. 7-14 shows the overall geometry. We know the initial position and velocity vectors of the interceptor, \vec{r}_{int} and \vec{v}_{int} , and the target, \vec{r}_{tgt} and \vec{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 7-14 shows this symbolically as the movement from \vec{r}_{tgt} and \vec{v}_{tgt} , to \vec{r}_{tgt_b} and \vec{v}_{tgt_b} after TOF, which is known as Kepler's problem (Sec. 2.3).

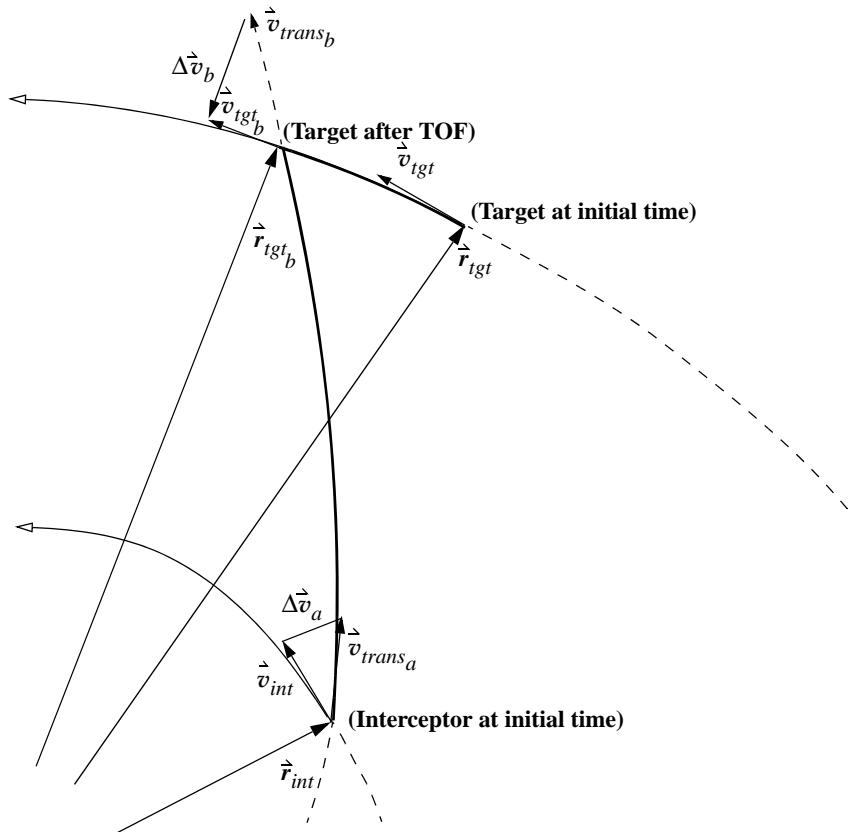


Figure 7-14. Geometry for Intercept Maneuvers. For the targeting maneuver to take place, we must account for the target's motion during the transfer.

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 58) or the Battin method (Algorithm 59) 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. 8 and 9 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 solution 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. A solution for all conic sections is especially useful in this phase because of the problem's geometry. Figure 7-15 shows that, as we increase the TOF for transfer, the transfer trajectory can change through each type of conic section.

Notice if the vectors are exactly 180° apart, we can't solve the problem because there's no distinction between the long and short ways. To solve this special case, we use the Battin, Thorne, or r -iteration technique discussed previously (see Escobal [1965] 1985:213-218). The 180° transfer case also represents a Hohmann-like transfer, which we can solve using techniques from the previous chapter. 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 7-16 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 the maneuver will give the final sign. If both signs are positive, the

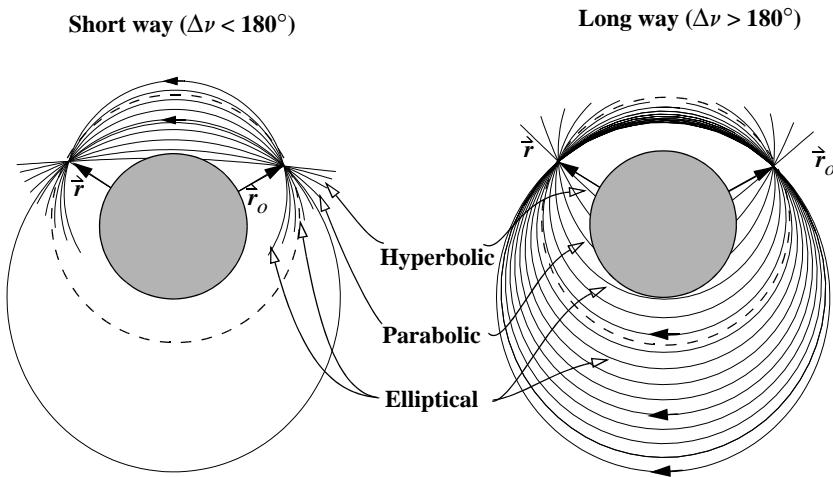


Figure 7-15. Varying Time of Flight for Intercept. As the time of flight increases for the transfer to the target (short way left, long way right), the transfer orbit becomes less eccentric until it reaches a minimum, shown as a dashed line. The eccentricity then begins to increase after this transfer. There is always a minimum eccentricity transfer, and a minimum change in velocity transfer. However, they generally don't occur at the same time. Finally, I've shown the initial and final positions with the same magnitude. This is not an additional requirement.

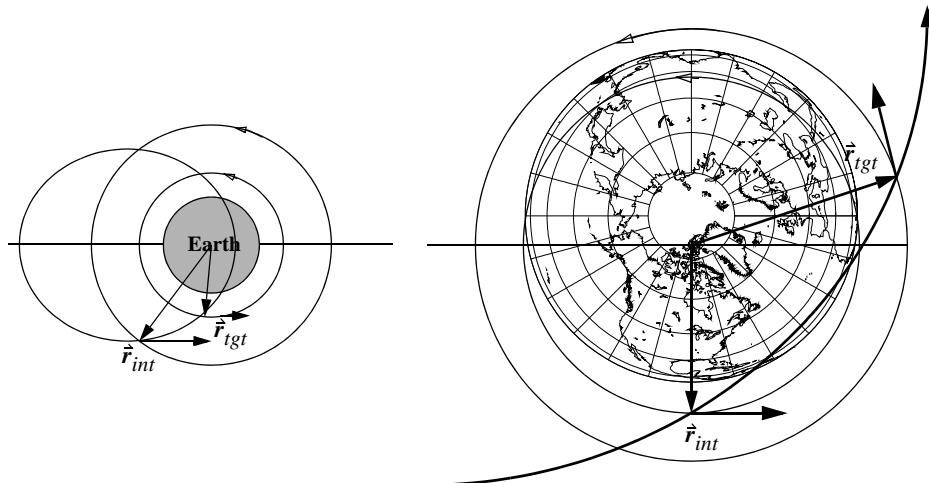


Figure 7-16. Geometry of Orbits that Can Hit the Earth. The orbit transfer on the left is valid because the transfer orbit intersects the Earth after the transfer. The transfer on the right would be unacceptable.

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. 7-16, 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 60: *HIT EARTH* ($\vec{r}_{int}, \vec{r}_{tgt}, \vec{v}_{trans_a}, \vec{v}_{trans_b}, a \Rightarrow r_p, \text{collision}$)

IF $((\vec{r}_{int} \cdot \vec{v}_{trans_a}) < 0.0)$ and $((\vec{r}_{tgt} \cdot \vec{v}_{trans_b}) > 0.0)$

$$\xi = \frac{v_{trans_a}^2}{2} - \frac{\mu}{r_{int}} = -\frac{\mu}{2a}$$

$$h_t = \left| \vec{r}_{int} \times \vec{v}_{trans_a} \right|$$

$$p = \frac{h_t^2}{\mu}$$

$$e = \sqrt{\frac{a-p}{a}}$$

$$r_p = a(1-e)$$

[IF $r_p \leq R_{\oplus}$ (6378.136 km)
 Collision
 ELSE
 No Collision]

With the basics identified, we can form an algorithm to use for each scenario we develop. Algorithm 61 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 61: *TARGET* ($\vec{r}_{int}, \vec{r}_{tgt}, \vec{v}_{int}, \vec{v}_{tgt}, \Delta t \Rightarrow \Delta \vec{v}_a, \Delta \vec{v}_b$)

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:

$$\text{LAMBERT—UNIVERSAL}(\vec{r}_{int}, \vec{r}_{tgt_b}, \Delta t, t_m \Rightarrow \vec{v}_{trans_a}, \vec{v}_{trans_b})$$

$$\Delta \vec{v}_a = \vec{v}_{trans_a} - \vec{v}_{int}$$

$$\Delta \vec{v}_b = \vec{v}_{tgt_b} - \vec{v}_{trans_b}$$

Check for Earth impact:

$$\text{HIT EARTH}(\vec{r}_{int}, \vec{r}_{tgt}, \vec{v}_{trans_a}, \vec{v}_{trans_b}, a \Rightarrow r_p, \text{collision})$$

To gain insight into the two steps for solving this problem, first examine the solution and some characteristics of the Lambert problem. Recall Fig. 7-15, 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. 6. Figure 7-17 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

$$\text{INTERCEPTOR } \vec{r}_{IJK} = -6518.1083 \hat{i} - 2403.8479 \hat{j} - 22.1722 \hat{k} \text{ km}$$

$$\dot{\vec{v}}_{IJK} = 2.604\,057 \hat{i} - 7.105\,717 \hat{j} - 0.263\,218 \hat{k} \text{ km/s}$$

$$\text{TARGET } \vec{r}_{IJK} = 6697.4756 \hat{i} + 1794.5832 \hat{j} + 0.0 \hat{k} \text{ km}$$

$$\dot{\vec{v}}_{IJK} = -1.962\,373 \hat{i} + 7.323\,674 \hat{j} + 0.0 \hat{k} \text{ km/s}$$

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.

As time increases, the velocity decreases to a local minimum. The minimum change in velocity for Fig. 7-17 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. 7-17) 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

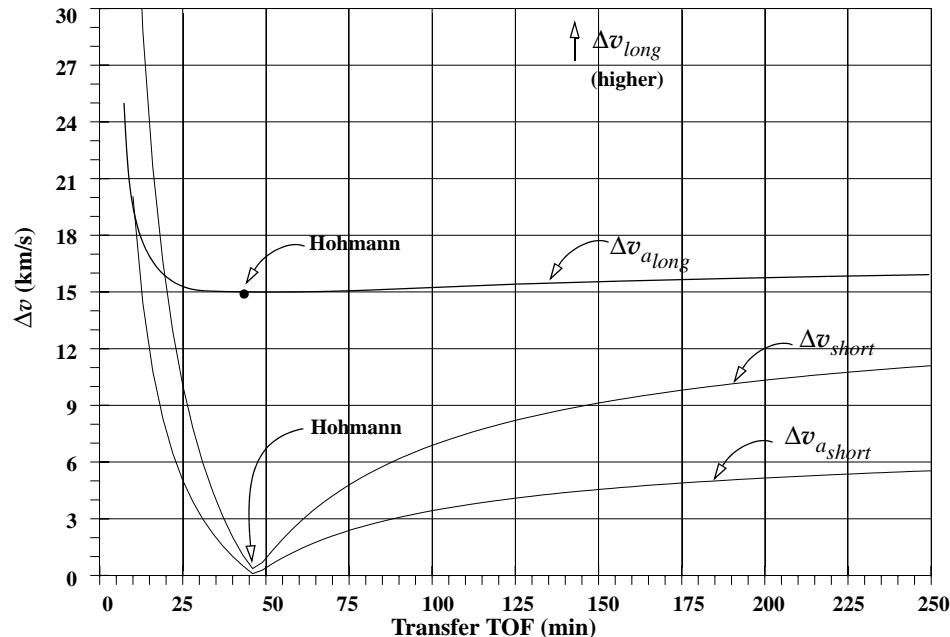


Figure 7-17. Short- and Long-Way Δv for Varying TOF, Target Fixed. As the time of flight increases, the change in velocity reaches a minimum at the point where the transfer mimics a Hohmann transfer. The first, and total change in velocities (Δv_a , Δv) are shown for both short and long transfers.

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 always has 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

$$\begin{aligned}
 \text{INTERCEPTOR} \quad & \dot{\vec{r}}_{IJK} = 5328.7862 \hat{i} + 4436.1273 \hat{j} + 101.4720 \hat{k} \text{ km} \\
 & \dot{\vec{v}}_{IJK} = -4.864\,779 \hat{i} + 5.816\,486 \hat{j} + 0.240\,163 \hat{k} \text{ km/s} \\
 \text{TARGET} \quad & \dot{\vec{r}}_{IJK} = 6697.4756 \hat{i} + 1794.5831 \hat{j} + 0.0000 \hat{k} \text{ km} \\
 & \dot{\vec{v}}_{IJK} = -1.962\,372 \hat{i} + 7.323\,674 \hat{j} + 0.000\,000 \hat{k} \text{ km/s}
 \end{aligned}$$

Figure 7-18 shows the short way and Fig. 7-19 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.

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. 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 computer code. Figure 7-20 shows two possible target locations for developing this test.

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 ($\vec{Tran}_n = \dot{\vec{r}}_{int} \times \dot{\vec{r}}_{tgt}$) and for the interceptor orbit prior to the burn ($\vec{h}_n = \dot{\vec{r}}_{int} \times \dot{\vec{v}}_{int}$). Then,

$$\text{IF } \vec{h}_n \cdot \vec{Tran}_n < 0, \text{ use long way} \quad \text{IF } \vec{h}_n \cdot \vec{Tran}_n > 0, \text{ use short way}$$

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

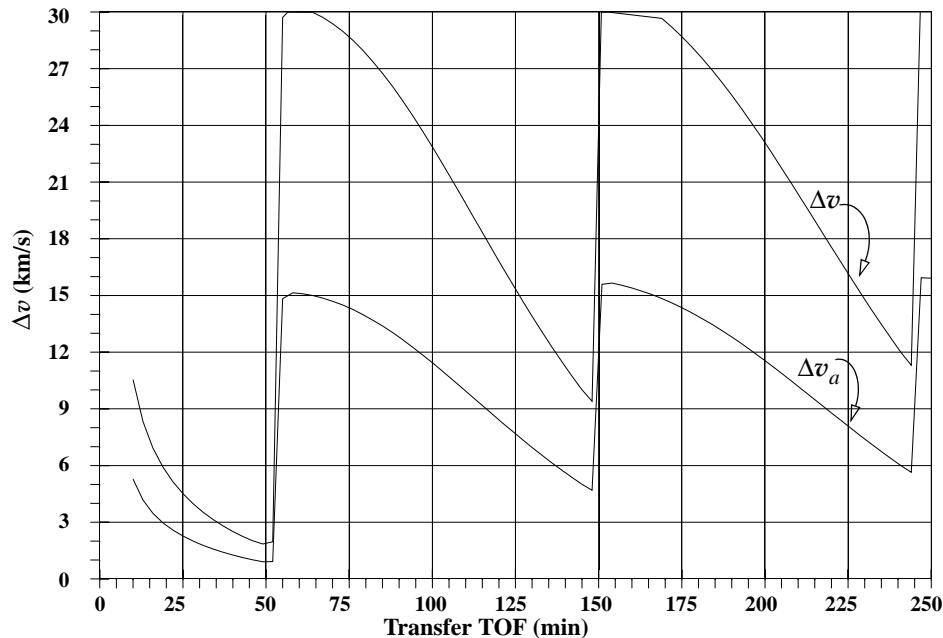


Figure 7-18. Δv for Short Way with Varying TOF, Target Moving. As the time of flight increases, the change in velocity decreases initially, but with more revolutions, the change in velocity tends to rise.

this addition yields a change in velocity graph for the combination of long- and short-way transfers, shown in Fig. 7-21.

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

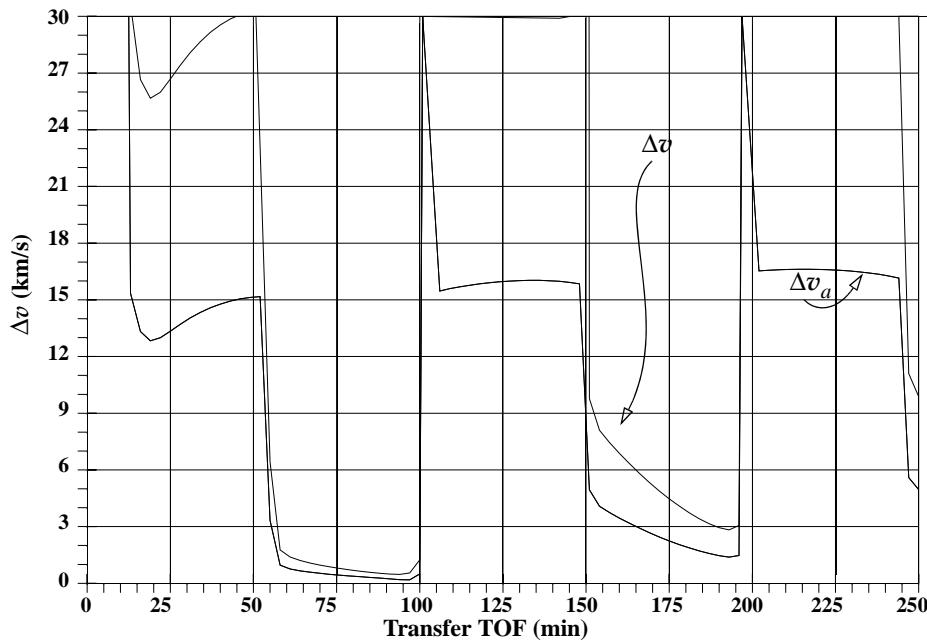


Figure 7-19. Δv for Long Way with Varying TOF, Target Moving. As the time of flight increases, the change in velocity varies significantly. The minimum still occurs at the point where the transfer is a Hohmann transfer.

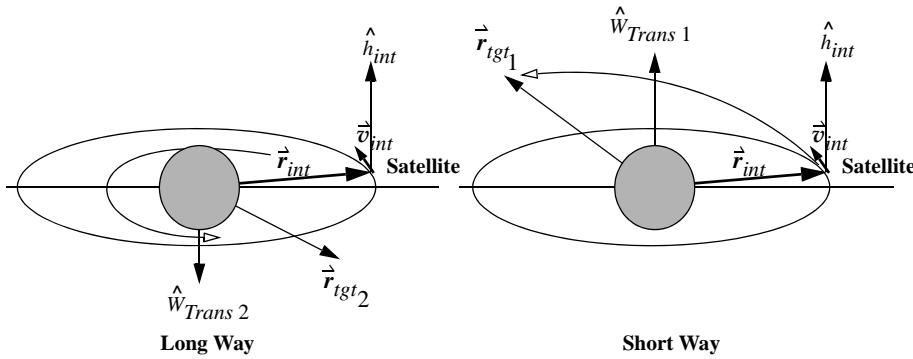


Figure 7-20. Orbital Geometry Showing Transfer Direction. We can determine if we're going the "shortest" way to the target by using cross products. The short-way trajectory is used when the interceptor angular momentum vector, h_{int} , is in the same direction as the transfer unit vector, $W_{Trans\ 1}$. The long way results when h_{int} and $W_{Trans\ 2}$ are in opposite directions.

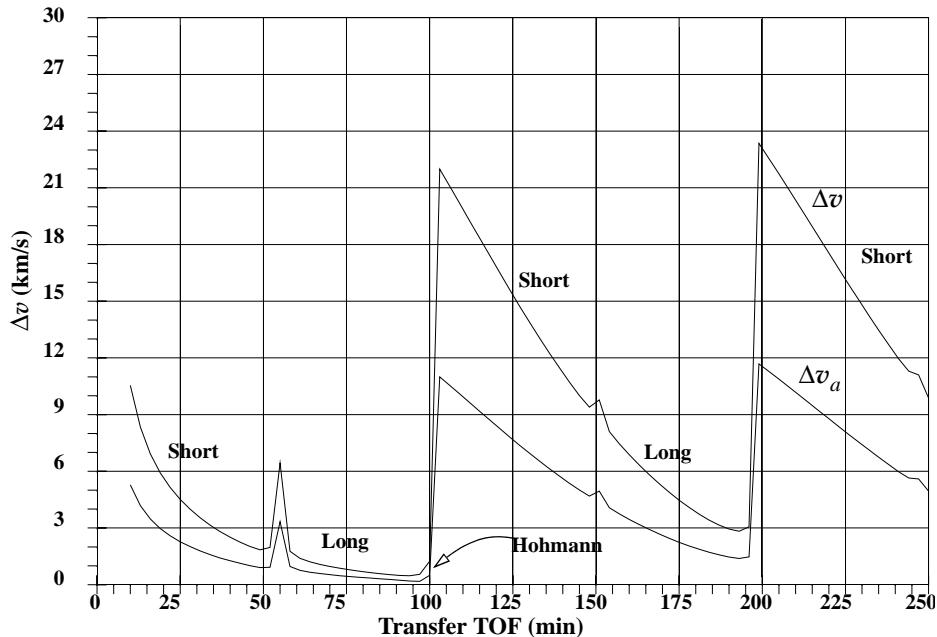


Figure 7-21. Δv for Short and Long Way with Varying TOF, Target Moving. The combination graph for the change in velocity takes the best of both transfers. Remember that the spikes occur because the inclinations differ ($i_{init} = 2^\circ$, $i_{final} = 0^\circ$), and the nodes are different ($\Omega_{init} = 15^\circ$, $\Omega_{final} = \text{undefined}$).

computation. Figure 7-22 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. 7-21. For clarity, Fig. 7-22 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. 7-22, 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. 7-19 and Fig. 7-21. 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. 7-21.

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 to determine when the rendezvous will be complete.

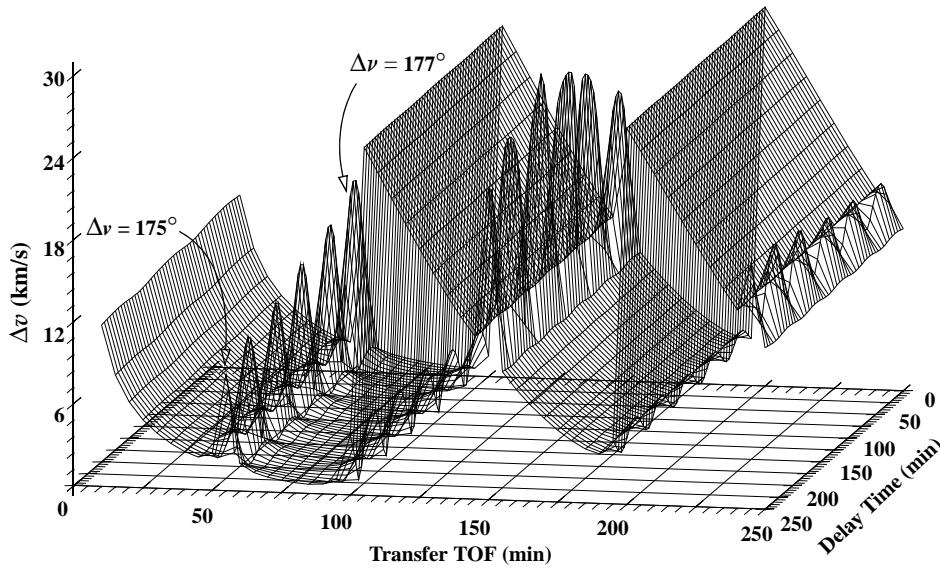


Figure 7-22. Location of Global Minimum. Determining the global minimum is difficult when a delay time is permitted. Even when the curve is rather well-behaved, the three-dimensional graph has too much information for accurate determinations. Notice how the inclination and node differences create large spikes near the $\Delta\nu = 180^\circ$ region.

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 7-23 shows a case with “minimum” separation. Initial vectors are

“MINIMUM” TEST CASE

$$\text{INTERCEPTOR} \quad \hat{r}_{IJK} = 4965.2226 \hat{i} - 1504.1795 \hat{j} + 4617.2886 \hat{k} \text{ km}$$

$$\hat{v}_{IJK} = 3.760\,556 \hat{i} - 3.882\,606 \hat{j} - 5.310\,687 \hat{k} \text{ km/s}$$

$$\text{TARGET} \quad \hat{r}_{IJK} = 5055.2385 \hat{i} - 1599.4069 \hat{j} + 4483.9878 \hat{k} \text{ km}$$

$$\hat{v}_{IJK} = 3.612\,200 \hat{i} - 3.836\,100 \hat{j} - 5.443\,000 \hat{k} \text{ km/s}$$

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

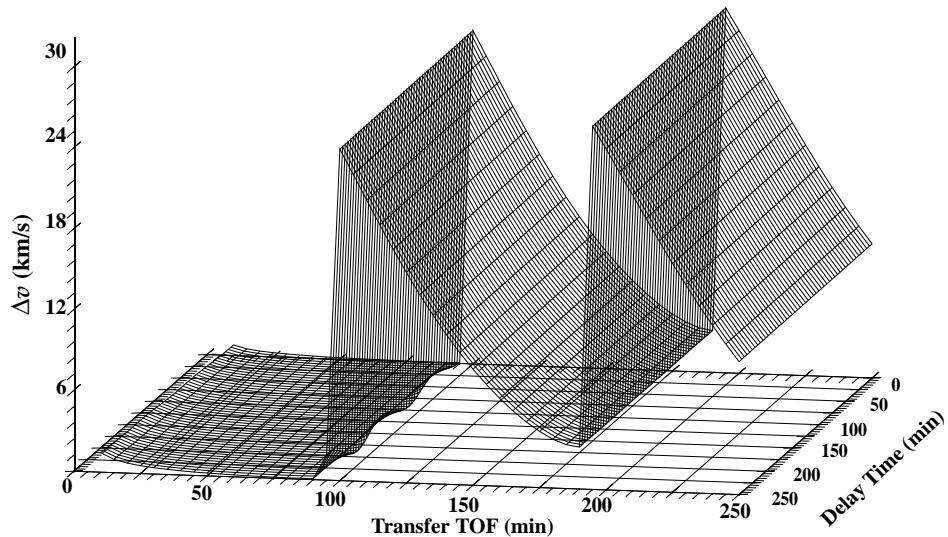


Figure 7-23. Location of Global Minimum (Min Case). For this case, determining the global minimum appears to be simple. But don't forget that a velocity change of 30 km/s is unrealistic! In this case, the inclinations and nodes are very similar; therefore, we don't see any spikes.

took 90 minutes. Because it's difficult to see variations on a large scale, I've enlarged the first section of Fig. 7-23 to produce Fig. 7-24.

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.

A "maximum" separation results in a different topography (Fig. 7-25). 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 tar-

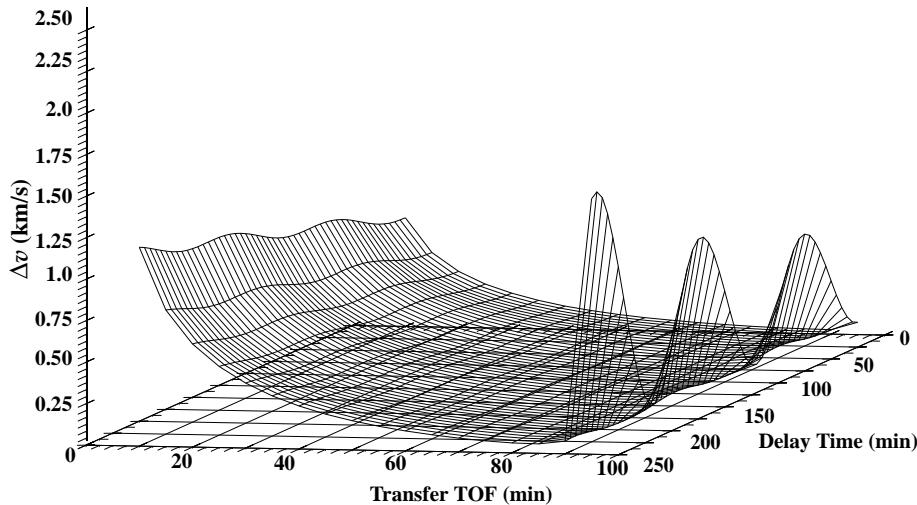


Figure 7-24. Location of Global Minimum (Min Case). With a larger scale, determining the global minimum is simpler, but notice the right-hand edge of the graph. Recall how quickly the curves rose to 30 km/s.

get 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.

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 is very sensitive to these initial parameters. Different initial vectors can result in large velocity change variations. 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.

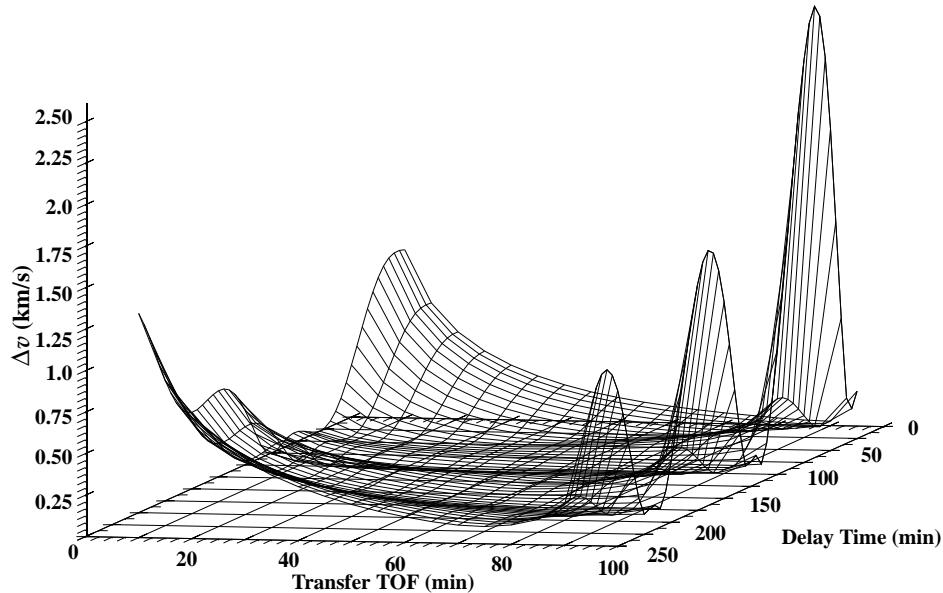


Figure 7-25. Location of Global Minimum (Max Case). Even with a larger scale, the determination of the global minimum is difficult for this “maximum” case. Again notice the right-hand edge of the graph. Recall how quickly the curve rose to 30 km/s. The inclinations and nodes are equal in this example.

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.

General Targeting Scenarios

The general problem of determining the minimum change in velocity needed for a specific orbital transfer is difficult to quantify because there are many variables in the analysis. Several broad solution techniques apply. The preceding discussion of sampling in certain dimensions is often used, but it is not efficient, and clearly is incapable of locating a minimum condition in many cases. We could also perform specific orbital changes to move the state of one satellite to provide better opportunities for the classical techniques. Again, this is not very efficient. A better choice could be the use of a least squares approach to iteratively select from several alternatives. Finally, a robust and structured approach is through the use of genetic algorithms. **Genetic algorithms** are an organized process to determine the optimum solution—in this case, a minimum change

in velocity. Haupt and Haupt (1998) define genetic algorithms as “a subset of evolutionary algorithms that model biological processes to optimize highly complex cost functions. A genetic algorithm allows a population composed of many individuals to evolve under specified selection rules to a state that maximizes the fitness.” Hence, they are perfect for applications like the targeting problem where multiple constraints affect the solution. Let’s consider another example to illustrate (conceptually), how these techniques would operate.

Suppose we have two satellites in different orbits (inclination and size) and we wish to rendezvous in a given amount of time. The preceding analysis (Sec. 7.7) can be time consuming to implement and determine an optimum maneuver to accomplish the transfer. We can simplify the process by combining a series of simple maneuvers to accomplish the desired goal. First, the interceptor must reach one of the two “common” points between the two orbits, although they may not be at the same altitude. At this point, a Δv is performed to rendezvous with the target satellite in a given number of revolutions. This can occur only if the “common” point is at the same altitude. If the altitudes are different, the Δv is used to adjust the size of the interceptor’s orbit to match the target satellite either $\frac{1}{2}$ a revolution, or a full revolution later. Notice this works only for the case of an intercept because the Δv for the final plane change to enable a rendezvous could be prohibitive depending on the inclination change that’s required. Thus, both of these techniques are intended for only the simplest of applications.

We could also solve this problem using a least squares technique [Chioma (2001)] (Least squares is discussed in Chap. 10). In this case, the errors in the “observations” for each trial are both the errors in the final position and velocity vectors (from the target’s state), as well as the magnitudes of each burn. The state is simply the components and timing of each burn. The partial derivatives contain the orbital dynamics and are often simple relations. You can weight the solution by imposing “restrictions” on the time, or the duration of any burns.

The genetic algorithm approach is potentially the most complex to set up, but it produces the most optimal answer. The state remains the initial position and velocity vector of the interceptor satellite, and the change in velocity (and / or time to rendezvous, etc.) is the constraints. The algorithm takes an initial solution to the problem and perturbs components of the state to determine their effect on the overall change in velocity. The difference with a genetic algorithm is the systematic selection and combination of trials, and the progression through the topography of the solution space. The power of this technique is slowly emerging, and applications are continuing to be found. Haupt and Haupt (1998) provide numerous examples to illustrate genetic algorithms. An excellent example for orbital dynamics is Smith (1999). He shows an application for the design and maintenance of satellite constellations through parallel genetic algorithms.

Problems

1. Gauss tried to locate Ceres with the following data (Montenbruck (1994, 245)):

Sept. 5, 1805	24:16:5	$\alpha = 6^{\text{h}} 23^{\text{m}} 57.54^{\text{s}}$	$\delta = 22^{\circ} 21' 27.08''$
Jan. 17, 1806	22:09:5	$\alpha = 6^{\text{h}} 45^{\text{m}} 14.69^{\text{s}}$	$\delta = 30^{\circ} 21' 24.20''$
May 23, 1806	20:39:9	$\alpha = 8^{\text{h}} 07^{\text{m}} 44.60^{\text{s}}$	$\delta = 28^{\circ} 02' 47.04''$

What are the orbital elements? (Hint: use Laplace's technique)

2. Use an angles-only technique to determine the orbit of the following satellite in Earth orbit beginning on May 14, 1996, and ending on May 15, 1996:

h	min	s	α_t (°)	δ_t (°)	ϕ_{sd} (°)	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
0	0	0.00	63.838 841 0	-7.293 504 6	68.073 414	-7.157 751	918.883 650
0	10	0.00	65.745 694 8	-8.219 761 7	72.521 101	-133.955 230	924.309 096
0	20	0.00	68.801 384 8	-5.353 928 7	39.708 368	-160.146 385	917.248 548
0	30	0.00	72.857 869 6	1.049 054 9	5.280 043	-169.644 991	910.429 439
0	40	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	0	0.00	81.299 487 6	11.250 311 9	-76.834 806	46.741 203	922.076 632
1	10	0.00	82.225 627 7	8.163 782 0	-45.063 420	8.959 525	908.584 733
1	20	0.00	83.148 682 5	3.554 594 9	-10.600 064	-1.284 745	896.563 207
1	30	0.00	84.652 559 4	-1.379 516 9	24.094 841	-9.783 934	899.905 334
1	40	0.00	87.261 216 4	-5.439 902 6	58.178 493	-23.462 257	914.684 420
1	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?

3. 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?
4. Use the following data to determine the position of a newly discovered asteroid. The data is from MSSS during a preliminary collection on asteroids. Are parallax and reduction techniques necessary? Hint: see Table 4-2 for the location of MSSS, and be sure to use the transformations in Chap. 4 to preprocess the data.

Year	Month	Days (UTC)	α_t (h min sec)	δ_t ($^{\circ}$ ' '')
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

5. Program Example 6-10 and explore options to decrease the change in velocity. How does your result compare to the solution of Example 6-10?
6. 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 1029), program the technique and compare your results to the examples in the application section (Sec. 7.7).
7. Use Table 10-2 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.
8. For the solution of Lambert’s problem using universal variables (Algorithm 58), we know from Fig. 2-5 that c_2 can equal zero. Is this a problem we must be careful of for computer implementations?
9. We know that Algorithm 58 can accommodate multiple revolution transfers by adjusting the iteration parameters. Does anything else need to be fixed? Does the same process apply to Algorithm 59?
10. Investigate the claim that Gauss’s angles-only method works best for observations that are less than about 10° apart. Use the data in Example 7-2 with points 6, 8, 10 and 6, 7, 8 to test your hypothesis.
11. Problem 6 in Chap. 6 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 9-4 has relevant orbital data for the Space Shuttle and MIR.)

CHAPTER 8 SPECIAL PERTURBATION TECHNIQUES

- 8.1 Historical Background
- 8.2 Introduction to Perturbations
- 8.3 Encke’s Formulation
- 8.4 Cowell’s Formulation
- 8.5 Numerical Integration Methods
- 8.6 Disturbing Forces
- 8.7 Forming Numerical Solutions
- 8.8 Practical Considerations

8.1 Historical Background

The fundamental problem in perturbation analysis is orbit propagation. Unlike the two-body analysis of Algorithm 7, the most accurate way to analyze perturbations is numerically. We often discuss the sheer complexity of numerical integration, 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 to perform 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 manipulating and dealing with larger numbers grew, so did the complexity of the theories. It quickly became apparent that the complexity of these theories could easily outpace the advances in computing technology. The problem then became how to efficiently evaluate complex theories for computational applications.

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. 9.) 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 described 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 was published between 1823 and 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 on elliptic functions and number theory. Legendre functions are the 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 local 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 attained enough precision with limited computing abilities. I'll only briefly introduce this method because computing technology has overcome the need to use such techniques.

Most 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 two of three return visits 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." Cowell's 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.

8.2 Introduction to Perturbations

Studying and modeling perturbations are key disciplines in astrodynamics. Although most of the solution techniques have been known for a long time, the literature only recently has begun to present (with increasing exactness) the precise methods needed to satisfactorily solve 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. 9. Blitzer was one of the early scientists to begin detailed analysis of the perturbative effects on near-Earth satellites, 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 undisturbed 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 or mean motions of the celestial bodies. Actual 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 accurately determined the data constants and coefficients for each central body.

Bate, Mueller, and White (1971:385–386) introduce perturbations as follows. We're familiar with random 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 that was based on previous measurements. In this example, the wind gust is a random perturbing force. Fortunately, most of the perturbing forces considered in orbital mechanics are more predictable and analytically tractable than a wind gust. The usual perturbing forces we consider for the orbit problem are both random and predictable. They include the asphericity of the central body, atmospheric drag and lift, third body effects, solar-radiation pressure, thrust, magnetic fields, solid-Earth and ocean tides, Earth re-radiation (albedo), relativistic effects, and others. I'll show how to analyze most of these later.

Don't get the idea that all deviations are small—they can be comparable to the primary attracting force (two-body gravitation). Whenever they're this large, they cease to be perturbations because the fundamental nature of the problem is different. 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 the three main approaches to examine the effects of perturbations: *special perturbation techniques* (using numerical methods—Sec. 8.5), *general perturbation techniques* (using analytical methods—Chap. 9), and *semianalytical techniques* (a combination of analytical and numerical techniques—Chap. 9).

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 can 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 lengthy computing times. NASA began the first complex numerical integrations during the late 1960s and early 1970s. Personal computers now compute sufficiently fast enough to perform complex perturbation analyses using numerical techniques. However, numerical methods suffer from errors that build up with truncation and round-off due to fixed computer wordlength. These errors can cause 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 8-1, which is simple, first order, and defined with a given initial condition. Find the value of this system when time, t , equals 0.5^s, assuming $\omega = 1 \text{ rad/s}$. Obviously, this example permits a direct solution if we recognize the integral as $(\text{SIN}(\omega t)/\omega)$.

▼ Example 8-1: Using Direct Integration.

GIVEN: $\dot{y}(t) = \cos(\omega t)$

$y(t = 0) = y_o = 0$

$\omega = 1 \text{ rad/s}$

FIND: $y(t = 0.5)$

Remembering to evaluate both limits of the 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 in Example 8-1. 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 8-2. Using a Numerical Technique.

GIVEN: $\dot{y}(t) = \cos(\omega t)$

$$y(t=0) = y_0 = 0, \omega = 1 \text{ rad/s}$$

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):

$$y(t=0.5) = y(t_o) + \cos(t_o)(t - t_o) = 0.0 + \cos(0.0)(0.5 - 0) = \underline{\underline{0.5}}$$

Notice how far off from the correct answer this result is. Now try a step size of 0.1 (Let $\Delta t = 0.1$).

$$t_k = 0$$

$$y(t_{k+1}=0.1) = y(t_k) + \cos(t_k)(t_{k+1} - t_k) = 0.0 + \cos(0.0)(0.1 - 0) = 0.1$$

$$t_k = t_{k+1}, t_{k+1} = t_k + (t_{k+1} - t_k) = t_k + \Delta t$$

$$y(t_{k+1}=0.2) = y(t_k) + \cos(t_k)(t_{k+1} - t_k) = 0.1 + \cos(0.1)(0.2 - 0.1) = 0.199\,500\,4$$

$$t_k = t_{k+1}, t_{k+1} = t_k + \Delta t$$

$$y(t_{k+1}=0.3) = y(t_k) + \cos(t_k)(t_{k+1} - t_k) = 0.199\,500\,4 + \cos(0.2)(0.3 - 0.2) = 0.297\,507\,1$$

$$t_k = t_{k+1}, t_{k+1} = t_k + \Delta t$$

$$y(t_{k+1}=0.4) = y(t_k) + \cos(t_k)(t_{k+1} - t_k) = 0.297\,507\,1 + \cos(0.3)(0.4 - 0.3) = 0.393\,040\,7$$

$$t_k = t_{k+1}, t_{k+1} = t_k + \Delta t$$

$$\blacktriangleleft y(t_{k+1}=0.5) = y(t_k) + \cos(t_k)(t_{k+1} - t_k) = 0.393\,040\,7 + \cos(0.4)(0.5 - 0.4) = \underline{\underline{0.485\,146\,8}}$$

Notice several 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 initial condition, we must recompute the entire sequence. 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. Some integrators used for solving the orbit problem employ this fixed step-size approach. Note the improvement with the smaller step size. The obvious conclusion from these two cases is that we should use the smallest possible step size of, say, 0.000 000 01, to obtain the most accurate answer of 0.479 425 539 216 302, right? Well, maybe. This choice *is* more accurate than the analytical method (see Sec. 9.2), but it uses too much computational time. At some point, the benefit of using a smaller and smaller step size is overwhelmed by the roundoff and truncation errors. The problem of determining an appropriate step size can be significant 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 heat exchange (friction) or external sources (thrust). 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.

You should 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 a specific 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. 1.3.3). 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 instead of simply the *potential*, to avoid confusion with *potential energy*.

The distinction between a specific force (used interchangeably with acceleration in this text) 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 $\hat{r} = x\hat{I} + y\hat{J} + z\hat{K}$ having a magnitude of $r = \sqrt{x^2 + y^2 + z^2}$, the acceleration is

$$\nabla U_{2\text{-body}} = -\frac{1}{2} \frac{\mu}{(x^2 + y^2 + z^2)^{3/2}} [2x\hat{I} + 2y\hat{J} + 2z\hat{K}] = -\frac{\mu}{r^3} \hat{r}$$

Notice the result, as expected, is the two-body acceleration for a satellite. Thus we can write the two-body equation [Eq. (1-14)] very compactly:

$$\ddot{\hat{r}}_{2\text{-body}} = \nabla U_{2\text{-body}} \quad (8-1)$$

If we add conservative perturbative forces to the simple Newtonian attraction, the equation of motion becomes

$$\ddot{\hat{r}} - \nabla U_{2\text{-body}} = \nabla R \quad \text{or} \quad \ddot{\hat{r}} = \nabla U \quad (8-2)$$

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 .

8.3 Encke's Formulation

As mentioned in the introduction, Encke's formulation* isn't very popular today because computers can easily perform a numerical solution to many problems with enough precision to retain the accuracy of the theories. Still, it is historically relevant and it illustrates a computational technique that we can apply to other problems. Refer

to Plummer (1918:222–224), as well as Brouwer and Clemence (1961:176–185), for classical presentations and details on the formulation. Kaplan (1976:345–348) presents a concise algorithm.

Encke's formulation 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 precision can actually be greater. The process continues until a *rectification point* where the osculating orbit is re-initialized. In Algorithm 62, this point is governed by a tolerance. Kaplan (1976:348) suggests about 1%. Battin (1987:450) notes that the rectification is advantageous in controlling the errors. This is especially useful for interplanetary trajectories. Figure 8-1 shows the geometry.

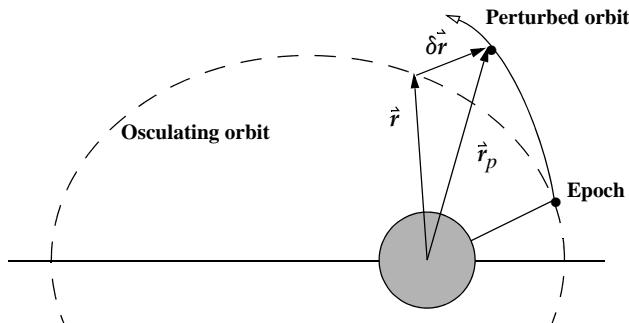


Figure 8-1. Encke's Formulation. The foundation for Encke's technique relies on the difference between the osculating (two-body) orbit and the actual perturbed motion, shown as δr . By integrating this difference, it can save many computations.

We know the two-body and perturbed accelerations:

$$\ddot{\vec{r}} = -\frac{\mu}{r^3} \hat{r} \quad \ddot{\vec{r}}_p = -\frac{\mu}{r_p^3} \hat{r}_p + \hat{a}_p$$

This means the difference between the two accelerations is

$$\ddot{\vec{\delta r}} = \hat{a}_p + \frac{\mu}{r^3} \left\{ \left(1 - \frac{r^3}{r_p^3} \right) \hat{r}_p - \hat{\delta r} \right\}$$

Notice 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 62, below (see also Sec. 8.6.3). We can use two-body relations to move the osculating orbit forward in time.

* I use the formulation terminology to be consistent with the discussion of Cowell's formulation. Encke did, however, use the phrase *neue methode* in Berliner Jahrbuch, c1857.

ALGORITHM 62: *ENCKE* ($\vec{r}_o, \vec{v}_o, t_{start}, t_{stop}, \Delta t \Rightarrow \vec{r}, \vec{v}$)

$$\delta\vec{r} = \vec{0} \quad \epsilon = 0 \quad f = 0 \quad \vec{r}_p = \vec{r}_o \quad \vec{v}_p = \vec{v}_o$$

$$t = t_{start}$$

LOOP

KEPLER ($\vec{r}_o, \vec{v}_o, t - t_{start} \Rightarrow \vec{r}, \vec{v}$)

$$\epsilon = \frac{\vec{r} \cdot \delta\vec{r}}{r^2} \quad f = \frac{1}{\epsilon} \left\{ 1 - \frac{1}{(1-2\epsilon)^{3/2}} \right\}$$

$$\ddot{\vec{r}} = \vec{a}_p + \frac{\mu}{r^3} \{ f\epsilon\vec{r}_p - \delta\vec{r} \}$$

$$\left[\begin{array}{l} \text{IF } \frac{\delta r}{r_p} > \text{tolerance } (\sim 1\%) \\ \quad \vec{r} = \vec{r}_p \quad \vec{v} = \vec{v}_p \end{array} \right.$$

ELSE

$$\vec{r}_p = \vec{r} + \delta\vec{r} \quad \vec{v}_p = \vec{v} + \delta\vec{v}$$

$$t = t + \Delta t$$

UNTIL $t > t_{stop}$

8.4 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. Although 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 to produce a more accurate equation of motion,

$$\ddot{\vec{r}} = -\frac{\mu}{r^3} \vec{r} + \vec{a}_{perturbed} \quad (8-3)$$

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 tides. Fortunately, the form of Eq. (8-3) permits us to add each effect linearly. Eq. (8-3) models the 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 (8-3) 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 re-form the three second-order differential equations of motion into six first-order differential equations because we can then use a broader class of integration methods. This first-order system is known as a variation of Cowell's formulation.

$$\bar{X} = \begin{bmatrix} \dot{\hat{r}} \\ \dot{\hat{v}} \end{bmatrix} \quad \dot{\bar{X}} = \begin{bmatrix} \dot{\hat{v}} \\ -\frac{\mu \hat{r}}{r^3} + \hat{a}_p \end{bmatrix}$$

8.5 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 numerical integration problems. We could spend considerable time discussing the numerous types of integrators and which is appropriate for a particular problem, so I've introduced only some of the more popular techniques. Sadler (1974) presents fundamental 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) method 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.

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. They are called *predictor-corrector* methods and they 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. The predictor-corrector methods are usually more accurate, but this improvement is achieved at the expense of increased complexity.

Step size techniques are employed with most numerical integration methods to limit error. **Fixed step-size** methods are most common for circular orbits because the step size is kept constant throughout the orbit. **Variable step-size** methods are better for eccentric orbits because they permit additional evaluations when the satellite is moving fastest, and fewer when the satellite is moving slowest. Variable step-size methods are more difficult to implement with multi-step techniques due to the use of back values, but it can be done in some situations.

Numerical integration methods are also classified by the equations they solve. **Single integration** techniques are designed for first-order differential equations. **Double integration** techniques solve second-order differential equations and produce direct answers. This is useful in astrodynamics [recall Eq. (8-3)], but when forces like atmospheric drag are included, the velocity is needed, requiring a single integration method.

It's also important to examine notation. For satellite calculations, we need $\ddot{\vec{r}}$ when $\dot{\vec{r}}$ is known. If we model non-conservative forces like drag, we also need $\ddot{\vec{r}}$. Therefore, I'll change the notation here from y to show the use of position vectors and their derivatives. Maury and Segal (1969) state that summed forms are a way to limit the accumulated error from the integrated variable. **Summed forms** sum the *back differences*, $\nabla_{i-1}(\ddot{\vec{r}}_n)$, to form *summed back differences*, $\tilde{\S}_n^{II}$, and $\tilde{\S}_n^{I}$ and they're only applicable for multi-step methods. The difference 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. Note the use of the binomial coefficient [Eq. (C-24)].

$$\begin{aligned}\nabla(\ddot{\vec{r}}_n) &= \ddot{\vec{r}}_n - \ddot{\vec{r}}_{n-1} & \nabla_2(\ddot{\vec{r}}_n) &= \ddot{\vec{r}}_n - 2\ddot{\vec{r}}_{n-1} + \ddot{\vec{r}}_{n-2} \\ \nabla_n(\ddot{\vec{r}}_n) &= \sum_{i=0}^j (-1)^i \binom{j}{i} \ddot{\vec{r}}_{n-i}\end{aligned}$$

Non-summed forms simply use the step-to-step back differences.

There are a few integrators that have gained popularity over the years and are included in existing orbit determination programs (See Table D-10 for a listing of programs). The Adams-Bashforth-Moulton integrator is widely used in these programs. The Gauss-Jackson integrator also has a long history of operational use. Berry and Healy (2001) present a detailed discussion of this method and the required equations. Herrick comments (1972) that 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. Burlisch-Stoer represents an alternative that has fewer calls to the force-model equations (Pohlen and Titus, 1995). It reduces computer runtime, but it appears to be unreliable as a general-purpose integrator. It works best in quadruple precision when FORTRAN is used. Shampine-Gordon (1975, and Berry and Healy, 2004) has received recent interest and could become a plausible alternative.

Most numerical integrators actually derive from the simplest form of numerical integration—the *Taylor series* integrator.* Recall the form of a Taylor series:

$$y(t) = y(t_o) + \dot{y}(t_o)(t - t_o) + \frac{\ddot{y}(t_o)(t - t_o)^2}{2!} + \frac{\dddot{y}(t_o)(t - t_o)^3}{3!} + \dots \quad (8-4)$$

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? A simple answer is the basis for an **Euler integrator** (after Leonhard Euler), which is simply a first-order Taylor series:

$$y(t) \equiv y(t_o) + \dot{y}(t_o)(t - t_o) \quad (8-5)$$

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 ($h = t - t_o$). 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. You've also assumed that you're operating in the near linear region.

Single-step: 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 also 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 method are

* Brook Taylor (1685–1731) published his famous theorem in 1715 in *Methodus incrementorum directa et inversa*.

$$\begin{aligned}
 \dot{y}_1 &= f(t_o, y_o) & \dot{y}_2 &= f\left(t_o + \frac{h}{2}, y_o + \frac{h}{2}\dot{y}_1\right) \\
 \dot{y}_3 &= f\left(t_o + \frac{h}{2}, y_o + \frac{h}{2}\dot{y}_2\right) & \dot{y}_4 &= f(t_o + h, y_o + h\dot{y}_3) \\
 y(t) &= y(t_o) + \frac{h}{6}(\dot{y}_1 + 2\dot{y}_2 + 2\dot{y}_3 + \dot{y}_4) + O(h^5)
 \end{aligned} \tag{8-6}$$

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,

$$f(t, y) = \begin{bmatrix} \dot{r} \\ \dot{v} \\ \ddot{r} \\ \ddot{v} \end{bmatrix} = \begin{bmatrix} \dot{r} \\ \dot{v} \\ \ddot{r} \\ \ddot{v} \end{bmatrix} \quad y(t_o) = \begin{bmatrix} r_o \\ v_o \end{bmatrix} \quad \therefore \dot{y}_1 = \begin{bmatrix} \dot{r} \\ \dot{v} \\ \ddot{r} \\ \ddot{v} \end{bmatrix}_{t=t_o} \tag{8-7}$$

The fourth-order terminology comes from the fact that our integrator is matched to a fourth-order Taylor series expansion about the initial value $y(t_o)$. 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. The fourth-order Runge-Kutta is really an Euler-type extrapolation using a weighted average of four rates (\dot{y}_i) in the interval $[t_o, t_o + h]$. Figure 8-2 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, we may distinguish several *orders* of numerical-integration routines: the order to which the method is locally or globally correct, or the order of its local or global error. Conventionally, a method is termed fourth-order if it's locally accurate to fourth-order, globally correct to third-order, has fifth-order local error and fourth-order global error (Berry, 2004).

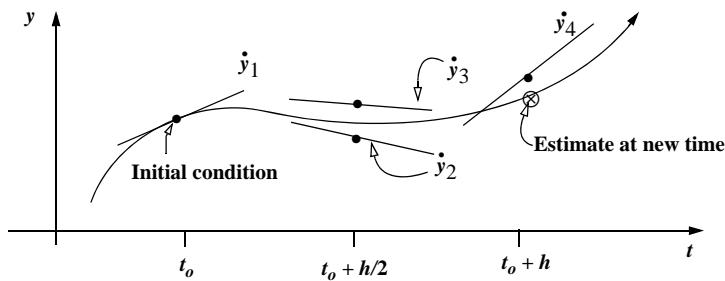


Figure 8-2. The Runge-Kutta Method. This algorithm works by evaluating the derivative (slope) at several different points (four for a fourth-order method) along the estimated trajectory. It's limited because all calculations depend on estimates of the function value. The intermediate estimates aid in 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):

$$\begin{aligned} y(t) &= y_o + h \sum_{i=0}^{j-1} b_i \dot{y}_i + O(h^{n+1}) \quad n = 0, 1, 2, \dots, n-1 \\ \dot{y}_o &= f(t_o, y_o) \quad c_{io} = p_i - \sum_{j=1}^{i-1} c_{ij} \\ \dot{y}_i &= f\left(t_n + p_i h, y_n + h \sum_{j=0}^{i-1} c_{ij} \dot{y}_j\right) \quad i = 1, 2, \dots, j-1 \end{aligned}$$

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 numerical integrations, of different order, 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 to maintain an even difference in the evaluations throughout the orbit. 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. Another 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 (for example time-regularized Cowell on page 537). We use Eq. (8-7) as the first-order equations of motion for satellite problems. The Runge-Kutta-Fehlberg algorithm employs the following values:

$$\begin{aligned} k_1 &= hf(t_o, y_o) \\ k_2 &= hf\left(t_o + \frac{1}{4}h, y_o + \frac{1}{4}k_1\right) \\ k_3 &= hf\left(t_o + \frac{3}{8}h, y_o + \frac{3}{32}k_1 + \frac{9}{32}k_2\right) \\ k_4 &= hf\left(t_o + \frac{12}{13}h, y_o + \frac{1932}{2197}k_1 - \frac{7200}{2197}k_2 + \frac{7296}{2197}k_3\right) \end{aligned}$$

$$k_5 = hf\left(t_o + h, y_o + \frac{439}{216}k_1 - 8k_2 + \frac{3680}{513}k_3 - \frac{845}{4104}k_4\right)$$

$$k_6 = hf\left(t_o + \frac{1}{2}h, y_o - \frac{8}{27}k_1 + 2k_2 - \frac{3544}{2565}k_3 + \frac{1859}{4104}k_4 - \frac{11}{40}k_5\right)$$

The two approximations use these values with a fourth-order and fifth-order Runge-Kutta algorithm. Their difference is

$$\Delta_{45} = \frac{1}{360}k_1 - \frac{128}{4275}k_3 - \frac{2197}{75240}k_4 + \frac{1}{50}k_5 + \frac{2}{55}k_6$$

If the absolute value of the delta 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 .

$$s \equiv 0.8408 \left[\frac{1 \times 10^{-8}h}{\Delta_{45}} \right]^{1/4}$$

Using predetermined minimum, $h_{min} = h/64$, and maximum, $h_{max} = h \cdot 64$, 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,

$$y = y_o + \frac{25}{216}k_1 + \frac{1408}{2565}k_3 + \frac{2197}{4104}k_4 - \frac{1}{5}k_5$$

Multi-step, Single Integration Techniques

The most popular methods here are the **Adams-Bashforth-Moulton** and **Shampine-Gordon**. Berry (2004) discusses these methods in detail, particularly Shampine-Gordon, which uses a divided difference (a similar operation to the back differences) to achieve a variable step-size. Adams-Bashforth-Moulton is well established in existing programs. It's a multi-step, fixed step-size method that 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 often require two evaluations per step. Because these methods require back values, many aren't self-starting. Berry and Healy (2001) show that you can start multi-step methods with iterative procedures that use the same method with the formulae shifted to correct the backpoints. You can also 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-Bashforth-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 and 245-302) and Sadler (1974) describe this traditional approach. Montenbruck and Pfleger (1994:90–91) show how to derive multi-step methods, specifically Adams-Bashforth’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,

$$\begin{aligned} p(t) &= \alpha_0 + \alpha_1 \tau + \alpha_2 \tau^2 + \alpha_3 \tau^3 & \tau &= \frac{\Delta t}{h} \\ \alpha_0 &= f(y(t_n)) = f_n & \alpha_1 &= \frac{1}{6}(-2f_{n-3} + 9f_{n-2} - 18f_{n-1} + 11f_n) \\ \alpha_2 &= \frac{1}{6}(-3f_{n-3} + 12f_{n-2} - 15f_{n-1} + 6f_n) & \alpha_3 &= \frac{1}{6}(-f_{n-3} + 3f_{n-2} - 3f_{n-1} + f_n) \end{aligned}$$

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-Bashforth 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 *ordinate* notations shortly but note that while back differences are also possible, ordinate forms are more popular. 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 values and repeat the correction process if necessary. As we increase the degree of the polynomial, accuracy usually rises. However, complexity also rises, and certain orders can produce unstable results. Eighth- and twelfth-degree 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-Bashforth-Moulton integrators.

Maury and Segal (1969) show how to determine the initial and recurring values of the summed back differences at each step.

$$\hat{S}_n^I = \hat{S}_{n-1}^I + \ddot{\hat{r}}_n \quad \hat{S}_n^{II} = \hat{S}_{n-1}^{II} + \hat{S}_{n-1}^I + \ddot{\hat{r}}_n = \hat{S}_{n-1}^{II} + \hat{S}_n^I$$

The initial values are found from the corrector relations (Adams-Moulton and Cowell), and the previous function evaluations.

$$\begin{aligned}\hat{S}_{n-1}^I &= \frac{1}{h} \ddot{\hat{r}}_{n-1} - \left\{ \frac{1}{2} \ddot{\hat{r}}_{n-1} + \alpha_{AM2} \nabla(\ddot{\hat{r}}_{n-1}) + \alpha_{AM3} \nabla_2(\ddot{\hat{r}}_{n-1}) + \alpha_{AM4} \nabla_3(\ddot{\hat{r}}_{n-1}) + \dots \right\} \\ \hat{S}_{n-1}^{II} &= \frac{1}{h} \ddot{\hat{r}}_{n-1} - \left\{ \frac{1}{12} \ddot{\hat{r}}_{n-1} + \alpha_{C2} \nabla(\ddot{\hat{r}}_{n-1}) + \alpha_{C3} \nabla_2(\ddot{\hat{r}}_{n-1}) + \alpha_{C4} \nabla_3(\ddot{\hat{r}}_{n-1}) + \dots \right\}\end{aligned}$$

Notice that this process will yield large numbers of values for storage, 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)

$$\dot{\hat{r}}_{n+1}^p = h \left\{ \alpha_{ABo} \hat{S}_n^I + \sum_{i=0}^j \beta_{ABi} \ddot{\hat{r}}_{n-i} \right\}$$

CORRECT - ADAMS-MOULTON SUMMED-ORDINATE FORM (order j)

$$\dot{\hat{r}}_{n+1} = h \left\{ \alpha_{ABo} \hat{S}_{n+1}^I + \sum_{i=0}^j \beta_{AMi} \dot{\hat{r}}_{n+1-i} \right\}$$

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 $\dot{\hat{r}}_{n+1}^p$ using those two results, $\dot{\hat{r}}_{n+1}$ and $\ddot{\hat{r}}_{n+1}$. This predicted value is used when $j = 0$ in the corrector phase. Varying the step-size is more difficult than with single-step methods because equal spacing is implicit in the formulation. This difficulty

* Berry and Healy (2001) show how to derive the ordinate forms by expressing the back difference operator in terms of binomial coefficients of the function points. This lets us write the back difference operator "in terms of the values of the functions."

can be overcome with simple interpolation routines and isn't difficult whenever we have to double the step size (Matthews, 1987:346–437).

Multi-step, Double Integration Techniques

I'll introduce three methods to implement this form of integration—a General Formulation, Störmer-Cowell, and Gauss-Jackson. These methods are sometimes called by different names so you should clearly identify which method is under consideration. I use the predictor-corrector nomenclature presented by Maury and Segal (1969), thus Störmer-Cowell implies a *Störmer predictor*, and a *Cowell corrector*. These techniques differ from single integration methods because they are designed especially for second-order systems. Maury and Segal (1969) present a detailed summary including tables of coefficients. Berry and Healy (2001) and Berry (2004) also give an excellent and expanded discussion. Herrick (1972:12–17, 260–275), Long et al. (1989:6-2 to 6-20), and Baker (1967:452–453) also introduce the formulations. Remember that the formulation is for second order systems—if we include drag, we must compute the first integral, or use a single integration method because the equations of motion depend on position and velocity vectors.

The same forms exist for these routines as before—non-summed and summed, and ordinate, back difference, and divided difference. Krogh's integrator (Krogh, 1974) uses the ***General Formulation*** in divided difference form. It includes a velocity term so it may not be a “true” double integration technique—also possibly why it is not as well known. It's a non-summed, variable step-size method that's self starting, although fixed step-size versions in back difference or ordinate forms are also possible. ***Störmer-Cowell*** is another non-summed method, but it has no velocity term. It can be written as a fixed step-size in back difference or ordinate form, and as a variable step-size in divided difference (Berry, 2004). ***Gauss-Jackson*** is simply the fixed step-size, summed form of Störmer-Cowell. There is no variable step-size divided difference form.

We usually prefer the summed-ordinate formulation for improved round-off error and ease of programming. The Gauss-Jackson method uses first and second, summed back-differences shown in the Adams-Bashforth-Moulton discussion. We use the predicted state in the corrector for $j = 0$.

PREDICT - GAUSS SUMMED-ORDINATE FORM (order j)

$$\hat{\vec{r}}_{n+1}^p = h \left\{ \alpha_{So} \hat{\vec{S}}_n^{II} + \alpha_{S1} \hat{\vec{S}}_n^{I} + \sum_{i=0}^j \beta_{Si} \ddot{\vec{r}}_{n-i} \right\}$$

CORRECT - JACKSON SUMMED-ORDINATE FORM (order j)

$$\hat{\vec{r}}_{n+1} = h \left\{ \alpha_{So} \hat{\vec{S}}_{n+1}^{II} + \{ \alpha_{So} + \alpha_{S1} \} \hat{\vec{S}}_{n+1}^{I} + \sum_{i=0}^j \beta_{Ci} \ddot{\vec{r}}_{n+1-i} \right\}$$

Table 8-1 lists coefficients needed to form different orders of numerical integrators. Consult Maury and Segal (1969) and Berry and Healy (2001) for complete tables.

TABLE 8-1. Numerical Integration Coefficients. I've listed sample coefficients for the Adams-Bashforth-Moulton (AB and AM), Störmer-Cowell (S and C), and Gauss-Jackson (G and J) integrators. The common denominators, D, assist computer implementations 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)

α_{ABi}	Non-summed Forms				Summed Ordinate Forms			
	α_{AMi}	α_{Si}	α_{Ci}		β_{ABi}	β_{AMi}	β_{Gi}	β_{Ji}
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				

8.5.1 Implementing an Integrator and Determining Step Size

The advantages of a numerical approach may seem to exclude other solution techniques, but numerical analysis doesn't reveal as much of the problem's qualitative behavior. Also remember the inherent difficulties of computer runtime, and perhaps less discussed, the choice of integrator and step-size. The *type* of orbit integrator can drastically affect performance and precision.

The first-order equations in Eq. (8-7) 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. 8.7.2. We'll need both the single and double integration 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.

Finding initial conditions for multi-step methods can be difficult. However, we can often 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. Another approach is to use a lower order integrator, but to also use a smaller step size (Maury and Segal, 1969). The important feature is for the error of the starting method and the multi-step method to be the same.

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. 8.2 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-size 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 precision 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.

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.

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 accelerations.

Variable step sizes are often used for highly elliptical or "difficult" orbits. Although they're accurate, they're not usually very efficient for most orbital problems. 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 can 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 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. 8.8.1.

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 a variety of 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 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**, also called ***s*-integration**,^{*} 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. From Eq. (2-17) and Eq. (9-17), we form

$$ndt = \frac{r}{a} dE = \frac{r^2}{a\sqrt{1-e^2}} d\nu \quad (8-8)$$

Selecting the best alternate variable to use depends on the size, nature, and frequency of the perturbation. One approach adopts the following time transformation [actually a **generalized Sundman transformation** because the index n is not 1 as in Eq. (2-40)].

$$dt = cr^n ds \quad (8-9)$$

Berry and Healy (2002 and 2003b) point out that if the constant $n = 1$ and $c = \sqrt{a/\mu}$ (as Sundman used), s is the eccentric anomaly, and that if $n = 2$ and $c = \sqrt{\mu a(1-e^2)}$, s is the true anomaly. Merson (1975) says that to evenly distribute the integration error throughout the orbit, $n = 1.5$ and $c = 1/\sqrt{\mu}$.

The 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-order differential equation [Eq. (8-9)]. 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.

* Some sources distinguish ***t*-integration** to refer to the [common] situation where time is the independent variable.

8.6 Disturbing Forces

To numerically integrate Cowell's formulation, we must have mathematical models for each perturbing force. 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.

8.6.1 Gravity Field of a Central Body

Given that the gradient of the potential for a spherical central body will yield the acceleration, we must examine how to form a potential function that includes the perturbing accelerations due to a nonspherical central body. The perturbing forces are mathematically described by an *aspherical-potential function*, U , which we'll develop following the methods of Roy (1988:171–177), Kaplan (1976:273–282), and Fitzpatrick (1970:288–308).

Deriving the Aspherical-Potential Function

Because the aspherical-potential function provides the means to determine the gravitational attraction on a satellite, it's logical to assume (from Newton) 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 at point P 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 8-3 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 infinitesimal element of mass, dm_{\oplus} , is

$$dU = G \frac{dm_{\oplus}}{\rho_Q}$$

In reality, a large but finite number of very small masses are within a body of fixed size; thus, the summation of each mass approaches an integral, and integrating over the entire body gives us the total potential.

$$U = G \int_{body} \frac{1}{\rho_Q} dm_{\oplus}$$

The geocentric distance (r) of the point (P) where we're evaluating the potential, and the distance from an infinitesimal mass, dm_{\oplus} , to the center of the Earth, r_Q , are

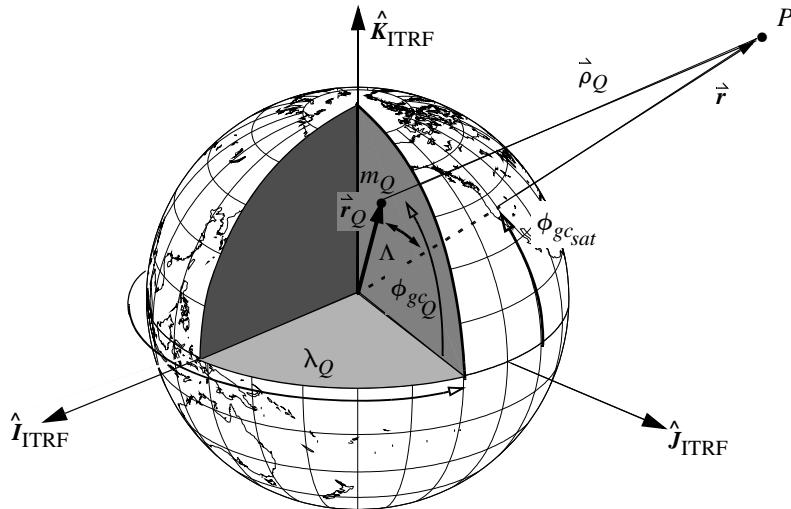


Figure 8-3. Deriving the Gravitational Potential. We consider each point in the Earth, m_Q , when determining the gravitational potential at P . The latitude is a geocentric value. Λ is the angle between the vectors r_Q and r .

$$r = \sqrt{x^2 + y^2 + z^2} \quad r_Q = \sqrt{\xi^2 + \eta^2 + \zeta^2}$$

Using the law of cosines [Eq. (C-21)], the slant range, ρ_Q , from an infinitesimal mass to the external point is

$$\rho_Q^2 = r^2 + r_Q^2 - 2rr_Q \cos(\Lambda)$$

We determine the angle between the two position vectors, Λ , by using the dot product:

$$\cos(\Lambda) = \frac{x\xi + y\eta + z\zeta}{rr_Q} = \frac{\hat{r} \cdot \hat{r}_Q}{rr_Q} \quad (8-10)$$

The slant range is now

$$\begin{aligned} \rho_Q &= r \sqrt{1 - 2\frac{r_Q}{r} \cos(\Lambda) + \left(\frac{r_Q}{r}\right)^2} = r \sqrt{1 - 2\alpha \cos(\Lambda) + \alpha^2} \\ \alpha &= \frac{r_Q}{r} \end{aligned}$$

The potential (of the body) then becomes

$$U = G \int_{body} \frac{dm_{\oplus}}{r \sqrt{1 - 2\alpha \cos(\Lambda) + \alpha^2}}$$

The 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:

$$\frac{1}{\sqrt{1 - 2\alpha \cos(\Lambda) + \alpha^2}} = \frac{1}{\sqrt{1 + x}} = \sum_{t=0} \alpha^t P_t[\cos(\Lambda)]$$

This expression is a series of *Legendre polynomials*, P_t . I'll use brackets to distinguish the argument of the Legendre polynomials, $\cos(\Lambda)$. Setting $\gamma = \cos(\Lambda)$, a form of Rodriguez's formula gives the *conventional* Legendre polynomials (Lambeck, 1988:10),

$$\begin{aligned} P_t[\gamma] &= \frac{1}{2^t t!} \frac{d^t (\gamma^2 - 1)^t}{d\gamma^t} \\ P_t[\gamma] &= \frac{1}{2^t} \sum_{j=0}^{\frac{t}{2}} \frac{(-1)^j (2t-2j)!}{j!(t-j)!(t-2j)!} \gamma^{t-2j} \end{aligned} \quad (8-11)$$

where $d^t() / d\gamma^t$ is the t^{th} order derivative. The first six Legendre polynomials are

$$\begin{array}{ll} P_0[\gamma] = 1 & P_3[\gamma] = \frac{1}{2}(5\gamma^3 - 3\gamma) \\ P_1[\gamma] = \gamma & P_4[\gamma] = \frac{1}{8}(35\gamma^4 - 30\gamma^2 + 3) \\ P_2[\gamma] = \frac{1}{2}(3\gamma^2 - 1) & P_5[\gamma] = \frac{1}{8}(63\gamma^5 - 70\gamma^3 + 15\gamma) \end{array}$$

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. (8-57).

For now, let's write the potential as

$$U = \frac{G}{r} \int_{\text{body}} \sum_{t=0} \alpha^t P_t[\cos(\Lambda)] dm_{\oplus} \quad (8-12)$$

To continue the analysis we can use two 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 moments 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. (8-12) 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 assume the central body is spherically symmetric and homogeneous, the potential reduces to the above expression.

The second term is slightly more difficult. If we use the Legendre polynomial terms and let $l = 1$ in Eq. (8-12), we can use Eq. (8-10) to find

$$\begin{aligned} U_1 &= \frac{G}{r} \int \cos(\Lambda) \alpha dm_{\oplus} = \frac{G}{r} \int \frac{x\xi + y\eta + z\zeta}{r^2} dm_{\oplus} \\ &= \frac{G}{r^3} \left(x \int \xi dm_{\oplus} + y \int \eta dm_{\oplus} + z \int \zeta dm_{\oplus} \right) \end{aligned}$$

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,

$$\int \xi dm_{\oplus} = \int \eta dm_{\oplus} = \int \zeta dm_{\oplus} = 0$$

$U_1 = 0$ because we chose the origin at the center of mass. In fact, the expressions above are a definition of the center of mass.

The third term in the expansion is

$$U_2 = \frac{1}{2r^3} \int (3\gamma^2 - 1) r_Q^2 dm_{\oplus}$$

Remembering that $\gamma^2 = \cos^2(\Lambda)$, we can substitute using Eq. (C-10).

$$U_2 = \frac{G}{2r^3} \int 2r_Q^2 dm_{\oplus} - \frac{G}{2r^3} \int 3r_Q^2 \sin^2(\Lambda) dm_{\oplus}$$

The left-hand term represents the *moments of inertia* about the three coordinate axes:

$$\left. \begin{aligned} A &\equiv \int (\eta^2 + \zeta^2) dm_{\oplus} \\ B &\equiv \int (\xi^2 + \zeta^2) dm_{\oplus} \\ C &\equiv \int (\xi^2 + \eta^2) dm_{\oplus} \end{aligned} \right\} \quad A + B + C = \int 2r_Q^2 dm_{\oplus}$$

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

$$U = \frac{Gm_{\oplus}}{r} + \frac{G}{2r^3}(A + B + C - 3I) + \dots \quad (8-13)$$

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. (8-12). The potential in the form of Eq. (8-12) isn't terribly useful because we can't directly find the angle, Λ .^{*} Spherical trigonometry allows us to develop an equation for this angle. Remember that the satellite's latitude [from Eq. (3-36)], will always be a *geocentric value* ($\phi_{gc_{sat}}$). The cosine law of spherical trigonometry yields

$$\begin{aligned} \cos(\Lambda) &= \cos(90^\circ - \phi_{gc_Q}) \cos(90^\circ - \phi_{gc_{sat}}) \\ &\quad + \sin(90^\circ - \phi_{gc_Q}) \sin(90^\circ - \phi_{gc_{sat}}) \cos(\lambda_Q - \lambda_{sat}) \end{aligned}$$

Reduction gives us

$$\cos(\Lambda) = \sin(\phi_{gc_Q}) \sin(\phi_{gc_{sat}}) + \cos(\phi_{gc_Q}) \cos(\phi_{gc_{sat}}) \cos(\lambda_Q - \lambda_{sat}) \quad (8-14)$$

The *addition theorem* (also called the *decomposition formula*) of spherical harmonics provides a way to expand the expressions for the Λ into Eq. (8-12) (Hobson, [1931] 1965:143):

$$\begin{aligned} P_l[\cos(\Lambda)] &= P_l[\sin(\phi_{gc_Q})] P_l[\sin(\phi_{gc_{sat}})] \\ &\quad + 2 \sum_{m=1}^l \frac{(l-m)!}{(l+m)!} \{ A_{l,m} A'_{l,m} + B_{l,m} B'_{l,m} \} \end{aligned} \quad (8-15)$$

$$\begin{aligned} A_{l,m} &= P_{l,m}[\sin(\phi_{gc_Q})] \cos(m\lambda_Q) & A'_{l,m} &= P_{l,m}[\sin(\phi_{gc_{sat}})] \cos(m\lambda_{sat}) \\ B_{l,m} &= P_{l,m}[\sin(\phi_{gc_Q})] \sin(m\lambda_Q) & B'_{l,m} &= P_{l,m}[\sin(\phi_{gc_{sat}})] \sin(m\lambda_{sat}) \end{aligned}$$

* This angle is often used in Chap. 11, so we'll call it the ground range, or the total range angle, Λ .

where the summation introduces “ l ” and “ m ” indices as *degree* and *order*, respectively.

Now, let's introduce the *associated* Legendre functions, $P_{l,m}$. 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

$$P_{l,m}[\gamma] = (1 - \gamma^2)^{m/2} \frac{d^m}{d\gamma^m} P_l[\gamma] \quad (8-16)$$

Table 8-2 shows a few examples. 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, which I'll discuss in Sec. 8.7.2.

TABLE 8-2. Associated Legendre Functions. This table gives a few sample expansions for the associated Legendre function, $P_{l,m}[\sin(\phi_{gc_{sat}})]$. For astrodynamics, we use the geocentric latitude. (Lambeck, 1988)

$P_{0,0}$	1	$P_{3,2}$	$15\cos^2(\phi_{gc_{sat}})\sin(\phi_{gc_{sat}})$
$P_{1,0}$	$\sin(\phi_{gc_{sat}})$	$P_{3,3}$	$15\cos^3(\phi_{gc_{sat}})$
$P_{1,1}$	$\cos(\phi_{gc_{sat}})$	$P_{4,0}$	$\frac{1}{8}\{35\sin^4(\phi_{gc_{sat}}) - 30\sin^2(\phi_{gc_{sat}}) + 3\}$
$P_{2,0}$	$\frac{1}{2}\{3\sin^2(\phi_{gc_{sat}}) - 1\}$	$P_{4,1}$	$\frac{5}{2}\cos(\phi_{gc_{sat}})\{7\sin^3(\phi_{gc_{sat}}) - 3\sin(\phi_{gc_{sat}})\}$
$P_{2,1}$	$3\sin(\phi_{gc_{sat}})\cos(\phi_{gc_{sat}})$	$P_{4,2}$	$\frac{15}{2}\cos^2(\phi_{gc_{sat}})\{7\sin^2(\phi_{gc_{sat}}) - 1\}$
$P_{2,2}$	$3\cos^2(\phi_{gc_{sat}})$	$P_{4,3}$	$105\cos^3(\phi_{gc_{sat}})\sin(\phi_{gc_{sat}})$
$P_{3,0}$	$\frac{1}{2}\{5\sin^3(\phi_{gc_{sat}}) - 3\sin(\phi_{gc_{sat}})\}$	$P_{4,4}$	$105\cos^4(\phi_{gc_{sat}})$
$P_{3,1}$	$\frac{1}{2}\cos(\phi_{gc_{sat}})\{15\sin^2(\phi_{gc_{sat}}) - 3\}$		

Now we make an important substitution to Eq. (8-15). By separating all the terms that are independent of the satellite's location in Eq. (8-12), we can arrive at a solution that isolates terms which depend only on the central body and those which relate the satellite's position. We define variables in Eq. (8-17).

$$\begin{aligned} C'_{l,m} &= \int_{\text{body}} r_Q^l \frac{(l-m)!}{(l+m)!} P_{l,m}[\sin(\phi_{gc_Q})] \cos(m\lambda_Q) dm_{\oplus} \\ S'_{l,m} &= \int_{\text{body}} r_Q^l \frac{(l-m)!}{(l+m)!} P_{l,m}[\sin(\phi_{gc_Q})] \sin(m\lambda_Q) dm_{\oplus} \end{aligned} \quad (8-17)$$

The coefficients in Eq. (8-17) represent the mathematical modeling for the Earth's shape using spherical harmonics. The special case for the *zonal harmonics* is

$$C'_{l,0} = \int_{\text{body}} r_Q^l P_l[\sin(\phi_{gc_Q})] dm_{\oplus}$$

which uses the conventional Legendre polynomials. Note that $S'_{l,0}$ is zero. When we substitute these values into the potential in Eq. (8-12),

$$\begin{aligned} U &= \frac{G}{r} \sum_{l=0}^{\infty} \frac{P_l[\sin(\phi_{gc_{sat}})]}{r^l} C'_{l,0} \\ &\quad + \frac{G}{r} \sum_{l=1}^{\infty} \sum_{m=1}^l \frac{P_{l,m}[\sin(\phi_{gc_{sat}})]}{r^l} \left\{ C'_{l,m} \cos(m\lambda_{sat}) + S'_{l,m} \sin(m\lambda_{sat}) \right\} \end{aligned}$$

Finally, a unit analysis suggests we remove units from the result. This leads to the C and S coefficients. We can non-dimensionalize the gravitational coefficients (Lambeck, 1988:11):

$$C'_{l,m} = C_{l,m} R_{\oplus}^l m_{\oplus} \quad S'_{l,m} = S_{l,m} R_{\oplus}^l m_{\oplus}$$

R_{\oplus} and m_{\oplus} are the Earth's mean equatorial radius and mass, respectively.

Using the definition of the gravitational parameter, find the gravitational potential as

$$\begin{aligned} U &= \frac{\mu}{r} \sum_{l=0}^{\infty} P_l[\sin(\phi_{gc_{sat}})] \left(\frac{R_{\oplus}}{r} \right)^l C_{l,0} \\ &\quad + \frac{\mu}{r} \sum_{l=1}^{\infty} \sum_{m=1}^l P_{l,m}[\sin(\phi_{gc_{sat}})] \left(\frac{R_{\oplus}}{r} \right)^l \left\{ C_{l,m} \cos(m\lambda_{sat}) + S_{l,m} \sin(m\lambda_{sat}) \right\} \end{aligned} \quad (8-18)$$

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 Lambeck (1988:11), Kaplan (1976:281), and others. First, a double summation encompasses the associated Legendre polynomials, $P_{l,m}$, and the harmonic coefficients, and $S_{l,0} = 0$.

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

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 $C_{1,0}$, $C_{1,1}$, and $S_{1,1}$ are all zero ($S_{1,0}$ is also zero by definition). This result corresponds to $l=1$ in MacCullagh's approach. It 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 terms are zero.

$$U = \frac{\mu}{r} \left[1 + \sum_{l=2}^{\infty} \sum_{m=0}^{l} \left(\frac{R_{\oplus}}{r}\right)^l P_{l,m}[\sin(\phi_{gc_{sat}})] \left\{ C_{l,m} \cos(m\lambda_{sat}) + S_{l,m} \sin(m\lambda_{sat}) \right\} \right] \quad (8-19)$$

This expression describes the gravitational attraction resulting from the irregular distribution of the Earth's mass using a potential function.

Some authors use a "J" notation for the zonal harmonics. The correct use of the "J" notation is that it's the negative of the C notation for zonal harmonics.

$$-C_{l,0} = J_l \quad (8-20)$$

Though this is merely a sign change, it can be very significant in forming the perturbation equations discussed in Sec. 9.6.

We can also separate the zonal and tesseral terms from Eq. (8-19) by recalling that $J_l = -C_{l,0}$ [Eq. (8-20)]. Thus,

$$U = \frac{\mu}{r} \left[1 - \sum_{l=2}^{\infty} J_l \left(\frac{R_{\oplus}}{r}\right)^l P_l[\sin(\phi_{gc_{sat}})] + \sum_{l=2}^{\infty} \sum_{m=1}^{l} \left(\frac{R_{\oplus}}{r}\right)^l P_{l,m}[\sin(\phi_{gc_{sat}})] \left\{ C_{l,m} \cos(m\lambda_{sat}) + S_{l,m} \sin(m\lambda_{sat}) \right\} \right] \quad (8-21)$$

This form has several important features. The *conventional* Legendre polynomials are merely special cases of the associated Legendre functions. Eq. (8-21) also depends on the C and S coefficients. Fortunately, the analysis of satellite motion allows us to empirically determine the C and S coefficients from observations. This is a common method in developing gravitational fields today.

Older computers sometimes used 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. (8-17). Computer trun-

tions may introduce errors so it's common to normalize in order to correct this deficiency. The following formula permits these transformations. The overall transformation, $\Pi_{l,m}$, is often used to abbreviate the relation. The overbar denotes normalized variables. Due to the size of the coefficients, $\Pi_{l,m}$ is usually determined with the maximum precision available on the computer. Don't confuse this process with non-dimensionalizing described earlier.

$$\begin{aligned}\Pi_{l,m} &= \sqrt{\frac{(l+m)!}{(l-m)!k(2l+1)}} \\ \bar{S}_{l,m} &= \Pi_{l,m} S_{l,m} \quad \bar{C}_{l,m} = \Pi_{l,m} C_{l,m} \quad \bar{P}_{l,m} = \frac{P_{l,m}}{\Pi_{l,m}} \\ k &= 1 \text{ if } m = 0 \\ k &= 2 \text{ if } m \neq 0\end{aligned}\quad (8-22)$$

When using normalized coefficients, you must also normalize the Legendre polynomials (conventional and associated) because the product must remain the same. Thus,

$$\bar{C}_{l,m} \bar{P}_{l,m} = C_{l,m} P_{l,m} \quad \bar{S}_{l,m} \bar{P}_{l,m} = S_{l,m} P_{l,m}$$

Note that most modern floating point computers don't require normalized values and can accurately process unnormalized coefficients. Normalizing requires an extra step in each calculation that *can* be eliminated. Also, the summation should always begin with the smallest coefficient, proceeding to the largest (as in the reduction calculations in Ch. 3) to preserve accuracy. An example will show how normalizing works.

▼ Example 8-3. Normalizing Gravitational Coefficients.

GIVEN: Gravitational coefficients in Table D-1 and Table D-2.

FIND: Normalized values for (4,0), (41,1), (70,65).

For $l = 4$ and $m = 0$,

$$\begin{aligned}\bar{C}_{4,0} &= \sqrt{\frac{(l+m)!}{(l-m)!k(2l+1)}} C_{4,0} = \sqrt{\frac{(4+0)!}{(4-0)!1(2(4)+1)}} 1.619\ 621\ 591 \times 10^{-6} \\ &= 5.398\ 738\ 638 \times 10^{-7}\end{aligned}$$

Notice that $k = 1$ and S don't exist for zonal harmonics ($m = 0$). For $l = 41$ and $m = 1$,

$$\begin{aligned}\bar{C}_{41,1} &= \sqrt{\frac{(l+m)!}{(l-m)!k(2l+1)}} C_{41,1} = \sqrt{\frac{(41+1)!}{(41-1)!2(2(41)+1)}} (-1.522\ 218\ 288 \times 10^{-9}) \\ &= -4.902\ 746\ 800 \times 10^{-9} \\ \bar{S}_{41,1} &= \sqrt{\frac{(l+m)!}{(l-m)!k(2l+1)}} S_{41,1} = \sqrt{\frac{(41+1)!}{(41-1)!2(2(41)+1)}} (-1.283\ 602\ 318 \times 10^{-9}) \\ &= -4.134\ 214\ 655 \times 10^{-9}\end{aligned}$$

Notice now that $k = 2$ because m isn't zero. For $l = 70$ and $m = 65$,

$$\begin{aligned}\bar{C}_{l,m} &= \sqrt{\frac{(l+m)!}{(l-m)!k(2l+1)}} C_{l,m} = \sqrt{\frac{(70+65)!}{(70-65)!2(2(70)+1)}} 1.431\ 238\ 529 \times 10^{-122} \\ &= 1.019\ 277\ 424 \times 10^{-9} \\ \bar{S}_{l,m} &= \sqrt{\frac{(l+m)!}{(l-m)!k(2l+1)}} S_{l,m} = \sqrt{\frac{(70+65)!}{(70-65)!2(2(70)+1)}} (1.951\ 368\ 590 \times 10^{-122}) \\ &= 1.775\ 623\ 037 \times 10^{-9}\end{aligned}$$

▲ Notice the coefficients' similar orders of magnitude. The normalization process also benefits the terms of higher degree and order.

We can identify the term which gives us an idea of the Earth's overall nonsphericity. Specifically, the ***normalized, second-degree, zonal gravitational coefficient***, $\bar{C}_{2,0}$, represents the Earth's equatorial bulge. Eq. (8-22) discusses normalization. The remaining parameters help represent the geopotential due to the central body's specific shape.

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

This value doesn't include the permanent tidal deformation (Sec. 8.6.5 and Sec. 9.5.3). The conventional (un-normalized) form of this equation uses the following relation:

$$C_{l,0} = \bar{C}_{l,0} \sqrt{2l+1} \quad (8-23)$$

This is simply a special case of general formulas. Thus the first zonal coefficient is (with a second-order zonal harmonic of $J_2 = 0.001\ 082\ 626\ 9$)

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

Spherical Harmonics

The trigonometric argument of the Legendre polynomials in Eq. (8-19) 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 states 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, sectoral, 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 $P_l[\sin(\phi_{gcsat})]$, there are l circles of latitude along which P_l is zero, and hence $(l+1)$ zones in which the function is alternately increasing (shaded) and decreasing (white). J_2 is by far the strongest perturbation due to the Earth's shape. As the table on the inside back 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 8-4 shows several depictions of the zonal harmonics.

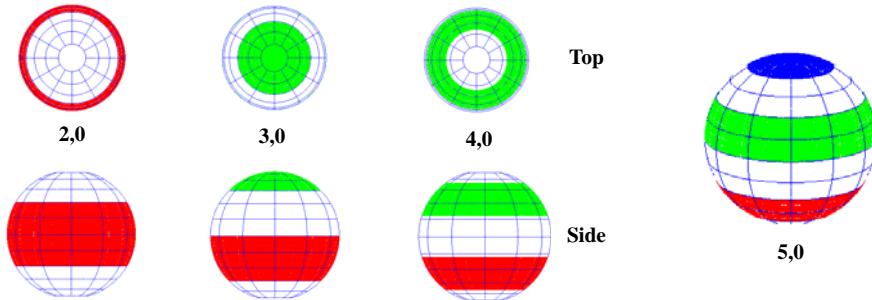


Figure 8-4. Zonal Harmonics. J_2 accounts for most of the Earth's gravitational departure from a perfect sphere. This band (and others) reflects the Earth's oblateness. The shading indicates regions of additional mass and the numbers link regions between the views. The third harmonic appears similar to the J_2 from the top but is reversed for the bottom view. J_5 is shown in a perspective view.

Sectoral harmonics occur when $l = m$ and represent bands of longitude. Figure 8-5 shows some sectoral harmonics. The polynomials $P_{l,l}[\sin(\phi_{gcsat})]$ are zero only at the poles ($\phi_{gcsat} = \pm 90^\circ$). On the other hand, the terms [$\cos(l\lambda)$ and $\sin(l\lambda)$] are zero for $2l$ different values of λ . Hence the lines along which the functions [$\cos(l\lambda)$ and $\sin(l\lambda)$] times $P_{l,l}[\sin(\phi_{gcsat})]$ vanish are meridians of longitude, which divide the sphere into $2l$ “orange-slice” sectors. Each sector represents l positive (shaded) and l negative (white) mass concentrations.

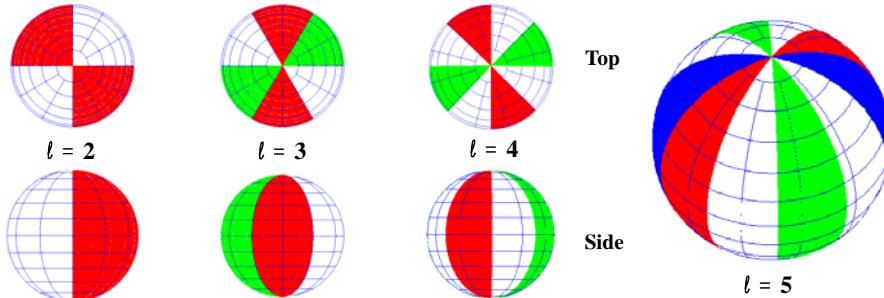


Figure 8-5. Sectoral Harmonics. Sectoral harmonics take into account the extra mass distribution in longitudinal regions.

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 $P_{l,m}[\sin(\phi_{gcsat})]$ are referred to as **tes-**

tesseral harmonics (because they look like tiles), for the sphere is divided up into a checkeredboard array. The number of circles of latitude along which $P_{l,m}[\sin(\phi_{gcsat})]$ is zero is equal to $(l - m)$, whereas the terms $[\cos(m\lambda)$ and $\sin(m\lambda)]$ vanish along $2m$ meridians of longitude. These zero lines represent the center of the latitude/longitude bands. Figure 8-6 shows the tesseral harmonics. Notice the similarities to, and differences from, the sectoral harmonics. Some texts refer to the sectorals as a subset of the tesserals. That's correct, but we can also keep them separate to distinguish their physical differences, as I do in this book.

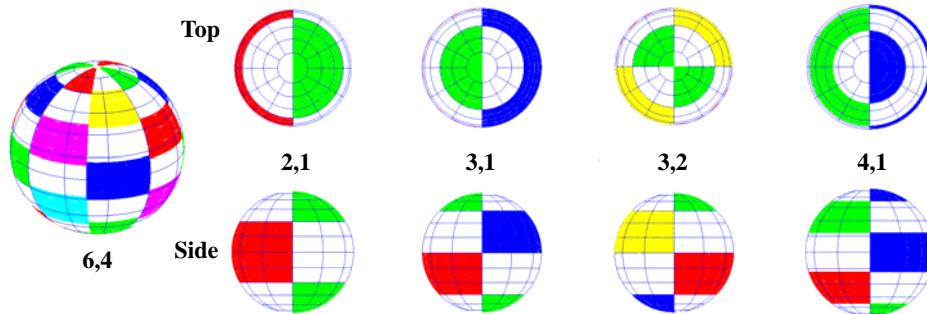


Figure 8-6. Tesseral Harmonics. Tesseral harmonics attempt to model specific regions on the Earth which depart from a perfect sphere. Depicting these and tying to the gravitational coefficients can be difficult.

Partial Derivatives

Now that we've developed a potential function for the central-body effects, we must take the gradient (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. 10. 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, ϕ_{gcsat} , 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 ITRF frame,* we can directly determine the partial derivatives of the aspherical-potential function (including the two-body attraction). We often differentiate the Legendre functions in spherical coordinates $(r, \phi_{gcsat}, \lambda)$, but find the acceleration in cartesian coordinates, so we have to apply the chain rule. The acceleration is

* Remember that we usually evaluate the accelerations in the body-fixed frame (ITRF), but we numerically integrate in an inertial frame (GCRF).

$$\ddot{a} = \frac{\partial U}{\partial r} \left(\frac{\partial r}{\partial \dot{r}} \right)^T + \frac{\partial U}{\partial \phi_{gc_{sat}}} \left(\frac{\partial \phi_{gc_{sat}}}{\partial \dot{r}} \right)^T + \frac{\partial U}{\partial \lambda_{sat}} \left(\frac{\partial \lambda_{sat}}{\partial \dot{r}} \right)^T \quad (8-24)$$

Now, using the nonspherical portion of Eq. (8-19), (notice there is no discontinuity at $\phi_{gc_{sat}} = 90^\circ$)

$$\begin{aligned} \frac{\partial U}{\partial r} &= -\frac{\mu}{r^2} \sum_{l=2m=0}^{\infty} \sum_{l=2m=0}^{\ell} \left(\frac{R_\oplus}{r} \right)^\ell (l+1) P_{l,m} [\sin(\phi_{gc_{sat}})] \\ &\quad \left\{ C_{l,m} \cos(m\lambda_{sat}) + S_{l,m} \sin(m\lambda_{sat}) \right\} \\ \frac{\partial U}{\partial \phi_{gc_{sat}}} &= \frac{\mu}{r} \sum_{l=2m=0}^{\infty} \sum_{l=2m=0}^{\ell} \left(\frac{R_\oplus}{r} \right)^\ell \{ P_{l,m+1} [\sin(\phi_{gc_{sat}})] - m \tan(\phi_{gc_{sat}}) P_{l,m} [\sin(\phi_{gc_{sat}})] \} \\ &\quad \times \{ C_{l,m} \cos(m\lambda_{sat}) + S_{l,m} \sin(m\lambda_{sat}) \} \\ \frac{\partial U}{\partial \lambda_{sat}} &= \frac{\mu}{r} \sum_{l=2m=0}^{\infty} \sum_{l=2m=0}^{\ell} \left(\frac{R_\oplus}{r} \right)^\ell m P_{l,m} [\sin(\phi_{gc_{sat}})] \left\{ S_{l,m} \cos(m\lambda_{sat}) - C_{l,m} \sin(m\lambda_{sat}) \right\} \end{aligned} \quad (8-25)$$

The derivatives of the position vector also proceed directly as

$$\begin{aligned} \frac{\partial r}{\partial \dot{r}} &= \frac{\dot{r}^T}{r} \\ \frac{\partial \phi_{gc_{sat}}}{\partial \dot{r}} &= \frac{1}{\sqrt{r_I^2 + r_J^2}} \left(-\frac{\dot{r}^T r_K}{r^2} + \frac{\partial r_K}{\partial \dot{r}} \right) \\ \frac{\partial \lambda_{sat}}{\partial \dot{r}} &= \frac{1}{r_I^2 + r_J^2} \left(r_I \frac{\partial r_J}{\partial \dot{r}} - r_J \frac{\partial r_I}{\partial \dot{r}} \right) \end{aligned} \quad (8-26)$$

Notice the three partial derivatives in Eq. (8-26) are unit vectors. Putting the acceleration terms together and breaking out individual components make Eq. (8-24)

$$\begin{aligned} a_I &= \left\{ \frac{1}{r} \frac{\partial U}{\partial r} - \frac{r_K}{r^2 \sqrt{r_I^2 + r_J^2}} \frac{\partial U}{\partial \phi_{gc_{sat}}} \right\} r_I - \left\{ \frac{1}{r_I^2 + r_J^2} \frac{\partial U}{\partial \lambda_{sat}} \right\} r_J - \frac{\mu r}{r^3} \\ a_J &= \left\{ \frac{1}{r} \frac{\partial U}{\partial r} - \frac{r_K}{r^2 \sqrt{r_I^2 + r_J^2}} \frac{\partial U}{\partial \phi_{gc_{sat}}} \right\} r_J + \left\{ \frac{1}{r_I^2 + r_J^2} \frac{\partial U}{\partial \lambda_{sat}} \right\} r_I - \frac{\mu r}{r^3} \\ a_K &= \frac{1}{r} \frac{\partial U}{\partial r} r_K + \frac{\sqrt{r_I^2 + r_J^2}}{r^2} \frac{\partial U}{\partial \phi_{gc_{sat}}} - \frac{\mu r}{r^3} \end{aligned} \quad (8-27)$$

8.6.2 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 dominant than those from the Earth's oblateness. For more distant satellites, third-body effects and solar-radiation pressure dominate more than oblateness and drag. Investigations of aerobraking and satellite tethers require accurate atmospheric models. We must also understand the troposphere and the ionosphere to predict the behavior of tracking and communications systems. The idea is to have a good enough model to account for the atmospheric density while accurately modeling the effects of drag on an orbit. 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 in this section.

The cause of drag is the atmospheric particles, which retard the satellite's motion. Calculating density is extremely complex for real-world problems. Drag is a nonconservative perturbation because total energy isn't conserved (loss of energy due to friction). In fact, all velocity-dependent perturbations are nonconservative because they change the total energy. 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, planetary sciences, as well as orbital mechanics. Accurately determining atmospheric properties is crucial to determining drag on satellites.

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 Russian GOST 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:

$$\ddot{a}_{drag} = -\frac{1}{2} \frac{c_D A}{m} \rho v_{rel}^2 \frac{\dot{\vec{v}}_{rel}}{|\dot{\vec{v}}_{rel}|} \quad (8-28)$$

The **coefficient of drag**, c_D , is a dimensionless quantity which reflects the satellite's susceptibility to drag forces. The drag coefficient for satellites in the upper atmosphere is often approximately 2.2 (using a flat plate model). Spheres have $c_D \sim 2.0$ to 2.1. The drag coefficient is satellite configuration-specific and is seldom approximated to more

than three significant digits. The ***atmospheric density***, ρ , indicates how dense the atmosphere is at the satellite altitude and is perhaps the most difficult parameter to determine. Another difficult parameter to estimate is the exposed ***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.

We usually call $m/(c_D A)$ the ***ballistic coefficient***, BC; it's another measure of a satellite's susceptibility to drag effects. With this definition, a low BC means drag will affect the satellite a lot—and vice versa. Note that *many* organizations use the reciprocal $(c_D A)/m$, so it's very important to understand which one is in use.

The velocity 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. Remember that although the atmosphere is rotating, it does so with a "profile" that follows a little behind the Earth. Due to friction with the Earth, the atmosphere closest to the Earth rotates a little faster than higher altitudes. The velocity vector relative to the rotating atmosphere is

$$\hat{\vec{v}}_{rel} = \frac{d\hat{\vec{r}}}{dt} - \hat{\vec{\omega}}_{\oplus} \times \hat{\vec{r}} = \begin{bmatrix} \frac{dx}{dt} + \omega_{\oplus} y & \frac{dy}{dt} - \omega_{\oplus} x & \frac{dz}{dt} \end{bmatrix}^T$$

Escobal ([1965] 1985:60) gives a more general expression including wind variations:^{*}

$$\hat{\vec{v}}_{rel} = \begin{bmatrix} \frac{dx}{dt} + \omega_{\oplus} y + v_w \{-\cos(\alpha)\sin(\delta)\cos(\beta_w) - \sin(\alpha)\sin(\beta_w)\} \\ \frac{dy}{dt} - \omega_{\oplus} x + v_w \{-\sin(\alpha)\sin(\delta)\cos(\beta_w) + \cos(\alpha)\sin(\beta_w)\} \\ \frac{dz}{dt} + v_w \{\cos(\delta)\cos(\beta_w)\} \end{bmatrix}$$

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.

* The signs differ from Escobal because this equation finds the contribution of the wind using v_w and β_w in the SEZ system, and then rotates to the geocentric system $\{\text{ROT3}(-\alpha) \text{ROT2}(-(90-\delta))\}$.

Determining 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 very important for precise models.

The gravitational attraction of molecules in the atmosphere mainly determines 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):

$$\rho = \frac{p_o M}{g_o R T} \quad (8-29)$$

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 relation is the ***hydrostatic equation***, which relates the change in pressure, Δp , to the density, gravity, and change in altitude, Δh :

$$\Delta p = -\rho g \Delta h \quad (8-30)$$

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.

To gain an idea of the importance of modeling atmospheric parameters for astrodynamics, consider Fig. 8-7. The complexity of the atmosphere is apparent from the sheer number of 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.
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

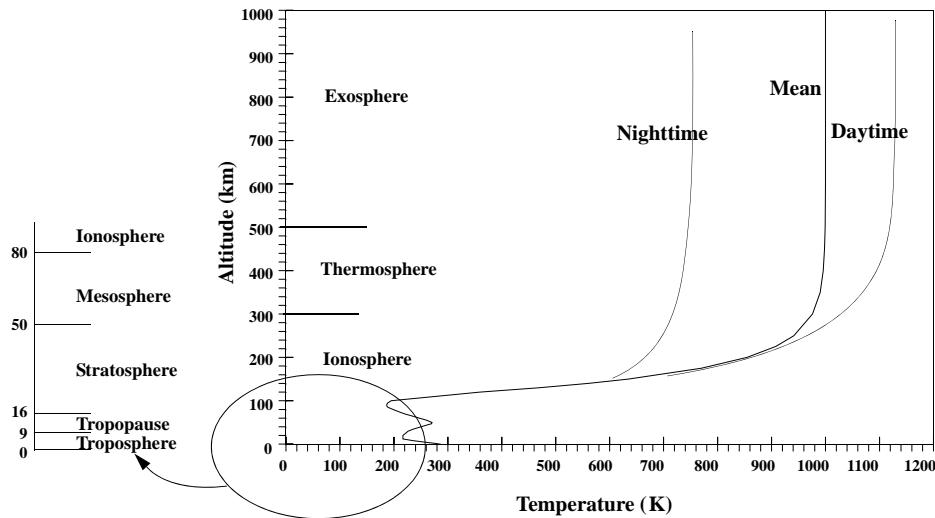


Figure 8-7. Earth's Atmosphere. The Earth's atmosphere is very complex. Notice how rapidly the temperature changes. The bands shown above are general areas where the atmosphere has similar properties. Use this figure carefully—values can change drastically over time.

completely symmetrical atmosphere to conform to every region on the Earth. Mountain ranges in the Himalayas cause significant differences in the wind, density, and temperature from the vast expanses of the open ocean at the same latitude but different longitudes.

Additional complexities appear in the *time-varying* models as we consider more real-world effects. For example, we must model

1. **Diurnal variations:** These variations occur every day as the Earth rotates. A pronounced atmospheric 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 2:00–2:30 P.M. It centers on the equator at the equinoxes but moves to higher latitudes depending on the Sun's declination, which varies throughout the year. 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, or *Solar Cycle** strongly varies the amount of incoming solar radiation that reaches Earth. 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. Recent cycles have experienced solar minima in March 1976, July 1986, August 1996, and November 2008 (Hathaway, 2010).
4. **Semi-annual / Seasonal variations:** These variations last about six months and are related to the varying distance of the Earth from the Sun and the Sun's declination 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 maximum 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 sunspot 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:** Ocean tides (and even atmospheric tides) also cause very small variations in the atmospheric density.

* The solar cycle is based primarily on the number of sunspots and there are Boulder and International Sunspot numbers (and some others). The Wolf 1848 formula is used by both groups, but the observational data is somewhat different. As a result, the dates of the minimums do not always agree.

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 disturbances cause ionization in the upper atmosphere, thereby affecting 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 noticeable, 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 and 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. Some atmospheric models, like the Jacchia-Roberts model, use 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. 8-8. 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

$$\text{one gamma} = 10^{-9} \text{ Tesla} = 10^{-9} \text{ kg}\cdot\text{s} / \text{m}$$

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 sunspots, although consistently large maxima of A_p usually occur in the declining phase of each 11-year cycle of $F_{10.7}$. 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 sunspot 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). Sample values of A_p are shown in Fig. 8-8.

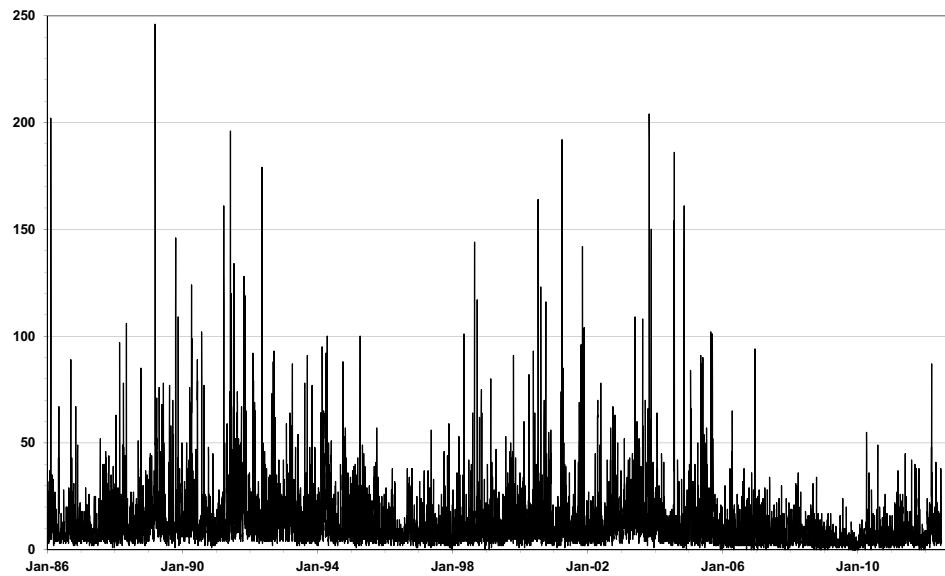


Figure 8-8. Daily Geomagnetic Planetary Index. Notice the seemingly chaotic behavior at almost all times for the average daily A_p . There may be some correlation to the solar cycle, but accurate prediction is very challenging! All units are gammas.

The 3-hourly index, a_p , is measured every three hours. It can be quite different from the daily values, and including it in calculations often results in substantially different (hundreds of meters) results. Some values are shown in Fig. 8-9.

Converting between K_p and a_p indices over their entire range, other than those actually measured is important. In over 50 years of recorded space weather data, there are only discrete recorded values of K_p and a_p quantities. Chapman and Bartels (1940) originally defined these discrete values (Table 8-3).

Examining a plot of the quantities, it's readily apparent that the relation is not linear.

Two practical situations arise. For the last month, only the 3-hourly K_p values are given, but any atmospheric model using a_p will need to convert the values, and when the values are finalized, they will be from the set of discrete values mentioned above. The second situation arises when assembling the predicted data, for which only a_p data is available. Specifically, values of a_p occur that do not have an exact corresponding K_p value per Chapman and Bartels (1940). To consistently use the data, a standard technique should be employed to convert the data to and from the a_p and K_p values, for both of these situations.

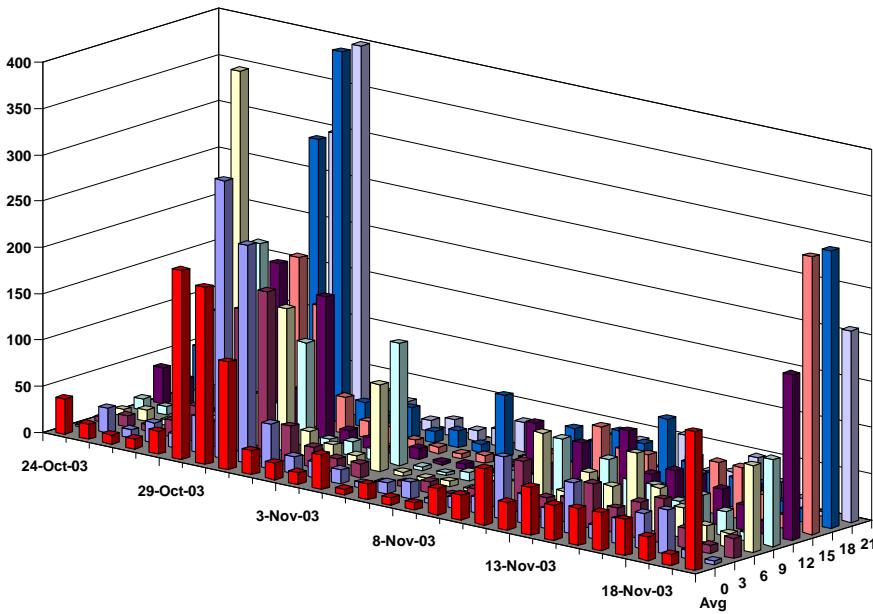


Figure 8-9. 3-hourly Geomagnetic Index. Notice the variability within a 24-hour day. The average daily values are shown in the “front” with each 3-hourly value behind it. All units are gammas.

TABLE 8-3. K_p and a_p Values. This table gives the discrete values for the geomagnetic indices. Note the plus and minus indicators represent 1/3 values and should not be rounded. Only two decimal places are shown for brevity.

K_p	0o	0+	1-	1o	1+	2-	2o	2+	3-	3o	3+	4-	4o	4+
K_p	0.0	0.33	0.67	1.0	1.33	1.67	2.0	2.33	2.67	3.0	3.33	3.67	4.0	4.33
a_p	0	2	3	4	5	6	7	9	12	15	18	22	27	32
K_p	5-	5o	5+	6-	6o	6+	7-	7o	7+	8-	8o	8+	9-	9o
K_p	4.67	5.0	5.33	5.67	6.0	6.33	6.67	7.0	7.33	7.67	8.0	8.33	8.67	9.0
a_p	39	48	56	67	80	94	111	132	154	179	207	236	300	400

A cubic splining technique seems to best replicate the observed data values, while simultaneously maintaining closure properties. The approach for the cubic spline is shown below (note this is modified to match all four points instead of 2 points and slope).

ALGORITHM 63: AP2Kp CONVERSION ($K_p \Rightarrow a_p$)

Setup defining arrays of a_p and K_p from Table 8-3.

Locate input a_{pin} or K_{pin} within the appropriate defining array input a_p or K_p (respectively)

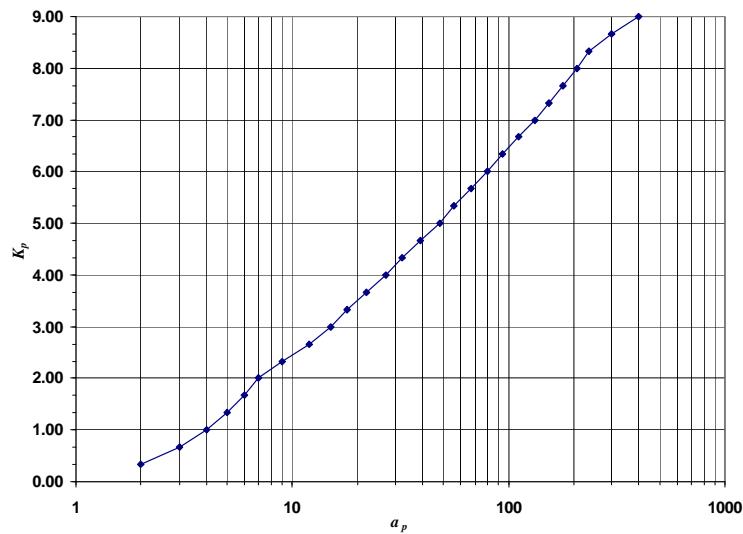


Figure 8-10. Relationship between a_p and K_p . The semi-logarithmic correlation between K_p and a_p is shown. Note that there are only distinct values that correspond between each scale and the K_p values are divided by 10 from published space weather values.

Find 4 adjacent points in the defining array for K_p and a_p (bracket the input K_{pin} and a_{pin} between points 2 and 3 ($K_{p1}, K_{p2}, K_{p3}, K_{p4}, a_{p1}, a_{p2}, a_{p3}, a_{p4}$))

Form coefficients for the cubic polynomials (one for a_p and one for K_p)

$$\alpha_{k0} = K_{p2}$$

$$\alpha_{a0} = a_{p2}$$

$$\alpha_{k1} = -\frac{K_{p1}}{3} - \frac{K_{p2}}{2} + K_{p3} - \frac{K_{p4}}{6}$$

$$\alpha_{a1} = -\frac{a_{p1}}{3} - \frac{a_{p2}}{2} + a_{p3} - \frac{a_{p4}}{6}$$

$$\alpha_{k2} = \frac{K_{p1}}{2} - K_{p2} + \frac{K_{p3}}{2}$$

$$\alpha_{a2} = \frac{a_{p1}}{2} - a_{p2} + \frac{a_{p3}}{2}$$

$$\alpha_{k3} = -\frac{K_{p1}}{6} + \frac{K_{p2}}{2} - \frac{K_{p3}}{2} + K_{p4}$$

$$\alpha_{a3} = -\frac{a_{p1}}{6} + \frac{a_{p2}}{2} - \frac{a_{p3}}{2} + a_{p4}$$

Solve the cubic polynomial for the real root (x) between 0.0 and 1.0 for either K_p or a_p (both are shown for completeness) using the input K_{pin} or a_{pin} value.

$$\alpha_{a3}x^3 + \alpha_{a2}x^2 + \alpha_{a1}x + \alpha_{a0} - a_{pin} = 0$$

$$\alpha_{k3}x^3 + \alpha_{k2}x^2 + \alpha_{k1}x + \alpha_{k0} - k_{pin} = 0$$

Solve for the remaining a_p or K_p value (both are shown for completeness)

$$\begin{aligned} a_p &= \alpha_{a3}x^3 + \alpha_{a2}x^2 + \alpha_{a1}x + \alpha_{a0} - K_{pin} \\ K_p &= \alpha_{k3}x^3 + \alpha_{k2}x^2 + \alpha_{k1}x + \alpha_{k0} - a_{pin} \end{aligned}$$

Solar Flux Data

Because the Sun's overall activity influences changes in 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 unpredictable perturbations in many satellites.

The contribution of solar flux to atmospheric density is mainly from incoming solar radiation. **Solar flux** (or **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. Today, there are a few space-borne systems to measure EUV flux, but no atmospheric-density models designed specifically to use these space-based measurements. However, scientists have determined that both EUV and incoming solar radiation with a wavelength of 10.7 cm, $F_{10.7}$ ($f = 2800$ MHz), originate in the same layers of the Sun's chromosphere and corona. Because the Earth's atmosphere is transparent to $F_{10.7}$ radiation, we can infer the relative strength of F_{EUV} from Earth-based measurements of 10.7 cm-length radio waves. Regular measurements of $F_{10.7}$ exist from about 1940. It's measured in *Solar Flux Units*, SFU, where

$$1 \text{ SFU} = 1 \times 10^{-22} \frac{\text{watt}}{\text{m}^2 \text{Hz}} \quad (8-31)$$

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 were routinely made at the Algonquin Radio Observatory in Ottawa, Ontario, Canada from 1947 until May 31, 1991 at 1700 UT. Since then, the measurements have been made at the Dominion Radio Astrophysical Observatory (DRAO) Penticton, British Columbia, Canada, at 2000 UT. Daily values are averaged to produce 81-day average values (3 solar rotations) denoted with a bar, $\bar{F}_{10.7}$. Figure 8-11 shows the values of $F_{10.7}$ for about one solar cycle.

The solar flux data may be either observed (at the true Sun-Earth distance), or adjusted to 1.0 AU (see Fig. 8-12). The raw observed data (and adjusted in DAILY-PLT.ADJ) may be found at NGDC (2013) and the conversion uses the current Earth-Sun distance ($r_{\oplus-Sun}$) and the AU distance (149,597,870.66 km).

$$F_{10.7}(obs) = \frac{F_{10.7}(adj)AU^2}{r_{\oplus-Sun}^2}$$

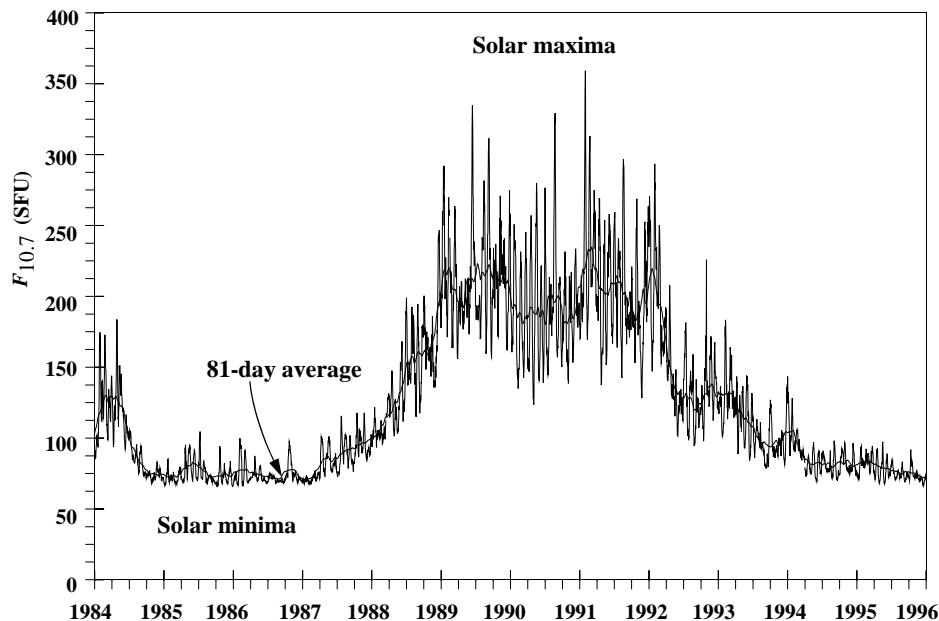


Figure 8-11. Values of Solar Flux. I've shown several years of $F_{10.7}$ values. Remember, $F_{10.7}$ is a radiation that originates at the same place as EUV which can't be measured on the Earth's surface. Notice the erratic behavior but a prominent rise during solar max. I've drawn the 81-day average, which we often use in density calculations. $F_{10.7}$ is also very difficult to predict.

Although the differences are small, some atmospheric models specifically require the solar flux at the true Earth-Sun distance. The difference is large enough to require separate values for the 81-day average, and because many programs use either the centered or trailing 81-day averages (about the date of interest), a total of six values are available for the solar flux. Also note that the values found in the GEOMAGNETIC and the SOLAR FLUX directories on NGDC do not agree completely because they are actually different data. The GEOMAGNETIC files contain Lenhart adjusted data while the SOLAR FLUX files contain observed and adjusted DRAO data (Knapp, 2005). Observed – adjusted values should show the cyclical variations of each solar cycle. Differences of the adjusted DRAO and Lenhart data should be zero. The data spikes in Fig. 8-12 appear to be random in that neither data source appears to be “correct” for all times.

Predicting Solar Flux

Solar flux receives a lot of attention because it is an important parameter in determining atmospheric density. The primary quantity, $F_{10.7}$, has been used as a proxy for the EUV radiation for many years. Several new indices have been under consideration for the past few years—Mg II (from NOAA and other satellites at about 280 nm), EUV (from the

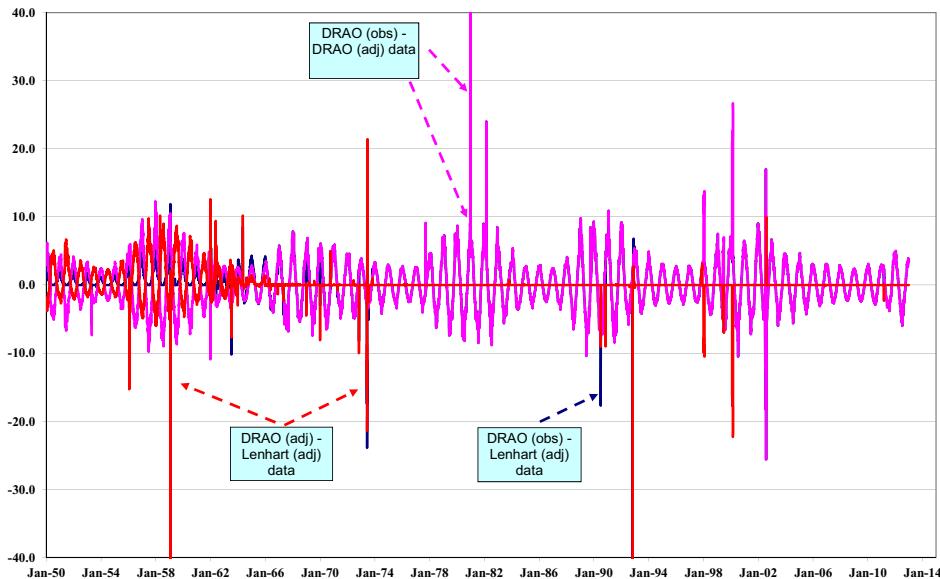


Figure 8-12. Difference of Observed and Adjusted Solar Flux Values. The solar cycles are evident in the observed minus adjusted solar flux differences (in SFU), along with the seasonal variations that drive the original variations over time. The data spikes occur in both observed and adjusted values, and in the DRAO and Lenhart data for a variety of reasons including multiple Sun-Earth distance corrections.

NASA/ESA SOHO satellite at the Lagrange point at 26-34 nm), etc.—but none has been unequivocally proven better than another and established as a new leader that will dramatically improve the accuracy of satellite operations. Moreover, predicting the $F_{10.7}$ values into the future poses significant challenges, and introduces significant uncertainty into operations requiring accurate forecasts of atmospheric behavior. Schatten and Sofia (1987) and Schatten (1988) have developed a monthly estimate of $F_{10.7}$ and a_p . Figure 8-13 shows some of those predictions and current data. There are early, mid and late timing, as well as min, mid, and maximum values for each prediction.

The difficulty of the prediction process becomes apparent for any cycle that departs significantly (i.e., cycle 24) from previous trends. Consider Fig. 8-14 for the plethora of predictions.

Even the short term predictions (3, 27 and 45 day) exhibit significant variations. Figure 8-15 shows the solar flux short term predictions.

The variability of the predictions is apparent. Many predictions rely on the solar sunspot activity. Routinely monitored since the 1700's, several relations exist between the sunspot number, R , averaged over a month or longer, and $F_{10.7}$ (Svalgaard, 2009).

$$F_{10.7} = 63.7 + 0.728 R + 0.000\ 89 R^2$$

$$F_{10.7} = 67 + 0.97 R + 17.6 (e^{-0.035R} - 1.0)$$

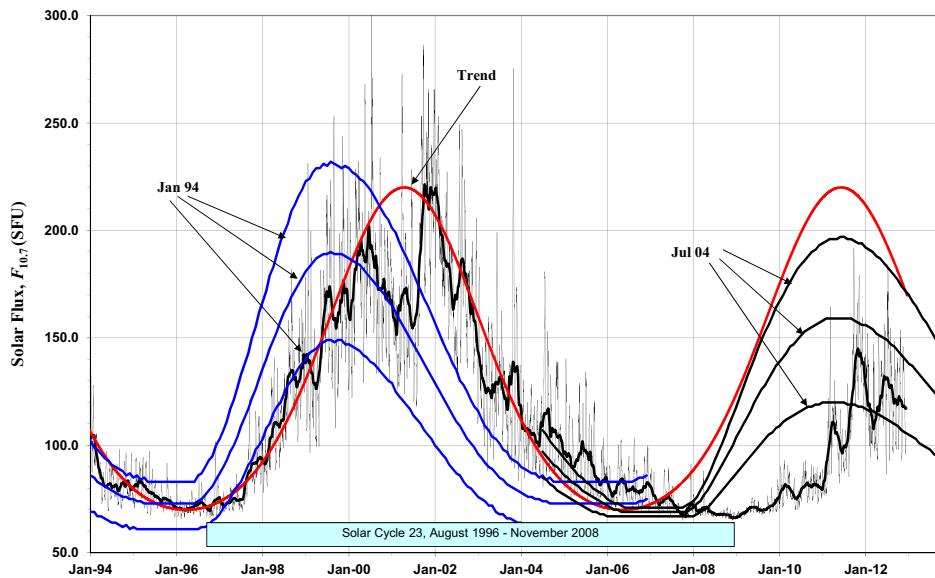


Figure 8-13. Schatten Predicted $F_{10.7}$. Actual data is shown for two Schatten predictions of $F_{10.7}$. Notice there are min, mid, and maximum predictions of the solar maxima (Schatten, 1988). Each prediction also includes early, mid, and late timing for the solar cycle (not shown). Comparing the predictions to actual data, the Jan 94 prediction was reasonably accurate while the Jul 04 prediction was significantly off. The 81-day average is close to the nominal prediction, although there is a lag. The polynomial is a simple heuristic expression that matches the last four solar cycles.

A polynomial intended to match the last few solar cycles, and assuming the next will not change dramatically from the previous, could also be used. In Eq. (8-32), t is the number of days from January 1, 1981. While matching cycle 23, it misses cycle 24.

$$F_{10.7} = 145 + 75 \cos(0.001\,696t + 0.35 \sin(0.000\,016\,95t)) \quad (8-32)$$

Another interesting approach is given by Oltrogge and Chao (2007). Their method for extrapolating and interpolating atmospheric density is based on adjusting or modifying actual density values, not the proxies. The rationale is that the dependence of density on the proxies is nonlinear; hence averaging or interpolating proxies does not yield correct average or interpolated density. They were examining the effect of solar flux prediction estimates on satellite lifetimes and noted that most orbital lifetime prediction errors were caused by (1) the poor performance of current mean solar activity predictions as compared to actual activity variations; (2) non-availability of predictions more than one solar cycle away; and (3) nonlinear density as a function of solar and geomagnetic activity. To circumvent these obstacles, they used the complete set of solar and geomagnetic data available (from Feb 1947). Thus, five solar cycles were used and combined into a single cycle of 10.82546 years (3954 days). Rather than attempting to produce a single

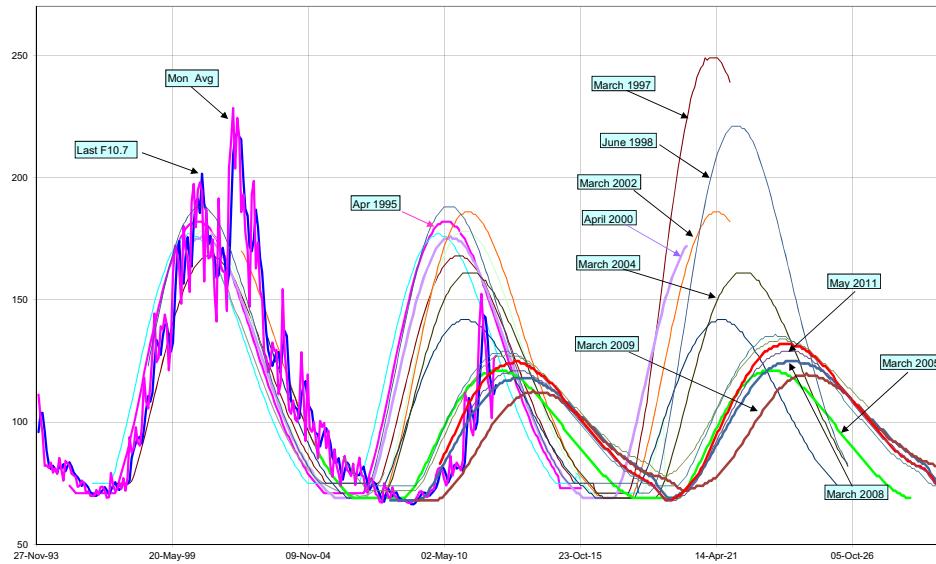


Figure 8-14. Recent Solar Flux Predictions. Several Schatten predictions of solar flux are shown. Notice that each Schatten prediction covers about two solar cycles and the newer predictions suggest a lower solar max for cycles 24 and 25. Units are SFU.

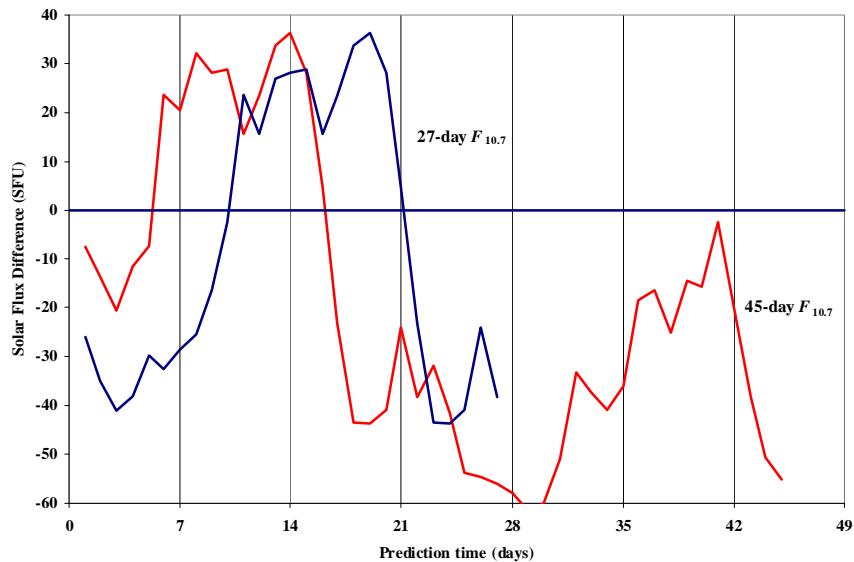


Figure 8-15. Short term $F_{10.7}$ Predictions. The 27 and 45 day predictions are shown when compared to the actual data. Notice the significant departures. The geomagnetic indices exhibit the same behavior.

mean solar cycle, they chose to directly use the solar flux and geomagnetic values from the previous cycles. Essentially, the technique creates a day of solar and geomagnetic data (a combined triad of values including the $\bar{F}_{10.7}$ 81-day average) from one of the five cycles at random, and does so for each day of the simulation. Thus, the existing data is used to construct a future solar cycle, while retaining all the variability and interrelations of the existing data.* This technique is certainly worthy of additional analysis and use as it avoids the vagaries of other prediction techniques that often vary widely for estimates just a few years apart.

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. 8-16.

There are various models to satisfy differing accuracy models. We could always use a time-varying model, like the Jacchia-Roberts atmosphere, but the computational requirements might be too great. Models of this fidelity are the most complete and need 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. Gaposchkin and Coster (1988) discuss various atmospheric models in detail. They conclude that no model is best for *all* applications. This is because there are physical reasons that the models have errors—winds, gravity waves, etc. They note that Jacchia-71 performs quite well, and is the fastest among the models tested (J71, DTM, J77, and three variations of J77). However, they prefer the J77 with certain updates for the best combination of speed, accuracy, and applicability.

One feature of the analytical density models is that their relatively simple structure allows analytical computation of the integrals required for analytical and semianalytical satellite propagation theories using these densities. Numerical implementations, and some models in semianalytical theories like the DSST require numerical quadrature for the same integrals.

Exponential Model (0–1000 km)

This simple, static model assumes the density of the atmosphere decays exponentially with increasing altitude. It also assumes a spherically symmetrical distribution of particles, in which the density, ρ , varies exponentially according to

$$\rho = \rho_o \exp\left[-\frac{h_{ellp} - h_o}{H}\right] \quad (8-33)$$

* Their resulting solar and geomagnetic activity computer code tdt2f10ap_60yr is available upon request from the author (Oltrogge@1EarthResearch.com) for users wishing to explore the model in greater detail.

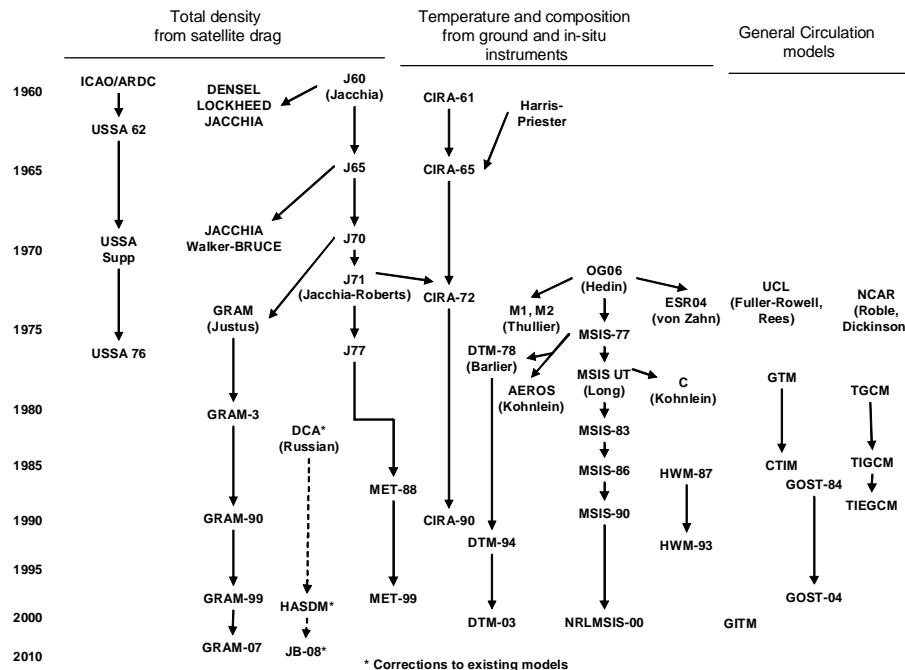


Figure 8-16. Development of Atmosphere Models. Notice the variety of models. Flow of information among the three overall categories is limited (Marcos, et al. 1993, 20). The main models in use today are the Standard Atmosphere, USSA76; variations of the Jacchia-Roberts, J71, J77, and GRAM99; COSPAR International Reference Atmosphere, CIRA90; Mass Spectrometer Incoherent Scatter, NRLMSIS-00; Drag Temperature Model (DTM), Marshall Engineering Thermosphere (MET), the Russian GOST, Global Ionosphere Thermosphere Model (GITM), and general circulation models. Dynamic Calibration of the atmosphere (DCA), High Accuracy Satellite Drag Model (HASDM), and Jacchia-Bowman (JB-08) are corrections to other models, usually J70.

where a *reference density*, ρ_o , is used with the *reference altitude*, h_o , the *actual altitude*, above the ellipsoid, h_{ellp} , and the *scale height*, H .* Table 8-4 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 simulating drag effects for design studies. Extensive tabulated data on atmospheric parameters (density, pressure, temperature,

* Scale height is the fractional change in density with height. It can be useful in determining numerical partial derivatives. Wertz (1978:108) shows that scale height is equal to $k \times \text{Temperature} / \text{molecular weight} \times \text{gravity}$, where k is the Boltzmann constant.

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.

TABLE 8-4. 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 exospheric temperature, $T_\infty = 1000$ 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	9.518×10^{-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

▼ Example 8-4. Calculating Atmospheric Density.

GIVEN: $r = 7125.3489$ km

FIND: ρ

Eq. (8-33) requires knowledge of the altitude, found by subtracting the Earth's radius (6378.137 km) from the satellite's radius r ($h_{ellp} = 747.2119$ km). Now, using values from Table 8-4, we use Eq. (8-33)

$$\rho = 3.614 \times 10^{-14} \exp\left[-\frac{747.2119 - 700}{88.667}\right] = 2.1219854 \times 10^{-14} \frac{\text{kg}}{\text{m}^3}$$

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 from being at a higher altitude. In practice, the height is found above the ellipsoid and does not assume a spherical Earth.

We can also determine pressure through a similar formula that approximates pressure at any altitude. The formula is

$$p = p_{SL} \text{EXP} \left[-\frac{h_{ellp}}{H} \right]$$

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 8-5]... [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 (70–2500 km)

Jacchia's models (J70, 1970; J71, 1971; J77, 1977) are popular in military operations (J70) and some analyses for astrodynamics. The Jacchia-Roberts atmosphere contains analytical expressions for determining exospheric temperature as a function of position, time, solar activity, and geomagnetic activity. With a computed temperature, 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 several regions: 90–125 km and above 125 km. He determined these regions 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 find, but predictions need some estimate such as the reasonably accurate Schatten approximation.

In 1971, Jacchia re-formulated his model using newer, and in many cases, more complete data. The changes are dramatic in comparison to the 1970 model. Jacchia (1971) remarks:

While overhauling the basic models, we have also tried to reanalyze these variations. In so doing, we have found that for some of them—the geomagnetic effect, the semiannual variation, and the Helium variation—the analytical formulation we had used was inadequate and had to be altered, or even drastically changed. In particular, the dissociation of the

TABLE 8-5. 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}	2.185×10^{-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}

semiannual variation from temperature variations has cleared up many puzzling results from the Helium-hydrogen region and eliminated the necessity of introducing ad hoc variations for these constituents.

The Jacchia 1977 revision incorporates satellite mass spectrometer data. The model again revised equations from the previous models, and there are several variations of this technique.

Roberts (1971) recognized that tabular determination of atmospheric density 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 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 $W(\nu)$ 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 Jacchia values. Sec. B.1 of the Appendixes details the formulas for the Jacchia-Roberts atmosphere. Oza and Freier (1995) suggest that the Jacchia-Roberts model is very responsive to changes in geomagnetic activity. Partial derivatives are available for the Jacchia-Roberts model in Long et al. (1989:4-54 to 4-57).

DTM (200–1200 km) and NRLMSIS-00 (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 Freier (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. Unfortunately, some operational programs routinely use the latest [trailing] 81-day average because the predicted values are not well known (Jablonski, 1992).

The MSIS models derive from the DTM models. They are quite accurate and have been successfully applied to many stressing problems. For instance, Hujasak (1999) shows results of using the MSIS model in determining reentry of a decaying Molniya satellite. Because the satellite is traveling so fast at perigee, it can cross several bands of the atmosphere during a single integration step. MSIS does not suffer these difficulties. A newer release is the NRLMSISE-00 model which is extremely popular for all applications. Complete source code is available for all the 1986, 1990, and 2000 models. The models run a little slower than the Jacchia models.

Russian GOST Model (120–1500 km)

The Russian GOST density model (Yurasov, 1999, Voiskovskii, 1973, Volkov, 1984, Jablonski [Boelitz], 1992, and Amelina et al. 1996) is an analytical method to obtain atmospheric density in an aspherical upper atmosphere from observations of Russian Cosmos satellites. It's been part of the Russian operational space surveillance system for

more than 30 years, was used with the Apollo-Soyuz Test Project in 1975, and it continues to incorporate updates from new satellite data. 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. Consequently, the model runs very fast. The model also includes long and short period variations in solar activity. The variability in the Russian 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 GOST model differed only 5–10% (Volkov, 1982) from the 1971 Jacchia density model.

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 120–1500 km. The current version is GOST 25645.166-2004. Appendix B further discusses this model.

Dynamic Calibration of the Atmosphere (DCA)

Unlike the previous models, this approach isn't really a model, but rather a technique for improving or correcting atmospheric density. It gives fundamental scientific information about the variations in the density and the statistics of these variations. The work was pioneered by Nazarenko in the early 1980's (Gorochov and Nazarenko, 1982) and researched by Draper Laboratory (many documents including Cefola and Nazarenko, 1999, Granholm, 2000, Bergstrom, 2002, Yurasov et al. 2005, and Wilkins et al. 2006). Granholm (2000) and Wilkins et al. 2006 provide an excellent description of the technique, and Bergstrom (2002) showed results of implementing and testing the theory. The concept relies on changing the density directly. Figure 8-17 shows the general process.

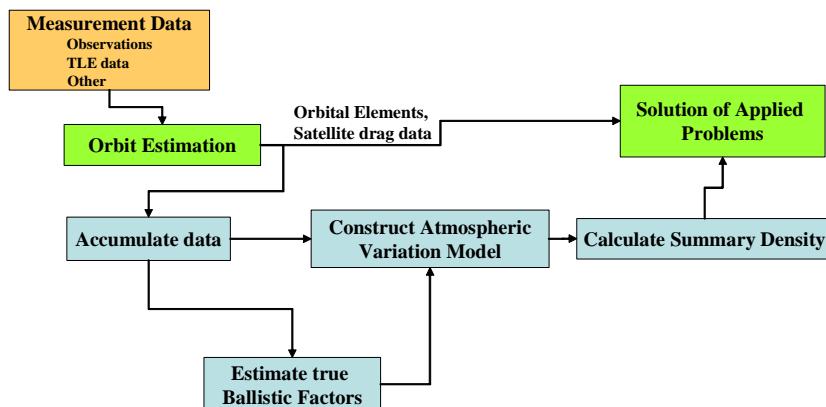


Figure 8-17. Process Flow for DCA models. This figure shows the interrelations for orbital estimation and DCA corrections. Note that the construction of the density model follows the estimation in parallel. Also, the determination of the true ballistic factors should be performed in parallel.

The model determines the density corrections about every 3 hours from a set of “calibration” satellites.* The three hour rate for density corrections is not unique. It came from a Russian consideration of determining the three hourly terms in the density from empirical inputs rather than observed geomagnetic data which they considered as unreliable in the early to mid 1980s. Note that in the recent work determining the density corrections from NORAD TLE information, the density corrections are determined once per day.

The “true” ballistic coefficient is used as an input, with processing about once every 20 days or so. This information becomes the basis of changing the atmospheric densities from an existing model—often Jacchia-Roberts-71 or an MSIS model.

The dynamic calibration work was later examined by the AFSPC (Storz, 1999). They investigated changing the temperature calculations and using spherical harmonics to extrapolate the results for global coverage. At the present time, there are a variety of AFSPC atmospheric model approaches implementing variations on the original DCA technique. Some are based on J70, while others use J71 with modifications. They use a variety of indices, some of which limit the time applicability of the model, and the performance is many times slower than the J70 model. See Marcos et al. (2006) and Bowman et al. (2006, 2008) for additional information.

Wright and Woodburn (2004) have independently shown that using a Kalman filter, you can separate the atmospheric density errors from the modeling errors and bypass the additional computations with the density, spherical harmonics, fit spans, etc. This work was limited to the simultaneous estimation of the ballistic coefficient and atmospheric density correction for each satellite independently without extension to additional satellites. The separation of estimates for two linearly dependent states was enabled by the filter formulation which allows for both variables to be unknown functions of time, but restricting their variability with time to have greatly different time scales. Thus, all the variation on short time scales becomes attributable to density corrections while a long term bias is moved into the ballistic coefficient. The filter formulation could be extended to simultaneously estimate corrections to parameters within the global density model to provide an analog to the dynamic calibration work mentioned above.

Most studies have focused on a perceived variability of the ballistic coefficient (independent from the attitude induced changes through the frontal area) because batch least squares techniques can not track the time-varying atmospheric model errors, nor the rapid and dynamic changes induced by atmospheric drag. The result is that the ballistic coefficient supposedly exhibits large variations caused by atmospheric density variations. In reality, the attitude contributes the largest portion of the error and the atmospheric density models simply add to that.

* The calibration satellites are an ad-hoc collection of LEO satellites that receive additional observational data, providing for slightly better than average orbit determination results. Selecting the calibration satellites for this technique is important because as the solar cycle raises and lowers the altitude at which atmospheric drag effects are prominent, the accuracy of the data from these calibration satellites will change.

The distinguishing feature of all the dynamic calibration approaches are that they permit the first breakthrough of the generally accepted 10-15% error in atmospheric models. All these approaches have significant applications to operational orbit determination.

Partial Derivatives

In general, we find the partial derivatives for drag using the chain rule and the density, ρ .

$$\frac{\partial \ddot{\vec{r}}_{drag}}{\partial \vec{r}} = \frac{\partial \ddot{\vec{r}}_{drag}}{\partial \rho} \frac{\partial \rho}{\partial \vec{r}}$$

The variety of atmospheric models would require significant space to list all the possibilities. Thus, I've only begun the general technique here. Consult Long et al. (1989:4-32) for additional information.

8.6.3 Third-Body Perturbations

In Chap. 1, 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, have a greater affect on satellites in higher altitude orbits. Their effects become noticeable about when the effects of drag begin to diminish. Because the 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 applications.

Accelerations Due to Third Body

I developed the equations of motion for a three-body system in Eq. (1-37). Let the third body be 3 and assume the mass of the satellite is negligible.

$$\ddot{\vec{r}}_{\oplus sat} = -\frac{\mu_{\oplus} \vec{r}_{\oplus sat}}{r_{\oplus sat}^3} + \mu_3 \left(\frac{\vec{r}_{sat3}}{r_{sat3}^3} - \frac{\vec{r}_{\oplus 3}}{r_{\oplus 3}^3} \right) \quad (8-34)$$

At this point, we can numerically integrate to determine the third-body's effect on the satellite; however, careful examination of Eq. (8-34) 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. 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) state 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. Roy (1978:202) shows how to find a solution directly.

$$Q = \frac{(r_{\oplus sat}^2 + 2(\dot{\vec{r}}_{\oplus sat} \cdot \dot{\vec{r}}_{sat3})) (r_{\oplus 3}^2 + r_{\oplus 3} r_{sat3} + r_{sat3}^2)}{r_{\oplus 3}^3 r_{sat3}^3 (\dot{\vec{r}}_{\oplus 3} + \dot{\vec{r}}_{sat3})} \quad (8-35)$$

$$\frac{\ddot{\vec{r}}_{sat3}}{r_{sat3}^3} - \frac{\dot{\vec{r}}_{\oplus 3}}{r_{\oplus 3}^3} = \dot{\vec{r}}_{sat3} Q - \frac{\dot{\vec{r}}_{\oplus sat}}{r_{\oplus 3}^3}$$

Geyling and Westerman (1971:113) give us another solution using a Taylor series expansion and neglect small terms because $r_{\oplus sat} \ll r_{\oplus 3}$. Including the last term in the parenthesis [Neta (1996)] the final result is

$$\ddot{\vec{r}}_{\oplus sat} \cong -\frac{\mu_{\oplus} \dot{\vec{r}}_{\oplus sat}}{r_{\oplus sat}^3} - \frac{\mu_3}{r_{\oplus 3}^3} \left(\dot{\vec{r}}_{\oplus sat} - 3\dot{\vec{r}}_{\oplus 3} \frac{\dot{\vec{r}}_{\oplus sat} \cdot \dot{\vec{r}}_{\oplus 3}}{r_{\oplus 3}^2} - \frac{15}{2} \left(\frac{\dot{\vec{r}}_{\oplus sat} \cdot \dot{\vec{r}}_{\oplus 3}}{r_{\oplus 3}^2} \right)^2 \dot{\vec{r}}_{\oplus 3} \right) \quad (8-36)$$

Although this form is numerically stable, it's approximate (the last term in the parenthesis is often omitted) and therefore less accurate. When the third body is the Moon, the distances from the Earth to the satellite and from the Earth to the Moon are much closer than that from the Earth to the Sun, we may need more terms from the Taylor series to preserve the accuracy.

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 [$\dot{\vec{r}}_{sat3} / r_{sat3}^3$ in Eq. (8-34)], we can find the magnitude of $\dot{\vec{r}}_{sat3}$ using the cosine law plus the sides of the triangle formed by the Sun, satellite, and center of the inertial coordinate system.

Designating ζ as the angle between the third body and the satellite as seen from the Earth, the cosine law gives

$$r_{sat3}^2 = r_{\oplus sat}^2 + r_{\oplus 3}^2 - 2r_{\oplus sat}r_{\oplus 3}\cos(\zeta)$$

and letting

$$h = \frac{r_{\oplus sat}}{r_{\oplus 3}} \quad B = \sum_{j=1}^{\infty} P_j[\cos(\zeta)]h^j$$

we expand $1/r_{sat3}$ in a series that is the generating function for Legendre polynomials

$$\frac{1}{r_{sat3}} = \frac{1}{r_{\oplus 3}} \{ P_0[\cos(\zeta)] + P_1[\cos(\zeta)]h + \dots \} = \frac{1+B}{r_{\oplus 3}}$$

We can substitute this into Eq. (8-34) and obtain an alternate form for the acceleration.

$$\ddot{\vec{r}}_{\oplus sat} = -\frac{\mu_{\oplus} \dot{\vec{r}}_{\oplus sat}}{r_{\oplus sat}^3} + \mu_3 \left\{ \dot{\vec{r}}_{sat3} \frac{(1+B)^3}{r_{\oplus 3}^3} - \frac{\dot{\vec{r}}_{\oplus 3}}{r_{\oplus 3}^3} \right\}$$

By expanding $(1+B)^3$

$$\ddot{\vec{r}}_{\oplus sat} = -\frac{\mu_{\oplus} \dot{\vec{r}}_{\oplus sat}}{r_{\oplus sat}^3} + \mu_3 \left\{ \dot{\vec{r}}_{sat3} \frac{(1 + 3B + 3B^2 + B^3)}{r_{\oplus 3}^3} - \frac{\dot{\vec{r}}_{\oplus 3}}{r_{\oplus 3}^3} \right\}$$

Recognizing that $\dot{\vec{r}}_{sat3} - \dot{\vec{r}}_{\oplus 3} = -\dot{\vec{r}}_{\oplus sat}$ and distributing signs, we have

$$\ddot{\vec{r}}_{\oplus sat} = -\frac{\mu_{\oplus} \dot{\vec{r}}_{\oplus sat}}{r_{\oplus sat}^3} - \mu_3 \left\{ \frac{-\dot{\vec{r}}_{sat3}(3B + 3B^2 + B^3) + \dot{\vec{r}}_{\oplus sat}}{r_{\oplus 3}^3} \right\} \quad (8-37)$$

Now, let's generalize this expression to multiple disturbing third bodies, $k = 1, 2, \dots, n$. Thus,

$$\ddot{\vec{r}}_{\oplus sat} = -\frac{\mu_{\oplus}}{r_{\oplus sat}^3} \dot{\vec{r}}_{\oplus sat} - \sum_{k=1}^n \frac{\mu_k}{r_{\oplus k}^3} (\dot{\vec{r}}_{\oplus sat} - \beta_k \dot{\vec{r}}_{satk}) \quad (8-38)$$

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. (8-34) 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.

We find the disturbing potential function by expressing 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 .

We use the addition theorem with which we developed the gravitational potential in Eq. (8-15), to obtain

$$P_l[\cos(\zeta)] = \sum_{m=0}^l k_m \frac{(l-m)!}{(l+m)!} P_{l,m}[\sin(\delta)] P_{l,m}[\sin(\delta_3)] \cos(m(\alpha - \alpha_3))$$

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. Using Eq. (C-16) for $\cos(m(\alpha - \alpha_3))$, define

$$A_{l,m} = \frac{\mu_3}{r_3^{l+1}} k_m \frac{(l-m)!}{(l+m)!} P_{l,m} [\sin(\delta_3)] \cos(m\alpha_3)$$

$$B_{l,m} = \frac{\mu_3}{r_3^{l+1}} k_m \frac{(l-m)!}{(l+m)!} P_{l,m} [\sin(\delta_3)] \sin(m\alpha_3)$$

Notice that we group the parameters relating to the third body into $A_{l,m}$ and $B_{l,m}$ as we did in the development of the Earth's gravitational potential. These coefficients are analogous to the $C_{l,m}$ and $S_{l,m}$ coefficients for the geopotential, but they depend on the physical location of the third body. Thus, they reflect the distribution of mass over the orbital motion of the third body, not the shape of the third body itself. We now have

$$R_{3-body} = \sum_{l=2}^{\infty} \sum_{m=0}^l r^l P_{lm} [\sin(\delta)] \{ A_{l,m} \cos(m\alpha) + B_{l,m} \sin(m\alpha) \} \quad (8-39)$$

Partial Derivatives

We require partial derivatives of the third-body acceleration for orbit-determination and estimation techniques in Chap. 10. Begin with Eq. (8-34) and assume the Sun is the third body. Consider first the partial derivative with respect to position $\dot{\vec{r}}_{\oplus sat}$:

$$\begin{aligned} \frac{\partial \ddot{\vec{r}}}{\partial \dot{\vec{r}}_{\oplus sat}} &= -\mu_{\oplus} \left[\frac{\partial(1/r_{\oplus sat}^3)}{\partial \dot{\vec{r}}_{\oplus sat}} \dot{\vec{r}}_{\oplus sat} + \frac{1}{r_{\oplus sat}^3} \frac{\partial \dot{\vec{r}}}{\partial \dot{\vec{r}}_{\oplus sat}} \right] \\ &\quad + \mu_{\odot} \left[\frac{\partial(1/r_{sat\odot}^3)}{\partial \dot{\vec{r}}_{\oplus sat}} \dot{\vec{r}}_{sat\odot} + \frac{1}{r_{sat\odot}^3} \frac{\partial \dot{\vec{r}}_{sat\odot}}{\partial \dot{\vec{r}}_{\oplus sat}} \right] \end{aligned} \quad (8-40)$$

Note that the indirect term ($\dot{\vec{r}}_{\oplus\odot}/r_{\oplus\odot}^3$) in Eq. (8-34) is independent of the satellite's position and velocity; thus, the partial derivative vanishes. Clearly, we'll need to use the identities $\partial r/\partial \dot{\vec{r}} = \dot{\vec{r}}/r$ and $\partial \dot{\vec{r}}/\partial \dot{\vec{r}} = \mathbf{I}$, where \mathbf{I} is a 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. (8-40) will produce a result proportional to $\dot{\vec{r}}_{\oplus sat}$ resulting in the apparent product of two vectors $\dot{\vec{r}}_{\oplus sat} \dot{\vec{r}}_{\oplus sat}^T$. 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 an $n \times 1$ column vector with a $1 \times m$ 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. (8-40) yields

$$\frac{\partial(1/r_{\oplus sat}^3)}{\partial \dot{\vec{r}}_{\oplus sat}} = \frac{-3}{r_{\oplus sat}^4} \frac{\partial r_{\oplus sat}}{\partial \dot{\vec{r}}_{\oplus sat}} = \frac{-3}{r_{\oplus sat}^5} \dot{\vec{r}}_{\oplus sat}$$

$$\frac{\partial(1/r_{sat\odot}^3)}{\partial \dot{\vec{r}}_{\oplus sat}} = \frac{-3}{r_{sat\odot}^4} \frac{\partial r_{sat\odot}}{\partial \dot{\vec{r}}_{\oplus sat}}$$

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. A single third body uses k and $i = 1$.

$$\begin{aligned}\frac{\ddot{\vec{r}}_{\oplus sat}}{\dot{\vec{r}}_{\oplus sat}} &= \left\{ -\frac{\mu_{\oplus}}{r_{\oplus sat}^3} - \sum_{i=1}^k \mu_i \frac{1}{r_{sat_i}^3} \right\} \mathbf{I} + 3 \left(\frac{\mu_{\oplus} \dot{\vec{r}}_{\oplus sat} \dot{\vec{r}}_{\oplus sat}^T}{r_{\oplus sat}^5} + \sum_{i=1}^k \mu_i \frac{\dot{\vec{r}}_{sat_i} \dot{\vec{r}}_{sat_i}^T}{r_{sat_i}^5} \right) \\ \frac{\ddot{\vec{r}}_{\oplus sat}}{\dot{\vec{r}}_{\oplus sat}} &= 0\end{aligned}$$

8.6.4 Solar-Radiation Pressure

Like drag, solar-radiation pressure is a nonconservative perturbation that becomes more pronounced 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.

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 exerts 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. The entire process involves determining the Sun's precise location; the correct satellite 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 that model the satellite's reflectivity. I've discussed the first two items 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) states 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 (Planck's constant is $h = 6.6256 \times 10^{-34} \text{ J.s}$). A *solar-radiation constant*—often called the *intensity, irradiance*, or solar flux is

$$SF = 1367 \text{ W/m}^2 \quad (8-41)$$

Many programs use this value (or another similar value) for the solar constant because determining the actual frequencies and energy is very difficult and varies over time. Sources of current values are listed in Appendix D, and Fig. 8-18 shows historical values. Values are sometimes given in Watts (e.g., 3.823×10^{26}) which we convert to Watts/m² by dividing by 4π times the Earth distance from the Sun squared— $(4\pi)(149,597,870,000)^2$.

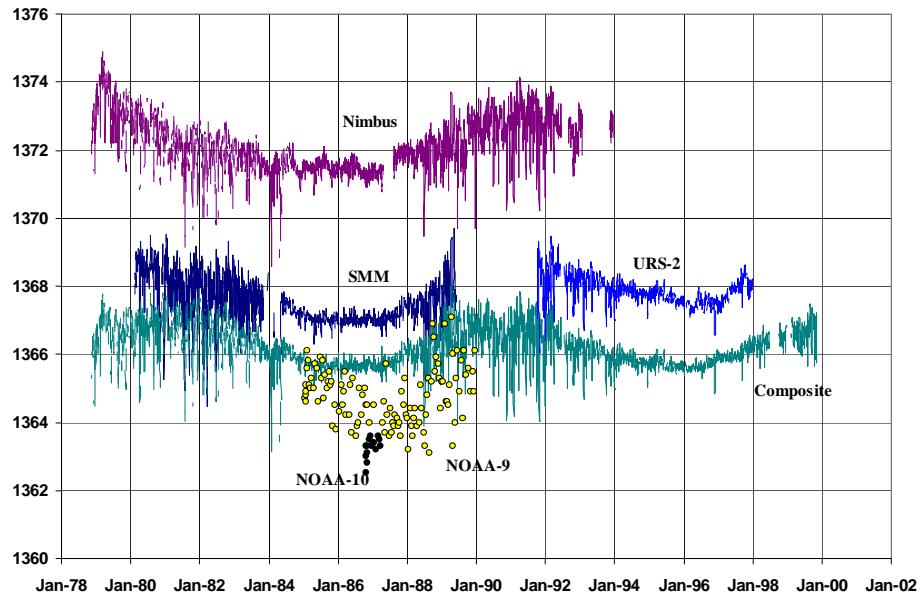


Figure 8-18. Solar Irradiance Values. Data from several satellites is shown for solar irradiance measurements. Notice the variability of values and lack of definitive trends. Solar Maximum Mission (SMM) and Upper Atmospheric Research Satellite (UARS) satellite data are also included. Units are Watts/m².

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 (notice the different solar irradiance value):

$$SF = \frac{1358}{1.004 + 0.0334 \cos(D_{aphelion})} \frac{\text{W}}{\text{m}^2} \quad (8-42)$$

where $D_{aphelion}$ is 2π times the days from when the Earth is at aphelion, as a fraction of the whole year. 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 effect of the pressure, or the 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 force of **solar pressure**, p_{srp} , per unit area, or the change in momentum:

$$p_{srp} = \frac{1367 \text{ W/m}^2}{3 \times 10^8 \text{ m/s}} = 4.57 \times 10^{-6} \frac{\text{W} \cdot \text{s}}{\text{m}^3} = 4.57 \times 10^{-6} \frac{\text{N}}{\text{m}^2}$$

Remember that this value will change if we use more precise values of the solar-radiation intensity, as is possible with Eq. (8-42).

Now we can develop expressions for the force. Using the **reflectivity**, c_R , the **solar-radiation pressure**, p_{srp} , and the **exposed area to the Sun**, A_\odot , the force is

$$\hat{\vec{F}}_{srp} = -p_{srp} c_R A_\odot \frac{\hat{\vec{r}}_{\oplus\odot}}{|\hat{\vec{r}}_{\oplus\odot}|} \quad (8-43)$$

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 momentum is transmitted (and therefore no force), but there may be some refraction. A value of 1.0 means that all the radiation is absorbed, and all the momentum is transmitted (i.e., it's a black body) resulting in a perturbing force. Finally, 2.0 indicates that all the radiation is reflected and twice the momentum is transmitted to the satellite (i.e., a flat mirror perpendicular to the light source). In this case, the resulting force is twice the absorption force due to Newton's first law. It turns out that determining c_R properly is extremely difficult. It changes over time and is virtually impossible to predict. This is especially true for complex satellites made of various materials, that enter and exit eclipse regions, and have a constantly changing orientation. For this reason, it's almost always a solution parameter, determined in differential correction (just like the drag coefficient).

It's common to assume the surface maintains a constant attitude perpendicular to the Sun, thereby setting $\phi_{inc} = 0^\circ$ (See Fig. 8-19). Though not very realistic, it does give us rough estimates for the problem of solar radiation. Burns et al. (2000) describe an attitude guidance law which yields a time-varying orientation. Newton's second law permits us to determine the acceleration with the given force:

$$a_{srp} = \frac{\vec{F}_{srp}}{m} = \frac{p_{srp} c_R A_\odot}{m}$$

Because the direction is always away from the Sun, a unit vector from the satellite to the Sun yields the correct direction (the sign is positive if $\hat{\vec{r}}_{\odot sat}$ is used instead). The acceleration commonly used for numerical simulations is

$$\ddot{\vec{a}}_{srp} = -\frac{p_{srp} c_R A_\odot}{m} \frac{\dot{\vec{r}}_{sat\odot}}{|\dot{\vec{r}}_{sat\odot}|} \quad (8-44)$$

In forming Eq. (8-44), we made several assumptions concerning satellite attitude and other 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 often ignore the torque and make the simplifying assumption that the disturbing acceleration is given by Eq. (8-44), where A_\odot is an average effective cross-section that implicitly incorporates c_R .

In reality though, the reflection process is two-fold: absorption followed by reflection. Figure 8-19 shows the forces imparted by absorption and reflection. Remember that the absorbed and reflected forces aren’t orthogonal, so they don’t combine the same as normal and tangential expressions. Burns et al. (2000) shows additional formulations for calculating different surfaces, and how to analyze these forms through VOP analysis (Sec. 9.3.2).

There are also two kinds of reflection—*specular* and *diffuse*. We won’t discuss absorbed reflection because this requires us to have complex three-dimensional models of the satellite. Now consider solar radiation incident on a surface element of area A_\odot , whose normal makes a *solar-incidence angle*, ϕ_{inc} , with the Sun-satellite line. The individual forces are (absorbed and reflected: specular and diffuse), and they occur in the normal \hat{n} and source incident \hat{s} directions.

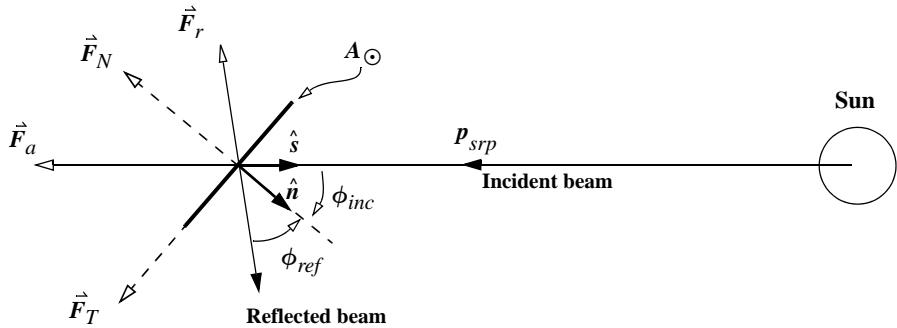


Figure 8-19. Incident Solar Radiation. The incident radiation produces one reflected beam resulting in reflected, F_r , and absorbed, F_a , forces. The reflected, ϕ_{ref} , and incident, ϕ_{inc} , angles are the same. We can also examine the interaction of the incoming radiation by looking at the normal, F_N , and tangential, F_T , force components. The satellite area that’s exposed to the Sun is depicted as a flat plate.

$$\begin{aligned}\vec{F}_{rs} &= -2p_{srp}c_{Rs}A_{\odot}\cos^2(\phi_{inc})\hat{n} \\ \vec{F}_a &= -p_{srp}c_{Ra}A_{\odot}\cos(\phi_{inc})\hat{s} \\ \vec{F}_{rd} &= -p_{srp}c_{Rd}A_{\odot}\cos(\phi_{inc})\left\{\frac{2}{3}\hat{n} + \hat{s}\right\}\end{aligned}$$

We can model diffuse and specular radiation forces by assuming a Lambertian diffusion (Luthcke et al. 1997). Summing the components of a macro-model

$$\vec{a}_{srp} = -\sum_{i=1} \frac{p_{srp}A_{\odot}c_{inc_i}}{m} \left\{ 2\left(\frac{c_{Rd_i}}{3} + c_{Rs_i}\cos(\phi_{inc_i})\right)\hat{n} + (1 - c_{Rs_i})\hat{s} \right\} \quad (8-45)$$

In this relation, we use the incoming solar pressure, p_{srp} , the angle between the surface normal and the incoming radiation, ϕ_{inc} , and the satellite mass, m . We also require the diffuse, c_{Rd} , and specular, c_{Rs} , reflectivities (Note that $c_{Ra} + c_{Rd} + c_{Rs} = 1.0$). The surface normal vector and the source (Sun) incidence vectors are needed for orienting the satellite. Macro-models approximate the shape and orientation of the satellite to more precisely find the acceleration. The summation adds each flat plate surface of the satellite model. The i -subscripts indicate that the area, angles, and coefficients will vary for each macro-model plate.

ROCK Solar Radiation Pressure Models

Well-known examples of the solar radiation pressure modeling discussed previously are the ROCK4 and ROCK42 models for GPS satellites (Fliegel et al. 1992). These models were originally developed by Rockwell (hence the name) and included methods to account for various surfaces on the satellites, and precise attitude. Because the programs were complex, Fourier series representations were developed to capture the essential behavior. Figure 8-20 shows the general orientation and nomenclature for the GPS body coordinate system. The angle formed by the Sun-satellite-Z axis is important with these models.

There are perturbing accelerations in both the X and Z directions. A y -bias also exists, which yields a smaller perturbing force. Several equations are relevant. Each block of GPS satellites have a specific set of equations. Standard (S_{10} and S_{20}) models match the AF implementation, and thermal models (T_{10} and T_{20}) also exist. The accelerations are as follows. The angle α is measured from the Sun to the Z -axis.

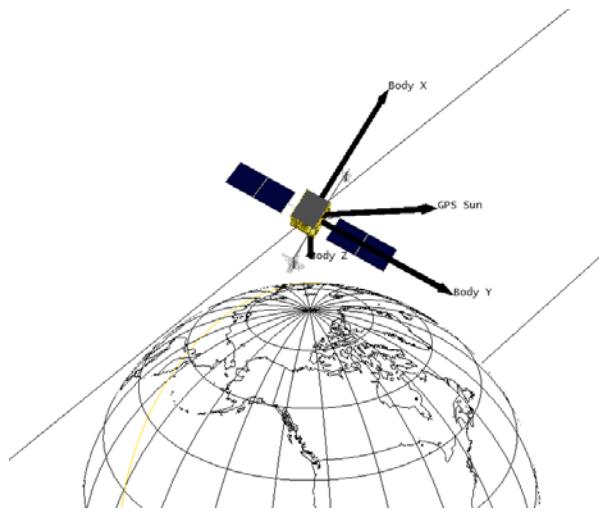


Figure 8-20. GPS body coordinate system. The orientation of a GPS satellite is shown with the satellite body coordinate system. The Z-axis points towards the Earth and the Y-axis lies along a solar panel arm, normal to the Sun-satellite vector. The X-axis completes a right handed coordinate system. (STK images courtesy of AGI).

$$S_{10}$$

$$X = -4.34 \sin(\alpha) + 0.1 \sin(2\alpha + 1.1) - 0.05 \cos(4\alpha) + 0.06$$

$$Z = -4.34 \cos(\alpha) + 0.17 \sin(2\alpha - 0.4) - 0.05 \sin(4\alpha) - 0.06$$

$$T_{10}$$

$$X = -4.55 \sin(\alpha) + 0.08 \sin(2\alpha + 0.9) - 0.06 \cos(4\alpha + 0.08) + 0.08$$

$$Z = -4.54 \cos(\alpha) + 0.20 \sin(2\alpha - 0.3) - 0.03 \sin(4\alpha)$$

(8-46)

$$S_{20}$$

$$X = -8.10 \sin(\alpha) + 0.05 \cos(2\alpha) - 0.056 \sin(4\alpha + 1.4) + 0.07$$

$$Z = -7.80 \cos(\alpha) + 0.024 \sin(2\alpha - 0.8) - 0.047 \sin(4\alpha + 0.9) + 0.02$$

$$T_{20}$$

$$X = -8.96 \sin(\alpha) + 0.16 \sin(3\alpha) + 0.10 \sin(5\alpha) - 0.07 \sin(7\alpha)$$

$$Z = -8.42 \cos(\alpha)$$

Earth Albedo and Infrared Radiation

Solar radiation that immediately reflects off the Earth back onto a satellite is called **albedo**. The amount of reflected radiation is about 30% of the incoming solar radiation. The remaining solar radiation is absorbed and re-emitted at a later time as infrared radiation, or **emissivity**. The infrared energy emanating at the Earth's surface (*IR*) is about 237 W/m^2 . Both effects can be measurable on some satellites. Unlike solar radiation

pressure, they are usually separated into specific wavelengths—each of which transfer through the atmosphere differently. We find the acceleration as before [Eq. (8-45)], but now we must first calculate the ground albedo and emissivity, and also consider each region of the Earth that is visible to the satellite, or a satellite plate model component.

The albedo and emissivity at the ground are shown by Knocke et al. (1988). They use a spherical harmonic expansion (limited to zeroth, first, and second degree, Table 8-2) to represent the functions worldwide. The generic expression is

$$\begin{aligned} p_{\text{radiation}} = & a_0 + [c_0 + c_1 \text{COS}(\omega(JD - t_o)) + c_2 \text{SIN}(\omega(JD - t_o))]P(\text{COS}(\phi)) + \\ & a_2 P(\text{SIN}(\phi)) \end{aligned}$$

where ϕ is the central latitude of the grid, and ω establishes the periodicity, $2\pi/365.25$.* The base epoch (t_o) is December 22, 1981 ($JD = 2,444,960.5$). They give expressions for albedo and emissivity.

$$p_{\text{alb}} = 0.34 + [0.10 \text{ COS}(\omega(JD - t_o))] \text{COS}(\phi) + 0.29 \text{ SIN}(\phi) \quad (8-47)$$

$$p_{\text{emiss}} = 0.68 + [-0.07 \text{ COS}(\omega(JD - t_o))] \text{COS}(\phi) - 0.18 \text{ SIN}(\phi)$$

To find the acceleration on a satellite, you must sum the accelerations induced on each component of a macro model, and from each grid location. Note the similarity with Eq. (8-45). For a single wavelength with albedo pressure (p_{alb})

$$\text{albedo} = -\sum_i \sum_j \frac{p_{\text{alb}} A_{ij} \text{COS}(\phi_{inc_{ij}})}{mc} \left\{ 2 \left(\frac{c_{Rd_i}}{3} + c_{Rs_i} \text{COS}(\phi_{inc_{ij}}) \right) \hat{n} + (1 - c_{Rs_i}) \hat{s} \right\}$$

The i components are the macro plate models, while the j indices represent grid points on the Earth. STK/HPOP documentation suggests that for calculations within about 2% of the actual magnitude on satellites above 200 km altitude, you can use a grid spacing on the Earth of about 10° , decreasing to about 1° below 300 km altitude. Note that the area, incident angle, and incoming radiation will vary for each Earth location (j) considered. Denser grids produce more accurate results, but at significantly higher processing times. You can account for the infrared radiation as an additional perturbing force using the same equation, and replacing p_{alb} with p_{emiss} .

A simple expression may be found from sample data in Wertz (1978:571). The albedo flux is $-197.65 \text{ LN}(h_{ellp}) + 1833.3 \text{ W/m}^2$, and the emissivity flux is $-39.451 \text{ LN}(h_{ellp}) + 391.44 \text{ W/m}^2$.

8.6.5 Other Perturbations

Many other forces affect a satellite's orbit, but most are very small and are often neglected. Tides are being modeled more often as computational resources expand and

* Note that longitude is not included as the variations for longitude are generally significantly less than those for latitude.

allow us to consider them. Thrust is a large perturbation, but unlike the other forms we've discussed, it's planned. Upper atmosphere winds are under investigation as accuracies have improved over the years (Sutton et al. 2006).

Tides

Although the first mention of tides is generally associated with ocean tides, the term also includes solid Earth tides, and to a lesser extent, pole tides, and atmospheric tides from air mass movement. Tides are complex and have only recently been studied in detail as computational and observational processes have become readily accessible. In general, tides are a result of a gravitational distortion caused by an external body—for the Earth, the Moon and Sun are the primary sources. The Earth rotation introduces periodicity into these effects.

We've seen that the orbits for the Sun and Moon are complex, and this requires us to consider many different periods in any 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.

Many sources define the tide generating potential (TGP) similar to the gravity potential we found in Eq. (8-21). The tidal potential is not directly observable, 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 . The Love numbers let us examine the Earth's response to perturbations from a frequency perspective. Although specific forms of equations differ, 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 deformations of the Earth due to perturbing forces of external gravitational attraction, particularly the Moon. They are generally the largest perturbing effects of tides discussed in this section. Internal forces also 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 beginning to be used routinely with numerical propagation. McCarthy and Petit (2003:59) show the basic formulae for implementation using corrections to the unnormalized gravitational coefficients. Note the use of the Love numbers ($k_{l,m}$) and summations for the Moon ($j = 2$) and Sun ($j = 3$).

$$\begin{aligned}\Delta C_{l,m} &= \frac{k_{l,m}}{2l+1} \sum_{j=2}^3 \frac{\mu_j}{\mu_{\oplus}} \left(\frac{R_{\oplus}}{r_j} \right)^{l+1} P_{l,m}[\sin(\phi_{gc_j})] \cos(m\lambda_j) \\ \Delta S_{l,m} &= \frac{k_{l,m}}{2l+1} \sum_{j=2}^3 \frac{\mu_j}{\mu_{\oplus}} \left(\frac{R_{\oplus}}{r_j} \right)^{l+1} P_{l,m}[\sin(\phi_{gc_j})] \sin(m\lambda_j)\end{aligned}\quad (8-48)$$

These corrections are determined for all values of m (0 to ℓ) for $\ell = 2$ and 3, and m (0 to 2) for $\ell = 4$. McCarthy and Petit (2003:60) list the nominal values. Note that for $\ell = 4$, alternate Love coefficients are used, and the Legendre polynomial has $\ell = 2$.

The centrifugal effect of the Earth's rotation also causes a solid Earth **pole tide**. A potential could be found, but the result is ultimately a change to the C_{21} and S_{21} coefficients. The parameters m_1 and m_2 are usually in arcseconds.

$$\Delta C_{21} = -1.333 \times 10^{-9} (m_1 - 0.0115 m_2)$$

$$\Delta S_{21} = -1.333 \times 10^{-9} (m_2 + 0.0115 m_1)$$

The tide generating potential has a zonal component ($\ell = 2$) that is non-zero. Thus, when using gravitational models such as JGM-3, you must first restore the permanent portion of the tidal effect in the model. See McCarthy and Petit (2003:67) for the procedure.

Ocean tides cause a large change in mass distribution as the water reacts to the various gravitational attractions. They are also the main reason for the differences between tidal theory and observational measurements. Casotto (1989) suggests that ocean tides are only about 10-15% of the magnitude of solid Earth tides. 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 perturbative effect of the oceans has been understood conceptually for quite some time; however, only recently have modern computers permitted realistic analysis of the perturbation. Developing correct tidal models is complicated because the resonant 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 ideal predictions of tidal models.

From McCarthy and Petit (2003:67), we find the corrections to the gravitational coefficients. Note the use of the density of seawater ($\rho_w = 1.025 \text{ gm/m}^3$), and the load deformation coefficients $k'_{t,m}$.

$$\begin{aligned}\Delta C_{t,m} &= \frac{4\pi R_\oplus^2 \rho_w}{m_\oplus} \left(\frac{1 + k'_{t,m}}{2t+1} \right) \sum_{j=1}^2 \frac{\mu_j}{\mu_\oplus} \left(\frac{R_\oplus}{r_j} \right)^{t+1} P_{t,m} \sin(\phi_{gc_j}) \cos(m\lambda_j) \\ \Delta S_{t,m} &= \frac{k'_{t,m}}{2t+1} \sum_{j=1}^2 \frac{\mu_j}{\mu_\oplus} \left(\frac{R_\oplus}{r_j} \right)^{t+1} P_{t,m} \sin(\phi_{gc_j}) \sin(m\lambda_j)\end{aligned}$$

A primary 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. Today, most measurements for tide models come from satellite altimeter data. A second major problem is our limited understanding of the world's oceans, especially the current effects and interactions at the coasts.

There is a coupling between solid Earth and ocean tides. The ocean tides cause mass redistribution, and a resultant loading on the crust. This causes a regional deformation to the crust. The Coriolis force of Earth rotation also induces a tidal deformation due to centrifugal acceleration. The exact dynamics of the Earth's core, mantle, and crust are not fully understood, but there are numerous wobbles that occur in each. Resonances occur in these wobbles, introducing additional phases to account in any response. Some of these periods range from hours to years. There is also a correlation with the nutation calculations discussed in Chap. 3.

Mathematically modeling tides usually involves a three-step process: (1) Calculate the tidal variations resulting from the Sun and Moon, (2) Update the corrections for smaller effects, and (3) Adjust for zero-tide solution if necessary.

Magnetic Field Effects

We get the field strength of the magnetic field, \vec{B} , if we want to calculate the perturbative effects of the charged particles on the satellite (such as magnetic torquing). It isn't needed for analyzing atmospheric density. Notice the use of a spherical harmonic representation of the magnetic-potential function, V . To be consistent with the geopotential function, I show the magnetic field as the positive gradient of the potential function.

$$\vec{B} = \nabla V$$

$$V = -R_{\oplus} \sum_{l=1}^k \left(\frac{R_{\oplus}}{r}\right)^{l+1} \sum_{m=0}^l \bar{P}_{l,m} [\cos(90^\circ - \phi_{gc})] \{ \bar{g}_{l,m} \cos(m\lambda) + \bar{h}_{l,m} \sin(m\lambda) \} \quad (8-49)$$

where the variables are as defined with the gravitational potential. The bars represent normalized parameters [as in Eq. (8-22)], which we'll discuss shortly. The *Schmidt functions*, $\bar{P}_{l,m}[\cos(90^\circ - \phi_{gc})]$, 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 8-6 approximate the *Schmidt coefficients*, $\bar{g}_{l,m}$ and $\bar{h}_{l,m}$. 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.

$$\begin{aligned} \Pi_{l,m}^{Schmidt} &= \sqrt{\frac{(2 - \delta_{0,m})(l-m)!}{(l+m)!} \frac{(2l-1)!!}{(l-m)!}} \\ \bar{g}_{l,m} &= \Pi_{l,m}^{Schmidt} g_{l,m} \quad \bar{h}_{l,m} = \Pi_{l,m}^{Schmidt} h_{l,m} \\ \delta_{l,m} &= 1 \text{ if } m = l \\ \delta_{l,m} &= 0 \text{ if } m \neq l \end{aligned}$$

TABLE 8-6. Gaussian Geomagnetic Coefficients. Example Schmidt coefficients are listed for a 13×13 magnetic field. Consider these values valid for 2010–2015. Appendix D lists Internet sites for current values.

l	m	$\bar{g}_{t,m}$	$\bar{h}_{t,m}$												
1	0	-29496.5	0.0	6	6	-77.9	54.9	9	8	-8.4	-6.1	12	1	-0.2	-0.8
1	1	-1585.9	4945.1	7	0	80.4	0.0	9	9	-10.1	7.0	12	2	0.3	0.3
2	0	-2396.6	0.0	7	1	-75.0	-57.8	10	0	-2.0	0.0	12	3	1.0	2.2
2	1	3026.0	-2707.7	7	2	-4.7	-21.2	10	1	-6.3	2.8	12	4	-0.7	-2.5
2	2	1668.6	-575.4	7	3	45.3	6.6	10	2	0.9	-0.1	12	5	0.9	0.5
3	0	1339.7	0.0	7	4	14.0	24.9	10	3	-1.1	4.7	12	6	-0.1	0.6
3	1	-2326.3	-160.5	7	5	10.4	7.0	10	4	-0.2	4.4	12	7	0.5	0.0
3	2	1231.7	251.7	7	6	1.6	-27.7	10	5	2.5	-7.2	12	8	-0.4	0.1
3	3	634.2	-536.8	7	7	4.9	-3.4	10	6	-0.3	-1.0	12	9	-0.4	0.3
4	0	912.6	0.0	8	0	24.3	0.0	10	7	2.2	-4.0	12	10	0.2	-0.9
4	1	809.0	286.4	8	1	8.2	10.9	10	8	3.1	-2.0	12	11	-0.8	-0.2
4	2	166.6	-211.2	8	2	-14.5	-20.0	10	9	-1.0	-2.0	12	12	0.0	0.8
4	3	-357.1	164.4	8	3	-5.7	11.9	10	10	-2.8	-8.3	13	0	-0.2	0.0
4	4	89.7	-309.2	8	4	-19.3	-17.4	11	0	3.0	0.0	13	1	-0.9	-0.8
5	0	-231.1	0.0	8	5	11.6	16.7	11	1	-1.5	0.1	13	2	0.3	0.3
5	1	357.2	44.7	8	6	10.9	7.1	11	2	-2.1	1.7	13	3	0.4	1.7
5	2	200.3	188.9	8	7	-14.1	-10.8	11	3	1.6	-0.6	13	4	-0.4	-0.6
5	3	-141.2	-118.1	8	8	-3.7	1.7	11	4	-0.5	-1.8	13	5	1.1	-1.2
5	4	-163.1	0.1	9	0	5.4	0.0	11	5	0.5	0.9	13	6	-0.3	-0.1
5	5	-7.7	100.9	9	1	9.4	-20.5	11	6	-0.8	-0.4	13	7	0.8	0.5
6	0	72.8	0.0	9	2	3.4	11.6	11	7	0.4	-2.5	13	8	-0.2	0.1
6	1	68.6	-20.8	9	3	-5.3	12.8	11	8	1.8	-1.3	13	9	0.4	0.5
6	2	76.0	44.2	9	4	3.1	-7.2	11	9	0.2	-2.1	13	10	0.0	0.4
6	3	-141.4	61.5	9	5	-12.4	-7.4	11	10	0.8	-1.9	13	11	0.4	-0.2
6	4	-22.9	-66.3	9	6	-0.8	8.0	11	11	3.8	-1.8	13	12	-0.3	-0.5
6	5	13.1	3.1	9	7	8.4	2.2	12	0	-2.1	0.0	13	13	-0.3	-0.8
l	m	dg/dt	dh/dt												
1	0	11.4	0.0	4	2	-8.9	3.2	6	2	-0.3	-2.1	7	6	-0.8	-0.3
1	1	16.7	-28.8	4	3	4.4	3.6	6	3	1.9	-0.4	7	7	0.4	0.2
2	0	-11.3	0.0	4	4	-2.3	-0.8	6	4	-1.6	-0.5	8	0	-0.1	0.0
2	1	-3.9	-23.0	5	0	-0.5	0.0	6	5	-0.2	0.8	8	1	0.1	0.0
2	2	2.7	-12.9	5	1	0.5	0.5	6	6	1.8	0.5	8	2	-0.5	0.2
3	0	1.3	0.0	5	2	-1.5	1.5	7	0	0.2	0.0	8	3	0.3	0.5
3	1	-3.9	8.6	5	3	-0.7	0.9	7	1	-0.1	0.6	8	4	-0.3	0.4
3	2	-2.9	-2.9	5	4	1.3	3.7	7	2	-0.6	0.3	8	5	0.3	0.1
3	3	-8.1	-2.1	5	5	1.4	-0.6	7	3	1.4	-0.2	8	6	0.2	-0.1
4	0	-1.4	0.0	6	0	-0.3	0.0	7	4	0.3	-0.1	8	7	-0.5	0.4
4	1	2.0	0.4	6	1	-0.3	-0.1	7	5	0.1	-0.8	8	8	0.2	0.4

Notice the Kronecker delta, $\delta_{l,m}$, and the double product in the formulation. Don't confuse this process with normalizing the gravitational coefficients as shown in Eq. (8-22). 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:

$$\begin{aligned}\Pi_{0,0}^{Schmidt} &= 1 \\ \Pi_{\ell,0}^{Schmidt} &= \Pi_{\ell-1,0}^{Schmidt} \frac{2\ell-1}{\ell} \quad \ell = 1, 2, \dots \\ \Pi_{\ell,m}^{Schmidt} &= \Pi_{\ell,m-1}^{Schmidt} \sqrt{\frac{(\ell-m+1)(\delta_{1,m}+1)}{\ell+m}} \quad m = 1, 2, \dots\end{aligned}$$

Notice these values are independent of satellite parameters, so we can calculate them once and place them in an external location.

South Atlantic Anomaly

The radiation arriving at the Earth is deflected by the Van Allen radiation belts which were discovered in 1958 (Van Allen et al. 1958) by James Van Allen (1914-2006). The belts are formed by the interaction of the particles with the magnetic field of the Earth. There are two primary belts—an inner one extending from about 1.1 to 3.3 ER, and an outer belt from about 3.3 to 4.0 ER. Figure 8-21 shows an approximate shape. A third belt was discovered in about 1990 from the Combined Release and Radiation Effects satellite (CRRES). Some sources suggest belts to 10.0 ER.

The South Atlantic Anomaly (SAA) is a dip in the Van Allen belts centered roughly over South America and the South Atlantic. Figure 8-22 shows the approximate location. Its altitude is about 250 km and varies over time.

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. 6. 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. 6. 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:

$$\begin{aligned}F_{thrust} &= I_{sp} \frac{dm}{dt} \quad \hat{a}_{thrust} = \frac{\vec{F}_{thrust}(t)}{m_c} \\ \Delta v &= g I_{sp} \text{LN} \frac{m_{init}}{m_{init} - m_{prop}}\end{aligned}\tag{8-50}$$

Now, find the vehicle's current mass, m_c , using

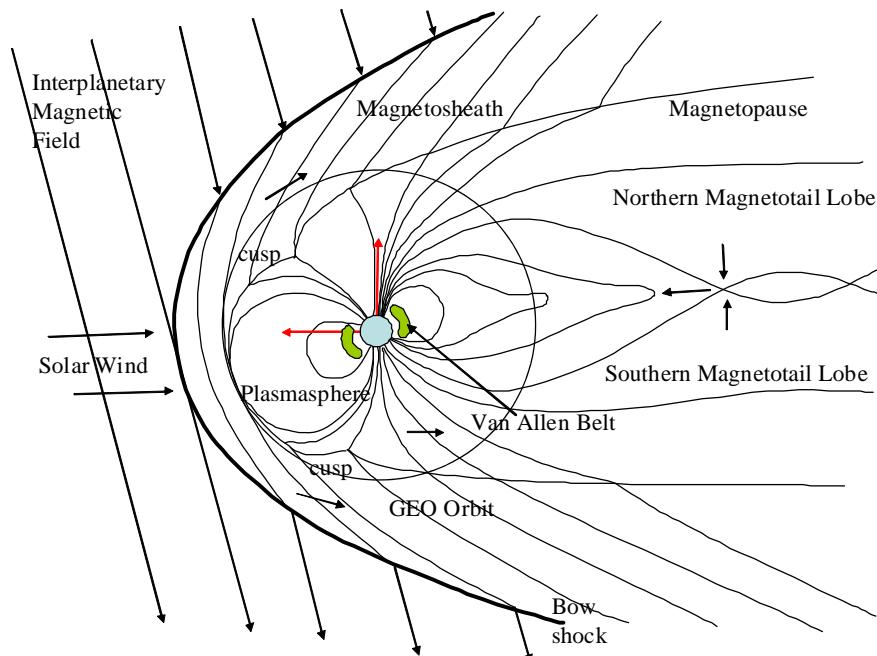


Figure 8-21. Van Allen Radiation Belts. The Earth's magnetic field distorts the incoming solar wind (radiation) from the Sun. The Van Allen radiation belts trap high energy particles and are located in a doughnut shape around the earth from about 1.1 to 2.0 ER, and 3.3 to 4.0 ER. The effect on satellites is primarily radiation. The figure is approximately to scale, and the GEO orbit is included for reference only. The basic figure is adapted from http://www.astro.phys.ethz.ch/theses/berthoud/main.htm#title_the_solar_wind, Dec 20, 2006.

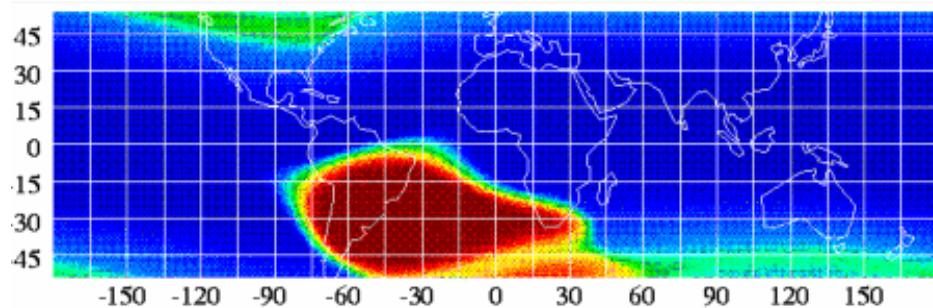


Figure 8-22. South Atlantic Anomaly. The approximate location of the SAA is over South America and the South Atlantic. The shape is variable over time, and is located at an average altitude of approximately 250 km. Source http://image.gsfc.nasa.gov/docs/ask_astro/answers/961004.html, Dec 20, 2006.

$$m_c = m_o - \int_{t_o}^{t_c} \frac{dm}{dt}(t) dt \quad (8-51)$$

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.

8.7 Forming Numerical Solutions

Special perturbations (a term 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 original model of 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 [Algorithm 64] 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. The vectors may be in any coordinate system, but typically, integration is done in ECI and force model evaluations are done in ECEF. The algorithm contains some shorthand—e.g., τ represents all the time parameters required for reduction operations, and the position and velocity vectors used for all acceleration calculations are in ECEF.

ALGORITHM 64: NUMERICAL INTEGRATION ($\vec{r}_o, \vec{v}_o, \tau_{start}, \tau_{stop}, \Delta\tau \Rightarrow \vec{r}, \vec{v}$)

$t = \tau_{start}$

LOOP

FK5($\vec{r}_{ECP}, \vec{v}_{ECP}, \vec{a}_{ECP}, yr, mo, day, UTC, \Delta UT1, \Delta AT, x_p, y_p \Rightarrow$
 $\vec{r}_{ECEF}, \vec{v}_{ECEF}, \vec{a}_{ECEF}$)

$$\begin{aligned}
U &= \frac{\mu}{r} \left[1 + \sum_{l=2}^{\infty} \sum_{m=0}^l \left(\frac{R_{\oplus}}{r} \right)^l P_{l,m} [\sin(\phi_{gc_{sat}})] \times \right. \\
&\quad \left. \{ C_{l,m} \cos(m\lambda_{sat}) + S_{l,m} \sin(m\lambda_{sat}) \} \right] \\
\hat{a}_{nonspherical} &= \nabla(U - U_{2-body}) = \nabla \left(U - \frac{\mu}{r} \right) = \nabla R \\
\hat{a}_{drag} &= -\frac{1}{2} \rho \frac{c_D A}{m} v_{rel}^2 \frac{\hat{v}_{rel}}{|\hat{v}_{rel}|} \\
\hat{a}_{3-body} &= \mu_3 \left(\frac{\hat{r}_{sat3}}{r_{sat3}^3} - \frac{\hat{r}_{\oplus3}}{r_{\oplus3}^3} \right) \\
\hat{a}_{srp} &= -\frac{p_{srp} c_R A_{\odot}}{m} \frac{\hat{r}_{sat\odot}}{|\hat{r}_{sat\odot}|} \\
\hat{a} &= -\frac{\mu}{r^3} \hat{r}_{ECEF} + \hat{a}_{nonspherical} + \hat{a}_{drag} + \hat{a}_{3-body} + \hat{a}_{srp} + \hat{a}_{tides} + \hat{a}_{other}
\end{aligned}$$

$$\begin{aligned}
&FK5(\hat{r}_{ECEF}, \hat{v}_{ECEF}, \hat{a}_{ECEF}, yr, mo, day, UTC, \Delta UT1, \Delta AT, x_p, y_p \Rightarrow \\
&\hat{r}_{ECI}, \hat{v}_{ECI}, \hat{a}_{ECI})
\end{aligned}$$

$$t = t + \Delta t$$

Numerically integrate (Runge-Kutta, Gauss Jackson, etc.)

UNTIL $t > \tau_{stop}$

Although this algorithm looks innocuous, its solution continues to challenge the greatest intellects. To account for all the major perturbations (central body, atmospheric drag, third body, solar radiation, thrust, other), we use expressions for the accelerations developed earlier in this chapter.

We can also numerically integrate the variation of parameter equations I'll introduce in Chap. 9 [Eq. (9-12), Eq. (9-14), and Eq. (9-24)]. This process is really a special perturbation solution, but I've kept it with the discussion in Chap. 9 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 of Algorithm 64. The simplified approach is useful for academic applications and it illustrates the general process. The technique is quite easy to adapt to the Runge-Kutta methods discussed in Sec. 8.5. The more complex method is a brief sketch of the steps required to fully implement Algorithm 64. Actually forming a complex numerical integration routine can take on several forms depending on the number of desired force models.

8.7.1 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 possible approximations.

It's possible to simplify the coordinate system transformations for certain applications, although ignoring the coordinate system altogether is not recommended. A common simplification is changing between just the PEF and TOD frames. This introduces additional error into the solution because the accelerations are applied at slightly different locations of the orbit, but for low-precision applications, this may be acceptable.

The central-body acceleration is simplified by excluding all tesseral terms (and therefore sectoral terms) in the aspherical potential and examining only the first few zonal harmonics. Although *PKEPLER* (Algorithm 65) does this analytically (for first-order J_2 secular effects), numerical integration provides the means to obtain the secular *and* periodic contributions of these zonal harmonics.

From the aspherical potential [Eq. (8-18)], we use the gradient to determine the 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 section beginning on page 549. 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 ITRF frame, 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

$$\begin{aligned}
 a_I &= \frac{\partial R_2}{\partial r_I} = -\frac{3J_2\mu R_{\oplus}^2 r_I}{2r^5} \left(1 - \frac{5r_K^2}{r^2}\right) \\
 a_J &= \frac{\partial R_2}{\partial r_J} = -\frac{3J_2\mu R_{\oplus}^2 r_J}{2r^5} \left(1 - \frac{5r_K^2}{r^2}\right) \\
 a_K &= \frac{\partial R_2}{\partial r_K} = -\frac{3J_2\mu R_{\oplus}^2 r_K}{2r^5} \left(3 - \frac{5r_K^2}{r^2}\right)
 \end{aligned} \tag{8-52}$$

The perturbations resulting from the third harmonic are developed from the disturbing potential resulting when $l=3$, so

$$R_3 = -\frac{J_3\mu}{2r} \left(\frac{R_{\oplus}}{r}\right)^3 \{5\sin^3(\phi_{gc}) - 3\sin(\phi_{gc})\}$$

Similar to J_2 , the accelerations due to J_3 are

$$\begin{aligned}
 a_I &= \frac{\partial R_3}{\partial r_I} = -\frac{5J_3\mu R_{\oplus}^3 r_I}{2r^7} \left(3r_K - \frac{7r_K^3}{r^2}\right) \\
 a_J &= \frac{\partial R_3}{\partial r_J} = -\frac{5J_3\mu R_{\oplus}^3 r_J}{2r^7} \left(3r_K - \frac{7r_K^3}{r^2}\right) \\
 a_K &= \frac{\partial R_3}{\partial r_K} = -\frac{5J_3\mu R_{\oplus}^3}{2r^7} \left(6r_K^2 - \frac{7r_K^4}{r^2} - \frac{3}{5}r^2\right)
 \end{aligned} \tag{8-53}$$

Express the perturbations resulting from the fourth harmonic, $l=4$, as

$$R_4 = -\frac{\mu J_4}{8r} \left(\frac{R_{\oplus}}{r}\right)^4 \{35\sin^4(\phi_{gc}) - 30\sin^2(\phi_{gc}) + 3\}$$

and the accelerations as

$$\begin{aligned}
 a_I &= \frac{\partial R_4}{\partial r_I} = \frac{15J_4\mu R_{\oplus}^4 r_I}{8r^7} \left(1 - \frac{14r_K^2}{r^2} + \frac{21r_K^4}{r^4}\right) \\
 a_J &= \frac{\partial R_4}{\partial r_J} = \frac{15J_4\mu R_{\oplus}^4 r_J}{8r^7} \left(1 - \frac{14r_K^2}{r^2} + \frac{21r_K^4}{r^4}\right) \\
 a_K &= \frac{\partial R_4}{\partial r_K} = \frac{15J_4\mu R_{\oplus}^4 r_K}{8r^7} \left(5 - \frac{70r_K^2}{3r^2} + \frac{21r_K^4}{r^4}\right)
 \end{aligned} \tag{8-54}$$

Escobal ([1965] 1985:50–51) shows the partials for the J_5 and J_6 zonal harmonics:

$$\begin{aligned} a_I &= \frac{\partial R_5}{\partial r_I} = \frac{3J_5\mu R_\oplus^5 r_I r_k}{8r^9} \left(35 - 210 \frac{r_k^2}{r^2} + 231 \frac{r_k^4}{r^4} \right) \\ a_J &= \frac{\partial R_5}{\partial r_J} = \frac{3J_5\mu R_\oplus^5 r_J r_k}{8r^9} \left(35 - 210 \frac{r_k^2}{r^2} + 231 \frac{r_k^4}{r^4} \right) \\ a_K &= \frac{\partial R_5}{\partial r_K} = \frac{3J_5\mu R_\oplus^5 r_k^2}{8r^9} \left(105 - 315 \frac{r_k^2}{r^2} + 231 \frac{r_k^4}{r^4} \right) - \frac{15J_5\mu R_\oplus^5}{8r^7} \end{aligned} \quad (8-55)$$

$$\begin{aligned} a_I &= \frac{\partial R_6}{\partial r_I} = -\frac{J_6\mu R_\oplus^6 r_I}{16r^9} \left(35 - 945 \frac{r_k^2}{r^2} + 3465 \frac{r_k^4}{r^4} - 3003 \frac{r_k^6}{r^6} \right) \\ a_J &= \frac{\partial R_6}{\partial r_J} = -\frac{J_6\mu R_\oplus^6 r_J}{16r^9} \left(35 - 945 \frac{r_k^2}{r^2} + 3465 \frac{r_k^4}{r^4} - 3003 \frac{r_k^6}{r^6} \right) \\ a_K &= \frac{\partial R_6}{\partial r_K} = -\frac{J_6\mu R_\oplus^6 r_K}{16r^9} \left(245 - 2205 \frac{r_k^2}{r^2} + 4851 \frac{r_k^4}{r^4} - 3003 \frac{r_k^6}{r^6} \right) \end{aligned} \quad (8-56)$$

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 \vec{v}_{rel} is the same as the velocity of the satellite. The result is an acceleration which computes quickly but can sacrifice considerable accuracy.

$$\vec{a}_{drag} = -\frac{1}{2}\rho \frac{c_D A}{m} v_{rel}^2 \frac{\vec{v}_{rel}}{|\vec{v}_{rel}|}$$

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. Unfortunately, some of these files can be very large. Several alternatives exist. First, we can calculate the position vectors of each body from analytical techniques presented in Sec. 5.1.1 and Sec. 5.2.1. 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.

Locking the position of the Sun or Moon—or any other time varying quantity—can introduce effects in the resulting calculations. 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 sys-

tems 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.

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 the previous discussions in Sec. 5.1.1 and Sec. 8.6.4). 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. A common approach assumes a spherical shape with no diffuse reflection. Simplified approaches also don't address shadowing often—a fact that slightly degrades accuracy.

8.7.2 Application: Complex Acceleration Model

Forming a complete, highly accurate, special perturbation method is a complex activity. With care, patience, and discipline, you can develop an extremely flexible and accurate tool. The equations are in Algorithm 64 just as for the simplified approach. The difference is that we rigorously determine each acceleration.

The first task is to determine the type of integration to perform. The variation of parameter solution is useful when analysis of the orbital elements and computational speed are important. Conventional integration processes the state vectors directly and is common in many flight dynamics programs. However, due to the additional processing of each complete force model, the computational requirements can be longer than the VOP approach.

The coordinate frame is very important in highly accurate programs and it requires correct EOP data (see Sec. 8.8.2). You usually integrate the equations of motion in an inertial frame, like the GCRF, which would allow you to avoid the complexity of having to account for rotating coordinates as discussed in Chap. 3 but other possibilities exist as in older programs that use the IAU-76/FK5, or TOD frames. These may not be desirable though for longer simulations because the true-of-date system continually moves. When applying the accelerations, remember that the satellite sees the “true” orientation of the Earth over time. The natural choice for coordinate systems is the Earth fixed (ITRF) system when you evaluate the accelerations. Therefore, you must use a coordinate frame that's consistent with where the satellite is at each instant of time. The pseudo-Earth fixed frame (PEF) is sometimes used for computational speed.

To start, use the aspherical potential in calculating high-precision ephemerides. Recall the form of the aspherical potential from Eq. (8-19). 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 small differences into the solution. Vallado (2005) showed that if truncation of the gravitational potential is necessary, additional zonal harmonic terms should be retained to improve the accuracy over simple ‘square’ truncations that are sometimes discussed.

Now consider the calculations for the associated Legendre functions. The equation [Eq. (8-16)] 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. (8-57) uses simple addition and multiplications instead of more costly trigonometric terms. You should see from Eq. (8-16) that subsequent values depend on previously computed values. Long et al. (1989:4–13) give specific relations:

$$\begin{aligned} P_{t,0}[\gamma] &= \frac{(2t-1)\gamma P_{t-1,0}[\gamma] - (t-1)P_{t-2,0}[\gamma]}{t} & t \geq 2 \\ P_{t,m}[\gamma] &= P_{t-2,m}[\gamma] + (2t-1)\cos(\phi_{gc})P_{t-1,m-1}[\gamma] & m \neq 0, \quad m < t \\ P_{t,t}[\gamma] &= (2t-1)\cos(\phi_{gc})P_{t-1,t-1}[\gamma] & t \neq 0 \end{aligned} \quad (8-57)$$

with the starting values of $[\cos(\phi_{gc})]$ can be replaced by $P_{1,1}[\gamma]$ in Eq. (8-57)]

$$\begin{aligned} P_{0,0}[\gamma] &= 1 \\ P_{1,0}[\gamma] &= \gamma = \sin(\phi_{gc}) \\ P_{1,1}[\gamma] &= \cos(\phi_{gc}) \end{aligned}$$

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 $P_{t-2,m}[\gamma]$. We sum t from 2 to the order desired, and then m goes from 0 to t . Remember to initialize each calculation with $P_{0,t-1} = 0$. 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. (8-57)] to find a smaller subset of the gravity field (say 12×12) and then use Eq. (8-16) to find the individual terms responsible for the resonance.

Notice that the matrix of associated Legendre functions can be calculated *a priori* to use with Eq. (8-12), but you can't perform this operation directly if you use normalized coefficients. In this case, you must store two matrices because $P_{t,m}$ and $P_{t,m+1}$ both multiply by the same gravitational coefficients.

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. 2.2.5). 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 recursion expressions for the latitude and longitude terms:

$$\begin{aligned}\sin(m\lambda) &= 2\cos(\lambda)\sin\{(m-1)\lambda\} - \sin\{(m-2)\lambda\} \\ \cos(m\lambda) &= 2\cos(\lambda)\cos\{(m-1)\lambda\} - \cos\{(m-2)\lambda\} \\ m\tan(\phi_{gc}) &= (m-1)\tan(\phi_{gc}) + \tan(\phi_{gc})\end{aligned}\quad (8-58)$$

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. Wright and Tanygin (2004) show the importance of interpolating the solar weather parameters and Vallado and Kelso (2005) show how to interpolate the geomagnetic and solar flux parameters. In fact, interpolation differences can sometimes exceed model differences. 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 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 Harris-Priester (page 568), which attempts to model certain tendencies during discrete intervals of the solar cycle. Schatten's (1988, 1993, etc.) 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.

The relative-velocity calculation is:

$$\hat{\vec{v}}_{rel} = \hat{\vec{v}}_{GCRF} - \hat{\vec{\omega}} \times \hat{\vec{r}}_{GCRF} + \hat{\vec{v}}_{wind}$$

For ease of implementation, all the vectors should be in an inertial coordinate system (I show GCRF here and recommend it). Accurate drag analysis (and solar radiation analysis) requires us to know the mass and 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 estimated 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. They're especially useful for atmospheric drag and solar-radiation calculations. Luthcke et al. (1997) used a macro model for TDRS to improve the orbit determination accuracy to a few meters. Knowing 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 the JPL recovery technique to recreate the planetary ephemerides as closely as possible. Of course, we must use planetary ephemerides (see Sec. 5.3.1) consistent with the inertial frame we use for integrating the equations of motion.

Lastly, the equations in Sec. 8.6.4 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 projection of 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 density models and for the third body. The acceleration due to solar radiation is also very sensitive to the time in which the Earth eclipses the satellite. A cylindrical shadow model is common (see page 299). 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.

8.8 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 several 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. 10, we'll concentrate on propagator-related issues in this section.

8.8.1 Verifying and Validating the Propagator

Perhaps the largest misconception in evaluating methods of propagation is what represents *truth*. We ***verify*** to existing data or results (ensure it's coded properly), while we ***validate*** that the model accurately reflects the truth (consistent and reliable results). Determining the “true” trajectory is ultimately tied to differential correction and involves the discussions of Sec. 10.4.2 and Sec. 10.9. It usually uses either existing or actual observational data. Although each type of data has advantages, both also have important drawbacks we must address. We must consider *both* types to completely ver-

ify and validate a propagation routine. Also important are precision and accuracy. **Precision** is the fineness (or reproducibility) with which a measurement is made while **accuracy** is how close the measurement is with respect to truth. These terms are sometimes confused.

First, let's introduce some terms that allow us to discuss how to verify and validate propagation algorithms. Because any propagator can coincidentally produce the 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). POE's are available for a few satellites such as TOPEX, LAGEOS, ETALON 1 and 2, GPS, TDRSS, and a few others. You should contact the appropriate organization to determine their availability. For instance, the Center for Space Research at the University of Texas at Austin produces POE's for TOPEX. See Appendix D for additional information.

When we use data to test a propagator, there are two options. If we use the same program, for instance during a system upgrade, we can use existing test cases and benchmarks to check our results (verification). 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 check. A necessary extension seeks an independent program to more fully verify the results. This technique overcomes the "familiarity" of the original program, and allows an independent evaluation of the propagation routine. Excellent examples of independent benchmarks are the sample orbits documented in Metzinger (1993) and Chao et al. (2000). The first document gives various orbits and their expected output from several different computing platforms. The propagation algorithm used by this study is the Draper version of the **Goddard Trajectory Determination System** (GTDS) with complete force models. GTDS is comparable with other legacy programs that I list in Appendix D (Vallado, 2005) and it is capable of producing results to well under a meter of the absolute truth. Because this document has been updated over many years, it contains a wealth of information for any validation effort. Be careful to use the same data (coordinate systems, gravitational constants, atmospheric parameters, etc.) when using different programs. It's especially difficult to ensure atmospheric data have identical actual and predicted values.

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 testing involves tests with real-world data and satellites (verification). 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. 10.9), this process isn't perfect. In fact, a review of Chap. 10 could lead us to think the truth is unachievable. Fortunately, this isn't the case and many satellite orbits (TOPEX, for instance) are known to centimeter accuracy (*post-processing*).

8.8.2 Physical Data and Sources

Collecting physical data for use with any type of perturbation analysis ensures consistent results. 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 (*post-processing*). However, we often need predicted values of unavailable current or future data to plan satellite missions (prediction). The accuracy of the predicted data strongly affects the final results. See Appendix D.5, "Data Sources," for additional information.

Gravitational Parameters

We get gravitational data from three main sources: measurements taken at the Earth's surface (terrestrial), altimetry (height above the Earth), 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 usually smaller. Other models also include the *Joint Gravity Model* (JGM-3, 70×70), developed through collaboration between NASA/Goddard and the University of Texas at Austin (Nerem et al. 1994). The JGM models updated NASA's *Goddard Earth Model* (GEM-T3) and have improvements using *International Earth Rotation Service* (IERS) constants, analysis of long-wavelength data, and TOPEX altimetry data. The Defense Mapping Agency's *World Geodetic Survey* (WGS-84, 41×41) has several different dates of release (1960, 1966, 1972, 1984). Each revision is referenced as WGS-##, with the appropriate year inserted. There are separate models from the University of Texas (UT/CSR TEG3), the NASA *Goddard Earth Model* (GEM), and others. A larger and combined model (EGM-96, 360×360) was published 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 at least a 50×50 field. Increasing the field size isn't always enough. For highly accurate missions, we usually need to consider other force models. Table D-1 and Table D-2 in Appendix D show representative coefficient values from the EGM-96 model.

The development of gravitational models over the last few decades shows a number of different models (Fig. 8-23). That's because it's hard to determine absolute gravitational coefficients, and having to resolve multiple data sources and satellite classes makes the process even more difficult (Table 8-7). Vetter (1994) concisely recounts the development of most commonly used gravitational models, whose use rests on two important facts:

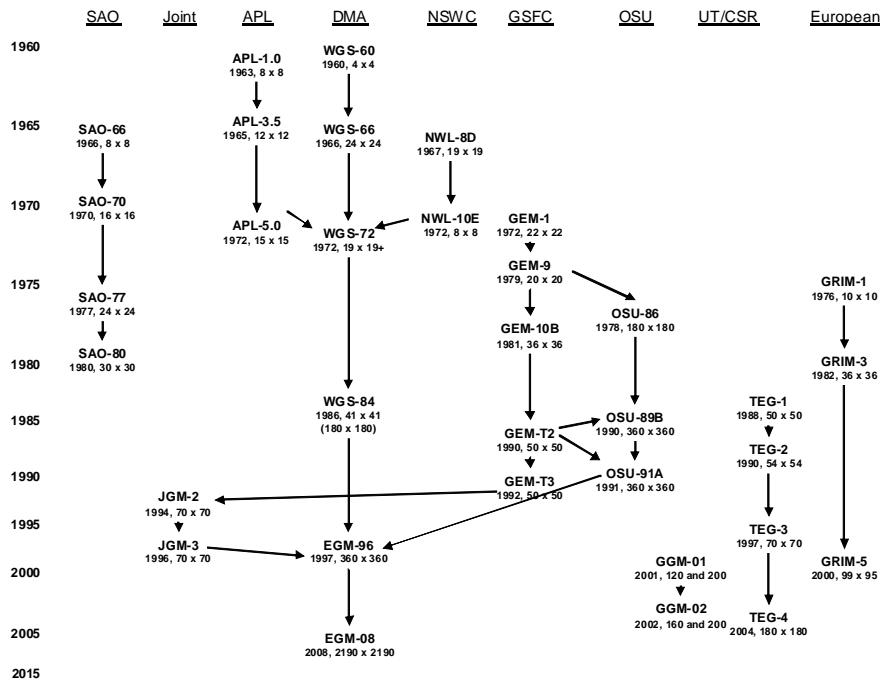


Figure 8-23. Development of Earth Gravity Models. The *Joint Gravity Models* (JGM) come from Goddard Space Flight Center (GSFC), Ohio State University (OSU), University of Texas at Austin (UT), and the European communities. The *Earth Gravity Model* (EGM) combines the JGM work with Defense Mapping Agencies efforts. The *Goddard Earth Models* (GEM) were produced annually beginning with GEM-1 in 1972. Even numbered models contain satellite and surface gravity data. Odd numbered models contain only satellite data. The Grace Gravity models (GGM-02) are derived solely from satellite data. *Standard Earth* (SAO) and *Applied Physics Laboratory* (APL) models were among the first models. The basic information is from Vetter (1994) and (2001).

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 back cover, for each model, 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

TABLE 8-7. Current Gravitational Model Data Sources. This table lists the major sources of data to create each gravity field (Vetter, 1994, 2001). The data types consist of Optical, Laser, Doppler, Range Rate (RR), Satellite to satellite tracking (S2), GPS, and DORIS information. Some models include satellite only, or combined results.

Model	Data Types	Perigee Alt	Inclination	Surface	Altimeter	# Sats
		km	°	Gravity		
JGM-3	O, L, D, RR, S2	600-2000	1-144	1 × 1 °MGA	Seasat, Geosat	30
	SLR, DORIS	5900, 35,000			GEOS-3, SPOT-2	
WGS-84	L, D, GPS	800-1100 5700, 20,500	50-90	1 × 1 °MGA	Seasat	10
EGM-96	All types	600-2000 5900, 35,000	1-144	1 × 1 °MGA 30 × 30' MGA	ERS-1, Geosat, TOPEX	30+
GEM T3	O, L, D, RR, S2	600-2000 5900, 35,000	1-144	1 × 1 °MGA	Seasat, Geosat	30
					GEOS-3	
OSU 91A	O, L, D, RR, S2	600-2000	1-144	1 × 1 °MGA	TOPEX	30
TEG-3	L, D, RR	780-1600 5900, 19200	15-108	1 × 1 °MGA	Seasat, Geosat	30
					GEOS-3	
GRIM 5	O, L, D	19,200	15-115	1 × 1 °MGA	SPOT-2	32

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) recommended the JGM-2 (Nerem et al. 1994) and the newer JGM-3 models because they're derived from many orbit classes. The EGM-96 effort is developed 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 is a complete 360×360 field. Any of these gravity fields may be *tuned* for a particular orbit, although tuning is uncommon because it's complex.

Earth Orientation Parameters

Earth orientation parameters let us precisely model the orientation of the Earth at a particular instant of time. First we need the precise location of the North Pole. Recall from page 223 that the pole precesses and nutates slightly over time. Highly accurate calculations must include corrections for polar motion. The data consists of x_p and y_p displacements from the **Celestial Intermediate Pole** (CIP), defined as the axis of rotation which is normal to the true equator of the Earth. The x_p and y_p angles represent displacements from the 0° and 270° longitude meridians, respectively. These coordinates are available from several sources.

The effect of using the EOP parameters is small but noticeable. Precise operations always use the proper EOP values. The EOP data should be interpolated to recover the

sub-diurnal values in UT and polar motion. These values are removed during processing, but should be re-added during operations. The values are valid at 0 hours, except the average values which are valid at noon. It can be particularly challenging to assemble a complete and consistent set of values for use in operations because the recent data is updated at irregular intervals.

Next, we must convert time. To convert between time systems (Algorithm 16), 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 Institute for Standards and Technologies (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 computers. See Sec. 3.5.4 for additional information.

NGA produces coefficients that will reproduce about one week of data for the polar motion coefficients, and the timing corrections. They are reasonably accurate. NGA re-introduced inclusion of zonal tide corrections into the data beginning with week 2 of 2005. This may change the accuracy slightly.

The EOP parameters, historical data, and analysis were discussed in detail in Vallado and Kelso (2005). That information is summarized in App. D.5, “Data Sources”.

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 solar flux measurements are taken at the Dominion Radio Astrophysical Observatory in Penticton, British Columbia. 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 (NGDC) 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. The solar flux data is generally adjusted to an Earth distance of 1.0 AU, however observed values are also required by some atmospheric models. Vallado and Kelso (2005) also studied the various sources of data and show how to produce a continuous file of data for use in operations. Their data is also available on the internet as shown in Appendix D.5, “Data Sources”. For studies discussing the accuracy of predictions, 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 are in Standish (1990:252–271). JPL creates these ephemerides by numerical inte-

grations accounting for the influence of the planets. The numerical integration uses a variable step-size that averages about 0.3 days. Once the ephemerides are created, the results are fit with Chebyshev polynomials and collected into 32-day blocks. For DE-245 and DE-405, each polynomial set covers 16 days for Venus, the Earth, and the Sun, 4 days for the Moon, 8 days for Mercury and the librations, and 32 days for the remaining objects. The general accuracy of the ephemerides is about 0.01". The Moon ephemerides are accurate to about 2 m (0.001") and the Sun is accurate to about 200 m (0.0003"). Long et al. (1989:3-84 to 3-88) describes how to retrieve the Chebyshev polynomials.

TABLE 8-8. Available Planetary and Lunar Ephemerides. The table shows available ephemerides and when they apply. Lunar ephemerides (LE-###) are produced with each development ephemerides. The nutations are the $\Delta\Psi$ and $\Delta\epsilon$ values for the IAU 1980 theory of nutation, and the librations are the 3 Euler velocities obtained by interpolating the Moon's position vector.

Designation	Basis	Year	Start	Stop	Comments
DE-102	FK4	1977	Apr 16, 1410	Dec 22, 3002	Long span
DE-118	FK4	1980	Dec 9, 1599	Mar 31, 2169	Previous standard
DE-200	FK5	1982	Dec 9, 1599	Mar 31, 2169	Includes nutations
DE-245	FK5	1993	Jun 16, 1750	Aug 1, 2051	Previous standard
DE-405	ICRF	1997	Dec 9, 1599	Dec 31, 2199	Includes nutations & librations
DE-406	ICRF	1997	Feb 23, -3000	May 6, 3000	Long span
DE-418	ICRF	2007	1899	2051	for New Horizons
DE-422	ICRF	2009	Feb 23, -3000	May 6, 3000	Incorporates new satellite data

There are also ephemerides available from NASA's Navigation and Ancillary Information Facility (NAIF). The so-called SPICE files (spacecraft, planet, instrument, c-matrix, events) are a collection of data and programs to aid navigation and trajectory solutions. The ephemerides come in variable and uniform time steps, as well as spice kernel (SPK), binary (BSP), and planetary constant kernel (PCK) formats. See <http://naif.jpl.nasa.gov/naif/about.html> for additional information.

IAU Resolution #3 (July 2006) recommends that TDB may be used for T_{Eph} as the independent time argument for the planetary ephemerides (DE-405).

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. 8.2? 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 Appendix D.5.
3. Develop coefficients for an eighth-order Runge-Kutta, Adams-Basforth, and Gauss-Jackson numerical integrator. Test the resulting propagators on a Molniya orbit.
4. Derive the form of third-body acceleration in Eq. (8-36). Also show that Eq. (8-36) is a simplified form of Eq. (8-37).
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.

$$\hat{r} = -26,175.1034 \hat{i} + 12,757.0706 \hat{j} + 14,626.6556 \hat{k} \text{ km}$$

$$\hat{v} = 2.376\,641 \hat{i} + 0.139\,677 \hat{j} + 2.078\,097 \hat{k} \text{ km/s}$$

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 an Earth Resources Technology mission. 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 data references the ECI (IAU-76/FK5) frame. The orbital data is as follows:

$$\hat{\vec{r}} = -2703.635\,550 \hat{i} + 1965.968\,884 \hat{j} - 6954.830\,225 \hat{k} \text{ km}$$

$$\hat{\vec{v}} = -5.779\,958\,666 \hat{i} - 4.125\,847\,273 \hat{j} + 1.079\,893\,238 \hat{k} \text{ km/s}$$

Epoch = December 8, 1996, 00:00:00.00 UTC

Research and locate the POE information for TOPEX after the week of interest and compare your answers.

9. a. In Sec. 3.3, we introduced the Solar System, and Earth barycenters. Draw a figure to show the magnitude and orientation of the barycenter locations on August 21, 2004 at 00:00 TDB.
 b. Use the following data from DE200 for August 21, 2004 at 00:00 TDB ($JD_{TDB} = 2453238.5$) to find the vector locations for the Sun and Earth barycenters from the center of each object. All units are in km.

Earth Center — Jupiter		
$\hat{\vec{r}}_I$	$\hat{\vec{r}}_J$	$\hat{\vec{r}}_K$
-941,461,060.3967	107,142,732.9047	66,072,182.1543
Sun Center — Jupiter		
-812,813,500.2447	34,019,092.0896	34,370,010.4158
Solar System Barycenter — Jupiter		
-812,196,708.2857	33,853,712.3193	34,283,473.3035
Earth-Moon Barycenter — Jupiter		
-941,456,961.8638	107,144,697.0723	66,072,980.4147
Sun Center — Earth		
128,647,560.1520	-73,123,640.8150	-31,702,171.7385
Sun Center — Earth-Moon Barycenter		
128,643,461.6191	-73,125,604.9826	-31,702,969.9989
Solar System Barycenter — Earth		
129,264,352.1110	-73,289,020.5854	-31,7887,08.8509
Solar System Barycenter — Earth-Moon Barycenter		
129,260,253.5782	-73,290,984.7530	-31,789,507.1112
Earth Center — Moon		
337,311.6606	161,652.1478	65,697.2955

- c. Does your answer in part *a* make sense in light of your answers to part *b*?

608

SPECIAL PERTURBATION TECHNIQUES

8.8

CHAPTER 9 GENERAL PERTURBATION TECHNIQUES

- 9.1 Historical Background
- 9.2 Introduction
- 9.3 Variation of Parameters
- 9.4 Hamilton's Formulation
- 9.5 Disturbing-Potential Formulations
- 9.6 Linearized Perturbations and Effects
- 9.7 Forming Analytical Solutions
- 9.8 Semianalytical Solutions
- 9.9 Practical Considerations
- 9.10 Summary of Perturbation Effects

9.1 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 precisely 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 (1926–2006) 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 Nep-

tune, 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)

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) 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. Vallado et al. (2006) recently updated this report, code, and test cases.

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 several technical advances.

The next decade saw a unique semianalytical theory from a team of scientists led by Paul Cefola (1942–). Remarkably, one of the 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. Studies (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.

9.2 Introduction

This chapter develops techniques that allow us to examine the effect of perturbations on the *orbital elements*. The numerical techniques of Chap. 8 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 orbital applications, its documentation, and the fidelity and the 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 Ex. 8-1, page 520, 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 9-1: Using an Analytical Solution.**

GIVEN: $\dot{y}(t) = \cos(\omega t)$ and $\omega = 1 \text{ rad/s}$

$$y(t = 0) = y_0 = 0$$

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 0.479 166 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. 9.2.1.

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 9-1 shows an example of each of these effects.

Secular changes in a particular element vary linearly over time, or in some cases, proportionally to some power of time, such as a quadratic. Remember that secular terms

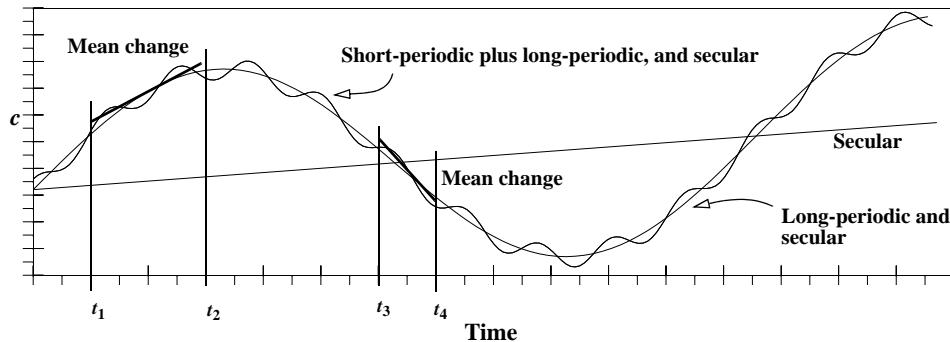


Figure 9-1. Effect of Perturbation Forces on Orbital Elements. The change in any orbital element, c , is illustrated. The straight line shows secular effects. The large oscillating line shows the secular plus long-periodic effects, and the small oscillatory line, which combines all three, shows the short-periodic effects. You should understand which elements are averaged when using mean elements. The two examples I've shown are greatly exaggerated but representative of possible errors.

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 simulation, 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. The corresponding position and velocity vectors define these ***osculating elements*** at any instant in time. “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

$$c = c_o + \dot{c}_1(t - t_o) + K_1 \cos(2\omega) + K_2 \sin(2\nu + \omega) + K_3 \cos(2\nu) \quad (9-1)$$

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. (9-1) 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 9-1 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 9-1 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.

9.2.1 The Method of Perturbations

The Method of Perturbations is not a 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

$$\frac{dc_i}{dt} = \epsilon_1 f_i + \epsilon_2 g_i \quad \epsilon_1 \ll 1 \quad \epsilon_2 \ll 1$$

and an approximation to the true solution can be developed as a power series in these small parameters:

$$\begin{aligned} c = c_o + \epsilon_1 \alpha_1(t) + \frac{\epsilon_1^2}{2!} \alpha_2(t) + \dots \\ + \epsilon_2 \beta_1(t) + \frac{\epsilon_2^2}{2!} \beta_2(t) + \dots \\ + \frac{\epsilon_1 \epsilon_2}{2!} \gamma_2(t) + \dots \end{aligned} \tag{9-2}$$

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. (2-69)].

Examples of small parameters are $J_2, J_3, C_{l,m}, S_{l,m}$. 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 $J_i \approx O(J_2^2)$; that is, the other zonal harmonic coefficients, J_3, J_4, J_5 are roughly of order J_2^2 .

Note from Eq. (9-2) 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 mécanique 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.

9.3 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 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. 1, 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. 2.

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,

$$\frac{d\vec{c}}{dt} = f(\vec{c}, t) \quad (9-3)$$

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:

$$\vec{r} = \dot{\vec{x}}(a, e, i, \Omega, \omega, M, t) = \dot{\vec{x}}(c_1, c_2, c_3, c_4, c_5, c_6, t) = \dot{\vec{x}}(\vec{c}, t)$$

$$\vec{v} = \dot{\vec{x}}(a, e, i, \Omega, \omega, M, t) = \dot{\vec{x}}(c_1, c_2, c_3, c_4, c_5, c_6, t) = \dot{\vec{x}}(\vec{c}, t)$$

The *six* constants, \hat{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.

$$\ddot{\hat{x}}(\hat{c}, t) + \frac{\mu \hat{x}(\hat{c}, t)}{|\hat{x}(\hat{c}, t)|^3} = 0 \quad \ddot{\hat{x}}(\hat{c}, t) + \frac{\mu \hat{x}(\hat{c}, t)}{|\hat{x}(\hat{c}, t)|^3} = \hat{a}_{pert} \quad (9-4)$$

where \hat{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 $dc_i/dt = 0$.

$$\begin{aligned} \dot{\hat{x}}(\hat{c}, t) &= \frac{d\hat{x}(\hat{c}, t)}{dt} = \frac{\partial \hat{x}(\hat{c}, t)}{\partial t} \\ \dot{\hat{x}}(\hat{c}, t) &= \frac{d\hat{x}(\hat{c}, t)}{dt} = \frac{\partial \hat{x}(\hat{c}, t)}{\partial t} + \sum_{i=1}^6 \frac{\partial \hat{x}(\hat{c}, t)}{\partial c_i} \frac{dc_i}{dt} \end{aligned} \quad (9-5)$$

We now differentiate the position vector in the perturbed case [Eq. (9-4)] 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. (9-5)b, and we impose the following constraint:

$$\sum_{i=1}^6 \frac{\partial \hat{x}(\hat{c}, t)}{\partial c_i} \frac{dc_i}{dt} \equiv 0 \quad (9-6)$$

Geyling and Westerman (1971:158) refer to Eq. (9-6) 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. (9-5)b, gives the same result as the unperturbed solution. A second derivative of the position vector, still using the condition of osculation, gives us

$$\frac{d^2 \hat{x}(\hat{c}, t)}{dt^2} = \frac{\partial^2 \hat{x}(\hat{c}, t)}{\partial t^2} + \sum_{i=1}^6 \frac{\partial \dot{\hat{x}}(\hat{c}, t)}{\partial c_i} \frac{dc_i}{dt}$$

We can now find the osculating element rates in terms of the original equations of motion by substituting into Eq. (9-4)b:

$$\frac{\partial^2 \dot{\vec{x}}(\vec{c}, t)}{\partial t^2} + \sum_{i=1}^6 \frac{\partial \dot{\vec{x}}(\vec{c}, t)}{\partial c_i} \frac{dc_i}{dt} + \frac{\mu \dot{\vec{x}}(\vec{c}, t)}{|\dot{\vec{x}}(\vec{c}, t)|^3} = \dot{\vec{a}}_{pert} \quad (9-7)$$

But we know that the unperturbed equations of motion [Eq. (9-4)a] hold for any time and can be written as

$$\frac{\partial^2 \dot{\vec{x}}(\vec{c}, t)}{\partial t^2} + \frac{\mu \dot{\vec{x}}(\vec{c}, t)}{|\dot{\vec{x}}(\vec{c}, t)|^3} = 0$$

so Eq. (9-7) simplifies to

$$\sum_{i=1}^6 \frac{\partial \dot{\vec{x}}(\vec{c}, t)}{\partial c_i} \frac{dc_i}{dt} = \dot{\vec{a}}_{pert} \quad (9-8)$$

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 ($\dot{\vec{a}}_{pert}$) 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. (9-6). Using matrix notation, the complete system of equations gives us

$$\begin{bmatrix} \sum_{i=1}^6 \frac{\partial \dot{\vec{x}}(\vec{c}, t)}{\partial c_i} \frac{dc_i}{dt} \\ \sum_{i=1}^6 \frac{\partial \dot{\vec{x}}(\vec{c}, t)}{\partial c_i} \frac{dc_i}{dt} \end{bmatrix} = \begin{bmatrix} \dot{\vec{0}} \\ \dot{\vec{a}}_{pert} \end{bmatrix} \quad (9-9)$$

This expression is not well suited for computations because it isn't of the form of Eq. (9-3). We can now examine two approaches for transforming Eq. (9-9) into a more desirable form—Lagrangian and Gaussian VOP.

9.3.1 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. (9-3) 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. (9-9) 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. (9-9) with $\left(\frac{\partial \dot{\vec{x}}(\vec{c}, t)}{\partial c_k}\right)$ and the top vector in Eq. (9-9) 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 (using the disturbing potential R)

$$\sum_{i=1}^6 \frac{dc_i}{dt} \left(\frac{\partial \dot{\vec{x}}(\vec{c}, t)}{\partial c_k} \cdot \frac{\partial \dot{\vec{x}}(\vec{c}, t)}{\partial c_i} - \frac{\partial \dot{\vec{x}}(\vec{c}, t)}{\partial c_i} \cdot \frac{\partial \dot{\vec{x}}(\vec{c}, t)}{\partial c_k} \right) = \frac{\partial R}{\partial c_k}$$

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:

$$[c_i, c_k] = \frac{\partial \dot{\vec{x}}(\vec{c}, t)}{\partial c_i} \cdot \frac{\partial \dot{\vec{x}}(\vec{c}, t)}{\partial c_k} - \frac{\partial \dot{\vec{x}}(\vec{c}, t)}{\partial c_k} \cdot \frac{\partial \dot{\vec{x}}(\vec{c}, t)}{\partial c_i} \quad (9-10)$$

After performing the matrix operations in Eq. (9-9) and invoking the Lagrange bracket notation, we can rewrite the matrix equation for the element rates.

$$\begin{bmatrix} [c_1, c_1] & \dots & [c_1, c_6] \\ \vdots & & \\ [c_6, c_1] & & [c_6, c_6] \end{bmatrix} \begin{bmatrix} \frac{d\vec{c}}{dt} \end{bmatrix} = \begin{bmatrix} \frac{\partial R}{\partial \vec{c}} \end{bmatrix}$$

We can express the time rate of change of each orbital element in matrix notation (let the matrix of Lagrange brackets be L):

$$\dot{\vec{c}} = L^{-1} \frac{\partial R}{\partial \vec{c}} \quad (9-11)$$

This is a general form of the Lagrange planetary equations.

A characteristic of Lagrange brackets is that they're time-independent, or constant. From Taff (1985:306), or Kaula (1966:27),

$$\frac{d[c_i, c_k]}{dt} \equiv 0$$

Some identities are obvious from the definition [Eq. (9-10)] and they reduce the number of Lagrange brackets we must evaluate to 15:

$$\begin{aligned} [c_i, c_i] &= 0 \\ [c_i, c_k] &= -[c_k, c_i] \end{aligned}$$

Evaluating the remaining Lagrange brackets isn't easy. It requires the partial derivatives

$$\frac{\partial \dot{\vec{x}}}{\partial c_i} \quad \text{and} \quad \frac{\partial \dot{\vec{x}}}{\partial c_i} \text{ where } i = 1 \dots 6$$

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 dependence of $\dot{\vec{x}}, \ddot{\vec{x}}$ on the orbital elements using the transformation from the geocentric (IJK) to the apsidal (PQW) frame. Letting $\dot{\vec{x}}, \ddot{\vec{x}}$ and $\dot{\vec{r}}_{PQW}, \ddot{\vec{v}}_{PQW}$ represent the position and velocity vectors in the geocentric and apsidal frames, respectively. From Eq. (2-104) and Eq. (2-106), the position and velocity vectors in the PQW frame are

$$\dot{\vec{r}}_{PQW} = \begin{bmatrix} r \cos(\nu) \\ r \sin(\nu) \\ 0 \end{bmatrix} \quad \dot{\vec{v}}_{PQW} = \begin{bmatrix} -na \sin(\nu) \\ \sqrt{1-e^2} \\ na(e + \cos(\nu)) \end{bmatrix}$$

Remember from Eq. (1-24) that

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

Eq. (3-30) allows us to transform between the IJK and PQW frames. If we let \mathbf{R} represent the transformation matrix from PQW to IJK,

$$\dot{\vec{x}} = \mathbf{R}_{[IJK]}^{[PQW]} \dot{\vec{r}}_{PQW} \quad \ddot{\vec{x}} = \mathbf{R}_{[IJK]}^{[PQW]} \dot{\vec{v}}_{PQW}$$

The previous relations show a useful 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,

$$\frac{\partial \dot{\vec{x}}}{\partial (\Omega, \omega, i)} = \frac{\partial \mathbf{R}_{[IJK]}^{[PQW]}}{\partial (\Omega, \omega, i)} (\dot{\vec{r}}_{PQW}, \dot{\vec{v}}_{PQW}) \quad \frac{\partial \dot{\vec{x}}}{\partial (a, e, M)} = \mathbf{R}_{[IJK]}^{[PQW]} \frac{\partial (\dot{\vec{r}}_{PQW}, \dot{\vec{v}}_{PQW})}{\partial (a, e, M)}$$

We can also determine the PQW position and velocity vectors at perigee.

$$\dot{\vec{r}}_{PQW} = \begin{bmatrix} a(1-e) \\ 0 \\ 0 \end{bmatrix} \quad \dot{\vec{v}}_{PQW} = \begin{bmatrix} 0 \\ na \sqrt{\frac{1+e}{1-e}} \\ 0 \end{bmatrix}$$

Three cases arise in the solution:

1. If both c_i and $c_k \in \{\Omega, i, \text{ or } \omega\}$, then

$$\frac{\partial \dot{\vec{x}}}{\partial c_i} = \begin{bmatrix} \frac{\partial x_1}{\partial c_i} \\ \frac{\partial x_2}{\partial c_i} \\ \frac{\partial x_3}{\partial c_i} \end{bmatrix} \quad \frac{\partial \dot{\vec{x}}}{\partial c_k} = \begin{bmatrix} \frac{\partial \dot{x}_1}{\partial c_k} \\ \frac{\partial \dot{x}_2}{\partial c_k} \\ \frac{\partial \dot{x}_3}{\partial c_k} \end{bmatrix}$$

where at perigee,

$$\begin{aligned} \frac{\partial x_j}{\partial c_i} &= \sum_{l=1}^3 \frac{\partial R_{jl}}{\partial c_i} \vec{r}_{PQW} = \frac{\partial R_{j1}}{\partial c_i} a(1-e) \\ \frac{\partial \dot{x}_j}{\partial c_k} &= \sum_{l=1}^3 R_{jl} \vec{v}_{PQW} = \frac{\partial R_{j2}}{\partial c_k} n a \sqrt{\frac{1+e}{1-e}} \end{aligned}$$

The indices j and l designate the matrix elements' row and column of the R matrix. The dot product is then

$$\frac{\partial \dot{\vec{x}}}{\partial c_i} \cdot \frac{\partial \dot{\vec{x}}}{\partial c_k} = \sum_{j=1}^3 \frac{\partial x_j}{\partial c_i} \frac{\partial \dot{x}_j}{\partial c_k} = \sum_{j=1}^3 n a^2 \sqrt{1-e^2} \frac{\partial R_{j1}}{\partial c_i} \frac{\partial R_{j2}}{\partial c_k}$$

Similarly,

$$\frac{\partial \dot{\vec{x}}}{\partial c_i} \cdot \frac{\partial \dot{\vec{x}}}{\partial c_k} = \sum_{j=1}^3 \frac{\partial \dot{x}_j}{\partial c_i} \frac{\partial x_j}{\partial c_k} = \sum_{j=1}^3 n a^2 \sqrt{1-e^2} \frac{\partial R_{j2}}{\partial c_i} \frac{\partial R_{j1}}{\partial c_k}$$

and we have

$$[c_i, c_k] = n a \sqrt{1-e^2} \sum_{j=1}^3 \left[\left\{ \frac{\partial R_{j1}}{\partial c_i} \frac{\partial R_{j2}}{\partial c_k} - \frac{\partial R_{j2}}{\partial c_i} \frac{\partial R_{j1}}{\partial c_k} \right\} \right]$$

2. If $c_i \in \{\Omega, i, \text{ or } \omega\}$ and $c_k \in \{a, e, M\}$, similar operations yield

$$[c_i, c_k] = \sum_{j=1}^3 \left[a(1-e) \frac{\partial R_{j1}}{\partial c_i} \left\{ R_{j1} \frac{\partial v_P}{\partial c_k} + R_{j2} \frac{\partial v_Q}{\partial c_k} \right\} - \frac{\sqrt{1-e^2} n a}{1-e} \frac{\partial R_{j2}}{\partial c_i} \left\{ R_{j1} \frac{\partial r_P}{\partial c_k} + R_{j2} \frac{\partial r_Q}{\partial c_k} \right\} \right]$$

3. If both c_i and $c_k \in \{a, e, M\}$, similar operations yield

$$[c_i, c_k] = \sum_{j=1}^3 \left[R_{j1}^2 \left\{ \frac{\partial r_P}{\partial c_i} \frac{\partial v_P}{\partial c_k} - \frac{\partial r_P}{\partial c_k} \frac{\partial v_P}{\partial c_i} \right\} + R_{j1} R_{j2} \left\{ \frac{\partial r_P}{\partial c_i} \frac{\partial v_Q}{\partial c_k} - \frac{\partial r_P}{\partial c_k} \frac{\partial v_Q}{\partial c_i} \right\} \right. \\ \left. + R_{j1} R_{j2} \left\{ \frac{\partial r_Q}{\partial c_i} \frac{\partial v_P}{\partial c_k} - \frac{\partial r_Q}{\partial c_k} \frac{\partial v_P}{\partial c_i} \right\} + R_{j2}^2 \left\{ \frac{\partial r_Q}{\partial c_i} \frac{\partial v_Q}{\partial c_k} - \frac{\partial r_Q}{\partial c_k} \frac{\partial v_Q}{\partial c_i} \right\} \right]$$

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. (3-30):

$$[\Omega, i] = \left[\begin{aligned} & \left\{ -\sin(\Omega)\cos(\omega) - \cos(\Omega)\cos(i)\sin(\omega) \right\} \sin(\Omega)\sin(i)\cos(\omega) \\ & - \left\{ \cos(\Omega)\cos(\omega) - \sin(\Omega)\cos(i)\sin(\omega) \right\} \cos(\Omega)\sin(i)\cos(\omega) \\ & - \left\{ \sin(\Omega)\sin(\omega) - \cos(\Omega)\cos(i)\cos(\omega) \right\} \sin(\Omega)\sin(i)\sin(\omega) \\ & - \left\{ \cos(\Omega)\sin(\omega) + \sin(\Omega)\cos(i)\cos(\omega) \right\} \cos(\Omega)\sin(i)\sin(\omega) \end{aligned} \right] n a^2 \sqrt{1-e^2}$$

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

$$[\Omega, a] = \frac{n a \sqrt{1-e^2}}{2} \cos(i) = [\omega, a] \cos(i)$$

$$[\Omega, e] = \frac{-n a^2 e}{\sqrt{1-e^2}} \cos(i) = [\omega, e] \cos(i)$$

$$[\omega, a] = \frac{na\sqrt{1-e^2}}{2}$$

$$[\omega, e] = \frac{-na^2e}{\sqrt{1-e^2}}$$

$$[a, M_o] = \frac{-na}{2}$$

We can now solve Eq. (9-11) to find the equations for the variation of parameters. The result is called the *Lagrange planetary equations*.

$$\begin{aligned} \frac{da}{dt} &= \frac{2}{na} \frac{\partial R}{\partial M_o} \\ \frac{de}{dt} &= \frac{1-e^2}{na^2e} \frac{\partial R}{\partial M_o} - \frac{\sqrt{1-e^2}}{na^2e} \frac{\partial R}{\partial \omega} \\ \frac{di}{dt} &= \frac{1}{na^2\sqrt{1-e^2}\sin(i)} \left\{ \cos(i) \frac{\partial R}{\partial \omega} - \frac{\partial R}{\partial \Omega} \right\} \\ \frac{d\omega}{dt} &= \frac{\sqrt{1-e^2}}{na^2e} \frac{\partial R}{\partial e} - \frac{\cot(i)}{na^2\sqrt{1-e^2}} \frac{\partial R}{\partial i} \\ \frac{d\Omega}{dt} &= \frac{1}{na^2\sqrt{1-e^2}\sin(i)} \frac{\partial R}{\partial i} \\ \frac{dM_o}{dt} &= -\frac{1-e^2}{na^2e} \frac{\partial R}{\partial e} - \frac{2}{na} \frac{\partial R}{\partial a} \end{aligned} \tag{9-12}$$

An alternate form of the Lagrange planetary equations is

$$\frac{dc_i}{dt} = -\sum_{k=1}^6 (c_i, c_k) \frac{\partial R}{\partial c_k}, \text{ where } i = 1 \dots 6$$

Here, the (c_i, c_k) are the Poisson brackets, which are intimately tied to the Lagrange brackets. Note that \mathbf{L}^{-1} is related to the negative matrix of *Poisson brackets*, $-\mathbf{P}$, because $\mathbf{LP}^T = \mathbf{I}$. 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, $R(a, e, i, \Omega, \omega)$, and the initial value of mean anomaly, M_o , we have the expressions in Eq. (9-12). Kaula's form of the potential [Eq. (9-28)] is well suited for use in Eq. (9-12).

Eq. (9-12) 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

$$M = n(t - T) = nt + \sigma \quad (9-13)$$

Differentiating, we have

$$\dot{M} = n + \dot{n}t + \dot{\sigma} = n + \frac{dn}{da}\dot{a}t + \dot{\sigma}$$

Eq. (9-12) gives us the time rate of change for a , and $dn/da = -3n/2a$ from Eq. (1-28). However, from Eq. (9-12), notice that $\partial R/\partial a$ has a contribution from a and an indirect contribution for Eq. (9-13). 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

$$\dot{M} = n - \frac{3n\dot{a}t}{2a} - \frac{1-e^2}{na^2e}\frac{\partial R}{\partial e} - \frac{2}{na}\left(\frac{\partial R}{\partial a}\right)_M + \frac{3t}{a^2}\frac{\partial R}{\partial M}$$

If we substitute

$$\dot{a} = \frac{2}{na}\frac{\partial R}{\partial M}$$

the terms with time appearing explicitly cancel, and we have

$$\dot{M} = n - \frac{1-e^2}{na^2e}\frac{\partial R}{\partial e} - \frac{2}{na}\left(\frac{\partial R}{\partial a}\right)_M$$

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. (9-12). 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. (9-12) 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:

$$\begin{aligned}
 B &= \sqrt{1 - h_e^2 - k_e^2} & C &= 1 + p_e^2 + q_e^2 \\
 \frac{da}{dt} &= \frac{2}{na} \frac{\partial R}{\partial \lambda_M} \\
 \frac{dh_e}{dt} &= \frac{B}{na^2} \frac{\partial R}{\partial k_e} - \frac{h_e B}{na^2(1+B)} \frac{\partial R}{\partial \lambda_M} + \frac{k C}{2na^2 B} \left\{ p_e \frac{\partial R}{\partial p_e} + q_e \frac{\partial R}{\partial q_e} \right\} \\
 \frac{dk_e}{dt} &= -\frac{B}{na^2} \frac{\partial R}{\partial h_e} - \frac{k_e B}{na^2(1+B)} \frac{\partial R}{\partial \lambda_M} - \frac{h_e C}{2na^2 B} \left\{ p_e \frac{\partial R}{\partial p_e} + q_e \frac{\partial R}{\partial q_e} \right\} \\
 \frac{dp_e}{dt} &= -\frac{p_e C}{2na^2 B} \left\{ k_e \frac{\partial R}{\partial h_e} - h_e \frac{\partial R}{\partial k_e} + \frac{\partial R}{\partial \lambda_M} \right\} + \frac{f_r C^2}{4na^2 B} \frac{\partial R}{\partial q_e} \\
 \frac{dq_e}{dt} &= -\frac{q_e C}{2na^2 B} \left\{ k_e \frac{\partial R}{\partial h_e} - h_e \frac{\partial R}{\partial k_e} + \frac{\partial R}{\partial \lambda_M} \right\} - \frac{f_r C^2}{4na^2 B} \frac{\partial R}{\partial p_e} \\
 \frac{d\lambda_M}{dt} &= n - \frac{2}{na} \frac{\partial R}{\partial a} + \frac{B}{na^2(1+B)} \left\{ h_e \frac{\partial R}{\partial h_e} + k_e \frac{\partial R}{\partial k_e} \right\} + \frac{C}{2na^2 B} \left\{ p_e \frac{\partial R}{\partial p_e} + q_e \frac{\partial R}{\partial q_e} \right\}
 \end{aligned} \tag{9-14}$$

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).

9.3.2 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. (9-9) expressing it as two separate equations and not using the matrix notation:

$$\sum_{i=1}^6 \frac{\partial \dot{\vec{x}}(\vec{c}, t)}{\partial c_i} \frac{dc_i}{dt} = \vec{0} \quad \sum_{i=1}^6 \frac{\partial \dot{\vec{x}}(\vec{c}, t)}{\partial c_i} \frac{\dot{dc}_i}{dt} = \vec{\dot{a}}_{pert}$$

Taking the dot product of the first equation with $\partial c_j / \partial \dot{\vec{x}}$ and the second by $\partial c_j / \partial \vec{x}$ and adding the two relations result in

$$\sum_{i=1}^6 \left(\frac{\partial c_j}{\partial \dot{\vec{x}}} \cdot \frac{\partial \dot{\vec{x}}}{\partial c_i} + \frac{\partial c_j}{\partial \vec{x}} \cdot \frac{\partial \dot{\vec{x}}}{\partial \dot{c}_i} \right) \frac{dc_i}{dt} = \frac{\partial c_j}{\partial \vec{x}} \cdot \vec{\dot{a}}_{pert} \quad (9-15)$$

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. (9-15) becomes

$$\sum_{i=1}^6 \delta_{j,i} \frac{dc_i}{dt} = \frac{\partial c_j}{\partial \vec{x}} \cdot \vec{\dot{a}}_{pert}$$

or more simply,

$$\frac{dc_j}{dt} = \frac{\partial c_j}{\partial \vec{x}} \cdot \vec{\dot{a}}_{pert}$$

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. 3 for additional information on coordinate systems (Sec. 3.3).

Gauss chose to develop the equations in the RSW system. I'll next present one geometrical derivation of the VOP equations. Summarizing from Sec. 3.4, let the components of the disturbing force (per unit mass) be \hat{R} along the radius vector, \hat{S} perpendicular to the 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 R , S , and W form a right-hand set of axes. The 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. (1-21)] 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. (1-24)] to get

$$\frac{dr}{d\nu} = \frac{r e \sin(\nu)}{1 + e \cos(\nu)}$$

From the angular momentum relations in Eq. (1-18) and Eq. (1-19), and the relation of mean motion [Eq. (2-77)],

$$h = r^2 \dot{\nu} = \sqrt{\mu p} = \sqrt{\mu a(1 - e^2)} = na^2 \sqrt{1 - e^2} \quad (9-16)$$

Equate the final two terms to get a relation that's useful throughout the development:

$$\sqrt{\mu a} = na^2$$

It's apparent from Eq. (9-16) that

$$\dot{\nu} = \frac{na^2}{r^2} \sqrt{1 - e^2} \quad (9-17)$$

Substitute the expressions as

$$\frac{da}{dt} = \frac{\mu}{2\xi^2} \left(\frac{na^2}{r^2} \sqrt{1 - e^2} \right) \left(\frac{re \sin(\nu)}{1 + e \cos(\nu)} F_R + r F_S \right)$$

or

$$\frac{da}{dt} = \frac{2e \sin(\nu)}{n \sqrt{1 - e^2}} F_R + \frac{2a \sqrt{1 - e^2}}{nr} F_S \quad (9-18)$$

To derive the remaining element variations, you'll need the time rate-of-change of angular momentum. Remember that in two-body motion, the angular momentum 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 ($\dot{h} = h \dot{W}$) 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 $d\hat{h}/dt$, you'll find that

$$\frac{dh}{dt} = \dot{h} = r F_S \quad (9-19)$$

As with the radial rate (\dot{r}), remember that $\dot{h} \neq d\hat{h}/dt$. To find \dot{e} , recall the expression for the eccentricity [Eq. (2-79)],

$$e = \sqrt{1 - \frac{p}{a}} = \sqrt{1 - \frac{h^2}{\mu a}}$$

Differentiate to get

$$\begin{aligned} \frac{de}{dt} &= \frac{1}{2} \left(1 - \frac{h^2}{\mu a}\right)^{-1/2} \left[\frac{-2h}{\mu a} \frac{dh}{dt} + (-h^2) \mu (-1) (\mu a)^{-2} \frac{da}{dt} \right] \\ &= \frac{1}{2e} \left[\frac{-2h}{\mu a} \frac{dh}{dt} + \frac{h^2}{\mu a^2} \frac{da}{dt} \right] \\ &= \left(-\frac{h}{\mu a e} \frac{dh}{dt} + \frac{h^2}{2\mu a^2 e} \frac{da}{dt} \right) \end{aligned}$$

Substitute (dh/dt) and (da/dt) into the equation above:

$$\frac{de}{dt} = -\frac{h}{\mu a e} r F_s + \frac{h^2}{2\mu a^2 e} \left(\frac{2e \sin(\nu)}{n \sqrt{1-e^2}} F_R + \frac{2a \sqrt{1-e^2}}{nr} F_S \right)$$

Now, use $(h = \sqrt{\mu a (1 - e^2)})$ to obtain

$$\frac{de}{dt} = -\frac{\sqrt{1-e^2}}{\sqrt{\mu a e}}(rF_S) + \frac{(1-e^2)\sin(\nu)}{an\sqrt{1-e^2}}F_R + \frac{(1-e^2)\sqrt{1-e^2}}{enr}F_S$$

but $\sqrt{\mu a} = na^2$, so

$$\frac{de}{dt} = -\frac{\sqrt{1-e^2}}{na^2 e}rF_S + \frac{\sqrt{1-e^2}\sin(\nu)}{na}F_R + \frac{\sqrt{1-e^2}}{na^2 e}\left(\frac{(1-e^2)a^2}{r}F_S\right)$$

and the final result is

$$\frac{de}{dt} = \frac{\sqrt{1-e^2}\sin(\nu)}{na}F_R + \frac{\sqrt{1-e^2}}{na^2 e}\left[\frac{a^2(1-e^2)}{r} - r\right]F_S \quad (9-20)$$

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. (2-82), recall that

$$\cos(i) = \frac{\hat{h} \cdot \hat{K}}{|\hat{h}| |\hat{K}|} = \frac{\hat{h} \cdot \hat{K}}{h}$$

Differentiating with respect to time to determine the rate of change gives us

$$-\sin(i)\frac{di}{dt} = \frac{h\left(\frac{d\hat{h}}{dt} \cdot \hat{K}\right) - (\hat{h} \cdot \hat{K})\frac{dh}{dt}}{h^2}$$

Using $(d\hat{h}/dt)$ and dh/dt from the development of Eq. (9-19) results in

$$-\sin(i)\frac{di}{dt} = \frac{h(rF_S\hat{W} - rF_W\hat{S}) \cdot \hat{K} - h\cos(i)rF_S}{h^2}$$

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

$$\frac{di}{dt} = \frac{r\cos(u)F_W}{na^2\sqrt{1-e^2}} \quad (9-21)$$

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. (2-83) and Eq. (2-84),

$$\cos(\Omega) = \frac{\hat{I} \cdot (\hat{K} \times \hat{h})}{|\hat{K} \times \hat{h}|}$$

Differentiating gives you

$$\begin{aligned} -\sin(\Omega) \frac{d\Omega}{dt} &= \frac{\hat{I} \cdot (\hat{K} \times \frac{d\hat{h}}{dt}) |\hat{K} \times \hat{h}| - \hat{I} \cdot (\hat{K} \times \hat{h}) \frac{d|\hat{K} \times \hat{h}|}{dt}}{|\hat{K} \times \hat{h}|^2} \\ &= \frac{\hat{I} \cdot [\hat{K} \times (rF_s \hat{W} - rF_w \hat{S})] h \sin(i) - h \cos(\Omega) \sin(i) \left(\frac{dh}{dt} \sin(i) + h \cos(i) \frac{di}{dt} \right)}{h^2 \sin^2(i)} \end{aligned}$$

Determine the vector cross products using the transformation matrix for PQW to IJK [Eq. (3-30)] and the alternate formulation, Eq. (3-19). 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. (9-16) and Eq. (2-5). Just as with di/dt , notice that only the W component of the force perturbs Ω . Therefore,

$$\frac{d\Omega}{dt} = \frac{r \sin(u) F_w}{na^2 \sqrt{1 - e^2} \sin(i)} \quad (9-22)$$

Next, obtain the argument of perigee rate. At this point, it's useful to recall the argument of latitude, u [Eq. (2-89)] as

$$\frac{(\hat{K} \times \hat{h}) \cdot \hat{r}}{|\hat{K} \times \hat{h}|} = r \cos(u)$$

Differentiating gives you

$$\frac{|\hat{K} \times \hat{h}| \left(\hat{K} \times \frac{d\hat{h}}{dt} \cdot \hat{r} \right) - (\hat{K} \times \hat{h}) \cdot \frac{d}{dt} |\hat{K} \times \hat{h}|}{|\hat{K} \times \hat{h}|^2} = -r \sin(u) \left(\frac{d\omega}{dt} + \frac{d\nu}{dt} \right)$$

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. (9-17)] because now perturbations are influencing the orbit. We know several relations,

$$\begin{aligned}\hat{K} \times \hat{W} \cdot \hat{r} &= r \sin(i) \cos(u) \\ \hat{K} \times \hat{S} \cdot \hat{r} &= r \hat{S} \times \hat{R} \cdot \hat{K} = -r \hat{W} \cdot \hat{K} = -r \cos(i) \\ \hat{K} \times \hat{h} \cdot \hat{r} &= rh \sin(i) \cos(u)\end{aligned}$$

and have already developed dh/dt [from Eq. (9-19)] and di/dt [Eq. (9-21)]. Therefore, you can find

$$\begin{aligned}\frac{d\omega}{dt} = \left\{ \frac{1}{h^2 \sin^2(i) r \sin(u)} \right\} &\left\{ -h \sin(i) [\hat{K} \times (r F_s \hat{W} - r F_w \hat{S}) \cdot \hat{r}] \right. \\ &\left. + (\hat{K} \times \hat{h} \cdot \hat{r}) \left[\frac{dh}{dt} \sin(i) + h \cos(i) \frac{di}{dt} \right] - \frac{d\nu}{dt} h^2 r \sin^2(i) \sin(u) \right\}\end{aligned}$$

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):

$$\begin{aligned}r \left(\frac{de}{dt} \cos(\nu) - e \sin(\nu) \frac{d\nu}{dt} \right) &= \frac{2h}{\mu} \frac{dh}{dt} \\ re \sin(\nu) \frac{d\nu}{dt} &= r \cos(\nu) \frac{de}{dt} - \frac{2h}{\mu} \frac{dh}{dt}\end{aligned}$$

At this point, you solve for $d\nu/dt$ and substitute for dh/dt [Eq. (9-19)] and de/dt [Eq. (9-20)], which gives you

$$\begin{aligned}\frac{d\nu}{dt} = \frac{\cot(\nu)}{e} \left[\frac{\sqrt{1-e^2}}{na} \sin(\nu) F_R + \frac{\sqrt{1-e^2}}{na^2 e} \left(\frac{a^2(1-e^2)}{r} - r \right) F_S \right] \\ - \frac{2h}{\mu r e \sin(\nu)} r F_S\end{aligned}$$

Collecting terms of F_S results in^{*}

$$\frac{d\nu}{dt} = \frac{\sqrt{1-e^2}}{nae} \cos(\nu) F_R - \frac{\sqrt{1-e^2}}{nae} \frac{2+e\cos(\nu)}{1+e\cos(\nu)} F_S$$

Now, let's substitute the perturbed rate of change of the true anomaly plus Eq. (9-19) and Eq. (9-21), 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

$$\begin{aligned} \frac{d\omega}{dt} = & \frac{\sqrt{1-e^2}}{nae} \left[-\cos(\nu) F_R + \frac{2+e\cos(\nu)}{1+e\cos(\nu)} \sin(\nu) F_S \right] \\ & - \frac{r \cot(i) \sin(u)}{h} F_W \end{aligned} \tag{9-23}$$

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. (2-7), $M_o = M - n(t - t_o)$]. Differentiating results in

$$\frac{dM_o}{dt} = \frac{dE}{dt} - \frac{de}{dt} \sin(E) - e \cos(E) \frac{dE}{dt} - \frac{dn}{dt} (t - t_o)$$

Find the rate of change of the mean motion using Eq. (1-27). (See also page 672).

$$\frac{dn}{dt} = -\frac{3n}{2a} \frac{da}{dt}$$

We know de/dt , Eq. (9-20) and $\frac{dE}{dt} = \frac{1}{r\sqrt{a}} \frac{\mu}{a}$ from Eq. (8-8), and you can get $\sin(E)$ and $\cos(E)$ from the expressions in Eq. (2-8) and Eq. (2-9). Bate, Mueller, and White (1971:406) give the result (for elliptical orbits only).

$$\frac{dM_o}{dt} = -\frac{1}{na} \left(\frac{2r}{a} - \frac{1-e^2}{e} \cos(\nu) \right) F_R - \frac{(1-e^2)}{nae} \left(1 + \frac{r}{a(1-e^2)} \right) \sin(\nu) F_S - \frac{dn}{dt} (t - t_o)$$

* 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.

You can simplify this equation as I've shown in Eq. (9-24).

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:

$$\begin{aligned}
 \frac{da}{dt} &= \frac{2}{n\sqrt{1-e^2}} \left\{ e \sin(\nu) F_R + \frac{p}{r} F_S \right\} \\
 \frac{de}{dt} &= \frac{\sqrt{1-e^2}}{na} \left\{ \sin(\nu) F_R + \left(\cos(\nu) + \frac{e + \cos(\nu)}{1+e\cos(\nu)} \right) F_S \right\} \\
 \frac{di}{dt} &= \frac{r \cos(u)}{na^2 \sqrt{1-e^2}} F_W \\
 \frac{d\Omega}{dt} &= \frac{r \sin(u)}{na^2 \sqrt{1-e^2} \sin(i)} F_W \\
 \frac{d\omega}{dt} &= \frac{\sqrt{1-e^2}}{nae} \left\{ -\cos(\nu) F_R + \sin(\nu) \left(1 + \frac{r}{p} \right) F_S \right\} - \frac{r \cot(i) \sin(u)}{h} F_W \\
 \frac{dM_o}{dt} &= \frac{1}{na^2 e} \left\{ (p \cos(\nu) - 2er) F_R - (p+r) \sin(\nu) F_S \right\} - \frac{dn}{dt} (t - t_o)
 \end{aligned} \tag{9-24}$$

Alternate expressions exist for the mean anomaly. Battin (1987:489) lists one.

$$\frac{d\nu}{dt} = \frac{h}{r^2} + \frac{1}{eh} \{ p \cos(\nu) \} F_R - (p+r) \sin(\nu) F_S$$

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. 8.5, 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.

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

$$F_R = \frac{\partial R}{\partial r} \quad F_S = \frac{1}{r} \frac{\partial R}{\partial u} \quad F_W = \frac{1}{r \sin(u)} \frac{\partial R}{\partial i} \quad (9-25)$$

Using the aspherical potential means we can illustrate the disturbing force's components for the zonal harmonics in Eq. (8-21) or Eq. (9-45) with the aid of Eq. (9-25). We can write the primed Legendre polynomials as combinations of other polynomials. Noting from Eq. (6-13) that $\sin(\phi_{gc}) = \sin(i) \sin(u)$, we have (for $m = 0$)

$$\begin{aligned} F_R &= \frac{\mu}{r^2} \sum_{l=2}^{\infty} J_l (l+1) \left(\frac{R_{\oplus}}{r} \right)^l P_l [\sin(i) \sin(u)] \\ F_S &= -\frac{\mu}{r^2} \sin(i) \cos(u) \sum_{l=2}^{\infty} J_l \left(\frac{R_{\oplus}}{r} \right)^l P'_l [\sin(i) \sin(u)] \\ F_W &= -\frac{\mu}{r^2} \cos(i) \sum_{l=2}^{\infty} J_l \left(\frac{R_{\oplus}}{r} \right)^l P'_{l-1} [\sin(i) \sin(u)] \end{aligned} \quad (9-26)$$

where the differential is (use $\alpha = \sin(i) \sin(u)$ for clarity)

$$(1 - \alpha^2) P'_l[\alpha] = (1 - \alpha^2) \frac{dP_l[\alpha]}{d\alpha} = -l\alpha P_l[\alpha] + lP_{l-1}[\alpha]$$

9.4 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

$$\mathcal{H} = KE - U = \frac{1}{2}mv^2 - R - U_{\text{2-body}}$$

where we've used the kinetic energy and the disturbing potential, R . Remember, $U = U_{\text{2-body}} + R$. Notice that Eq. (9-11) 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 orbital elements (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 elements ($L_d, h, H_d, M, \omega, \Omega$), described in Eq. (2-102), 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. (9-12) take the particularly simple form

$$\begin{aligned}\dot{L}_d &= \frac{\partial \mathcal{H}}{\partial M} & \dot{M} &= -\frac{\partial \mathcal{H}}{\partial L_d} \\ \dot{h} &= \frac{\partial \mathcal{H}}{\partial \omega} & \dot{\omega} &= -\frac{\partial \mathcal{H}}{\partial h}\end{aligned}\tag{9-27}$$

$$\dot{H}_d = \frac{\partial \mathcal{H}}{\partial \Omega} \quad \dot{\Omega} = -\frac{\partial \mathcal{H}}{\partial H_d}$$

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 very useful 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 elements 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 right ascension 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 elements. The generalized momentum corresponding to ignor-

able 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 elements—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 elements 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 elements:

$$\mathcal{H}^*(L'_d, h', H'_d, \underline{\omega}', \underline{\underline{\omega}}) = \mathcal{H}^{**}(L''_d, h'', H''_d, \underline{\omega}'', \underline{\underline{\omega}}'')$$

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 L''_d , h'' , and H''_d are constants. This follows from the vanishing rates above. Thus,

$$\begin{aligned}L''_d &= \text{const}_1 & M'' &= k_1 \Delta t + M''_o \\ h'' &= \text{const}_2 & \omega'' &= k_2 \Delta t + \omega''_o \\ H''_d &= \text{const}_3 & \Omega'' &= k_3 \Delta t + \Omega''_o\end{aligned}$$

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.

9.5 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.

9.5.1 Gravity Potential in Terms of the Satellite's Orbital Elements

Although the aspherical potential at point P in Fig. 8-3 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. 9-2, 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 (ROT1[i] ROT3[Ω]). This eliminates ϕ_{sat} in favor of u , i , Ω , and θ_{GMST} . Recognize that θ_{GMST} equals $\omega_{\oplus}(t - t_o)$, as shown in Fig. 9-2. The next step transforms r and u to semimajor axis and mean anomaly using Eq. (2-52). 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:

$$U = \frac{\mu}{r} + \sum_{l=2}^{\infty} \sum_{m=0}^{l} \sum_{p=0}^{l} \sum_{q=-\infty}^{\infty} \frac{\mu}{a} \left(\frac{R_{\oplus}}{a} \right)^l F_{tmp}(i) G_{lpq}(e) S_{tmpq}(\omega, M, \Omega, \theta_{GMST}) \quad (9-28)$$

The S functions given by Kaula (1966:37) contain the gravitational coefficients:

$$S_{tmpq}(\omega, M, \Omega, \theta_{GMST}) = \begin{cases} C_{l,m} \cos(\Theta_{tmpq}) + S_{l,m} \sin(\Theta_{tmpq}) & \text{if } (l-m) \text{ even} \\ -S_{l,m} \cos(\Theta_{tmpq}) + C_{l,m} \sin(\Theta_{tmpq}) & \text{if } (l-m) \text{ odd} \end{cases} \quad (9-29)$$

$$\Theta_{tmpq} = (l-2p)\omega + (l-2p+q)M + m(\Omega - \theta_{GMST})$$

The F and G functions are simply expressions of the inclination and eccentricity, respectively: (note the use of binomial coefficients)

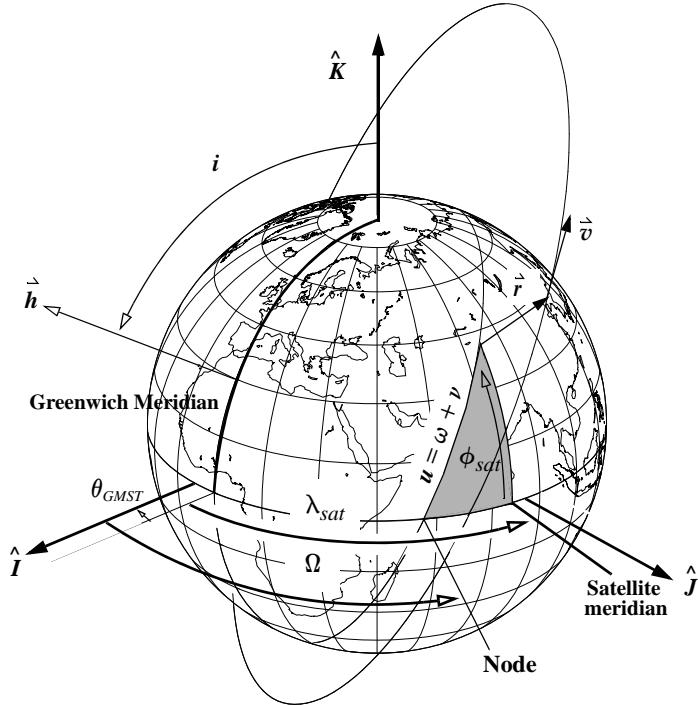


Figure 9-2. Potential Transformation. Converting the potential relies on the relation of the classical orbital elements and spherical coordinates (ϕ_{sat} , λ_{sat} , r).

$$F_{tmp}(i) = \sum_{t=0}^{\min(p, k)} \left\{ \frac{(2t-2t)!}{t!(t-t)!(t-m-2t)!2^{2t-2t}} \sin^{t-m-2t}(i) \right. \\ \times \sum_{s=0}^m \binom{m}{s} \cos^s(i) \sum_c \binom{t-m-2t+s}{c} \binom{m-s}{p-t-c} (-1)^{c-k} \left. \right\} \quad (9-30)$$

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

$$G_{lpq}(e) = \frac{1}{(1-e^2)^{l-\frac{1}{2}}} \sum_{d=0}^{p'-1} \binom{l-1}{2d+l-2p'} \binom{2d+l-2p'}{d} \left(\frac{e}{2}\right)^{2d+l-2p'} \quad (9-31)$$

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 oth-

ers express the inclination, $F_{tmp}(i)$, and eccentricity functions $G_{lpq}(e)$. Table 9-1 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.

9.5.2 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 differences.

Start with Eq. (8-39) which 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 $\text{SIN}(\delta) = \text{SIN}(i) \text{ SIN}(\omega + \nu)$. We also introduce an inclination function F_{tmp} defined in Eq. (9-30). The result is

$$R_{3-body} = \sum_{l=2}^{\infty} r^l \sum_{m=0}^l \sum_{p=0}^l F_{tmp}(i) \left\{ \begin{aligned} & \left[\begin{matrix} A_{l,m} \\ -B_{l,m} \end{matrix} \right]_{l-m \text{ odd}}^{l-m \text{ even}} \cos[(l-2p)u + m\Omega] \\ & + \left[\begin{matrix} B_{l,m} \\ A_{l,m} \end{matrix} \right]_{l-m \text{ odd}}^{l-m \text{ even}} \sin[(l-2p)u + m\Omega] \end{aligned} \right\} \quad (9-32)$$

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.

$$\begin{aligned} A_{l,m} &= \frac{\mu_3}{r_3^{l+1}} k_m \frac{(l-m)!}{(l+m)!} \sum_{h=0}^l F_{tmb}(i_3) \left[\begin{matrix} \text{COS} \\ \text{SIN} \end{matrix} \right]_{l-m \text{ odd}}^{l-m \text{ even}} [(l-2h)u_3 + m\Omega_3] \\ B_{l,m} &= \frac{\mu_3}{r_3^{l+1}} k_m \frac{(l-m)!}{(l+m)!} \sum_{h=0}^l F_{tmb}(i_3) \left[\begin{matrix} \text{SIN} \\ \text{COS} \end{matrix} \right]_{l-m \text{ odd}}^{l-m \text{ even}} [(l-2h)u_3 + m\Omega_3] \end{aligned} \quad (9-33)$$

Substituting Eq. (9-33) into Eq. (9-32) and simplifying yields

$$R_{3-body} = \frac{\mu_3}{r_3} \sum_{l=2}^{\infty} \left(\frac{r}{r_3^3} \right)^l \sum_{m=0}^l k_m \frac{(l-m)!}{(l+m)!} \sum_{p=0}^l F_{tmp}(i) \sum_{h=0}^l F_{tmb}(i_3) \cos[(l-2p)(\omega + \nu) - (l-2h)(\omega_3 + \nu_3) + m(\Omega - \Omega_3)]$$

Next, we introduce the eccentricity and mean anomaly because the rotations are related to time. This requires a Fourier series expansion of the product

TABLE 9-1. 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	l	p	q	Eccentricity Function, $G(e)$
2	0	0	$-(3/8)\sin^2(i)$	2	1	-2	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	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	2	2	2	0
2	1	1	$-(3/2)\sin(i)\cos(i)$	2	0	-1	2	2	1	$-(1/2)e + (1/16)e^3 + \dots$
2	1	2	$-(3/4)\sin(i)(1 - \cos(i))$	2	0	0	2	2	0	$1 - (5/2)e^2 + (13/16)e^4 + \dots$
2	2	0	$(3/4)(1 + \cos(i))^2$	2	0	1	2	2	-1	$(7/2)e - (123/16)e^3 + \dots$
2	2	1	$(3/2)\sin^2(i)$	2	0	2	2	2	-2	$(17/2)e^2 - (115/6)e^4 + \dots$
2	2	2	$(3/4)(1 - \cos(i))^2$							
3	0	0	$-(5/16)\sin^3(i)$	3	1	-2	3	2	2	$(11/8)e^2 + (49/16)e^4 + \dots$
3	0	1	$(15/16)\sin^3(i) - (3/4)\sin(i)$	3	1	-1	3	2	1	$e(1 - e^2)^{-5/2}$
3	0	2	$-F_{301}$	3	1	0	3	2	0	$1 + 2e^2 + (239/64)e^4 + \dots$
3	0	3	$-F_{300}$	3	1	1	3	2	-1	$3e + (11/4)e^3 + \dots$
3	0	3		3	1	2	3	2	-2	$(53/8)e^2 + (39/16)e^4 + \dots$
3	1	0	$-(15/16)\sin^2(i)(1 + \cos(i))$	3	0	-2	3	3	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	3	3	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	3	3	0	$1 - 6e^2 + (423/64)e^4 + \dots$
3	1	3	$-(15/16)\sin^2(i)(1 - \cos(i))$	3	0	1	3	3	-1	$5e - 22e^3 + \dots$
3	2	0	$(15/8)\sin(i)(1 + \cos(i))^2$	3	0	2	3	3	-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	$(15/8)(1 - \cos(i))^3$							

$$\left(\frac{r}{a}\right)^l \cos[(l-2p)(\omega + \nu) - (l-2h)(\omega_3 + \nu_3) + m(\Omega - \Omega_3)] = \sum_{q=-\infty}^{\infty} H_{lpq}(e) \Theta_{3\text{-bodysat}}$$

$$\Theta_{3\text{-bodysat}} = \cos(l-2p)\omega + (l-2p+q)M - (l-2h)(\omega_3 + \nu_3) + m(\Omega - \Omega_3)$$

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_{lpq} eccentricity functions introduced with the geopotential conversion because the H_{lhq} functions result from a Fourier

expansion of functions of the form $(r/a)^l \cos(m\nu)$, in contrast to G_{lpq} 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 :

$$H_{lp, 2p-l}(e) = \frac{-\beta^{l-2p'}}{(1+\beta^2)^{l+1}} \binom{2l+1-2p'}{l-2p'} \sum_{k=0} \binom{l+1}{k} \binom{2p'+1}{k} \frac{\beta^{2k}}{\binom{l-2p'+k}{k}}$$

where $p' = p$ for $p \leq l/2$ and $p' = l - p$ for $p \geq l/2$. Similarly, we can expand the product

$$\begin{aligned} \left(\frac{a}{r_3}\right)^{l+1} \cos(l-2p)\omega + \underset{\infty}{\text{...}} &= (l-2p+q)M - (l-2h)(\omega_3 + \nu_3) + m(\Omega - \Omega_3) \\ &= \sum_{q=-\infty} G_{l+1, hq'}(e_3) \cos(\Theta_{3-body}) \end{aligned}$$

$$\Theta_{3-body} = (l-2p)\omega + (l-2p+q)M - (l-2h)\omega_3 - (l-2h+q')M_3 + m(\Omega - \Omega_3)$$

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:

$$\begin{aligned} R_{3-body} = & \frac{\mu_3}{a_3} \sum_{l=2}^{\infty} \left(\frac{a}{a_3}\right)^l \sum_{m=0}^l k_m \frac{(l-m)!}{(l+m)!} \sum_{p=0}^l F_{tmp}(i) \sum_{h=0}^l F_{lmh}(i_3) \times \\ & \sum_{q=-\infty}^{\infty} \sum_{q'=-\infty}^{\infty} H_{lpq}(e) G_{l+1, hq'}(e_3) \cos(\Theta_{3-body}) \end{aligned} \quad (9-34)$$

This expression is very similar to Eq. (9-28) 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, $A_{l,m}$, $B_{l,m}$, 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 $A_{l,m}$, $B_{l,m}$ functions don't need to be updated as often.

9.5.3 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:

$$\begin{aligned}
 U_{tides} &= \sum_{l=2}^{\infty} \sum_{m=0}^l \sum_{h=0}^l \sum_{q'=-\infty}^{\infty} k_l[\phi_{sat}, \lambda] r^l B_{l,m} C_{lmhq'} P_{l,m} \sin[\phi_{sat}] S_{lmhq'}(\omega, M, \Omega, \theta_{GMST}) \\
 S_{lmhq'}(\omega, M, \Omega, \theta_{GMST}) &= \begin{cases} \cos(\Theta_{tides}) & \text{if } (l-m) \text{ even} \\ \sin(\Theta_{tides}) & \text{if } (l-m) \text{ odd} \end{cases} \\
 \Theta_{tides} &= (l-2h)\omega + (l-2h+q')M - \epsilon_{tmpq}[\phi_{sat}, \lambda] + m\Omega - m(\lambda + \theta_{GMST})
 \end{aligned} \tag{9-35}$$

The B and C functions are from the third-body potential [Eq. (9-34)]. They include the third-body inclination and eccentricity functions.

$$\begin{aligned}
 B_{l,m} &= \mu_3 k_m \frac{(l-m)!}{(l+m)!} \\
 C_{lmhq'} &= \left(\frac{1}{a_3}\right)^{l+1} F_{lmh}(i_3) G_{lhq'}(e_3)
 \end{aligned}$$

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.

9.6 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.

9.6.1 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. (8-21)]. 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. (8-21). Let's consider the case where $\ell = 2$, 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,

$$R = -\frac{\mu}{r} \left[J_2 \left(\frac{R_\oplus}{r} \right)^2 P_2[\sin(\phi_{gc})] + \left(\frac{R_\oplus}{r} \right)^2 \left\{ P_{2,1}[\sin(\phi_{gc})][C_{2,1}\cos(\lambda) + S_{2,1}\sin(\lambda)] + P_{2,2}[\sin(\phi_{gc})][C_{2,2}\cos(2\lambda) + S_{2,2}\sin(2\lambda)] \right\} \right]$$

The Legendre polynomial for $\ell = 2$ 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. (6-13), 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-14) 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 that ignoring the appearance of ω and ν changes the disturbing function. 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)

$$\begin{aligned} \overline{\left(\frac{a}{r} \right)^3} &= \frac{1}{2\pi} \int_0^{2\pi} \left(\frac{a}{r} \right)^3 dM = \frac{1}{2\pi} \int_0^{2\pi} \left(\frac{a}{r} \right)^3 \frac{r}{a} dE \\ &= \frac{1}{2\pi} \int_0^{2\pi} \left(\frac{a}{r} \right)^3 \frac{r^2}{a^2 \sqrt{1-e^2}} d\nu = \frac{1}{2\pi(1-e^2)^{3/2}} \int_0^{2\pi} (1+e\cos(\nu)) d\nu \end{aligned}$$

where we transform variables $dM = \frac{r}{a} dE$ and $dE = \frac{r}{a\sqrt{1-e^2}} d\nu$. Thus,

$$\overline{\left(\frac{a}{r} \right)^3} = \frac{1}{(1-e^2)^{3/2}}$$

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

$$R_{avg} = \frac{1}{2\pi} \int_0^{2\pi} R(M) dM = -\frac{3}{2} n^2 R_\oplus^2 J_2 \left(\frac{1}{(1-e^2)^{3/2}} \right) \left(\frac{\sin^2(i)}{2} - \frac{1}{3} \right) \quad (9-36)$$

We can get the same result using Kaula's form of the potential, Eq. (9-28). In this case, $m = 0$ because of zonal harmonics, $q = 0$, and $p = l/2$. 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!

$$\begin{aligned} R_{avg} &= \frac{\mu}{a} \left(\frac{R_\oplus}{a} \right)^2 F_{201} G_{210} \{ C_{20} \cos(\Theta_{2010}) + S_{20} \sin(\Theta_{2010}) \} \\ R_{avg} &= -\frac{\mu J_2}{a} \left(\frac{R_\oplus}{a} \right)^2 \left(\frac{3}{4} \sin^2(i) - \frac{1}{2} \right) \left(\frac{1}{(1-e^2)^{3/2}} \right) \end{aligned}$$

The formula is reconciled with Eq. (9-36) by substituting $n^2 a^3 = \mu$.

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. (9-36) above, do the partial differentiation to get

$$\frac{\partial R_{avg}}{\partial i} = -\frac{3}{2} n^2 R_\oplus^2 J_2 \left(\frac{1}{(1-e^2)^{3/2}} \right) \left(\frac{2 \sin(i) \cos(i)}{2} \right) = -\frac{3}{2} n^2 R_\oplus^2 J_2 \frac{\sin(i) \cos(i)}{(1-e^2)^{3/2}}$$

Combine this expression with the Lagrange VOP to find the secular rate of change,

$$\frac{d\Omega}{dt} = \frac{1}{na^2} \left(-\frac{3}{2} \right) n^2 R_\oplus^2 J_2 \left(\frac{1}{(1-e^2)^2} \right) \cos(i) = -\frac{3nR_\oplus^2 J_2}{2p^2} \cos(i)$$

The answer is

$$\dot{\Omega}_{sec} = -\frac{3nR_\oplus^2 J_2}{2p^2} \cos(i) \quad (9-37)$$

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. 9.2 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. (9-37) 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. (9-37). 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 Ω_{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 right ascension of the ascending node. From Merson (1961), the right ascension of the ascending node rate is

$$\begin{aligned} \dot{\Omega}_{sec} = & -\frac{3J_2R_\oplus^2n\cos(i)}{2p^2} + \frac{3J_2^2R_\oplus^4n\cos(i)}{32p^4} \left\{ 12 - 4e^2 - (80 + 5e^2)\sin^2(i) \right\} \\ & + \frac{15J_4R_\oplus^4n\cos(i)}{32p^4} \left\{ 8 + 12e^2 - (14 + 21e^2)\sin^2(i) \right\} \\ & - \frac{105}{1024} \frac{J_6R_\oplus^6n\cos(i)}{p^6} \left\{ 64 + 160e^2 + 120e^4 - (288 + 720e^2 + 540e^4)\sin^2(i) \right. \\ & \quad \left. + (264 + 660e^2 + 495e^4)\sin^4(i) \right\} \end{aligned} \quad (9-38)$$

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. (9-36). Applying the Lagrange VOP equation, Eq. (9-12), 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. (9-37), 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:

$$\dot{\omega}_{sec} = \frac{3nR_{\oplus}^2J_2}{4p^2} \left\{ 4 - 5\sin^2(i) \right\} \quad (9-39)$$

We can also derive a more complex expression for the argument of perigee, including the first six zonal harmonics:

$$\begin{aligned}
\dot{\omega}_{sec} = & \frac{3nJ_2R_{\oplus}^2}{4p^2} \left\{ 4 - 5\sin^2(i) \right\} \\
& + \frac{9nJ_2^2R_{\oplus}^4}{384p^4} \left\{ 56e^2 + (760 - 36e^2)\sin^2(i) - (890 + 45e^2)\sin^4(i) \right\} \\
& - \frac{15J_4R_{\oplus}^4n}{128p^4} \left\{ 64 + 72e^2 - (248 + 252e^2)\sin^2(i) + (196 + 189e^2)\sin^4(i) \right\} \quad (9-40) \\
& + \frac{105J_6R_{\oplus}^6n}{2048p^6} \left\{ 256 + 960e^2 + 320e^4 - (2048 + 6880e^2 + 2160e^4)\sin^2(i) \right. \\
& \left. + (4128 + 13080e^2 + 3960e^4)\sin^4(i) - (2376 + 14520e^2 + 2145e^4)\sin^6(i) \right\}
\end{aligned}$$

Deriving the mean anomaly introduces some subtle differences. As with the previous derivations, let's begin with the Lagrange planetary equations [Eq. (9-12)]:

$$\frac{dM_o}{dt} = -\frac{1-e^2}{na^2e} \frac{\partial R}{\partial e} - \frac{2}{na} \frac{\partial R}{\partial a}$$

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 $l = 2$ case. Therefore, combining gives us

$$\frac{dM_o}{dt} = -\frac{1-e^2}{na^2e} \left\{ \frac{9}{2} n^2 R_{\oplus}^2 J_2 \frac{e}{(1-e^2)^{5/2}} \left(\frac{\sin^2(i)}{2} - \frac{1}{3} \right) \right\} - \frac{6nR_{\oplus}^2 J_2}{a^2} \left(\frac{1}{(1-e^2)^{3/2}} \right) \left(\frac{3\sin^2(i)}{4} - \frac{1}{2} \right)$$

Simplifying yields

$$\begin{aligned}
\frac{dM_o}{dt} = & \frac{9nR_{\oplus}^2 J_2}{2a^2(1-e^2)^{3/2}} \left(\frac{\sin^2(i)}{2} - \frac{1}{3} \right) - \frac{9nR_{\oplus}^2 J_2}{a^2(1-e^2)^{3/2}} \left(\frac{\sin^2(i)}{2} - \frac{1}{3} \right) \\
\dot{M}_o = & \frac{-3nR_{\oplus}^2 J_2 \sqrt{1-e^2}}{4p^2} \{ 3\sin^2(i) - 2 \} \quad (9-41)
\end{aligned}$$

The mean anomaly's rate of change due to the first six secular zonal harmonics is

$$\begin{aligned}
\dot{M}_o = & \frac{3nR_{\oplus}^2 J_2 \sqrt{1-e^2}}{4p^2} \left\{ 2 - 3\sin^2(i) \right\} \\
& + \frac{3nR_{\oplus}^4 J_2^2}{512p^4 \sqrt{1-e^2}} \left\{ 320e^2 - 280e^4 + (1600 - 1568e^2 + 328e^4)\sin^2(i) \right. \\
& \quad \left. + (-2096 + 1072e^2 + 79e^4)\sin^4(i) \right\} \\
& - \frac{45J_4 R_{\oplus}^4 e^2 n \sqrt{1-e^2}}{128p^4} \left\{ -8 + 40\sin(i) - 35\sin^2(i) \right\} \\
& + \frac{35J_6 R_{\oplus}^6 n \sqrt{1-e^2}}{2048p^6} \left\{ -128 + 320e^2 + 240e^4 + (1344 - 3360e^2 - 2520e^4)\sin(i) \right. \\
& \quad \left. + (-1512 + 3780e^2 + 2835e^4)\sin^2(i) - (-1848 + 4620e^2 + 3465e^4)\sin^4(i) \right\}
\end{aligned} \tag{9-42}$$

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. 9.7.2 and Sec. 9.7.3).

$$\begin{aligned}
\Delta a_{SP} = & \frac{J_2 R_{\oplus}^2}{a} \left[\left(\frac{a}{r} \right)^3 - \frac{1}{(1-e^2)^{3/2}} \right. \\
& \quad \left. + \left\{ -\left(\frac{a}{r} \right)^3 + \frac{1}{(1-e^2)^{3/2}} + \left(\frac{a}{r} \right)^3 \cos(2\omega + 2\nu) \right\} \frac{3\sin^2(i)}{2} \right] \\
\Delta e_{SP} = & \frac{J_2 R_{\oplus}^2}{4} \left[\frac{-2}{a^2 e \sqrt{1-e^2}} + \frac{2a(1-e^2)}{er^3} + \left\{ \frac{3}{a^2 e \sqrt{1-e^2}} - \frac{3a(1-e^2)}{er^3} \right. \right. \\
& \quad \left. - \frac{3(1-e^2)\cos(\nu + 2\omega)}{p^2} - \frac{3\cos(2\nu + 2\omega)}{a^2 e (1-e^2)} \right. \\
& \quad \left. + \frac{3a(1-e^2)\cos(2\nu + 2\omega)}{er^3} - \frac{(1-e^2)\cos(3\nu + 2\omega)}{p^2} \right\} \sin^2(i) \left. \right] \\
\Delta i_{SP} = & \frac{J_2 R_{\oplus}^2 \sin(2i)}{8p^2} \left[3\cos(2\omega + 2\nu) + 3e\cos(2\omega + \nu) + e\cos(2\omega + 3\nu) \right]
\end{aligned}$$

$$\begin{aligned}
\Delta\Omega_{SP} &= -\frac{J_2 R_\oplus^2 \cos(i)}{4p^2} \left[6(\nu - M + e \sin(\nu)) \right. \\
&\quad \left. - 3 \sin(2\omega + 2\nu) - 3e \sin(2\omega + \nu) - e \sin(2\omega + 3\nu) \right] \\
\Delta\omega_{SP} &= \frac{3J_2 R_\oplus^2}{2p^2} \left[\left(2 - \frac{5}{2} \sin^2(i) \right) (\nu - M + e \sin(\nu)) \right. \\
&\quad + \left(1 - \frac{3}{2} \sin^2(i) \right) \left\{ \frac{1}{e} \left(1 - \frac{1}{4} e^2 \right) \sin(\nu) + \frac{1}{2} \sin(2\nu) + \frac{e}{12} \sin(3\nu) \right\} \\
&\quad - \frac{1}{e} \left\{ \frac{1}{4} \sin^2(i) + \left(\frac{1}{2} - \frac{15}{16} \sin^2(i) \right) e^2 \right\} \sin(\nu + 2\omega) \\
&\quad + \frac{e}{16} \sin^2(i) \sin(\nu - 2\omega) - \frac{1}{2} \left(1 - \frac{5}{2} \sin^2(i) \right) \sin(2\nu + 2\omega) \\
&\quad + \frac{1}{e} \left\{ \frac{7}{12} \sin^2(i) - \frac{1}{6} \left(1 - \frac{19}{8} \sin^2(i) \right) e^2 \right\} \sin(3\nu + 2\omega) \\
&\quad \left. + \frac{3}{8} \sin^2(i) \sin(4\nu + 2\omega) + \frac{e}{16} \sin^2(i) \sin(5\nu + 2\omega) \right] \\
\Delta M_{SP} &= \frac{3J_2 R_\oplus^2 \sqrt{1-e^2}}{2ep^2} \left[\left(1 - \frac{3}{2} \sin^2(i) \right) \left\{ \left(1 - \frac{e^2}{4} \right) \sin(\nu) + \frac{e}{2} \sin(2\nu) + \frac{e^2}{12} \sin(3\nu) \right\} \right. \\
&\quad + \sin^2(i) \left\{ \frac{1}{4} \left(1 + \frac{5}{4} e^2 \right) \sin(\nu + 2\omega) - \frac{e^2}{16} \sin(\nu - 2\omega) - \frac{7}{12} \left(1 - \frac{e^2}{28} \right) \sin(3\nu + 2\omega) \right. \\
&\quad \left. \left. - \frac{3e}{8} \sin(4\nu + 2\omega) - \frac{e^2}{16} \sin(5\nu + 2\omega) \right\} \right]
\end{aligned}$$

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. (8-21) 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. 1 for two-body motion (Sec. 1.3.3). 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,

$$h \cos(i) = \text{constant} \quad (9-43)$$

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. (9-43). 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 $\Omega = 0$). We can also write Eq. (9-43) as

$$h \cos(i) = \sqrt{\mu a(1-e^2)} \cos(i) = \text{constant} \quad (9-44)$$

Because changes in semimajor axis and inclination can only be periodic, Eq. (9-44) 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. (9-28), which better enables us to categorize the specific effects. According to Fig. 9-1, only three effects require discussion: secular, short-periodic, and long-periodic. But we can distinguish two other types of perturbations: *m*-dailies and resonance. The starting point for each category is the transformed disturbing-potential function in Eq. (9-28). It's written here with the *S*-function explicitly defined because the analysis proceeds from this point:

$$R = \frac{\mu}{a} \sum_{l=2m=0}^{\infty} \sum_{p=0}^l \left(\frac{R_\oplus}{a}\right)^l \sum_{q=-\infty}^l F_{tmp}(i) G_{lpq}(e) \begin{cases} C_{l,m} \cos(\Theta) + S_{l,m} \sin(\Theta) & \text{if } (l-m) \text{ even} \\ -S_{l,m} \cos(\Theta) + C_{l,m} \sin(\Theta) & \text{if } (l-m) \text{ odd} \end{cases} \quad (9-45)$$

$$\Theta = \Theta_{tmpq} = (l-2p)\omega + (l-2p+q)M + m(\Omega - \theta_{GMST})$$

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. (9-29), 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 Mean Sidereal Time, θ_{GMST} ; 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 9-2 shows the requirements for each category.

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_{tmpq} . Here, we have

$$\dot{\Theta}_{tmpq} = (l-2p)\dot{\omega} + (l-2p+q)\dot{M} + m(\dot{\Omega} - \dot{\theta}_{GMST}) \quad (9-46)$$

TABLE 9-2. 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 ρ	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 Tesselar 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

Now we can recognize \dot{M} is dominated by the mean motion (n) and take $\dot{\theta}_{GMST}$ as the Earth's angular velocity, ω_{\oplus} . We developed sample values for the secular rates of change earlier [$\dot{\omega}$ in Eq. (9-39), $\dot{\Omega}$ in Eq. (9-37), and \dot{M} in Eq. (9-41)]. The equations permit initial analysis of the effects of each element on the trigonometric term. For a typical satellite in low-Earth orbit,

$$\dot{\Omega} = 5^\circ \text{ to } 10^\circ / \text{day}, \dot{\omega} = 5^\circ \text{ to } 10^\circ / \text{day} \quad (9-47)$$

$$\dot{M}_{2-body} = n = 13 \text{ to } 15 \text{ revs / day} = 4700^\circ \text{ to } 5400^\circ / \text{day}$$

$$\dot{\theta}_{GMST} = \omega_{\oplus} \approx 361^\circ / \text{day}$$

These values are used in evaluating the dominant effects on the satellite. Notice that secular, then resonant, and long-periodic terms produce ever smaller effects on the motion. Consider a modified element rate of

$$\dot{c} = \alpha + \beta \cos(\omega_{\oplus} t) + \gamma \cos(\dot{\omega} t) + \delta \cos(nt)$$

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. (9-45), with $m = 0$ and l even. Therefore,

1. $m = 0$ by definition, so $m(\Omega - \theta_{GMST}) = 0$.
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. $\Theta_{tmpq} = 0$ 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. 9.3.1.

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. 9-3). This phenomenon causes regression of the node (for direct orbits) or a counter rotation of the orbital plane around the polar axis (Fig. 9-3). The arrows show the direction of force due to equatorial bulge. An expression for the node's secular rate of change appears in Eq. (9-37).

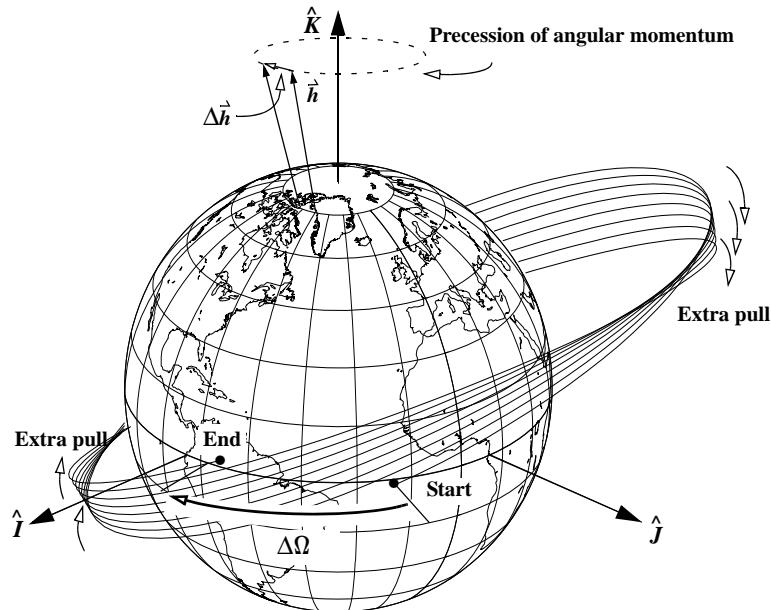


Figure 9-3. How Nodal Regression Works. The effect of a J_2 perturbation that is 20 times the real value shows the precession of the right ascension of the ascending node. The perturbation manifests itself through a change in the angular-momentum vector, and the node regresses much like a precessing top. For this example, the correct interpretation is to say that the satellite's line of nodes experiences a rotation of 50/8° 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, $\Delta\hat{h}$, is parallel to the torque, the precession is parallel to the line of nodes and perpendicular to the angular momentum (Fig. 9-3). 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 I - J direction (at the instant shown, t_{end}).

Figure 9-4 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. (9-37) to determine the change in the node over time (where Ω_o is the initial value of the node):

$$\dot{\Omega} = \Omega_o + \Omega \Delta t$$

Note that Eq. (9-37) is zero for an inclination of 0° which may seem to contradict Fig. 9-4, but remember Ω doesn't exist for equatorial orbits. As we saw with Eq. (9-37) and Eq. (9-38), 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.

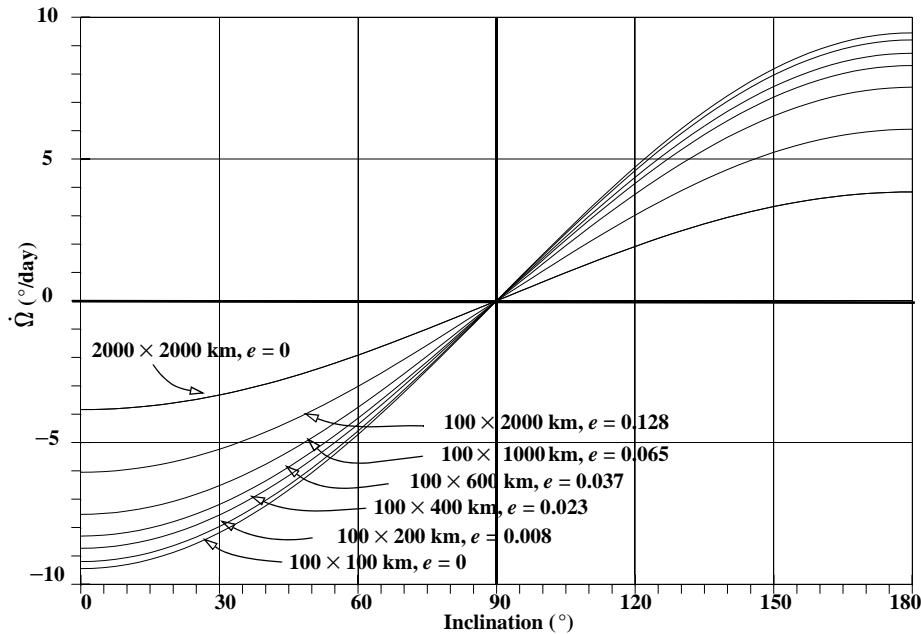


Figure 9-4. Daily Nodal Regression in °/day. For the eccentric orbits, I used a perigee altitude of 100 km with the apogee values as indicated. The 100×2000 km orbit shows that the perturbing effect for each of the *eccentric* orbits would be smaller if the orbit were circular at that apogee altitude (2000×2000 km).

The next major effect of the secular variations caused by J_2 deals with apsidal motion, as Fig. 9-5 shows. Again, it's useful to visualize this effect by examining Fig. 9-6, which I created using Eq. (9-39) and which shows the rate of change for the location of perigee.

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

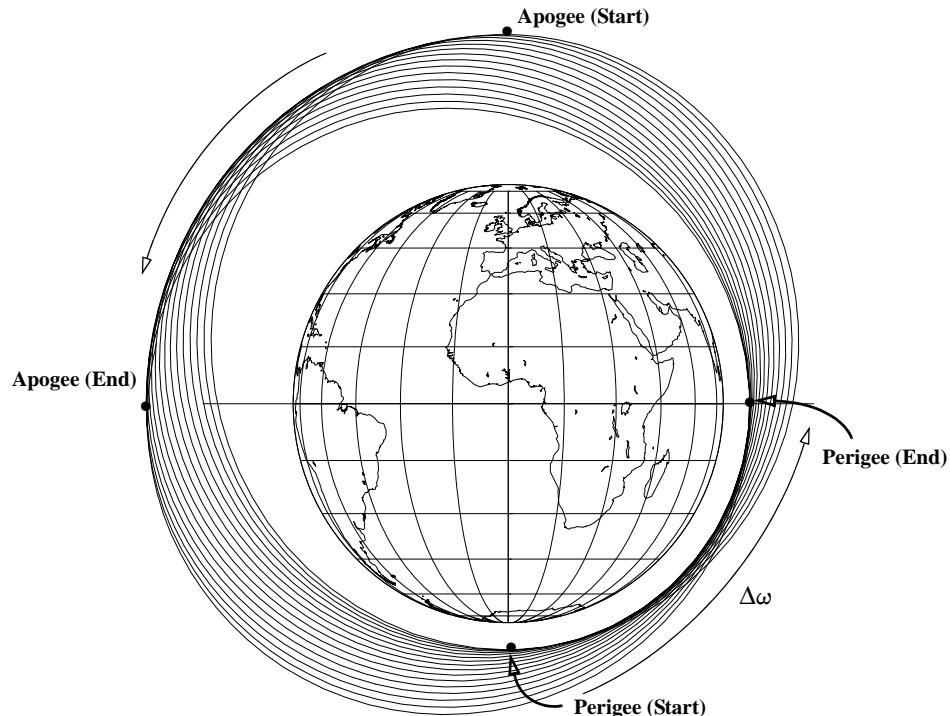


Figure 9-5. How Apsidal Rotation Works. We can see the effect of apsidal rotation after the J_2 perturbation effect increases by a factor of 40 for a polar orbit (to eliminate the nodal regression). Notice how perigee (and apogee) locations change dramatically in a few revolutions.

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 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 40 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

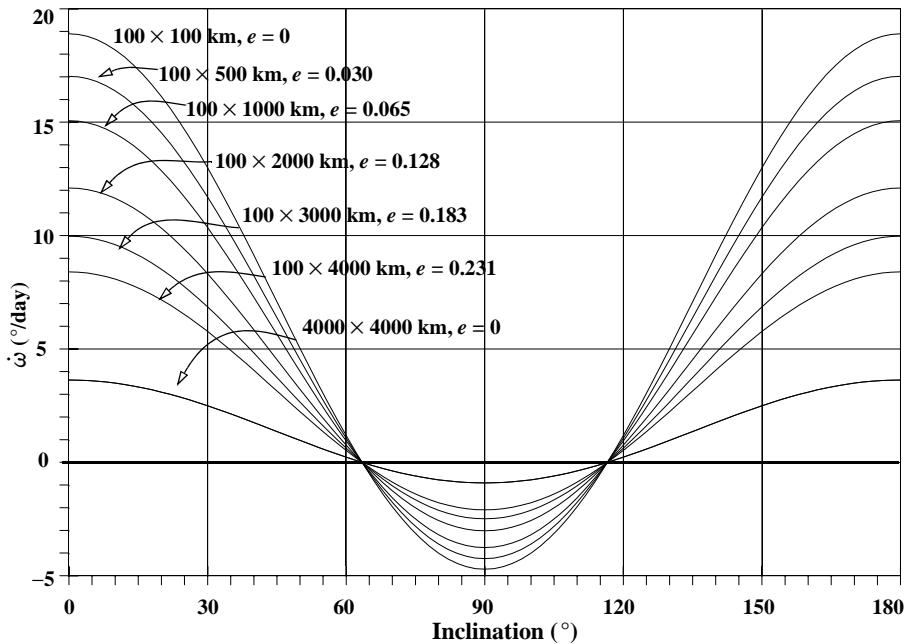


Figure 9-6. Daily Apsidal Regression in °/day. As with nodal regression, circular orbits at an altitude (say, 4000×4000) would have a smaller daily change than an eccentric orbit with apogee at the same altitude (100×4000 km).

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. (9-41). 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. 11.4.2.

Long-Periodic Effects from Zonal Harmonics

Long-periodic variations are especially useful in planning lengthy missions. Remember that the governing equation is Eq. (9-45), 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 $1/2$, $1/3$, $1/4$, 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. (9-39)]. Note that larger or smaller values of the fundamental period can be obtained depending on the value of i .

The magnitude of these effects is 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. 9-7. 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 .

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 $l - 2p = 0$ and $l - 2p + q = 0$. Consequently, only even values of l apply, $p = l/2$, and $q = 2p - l$. To gain insight into m -daily effects, recall the transformed aspherical potential, written here in Kaula's form, Eq. (9-45), with

$$\Theta_{m-daily} = m(\Omega - \theta_{GMST})$$

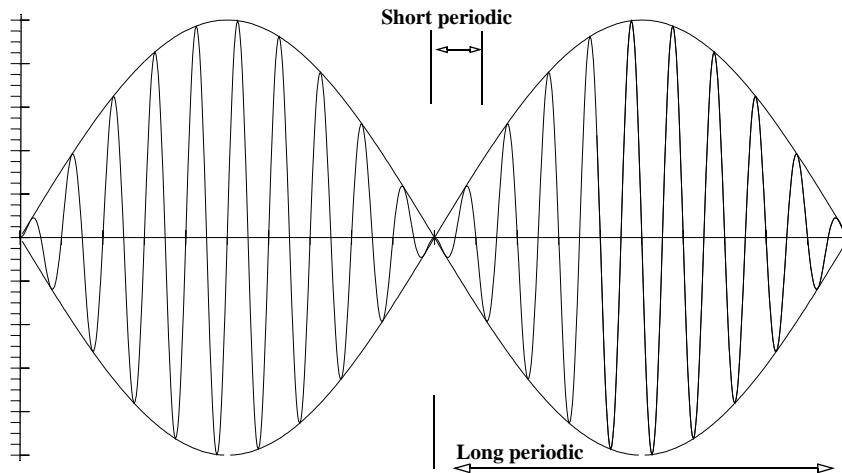


Figure 9-7. Beat-Period Coupling. This figure shows the coupling of short-periodic (SP) and long-periodic (LP) variations for LP/SP beat variations due to zonal harmonics. Over time, a beat phenomenon occurs as a result of combining the two rates. This figure can also represent the coupling of long-periodic and m -daily periods for long-periodic and m -daily beat periods contributed by sectorial and tesseral harmonics.

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 ($\omega_{\oplus} \gg \Omega$), and with Eq. (9-46), the period is approximately

$$\beta_{m\text{-daily}} = \frac{2\pi}{m|\dot{\Omega} - \omega_{\oplus}|} \approx \frac{2\pi}{m\omega_{\oplus}}$$

Only geopotential terms of even degree (ℓ) and all orders ($m \leq \ell$) 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 $q = 2p - \ell$. 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. 9-7 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. 9-7. 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 geosynchronous 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

$$\dot{\Theta}_{lmpq} \approx 0 \Rightarrow (l - 2p)(\dot{\omega} + \dot{M}) \approx m(\dot{\theta}_{GMST} - \dot{\Omega}) = m(\omega_{\oplus} - \dot{\Omega})$$

or

$$\frac{\dot{\omega} + \dot{M}}{\omega_{\oplus} - \dot{\Omega}} \approx \frac{P}{Q}$$

where P and Q are integers, and $Q \neq 0$. We'll see in Ch. 11 that these integers represent the revolutions to repeat (P , an integer), and the days to repeat (Q , also an integer for exact resonance, but it can be a real). The cases where Q is an integer occur very rarely. Those values of l, m, p which satisfy the condition

$$\frac{P}{Q} = \frac{m}{l - 2p}$$

or $m = jP, jQ = (l - 2p)$, $j = 1, 2, 3, \dots$ 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 $l - 2p = 0$ and $q \neq 0$. 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 $m = jP, jQ = q, j = 1, 2, 3, \dots$

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 $G_{lmpq} \sim Oe^{|q|}$ 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,

$$\epsilon_\Theta = Q(\dot{\omega} + \dot{M}) - P(\omega_\oplus - \dot{\Omega}) \quad (9-48)$$

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

$$\rho_{res} = \frac{2\pi}{|\epsilon_\Theta|}$$

This resonance causes periodic variations in the motion of this period and its harmonics with periods ρ_{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, $\Theta \approx \epsilon_\Theta$ in Eq. (9-48). 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, $C_{l,m}$ and $S_{l,m}$, 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 $C_{l,m}$ and $S_{l,m}$ 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 :

$$\frac{P_o}{Q_o} = \alpha_o = \frac{\alpha_o}{1}$$

$$\frac{P_1}{Q_1} = \alpha_o + \frac{1}{\alpha_1} = \frac{1 + \alpha_o \alpha_1}{\alpha_1}$$

$$\frac{P_2}{Q_2} = \alpha_o + \frac{1}{\alpha_1 + \frac{1}{\alpha_2}} = \frac{\alpha_o(1 + \alpha_1 \alpha_2) + \alpha_2}{1 + \alpha_1 \alpha_2}$$

$$\frac{P_3}{Q_3} = \frac{\alpha_o[\alpha_1(1 + \alpha_2\alpha_3) + \alpha_3] + 1 + \alpha_2 + \alpha_3}{\alpha_1(1 + \alpha_2\alpha_3) + \alpha_3}$$

Combining the previous discussions, once P_o , Q_o and P_1 , Q_1 are found for a given revs per day (x)

$$\begin{aligned}\alpha_o &= \text{int}(x) \\ P_o &= \alpha_o \\ Q_o &= 1 \\ rem &= x - P_o \\ \alpha_1 &= \text{int}\left(\frac{1}{rem}\right) \\ P_1 &= 1 + \alpha_o\alpha_1 \\ Q_1 &= \alpha_1\end{aligned}$$

we can compute P_i and Q_i recursively by ($n = 2, 3, \dots$)

$$\begin{aligned}\alpha_n &= \text{int}\left(\frac{1}{rem - \alpha_{n-1}}\right) \\ P_n &= \alpha_n P_{n-1} + P_{n-2} \\ Q_n &= \alpha_n Q_{n-1} + Q_{n-2}\end{aligned}$$

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 $C_{2,2}$ term corresponding to equatorial ellipticity. Recall the zonal harmonics in Fig. 8-5, but in a rotating reference frame with the Earth (Fig. 9-8). 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. 9-8).

On first examination, we might expect the satellite to accelerate in the direction of F . However, the drag paradox (see Sec. 9.6.2 and Eq. (9-51) where satellite velocity opposes the force) 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

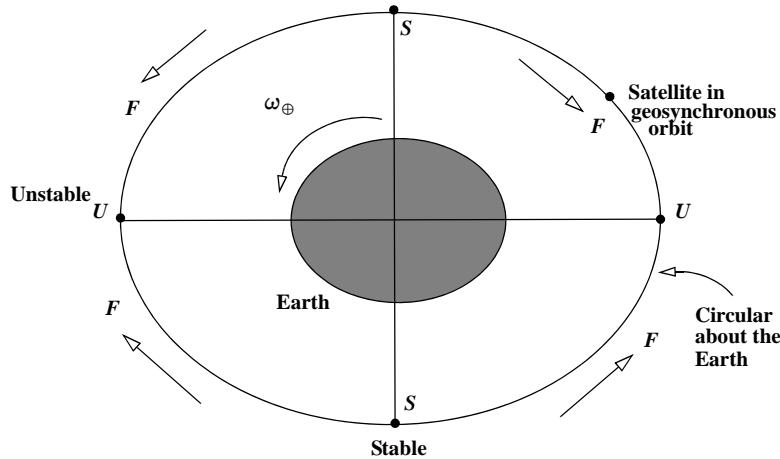


Figure 9-8. Polar View of an Equatorial Section of the Earth. ($C_{2,2}$ only) F is the net tangential force on the satellite at the positions shown. $C_{2,2}$ models a longitudinal asymmetry of the Earth. Both stable (S) and unstable (U) positions are identified.

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 useful 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. 11.2.1) 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.

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. 9-9. 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.

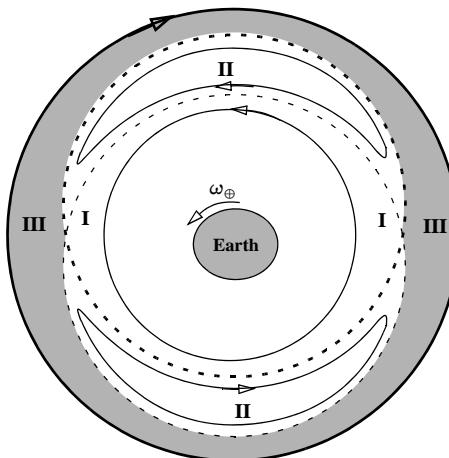


Figure 9-9. Motion of a Geosynchronous Satellite. This figure shows paths of nearly circular, small-inclination orbits in a coordinate frame rotating with Earth, as seen from the North Pole. Arrows show the satellite's direction of motion. Region I: orbital periods are less than 24^h; Region II: periods are 24^h; Region III: periods are greater than 24^h (Kaula, 1966, 54).

2. The main effects on the orbit are the secular motions of the node and perigee. Eq. (9-39) 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.
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. (9-39)] 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 .

9.6.2 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

$$a = a_o - \frac{2a_o}{3n_o} \dot{n} \Delta t \quad (9-49)$$

You can combine this expression with the rate of change for eccentricity to yield an approximate relation:

$$e = e_o - \frac{2(1 - e_o) \dot{n}_o}{3n_o} \Delta t \quad (9-50)$$

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. (9-24)]:

$$\frac{da}{dt} = \frac{2e \sin(\nu)}{n\sqrt{1-e^2}} F_R + \frac{2p}{nr\sqrt{1-e^2}} F_S$$

Drag opposes the velocity vector. Thus, we need to find the drag components in the NWT coordinate system because the T -axis is always aligned with the velocity vector. Unfortunately, the VOP equations are in the RSW system [Eq. (9-24)], requiring a transformation through the flight-path angle. Roy (1988:325) shows the transformation using the relations for the flight-path angle from Eq. (2-95) and Eq. (2-96):

$$\begin{aligned} \frac{da}{dt} &= \frac{2e \sin(\nu)}{n\sqrt{1-e^2}} \left(-\frac{1}{2} \rho \frac{c_D A}{m} v_{rel}^2 \right) \sin(\phi_{fpa}) \\ &\quad + \frac{2p}{nr\sqrt{1-e^2}} \left(-\frac{1}{2} \rho \frac{c_D A}{m} v_{rel}^2 \right) \cos(\phi_{fpa}) \end{aligned}$$

Substituting Eq. (2-95) and Eq. (2-96) 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. (9-16) and the trajectory equation, Eq. (1-24):

$$\frac{dr}{dt} = \dot{r} = \frac{dr}{d\nu} \frac{d\nu}{dt} = \frac{pe \sin(\nu)}{(1+e \cos(\nu))^2 r^2} \frac{h}{p} = \frac{h}{p} e \sin(\nu)$$

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. (9-17):

$$\frac{dt}{d\nu} = \frac{r^2}{na^2\sqrt{1-e^2}} = \frac{a^2(1-e^2)^2}{na^2(1+e\cos(\nu))^2\sqrt{1-e^2}}$$

$$\frac{dt}{d\nu} = \frac{(1-e^2)^{3/2}}{n(1+e\cos(\nu))^2}$$

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 a , and e assuming the drag force components are only in the orbital plane. 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. (2-10) 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 symbol for Bessel functions with the zonal harmonics). Note that the node and inclination have expressions because the drag force is assumed to have a small out of plane component. Remember expressions are based on an exponential model of atmospheric density. Thus,

$$\begin{aligned}\Delta a_{rev} &= -2\pi\delta a^2 \rho_p \left\{ J_o + 2eJ_1 + \frac{3e^2}{4}(J_0 + J_2) + \frac{e^3}{4}(3J_1 + J_3) + 0(e^4) \right\} \text{EXP}[-c] \\ \Delta e_{rev} &= -2\pi\delta a \rho_p \left\{ J_1 + \frac{e}{2}(J_0 + J_2) - \frac{e^2}{8}(5J_1 - J_3) - \frac{e^3}{16}(5J_0 + 4J_2 - J_4) + 0(e^4) \right\} \text{EXP}[-c] \\ \Delta i_{rev} &= -\frac{\pi a \omega_{\oplus} \delta \rho_p}{2n\sqrt{Q}} \text{SIN}(i) \left\{ (J_0 - 2eJ_1 + (J_2 - 2eJ_1)\cos(2\omega) + 0(e^2)) \right\} \text{EXP}[-c] \\ \Delta \Omega_{rev} &= -\frac{\pi a \omega_{\oplus} \delta \rho_p}{2n\sqrt{Q}} \left\{ J_2 - 2eJ_1 + 0(e^2) \right\} \text{SIN}(2\omega) \text{EXP}[-c] \\ \Delta \omega_{rev} &= -\Delta \Omega_{rev} \text{COS}(i)\end{aligned}$$

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, determining a satellite's lifetime can predict how long a mission will last.

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 atmo-

sphere (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, ρ ; 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,

$$\begin{aligned}\xi &= -\mu/2a & \rho &= 2\pi\sqrt{a^3/\mu} \\ KE_{avg} &= \mu/2a = -\xi & PE_{avg} &= -\mu/a = 2\xi \\ n &= \sqrt{\mu/a^3} & v &= \sqrt{\mu/a}\end{aligned}$$

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

$$\Delta\xi = \vec{F} \cdot \vec{dist} = \vec{F} \cdot \vec{v}\Delta t = -F_T v \Delta t \quad (9-51)$$

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:

$$\begin{aligned}\Delta\xi &= \frac{\mu}{2a^2}\Delta a & \Delta a &= \frac{2a^2}{\mu}\Delta\xi \\ \Delta\rho &= 3\pi\sqrt{\frac{a}{\mu}}\Delta a = (3a\rho/\mu)\Delta\xi \\ \Delta KE_{avg} &= -\Delta\xi & \Delta PE_{avg} &= 2\Delta\xi \\ \Delta n &= -\frac{3an}{\mu}\Delta\xi & \Delta n &= -\frac{3v_{cs}}{2a^2}\Delta a \quad (\text{circular orbits only}) \\ \Delta v &= -\sqrt{a/\mu}\Delta\xi\end{aligned} \quad (9-52)$$

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 9-3 summarizes the results.

In the case of drag, the perturbing force directly opposes the motion, so $\Delta\xi < 0$. Accordingly, from Table 9-3, the orbit's size and period will decrease while the mean velocity increases. Because drag is strongest at perigee, where the velocity and atmospheric 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. 9-10. Hence, the *drag paradox* is shown as the effect of atmospheric friction, which *increases* the satellite's average motion as it spirals inward.

TABLE 9-3. 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

Another way to visualize this is as a consequence of Kepler's 2nd Law (equal areas in equal times) and the conservation of angular momentum. As objects are closer to the Earth, they move faster, irrespective of eccentricity. In addition, drag diminishes the kinetic energy so the centrifugal acceleration is no longer sufficient to balance gravity, and the satellites gets pulled towards the Earth, trading potential for kinetic energy until the centrifugal acceleration again balances. This process repeats over and over again.

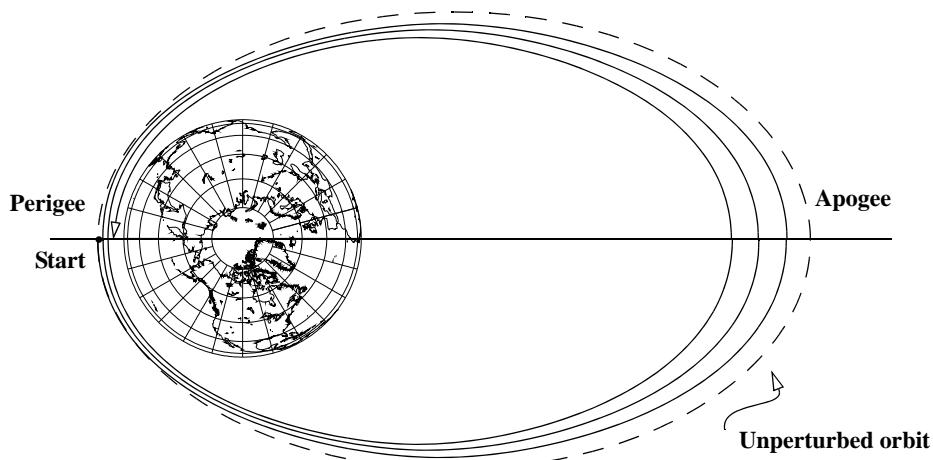


Figure 9-10. Contraction of Orbit Under Drag. This figure is greatly exaggerated but shows the general effect of drag on the satellite's orbit. The radius of perigee tends to remain constant, while apogee shrinks and the eccentricity approaches 0.0.

Obviously, the period can change rapidly for a highly drag-perturbed satellite. The change in orbital period, $\dot{\rho}$, may be calculated to first order from the change in energy or semimajor axis in Eq. (9-52) (remembering that $\rho = 2\pi/n$):

$$\Delta\rho = \frac{3}{2} \frac{\rho}{a} \Delta a = \frac{3\pi}{na} \Delta a \quad (9-53)$$

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 direct implications for a satellite's lifetime. Figure 9-11 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.

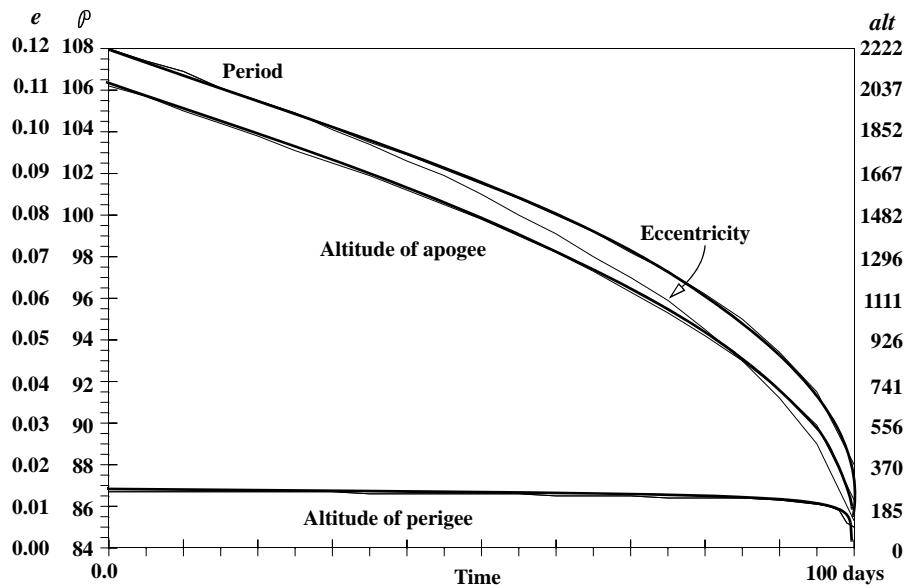


Figure 9-11. Variations of Orbital Properties During a Satellite's Lifetime. In general, certain orbital parameters tend toward “zero,” and eventually the satellite will reenter and crash. The left-hand axes are eccentricity and period (min) respectively, whereas the right-hand scale is altitude for 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 orbit determination (Chap. 10). 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. (8-28). 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.

9.6.3 Third-Body Analysis

Analyzing the third-body perturbations on a satellite's orbit is difficult because the periods are often very long. We assume the use of 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. 9-12, find the direction cosines, A , B , and C :

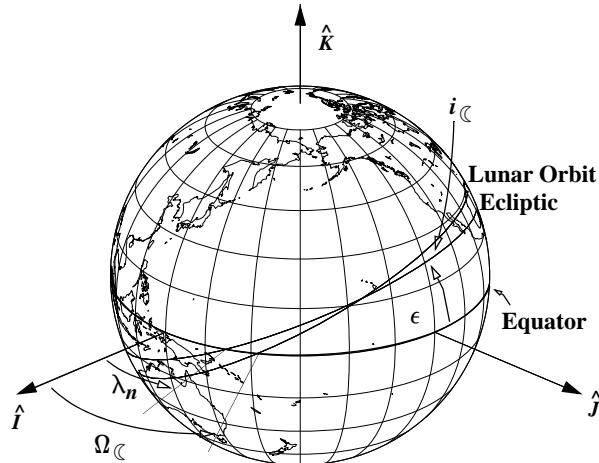


Figure 9-12. Parameters of the Lunar Orbit. This diagram shows the groundtrack of the Moon's orbit about the Earth, along with many of the angles needed to specify its location. Remember that the obliquity, ϵ , is about 23.5° and the Moon is inclined about 5.1° to the ecliptic.

$$\begin{aligned} A &= \cos(\Omega - \Omega_3) \cos(u_3) + \cos(i_3) \sin(u_3) \sin(\Omega - \Omega_3) \\ B &= \cos(i)[-\sin(\Omega - \Omega_3) \cos(u_3) + \cos(i_3) \sin(u_3) \cos(\Omega - \Omega_3)] \\ &\quad + \sin(i) \sin(i_3) \sin(u_3) \\ C &= \sin(i)[\cos(u_3) \sin(\Omega - \Omega_3) - \cos(i_3) \sin(u_3) \cos(\Omega - \Omega_3)] \\ &\quad + \cos(i) \sin(i_3) \sin(u_3) \end{aligned}$$

For the Sun, $i_3 = \epsilon \approx 23.5^\circ$, $\Omega_3 = 0.0^\circ$, $u_3 = \lambda_{M\odot}$, and for the Moon, $i_{\odot} \approx 5.1^\circ$, $\cos(i_3) = \cos(\epsilon) \cos(i_{\odot}) - \sin(\epsilon) \sin(i_{\odot}) \cos(\Omega_{\odot})$, and

$$\sin(\Omega_3) = \frac{\sin(i_{\odot}) \sin(\Omega_{\odot})}{\sin(i_{\odot})}$$

$$u_3 = \lambda_{M\odot} - \Omega_{\odot} + \csc\left(\frac{\sin(\epsilon) \sin(\Omega_{\odot})}{\sin(i_3)}\right)$$

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 (deg/day) of the elements are

$$\begin{aligned}\dot{a} &= 0 & \dot{r}_p &= -a\dot{e} \\ \dot{e} &= -\frac{15\mu_3 e \sqrt{1-e^2}}{4r_3^3 n} \{ 2AB\cos(2\omega) - (A^2 - B^2)\sin(2\omega) \} \\ \dot{i} &= \frac{3\mu_3 C}{4r_3^3 n \sqrt{1-e^2}} \{ A[2 + 3e^2 + 5e^2\cos(2\omega)] + 5Be^2\sin(2\omega) \} \\ \dot{\Omega} &= \frac{3\mu_3 C}{4r_3^3 n \sqrt{1-e^2} \sin(i)} \{ 5Ae^2\sin(2\omega) + B[2 + 3e^2 - 5e^2\cos(2\omega)] \} \\ \dot{\omega} &= -\dot{\Omega}\cos(i) + \frac{3\mu_3 \sqrt{1-e^2}}{2r_3^3 n} \left\{ 5ABS\sin(2\omega) + \frac{5}{2}(A^2 - B^2)\cos(2\omega) \right. \\ &\quad \left. - 1 + \frac{3(A^2 + B^2)}{2} \right\} + \frac{15\mu_3 a(A\cos(\omega) + B\sin(\omega))}{4r_3^3 n e r_3} \left\{ 1 - \frac{5}{4}(A^2 + B^2) \right\}\end{aligned}\tag{9-54}$$

For moderately small eccentricities, second-order terms become noticeable 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)

$$\begin{aligned}\dot{\Omega}_{sec} &= -\frac{3\mu_3(2+3e^2)[2-3\sin^2(i_3)]}{16r_3^3 n \sqrt{1-e^2}} \cos(i) \\ \dot{\omega}_{sec} &= \frac{3\mu_3[2-3\sin^2(i_3)]}{16r_3^3 n \sqrt{1-e^2}} \left\{ e^2 + 4 - 5\sin^2(i) \right\}\end{aligned}\tag{9-55}$$

Smith's (1962) equations, which include terms in e_3 , are

$$\begin{aligned}\dot{\Omega}_{sec} &= -\frac{3\mu_3(1-e^2)^3(1+5e^2)[2-3\sin^2(i_3)]}{8r_3^3n(1-e_3^2)^{3/2}}\cos(i) \\ \dot{\omega}_{sec} &= \frac{3\mu_3(1-e^2)^3[2-3\sin^2(i_3)]}{16r_3^3n(1-e_3^2)^{3/2}}\left\{4-5\sin^2(i)+5e^2\left(3-\frac{7}{2}\sin^2(i)\right)\right\}\end{aligned}\quad (9-56)$$

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. 9-13). 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.

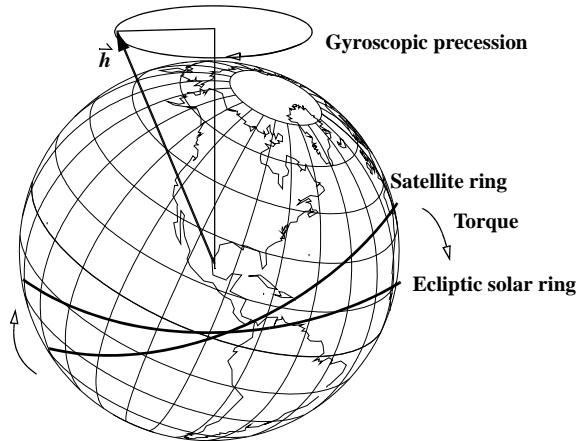


Figure 9-13. Third-Body Interactions. Imagine that the entire mass of a third body (the Sun, for instance) occupies a band about the planet. The resulting torque causes the satellite's orbit to precess like a gyroscope.

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. (9-56) 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 Earth's rotation and the combined perturbative effects of the Sun and Moon cause the tidal force. 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 operational application 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 for one cycle of 0-15° and back in about 53 years. Many geosynchronous satellites launched 20 and 30 years ago that began in equatorial orbits, now have inclinations of up to 15° (in about 26 years). This phenomena poses special challenges for collision avoidance as the satellites drift aimlessly back through the GEO belt. They can easily collide with active satellites. You can consult Chao (2005:62) for additional information.

For satellite orbits that dip into the Earth's atmosphere, a coupling occurs between the luni-solar and drag perturbations. The variation in eccentricity, Eq. (9-54), 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

$$\Delta r_p = \frac{15\mu_{\mathbb{C}}ae\sqrt{1-e^2}}{8nr_{\mathbb{C}}^3} \sin(2\omega_{\mathbb{C}})\sin(i_{\mathbb{C}}) \quad (9-57)$$

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 luni-solar changes of perigee are in the same direction, they increase the effect on a satellite's lifetime. For accurate results, evaluate Eq. (9-54) 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 picking one unit 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) 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. (9-54) 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.

9.6.4 Solar-Radiation Analysis

As mentioned earlier, solar-radiation effects depend on various parameters. We must 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 Cook (1962). He assumes that the disturbing force, F_{SR} , is constant and given by Eq. (8-43) while the satellite is in sunlight and that it is zero in the Earth's shadow. Hoots (2013) and Chao (2005:29-31) also provide these equations.

Cook gives the perturbations in terms of radial, transverse, and normal (R_p , S_p , W) components of the disturbing force, evaluated at perigee as if perigee is in sunlight.*

$$\begin{aligned} R_p &= F_{SR} \left\{ \left[\cos^2\left(\frac{\epsilon}{2}\right) \cos(\omega + \Omega - \lambda_\odot) + \sin^2\left(\frac{\epsilon}{2}\right) \cos(\omega + \Omega + \lambda_\odot) \right] \cos^2\left(\frac{i}{2}\right) \right. \\ &\quad + \left[\cos^2\left(\frac{\epsilon}{2}\right) \cos(\omega - \Omega + \lambda_\odot) + \sin^2\left(\frac{\epsilon}{2}\right) \cos(\omega - \Omega - \lambda_\odot) \right] \sin^2\left(\frac{i}{2}\right) \\ &\quad \left. + \frac{1}{2} \left[\cos(\omega - \lambda_\odot) - \cos(\omega + \lambda_\odot) \right] \sin(i) \sin(\epsilon) \right\} \\ S_p &= -F_{SR} \left\{ \left[\cos^2\left(\frac{\epsilon}{2}\right) \sin(\omega + \Omega - \lambda_\odot) + \sin^2\left(\frac{\epsilon}{2}\right) \sin(\omega + \Omega + \lambda_\odot) \right] \cos^2\left(\frac{i}{2}\right) \right. \\ &\quad + \left[\cos^2\left(\frac{\epsilon}{2}\right) \sin(\omega - \Omega + \lambda_\odot) + \sin^2\left(\frac{\epsilon}{2}\right) \sin(\omega - \Omega - \lambda_\odot) \right] \sin^2\left(\frac{i}{2}\right) \\ &\quad \left. - \frac{1}{2} \left[\sin(\omega + \lambda_\odot) - \sin(\omega - \lambda_\odot) \right] \sin(i) \sin(\epsilon) \right\} \end{aligned}$$

* Because the normal component, W , is constant over the orbit, there is no "p" subscript.

$$\begin{aligned}
W \sin(\omega) &= -\frac{F_{SR}}{2} \left\{ \left[\cos(\omega + \Omega - \lambda_{\odot}) - \cos(\omega - \Omega + \lambda_{\odot}) \right] \sin(i) \cos^2\left(\frac{\epsilon}{2}\right) \right. \\
&\quad + \left[\cos(\omega + \Omega + \lambda_{\odot}) - \cos(\omega - \Omega - \lambda_{\odot}) \right] \sin(i) \sin^2\left(\frac{\epsilon}{2}\right) \\
&\quad \left. + \left[\cos(\omega + \lambda_{\odot}) - \cos(\omega - \lambda_{\odot}) \right] \cos(i) \sin(\epsilon) \right\} \\
W \cos(\omega) &= \frac{F_{SR}}{2} \left\{ \left[\sin(\omega + \Omega - \lambda_{\odot}) - \sin(\omega - \Omega + \lambda_{\odot}) \right] \sin(i) \cos^2\left(\frac{\epsilon}{2}\right) \right. \\
&\quad + \left[\sin(\omega + \Omega + \lambda_{\odot}) - \sin(\omega - \Omega - \lambda_{\odot}) \right] \sin(i) \sin^2\left(\frac{\epsilon}{2}\right) \\
&\quad \left. + \left[\sin(\omega + \lambda_{\odot}) - \sin(\omega - \lambda_{\odot}) \right] \cos(i) \sin(\epsilon) \right\}
\end{aligned}$$

Notice these expressions use the obliquity of the ecliptic, $\epsilon = 23.5^\circ$, and the ecliptic longitude of the Sun, λ_{\odot} . Burns et al. (2000) note that trigonometric substitutions and half-angle transformations are used to find these equations. They also provide the conditions for resonance,

$$\dot{\omega} \pm \dot{\Omega} \pm \dot{\lambda}_{\odot} = 0 \quad \dot{\omega} \pm \dot{\Omega} = 0$$

Hoots (2013) shows how to find the secular and periodic changes in the elements. Beginning with the Lagrange planetary equations [Eq. (9-24)], the true anomaly is changed to eccentric anomaly. Averaging is applied and the secular equations are found by integrating over one revolution with the mean anomaly at the exit and entrance from the shadow. For cases where the satellite enters and exits the Earth's shadow, it's necessary to find the eccentric anomaly (E_{ent} , E_{exit}). Some terms are abbreviated as follows:

$$\begin{aligned}
(\text{COS}\Delta E) &= \cos(E_{ent}) - \cos(E_{exit}) & (\text{SIN}\Delta E) &= \sin(E_{ent}) - \sin(E_{exit}) \\
(\text{COS}2\Delta E) &= \cos(2E_{ent}) - \cos(2E_{exit}) & (\text{SIN}2\Delta E) &= \sin(2E_{ent}) - \sin(2E_{exit}) \\
(\Delta E) &= E_{ent} - E_{exit}
\end{aligned}$$

Cook (1962) used a mix of true and eccentric anomaly in each equation, while Hoots (2013) uses primarily eccentric anomaly. The secular rates of change in the orbital elements and perigee height are taken from Hoots (2013).

$$\begin{aligned}
\dot{a} &= \frac{1}{\pi n} \left[\sqrt{1-e^2} (\text{SIN}\Delta E) S_p + (\text{COS}\Delta E) R_p \right] \\
\dot{e} &= \frac{\sqrt{1-e^2} S_p}{2\pi n a} \left[\frac{3}{2}(\Delta E) + \frac{1}{4}(\text{SIN}2\Delta E) - 2e(\text{SIN}\Delta E) \right] + \frac{\sqrt{1-e^2} R_p}{8\pi n a} [(\text{COS}2\Delta E)]
\end{aligned}$$

$$\begin{aligned}
\dot{r}_p &= \frac{n}{2\pi}[(1-e)\dot{a} - a\dot{e}] \\
\dot{i} &= \frac{\cos(\omega)W}{2\pi na\sqrt{1-e^2}} \left[-\frac{3}{2}e(\Delta E) + (1+e^2)(\sin\Delta E) - \frac{1}{4}e(\sin 2\Delta E) \right] \\
&\quad - \frac{\sin(\omega)W}{2\pi na} \left[-(COS\Delta E) + \frac{1}{4}e(COS 2\Delta E) \right] \\
\dot{\Omega} &= \frac{\cos(\omega)W}{2\pi n a \sin(i)} \left[-(COS\Delta E) + \frac{1}{4}e(COS 2\Delta E) \right] \\
&\quad + \frac{\sin(\omega)W}{2\pi n a \sqrt{1-e^2} \sin(i)} \left[-\frac{3}{2}e(\Delta E) + (1+e^2)(\sin\Delta E) - \frac{1}{4}e(\sin 2\Delta E) \right] \\
\dot{\omega} &= -\frac{\sqrt{1-e^2}R_p}{2\pi n a e} \left[\frac{3}{2}(\Delta E) - e(\sin\Delta E) - \frac{1}{4}(\sin 2\Delta E) \right] \\
&\quad - \frac{S_p}{2\pi n a e} \left[\frac{1}{4}(COS 2\Delta E) - e(COS\Delta E) \right] - \dot{\Omega} \cos(i) \\
\dot{M} &= n - \frac{R_p}{2\pi n a} \left[-\frac{9}{2}e(\Delta E) - \frac{5}{4}e(\sin 2\Delta E) + (5+2e^2)(\sin\Delta E) \right] \\
&\quad - \sqrt{1-e^2} \int_{M_{exit}}^{M_{entry}} (\dot{\omega} + \dot{\Omega} \cos(i)) dM + \frac{5\sqrt{1-e^2}S_p}{2\pi n a} \left[(COS\Delta E) - \frac{1}{4}e(COS 2\Delta E) \right]
\end{aligned}$$

For a satellite in sunlight during the entire orbit, set $E_{ent} = E_{exit} + 360^\circ$. The mean anomaly rate is from Hoots (2013).

$$\begin{aligned}
\dot{a} &= 0 \\
\dot{e} &= \frac{3\sqrt{1-e^2}}{2na} S_p \quad \dot{r}_p = -a\dot{e} \\
\dot{i} &= -\frac{3e\cos(\omega)W}{2na\sqrt{1-e^2}} \\
\dot{\Omega} &= -\frac{3e\sin(\omega)W}{2na\sqrt{1-e^2} \sin(i)} \tag{9-58} \\
\dot{\omega} + \dot{\Omega} \cos(i) &= -\frac{3\sqrt{1-e^2}}{2nae} R_p \\
\dot{M} &= \frac{9eR_p}{2na} - \sqrt{1-e^2}(\dot{\omega} + \dot{\Omega} \cos(i))
\end{aligned}$$

The presence of $\sin(i)$ and eccentricity in the denominator often requires equinoctial (or another non-singular element set) be used in programs. Appropriate methods for determining the eccentric anomaly on entering and exiting the shadow in terms of the satellite's position and the Sun's geocentric coordinates are in Sec. 5.1.1.

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. (9-52),

$$\Delta a = \frac{2a^2}{\mu} \Delta \xi = \frac{2a^2 \hat{\vec{F}} \cdot \Delta \hat{\vec{r}}}{\mu} = -\frac{2a^2}{\mu} F \Delta r_{\odot}$$

where F is given by Eq. (8-43), 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} ,

$$\Delta a = -\frac{2a^2}{\mu} F \Delta r_{\odot} \quad (9-59)$$

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 8-4, 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. 8.6.4). Traditionally, 800 km is regarded as an approximate transition altitude between drag and solar-radiation pressure (4.5×10^{-6} kg·m/s²). These numbers don't match exactly due to the variability in atmospheric density. Many simulations include solar radiation pressure effects to much lower altitudes.

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. However, you can include its effect for much lower altitudes.
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.

9.7 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 set of equations (like Algorithm 64) 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. 9.8.

Realize that each averaging process will produce “mean” elements that are specific to that process. Most analytical techniques use mean elements, but all mean elements aren't the same. In Chap. 8, we saw that the numerical techniques assume osculating vectors that capture the true time-varying behavior (Fig. 9-1) 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.

9.7.1 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. Two variations exist depending 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 a simplified gravitational model (J_2). The overall idea of **PKEPLER** is to determine the classical orbital elements, update the elements through the desired time, and reform the future state vector. Its major difference from Algorithm 7 is that 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:

$$M = M_o + n_o \Delta t + \frac{\dot{n}_o}{2} \Delta t^2 + \frac{\ddot{n}_o}{6} \Delta t^3$$

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. (9-49) and Eq. (9-50).

Secular changes in right ascension of the ascending node and argument of perigee are determined through the simplified equations developed earlier in this chapter [Eq. (9-37) and Eq. (9-39)]. With these newly created elements, we can determine the future (or past) state vector.

Implementing the PKEPLER Solution

We must design the algorithm so special orbits (circular-inclined and equatorial) are possible. The solution is the same as in Algorithm 7. If the orbit is circular, we assign values to the eccentric anomaly—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.

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}_o and \ddot{n}_o , but we often use two-line element sets because of their current availability. Be very careful when using these relations. For decades TLEs from Space Command have transmitted the mean-motion rate *already divided by two*, and the acceleration *divided by six*. Don't divide twice! The algorithm incorporates traditional values, *not* those contained in the TLE. If we don't have an estimate of \dot{n}_o and \ddot{n}_o , we can simply omit the equations using those quantities. But remember, we'll no longer have any approximation for drag in the solution.

I've shown the algorithm with \dot{n}_o and \ddot{n}_o included for general purpose. If these values are unknown, we set them to zero and find a J_2 -only solution using the same equations, less the update for the semimajor axis and eccentricity.

ALGORITHM 65: PKEPLER ($\hat{r}_o, \hat{v}_o, \Delta t, \dot{n}_o, \ddot{n}_o \Rightarrow \hat{r}, \hat{v}$)
RV2COE ($\hat{r}_o, \hat{v}_o \Rightarrow a_o, e_o, i_o, \Omega_o, \omega_o, \nu_o u, \lambda_{true}, \tilde{\omega}_{true}$)
IF $e \neq 0$
vTOANOMALY ($\nu, e \Rightarrow E_o$)

ELSE

$$E_o = u \text{ or } E_o = \lambda_{true}$$

$$M_o = E_o - e_o \sin(E_o)$$

$$p_o = a_o (1 - e_o^2), n_o = \sqrt{\frac{\mu}{a_o^3}}$$

Update for perturbations

$$a = a_o - \frac{2a_o}{3n_o} n_o \Delta t \quad e = e_o - \frac{2(1 - e_o)}{3n_o} n_o \Delta t$$

$$\Omega = \Omega_o - \frac{3n_o R_\oplus^2 J_2}{2p_o^2} \cos(i) \Delta t$$

$$\omega = \omega_o + \frac{3n_o R_\oplus^2 J_2}{4p_o^2} \left\{ 4 - 5 \sin^2(i) \right\} \Delta t$$

$$M = M_o + n_o \Delta t + \frac{\dot{n}_o}{2} \Delta t^2 + \frac{\ddot{n}_o}{6} \Delta t^3$$

$$p = a (1 - e^2)$$

$$\mathbf{KEPEQTNE}(M, e \Rightarrow E)$$

IF $e \neq 0$

$$\mathbf{ANOMALYTor}(E, e \Rightarrow \nu)$$

ELSE

$$u = E \text{ or } \lambda_{true} = E$$

$$\mathbf{COE2RV}(p, e, i_o, \Omega, \omega, \nu(u, \lambda_{true}, \tilde{\omega}_{true}) \Rightarrow \vec{r}, \vec{v})$$

9.7.2 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 AFSPC 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 7: 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. Be aware that the expression for the semimajor axis was chosen because it is “convenient”. 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. (9-41)].

$$\begin{aligned}
 a &= \bar{a} + \Delta a_{SP} & \bar{a} &= a_o \left\{ 1 - \frac{3R_\oplus^2 J_2 \sqrt{1-e_o^2}}{4p^2} (2 - 3\sin^2(i_o)) \right\} \\
 e &= e_o + \Delta e_{LP} + \Delta e_{SP} \\
 i &= i_o + \Delta i_{LP} + \Delta i_{SP} \\
 \omega &= \omega_o + \dot{\omega} \Delta t + \Delta \omega_{LP} + \Delta \omega_{SP} \\
 \Omega &= \Omega_o + \dot{\Omega} \Delta t + \Delta \Omega_{LP} + \Delta \Omega_{SP} \\
 M &= M_o + \bar{n} \Delta t + \Delta M_{SP} & \bar{n} &= n_o \left\{ 1 + \frac{3R_\oplus^2 J_2 \sqrt{1-e_o^2}}{4p^2} (2 - 3\sin^2(i_o)) \right\} \\
 n_o^2 a_o^3 &= \mu
 \end{aligned}$$

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. (9-38) and Eq. (9-40)]; 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

$$\begin{aligned}
 \dot{\Omega}_{J_2^2} \Rightarrow & \frac{3J_2^2 R_\oplus^4 n \cos(i)}{32p^4} \left\{ -36 - 4e^2 + 48\sqrt{1-e^2} \right. \\
 & \left. + (40 - 5e^2 - 72\sqrt{1-e^2}) \sin^2(i) \right\}
 \end{aligned}$$

$$\dot{\omega}_{J_2} = \frac{9nJ_2^2 R_\oplus^4}{384p^4} \left\{ 192 + 56e^2 - 192\sqrt{1-e^2} \right. \\ \left. + (-172 + 288\sqrt{1-e^2})\sin^2(i) + e^2\sin^4(i) \right\}$$

Kozai's expressions for short-periodic variations are in Sec. 9.6.

9.7.3 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 elements, 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. 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 elements 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. 9.7.2 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 elements 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 variations. 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. 9.3), from Kozai theory (Sec. 9.7.2), 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. (9-27)] 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):

$$\begin{aligned} \eta &= \sqrt{1 - \bar{e}^2} & \theta &= \cos(\bar{i}) \\ \gamma_2 &= \frac{J_2 R_\oplus^2}{2\bar{a}^2} & \gamma'_2 &= \frac{\gamma_2}{\eta^4} = \frac{J_2 R_\oplus^2}{2\bar{p}^2} \end{aligned}$$

$$\begin{aligned}\gamma_3 &= \frac{-J_3 R_{\oplus}^3}{\bar{a}^3} & \gamma'_3 &= \frac{\gamma_3}{\eta^6} = \frac{-J_3 R_{\oplus}^3}{\bar{p}^3} \\ \gamma_4 &= \frac{-3J_4 R_{\oplus}^4}{8\bar{a}^4} & \gamma'_4 &= \frac{\gamma_4}{\eta^8} = \frac{-3J_4 R_{\oplus}^4}{8\bar{p}^4} \\ \gamma_5 &= \frac{-J_5 R_{\oplus}^5}{\bar{a}^5} & \gamma'_5 &= \frac{\gamma_5}{\eta^{10}} = \frac{-J_5 R_{\oplus}^5}{\bar{p}^5}\end{aligned}$$

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. (9-38) to Eq. (9-42)] 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 right ascension of the ascending node and argument of perigee (respectively) are

$$\begin{aligned}\dot{\Omega}_{J_2^2} &\Rightarrow \frac{3J_2^2 R_{\oplus}^4 n \cos(i)}{32p^4} \left\{ -36 - 4e^2 - 24\sqrt{1-e^2} + (40 - 5e^2 + 36\sqrt{1-e^2}) \sin^2(i) \right\} \\ \dot{\omega}_{J_2^2} &\Rightarrow \frac{9n J_2^2 R_{\oplus}^4}{384p^4} \left\{ 1062 + 168e^2 + 576\sqrt{1-e^2} + (-2382 - 108e^2 - 1584\sqrt{1-e^2}) \sin^2(i) \right. \\ &\quad \left. + (1290 - 135e^2 + 1080\sqrt{1-e^2}) \sin^4(i) \right\}\end{aligned}$$

The long-periodic variations are identical to Kozai's formulas.

9.7.4 Operational Applications

The U.S. Air Force began the *Simplified General Perturbations* (SGP) model series development in the 1960s (Lane 1965), and it became operational in the early 1970s (Lane and Cranford, 1969). The development culminated in *Simplified General Perturbations 4* (SGP4), and although the name is similar, the mathematical technique is very different from the original SGP technique. The first release of the refined SGP4 propagator source code was Spacetrack Report Number 3 (STR#3, Hoots and Roehrich, 1980). That release resulted from a user compatibility survey of space surveillance operational sites and official users. The magnitude of the resulting variations spurred an effort to promote better compatibility for users. The intent was to get the operational community, as well as ordinary users, synchronized with respect to the implementation. The best vehicle for this was a technical report, including the computer source code. It was designed for the widest possible dissemination. Although most of the equations were given, the use of the source code became common practice for using *Two-line Element* (TLE) sets.

Spacetrack Report Number 3 officially introduced five orbital propagation models to the user community—SGP, SGP4, SDP4, SGP8 and SDP8—all “generally” compatible

with the TLE data. At the time, SGP had just been replaced by SGP4/SDP4 (the latter having included deep-space perturbations). The SGP8/SDP8 model was developed to alleviate deficiencies of SGP4/SDP4 for the special cases of orbital decay and reentry. The approach provided a closed-form solution based on the general trends of orbital elements as they neared reentry, and was quite successful. However, there is no evidence to suggest that SGP8/SDP8 was implemented for operational TLE formation.

After STR#3, Spacetrack Report Number 6 (Hoots, 1986) was publicly released by North American Aerospace Defense Command (NORAD). Some researchers initially assumed this release was intended to update portions of the SDP4 deep-space routines, but the actual intention was to document HANDE* and had little to do with SGP4/SDP4. Nevertheless, it provided amateur satellite trackers and researchers with a confirmation of identified deficiencies in the original validation and verification efforts. This report has not been as widely circulated as STR#3, which benefited from its early electronic availability (Kelso, 1988).

In the early 1990s, the NASA Goddard Space Flight Center (GSFC) obtained a copy of the 1990 standalone SGP4 code from project SpaceTrack as part of a study on orbit propagation models for the SeaWiFS Mission (Patt et al. 1993). In 1996-7 they released the unrestricted code on the Internet and to numerous organizations around the world involved in the SeaWiFS Mission. It confirmed changes already discovered by many independent researchers.

In 1998, Hoots published a history of the equations, background, and technical information on SGP4. In 2004, Hoots et al. published the complete equations (including the deep-space portion). These publications cover the incorporation of resonances, third-body forces, atmospheric drag, and other perturbations into the mathematical technique. We note that all published reports on SGP4 have suggested only improvements in the code used to implement it, and not any changes to the underlying theory. Thus, the equations in Hoots (2004) should be representative of the current mathematical theory. Finally in 2006, Vallado et al. released an updated and combined set of code, test cases, results and analysis for the SGP4 routine. It is available on the Internet at <http://www.centerforspace.com/downloads/>.

The US Navy also leveraged the work of Brouwer and pursued a similar, but different course, with a resulting theory called **Position Partials and Time (PPT2)**. In general, PPT includes additional terms over SGP4 in the expansions. The same modifications for third body and resonance effects from SGP4 were incorporated in the current version, PPT3. Hoots, et al. (2004) discuss some of the specific details. A significant modification appears in the work of Cefola, and Fonte (1996), who have enhanced the U.S. Navy's analytical theory. They include techniques from the Draper Semianalytical Satellite Theory and modify the treatment of *m*-daily effects on the orbits. This

* The HANDE model was intended to replace the analytical SGP4/SDP4 model. It incorporated the effects of the Jacchia dynamic atmosphere models for the average solar flux during the propagation interval, while retaining the speed and character of an analytic general perturbations model. It also included the full Brouwer gravity solution, much of which had been dropped for the SGP4 simplification. The code was implemented in the operational system, but its use is unknown.

change improves the average accuracy of the PPT3 theory for LEO satellites from about 1 km, to about 200-300 m.

Both theories use mean orbital elements and their associated partial derivatives expressed in series expansions of the initial conditions of the differential equations. These analytical theories were designed to operate efficiently on the earliest computers, but the elements are *not* interchangeable. Assumptions were used to simplify these theories. The Earth's gravitational model is truncated to a few zonal terms. The atmospheric model is a static density field with exponential decay, and third-body influences and resonance effects are partially modeled. The coordinate system assumes a "true equator, mean equinox" system—see Sec. 3.7.4 for additional information.

9.8 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). The Russians developed a semianalytical theory about the same time as the Draper theory below. It's comparable to the Draper technique, but we won't discuss it here. Refer to Yurasov (1996) and Nazarenko (1999) for additional information. HANDE and SALT are designed for satellites in very low orbits around the Earth and they are well-suited for that purpose. DSST and the Russian theory are general purpose routines. We'll focus on the Draper Semianalytical Satellite Theory, which is complete, applicable to all orbit classes, and highly flexible. Code is available from Open Source Software Initiatives.*

9.8.1 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. (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.

DSST's basic philosophy is rather simple. We first establish equations of motion for the osculating elements. We then model conservative perturbations through Lagrange's VOP equations [Eq. (9-12)] and nonconservative forces through Gauss's VOP equations [Eq. (9-25)]. 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

$$\begin{aligned} \frac{d\hat{a}_i}{dt} &= n(\hat{a})\delta_{i6} + \epsilon F_i(\hat{a}_1, \hat{a}_2, \dots, \hat{a}_5, \lambda_M) \\ \delta_{i6} &= 0 \quad i \neq 6 \\ \delta_{i6} &= 1 \quad i = 6 \end{aligned} \quad i = 1, 2, \dots, 6 \quad (9-60)$$

We can write the osculating elements as

* Basic DSST computer code is available from <http://www.orekit.org> and <http://tastrody.unirioja.es>.

$$\hat{a}_i = a_i + \eta_i(a_1, a_2, \dots, a_5, \lambda_M) \quad i = 1, 2, \dots, 6 \quad (9-61)$$

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. (9-62). This transformation expresses the osculating elements in terms of mean elements

$$\hat{a}_i = a_i + \sum_{j=1}^m \epsilon^j \eta_{ij}(a_1, a_2, \dots, a_5, \lambda_M) \quad (9-62)$$

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.

We assume the *mean* elements are governed by VOP equations of the form

$$\frac{da_i}{dt} = n(a_1) \delta_{i6} + \sum_{j=1}^m \epsilon^j A_{ij}(a_1, a_2, \dots, a_5, \lambda_M) \quad (9-63)$$

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

$$\begin{aligned} F_i(\hat{a}, \hat{\lambda}_M) &= F_i(a, \lambda_M) + \sum_{j=1}^m \epsilon^j f_{ij}(a_1, a_2, \dots, a_5, \lambda_M) \\ n(\hat{a}_1) &= n(a_1) + \sum_{j=1}^m \epsilon^j N_j(a_1) \end{aligned}$$

$$\begin{aligned}
f_{i1} &= \sum_{j=1}^6 \frac{\partial F_i}{\partial a_j} \eta_{j1} & f_{i2} &= \sum_{j=1}^6 \frac{\partial F_i}{\partial a_j} \eta_{j2} + \frac{1}{2} \sum_{j=1}^6 \sum_{k=1}^6 \frac{\partial F_i}{\partial a_j \partial a_k} \eta_{j1} \eta_{k1} \\
&\dots \\
N_1 &= -\frac{3\eta_{11}}{2a_1} n(a_1) & N_2 &= \left\{ -\frac{3\eta_{12}}{2a_1} + \frac{15}{8} \left(\frac{\eta_{11}}{a_1} \right)^2 \right\} n(a_1) \\
&\dots
\end{aligned}$$

Thus the osculating VOP equations are expressed in the mean elements as

$$\frac{d\hat{a}_i}{dt} = \delta_{i6} \left(n + \sum_{j=1}^m \epsilon^j N_j \right) + \epsilon F_i(a, \lambda_M) + \sum_{j=1}^m \epsilon^{(j+1)} f_{ij}(a_1, a_2, \dots, a_5, \lambda_M) \quad (9-64)$$

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

$$\frac{d\hat{a}_i}{dt} = n(a_1) \delta_{i6} + \sum_{j=1}^m \epsilon^j \left\{ A_{ij} + n(a_1) \frac{\partial \eta_{ij}}{\partial \lambda_M} + \sum_{\ell=1}^{m-j} \sum_{k=1}^6 A_{k\ell} \frac{\partial \eta_{ij}}{\partial a_k} \right\}$$

We now set the right-hand-side of the equation equal to the right-hand-side of Eq. (9-64). 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:

$$\begin{aligned}
A_{i1} + \frac{\partial \eta_{i1}}{\partial \lambda_M} n(a_1) &= F_i(a_1, a_2, \dots, a_5, \lambda_M) + N_1 \delta_{i6} \\
A_{i2} + \frac{\partial \eta_{i1}}{\partial \lambda_M} n(a_1) &= f_{i1}(a_1, a_2, \dots, a_5, \lambda_M) + N_2 \delta_{i6} - \sum_{k=1}^6 \frac{\partial \eta_{i1}}{\partial a_k} A_{k1}
\end{aligned} \quad (9-65)$$

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. (9-65) gives

$$\begin{aligned} A_{i1} &= \langle N_1 \rangle \delta_{i6} + \langle F_i(a_1, a_2, \dots, a_5, \lambda_M) \rangle \quad j = 1 \\ A_{i2} &= \langle N_2 \rangle \delta_{i6} + \langle f_{i1}(a_1, a_2, \dots, a_5, \lambda_M) \rangle \end{aligned} \quad (9-66)$$

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. (9-66) from Eq. (9-65) yields a partial differential equation for the η_{ij} . For $j = 1$, we obtain

$$n(a_1) \frac{\partial \eta_{i1}}{\partial \lambda_M} = N_1^S \delta_{i6} + F_i^S(a, \lambda_M)$$

$$\begin{aligned} F_1^S(a, \lambda_M) &= F_i(a, \lambda_M) - \langle F_i(a, \lambda_M) \rangle \\ N_1^S &= N_i - \langle N_i \rangle \end{aligned}$$

Note that F_1^S and N_1^S have zero mean by definition (i.e., $\langle F_1^S \rangle = 0$). 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. (9-65). 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.

$$\epsilon F_i \Leftarrow \sum_{\alpha} \nu_{\alpha} F_{i\alpha} \quad (9-67)$$

The sum is taken over all the perturbations of interest. Each ν_{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. (9-63)] 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. (9-61), 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 (1983) 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. 10.4.1). 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

$$\partial \hat{\vec{a}} / \partial \hat{\vec{a}}_o = [\mathbf{I} + \mathbf{B}_1] [\mathbf{B}_2 : \mathbf{B}_3] + [0 : \mathbf{B}_4]$$

- \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.

9.9 Practical Considerations

We've examined many techniques to solve for the effect of perturbations on a satellite orbit. It's useful to look at some concepts that transcend the techniques of Chap. 8 and this chapter. In particular, we must address the issue of expected accuracy from any technique including how the force models affect the orbits over time. Remember that *accuracy* represents how good the result is while *precision* is the variability of each result.

A note about units is necessary for this discussion because several conventions exist. When we compare satellite orbits, the units are often simply differences (in meters or kilometers). However, differences in ground station distances to the satellite orbits are also used, presenting other options. The results can also be presented as angular separations seen from a site on the ground. Moreover, 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 9-14 shows the interrelationship among errors in linear units, angular units, and the satellite's altitude.

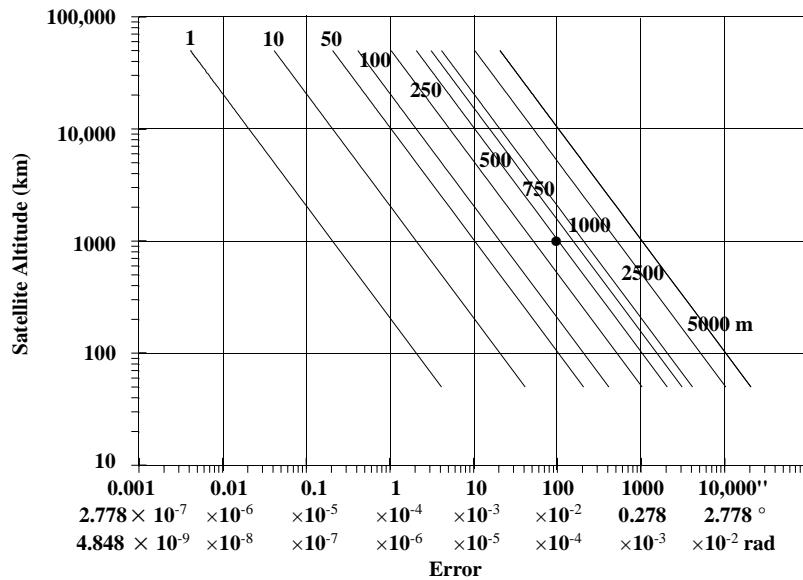


Figure 9-14. Error Values for Various Altitude and Angular Separations. This chart helps us determine the error at the satellite's altitude for a given angular separation in arcseconds, degrees, or radians. For example, a $100''$ (0.0278° or 4.848×10^{-4} rad) angular error for a satellite at 1000 km is about 500 m.

As we design a satellite mission, success requires a certain accuracy from the orbit determination process. This accuracy helps us develop an **error budget** which summarizes acceptable errors that won't negatively impact the mission. Determining an error budget for a particular satellite is a challenging task. Among the inputs are the availability of data, the *true* mission requirements, how the perturbing forces will affect the satellite orbit over time, and how much variability each force model has. Often, requirements are written based on the current system because no other data exists, and rather than flag deficiencies, it's sometimes easier to show compliance than shortcomings.

In determining an orbit from observational data, we should always strive to match the propagator to the precision and accuracy of the available data. Knowles (1995) shows a chart (see Fig. 9-15) which shows notional 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. 10.8.

From the discussions of differential-correction and propagation techniques in Chap. 10, 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 propaga-

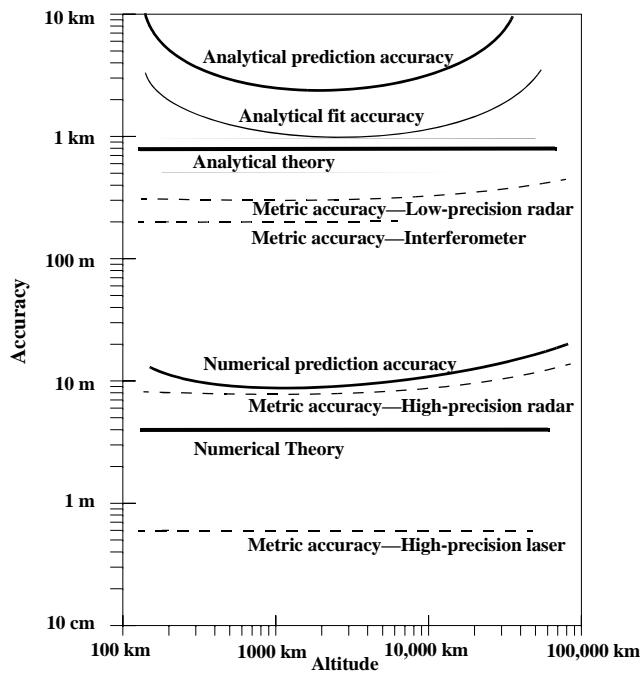


Figure 9-15. Errors in Determining a Satellite's Position. This figure (Knowles, 1995) shows general trends for typical accuracies of numerical and analytical theories (solid lines) for modest propagation times of a few days. Semianalytical theories range from nearly numerical to nearly analytical, depending on their force models and configurations. 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.

tions will further degrade the accuracy, as we'll see in Sec. 10.9. [See also Junkins et al. (1996).] To illustrate the use of Fig. 9-15, 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.

9.9.1 Initial Data

Initial conditions are just as important as the environmental data; however, they usually receive much less attention in the literature. First and foremost, any simulation must have an initial state to use as input data. Of course, the exact form of this state varies widely. Generally, we may view it as being 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 calculate, and the NORAD two-line elements sets. Appendix D.5 lists some sources. Many precise 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, TT, ...); coordinate systems (IAU-76/FK5, GCRF, ...); theories (Kozai, Brouwer, numerical, ...); and assumptions (mean or osculating, two-body, ...). Everything should be consistent!

Confusion often arises concerning the theories. Unfortunately, some codes were developed with little documentation and few configuration controls. As a result, these 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 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 distinct differences. **Two-body elements** 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 potential conversions between osculating vectors, TLE, Kozai, and Brouwer mean elements, and so on. A commonly available data set is the TLE discussed previously (Sec. 2.4.2). For the casual user, these element sets work well with SGP4 [described in Vallado et al. (2006)] and are an easy way to simulate many satellites with some precision. This theory simulates some of the effects of drag using a ballistic coefficient term, 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 frequency contributions), while osculating elements capture all time-varying effects. When the epoch of the element set and the ephemeris generation 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. AFSPC usually uses a simplified version of Brouwer's theory

(SGP4) to determine the TLE, 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. Fortunately, most existing versions of SGP4 contain this 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 that differ substantially from the satellite's instantaneous state vector. However, 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). A subtlety exists. Numerical techniques require precise inputs to generate precise outputs, but the frequency of the perturbations changes as the number and complexity of force models change. Thus, numerical propagators should proceed with the same force models used to generate (estimate) them. 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 semimajor 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

$$\Delta M = -\frac{3n}{2a} \Delta a \Delta t$$

Mean to Osculating Element Conversions

The conversion from *mean* to *osculating* elements is well defined [Eq. (9-61)] for complete analytical theories with short periodic variations, but remember there is a different approach for *every* analytical theory. Kwok (1987) presents a technique for converting mean to osculating elements, considering only J_2 perturbations. Remember that this approach is applicable only for mean elements generated with a J_2 -only theory! While this may seem limiting at first, many instances of this routine exist within the operational community due to the availability of source code and technical documentation.

* In *some* cases an analytical technique may actually be more accurate over a long period of time (many months) than a numerical technique. This is because the mean elements may better capture the long term dominant perturbations than numerical techniques which rely on estimated inputs (EOP, space weather, etc.) to determine all the periodic variations at each time from epoch.

The approach uses Hill's variables and evaluates many of the expressions from Kozai and Brouwer theory (1959).

The short-period variations for inclination, semiparameter, and right ascension of the node are found first. Note that the semimajor axis is not used to bypass potential computational difficulties. All input variables are considered mean elements.

$$\begin{aligned}\Delta i_{SP} &= \frac{J_2 R_\oplus^2 \sin(i) \cos(i)}{4p^2} \left[3 \cos(2u) + 3e \cos(2\omega + \nu) + e \cos(2\omega + 3\nu) \right] \\ \Delta p_{SP} &= \frac{J_2 R_\oplus^2 \sin^2(i)}{2p} \left[3 \cos(2u) + 3e \cos(2\omega + \nu) + e \cos(2\omega + 3\nu) \right] \\ \Delta \Omega_{SP} &= - \frac{J_2 R_\oplus^2 \cos(i)}{4p^2} \left[6(\nu - M + e \sin(\nu)) \right. \\ &\quad \left. - 3 \sin(2u) - 3e \sin(2\omega + \nu) - e \sin(2\omega + 3\nu) \right]\end{aligned}$$

Kwok then finds the short periodic variations in position, radial velocity and argument of latitude.

$$\begin{aligned}\Delta r_{SP} &= - \frac{J_2 R_\oplus^2}{4p} \left[(3 \cos^2(i) - 1) \left\{ \frac{2\sqrt{1-e^2}}{1+e \cos(\nu)} + \frac{e \cos(\nu)}{1+\sqrt{1-e^2}} + 1 \right\} - \sin^2(i) \cos(2u) \right] \\ \Delta \dot{r}_{SP} &= \frac{J_2 \sqrt{\mu} R_\oplus^2}{4p^{5/2}} \left[(3 \cos^2(i) - 1) e \sin(\nu) \left\{ \sqrt{1-e^2} + \frac{(1+e \cos(\nu))^2}{1+\sqrt{1-e^2}} \right\} \right. \\ &\quad \left. - 2 \sin(i)^2 (1-e \cos(\nu))^2 \sin(2u) \right] \\ \Delta u_{SP} &= \frac{J_2 R_\oplus^2}{8p^2} \left[(6 - 30 \cos^2(i))(\nu - M) + 4e \sin(\nu) \left((1 - 6 \cos^2(i)) + \frac{1 - 3 \cos^2(i)}{1 + \sqrt{1 - e^2}} \right) \right. \\ &\quad \left. + \frac{1 - 3 \cos^2(i)}{1 + \sqrt{1 - e^2}} e^2 \sin(2\nu) + (5 \cos^2(i) - 2) 2e \sin(\nu + 2\omega) \right. \\ &\quad \left. + (7 \cos^2(i) - 1) \sin(2u) + 2 \cos^2(i) e \sin(3\nu + 2\omega) \right]\end{aligned}$$

Next, the corrections are added to the original mean elements. Note that a few miscellaneous conversions are needed to recover the classical osculating orbital elements.

$$\begin{aligned}
r &= \frac{p}{1 + e \cos(\nu)} & \dot{r} &= \sqrt{\frac{\mu}{p}} e \sin(\nu) \\
r_{osc} &= r + \Delta r_{SP} & \dot{r}_{osc} &= \dot{r} + \Delta \dot{r}_{SP} & p_{osc} &= p + \Delta p_{SP} \\
A &= \frac{p_{osc}}{r_{osc}} - 1 & B &= \sqrt{\frac{p_{osc}}{\mu}} \dot{r}_{osc} & e_{osc} &= \sqrt{A^2 + B^2} \\
a_{osc} &= \frac{p_{osc}}{1 - e_{osc}^2} & i_{osc} &= i + \Delta i_{SP} & \Omega_{osc} &= \Omega + \Delta \Omega_{SP} \\
u_{osc} &= u + \Delta u_{SP} & \nu_{osc} &= \text{ATAN}\left(\frac{B}{A}\right) \\
\omega_{osc} &= u_{osc} - \nu_{osc}
\end{aligned}$$

Osculating to Mean Element Conversions

Unfortunately, it's more difficult to find mean elements when osculating values are known (you could approximate the equations in the previous section finding the short periodic terms using the osculating input values, but iteration is required). In Chap. 8 we saw that numerical techniques are becoming more popular. Thus, we must understand how to convert an osculating state vector to a mean element set. This situation usually arises because we need the speed of an analytical or semianalytical technique to perform accurate and long term studies. I'll introduce two options.

A **single-point conversion** (or **epoch point**) solves Eq. (9-61) 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. (9-61). 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 formulas 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. 10 for details on estimation). The propagator (numerical, analytical, or semianalytical) generates an ephemeris for a day or two (longer for GEO). The vectors are treated as observations in the least squares process. The differences (residuals) between the numerical, 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. See Sec 10.10 for additional information.

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. (2-99), 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.

9.9.2 Perturbing Force Effects on Satellites

Some organizations use a single value to represent the "error" of a satellite ephemeris—and sometimes at a single time. The most accurate representations use the covariance, and the resulting propagation shows the expected performance over time. In the absence of covariance data, you should understand some general trends. Consider an analysis for two highly elliptical satellites for about 1 revolution in the future, vs. the planning at 3 days ahead for a close conjunction. Because the orbital uncertainty in a single period of the satellite can be about a km or more, the difference at 3 days will be considerably more—definitely important if a maneuver decision is to be made from the data.

Some sensitivity studies exist in the literature. Schutz et al. (1980) examined a comparison of propagation techniques for orbit determination between UT Austin's UTO-PIA and GEODYN. While they focus more on the orbit determination differences, it provided an initial look into the forces affecting propagation. Barker et al. (1996) showed results of gravitational truncation, averaged (RMS) over several days. While this may be useful for long-term trends or viewgraphs, it is inadequate for the mission planner who is concerned about an upcoming maneuver and the selection of which force models to include in any analyses. In addition, statistics of overall satellite catalog per-

formance make simple viewgraph presentations, but they are not sufficient for an operational planner. A more recent study effectively examined the accelerations required for various orbital classes (Register, 2003). Vallado (2005) also studied several orbit classes and gave detailed results—most of which is summarized and expanded in this section.

First, we must understand the propagation setup and modeling capabilities. It's difficult to find consistently accurate POE's for all orbital classes. The following graphs all use Analytical Graphics Inc. Satellite Toolkit High Precision Orbit Propagator (HPOP). The ephemeris generation used a 70×70 WGS-84/EGM-96 gravity field, a NRLMSIS-00 atmosphere with actual EOP and solar weather indices, the complete JPL DE-245 solar and lunar ephemerides, solar-radiation pressure, solid and ocean tides, and albedo to calculate the reference orbits. The initial epoch for all simulations is January 4, 2003, at 00:00 UTC. This time is during moderate solar activity ($F_{10.7} = 140$) to enhance the effects of drag. The coordinate system is J2000 (IAU-76/FK5), and the nutation parameters are calculated at each step. The propagation used a Runge-Kutta 7/8 integrator with relative error set to 1×10^{-15} and a step size of 10 seconds.

We begin with low Earth orbits. Figure 9-16 shows data for the International Space Station at about 380 km altitude. For these orbits, several characteristics are noted. The

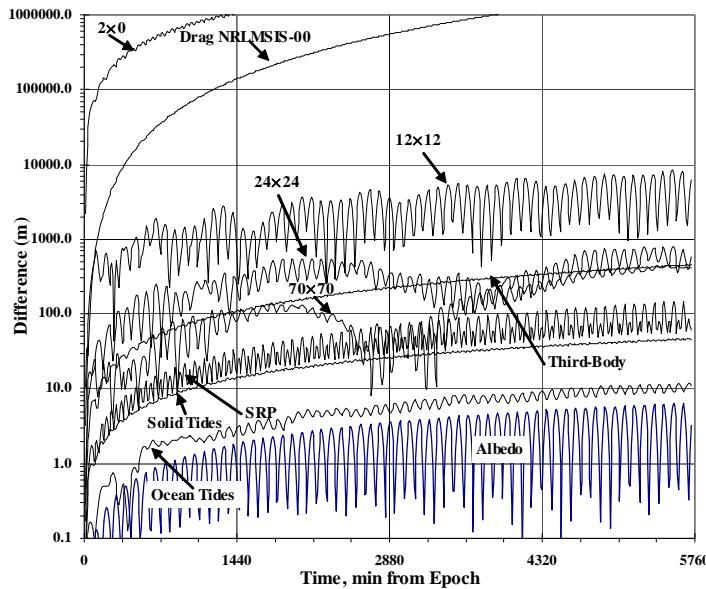


Figure 9-16. Force Model Comparisons — LEO 380 \times 390 km, 51.6°. This figure shows the positional difference over time (four days) from using various force models on the same initial state. Each comparison is made with respect to a two-body ephemeris, except for the gravity runs which compare to the nearest gravity case. Thus, “third-body” is a comparison of a third-body ephemeris to a two-body ephemeris, and “24 \times 24” is a comparison of a WGS-84/EGM-96 24 \times 24 gravity field to a WGS-84/EGM-96 12 \times 12 gravity field ephemeris. This is the ISS, NORAD# 25544.

high contribution of drag to the results are especially prominent. The gravity field also has a large effect. Truncating the gravity field is not recommended as neglecting even the portion from 24×24 to 70×70 can contribute near km-level differences. In fact, third-body forces contribute about as much perturbation as the 70×70 gravity field. Note that tides represent a small contribution, but one that is needed for precise (cm-level) work. Although each line appears to level-off over time, this is only an artifact of the log scale—the errors do continue to grow.

The next satellite is in a slightly higher (about 500 km altitude), sun-synchronous, repeat groundtrack orbit. Many of the same trends exist as in the ISS case, but notice the larger effect of the solid tides. Because the orbit is approximately sun-synchronous, the satellite experiences additional contributions from solid tides as opposed to a satellite at the same altitude, but different inclination. Essentially, the satellite is orbiting over the same locations on the Earth, at similar times during each day. The repeated exposure to the gravitational attraction of the non-spherical Earth produces this effect. Ocean tides do not exhibit the same effect, and the effects are more similar to the ISS results.

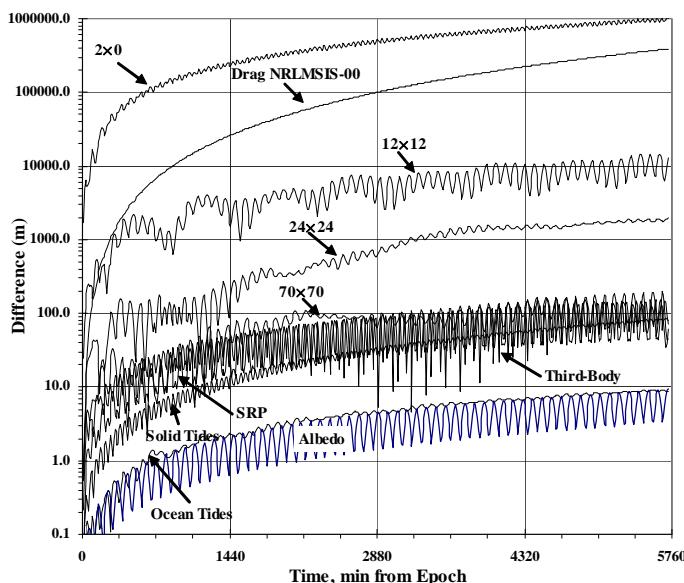


Figure 9-17. Force Model Comparisons — LEO 500×500 km, 97.6° . This figure shows RSS position differences for several forces. This is the JERS satellite, NORAD# 21867 which is also Sun-synchronous.

As satellite altitude increases, the drag contribution drops significantly. Vanguard II was launched in the 1960's to determine cloud cover and atmospheric densities. The orbit is interesting because it is somewhat outside the atmosphere, and it experiences the effects of solar radiation pressure at apogee, and atmospheric drag at perigee. This effect is most dramatic with Molniya satellites, but also shows with this LEO satellite ($e =$

0.15). As the apogee gets larger, the third body and solar radiation pressure forces increase.

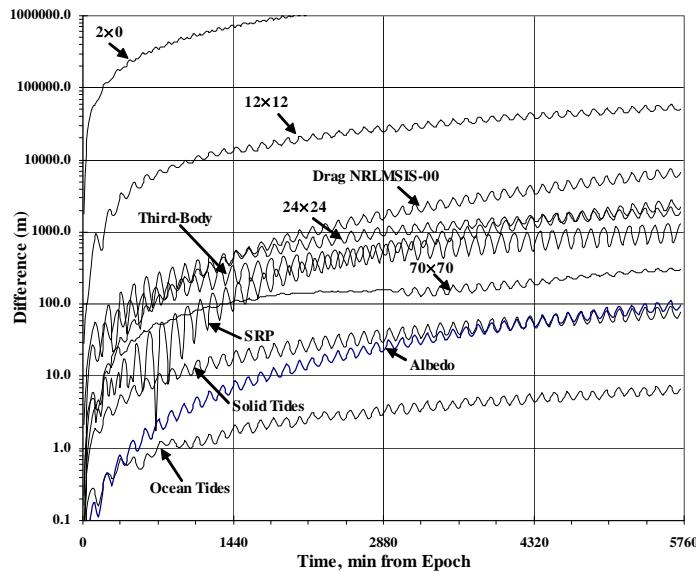


Figure 9-18. Force Model Comparisons — 550 × 3020 km, 32.8°. This figure shows RSS position differences for several forces. This is Vanguard II, NORAD# 11. $e = 0.15$.

Next for TOPEX, in about a 1300 km altitude orbit, the perturbation contributions are much smaller, but non-negligible. The TOPEX satellite, and its follow-on JASON, are used primarily to determine the height of sea-level. This requires very precise orbits, and several forms of tracking and observational measurements are used to determine the orbits. Fortunately, the satellite is above much of the atmosphere, although it still contributes an effect. Solar radiation pressure begins to be more important, and the higher order gravity field begins to be smaller than third-body and solar radiation pressure forces.

Global positioning satellites show the increase in effects from third-body and solar radiation pressure. Note that drag modeling is essentially negligible at this altitude, and much of the gravity field may be ignored. The solar radiation pressure model becomes very important. Several models are used, with the most popular being a flat plate model with various factors (Fliegel et al. 1992). This model is generally called the ROCK 4 model, and it often has subscripts for the Block I and Block II satellites. The ROCK models also account for eclipses (satellites behind the Earth) which can dramatically affect the orbit results.

The Molniya orbit has been of interest ever since it was first used in the early 1960's. The Former Soviet Union used this orbit for communications over its northern latitude landmass. The inclination, near the critical inclination (63.4°), yields little to no preces-

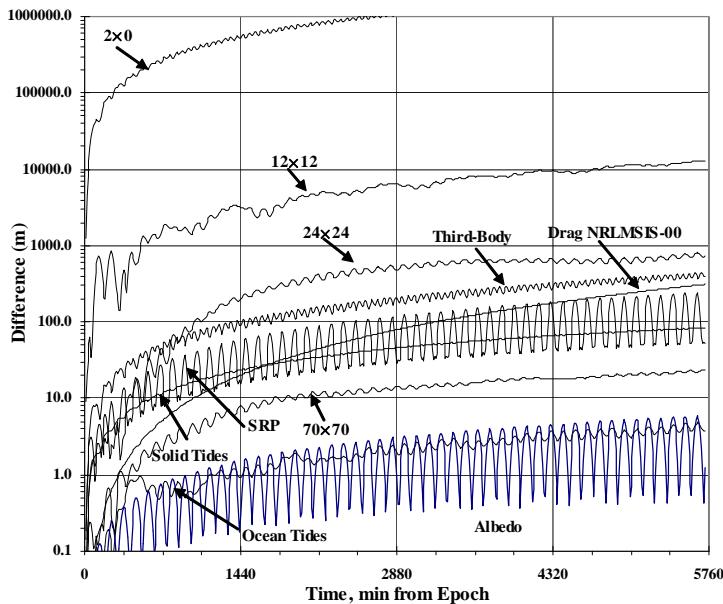


Figure 9-19. Force Model Comparisons — 1300×1300 km, 66° . This figure shows RSS position differences for several forces. This is the TOPEX satellite, NORAD# 22076.

sion of the nodes, and the eccentric orbit means that the satellite apogee could remain over the country for long periods of time. These are among the most difficult orbits to model because they have very high velocities in the atmosphere, increasing the amount of atmospheric drag, and they spend long periods of time near apogee where the third body and solar radiation pressure forces can influence the orbit. The peaks/valleys are associated with atmospheric entry and satellite eclipse. Notice that even the gravitational effects become important with these satellites. Figure 9-21 shows this highly eccentric orbit.

Geosynchronous satellites show a predominant effect from third body and solar radiation. They are also influenced by Earth shadowing and sometimes undergo an “eclipse season”—a period where the satellite spends an appreciable portion of its orbit in the shadow. Notice the smoother behavior of the forces because of the more circular nature of the orbit. Over periods of weeks and months, you can observe slowly varying motion in the orbital elements due to the m -monthly and m -yearly periodic variations and other related terms.

9.9.3 Perturbing Force Variability

In the previous section, we saw that the force models could produce dramatic and dynamic results on the satellite orbits. Here, we'll examine gravity, atmospheric drag, and solar radiation pressure. Note that this section shows prediction differences based

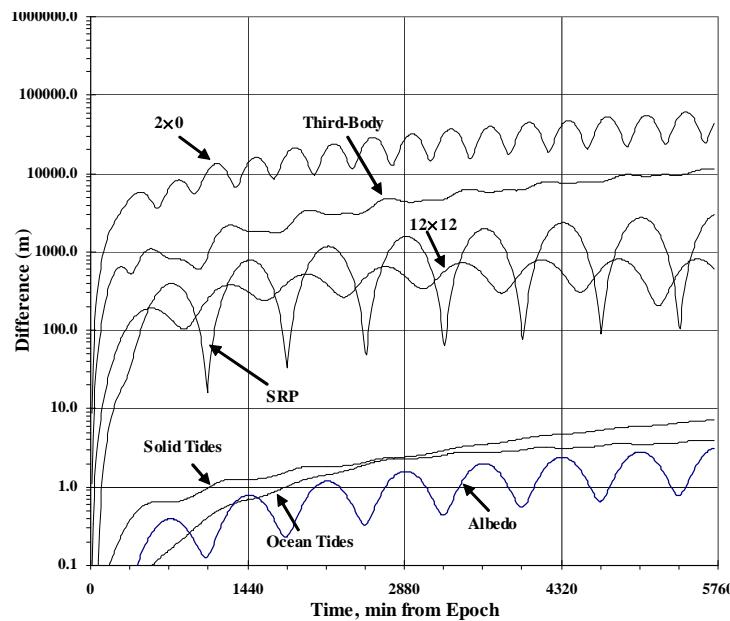


Figure 9-20. Force Model Comparisons — 20000×20000 km, 55° . This figure shows RSS position differences for several forces. This is a typical GPS satellite, NORAD# 26690.

on the propagation of identical state vectors with differing implementations of acceleration models and input data.

Modern gravitational models no longer exhibit the “large” variations sometimes associated with historical models. However, differences can still arise and care should be taken when choosing a particular model. An important difference involves truncating a particular field. While a rigorous approach requires the complete gravitational field, many applications use reduced gravity fields to speed computational processing and to satisfy program limitations. Some operational systems use a single 24×24 field for LEO orbits, rapidly truncating the gravity field as the orbits get higher in altitude. This may not be the best approach to accurately determine the orbit. Barker et al. (1996) suggested a link between accuracy and the zonal truncation. Most studies examine or use a square gravity field in satellite operations.

However, gravity fields are being generated with ever larger sizes (180×180 , 2160×2160 , etc). For “practical” applications, we often truncate these fields (say 70×70 for LEO’s). This truncation works, but the greatest accuracy is achieved only when using the full field. The coefficients will be different if their determination is constrained to 70×70 instead of 2160×2160 . The contributions of the higher order coefficients are small, but they can be noticeable in precise applications. Another contributing factor is the orbits from which the gravity field is determined. If an orbit has an inclination that isn’t well represented in the gravity field solution, the results may differ, and a smaller field may even be better in some cases.

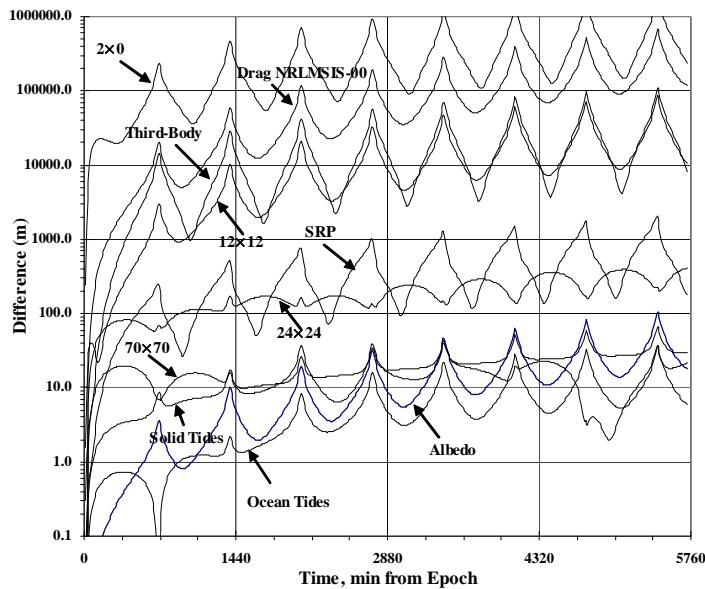


Figure 9-21. Force Model Comparisons — 285 × 38000 km, 62.1°. This figure shows RSS position differences for several forces. This is a Molniya satellite, NORAD# 20052.

Digital gravity models* have been proposed and successfully implemented as an efficient way to use the entire gravity field while simultaneously reducing computational burden. By using the full gravity field, an analyst can avoid the pitfalls of arbitrary gravity field truncation commonly used. They are typically implemented as an overlay to the central body and J_2 zonal acceleration of all remaining gravity components.

One technique (Oltrogge, 2008) uses a multi-purpose subdivided equal-angles grid scheme to minimize the number of grid points (and storage), maps transformations to and from the grid representation, and uses spherical trigonometry and data weighting to interpolate between the nearest three grid points. Tenth-of-a-degree grid spacing ensures sufficient fidelity to accommodate gravity fields with degree and order of up to 3600. The current implementation has been demonstrated to work well with the Runge-Kutta 7-8 variable step integrators.

While the first approach ensures continuity of acceleration at the cell boundaries, some of the more involved multi-step integrators such as Adams-Bashforth may depend upon functional continuity in the third, fourth, and even fifth derivatives as well. A second interpolated gravity representation (Arora and Russell 2011) generates variable-order fitting functions on a rectangular mesh to maintain continuity for the higher-order derivatives. Although multi-step integrators are generally more efficient than single- or

* Digital may also be construed here as interpolated, differential, or gridded because they determine the acceleration at discrete points.

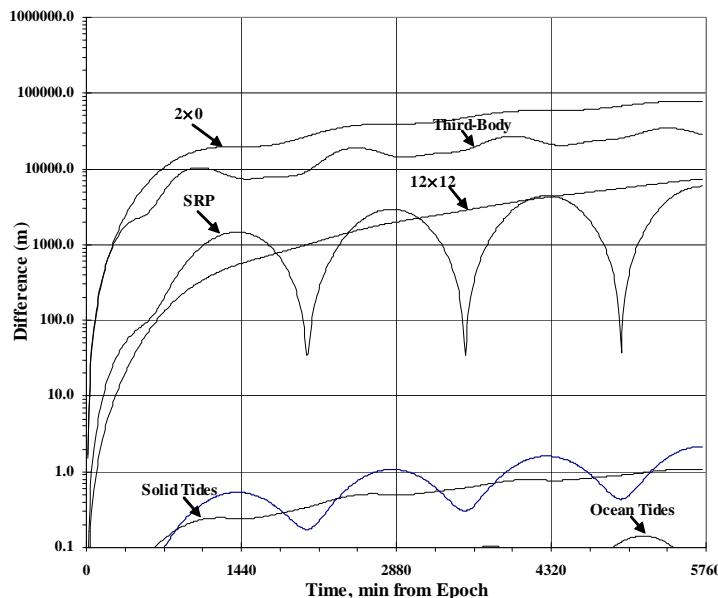


Figure 9-22. Force Model Comparisons — 35780×35780 km, 0° . This figure shows RSS position differences for several forces. This is the Galaxy 4 geosynchronous satellite, NORAD# 26038. Note how the eclipses affect the solar radiation pressure contribution.

intermediate-step methods, such multi-step integrator continuity requirements and other stability considerations may prove problematic depending on the forcing function components being integrated. Both of these techniques show great potential, and a hybrid approach that provides the benefits of both is currently being explored.

Atmospheric density encounters an entirely different set of variations, and from a much larger number of sources. Unlike gravitational models, the various atmospheric density models can vary widely and are strongly dependent on the particular satellite orbit. In addition, the satellite physical parameters (c_D , A , m) can vary substantially, and may not be known for some satellites. Finally, the indices used by the atmospheric models are reasonably well known, but no consensus exists for their use (interpolation, etc.), and prediction is very difficult. The major error sources are listed below (note that density, BC, etc., are not listed as they are derivative effects from the items listed below). This list is generally ordered in decreasing magnitude of effect, although the exact effect will differ for different orbital regimes. (Vallado and Finkleman, 2008)

- Using predicted values of $F_{10.7}$, K_p , a_p for real-time operations
- Lack of satellite attitude in determining the time-varying cross-sectional area
- Not using the actual measurement time for the values ($F_{10.7}$ in particular at 2000 UTC)
- Using step functions for the atmospheric parameters vs interpolation

- Using the trailing or last 81-day average $F_{10.7}$ vs. the central 81-day average
- Using observed or adjusted space weather parameter values
- Using undocumented differences from the original atmospheric model definition
- Not accounting for [possibly] known dynamic effects—changing attitude, molecular interaction with the satellite materials, etc.
- Inherent limitations of the atmospheric models
- Use of differing interpolation techniques for the atmospheric indices
- Using approximations for the satellite altitude, solar position, etc.
- Using K_p or a_p and converting between these values
- Use of $E_{10.7}$ vs $F_{10.7}$ in the atmospheric models (this is not well characterized yet)

Figure 9-23 shows the variability of changing the treatment of the input data and processing within a single (Jacchia-Roberts) model. These particular options come from STK's choices for modeling parameters within HPOP. The values can be held constant (F107DayCon: $F_{10.7}$ only, ConAll: $F_{10.7}$ and K_p , F107AvgCon: $F_{10.7}$ average value only, F107AllCon), the last or trailing $F_{10.7}$ data may be used at different measurement times (Last81 at 1700 UTC, Last81 2000:at 2000 UTC), and the K_p data may use different intervals (Daily:at 0 hrs, 3-hourly-every three hours).

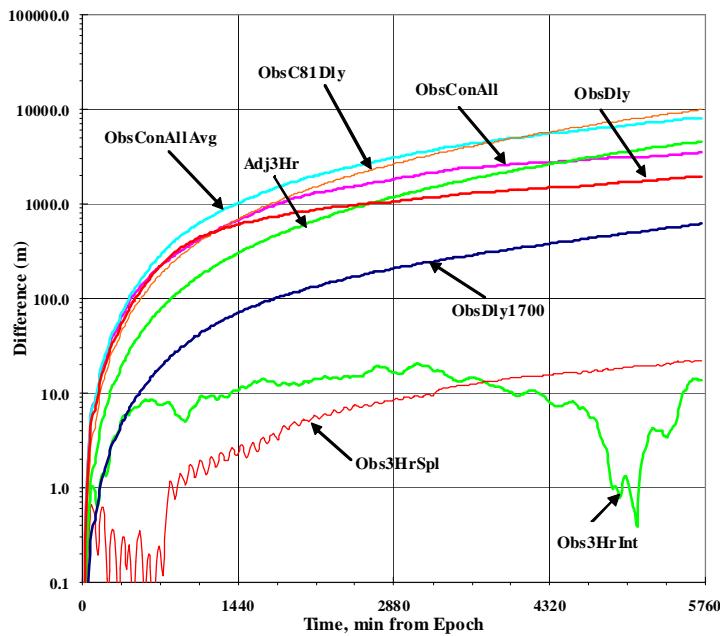


Figure 9-23. Atmospheric Density Data Sensitivity: Positional differences are shown for satellite 21867. Jacchia-Roberts is the baseline for all runs using 3-hourly K_p and interpolation for both K_p and $F_{10.7}$. The plot shows the effect of various options for treating solar weather data. Specific options are discussed in the text.

Equally important is the atmospheric model itself, often having the same or larger variation than the factors listed above. Figure 9-24 shows variations between atmospheric models. The basis for each test was the Jacchia-Roberts model in HPOP. Notice the similarity in the magnitude of the differences to Fig. 9-23.

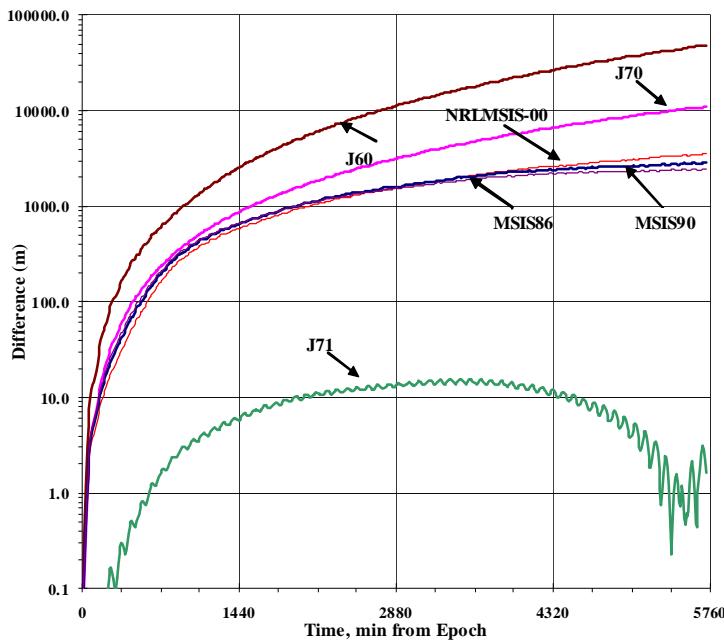


Figure 9-24. Atmospheric Density Model Sensitivity: Positional differences are shown for satellite 21867 for various atmospheric models. Jacchia-Roberts is the baseline for all runs with 3-hourly interpolation. Note the large differences in simply selecting a different model. Also after about 1 day, the smaller transient variations are overcome by the atmospheric differences, and the curves become much smoother.

For most of the comparisons, the MSIS models were quite close, as expected by the model descriptions. The J60 model appeared to be significantly different in all cases from the other models and J70 seemed to differ most from the J71 and JRob models. Note that these tests did not examine comparisons with POEs, so it's most important to come away with the overall level of variability within the different models.

From Fig. 9-23, note that simply changing the solar flux to be a constant—a difference of only a few SFU's—causes nearly a 10 km difference at 4 days (yet we saw earlier that both long and short term predictions can result in Solar Flux changes of 30-50 SFU's!). This indicates that the prediction of solar flux is very important, and if it differs by more than just a few SFU's, it can have dramatic effects on the results of a propagation. The important point to understand from Fig. 9-23 and Fig. 9-24 is that the choice of atmospheric model has as much effect on the overall results as the various ways to use the input data. This is remarkable as many organizations “go to the mat” for a particular

model. However, based on how they use the input data parameters, they may actually obtain worse results than had they simply used the input parameters differently!

Accurate modeling of solar radiation pressure is also challenging for several reasons. A significant factor is the use of a model to include times when the satellite passes through the Earth's shadow. The model may take several forms, including a simple cylindrical model, and one that models precise umbral and penumbral regions. Models also consider the attenuation effect that the Earth's atmosphere has on the shadow region. Some programs accomplish this with an effective Earth radius that is slightly larger (say 23 km) than the Earth's physical size. Depending on the length of time in the shadow, these options can have a dramatic effect on propagations.

Determining the exact times of entry and exit from the Earth's shadow is important as the satellite will usually cross from sunlight to being in the shadow in the middle of an integration step. This usually requires some type of iteration to obtain the precise entry and exit times.

Using a single value for the incoming solar luminosity, or equivalent flux at 1 AU can also alter the resulting perturbing values. There are seasonal variations in the amount of incoming luminosity that can be included in these values.

Finally, using the time for the light to travel from the Sun to modify the satellite position can affect calculations. There are three primary approaches: treating the time from the Sun to the satellite as instantaneous, using the light time delay to the central body, and using the light time delay to the satellite.

Solar radiation pressure (Fig. 9-25) also exhibits some variability, although not as large as the atmospheric density.

9.9.4 Aligning Numerical Propagation Programs

It is often important to transfer data between organizations (e.g., for conjunction analysis or radio frequency interference mitigation). To do so requires understanding of the propagation, data formats, and others. Schutz (1980) lists several factors that are important. Each of these can contribute to differences encountered when comparing programs:

- Inaccurate or mis-applied mathematical models
- Measurement errors
- Truncation error when approximating a mathematical function
- Round-off error resulting from computational precision
- Mathematical model simplifications and approximations
- Human error
- Identifying the precise parameters used in the setup of an individual program
- Treatment of input data for various force models

To effectively perform program comparisons, many runs are made to lay a foundation from which any differences can be understood. However, a building block approach must be used to avoid missing differences that may accumulate from the first tests.

First we examine the propagation techniques as shown in Sec. 9.9.2. The effect of each force model on a particular satellite orbit is examined. This shows the “macro-level” importance that could be attached to each force model for a particular satellite—

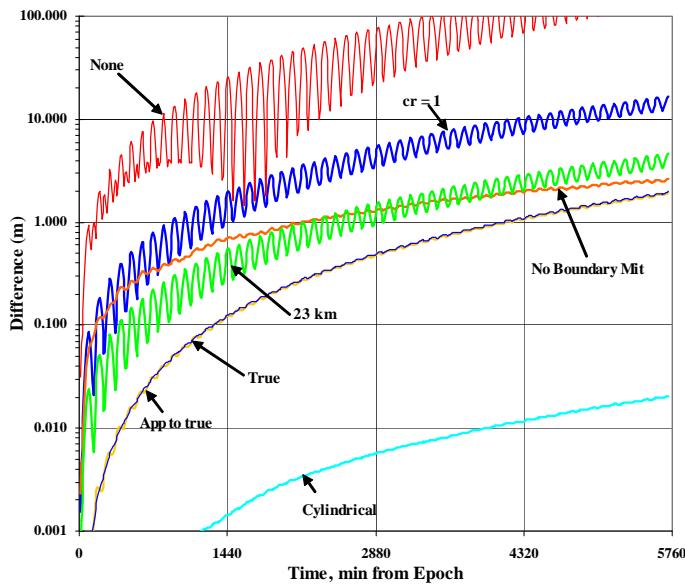


Figure 9-25. Sample Solar Radiation Pressure Sensitivity: Positional differences are shown for the Starlette satellite (NORAD 7646) using options from HPOP. The baseline is a dual-cone (umbra/penumbra) shadow model. Using no shadow model (none) produces the largest differences. A simple cylindrical model introduces modest differences. Shadow boundary mitigation (no boundary) and the effective Earth size (23 km) contribute noticeable differences. The treatment of light travel time between the Sun and central body (app to true) and instantaneous travel (true) produce smaller, but still detectable results.

e.g., examining solid Earth tides for a sun-synchronous orbit may be more important than on a lower altitude, less inclined orbit.

Next, a series of sensitivity studies examining gravitational field truncations, atmospheric drag, and solar radiation pressure differences is performed (Sec. 9.9.3). Ideally, these studies would include Orbit Determination for each test because different force models affect the overall orbital energy.

With an understanding of the preceding tests, the ephemerides themselves are compared. The ephemeris comparisons were easiest to pursue when a series of tests were conducted to build up the force models. In this way, we could determine if a particular force model was contributing the majority of the difference to the overall solution. In general, each of the following tests check a different portion of the programs under consideration. Note that for some of the comparisons, the gravity field was included with other forces once it was examined separately. This also permitted an evaluation of any coupling between force models (such as drag and gravitational acceleration).

- Two-body—Checks numerical integrators, coordinate and time systems
- Gravity field—Checks same Earth radius, gravitational coefficients, etc.

- Two-body plus atmospheric drag—Checks atmospheric density model, solar weather parameter treatment. This can be complex.
- Two-body plus third-body—Checks incorporation of JPL DE/LE files, other files and constants
- Two-body plus solar radiation pressure—Checks Earth shadow assumptions, solar constants, etc.

Conservative forces show remarkable agreement between programs (Fig. 9-26), despite being the largest perturbation causes. Data for several satellites were made available for many of the programs, including some challenging cases. In the cases of GTDS and Raytheon TRACE, the force models were built up in a step-wise manner. This was extremely helpful, and it didn't add very much analysis time. Note that the position and velocity curves essentially "mirror" each other—something that requires equal significant digits in each ephemeris.

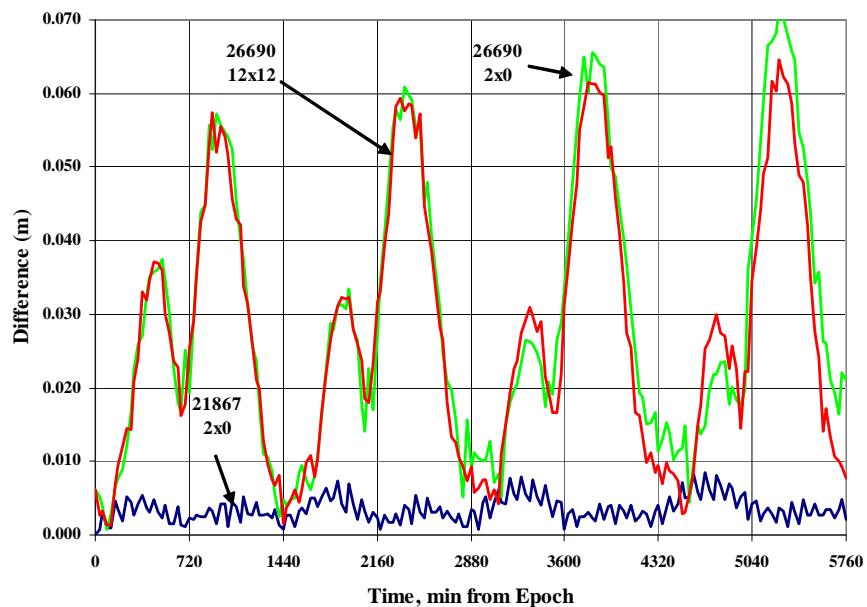


Figure 9-26. GTDS Gravitational Comparisons to STK/HPOP. Positional differences are shown for two satellites (26690-GPS, 21867-JERS). Note that all the results are less than 70 cm after 4 days of propagation.

Although solar radiation pressure is a smaller magnitude effect, it was a non-conservative force that could "easily" be examined and "tweaked" to analyze potential variations. Figure 9-27 shows sample results for comparisons of Raytheon TRACE and HPOP. All runs used dual-cone shadow models. Note that some of these orbits experienced no shadowing during the propagation interval. This makes analysis more difficult as many of the options had no effect because they treat the eclipse periods differently.

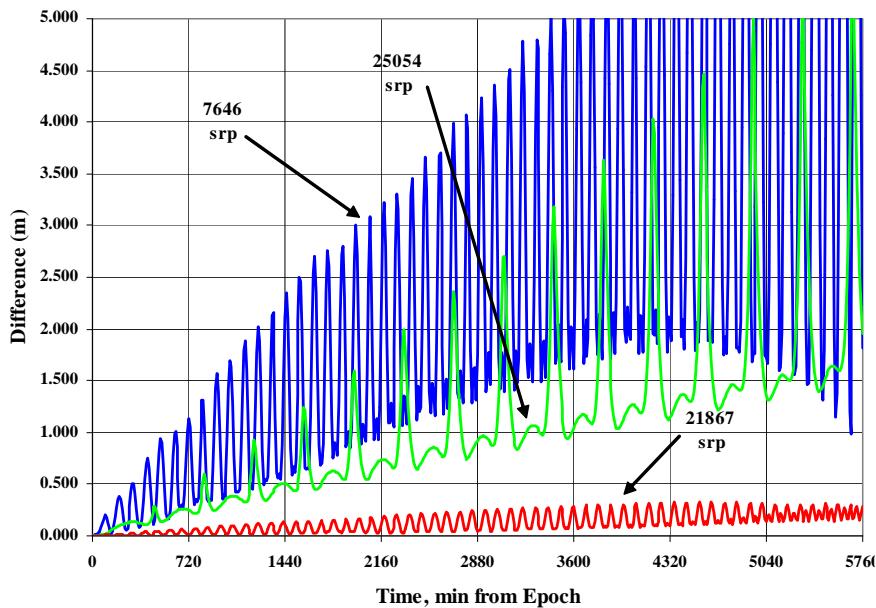


Figure 9-27. TRACE Solar Radiation Pressure Comparisons to STK/HPOP. Several satellite positional differences are shown for solar radiation pressure. Although the results show meter-level differences, these are orders of magnitude below the effects of the dominant other forces.

9.10 Summary of Perturbation Effects

If we categorize various satellite orbits, we can accurately evaluate the perturbing forces effect on each. Understanding which orbital elements are affected by certain types of perturbations shows how the satellite's orbit evolves over time. The combination lets us more efficiently apply perturbations for individual operations, and enables us to better meet mission requirements.

Figure 9-28 shows several types of orbits, to scale, that will be useful in this process.

Most practical orbit types are in Table 9-4. The International Space Station (ISS) is listed with other LEO satellites because the ISS executes numerous station keeping maneuvers to maintain its orbit, while many other LEO satellites do not. Most operational satellites usually drift until the positional error becomes unacceptable for their particular mission. Non-functioning satellites drift continually. Notice some categories aren't very common today, but future technology may make these less popular classes more accessible.

Table 9-5 shows the orbital elements and which perturbing forces affect their values. Notice I've divided the perturbations only into secular, and periodic effects. The method of calculation may not be intuitive at first. The secular values represent the average difference at the end of the 4-day interval of time. Thus over the last orbit, all the difference values are averaged, and this becomes the secular effect for that time interval. For

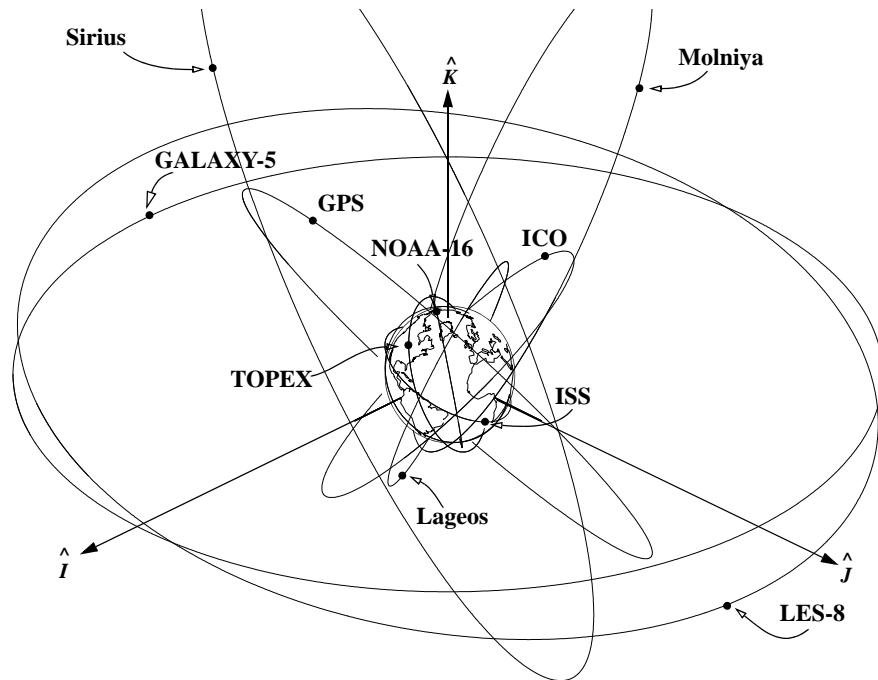


Figure 9-28. Representative Satellite Orbits. This figure shows various satellite orbits to the correct scale. Galaxy-5 is in a GEO orbit that is essentially equatorial, whereas the LES-8 orbit is inclined about 11° due to luni-solar perturbations over many years. Note I've shown only one GPS orbit plane. In reality, 24 satellites are deployed in six planes. Notice the extreme height of apogee for the Molniya orbit with its apogee location over Russia.

the periodic values, the standard deviation of the difference values is found. At one-sigma, this should encompass about 68% of the effect. Other levels are possible if needed.

Table 9-6 shows the effects of perturbations on a variety of satellite orbits. The effects are averaged over each orbital period to give an indication of the relative effect of the perturbing force.

TABLE 9-4. Typical Satellite Classes. This table shows some orbits I've used throughout this book. The shaded cells generally have SLR tracking data available for reference orbit formation. The Molniya, Prognoz, and GEO transfer (GTO) orbits are the only eccentric orbits I've shown because there's too much variety when looking at all possible combinations. SOHO is an interplanetary satellite at the Lagrange point. IKONOS is a 1 m class resolution, color imagery satellite. The categories are often subject to interpretation. The Space Satellite Catalog (SSC) number refers to the designator given by the US Air Force.

Name	Category	Apogee Alt (km)	Perigee Alt (km)	e	i (°)	n (rev/day)	SSC #
ISS (ZARYA)	LEO	406	392	0.001033	51.600	15.560843	25544
Grace 1	LEO	457	439	0.001318	89.000	15.393594	27391
Grace 2	LEO	458	438	0.001465	89.000	15.393594	27392
Terra SAR X	LEO SS	510	507	0.000218	97.500	15.191188	31698
ERS 2	LEO	559	554	0.000361	98.500	15.033736	23560
HST	LEO	561	557	0.000288	28.500	15.025610	20580
SBSS (USA 216)	LEO	633	630	0.000214	97.900	14.793101	37168
Geoeye 1	LEO SS	685	673	0.000850	98.100	14.643999	33331
IKONOS 2	LEO SS	680	678	0.000142	98.100	14.643999	25919
Larets	LEO SS	692	676	0.001133	97.800	14.628450	27944
Landsat 7	LEO SS	704	701	0.000212	98.200	14.571157	25682
Envisat	LEO SS	768	766	0.000140	98.500	14.374299	27386
ERS 1	LEO SS	796	746	0.003497	98.200	14.362237	21574
Stella	LEO SS	805	796	0.000627	98.600	14.273798	22824
SPOT 5	LEO SS	826	824	0.000139	98.700	14.201035	27421
NOAA 16	LEO SS	858	842	0.001107	99.100	14.127423	26536
Starlette	LEO	1107	805	0.020589	49.800	13.822258	7646
Explorer 27	LEO	1296	936	0.024019	41.200	13.381971	1328
Topex	Hi LEO	1343	1332	0.000713	66.000	12.809873	22076
Jason	Hi LEO	1331	1320	0.000714	66.000	12.839816	26997
Jason 2	Hi LEO	1344	1332	0.000778	66.000	12.808628	33105
Ajisai	Hi LEO	1497	1479	0.001144	50.000	12.444007	16908
Vanguard 2	Hi LEO	2960	552	0.148018	32.900	11.834102	11
Lageos 1	MEO	5948	5838	0.004482	109.800	6.386682	8820
Lageos 2	MEO	5951	5616	0.013773	52.700	6.473132	22195
ICO f2	MEO	10622	10578	0.001296	45.000	3.924348	26857
ARIANE 5 R/B	GTO	35383	214	0.727334	2.900	2.309451	37811
Etalon 1	MEO	19169	19082	0.001706	64.400	2.131569	19751
Etalon 2	MEO	19161	19082	0.001549	64.400	2.132070	20026
MOLNIYA 1-93	HEO	38567	1794	0.692298	64.800	2.005829	28163
NAVSTAR 36	GPS	20380	19985	0.007436	53.800	2.005602	23027
NAVSTAR 66	GPS	20200	20164	0.000678	55.000	2.005659	37753
GIOVE-B	GPS	23237	23095	0.002403	55.900	1.709604	32781
Intelsat 906	GEO	35801	35772	0.000344	0.000	1.002721	27513
SES 4	GEO	35792	35782	0.000119	0.000	1.002703	38087
TDRS 10	GEO	35816	35756	0.000712	1.500	1.002739	27566
Sirius 1	SGEO	47138	24438	0.269173	60.600	1.002667	26390
Prognoz 1	Lunar	200002	469	0.935776	64.900	0.249392	5941
Geotail	Lunar	187221	54039	0.524305	35.100	0.191803	22049
SOHO	Heliocentric	Launched December 2, 1995					23726

TABLE 9-5. Summary of Perturbation Effects. This table lists the overall results from the force model comparisons of several satellites (all values are in meters). Two sections are provided—secular and periodic. The secular is the average over the last revolution at the end of the time span, in this case 4 days. The periodic values are the standard deviations of the differences over the last revolution. The baseline for comparison is a two-body orbit. Because the differences can be so large, the gravity cases refer to the previous case, thus “vs 24 × 24” is the difference of a 24 × 24 gravity field, and a 12 × 12 gravity field propagation.

Secular (Average over the last period)								
Name	Apogee alt (km)	Perigee alt (km)	Incl (deg)	vs 2 × 0	vs 12 × 12	vs 24 × 24	vs 70 × 70	vs MSIS00
ISS	389	377	51.6	4179448.02	4475.43	558.03	418.60	2200700.77
JERS	490	475	97.6	559985.25	25983.24	325.83	16.85	222855.71
Starlette	1100	800	49.8	1649332.00	14790.04	1505.60	5.62	1868.78
Vanguard 2	3023	550	32.86	2799810.40	54047.23	1956.43	305.60	6727.46
TOPEX	1347	1340	66.06	2140804.23	12809.29	768.07	22.66	309.47
NAVSTAR 50	20282	20282	55.24	43563.66	521.52	0.00	0.00	0.05
SL-12 R/B	21371	186	46.7	600259.59	17167.00	47.24	372.21	4981564.90
Molnyia 3-35	38026	285	62.05	1649332.00	14790.04	1505.60	5.62	1868.78
Galaxy 11	35790	35785	0.03	69198.46	5779.70	0.06	49125.83	0.00
Periodic (std dev over the last period)								
Name	Apogee alt (km)	Perigee alt (km)	Incl (deg)	vs 2 × 0	vs 12 × 12	vs 24 × 24	vs 70 × 70	vs MSIS00
ISS	389	377	51.6	160767.20	2061.85	113.52	7.85	22565.24
JERS	490	475	97.6	429353.68	18980.24	67.95	7.39	142934.82
Starlette	1100	800	49.8	261369.53	1704.03	22.70	1.73	43.25
Vanguard 2	3023	550	32.86	92256.76	4209.46	218.57	14.44	806.20
TOPEX	1347	1340	66.06	93836.85	100.98	42.77	0.53	4.12
NAVSTAR 50	20282	20282	55.24	12064.61	213.37	0.00	0.00	0.00
SL-12 R/B	21371	186	46.7	184280.90	7621.10	16.55	152.63	2514036.34
Molnyia 3-35	38026	285	62.05	261369.53	1704.03	22.70	1.73	43.25
Galaxy 11	35790	35785	0.03	7567.86	931.75	0.01	4402.17	0.00
Secular (Average over the last period)								
Name	Apogee alt (km)	Perigee alt (km)	Incl (deg)	vs ThirdBody	vs SRP	vs SolidTides	vs OceanTides	vs Albedo
ISS	389	377	51.6	455.60	95.74	18.47	10.60	3.62
JERS	490	475	97.6	29925.24	725.71	29.23	12.11	34.99
Starlette	1100	800	49.8	200.95	121.35	76.73	5.10	11.97
Vanguard 2	3023	550	32.86	1008.79	2443.25	55.92	6.52	97.51
TOPEX	1347	1340	66.06	414.42	142.41	67.08	4.31	3.54
NAVSTAR 50	20282	20282	55.24	10137.84	1632.52	3.93	6.58	1.95
SL-12 R/B	21371	186	46.7	5996.97	407.38	11.28	1.50	6.20
Molnyia 3-35	38026	285	62.05	200.95	121.35	76.73	5.10	11.97
Galaxy 11	35790	35785	0.03	27247.77	2849.94	1.52	0.08	1.21
Periodic (std dev over the last period)								
Name	Apogee alt (km)	Perigee alt (km)	Incl (deg)	vs ThirdBody	vs SRP	vs SolidTides	vs OceanTides	vs Albedo
ISS	389	377	51.6	3.64	33.29	5.04	1.03	2.23
JERS	490	475	97.6	26313.26	487.30	5.42	7.85	22.69
Starlette	1100	800	49.8	32.78	33.94	1.97	0.70	2.62
Vanguard 2	3023	550	32.86	222.20	311.03	11.72	0.71	12.05
TOPEX	1347	1340	66.06	24.41	79.76	1.81	0.44	2.03
NAVSTAR 50	20282	20282	55.24	832.04	972.10	0.18	0.42	0.80
SL-12 R/B	21371	186	46.7	2242.38	168.06	1.66	0.63	4.10
Molnyia 3-35	38026	285	62.05	32.78	33.94	1.97	0.70	2.62
Galaxy 11	35790	35785	0.03	4819.47	1840.33	0.17	0.05	0.56

TABLE 9-6. Summary of Perturbation Effects. This table shows the general effects for the perturbing forces on the orbital elements. Notes are also listed with each perturbing force.

Element	All zonal	Even zonals	Sectorial Tesseral	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 		
m-daily						
<ul style="list-style-type: none"> - Even values of l only - if $m = 1$, $\theta = 24^\circ$; $m = 2$, $\theta = 12^\circ$ - Higher harmonics ($m > 14$) are hard to distinguish from Zonal SP 						
LP/SP/Beat						
<ul style="list-style-type: none"> - LP drives amplitude, SP oscillates within 						
SECTORIAL/TESSERAL:						
<ul style="list-style-type: none"> - Most numerous - SP drives LP amplitude, MD oscillates within 						
Resonance						
<ul style="list-style-type: none"> - 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 tesserals only -- Stable GEO, $\lambda = 75.3^\circ, 255.3^\circ$ -- Unstable GEO, $\lambda = -14.7^\circ, 165.3^\circ$ 						

Problems

1. Herrick (1972, Chapters 16 and 17) introduces a grave symbol (e.g., \grave{a}) 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. 8 and 9, 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 altitude, circular, equatorial orbit? Does the inclination make a difference? Why? Is your answer different for an eccentric orbit ($e = 0.005$)? 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 Ω 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 7 in Chap. 2?
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. 10)
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 TLE 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)].

730

GENERAL PERTURBATION TECHNIQUES

9.10

CHAPTER 10 ORBIT DETERMINATION AND ESTIMATION

- 10.1 Historical Background
- 10.2 Linear Least Squares
- 10.3 Nonlinear Least Squares
- 10.4 Application: Orbit Determination With Differential Correction
- 10.5 Sequential-Batch Least Squares
- 10.6 Kalman Filtering
- 10.7 Calculating the Matrices
- 10.8 Implementing Differential-Correction Solutions
- 10.9 Practical Considerations

10.1 Historical Background

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 innovations took place between the original foundations and current theories, but the fundamental principles are the same.

Gauss's theory, Kalman filters, and so forth, ultimately needed more precise data. 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 estimating the orbit of a distant minor planet about the Sun, he may never have published his results. The observations wouldn't have been accurate enough to support his calculations. Fortunately, Ceres was an appropriate subject for the observational techniques of the time.

Much is made of the rediscovery of the asteroid Ceres in 1801. This event 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 margin of error.

The events leading to this discovery are 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) independently formulated his own results in 1806 and published them 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 least squares, 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! Gauss 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 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 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 [least squares] method 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 Wiener (1894–1964) in 1941 and 1942, respectively. According to Sorenson, the independent approach of Kolmogorov and Wiener 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 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, no one made anything of their similarities at first. But later, Kalman filtering became more popular, and many researchers have 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.

The United States' first modern event in space *surveillance* occurred when the Millstone Hill radar detected signals of Sputnik I. It's reasonable to assume that the Russians had orbit determination methods in place on, or about the same time. 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 to clarify J_2 just two weeks after launch. Surveillance *tracking** of a satellite began for the United States when the Millstone Hill radar successfully tracked Sputnik II. Over the next decades, many types of radars and sensor devices were designed and built, mainly to detect ballistic missiles. During the Cold War, apprehension about the possibility of a nuclear first strike made early warning a priority mission. Even today, some tracking systems are better suited for missile detection than for space surveillance. Still, the designers of the

* Tracking is detecting signals, finding the orbit, and predicting a future position.

original systems over-designed the sensors, so they continue to work remarkably well for both missions.

Orbit determination from observations remains a key operation today. Because objects as small as flecks of paint have damaged the Space Shuttle's windows, accurately determining the orbits of small objects is becoming increasingly important. The use of laser technology for tracking satellites, resupplying power, and communicating also demands accuracy. 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 orbit determination and some practical constraints. A modern challenge is the ability to accurately predict an object's position and velocity, including realistic measures of uncertainty. Orbit determination requires estimation, which is intimately tied to *initial orbit determination* (Chap. 7), *prediction* (Chap. 8 and Chap. 9), and *uncertainty estimates*. We'll begin with some general terms.

Regardless of the observational data source used for orbit determination, we should make our algorithms general enough to process them. 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. This process incorporates the propagation techniques of Chap. 2. 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 and sequential-batch methods that update the state and covariance matrices. Finally, ***smoothing*** techniques improve previous state solutions by combining them with future data. We sometimes apply this process to the observations themselves. Least squares and traditional sequential-batch least squares are smoothing techniques because they determine the state at a particular epoch time.*

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 distinction 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 are incompletely known, so this method is unacceptable for highly accurate missions. On the other hand, the ***stochastic*** approach (involving or containing a random variable, a chance, or a probability) 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. The stochastic approach combines 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 ways to determine an orbit. The underlying approach is stochastic because we use additional observations which include noise to estimate the

* For Kalman-filter smoothing, additional considerations arise. We won't consider these in detail. Consult Gelb (1989:200-203) and Bierman (1977:211-232) for more information.

best answer. *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 directly, so we modify the linear least squares approach with an iterative processing solution called differential correction. *Sequential least squares* is based on least-squares evaluations of sequential batches of observations. It uses statistical information from previous least-squares solutions to 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 the term "differential correction" is used interchangeably to describe 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 terminology with the sequential least-squares case to distinguish it from Kalman-filtering applications which are more sequential in nature.

10.2 Linear Least Squares

Linear unweighted least squares is the simplest estimation technique. It assumes that all data is given equal weighting, or importance. We'll use an example problem to discuss the least squares 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. 10-1.

The eight data points shown are our observations, but the data contain errors. How do we find $y = f(x)$? Two possible solutions are

1. Fit a seventh-order polynomial through the eight points (deterministic).
2. By understanding the problem's dynamical nature, choose an appropriate mathematical model and fit the measurements "as well as possible."

The latter solution is essentially an optimization problem, defining "as well as possible" to mean "minimize the sum of the squares of the residuals" (least squares). The residuals will be the difference in the actual observations and those obtained using the state vector solution. The first step in the solution is to define all variables:

x_o	\Rightarrow	Observed values of the <i>independent</i> variable
y_o	\Rightarrow	Observed values of the <i>dependent</i> variable
N	\Rightarrow	Number of data points or observations
i	\Rightarrow	i^{th} data point subscript
y_c	\Rightarrow	Computed value of the dependent variable, based on the independent variable, x . $y = f(x)$

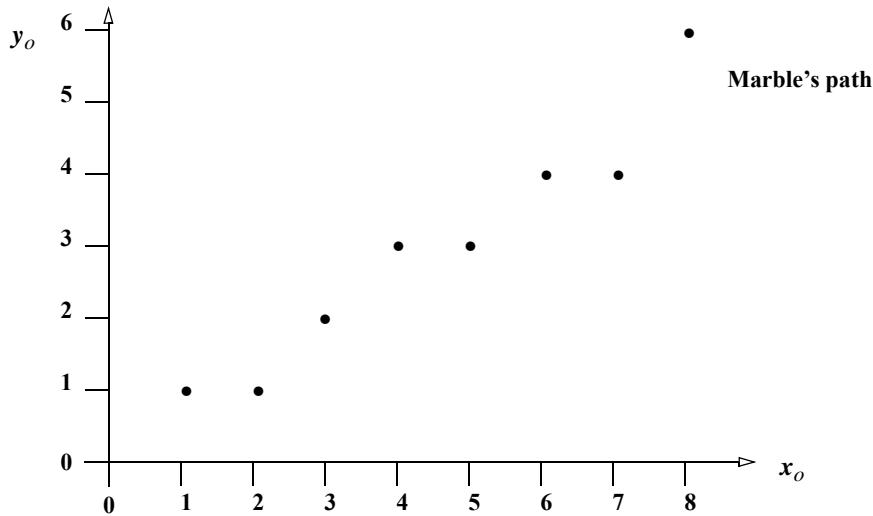


Figure 10-1. Path of a Marble Across a Table. To illustrate the process of least squares, consider a randomly moving marble on a table. The x and y locations are the actual measurements over time.

$\bar{r} = y_o - y_c \Rightarrow \text{residual}$, the difference between observed and computed values of the dependent variable (the difference between the model and reality)

Let's 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

$$J = \sum_{i=1}^N \bar{r}_i^2 = \text{a minimum} \quad (10-1)$$

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 estimate values for the parameters α and β such that the sum of the squares of the residuals, \bar{r}_i^2 , is a minimum. These parameters are known as the *state* for this problem. 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 cost function to zero and solve for the unknowns α and β . The results indicate the extreme values of the function. The function [Eq. (10-1)] is a parabola (actually a bowl because there are multiple parameters) with no maximum and only one minimum; thus, the first derivative will provide a global minimum. The computed y_{c_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, you could simply differentiate the cost function and determine the answer directly. Let's take a different approach and 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:

$$\begin{aligned} \frac{\partial}{\partial \alpha} \sum_{i=1}^N \bar{r}_i^2 &= \sum_{i=1}^N \frac{\partial \bar{r}_i^2}{\partial \alpha} = \sum_{i=1}^N 2 \bar{r}_i \frac{\partial \bar{r}_i}{\partial \alpha} = 0 & \frac{\partial}{\partial \beta} \sum_{i=1}^N \bar{r}_i^2 &= \sum_{i=1}^N \frac{\partial \bar{r}_i^2}{\partial \beta} = \sum_{i=1}^N 2 \bar{r}_i \frac{\partial \bar{r}_i}{\partial \beta} = 0 \\ \sum_{i=1}^N \bar{r}_i \frac{\partial \bar{r}_i}{\partial \alpha} &= 0 & \sum_{i=1}^N \bar{r}_i \frac{\partial \bar{r}_i}{\partial \beta} &= 0 \end{aligned}$$

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 , you must separate them from everything else. In matrix notation (and dividing 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

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ x_{o_1} & x_{o_2} & \dots & x_{o_8} \end{bmatrix} \begin{bmatrix} y_{o_1} \\ y_{o_2} \\ \vdots \\ y_{o_8} \end{bmatrix} - \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_{o_1} & x_{o_2} & \dots & x_{o_8} \end{bmatrix} \begin{bmatrix} 1 & x_{o_1} \\ 1 & x_{o_2} \\ \vdots & \vdots \\ 1 & x_{o_8} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

The new matrix is the transpose of the left-hand matrix. Rearrange the matrix addition to place the state vector on one side:

$$\underbrace{\begin{bmatrix} 1 & 1 & \dots & 1 \\ x_{o_1} & x_{o_2} & \dots & x_{o_8} \end{bmatrix}}_{\mathbf{A}^T} \underbrace{\begin{bmatrix} 1 & x_{o_1} \\ 1 & x_{o_2} \\ \vdots & \vdots \\ 1 & x_{o_8} \end{bmatrix}}_{\hat{\mathbf{X}}} = \underbrace{\begin{bmatrix} 1 & 1 & \dots & 1 \\ x_{o_1} & x_{o_2} & \dots & x_{o_8} \end{bmatrix}}_{\mathbf{A}^T} \underbrace{\begin{bmatrix} y_{o_1} \\ y_{o_2} \\ \vdots \\ y_{o_8} \end{bmatrix}}_{\mathbf{b}}$$

Now define these matrices in a symbolic form. Let N be the number of observations (in this case $N = 8$).

\mathbf{A} \Rightarrow partial-derivative matrix ($N \times 2$)

$\mathbf{A}^T \Rightarrow$ \mathbf{A} transpose ($2 \times N$)

$\hat{\mathbf{X}}$ \Rightarrow solution, state vector, or state space (2×1)

\mathbf{b} \Rightarrow observation matrix ($N \times 1$)

In symbolic form then, the equation is

$$\mathbf{A}^T \mathbf{A} \hat{\mathbf{X}} = \mathbf{A}^T \mathbf{b}$$

These equations are called the **normal equations**, and they trace their history back to Gauss. Although \mathbf{A} is an $N \times 2$ matrix, $\mathbf{A}^T \mathbf{A}$ is a 2×2 matrix. Thus if we form $\mathbf{A}^T \mathbf{A}$ (the *information matrix*) directly, we don't have to handle large dimension 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 want to solve for $\hat{\mathbf{X}}$ (α and β), the parameters that define the trajectory's orientation (state). These don't define the *shape* of the trajectory because we already did that using our selection of the mathematical model 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 \mathbf{A} and \mathbf{A}^T aren't usually square matrices, the matrix product $\mathbf{A}^T \mathbf{A}$ is always square. Thus, we may invert it, provided it's *positive definite* (not singular).† Solving for $\hat{\mathbf{X}}$ gives us

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \hat{\mathbf{X}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$$

and the general solution of $\hat{\mathbf{X}}$ for the linear case is

$$\hat{\mathbf{X}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b} \quad (10-2)$$

* Accumulating the information matrix is complex for practical applications like gravity field estimation. Orthogonal transformations and single value decomposition can help these operations.

† A matrix \mathbf{a} is positive definite when $\mathbf{x}^T \mathbf{a} \mathbf{x} > 0$ for all \mathbf{x} (See Bierman [1977:33-37]). This is also the observability requirement.

When we put in data for each variable, we find the estimate of the state, $\hat{\mathbf{X}}$. The overall process is sometimes called **parameter estimation** because we are ultimately trying to determine the parameters, α and β .

▼ **Example 10-1: Using Linear Least Squares (1).**

GIVEN: Data from Fig. 10-1.

FIND: Best estimate of the state and a mathematical model for the equations of motion

From Fig. 10-1, the observational data is:

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 ($y = \alpha + \beta x$). The calculated values are then

$$y_{c_i} = \alpha + \beta x_{o_i}$$

The $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}^T\mathbf{b}$ matrices are found by summing the following relations:

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} N & N \\ \sum_{i=1}^N 1 & \sum_{i=1}^N x_{o_i} \\ N & N \\ \sum_{i=1}^N x_{o_i} & \sum_{i=1}^N x_{o_i}^2 \end{bmatrix} = \begin{bmatrix} 8 & 36 \\ 36 & 204 \end{bmatrix}, \quad \mathbf{A}^T\mathbf{b} = \begin{bmatrix} N \\ \sum_{i=1}^N y_{o_i} \\ N \\ \sum_{i=1}^N x_{o_i} y_{o_i} \end{bmatrix} = \begin{bmatrix} 24 \\ 136 \end{bmatrix}$$

Finding the inverse is sometimes the most difficult part of the problem and can cause many numerical difficulties as the model becomes more complex. For this example, we can solve simultaneous equations through $(\mathbf{A}^T\mathbf{A})^{-1}(\mathbf{A}^T\mathbf{A}) = \mathbf{I}$. An even simpler approximation uses the matrix rules in Sec. C.3 to find the inverse. The determinant is $(8)(204) - (36)(36) = 336$. With this value, find

$$(\mathbf{A}^T\mathbf{A})^{-1} = \frac{\begin{bmatrix} 204 & -36 \\ -36 & 8 \end{bmatrix}}{336} = \begin{bmatrix} 0.6071 & -0.1071 \\ -0.1071 & 0.0238 \end{bmatrix}$$

The linear least-squares solution using Eq. (10-2) is

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$$

Solving results in the best estimate of the state

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 0.6071 & -0.1071 \\ -0.1071 & 0.0238 \end{bmatrix} \begin{bmatrix} 24 \\ 136 \end{bmatrix} = \begin{bmatrix} 0.0 \\ 0.667 \end{bmatrix}$$

and the mathematical model for the equation of motion is

▲ $y_{c_i} = 0.0 + 0.667x_{o_i}$.

10.2.1 Error Analysis

If our measurements and models were perfect, we wouldn't need estimation at all! Unfortunately, our measurements and models are imperfect with errors arising from

many sources. By examining the estimation process from observation collection through orbit determination, we can mitigate the major sources of observation 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 (or not even being taken!). Other factors may corrupt data. Consider a satellite's pass over a ground station and recall the coverage limitations discussed in Fig. 4-4. 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 of greater atmospheric distortion from looking through a larger slice of the atmosphere.

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 non-random time-varying errors (drift)—as shown in Fig. 10-2. ***Biases*** are a *constant* offset from the “true” values, whereas ***noise*** is a statistical measure (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 and unpredictable variation of the observed mean value over an interval of time. It can also have an associated statistical variation (drift noise). If we know the biases, we can simply subtract them from each observation. We can estimate measurement biases in estimation, but it's better to have an initial value *a priori* to streamline the process. We can estimate the time-varying errors of drift with some Kalman filters, but we still can't predict them.

To illustrate the effect of a bias, let's continue with Example 10-1. 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 mathematical equation of motion for Example 10-1 would then be $y_{c_i} = 1 + 0.667x_{o_i}$. Clearly, a systematic error in the observations can change the state estimate. This does not indicate that estimation will detect the bias (this is only possible when the system is linear *and* the y observations aren't coupled to the x observations). In general, least-squares techniques *won't* remove instrument bias, although we can *solve for* biases as part of the state space.

Time-varying errors are discussed less often. The inaccuracy of clocks over time is one source of these errors. Tracking systems such as GPS (and the older Navy Transit system) rely on separate transmitting and receiving clocks (when less than 4 satellites are available). 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. Each of these errors affects range and Doppler measurements. Other time-varying errors result from local tropospheric effects, such as changing humidity levels. 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.

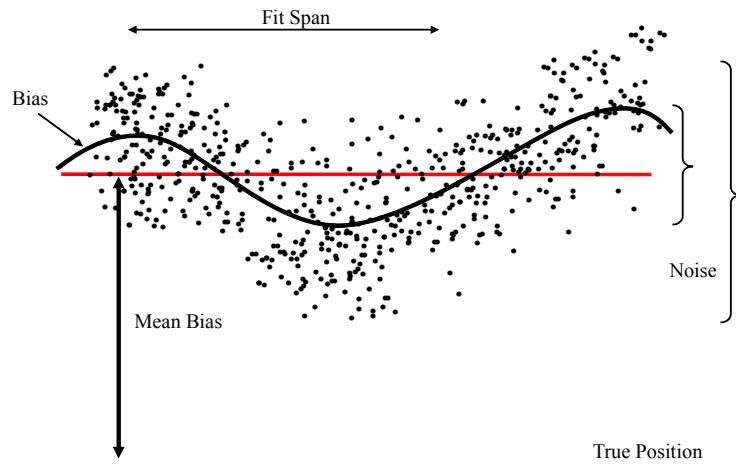


Figure 10-2. Representation of Noise, Biases, and Drift. The mean of the measured values is the average of all observations but this may not be the same as the true position. The difference is the mean bias. The noise is the variation about the mean (shown larger here). There is also noise representing the variation in the bias over time (shown smaller here). This motion is called drift and it represents the change in sensor accuracy over time. Recognize that the noise here is for illustration only and should not be viewed as the standard deviation which would be smaller.

Mathematical Modeling

The other main source of error is the mathematical modeling used during data processing. Although there are some very precise models for propagation (including complex atmospheric 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 the radial direction if they use the JGM-3 Gravity Model. *Pre-processed* solutions are typically 100 times worse! The differences arise partly because our knowledge of future input parameters (k_p , $F_{10.7}$, etc.) is limited, while post-fit solutions can use measured quantities. In addition, we can solve for time-varying parameters like BC in post-fit solutions, while we usually use fixed values in predictions. For low-Earth satellites, the inability to model the BC in predictions can cause large errors.

Even our best theories can't exactly model the atmosphere, the Earth's gravity field, or the satellite's attitude. **Process noise**, v , is a mathematical model of the errors in the system dynamics.* It quantifies our ability to model accurately the actual dynamics before and after a given epoch. 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

* Bierman (1977:113) notes that in a general sense, process noise also represents effects of unmodeled parameters, linearization, and even "effects such as leaky attitude controls or solar wind."

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.

The effect of process-noise is that the dynamics may introduce measurable error into the solution. 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. 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 this in Sec. 10.6.

Statistical Concepts

Several concepts from probability theory will help us understand the estimation process. You may need to review some of the fundamental concepts because I'll explain only briefly those relevant to astrodynamics. Consult texts such as Walpole (1974) for more information and formal derivations.

The probability of an event, $\xi < x < \xi + d\xi$, occurring for a random variable, x , 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 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 [$p(\xi)$].” We distinguish between a **sample mean**, which contains only a subset of the possible outcomes (e.g., ten trials of flipping a coin resulting in six heads and four tails), and the **population mean**, which is the totality of all possible outcomes (a coin toss will result in half head, and half tail outcomes). The averages for a finite sample can be shown, for large N , to approximate the corresponding integral expectations. For example, we define the *sample mean* as

$$\bar{x} \equiv E(x) = \int_{-\infty}^{\infty} \xi p(\xi)d\xi \approx \frac{1}{N} \sum_{i=1}^N \xi_i \quad (10-3)$$

The *Law of large numbers* permits us to approximate \bar{x} using the sample mean. This is the basis for *Monte-Carlo methods*. As the sample size grows, distributions become Gaussian.

Recalling Eq. (10-1), it makes sense to examine the expected value of x^2 , $E(x^2)$. If there is only one measurement, the expected value can't vary. The expected values for a population and a sample estimate, respectively, are

$$E(x^2) \equiv \int_{-\infty}^{\infty} \xi^2 p(\xi) d\xi \cong \frac{1}{N} \sum_{i=1}^N \xi_i^2$$

From this expression we can also determine the **root mean square** (RMS),* which is simply the square root of the mean expected squared value. Analysis of RMS values is widely used—often in place of standard deviations, which we'll introduce shortly. There are some practical issues here for orbits (distributed dispersions, 3-dimensional metrics, etc.). See Seago, et al. (2000) for more information.

$$\text{RMS} \equiv \sqrt{E(x^2)} \cong \sqrt{\frac{1}{N} \sum_{i=1}^N \xi_i^2} \quad (10-4)$$

For Example 10-1, the RMS of the residuals is equal to 0.408 248. We find this using

$$\sqrt{\frac{1}{8} \sum_{i=1}^N (y_{o_i} - y_{c_i})^2} = \sqrt{\frac{1}{8} \sum_{i=1}^N (y_{o_i} - (0 + 0.667x_{o_i}))^2}$$

The **sample variance**, σ^2 , represents the variability of the expected value of each variable about the sample mean:

$$\begin{aligned} \sigma^2 &\equiv E\{[x - E(x)]^2\} = E\{[x - \bar{x}]^2\} \\ &= \int_{-\infty}^{\infty} (\xi - \bar{x})^2 p(\xi) d\xi \cong \frac{1}{N-1} \sum_{i=1}^N (\xi_i - \bar{x})^2 \end{aligned} \quad (10-5)$$

The **sample 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. Some programs use these quantities interchangeably, but it's very important to keep the two distinct.[†]

The likelihood of an event occurring is described by a *probability density function*, $p(\xi)$. Figure 10-3 illustrates a normal (Gaussian) probability density function. The general formula is

* A **root sum square** (RSS) is also sometimes used in cases where one component dominates the overall result: for instance the in-track component of perturbations resulting from atmospheric

drag. The formula is RSS = $\sqrt{\sum_{i=1}^N \xi_i^2}$

† The literature is somewhat confusing about the formula for the RMS. Selby (1975:573) clearly states the divisor is N , whereas Neter (1982:61) says notations vary. When we use the sample estimate of the mean, we lose one degree of freedom, and use $N - 1$. Many computer programs assume the sample size (N) is very large, so that using N in place of $N - 1$ matters little.

$$p(\xi) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(\xi - \bar{x})^2}{2\sigma^2}\right] \quad (10-6)$$

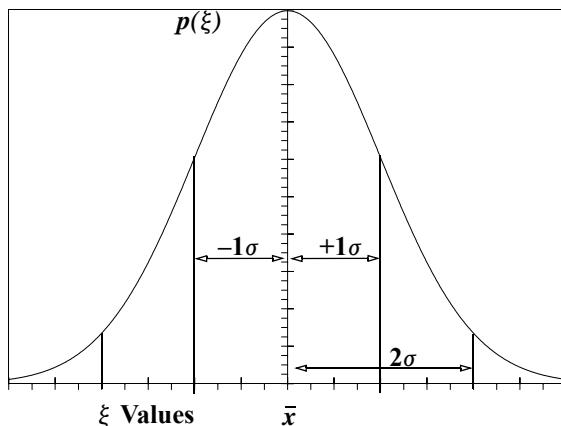


Figure 10-3. Gaussian Distribution. I've shown the mean, \bar{x} , and two vertical bands showing the standard deviation, σ . Each standard deviation consists of positive and negative components. The total area under the curve is unity. The area between any two points gives us the probability; thus, a $\pm 1\sigma$ region encompasses 0.683, or 68.3% (see also Table 10-1). Each band of standard deviation occurs at equal displacements from the mean (on the ξ axis).

The Gaussian probability density function 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 that a large number of individual estimates having independent errors will converge to a normal [Gaussian] distribution. If the errors are truly random and the model accurately captures the dynamics, more data will produce better estimates. Of course, for cases with large errors in the system model, such as using a linear model for parabolic-shaped data, the result won't improve with more data.

We've introduced the sample mean and standard deviations for a single scalar random variable. The results also generalize to vectors. Given a vector with random variables as components, its sample mean is the vector with components given by the mean of the individual random variables. The concept of sample standard deviation generalizes to a square covariance matrix, P , which we'll discuss shortly. A mass distribution is an excellent analogy to the probability distribution for a vector of two or more scalar functions. The mean is analogous to the location of the center of mass, and the covari-

ance matrix is analogous to the matrix for the moments of inertia. We saw examples of this in Chap. 8. Table 10-1 shows probabilities of a given data point being within a certain percentage standard deviation of the mean, for one- to three-dimensional Gaussian probability distributions. A “ 3σ ” or “ 4σ ” reference is common because each probability is approximately equal.

TABLE 10-1. 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), and Wiesel (1983:24)]. Regardless of dimension, the probability distribution is normalized to 1.0.*

Dimension (N)	$z = 1\sigma$	2σ	3σ	4σ
1	0.6827	0.9545	0.9973	0.9999
2	0.3935	0.8647	0.9889	0.9996
3	0.1987	0.7385	0.9707	0.9989

* We find these probabilities using the following equations for $N = 1, 2$, and 3 (Wiesel, 1983, 23).

$$\text{ERF}\left(\frac{z}{\sqrt{2}}\right) \quad 1 - \exp\left[\frac{-z^2}{2}\right] \quad \text{ERF}\left(\frac{z}{\sqrt{2}}\right) - \sqrt{\frac{2}{\pi}} z \exp\left[\frac{-z^2}{2}\right]$$

The two-dimensional case represents the volume formed by rotating the bell curve (in Fig. 10-3) about the vertical axis associated with the mean value, \bar{x} . This gives a bell shaped area with the maximum probability at the peak. The case of three dimensions results in a spherical density “cloud” in which the probability is highest at the center. Fig. 10-4 depicts several forms.

Covariance Matrix

When we consider random variables as vectors, we form the sample mean using Eq. (10-3), only now we have a matrix solution. Suppose we have a system in which the observations are \mathbf{y} , the state is \mathbf{x} , and we have arbitrary constants, m and \mathbf{b} . Using a simple linear relation,

$$\mathbf{y} = \mathbf{b} + m\mathbf{x} \quad (10-7)$$

Using Eq. (10-3), we can determine the mean of the observations,

$$\bar{\mathbf{y}} \equiv E(\mathbf{y}) = E[\mathbf{b} + m\mathbf{x}] = \mathbf{b} + mE(\mathbf{x}) = \mathbf{b} + m\bar{\mathbf{x}}$$

From Eq. (10-5), we find the sample variance as follows. It's a square matrix due to the squared term in the expectation operator (\mathbf{b} cancels when we substitute \mathbf{y} and $\bar{\mathbf{y}}$):

$$\mathbf{P}_y \equiv E[(\mathbf{y} - \bar{\mathbf{y}})(\mathbf{y} - \bar{\mathbf{y}})^T] = E[m(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T m^T]$$

Using properties from linear-error theory gives us

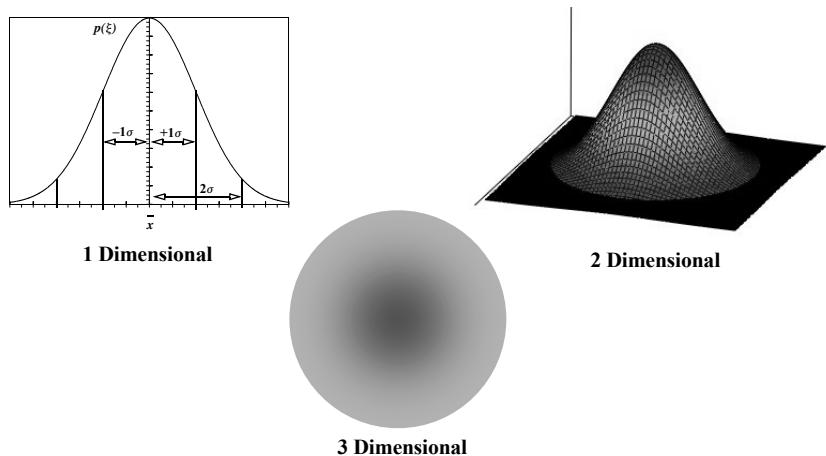


Figure 10-4. Multi-dimensional Probability Distribution. I've shown multiple probability distributions. The 3-dimensional case depicts a spherical cloud that contains the maximum probability.

$$\mathbf{P}_y = mE[(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T]m^T \quad \mathbf{P}_y = m\mathbf{P}_x m^T \quad (10-8)$$

This result is usually known as the *similarity transformation*; it allows us mathematically to describe the covariance matrix, \mathbf{P}_y .

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 uncertainty of the answers. The **covariance matrix**, $\mathbf{P} = (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1}$, contains the estimates for the closeness of the fit with the chosen dynamics.* 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”.† For the simple least-squares problem, we calculate the covariance matrix as the matrix $(\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1}$. An example state having three estimated parameters— α, β, γ —gives us a covariance matrix of the form

* If we do not have the weighting matrix, \mathbf{W} , and have only $(\mathbf{A}^T \mathbf{A})^{-1}$, additional processing is required. A \mathbf{W} value of unity (i.e., no knowledge of the sensor weights) requires us to examine the scale of the variances. You can estimate the sample observation variance from the scatter of the residuals. Then, you multiply this value with the standard deviation from the covariance matrix. Mathematically, derive this as follows. (use $E\{\mathbf{b}\mathbf{b}^T\} = \sigma_m^2$ for the variance of the measurements)

$$\mathbf{P} = E\{(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b} \mathbf{b}^T \mathbf{A} ((\mathbf{A}^T \mathbf{A})^{-1})^T\} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \sigma_m^2 \mathbf{A} ((\mathbf{A}^T \mathbf{A})^{-1})^T = \sigma_m^2 (\mathbf{A}^T \mathbf{A})^{-1}$$

† Most of our examples include weighting information, thus avoiding this additional calculation.

† 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.

$$\mathbf{P} \equiv (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} = E([\mathbf{X} - \bar{\mathbf{X}}]^2) = \begin{bmatrix} \sigma_\alpha^2 & \mu_{\alpha\beta}\sigma_\alpha\sigma_\beta & \mu_{\alpha\gamma}\sigma_\alpha\sigma_\gamma \\ \mu_{\beta\alpha}\sigma_\alpha\sigma_\beta & \sigma_\beta^2 & \mu_{\beta\gamma}\sigma_\beta\sigma_\gamma \\ \mu_{\gamma\alpha}\sigma_\alpha\sigma_\gamma & \mu_{\gamma\beta}\sigma_\beta\sigma_\gamma & \sigma_\gamma^2 \end{bmatrix}$$

where σ_α is the standard deviation defined in Eq. (10-5), 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 sample standard deviations for each element of the state space. You should include these values when discussing the results of an estimation calculation because they establish a level of uncertainty in each element. AF Space Command didn't use this data for many decades because it was deemed inaccurate when obtained from sparse data with limited dynamics. The covariance may indeed be inaccurate due to poorly known parameters (noise, biases, and drift) and inadequate 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 standard deviations represent 1σ values, but because we are using a three-dimensional state vector, the 1σ value represents only a 20% probability (See Table 10-1). Most analyses for satellites rely on a 3σ value, so the uncertainty is much larger than the original 1σ value.

The off-diagonal elements are called covariance terms; they contain the *correlation coefficients* ($\mu_{xy}^2 = (1/\sigma_x\sigma_y) E\{(x-\bar{x})(y-\bar{y})\}$). The **correlation coefficients** represent the degree of correlation among the elements, with zero being no correlation. Positive signs indicate a direct correlation, while negative signs imply an inverse relationship. We want the correlations to be near zero or unity for ease of interpretation. Consider solving for a satellite's classical orbital elements and the drag coefficient, c_D . We've explored c_D in Sec. 8.6.2; it scales 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.

The standard deviations give the statistical accuracy of the state in the satellite coordinate frame (usually geocentric equatorial). However, the along-track direction usually contains the largest uncertainty because non-conservative forces (drag, solar radiation pressure) tend to affect mean motion, but not inclination. Section 10.8.5 shows how to transform the covariance to alternate coordinate systems and orbital state formats.

Collision probability studies require the standard deviation for the three-dimensional error ellipsoid around a satellite. To find the exact orientation of the error ellipsoid, we can calculate the eigenvectors ($\hat{\mathbf{x}}_{eig}$). The eigenvalues (λ_{eig}) give us the dimensions of the error ellipsoid. We often examine only the position vector uncertainty (the upper 3×3 matrix that we'll call \mathbf{B}) because most applications use only positional information.

There are several techniques to determine the eigenvalues and eigenvectors. One approach finds the eigenvalues using a determinant (a_{ii} are the terms of \mathbf{B}).

$$\det|\mathbf{B} - \lambda_{eig}\mathbf{I}| = 0$$

After expanding the above determinant (because the matrix is symmetric [$a_{ij} = a_{ji}$] several terms can be simplified),

$$\begin{aligned} & -\lambda_{eig}^3 + \lambda_{eig}^2 (a_{11} + a_{22} + a_{33}) \\ & + \lambda_{eig}(-a_{11}a_{33} - a_{11}a_{22} - a_{22}a_{33} + a_{32}a_{23} + a_{12}a_{21} + a_{13}a_{31}) \\ & + a_{11}a_{22}a_{33} - a_{11}a_{32}a_{23} - a_{12}a_{21}a_{33} + a_{12}a_{31}a_{23} + a_{13}a_{21}a_{32} - a_{13}a_{31}a_{22} = 0 \end{aligned}$$

solve the resulting cubic equation with analytical techniques (Appendix C). The answers let us write simultaneous equations for the eigenvectors. We find the eigenvectors by substituting each λ_{eig} into

$$|\mathbf{B} - \lambda_{eig}\mathbf{I}| \hat{\mathbf{x}}_{eig} = 0$$

See Press et al. (1992:449-489) for several solution methods. Iteration is common. We can check the eigenvalues using the newly found eigenvectors.

$$\hat{\mathbf{x}}_{eig}^T \mathbf{B} \hat{\mathbf{x}}_{eig} = \lambda_{eig}$$

Consider Parameters

For least squares, we can include the effect of the uncertainty of unestimated parameters on the covariance through **consider parameters**. These parameters appear as extra variables in the state space and covariance, but they aren't estimated through processing. Consider parameters provide a way to incorporate *a priori* estimates of certain parameters and 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. See Bierman (1977:162–182) and Long et al. (1989:Chap. 8) for more information.

Editing Data

We can reject certain data by testing those with residuals outside some number of standard deviations from the mean. Of course, we must already know the standard deviation before rejecting data. Eliminating data can be very tricky, especially if the data is sparse. We don't reject data on each iteration because this process will often edit out all the data (Wiesel, 1983:69)! As an example of spurious data points, consider Example 10-2.

▼ Example 10-2. Using Linear Least Squares (2).

Suppose observation 7 differs from Example 10-1:

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 \mathbf{A} matrix is the same as before, as is \mathbf{A}^T and $(\mathbf{A}^T \mathbf{A})^{-1}$. But the \mathbf{b} matrix has changed, $\mathbf{b}^T = [1 \ 1 \ 2 \ 3 \ 3 \ 4 \ 7 \ 6]$.

$$A^T b = \begin{bmatrix} N \\ \sum_{i=1}^N y_{o_i} \\ \sum_{i=1}^N x_{o_i} y_{o_i} \end{bmatrix} = \begin{bmatrix} 27 \\ 157 \end{bmatrix}$$

The $A^T A$ 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 curve*. The residuals are the estimated or assumed errors off the line. Next, calculate the residuals:

$$\bar{r}_i = y_{o_i} - y_{c_i} = y_{o_i} - (\alpha + \beta x_{o_i}) = y_{o_i} - (-0.429 + 0.845x_{o_i})$$

The residuals for each observation are:

$$y_{c_i} \quad 0.417 \quad 1.262 \quad 2.107 \quad 2.952 \quad 3.797 \quad 4.643 \quad 5.488 \quad 6.333$$

$$\bar{r}_i \quad 0.583 \quad -0.262 \quad -0.107 \quad 0.048 \quad -0.797 \quad -0.643 \quad 1.512 \quad -0.333$$

Using the residuals, the RMS is

$$\text{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^N \bar{r}_i^2} = \sqrt{\frac{3.869}{8}} = 0.6954$$

After finding the residuals based on the initial estimate, we might reject observations whose residuals are larger than a specific value as being unreasonable, or invalid. We can check for points with a residual outside, say, a 2σ range. The 2σ ($2(0.695)$) limit is for illustration only. It's sometimes 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 new answer: $y_{c_i} = -0.131 + 0.721x_{o_i}$.

Let's examine what detecting and eliminating a bad data point really means. If we need to determine the confidence in each answer, we examine the standard deviation from the covariance matrix.

$$y_{c_i} = -0.429 + 0.845x_{o_i} \quad \text{RMS} = 0.6954$$

$$\sigma_\alpha = \sqrt{0.6071} = 0.7792 \quad \sigma_\beta = \sqrt{0.0238} = 0.1543$$

Because there is no weighting matrix, we must multiply the standard deviations from the covariance matrix, with the sample observation variances. We estimate these variances as the square root of the sum of the residuals squared, divided by one less than the number of observations. The sample observation variance is $\sqrt{3.869/7} = 0.7435$. Multiplying with each standard deviation from the covariance matrix

$$\alpha = -0.429 \pm 0.5793 \text{ and } \beta = 0.845 \pm 0.1147$$

After eliminating point 7, the values become

$$y_{c_i} = -0.131 + 0.721x_{o_i} \quad \text{RMS} = 0.3210$$

$$\alpha = -0.131 \pm 0.2763 \text{ and } \beta = 0.721 \pm 0.0587$$

Notice that the RMS decreased, and our confidence in the answer got better because there was less “bad” data. Usually, we want greater confidence in the answer. In the example, this is the best we can do with only sparse data.

A word of caution about rejecting data. Although we can use the RMS to edit the data, too tight a tolerance may cause us to lose observations that are crucial to the results, hence the larger ranges for standard deviations used to edit data. ***Chauvenet’s criterion*** formalizes this process and suggests that we eliminate data when the probability of its occurrence is less than 50%. Essentially, we find the probability that the given observations could occur within the sample set, and compare the answer to the rejection probability (50%).

Finally, we must distinguish between a curve fit and an estimate using physical models. With an arbitrary curve, we may fit the data well but obtain coefficients to a model with no value for prediction. A difficult part of estimation is getting the correct mathematical 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.”

10.2.2 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. We expect the residuals for Set *B* will be larger than those from Set *A*, so we want to weight the fit toward Set *A*, because its data is better. We *weight the residuals* using the inverses of the standard deviations. In reality, we often don’t 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.* We use the inverse so that a small standard deviation (more accurate observations) will weight the sensor more than a large one.

$$w_A = \frac{1}{\sigma_A} \text{ and } w_B = \frac{1}{\sigma_B} \text{ so that } w_A > w_B$$

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:

* Gauss placed this heuristic choice (observations having similar characteristics) on rigorous theoretical grounds, and under reasonable assumptions, we can prove this choice leads to the most likely trajectory estimate. Later, we’ll use a measurement noise matrix, \mathbf{R} , that is equal to the square of the noise values.

$$y_{c_i} = \alpha + \beta x_{o_i} \quad y_c = f(x) = \alpha + \beta x$$

$$w_i \bar{r}_i = w_i(y_{o_i} - y_{c_i}) = w_i(y_{o_i} - (\alpha + \beta x_{o_i}))$$

Now, write the weighted residuals using matrix notation, remembering that each observation will have weights associated with the state-vector parameters. For N observations,

$$w_i = \begin{bmatrix} \frac{1}{\sigma_A} & 0 \\ 0 & \frac{1}{\sigma_B} \end{bmatrix} = \begin{bmatrix} w_A & 0 \\ 0 & w_B \end{bmatrix} \quad i = 1 \dots N \quad \bar{r} = \begin{bmatrix} y_{o_1} - y_{c_1} \\ y_{o_2} - y_{c_2} \\ \vdots \\ y_{o_N} - y_{c_N} \end{bmatrix}$$

$$\bar{r} = \mathbf{b} - \mathbf{AX}$$

$$\mathbf{A} = \begin{bmatrix} 1 & x_{o_1} \\ 1 & x_{o_2} \\ \vdots & \vdots \\ 1 & x_{o_N} \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \quad \mathbf{W} = w_i^T w_i = \begin{bmatrix} w_1^2 & 0 & \dots & 0 \\ 0 & w_2^2 & & \\ \vdots & & \ddots & \\ 0 & \dots & 0 & w_N^2 \end{bmatrix}$$

We denote the *weighting matrix*, \mathbf{W} , to simplify the notation. Note that \mathbf{W} contains the weights (as appropriate) for each measurement. In practice, we accumulate measurements so the size of the \mathbf{W} matrix is smaller. Applying the least-squares criterion [Eq. (10-1)] to the weighted residuals produces the cost function:

$$J = \sum_{i=1}^N w_i^2 \bar{r}_i^2 = \bar{r}^T \mathbf{W} \bar{r} = (\mathbf{b} - \mathbf{AX})^T \mathbf{W} (\mathbf{b} - \mathbf{AX})$$

To find the minimum, we first expand the cost function:

$$J = (\mathbf{b} - \mathbf{AX})^T \mathbf{W} (\mathbf{b} - \mathbf{AX}) = \mathbf{b}^T \mathbf{W} \mathbf{b} - 2 \mathbf{b}^T \mathbf{W} \mathbf{A} \mathbf{X} + \mathbf{X}^T \mathbf{A}^T \mathbf{W} \mathbf{A} \mathbf{X}$$

Now, 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.

$$\frac{\partial J}{\partial \hat{\mathbf{X}}} = -2 \mathbf{b}^T \mathbf{W} \mathbf{A} + 2 \hat{\mathbf{X}}^T \mathbf{A}^T \mathbf{W} \mathbf{A} = 0$$

Let's simplify by equating and taking the transpose of both sides and distributing the transpose operator. Noting that $\mathbf{W}^T = \mathbf{W}$, we get

$$\begin{aligned}\hat{\mathbf{X}}^T \mathbf{A}^T \mathbf{W} \mathbf{A} &= (\mathbf{b}^T \mathbf{W} \mathbf{A})^T \\ \mathbf{A}^T \mathbf{W} \mathbf{A} \hat{\mathbf{X}} &= \mathbf{A}^T \mathbf{W} \mathbf{b}\end{aligned}$$

Assuming that $\mathbf{A}^T \mathbf{W} \mathbf{A}$ is invertible (the observability criteria), the solution state is

$$\hat{\mathbf{X}} = (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W} \mathbf{b} = \mathbf{P} \mathbf{A}^T \mathbf{W} \mathbf{b} \quad (10-9)$$

If we let \mathbf{W} be a matrix with standard deviations of 0.1 and 0.02 respectively (the weights are then 10 and 50), and use the covariance matrix and state in Example 10-1, we get

$$(\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} = \begin{bmatrix} 0.0231 & -0.0039 \\ -0.0039 & 0.0008 \end{bmatrix} \quad y_c = -0.3243 + 0.7568x_o$$

The answer represents a line that minimized the residuals. The covariance matrix tells us the uncertainty in the answer.

$$\alpha = -0.3243 \pm 0.1520 \quad \beta = 0.7568 \pm 0.0285$$

These are both one-sigma confidence intervals. Notice how much uncertainty there is in the answer! Without the standard deviations, the user would never know how much uncertainty exists in the solution parameters. The off-diagonal terms show the two coefficients have an inverse correlation, (-0.0039), but they're not completely independent.

10.3 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 the least-squares method to nonlinear problems by linearizing the problem, obtaining an approximate solution, and iterating to refine the answer. As an example, assume an ant is rolling a marble across a table in a sine-wave pattern. Figure 10-5 shows the expected path of the marble, along with many observations. Our job is to determine the amplitude, α , and the phase shift, β , of the sine wave.

Again, assume our sensors measure y as a function of x but y gives us noisy data. 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 a nonlinear function of the state, α and β . Let's proceed through the least-squares development and see if this nonlinearity makes any difference. Apply the least-squares criterion of Eq. (10-1) to get

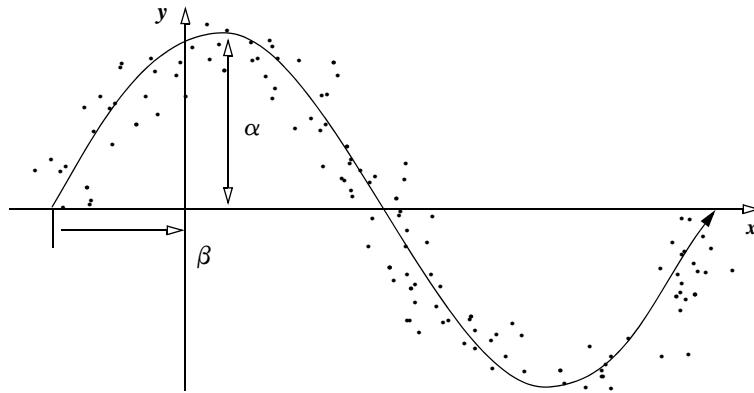


Figure 10-5. Sinusoidal Path of a Marble. If we choose a sinusoidal path $y = f(x) = \alpha \sin(x + \beta)$ as the mathematical model, the estimation problem becomes nonlinear because we can't separate α and β .

$$\begin{aligned} \sum_{i=1}^N \bar{r}_i \frac{\partial \bar{r}_i}{\partial \alpha} &= 0 & \sum_{i=1}^N \bar{r}_i \frac{\partial \bar{r}_i}{\partial \beta} &= 0 \\ \frac{\partial \bar{r}_i}{\partial \alpha} &= -\sin(x_{o_i} + \beta) & \frac{\partial \bar{r}_i}{\partial \beta} &= -\alpha \cos(x_{o_i} + \beta) \end{aligned}$$

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 let the A^T be the left-hand $2 \times N$ matrix (similar to page 738). But now the partial-derivative matrix results in expressions:

$$A^T \begin{bmatrix} y_{o_1} - \alpha \sin(x_{o_1} + \beta) \\ y_{o_2} - \alpha \sin(x_{o_2} + \beta) \\ \vdots \\ y_{o_N} - \alpha \sin(x_{o_N} + \beta) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Now split the residual matrix to find

$$\mathbf{A}^T \begin{bmatrix} y_{o_1} \\ y_{o_2} \\ \vdots \\ y_{o_N} \end{bmatrix} - \mathbf{A}^T \begin{bmatrix} \alpha \sin(x_{o_1} + \beta) \\ \alpha \sin(x_{o_2} + \beta) \\ \vdots \\ \alpha \sin(x_{o_N} + \beta) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (10-10)$$

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 relationship, $\alpha \sin(x + \beta)$. 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 a method of handling a nonlinear relationship for the measurement state. Fortunately, a Taylor series allows you to approximate the nonlinear equations with linear equations, provided you can neglect the higher-order terms in the Taylor series. So, if you compute $y = f(\alpha, \beta)$ 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

$$\begin{aligned} y_c &= y|_{\alpha_n \beta_n} + (\alpha - \alpha_n) \frac{\partial y}{\partial \alpha} \Big|_{\alpha_n \beta_n} + (\beta - \beta_n) \frac{\partial y}{\partial \beta} \Big|_{\alpha_n \beta_n} + \frac{(\alpha - \alpha_n)^2}{2!} \frac{\partial^2 y}{\partial \alpha^2} \Big|_{\alpha_n \beta_n} \\ &\quad + \frac{(\beta - \beta_n)^2}{2!} \frac{\partial^2 y}{\partial \beta^2} \Big|_{\alpha_n \beta_n} + \dots \end{aligned}$$

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, you can calculate the values of the observations and the partial derivatives using the initial estimates of the state (α_n, β_n) from above. Thus, Eq. (10-10) becomes

$$\mathbf{A}^T \begin{bmatrix} y_{o_1} \\ y_{o_2} \\ \vdots \\ y_{o_N} \end{bmatrix} - \mathbf{A}^T \begin{bmatrix} y_{n_1} + \Delta\alpha \frac{\partial y_{n_1}}{\partial \alpha} + \Delta\beta \frac{\partial y_{n_1}}{\partial \beta} \\ y_{n_2} + \Delta\alpha \frac{\partial y_{n_2}}{\partial \alpha} + \Delta\beta \frac{\partial y_{n_2}}{\partial \beta} \\ \vdots \\ y_{n_N} + \Delta\alpha \frac{\partial y_{n_N}}{\partial \alpha} + \Delta\beta \frac{\partial y_{n_N}}{\partial \beta} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

where

$$\frac{\partial y_{n_i}}{\partial \alpha} = \sin(x_{o_i} + \beta_n) \quad \frac{\partial y_{n_i}}{\partial \beta} = \alpha_n \cos(x_{o_i} + \beta_n)$$

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:

$$\mathbf{A}^T \begin{bmatrix} y_{o_1} - y_{n_1} \\ y_{o_2} - y_{n_2} \\ \vdots \\ y_{o_N} - y_{n_N} \end{bmatrix} - \mathbf{A}^T \begin{bmatrix} \frac{\partial y_{n_1}}{\partial \alpha} & \frac{\partial y_{n_1}}{\partial \beta} \\ \frac{\partial y_{n_2}}{\partial \alpha} & \frac{\partial y_{n_2}}{\partial \beta} \\ \vdots & \vdots \\ \frac{\partial y_{n_N}}{\partial \alpha} & \frac{\partial y_{n_N}}{\partial \beta} \end{bmatrix} \begin{bmatrix} \Delta\alpha \\ \Delta\beta \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Consider some interesting features of this equation. The matrix containing $y_{o_i} - y_{n_i}$ looks like the \mathbf{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 $\tilde{\mathbf{b}}$. The matrix containing the observation partials is the transpose of \mathbf{A}^T , so it's the \mathbf{A} matrix. We evaluate the \mathbf{A} matrix using the nominal state vector. The matrix containing $\Delta\alpha$ and $\Delta\beta$ corresponds to the $\hat{\mathbf{X}}$ matrix, except it's now the *corrections* to α and β . We'll call it $\delta\hat{\mathbf{x}}$. Substituting the newly defined matrices gives us

$$\mathbf{A}^T \tilde{\mathbf{b}} - \mathbf{A}^T \mathbf{A} \delta\hat{\mathbf{x}} = 0$$

and assuming observability, the estimated corrections to the state are

$$\delta\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \tilde{\mathbf{b}} \quad (10-11)$$

We can keep the dimensions of the matrices small by applying the same summation principles as in the linear case. Eq. (10-11) is identical to the linear least-squares equation, except that

1. \mathbf{A} , $\tilde{\mathbf{b}}$, and $\delta\hat{\mathbf{x}}$ are defined in the derivation
2. \mathbf{A} , \mathbf{A}^T , and $\tilde{\mathbf{b}}$ use the previous estimates of α_n and β_n
3. It's an approximate solution because we've used a truncated Taylor series.
4. We must 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 global 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.

Points (2) and (3) imply the need 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. Compute y_{n_i} corresponding to each x_{o_i}
2. Compute each residual $\bar{r}_i = y_{o_i} - y_{n_i}$
3. Compute each partial derivative, $\partial y_{n_i} / \partial \alpha$ and $\partial y_{n_i} / \partial \beta$, using α_n , β_n
4. Form \mathbf{A} , \mathbf{A}^T and $\tilde{\mathbf{b}}$
5. Solve for $\Delta\alpha$ and $\Delta\beta$ using Eq. (10-11)
6. Find $\alpha_{n_{new}} = \alpha_{n_{old}} + \Delta\alpha$ and $\beta_{n_{new}} = \beta_{n_{old}} + \Delta\beta$
7. If you meet a stopping criterion, quit. Otherwise, return to step (1).

Because we're correcting the state on each iteration, this process is termed ***differential correction***.

The last step is knowing when to quit. 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. In practice, the changes to the state rarely become identically zero, so the RMS is a better choice. We quit when the RMS of the residuals stops changing, within a certain tolerance:

$$\text{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^N \bar{r}_i^2} = \sqrt{\frac{\tilde{\mathbf{b}}^T \tilde{\mathbf{b}}}{N}} \quad (10-12)$$

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

$$\left| \frac{RMS_{old} - RMS_{new}}{RMS_{old}} \right| \leq \epsilon \quad (10-13)$$

where the convergence tolerance, ϵ , is a function of the problem's nonlinearity and the measurement error. Program the test condition in software 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 weighting the observations. 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

$$\hat{\delta x} = (A^T W A)^{-1} A^T W \tilde{b} = P A^T W \tilde{b} \quad (10-14)$$

where again, for our state-space dimension of 2, with N observations,

- $A \Rightarrow$ partial-derivative matrix ($N \times 2$)
- $A^T \Rightarrow$ A transpose ($2 \times N$)
- $\hat{\delta x} \Rightarrow$ corrections to the state vector (2×1)
- $b \Rightarrow$ residual matrix ($N \times 1$)
- $W \Rightarrow$ weighting matrix ($N \times N$)
- $P = (A^T W A)^{-1} \Rightarrow$ covariance matrix (2×2)

Error analysis (Sec. 10.2.1) is also the same as for the linear problem. The covariance matrix $(A^T W A)^{-1}$ contains the variances and covariances of the solution parameters (state space) and indicates observability from the available data types, as well as the effect of noisy data on the estimated parameters. It is not dependent on actual data. These statistical values will be accurate *only* if our choice of models matches reality and the calibration values are correct. Our model 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

* You can show that the least squares method is a minimum variance estimator, but the residuals are still a better exit criteria.

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 10-3 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 10-3. Using Nonlinear Least Squares.**

GIVEN:

x_{o_i}	1	2	3	4
y_{o_i}	2.5	8.0	19.0	50.0

FIND: Best estimate of the trajectory

Assume the data above represents a parabolic function [a quick plot suggests this model],

$$y_{c_i} = \alpha(x_{o_i})^\beta$$

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:

$$\frac{\partial y_{n_i}}{\partial \alpha_n} = (x_{o_i})^{\beta_n} \quad \frac{\partial y_{n_i}}{\partial \beta_n} = \alpha_n(x_{o_i})^{\beta_n} (\ln x_{o_i})$$

The solution rests in solving the equation

$$\hat{\delta x} = (A^T W A)^{-1} A^T W \tilde{b} \text{ or } (A^T A)^{-1} A^T \tilde{b}$$

We assume the weights are unity values. For equally weighted measurements, this solution simplifies to the linear form of Eq. (10-2), 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 $(1)^2 + (10.2915)^2 + (40.2476)^2 + (105.9147)^2 = 12,944.71$.

$$A^T A = \sum_{i=1}^4 \begin{bmatrix} \left(\frac{\partial y_{n_i}}{\partial \alpha_n}\right)^2 & \frac{\partial y_{n_i}}{\partial \alpha_n} \frac{\partial y_{n_i}}{\partial \beta_n} \\ \frac{\partial y_{n_i}}{\partial \alpha_n} \frac{\partial y_{n_i}}{\partial \beta_n} & \left(\frac{\partial y_{n_i}}{\partial \beta_n}\right)^2 \end{bmatrix} = \begin{bmatrix} 12,944.71 & 8216.22 \\ 8216.22 & 5251.58 \end{bmatrix}$$

$$(A^T A)^{-1} = \begin{bmatrix} 0.011\ 08 & -0.017\ 34 \\ -0.017\ 34 & 0.027\ 31 \end{bmatrix}$$

Now, let's find the $A^T \tilde{b}$ 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.

$$A^T \tilde{b} = \sum_{i=1}^4 \begin{bmatrix} \frac{\partial y_{n_i}}{\partial \alpha_n} (y_{o_i} - y_{n_i}) \\ \frac{\partial y_{n_i}}{\partial \beta_n} (y_{o_i} - y_{n_i}) \end{bmatrix} = \begin{bmatrix} 34.3598 \\ 10.5796 \end{bmatrix}$$

Putting everything together gives us

$$\hat{\delta x} = \begin{bmatrix} \Delta\alpha \\ \Delta\beta \end{bmatrix} = \begin{bmatrix} 0.011\ 08 & -0.017\ 34 \\ -0.017\ 34 & 0.027\ 31 \end{bmatrix}^{-1} \begin{bmatrix} 33.9071 \\ 10.2919 \end{bmatrix} = \begin{bmatrix} 0.1973 \\ -0.3067 \end{bmatrix}$$

Therefore, the updated values for the state are

$$\begin{aligned} \alpha_{new} &= 0.4721 + 0.1973 = 0.6692 \\ \beta_{new} &= 3.3634 - 0.3067 = 3.0567 \end{aligned}$$

Using the updated values, the b matrix becomes $y_{o_i} - y_{n_i} = y_{o_i} - \alpha x_{o_i}^\beta$ and

$$RMS_{new} = \sqrt{\frac{(1.8306)^2 + (2.4302)^2 + (-0.2350)^2 + (3.6568)^2}{4}} = 2.3813$$

After the second iteration,

$$\hat{\delta x} = \begin{bmatrix} \Delta\alpha \\ \Delta\beta \end{bmatrix} = \begin{bmatrix} 0.0633 \\ -0.0180 \end{bmatrix}, \hat{X} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 0.7327 \\ 3.0387 \end{bmatrix}$$

The new $RMS_{new} = 1.5820$.

Checking the percentage of change of the RMS, we find

$$\Delta_{RMS} = \left| \frac{2.3813 - 1.5820}{2.3813} \right| = 0.3357$$

We haven't converged, so we need another iteration. After the third iteration,

$$\hat{\delta x} = \begin{bmatrix} \Delta\alpha \\ \Delta\beta \end{bmatrix} = \begin{bmatrix} 0.0021 \\ -0.0008 \end{bmatrix}, \hat{X} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 0.7348 \\ 3.0379 \end{bmatrix}$$

$$RMS_{new} = 1.5818 \text{ and the } \Delta_{RMS} = \frac{1.5820 - 1.5818}{1.5820} = 0.0004$$

Both the RMS_{new} has gone down and the percentage change is small, so we have converged on the solution. The best estimate of the trajectory is

$$\blacktriangle \quad y_i = 0.7349(x_i)^{3.0378}$$

10.4 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 radar, optical, or other measurements of the motion. A minimum of six elements completely define the orbit—

either position and velocity vectors, or a set of orbital elements. We often augment them with solve-for parameters. The **solve-for** parameters allow us to estimate dynamical and measurement model parameters like the drag coefficient and measurement biases. In orbit determination, **state space** refers to the set of six orbital elements (or position and velocity vectors) and all the solve-for parameters. When applying differential correction to orbit determination, however, we must handle several situations:

1. We can choose from several element sets: $\{a, e, i, \Omega, \omega, \nu\}$, $\{k_e, h_e, n, \lambda_{mean}, \rho_e, q_e\}$, $\{r_I r_J r_K, v_I v_J v_K\}$, and so on (Refer to Sec. 2.4.). Each state space completely describes an orbit and is therefore valid. Certain sets of orbital elements have conditions for which at least one element is undefined (they have **singularities**). There are nonsingular element sets, including position and velocity vectors. The equinoctial elements also have nonsingular partial derivatives and vary slower than cartesian coordinates which make them a very attractive choice when designing a complete orbit determination system. 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 each observation time (ρ, β, el , etc.). Remember β is the azimuth.
3. The measurements are nonlinear, complex functions of the state. Recall Sec. 4.4.3 (**RAZEL**).

Assume a sensor can measure range, azimuth, and elevation at each time, t_i , so that

$$y_{o_i} = \begin{bmatrix} \rho_o \\ \beta_o \\ el_o \end{bmatrix} \text{ at times } = t_i$$

To calculate residuals, we need to compute some predicted measurements from the position and velocity vectors:

$$y_{c_i} = \begin{bmatrix} \rho_c \\ \beta_c \\ el_c \end{bmatrix} \text{ at time } = t_i$$

Because 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 a nominal trajectory. For this problem, if we ignore higher order terms, the vector expression at $t = t_i$ is

$$y_{c_i} = y_{n_i} + \Delta r_I \frac{\partial y_{n_i}}{\partial r_I} + \Delta r_J \frac{\partial y_{n_i}}{\partial r_J} + \Delta r_K \frac{\partial y_{n_i}}{\partial r_K} + \Delta v_I \frac{\partial y_{n_i}}{\partial v_I} + \Delta v_J \frac{\partial y_{n_i}}{\partial v_J} + \Delta v_K \frac{\partial y_{n_i}}{\partial v_K}$$

The nominal trajectory is $y_{n_i} = f(r_I r_J r_K, v_I v_J v_K, t_i)$, a function of the nominal vector at each observation time. In addition, each position and velocity vector component is at time t_i .^{*} Define the residuals, \bar{r}_i . I'll always use the overbar to indicate residuals.

$$\bar{r}_i = y_{o_i} - y_{c_i} = y_{o_i} - \left(y_{n_i} + \Delta r_I \frac{\partial y_{n_i}}{\partial r_I} + \dots + \Delta v_K \frac{\partial y_{n_i}}{\partial v_K} \right)$$

Assume each measurement is weighted using its appropriate standard deviation. If you have data from AFSPC, you can use Table 4-4 for example noise (standard deviation) estimates.

$$w_\rho = \frac{1}{\sigma_\rho} \quad w_\beta = \frac{1}{\sigma_\beta} \quad w_{el} = \frac{1}{\sigma_{el}}$$

You can subtract out the biases (again from Table 4-4) 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:

$$w_i = \begin{bmatrix} w_\rho & 0 & 0 \\ 0 & w_\beta & 0 \\ 0 & 0 & w_{el} \end{bmatrix}, i = 1 .. N$$

Set the first derivatives equal to zero. The two vector equations produce six scalar equations:

$$\sum_{i=1}^N w_i^T w_i \bar{r}_i \frac{\partial \bar{r}_i}{\partial \vec{r}} = 0 \quad \sum_{i=1}^N w_i^T w_i \bar{r}_i \frac{\partial \bar{r}_i}{\partial \vec{v}} = 0$$

Because $\frac{\partial \bar{r}_i}{\partial r_I} = -\frac{\partial y_{n_i}}{\partial r_I}$, etc., you can put these equations in matrix form and, depending on the number of observations, N :

^{*} The requirement for nominal vectors often means that we must determine the initial orbit (Chap. 7, *SITE-TRACK*, *GIBBS*, *LAMBERT*, etc.) to form each vector from the observations. The vectors are then propagated to a common epoch, where the nominal vector is formed.

$$-1 \begin{bmatrix} \frac{\partial y_{n_1}}{\partial r_I} & \frac{\partial y_{n_2}}{\partial r_I} & \cdots & \frac{\partial y_{n_N}}{\partial r_I} \\ \frac{\partial y_{n_1}}{\partial r_J} & & \frac{\partial y_{n_N}}{\partial r_J} & \\ \vdots & & \vdots & \\ \frac{\partial y_{n_1}}{\partial v_K} & \cdots & \cdots & \frac{\partial y_{n_N}}{\partial v_K} \end{bmatrix} \begin{bmatrix} w_1^2 \bar{r}_1 \\ w_2^2 \bar{r}_2 \\ \vdots \\ w_N^2 \bar{r}_N \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

If you define the *partial-derivative matrix* as A^T and the weighting matrix, \mathbf{W} , as the product of $w_i T_{\mathbf{w}_i}$, the matrix algebra is the same as for the nonlinear least-squares case. Thus, you have the same solution as for the nonlinear case [Eq. (10-14)], with the new definitions. I've shown a single observation set for the \mathbf{b} matrix.

$$\delta \bar{x} = \begin{bmatrix} \Delta r_I \\ \Delta r_J \\ \Delta r_K \\ \Delta v_I \\ \Delta v_J \\ \Delta v_K \end{bmatrix} \quad \tilde{\mathbf{b}} \equiv \begin{bmatrix} \bar{r}_1 \\ \bar{r}_2 \\ \vdots \\ \bar{r}_N \end{bmatrix} \quad \bar{r}_i = \begin{bmatrix} \rho_{o_i} - \rho_{n_i} \\ \beta_{o_i} - \beta_{n_i} \\ el_{o_i} - el_{n_i} \end{bmatrix}$$

10.4.1 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, \mathbf{A} , looks like this (I've shown the partials for a single observation):

$$\mathbf{A} = \frac{\partial \text{observations}}{\partial \mathbf{X}_o} \quad \frac{\partial \text{obs}_i}{\partial \mathbf{X}_o} = \begin{bmatrix} \frac{\partial \rho_i}{\partial r_{I_o}} & \frac{\partial \rho_i}{\partial r_{J_o}} & \frac{\partial \rho_i}{\partial r_{K_o}} & \frac{\partial \rho_i}{\partial v_{I_o}} & \frac{\partial \rho_i}{\partial v_{J_o}} & \frac{\partial \rho_i}{\partial v_{K_o}} \\ \frac{\partial \beta_i}{\partial r_{I_o}} & \frac{\partial \beta_i}{\partial r_{J_o}} & \frac{\partial \beta_i}{\partial r_{K_o}} & \frac{\partial \beta_i}{\partial v_{I_o}} & \frac{\partial \beta_i}{\partial v_{J_o}} & \frac{\partial \beta_i}{\partial v_{K_o}} \\ \frac{\partial el_i}{\partial r_{I_o}} & \frac{\partial el_i}{\partial r_{J_o}} & \frac{\partial el_i}{\partial r_{K_o}} & \frac{\partial el_i}{\partial v_{I_o}} & \frac{\partial el_i}{\partial v_{J_o}} & \frac{\partial el_i}{\partial v_{K_o}} \end{bmatrix} \quad (10-15)$$

The position and velocity vectors are the nominal state at epoch t_o ; thus, these partials relate observations at many different times to a state vector at one *epoch* time. In essence, then, the \mathbf{A} matrix describes how changes in the initial state (position and

velocity vectors) affect the computed 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 need to include how perturbations affect this overall motion.

To find these nonlinear, time-dependent derivatives, the first step is to break up the partial derivative using the chain rule:

$$\mathbf{A} = \frac{\partial \text{observations}}{\partial \hat{\mathbf{X}}_o} = \frac{\partial \text{observations}}{\partial \hat{\mathbf{X}}} \frac{\partial \hat{\mathbf{X}}}{\partial \hat{\mathbf{X}}_o} = \mathbf{H}\Phi \quad (10-16)$$

Eq. (10-16) distinguishes the observation partial derivatives (\mathbf{H}) from the partial derivative of the state over time. This latter matrix is called the matrix of *variational equations*, or the *error state transition matrix*, Φ and it relates the state errors at time t ($\delta\hat{\mathbf{x}}$), to the state errors at time t_o ($\delta\hat{\mathbf{x}}_o$). I'll use a subscript "o" for the epoch time only to simplify the notation. I also don't specify the time dependence (t) with the state vector for clarity. We calculate the two matrices in Eq. (10-16) using analytical or numerical integration techniques, or by finite differencing. We'll discuss finite differencing here and delay the remaining techniques until Sec. 10.7.

It's fairly simple to approximate the complete \mathbf{A} matrix [Eq. (10-16)] using ***finite differencing*** because we take small differences of the state to determine their effect on the system. The technique is independent of any particular propagation method.

$$\frac{\partial \text{observations}}{\partial \hat{\mathbf{X}}_o} \approx \frac{f(\hat{\mathbf{X}} + \delta_i) - f(\hat{\mathbf{X}})}{\delta_i}$$

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 one of the nominal and one of the varied 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 the vector magnitude rather than a fixed delta or a percentage of each element when using position and velocity vectors in the state vector. This technique eliminates the sensitivity to large and small values within the state vector elements. But it's really not appropriate for orbital elements like 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 \mathbf{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,

$$\delta_i = \hat{\mathbf{X}}_{mod_i} - \hat{\mathbf{X}}_{nom_i}$$

you must also calculate each observation as the modified observation minus the nominal observation. The approximation of the derivative of each observation with respect to an element of the state becomes

$$\frac{\partial \text{observations}}{\partial \hat{\mathbf{X}}_o} \approx \frac{\text{obs}_{mod} - \text{obs}_{nom}}{\delta_i}$$

Another approach, ***central differencing***, takes both positive and negative deviations from the nominal state vector to determine the approximation:

$$\frac{\partial \text{observations}}{\partial \hat{\mathbf{X}}_o} \approx \frac{f(\hat{\mathbf{X}} + \delta_i) - f(\hat{\mathbf{X}} - \delta_i)}{2\delta_i} \quad (10-17)$$

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 a depiction of the propagations of the state at epoch and an observation time in Fig. 10-6.

In algorithm form, finite differencing looks like this (you can use any propagator):

FOR $j = 1$ to the number of states

$$\begin{aligned} & \mathbf{PKEPLER}(\hat{\mathbf{r}}_{nom}, \hat{\mathbf{v}}_{nom}, \Delta t \Rightarrow \hat{\mathbf{r}}_i, \hat{\mathbf{v}}_i) \\ & \mathbf{RAZEL}(\hat{\mathbf{r}}_{ECI_i}, \hat{\mathbf{v}}_{ECI_i}, yr, mo, day, UTC, \Delta UT1, \Delta AT, x_p, y_p \\ & \quad \phi_{gd}, \lambda, h_{ellp} \Rightarrow \rho_n, \beta_n, el_n, \dot{\rho}_n, \dot{\beta}_n, \dot{el}_n) \end{aligned}$$

Modify each component of the nominal state vector

$$\delta_j = \hat{\mathbf{X}}_{nom_j}(0.01) \text{ modify by 1% of the original value}$$

$$\hat{\mathbf{X}}_{mod_j} = \hat{\mathbf{X}}_{nom_j} + \delta_j$$

$$\begin{aligned} & \mathbf{PKEPLER}(\hat{\mathbf{r}}_{mod}, \hat{\mathbf{v}}_{mod}, \Delta t \Rightarrow \hat{\mathbf{r}}_i, \hat{\mathbf{v}}_i) \\ & \mathbf{RAZEL}(\hat{\mathbf{r}}_{ECI_i}, \hat{\mathbf{v}}_{ECI_i}, yr, mo, day, UTC, \Delta UT1, \Delta AT, x_p, y_p \\ & \quad \phi_{gd}, \lambda, h_{ellp} \Rightarrow \rho_m, \beta_m, el_m, \dot{\rho}_m, \dot{\beta}_m, \dot{el}_m) \end{aligned}$$

Find A matrix elements 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 $\hat{\mathbf{X}}_{mod_j}$ to its original value $\hat{\mathbf{X}}_{nom_j}$

The least-squares process minimizes 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 obser-

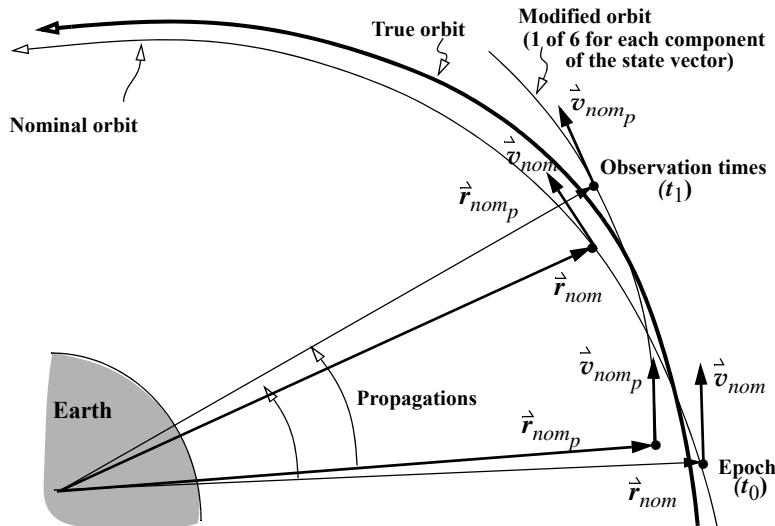


Figure 10-6. Propagations to Calculate the A Matrix by Finite Differencing. We find the partial derivative of the observations with respect to the state by modifying the initial state, propagating the nominal and modified states to each observation time and finding the difference of the observations. This figure shows movement of one nominal and modified state vector to an observation time. Although we can choose any epoch time, the beginning or the end of a batch of data are most common.

vations according to their assumed accuracy. The cost function is a scalar function that depends on the computed observations which comes from the state vector. We treat the actual observations as constants. We can think of the cost function as a manifold (topological surface) in an $n+1$ dimensional space with wells corresponding to the minima of the cost function. In this analogy, n is the number of parameters in the state vector. The initial conditions give 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.

10.4.2 Implementing Least-Squares Techniques

Developing a generic computer routine for differential correction requires us to take into account many issues:

1. Form of the state vector: position and velocity vectors or orbital elements, and which solve-for parameters to include—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 and force models
6. Ability to consider dynamic modeling errors—*consider parameters*
7. Tolerances and criteria for ending the computation

We've introduced the importance of the radius of convergence to a solution—the initial orbit is very important. In the case of initial orbit determination, we know little about the orbit and we have only observational data. In these cases, the initial (nominal) orbit comes from one of the techniques in Chap. 7. Although the **SITE-TRACK** algorithm suggests we can form position and velocity vectors from each observation, it's more common to have data that allows us to form only the position vector.

Algorithm 66 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 an angles-only method, and then are moved to a common time and divided by the number of observations (all in ECI). Although I show a simplified propagator (**PKEPLER**), I recommend a more accurate scheme for practical problems. For cases when the data spans a modest time interval (days or weeks), **PKEPLER** is probably sufficient to form an initial guess.* The initial nominal orbit will converge on the correct solution as long as the local linearity assumptions in the region around the reference trajectory hold over the fit span. Some orbits won't converge without a fairly close initial estimate, especially with limited data. Note that t_i represents the UTC time and the Earth Orientation Parameters to change between ECI and ECEF.

ALGORITHM 66: *NOMINAL STATE* ($\phi_{gd}, \lambda, h_{ellp}, obs, t_i \Rightarrow \dot{\vec{r}}_{nom}, \dot{\vec{v}}_{nom}$)

SITE-TRACK at each t_i

($\phi_{gd}, \lambda, h_{ellp}, \rho, \beta, el, yr, mo, day, UTC, \Delta UT1, \Delta AT, x_p, y_p \Rightarrow \dot{\vec{r}}_{ECI}, \dot{\vec{v}}_{ECI}$)

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 **ANGLES-ONLY**)

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 $\bar{X}_{nom} = \begin{bmatrix} \dot{\vec{r}}_{nom} \\ \dot{\vec{v}}_{nom} \end{bmatrix}$

Estimating a state vector depends on the quantity, availability, and type of observational data. In general, to assure observability, 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

* Although Algorithm 66 gives an initial estimate of the state vector, it's more convenient to already have an initial estimate of the satellite's state vector from another source so we can proceed directly. Except for new launches, this is usually the case.

the confidence of the result. Seven basic types or combinations of observational data are usually available (see Sec. 4.2). Each type of data requires us to find observation partial derivatives. We should also use relative weighting (noise and bias, see Table 4-4 for sample values) during the differential correction process for improved accuracy.

Choosing between analytical and numerical partial derivatives rests with the treatment of Eq. (10-16). For improved accuracy, we determine the observation partial derivatives analytically and evaluate the variational equations either numerically or analytically. Analytical formulations for the variational equations have an advantage over numerical techniques due to the increased speed. However, numerical approaches can be more accurate because the analytical model truncates series expansions that would be infinite even for a modest set of perturbations.

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 just a few more iterations to converge. Despite the flexibility in evaluating the partial-derivative matrix, we must be careful. Remember that the state gives us the most likely position of the satellite, and the covariance gives us the orientation and size of the error ellipsoid that contains the satellite, within a certain confidence level (1σ , 2σ , etc.). For mission applications, we must propagate the state and covariance into the future. We propagate the state with complete force models, consistent with those used in the differential correction, to preserve as much accuracy as possible. We propagate the covariance with reduced force models because we're only moving the [small] uncertainty in the state. Because we use the error state transition matrix to propagate the covariance, and Φ is found using linearized dynamics, the rigorous force models aren't as important, but they may influence the statistical validity of the propagated covariance.

The choice of propagator should also be tied to the accuracy of the observations. A good rule is to use force models that are an order of magnitude more accurate than your best observations (if possible). 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!

On the computer, we can accumulate $\mathbf{A}^T \mathbf{W} \mathbf{A}$ and $\mathbf{A}^T \mathbf{W} \tilde{\mathbf{b}}$ as in the linear and nonlinear examples. Suppose we have 1000 observations of ρ , β , and el —for orbit estimation. For a six-parameter state space, our \mathbf{A} matrix will be 6×3000 , \mathbf{W} will be 3000×3000 , and $\tilde{\mathbf{b}}$ will be 3000×1 . This imposes significant memory requirements on a program. However, we can evaluate each observation individually and sum the result. This process reduces the matrix sizes to 6×3 , 3×3 , and 3×1 respectively—a significant savings!

The most difficult question may be “when to quit”. Because orbit determination isn't linear, we change the orbital elements for each step and iterate until the residuals stop changing. However, the units of measure for each observation may differ vastly from the others; therefore, we must do some scaling to keep the largest ones from controlling the process. We can do this with the constant weights (\mathbf{W}_i matrices); thus, if

$$RMS_{new} = \sqrt{\frac{1}{N} \sum_{i=1}^N w_i^2 \bar{r}_i^2}$$

we define the weighted RMS as

$$RMS_{new} = \sqrt{\frac{1}{N} \frac{\sum_{i=1}^N (\Delta\rho_i)^2 + \sum_{i=1}^N (\Delta\beta_i)^2 + \sum_{i=1}^N (\Delta el_i)^2}{\frac{\sigma_\rho^2}{3} + \frac{\sigma_\beta^2}{3} + \frac{\sigma_{el}^2}{3}}} = \sqrt{\frac{\tilde{\mathbf{b}}^T \tilde{\mathbf{W}} \tilde{\mathbf{b}}}{n_{meas}(N)}} \quad (10-18)$$

We've added n_{meas} to this expression to account for the total number of different types of measurements associated with an observation (for example $n_{meas} = 3$ as shown). We stop iterating whenever the change in the weighted RMS becomes small, as in Eq. (10-13). 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}) can cause the iterations to eventually become unstable. A percentage change of the RMS values usually eliminates this difficulty when coupled with a limit of, say, 1×10^{-3} .

The principles in Algorithm 67 are valid for a number of different applications. I have not shown a specific propagation technique in the algorithm because the choice of a propagator is tied to the desired accuracy of the differential correction process. Finally, as with the **HGIBBS** method, you should be careful to preserve accuracy if using the JD as the time argument. The Modified Julian Date is recommended in these cases.

ALGORITHM 67: Differential Correction (Obs at t_i , $\hat{\mathbf{X}}_{nom} \Rightarrow \hat{\mathbf{X}}_o$)

FOR $i = 1$ to the number of observations N

 Propagate the nominal state to each t_i and find computed observations

$$\hat{\mathbf{X}}_{nom_i} = \int_{t_o}^{t_i} \dot{\hat{\mathbf{X}}}_{nom} dt + \hat{\mathbf{X}}_{nom}$$

RAZEL (\vec{r}_{nomECP} , \vec{v}_{nomECP} , yr, mo, day, UTC,

$\Delta UT1$, ΔAT , x_p , y_p , ϕ_{gd} , λ , $h_{ellp} \Rightarrow \rho$, β , el , $\dot{\rho}$, $\dot{\beta}$, \dot{el})

 Find the $\tilde{\mathbf{b}}$ matrix as observed minus computed data at each t_i

 Perform Finite Differencing for the A matrix, or:

$$\frac{d\hat{\mathbf{X}}}{dt} = \dot{\hat{\mathbf{X}}}_{2-body} + \dot{\hat{\mathbf{X}}}_{nonspherical} + \dot{\hat{\mathbf{X}}}_{drag} + \dot{\hat{\mathbf{X}}}_{3-body} + \dot{\hat{\mathbf{X}}}_{SR} + \dot{\hat{\mathbf{X}}}_{other}$$

$$\begin{aligned} \dot{\hat{\mathbf{F}}} &= \frac{\partial \dot{\hat{\mathbf{X}}}}{\partial \hat{\mathbf{X}}} \quad \Phi(t, t_o) = \int_{t_o}^t \mathbf{F}(t) \Phi(t, t_o) dt \\ \mathbf{A} &= \frac{\partial \text{observations}}{\partial \hat{\mathbf{X}}} \frac{\partial \dot{\hat{\mathbf{X}}}}{\partial \hat{\mathbf{X}}_o} = \mathbf{H}\Phi \end{aligned}$$

Accumulate $\mathbf{A}^T \mathbf{W} \mathbf{A}$ and $\mathbf{A}^T \mathbf{W} \tilde{\mathbf{b}}$

END LOOP

$$\delta\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W} \tilde{\mathbf{b}} = \mathbf{P} \mathbf{A}^T \mathbf{W} \tilde{\mathbf{b}}$$

$$\text{Check RMS for convergence} = \sqrt{\frac{\tilde{\mathbf{b}}^T \mathbf{W} \tilde{\mathbf{b}}}{n_{\text{meas}}(N)}}$$

Update the state vector and repeat if not converged:

$$\hat{\mathbf{X}}_{\text{nom}} = \hat{\mathbf{X}}_{\text{nom}} + \delta\hat{\mathbf{x}}$$

Here's an example to demonstrate this technique.

▼ Example 10-4. Using Differential Correction for Least-Squares Orbit Determination.

GIVEN: Observations from Kaena Point on January 29, 1995 for GEOS-III (see Table 10-2).

$\Delta AT = 29$ s, $\Delta UT1 = 0.326\ 106\ 8$ s, $x_p = -0.115\ 54''$, $y_p = 0.481\ 87''$, $m = 341$ kg, $A = 1.1\ m^2$. Assume LOD = 0.0 sec and use the site coordinates from Table 4-2.

FIND $\hat{\mathbf{X}}, \hat{\mathbf{P}}$ at an epoch at the first observation, using the first ten observations.

Raw observations may arrive in a condensed format which you can expand or parse. 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 only minimal information, and discuss the final answer. If we use Algorithm 66, (use Herrick Gibbs because the observations are closely spaced) and subtract the biases from Table 4-4 from each observation, the first ten observations (at the epoch time [first observation] of JD = 2,449,746.610 150) give us a nominal vector in J2000 (IAU-76/FK5) $\hat{\mathbf{X}}_{\text{nom}}^T = 5748.5350\ 2679.6404\ 3442.8654\ 4.328\ 274 - 1.918\ 662 - 5.727\ 629$ km, km/s. This vector is close to the answer, so let's use a vector that is a little farther from the answer so we can better see the iterations.

$$\hat{\mathbf{X}}_{\text{nom}}^T = [5975.2904\ 2568.6400\ 3120.5845\ 3.983\ 846 - 2.071\ 159 - 5.917\ 095] \text{ km, km/s}$$

We'll use finite differencing to find the partial-derivative matrix. The first iteration using $N = 10$ observations produces accumulated matrices of

$$\mathbf{A}^T \mathbf{W} \mathbf{A} = \sum_{i=1}^{10} \begin{bmatrix} \mathbf{A}^T \\ \mathbf{W} \\ \mathbf{A} \end{bmatrix}_{6 \times 3} \begin{bmatrix} \mathbf{W} \\ \mathbf{A} \end{bmatrix}_{3 \times 3} \begin{bmatrix} \mathbf{A}^T \\ \mathbf{W} \\ \mathbf{A} \end{bmatrix}_{3 \times 6}$$

TABLE 10-2. Example Observations. Data for GEOS-III (#7734) is from Kaena Point, Hawaii.

Sat	Yr Mo D	Time (UTC)	Range (km)	Az (°)	El (°)
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

$$A^T W A = \begin{bmatrix} 136.7 & 240.2 & 73.7 & 9290.2 & 16859.2 & 3981.9 \\ 240.2 & 1006.2 & 300.1 & 15580.9 & 56523.9 & 11843.9 \\ 73.7 & 300.1 & 145.4 & 3906.1 & 12469.0 & 5665.1 \\ 9290.2 & 15580.9 & 3906.1 & 767027.6 & 1322814.7 & 278780.4 \\ 16859.2 & 56523.9 & 12469.0 & 1322814.7 & 4368107.9 & 764948.9 \\ 3981.9 & 11843.9 & 5665.1 & 278780.4 & 764948.9 & 365885.3 \end{bmatrix}$$

$$A^T W \tilde{b} = \sum_{i=1}^{10} \begin{bmatrix} A^T \\ 6 \times 3 \end{bmatrix} \begin{bmatrix} W \\ 3 \times 3 \end{bmatrix} \begin{bmatrix} \tilde{b} \\ 3 \times 1 \end{bmatrix} = \begin{bmatrix} 38,634.2 \\ 215,383.4 \\ 79,253.9 \\ 2,260,058.9 \\ 10,504,371.5 \\ 2,995,933.5 \end{bmatrix}$$

where we've used the weights (noise) from Table 4-4. Be careful to match the units here. I'm using km but Table 4-4 is in meters. I show degrees for the angles, but programs would use radians.

$$\mathbf{W} = \begin{bmatrix} \frac{1}{\sigma_\rho^2} & 0 & 0 \\ 0 & \frac{1}{\sigma_\beta^2} & 0 \\ 0 & 0 & \frac{1}{\sigma_{el}^2} \end{bmatrix} = \begin{bmatrix} \frac{1}{0.0925^2} & 0 & 0 \\ 0 & \frac{1}{0.0224^2} & 0 \\ 0 & 0 & \frac{1}{0.0139^2} \end{bmatrix}$$

The update to the state is

$$\hat{\delta x} = (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W} \tilde{\mathbf{b}} = \mathbf{P} \mathbf{A}^T \mathbf{W} \tilde{\mathbf{b}}$$

$$\hat{\delta x}^T = [-196.648 \ 153.442 \ 299.313 \ 0.235 \ 155 \ 0.262 \ 115 \ -0.000 \ 201] \text{ (km, km/s)}$$

and the nominal state becomes

$$\hat{X}_{nom} = \hat{X}_{nom} + \hat{\delta x}$$

$$\hat{X}_{nom}^T = [5778.642 \ 2722.082 \ 3419.897 \ 4.219 \ 001 \ -1.809 \ 044 \ -5.917 \ 296] \text{ (km, km/s)}$$

Subsequent iterations produce

$$\hat{\delta x} = \begin{bmatrix} -30.6618 \\ -41.9439 \\ 24.2692 \\ 0.113 \ 001 \\ -0.108 \ 003 \\ 0.197 \ 049 \end{bmatrix}, \begin{bmatrix} 0.6172 \\ -0.4073 \\ -1.1642 \\ -0.0015 \\ -0.0058 \\ -0.0063 \end{bmatrix}, \begin{bmatrix} 0.0037 \\ -0.0021 \\ 0.0051 \\ -0.0000 \\ 0.0000 \\ -0.0000 \end{bmatrix}, \begin{bmatrix} -0.000 \ 005 \\ 0.000 \ 095 \\ -0.000 \ 005 \\ 0.000 \ 000 \\ 0.000 \ 000 \\ 0.000 \ 000 \end{bmatrix} \text{ (km and km/s)}$$

The final answer and covariance matrix are (ECI)

$$\hat{r}_1 = 5748.6011 \hat{i} + 2679.7287 \hat{j} + 3443.0073 \hat{k} \text{ km}$$

$$\hat{v}_1 = 4.330 \ 462 \hat{i} - 1.922 \ 862 \hat{j} - 5.726 \ 564 \hat{k} \text{ km/s}$$

$$\hat{\mathbf{P}} = \begin{bmatrix} 0.073 \ 098 & 0.004 \ 509 & -0.008 \ 370 & -8.6193 \times 10^{-4} & -3.8277 \times 10^{-5} & 4.1945 \times 10^{-5} \\ 0.004 \ 509 & 0.040 \ 444 & -0.062 \ 260 & -1.9932 \times 10^{-6} & -4.4565 \times 10^{-4} & 6.6809 \times 10^{-4} \\ -0.008 \ 370 & -0.062 \ 260 & 0.107 \ 335 & 5.2887 \times 10^{-5} & 6.9359 \times 10^{-4} & -1.2223 \times 10^{-3} \\ -8.6193 \times 10^{-4} & -1.9932 \times 10^{-6} & 5.2887 \times 10^{-5} & 1.3926 \times 10^{-5} & -3.7863 \times 10^{-7} & -9.5780 \times 10^{-7} \\ -3.8277 \times 10^{-5} & -4.4565 \times 10^{-4} & 6.9359 \times 10^{-4} & -3.7863 \times 10^{-7} & 5.9647 \times 10^{-6} & -9.5101 \times 10^{-6} \\ 4.1945 \times 10^{-5} & 6.6809 \times 10^{-4} & -1.2223 \times 10^{-3} & -9.5780 \times 10^{-7} & -9.5101 \times 10^{-6} & 1.9962 \times 10^{-5} \end{bmatrix}$$

We first look at the confidence in the answer using the covariance matrix. Our answer has a standard deviation (1σ) r_I of 270 m ($\sqrt{0.0731 \times 1000}$), r_J is 201 m and r_K is 328 m for a total of 470 m. A more complete evaluation is found in the eigenvalues and eigenvectors which give the error ellipse orientation and dimensions. Numerical programs (such as Matlab) may sort the eigenvectors. The eigenvectors are as follows.

$$\begin{aligned} e1 &= -0.129 \ 514 & -0.508 \ 518 & 0.851 \ 183 & 0.001 \ 081 & 0.005 \ 638 & -0.009 \ 504 \\ e2 &= -0.991 \ 487 & 0.071 \ 737 & -0.107 \ 992 & 0.011 \ 812 & -0.000 \ 959 & 0.001 \ 925 \\ e3 &= 0.006 \ 210 & 0.857 \ 928 & 0.513 \ 350 & 0.006 \ 283 & -0.008 \ 218 & -0.016 \ 904 \\ e4 &= -0.001 \ 995 & -0.011 \ 735 & -0.005 \ 090 & -0.158 \ 071 & -0.918 \ 045 & -0.363 \ 371 \\ e5 &= 0.011 \ 551 & -0.006 \ 382 & -0.000 \ 278 & 0.965 \ 401 & -0.220 \ 819 & 0.138 \ 076 \end{aligned}$$

$$\mathbf{e}_6 = -0.001\ 668 \quad 0.006\ 675 \quad 0.016\ 462 \quad -0.206\ 963 \quad -0.329\ 138 \quad 0.921\ 150$$

The eigenvalues are (for position and velocity)

$$381.866 \quad 268.086 \quad 56.690 \text{ m and } 0.4487 \quad 1.8869 \quad 2.5488 \text{ m/s.}$$

The magnitude of the three position eigenvalues is also 470 m. The eigenvectors are not generally aligned with the Cartesian position vectors. If you plot them, you'll see that they do not coincide with the position or velocity vectors. As a general rule, the longest error dimension generally lies along the velocity vector (in-track). However, the small quantity of data we used in this example causes the error ellipse to be aligned in arbitrary directions. As we process more data, the ellipse conforms better to the in-track generality.

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 laser program. The test demonstrated increased accuracy by nearly a full order of magnitude over existing operational methods (Phillips Laboratory, 1995). The AF Space Command used only 3-5 observations per pass from a few sensors giving 400 m accuracy (note how close this is to our simple answer above!). Phillips Laboratory used 30-50 observations per pass from a single station, giving about 10 m accuracy [the original TOD value has been converted to J2000 (IAU-76/FK5)].

$$\hat{\mathbf{x}}_{nom}^T = [5748.6070 \ 2679.9324 \ 3442.7902 \ 4.330\ 460 \ -1.922\ 874 \ -5.726\ 562] \text{ (km, km/s)}$$

Notice the similarity between the answers (only about 297 m difference). The standard-deviation RMS from the actual experiment (combined cross-track and along-track only) was about 4 m. This coincided with the sensor bias solved for during the differential correction. Recall we used a range weight of 92.5 m. If we rerun the simple example using a range bias of say 8 m, our answers change slightly, and they better represent the "true" answer. The standard deviation is the same, but our answer is about 45 m closer. The main differences between the simple and actual experiments arise from using precise propagation formulas in the differential correction, and the quantity of data included in the problem.

Finally, if we assume a generic inertial frame and discount the ECI-ECEF conversions, we arrive at a very different answer, but actually, this is close to a TOD answer since nutation and precession are not included. Nevertheless, the answer is only about 740 m off from the TOD "actual" answer.

$$\blacktriangle \quad \hat{\mathbf{x}}_{nom}^T = [5753.173 \ 2673.361 \ 3440.304 \ 4.324\ 207 \ -1.924\ 299 \ -5.728\ 216] \text{ (km, km/s)}$$

10.5 Sequential-Batch Least Squares

Suppose we have 1,000 observations and we use Eq. (10-14) 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 method processes successive batches of data. We use the statistical information from the least squares processing of the first 1000 observations, and combine it with the new information. The operation seeks a conditional probability density function, and we evaluate it using **Bayes Theorem** which for two events, x and z , is stated as

$$p(x|z) = \frac{p(z|x)p(x)}{p(z)}$$

The notation of $p(z|x)$ indicates the conditional probability of z occurring given that x has already occurred. This sequential-batch technique is sometimes called a **Bayes filter** or **Bayesian estimation** because it processes segments of data and then updates the state

and covariance using the *a priori* data (Gelb, 1989:103-105). Essentially, the previous state estimate becomes the given event (x above), and the second event (z) is the new data.

We must address two main issues: how to incorporate new data and how to propagate the state estimate and covariance to the current time. Each technique mentioned above will incorporate new data; however, it's possible to formulate 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:

$\hat{\mathbf{X}}(0|k) \Rightarrow$ the *best state estimate* (or answer) at t_o , based on observations through t_k

$\delta\hat{\mathbf{x}}(0|k) \Rightarrow$ 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.

$\tilde{\mathbf{b}}_k \Rightarrow$ the *residual matrix* containing residuals through t_k

$\mathbf{A}_k \Rightarrow$ the *partial-derivative matrix* relating all observations through t_k to the state at t_o

$\hat{\mathbf{P}} \equiv (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \Rightarrow$ the *covariance matrix*. It's expressed as $\hat{\mathbf{P}}(0|k)$, 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, \mathbf{W}_{new} . Summarizing, we have a state-update estimate, $\delta\hat{\mathbf{x}}(0|k)$, based on observations through time t_k ; the accuracy estimate, $\hat{\mathbf{P}}(0|k)$; some observations, y_n , at time, t_n ; and some statistical values estimating the accuracy of the new observations, \mathbf{W}_{new} . One 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 divided residual equation, we see the residual matrix and the partial-derivative matrix are now

$$\tilde{\mathbf{b}} = \begin{bmatrix} [\tilde{\mathbf{b}}]_k \\ \vdots \\ [\tilde{\mathbf{b}}]_{new} \end{bmatrix} \quad \mathbf{A} = \begin{bmatrix} [\mathbf{A}]_k \\ \vdots \\ [\mathbf{A}]_{new} \end{bmatrix}$$

The $[\mathbf{A}]_{new}$ matrix accounts for the new observations, and we find it as before [Eq. (10-16)]. The partial derivatives are for the new observation with regard to the state at the original epoch, t_o , to account for the new data. The weighting matrix is as before:

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_k & 0 \\ 0 & \mathbf{W}_{new} \end{bmatrix}$$

with the components of the \mathbf{W}_{new} matrix being the reciprocal of the observation measurement noise. The nonlinear update equation [Eq. (10-14)] now becomes

$$\hat{\delta}\mathbf{x}(0|k+n) = \left([\mathbf{A}_k^T \mathbf{A}_{new}^T] \begin{bmatrix} \mathbf{W}_k & 0 \\ 0 & \mathbf{W}_{new} \end{bmatrix} \begin{bmatrix} \mathbf{A}_k \\ \mathbf{A}_{new} \end{bmatrix} \right)^{-1} [\mathbf{A}_k^T \mathbf{A}_{new}^T] \begin{bmatrix} \mathbf{W}_k & 0 \\ 0 & \mathbf{W}_{new} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{b}}_k \\ \tilde{\mathbf{b}}_{new} \end{bmatrix}$$

Expanding terms gives us

$$\hat{\delta}\mathbf{x}(0|k+n) = (\mathbf{A}_{new}^T \mathbf{W}_{new} \mathbf{A}_{new} + \mathbf{A}_k^T \mathbf{W}_k \mathbf{A}_k)^{-1} (\mathbf{A}_{new}^T \mathbf{W}_{new} \tilde{\mathbf{b}}_{new} + \mathbf{A}_k^T \mathbf{W}_k \tilde{\mathbf{b}}_k)$$

The covariance matrix is, $\hat{\mathbf{P}}_k^{-1} = \mathbf{A}_k^T \mathbf{W}_k \mathbf{A}_k$, in Eq. (10-14). So:

$$\hat{\delta}\mathbf{x}(0|k+n) = (\mathbf{A}_{new}^T \mathbf{W}_{new} \mathbf{A}_{new} + \hat{\mathbf{P}}_k^{-1})^{-1} (\mathbf{A}_{new}^T \mathbf{W}_{new} \tilde{\mathbf{b}}_{new} + \mathbf{A}_k^T \mathbf{W}_k \tilde{\mathbf{b}}_k) \quad (10-19)$$

and the updated covariance matrix is

$$\hat{\mathbf{P}}_{k+n} = \hat{\mathbf{P}}(0|k+n) = (\mathbf{A}_{new}^T \mathbf{W}_{new} \mathbf{A}_{new} + \hat{\mathbf{P}}_k^{-1})^{-1} \quad (10-20)$$

so $\hat{\delta}\mathbf{x}(0|k+n)$ is the improved correction to $\hat{\mathbf{X}}(0|k)$ due to additional data through time t_n . We still have a nonlinear problem and a linearized solution, 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, a computer implementation should require larger \mathbf{A} and \mathbf{b} matrices. But careful examination of $\hat{\delta}\mathbf{x}$ in Eq. (10-19) and Eq. (10-20) shows that although \mathbf{A} and \mathbf{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 67 also applies to sequential batch formulations.

We can rearrange the inverse in Eq. (10-20) to reduce the complexity of the matrix inversion so it has the same dimension as the next subset of measurements. This form is similar to Eq. (10-29), which we'll discuss in the Kalman-filter section. \mathbf{W} usually doesn't require calculation because it's the matrix for known measurement errors.

$$\hat{\mathbf{P}}_{k+n} = \hat{\mathbf{P}}_k - \hat{\mathbf{P}}_k \mathbf{A}_{new}^T (\mathbf{W}_{new}^{-1} + \mathbf{A}_{new} \hat{\mathbf{P}}_k \mathbf{A}_{new}^T)^{-1} \mathbf{A}_{new} \hat{\mathbf{P}}_k \quad (10-21)$$

Eq. (10-21) is the solution, and it's 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.

Now suppose we get another batch of observations at a new time, t_{n+1} . We can use the sequential batch [Eq. (10-19) and Eq. (10-20)] just by saving the matrix on the left-hand brackets (before inverting) as $\hat{\mathbf{P}}(0|k)$ 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).

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 *error state transition matrix* [Eq. (10-29)] is used to propagate our covariance.

$$\hat{\mathbf{P}} \approx \Phi \hat{\mathbf{P}}_o \Phi^T + \mathbf{Q}$$

This is really a sequential estimation technique and it also invokes an additional component (process noise, \mathbf{Q}) in the formulation. We'll discuss the Kalman filter shortly. The result is a predicted estimate using the new data. Remember that Φ is often a linearized approximation in the differential correction. When we use it in propagating the covariance matrix, we may introduce unnecessary errors.

Continuing with Example 10-4, let's look at Example 10-5, which presents the sequential-batch solution:

▼ Example 10-5. Using Sequential-Batch Estimation for Differential Correction.

GIVEN: Eight new observations from Kaena Point on January 29, 1995 for GEOS-III, and the solution from Example 10-4.

FIND: $\hat{\mathbf{X}}, \hat{\mathbf{P}}$

We examine the last eight observations in Table 10-2 on page 770. We know the predicted value of the state from Example 10-4 but must find the accumulated matrices for the new data. For the first iteration, we could choose any epoch (and would probably choose a later time), but let's keep the original epoch from Example 10-4. The weighting matrix remains the same as in Example 10-4 because the observations are from the same site. If we had observations from a different station, we could easily insert the noise parameters during the accumulation process.

$$\mathbf{A}_{new}^T \mathbf{W}_{new} \mathbf{A}_{new} = \sum_{i=1}^8 [\mathbf{A}_{new}^T]_i [\mathbf{W}_{new}]_i [\mathbf{A}_{new}]_i$$

$i = 1$	6×3	3×3	3×6
---------	--------------	--------------	--------------

$$\mathbf{A}_{new}^T \mathbf{W}_{new} \mathbf{A}_{new} = \begin{bmatrix} 216.7346 & 327.1327 & 42.7977 & 3.541 \times 10^4 & 5.508 \times 10^4 & 4684.2902 \\ 327.1327 & 785.4208 & 87.5320 & 5.119 \times 10^4 & 1.280 \times 10^5 & 8758.9867 \\ 42.7977 & 87.5320 & 62.3664 & 5791.8322 & 1.173 \times 10^4 & 9613.1211 \\ 3.541 \times 10^4 & 5.119 \times 10^4 & 5791.8322 & 5.929 \times 10^6 & 8.845 \times 10^6 & 5.980 \times 10^5 \\ 5.508 \times 10^4 & 1.280 \times 10^5 & 1.173 \times 10^4 & 8.845 \times 10^6 & 2.145 \times 10^7 & 1.038 \times 10^6 \\ 4684.2902 & 8758.9867 & 9613.1211 & 5.980 \times 10^5 & 1.038 \times 10^6 & 1.555 \times 10^6 \end{bmatrix}$$

and

$$\mathbf{A}_{new}^T \mathbf{W}_{new} \tilde{\mathbf{b}}_{new} = \sum_{i=1}^8 [\mathbf{A}_{new}^T]_{6 \times 3} [\mathbf{W}_{new}]_{3 \times 3} [\tilde{\mathbf{b}}_{new}]_{3 \times 1} = \begin{bmatrix} 19.8969 \\ -18.1124 \\ -0.2737 \\ 3815.5860 \\ -3395.8956 \\ 69.3738 \end{bmatrix}$$

Using Eq. (10-19) and Eq. (10-20), we find the updated state. Although these parameters contain additional observations, they refer to the original epoch. The $\mathbf{A}_k^T \mathbf{W}_k \mathbf{A}_k$ and $\mathbf{A}_k^T \mathbf{W}_k \tilde{\mathbf{b}}_k$ matrices remain fixed for the iteration process.

$$\begin{aligned} \hat{\delta \mathbf{x}}_{k+n} &= (\mathbf{A}_{new}^T \mathbf{W}_{new} \mathbf{A}_{new} + \mathbf{A}_k^T \mathbf{W}_k \mathbf{A}_k)^{-1} (\mathbf{A}_{new}^T \mathbf{W}_{new} \tilde{\mathbf{b}}_{new} + \mathbf{A}_k^T \mathbf{W}_k \tilde{\mathbf{b}}_k) \\ \hat{\mathbf{P}}_{k+n} &= (\mathbf{A}_{new}^T \mathbf{W}_{new} \mathbf{A}_{new} + \hat{\mathbf{P}}_k^{-1})^{-1} \end{aligned}$$

The iterations produce

$$\hat{\delta \mathbf{x}}_{k+n} = \begin{bmatrix} -0.1180 \\ 0.1733 \\ -0.2102 \\ 0.002846 \\ -0.001967 \\ 0.001220 \end{bmatrix} = \begin{bmatrix} -0.0311 \\ 0.0773 \\ -0.0966 \\ 0.0005 \\ -0.0006 \\ 0.0003 \end{bmatrix} \text{ (km and km/s)}$$

The additional data adds only a little to the answer from Example 10-4. The final answer and covariance matrix are

$$\hat{r}_1 = 5748.4520 \hat{I} + 2679.9794 \hat{J} + 3442.7005 \hat{K} \text{ km}$$

$$\hat{\dot{v}}_1 = 4.333795 \hat{I} - 1.925435 \hat{J} - 5.725071 \hat{K} \text{ km/s}$$

$$\hat{\mathbf{P}} = \begin{bmatrix} 0.506797 & -0.166213 & -0.070408 & -2.9334 \times 10^{-3} & 9.2162 \times 10^{-4} & 3.5696 \times 10^{-4} \\ -0.166213 & 0.133276 & -0.081572 & 1.0249 \times 10^{-3} & -7.6469 \times 10^{-4} & 3.7110 \times 10^{-4} \\ -0.070408 & -0.081572 & 0.779042 & 3.4183 \times 10^{-4} & 3.1813 \times 10^{-4} & -4.4879 \times 10^{-4} \\ -2.9334 \times 10^{-3} & 1.0249 \times 10^{-3} & 3.4183 \times 10^{-4} & 1.7474 \times 10^{-5} & -5.8879 \times 10^{-6} & -1.8332 \times 10^{-6} \\ 9.2162 \times 10^{-4} & -7.6469 \times 10^{-4} & 3.1813 \times 10^{-4} & -5.8879 \times 10^{-6} & 4.5574 \times 10^{-6} & -1.2167 \times 10^{-6} \\ 3.5696 \times 10^{-4} & 3.7110 \times 10^{-4} & -4.4879 \times 10^{-4} & -1.8332 \times 10^{-6} & -1.2167 \times 10^{-6} & 2.6736 \times 10^{-5} \end{bmatrix}$$

Notice the similarity to the answer from Example 10-4. The eigenvalue standard deviation (1σ) has worsened and is now about 894 m, r_J is 751 m and r_K is 236 m, for a total of about 1191 m. This is

▲ caused by processing so few observations. If we had processed an additional 18 observations, the eigenvalue uncertainty would have been about 439 m.

10.6 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, weeks, or longer. 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 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 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. 10.6.5 because it has successfully demonstrated performance in over a decade of operational use with no difficulties.

* The principle of **Maximum likelihood** seeks to find an estimate of the state that maximizes the probability of having obtained the given observational data set.

Using Fig. 10-7 to illustrate multiple passes, 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 ($\hat{\mathbf{X}}_k$ and $\hat{\mathbf{P}}_k$) from data at the original epoch, t_k .

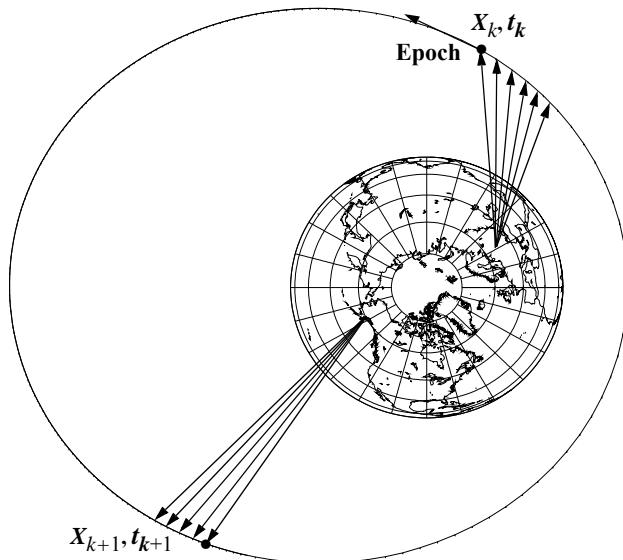


Figure 10-7. Observations from Multiple Passes. Differential correction minimizes the sum of the squares of the residuals by referring all corrections to a single state at a *fixed epoch*, t_k . The Kalman filter finds a state update at each observation time, t_{k+1} .

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 use 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, $\delta\mathbf{x}$, about a nominal trajectory. This concept is very similar to Encke's method. The extended Kalman filter uses the current state estimate to generate a new reference trajectory at each observation time. In both cases, we use a two-step process because we have a nonlinear system. We *predict* the new state estimate based on previous data, and then *update* that result with the new observations.

Linearizing the Dynamics—The error 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. The only constraint is that the two trajectories remain close enough so we can neglect the higher order terms [Eq. (10-24)]. If the two trajectories are \mathbf{X} and \mathbf{Y} , we can express the initial conditions and derivatives of the states as

$$\begin{aligned}\mathbf{X}(t_o) &= \mathbf{X}_o & \dot{\mathbf{X}} &= f(\mathbf{X}) \\ \mathbf{Y}(t_o) &= \mathbf{Y}_o & \dot{\mathbf{Y}} &= f(\mathbf{Y})\end{aligned}\quad (10-22)$$

where f simply denotes a function of the state. Letting the difference in the two trajectories be $\delta\mathbf{x}$ gives us

$$\mathbf{Y} = \mathbf{X} + \delta\mathbf{x} \quad (10-23)$$

Substituting Eq. (10-23) into Eq. (10-22) results in $\dot{\mathbf{Y}} = f(\mathbf{X} + \delta\mathbf{x})$. Now, because $f(\mathbf{X} + \delta\mathbf{x})$ is nonlinear, let's expand the function in a Taylor series about \mathbf{X} , ($\delta\mathbf{x} = \mathbf{Y} - \mathbf{X}$):

$$\dot{\mathbf{Y}} = f(\mathbf{X}) + \frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} \delta\mathbf{x} + \frac{\partial f^2(\mathbf{X})}{2! \partial \mathbf{X}^2} \delta\mathbf{x}^2 + \dots \quad (10-24)$$

Here, the partial derivative of $f(\mathbf{X})$ is a time varying square matrix. Next, take the time derivative of Eq. (10-23) and substitute it into the left-hand side of Eq. (10-24):

$$\dot{\mathbf{X}} + \delta\dot{\mathbf{x}} = f(\mathbf{X}) + \frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} \delta\mathbf{x} + \frac{\partial f^2(\mathbf{X})}{2! \partial \mathbf{X}^2} \delta\mathbf{x}^2 + \dots$$

Using Eq. (10-22), we can simplify this expression as

$$\delta\dot{\mathbf{x}} = \frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} \delta\mathbf{x} + \mathbf{u} = \mathbf{F}(t) \delta\mathbf{x} + \mathbf{u} \quad (10-25)$$

I've introduced two new quantities here. The neglected terms of the second order and higher are represented by a vector, \mathbf{u} , and $\mathbf{F}(t)$ is the matrix of partial derivatives (also called the **Jacobian matrix**) of the state rates as indicated (See Sec. 10.7 for formulations of \mathbf{F}). Eq. (10-25) represents the linearized dynamics and its solution is the time-varying difference between the two neighboring trajectories. Let's examine the formation of the \mathbf{F} matrix and the \mathbf{u} matrix.

We can find a relationship for \mathbf{F} by assuming a solution to Eq. (10-25) of the form

$$\delta\mathbf{x} = \Phi(t, t_o) \delta\mathbf{x}_o \quad (10-26)$$

Notice we've introduced the **error state transition matrix**, Φ , which moves 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. (10-26) and substitute into Eq. (10-25), allowing us to find a first-order differential equation and an initial condition:

$$\begin{aligned}\dot{\Phi}(t, t_o) &= \mathbf{F}(t)\Phi(t, t_o) \\ \Phi(t_o, t_o) &= \mathbf{I}\end{aligned}\tag{10-27}$$

The initial condition results from Eq. (10-26) if the time interval is zero. The following matrix identities apply to the error state transition matrix:

$$\begin{aligned}\Phi(t, t_o) &= \Phi^{-1}(t_o, t) \\ \Phi(t_1, t_o) &= \Phi(t_1, t)\Phi(t, t_o)\end{aligned}$$

An important restriction is that Φ be invertible; thus, the determinant of Φ can't equal zero.

For the orbit problem, the error state transition matrix is often obtained by numerically integrating Eq. (10-27) along with the equations of motion. The fidelity of the integration for Eq. (10-27) during the differential correction 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 error 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.

We must also evaluate the truncation term from Eq. (10-25). Over small time intervals, we will assume \mathbf{u} is essentially constant. Thus, using Euler's method (first-order approximation) results in

$$\delta\mathbf{x} = \mathbf{u}\delta t$$

The error 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 \mathbf{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

$$\mathbf{v} = \int_{t_o}^t \Phi(t, t_o)\mathbf{u} dt$$

This equation defines process noise. The ***process noise***, \mathbf{v} , is the uncertainty in the dynamics model used in the propagation process.

Estimating the Covariance and State Error at the Current Time

Using the linearized dynamics, the resulting updates can be much faster. Don't confuse the nonlinear propagation of the state with the propagation of the linearized error in the state. When you propagate the state by numerical integration, this is a nonlinear process. When you use Φ to propagate the state error, you've assumed linearized dynamics for

the solution and a subsequent update is required. Remember that v accounts for stochastic effects and it has a zero mean. The resulting linear approximation for the propagated state error becomes

$$\delta\bar{x}_{k+1} = \Phi\hat{x}_k + v^* \quad (10-28)$$

We use the error state transition matrix, Φ , to propagate the covariance matrix so we can see how good our state is at the new time. Gelb (1989:75–76) defines the covariance matrix as the expected value of the square of the errors in the state, so we can estimate it by

$$\hat{P}_k = E(\hat{\delta x}_k \hat{\delta x}_k^T)$$

The predicted value of the covariance is based on the predicted state error at the new time, t_{k+1} . This includes data only through t_k . Thus,

$$\bar{P}_{k+1} = E(\hat{\delta x}_{k+1} \hat{\delta x}_{k+1}^T)$$

Because we've already found the predicted value for the state error at the new time, we can substitute for it:

$$\bar{P}_{k+1} = E\{(\Phi\hat{x}_k + v)(\Phi\hat{x}_k + v)^T\}$$

$$\bar{P}_{k+1} = E(\Phi\hat{\delta x}_k \hat{\delta x}_k^T \Phi^T + \Phi\hat{\delta x}_k v^T + v\hat{\delta x}_k^T \Phi^T + vv^T)$$

Let's assume the state error and the error in the dynamics are uncorrelated. This means the expected value of their product is zero, and

$$\bar{P}_{k+1} = \Phi E(\hat{\delta x}_k \hat{\delta x}_k^T) \Phi^T + E(vv^T)$$

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 error in propagating the covariance matrix through time.[†]

$$\bar{P}_{k+1} = \Phi \hat{P}_k \Phi^T + Q \quad (10-29)$$

* Because v is random, we generally perform a deterministic propagation of the state. More detailed methods include process noise, but that's beyond our current scope.

† Gelb (1989:121–122) shows the development of the Matrix Riccati equation which propagates the covariance matrix for continuous systems. GPS signal processing uses continuous data.

$$\dot{\bar{P}}_{k+1} = F\hat{P}_k + \hat{P}_k F^T + \Phi Q \Phi^T$$

We've reached the first goal: propagating the state-error estimate, and covariance matrix to the new time. Eq. (10-28) and Eq. (10-29) perform this propagation, which results in *predictions* of the state error and covariance at the current time. These equations 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.

Updating the State Estimates

We can refine the predicted results and find generalized expressions by rewriting the result for sequential differential correction—[Eq. (10-19) and Eq. (10-20)]. Doing so gives us equations to correct our estimate for the new data at time t_{k+1} :

$$\begin{aligned}\delta\hat{\mathbf{x}}_{k+1} &= (\mathbf{H}^T \mathbf{W} \mathbf{H} + \bar{\mathbf{P}}_{k+1}^{-1})^{-1} (\mathbf{H}^T \mathbf{W} \tilde{\mathbf{b}} + \bar{\mathbf{P}}_{k+1}^{-1} \delta\bar{\mathbf{x}}_{k+1}) \\ \hat{\mathbf{P}}_{k+1} &= (\mathbf{H}^T \mathbf{W} \mathbf{H} + \bar{\mathbf{P}}_{k+1}^{-1})^{-1}\end{aligned}\quad (10-30)$$

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 (See Sec. 10.7 for formulations of \mathbf{H}). 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 also 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. (10-30)] directly, but examining the covariance update shows the need to invert a 6×6 matrix (for a 6×1 state vector). It's desirable 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. (10-20)] (notice I've deleted the subscripts for clarity, and all the data is at the current time):

$$\hat{\mathbf{P}} = (\mathbf{H}^T \mathbf{W} \mathbf{H} + \bar{\mathbf{P}}^{-1})^{-1}$$

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

$$\hat{\bar{P}}\hat{P}^{-1}\bar{P} = \hat{P}H^TWH\bar{P} + \hat{P}\bar{P}^{-1}\bar{P}$$

and simplify to produce

$$\bar{P} = \hat{P}H^TWH\bar{P} + \hat{P} \quad (10-31)$$

Although we won't use this expression now, notice you could rearrange Eq. (10-31) as $\bar{P} = \bar{P} - PH^TWH\bar{P}$. Post-multiplying Eq. (10-31) by H^TW to get

$$\bar{P}H^TW = \hat{P}H^TWH\bar{P}H^TW + \hat{P}H^TW$$

and factoring yields

$$\bar{P}H^TW = \hat{P}H^TW(\bar{P}H^TW + I)$$

The key now is to recognize the need to factor out a W matrix. Therefore, replace the identity matrix with $W^{-1}W$:

$$\bar{P}H^TW = \hat{P}H^TW(\bar{P}H^TW + W^{-1}W)$$

Factor a W out of $\bar{P}H^TW + W^{-1}W$, remembering the *order* of multiplication is important:

$$\bar{P}H^TW = \hat{P}H^TW(\bar{P}H^T + W^{-1})W$$

Post-multiply both sides by W^{-1} :

$$\bar{P}H^T = \hat{P}H^TW(\bar{P}H^T + W^{-1})$$

Post-multiply both sides by $(\bar{P}H^T + W^{-1})^{-1}$:

$$\bar{P}H^T(\bar{P}H^T + W^{-1})^{-1} = \hat{P}H^TW(\bar{P}H^T + W^{-1})(\bar{P}H^T + W^{-1})^{-1} = \hat{P}H^TW$$

Solve for the right-hand side:

$$\hat{P}H^TW = \bar{P}H^T(\bar{P}H^T + W^{-1})^{-1}$$

Now, define $K \equiv \hat{P}H^TW$ as the **Kalman gain matrix** (K), so that

$$K = \bar{P}H^T(\bar{P}H^T + W^{-1})^{-1} \quad (10-32)$$

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 PH^TW in Eq. (10-31) identified several steps back ($P = \bar{P} - PH^TWH\bar{P}$) results in

$$\hat{P} = \bar{P} - K\bar{H}\bar{P} = [I - KH]\bar{P} \quad (10-33)$$

This last formula is the basic covariance relation we want. It can exhibit numerical problems (a loss of positive definiteness) and 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):

$$\hat{\mathbf{P}} = (\mathbf{I} - \mathbf{K}\mathbf{H})\bar{\mathbf{P}}(\mathbf{I} - \mathbf{K}\mathbf{H})^T + \mathbf{K}\mathbf{R}\mathbf{K}^T \quad (10-34)$$

Thus, to guarantee positive definiteness requires the cost of additional computations.*

Although no inversions appear explicitly in the final answer [Eq. (10-33) or Eq. (10-34)], 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 the 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 ! Scalar division compared to even a 3×3 matrix inversion is a tremendous computational savings and one of the real benefits of using a Kalman filter. These observations mask an important issue—minimal stability or the performance of the filter when only a few observations are used. We'll discuss data requirements in Sec. 10.9.1.

Our final task is to update the state errors using the corrected value of $\hat{\mathbf{P}}$ and the predicted value of $\delta\bar{x}$. If we substitute the expression for $\hat{\mathbf{P}}$ into the state update in Eq. (10-30), we may write the estimate of the state errors as

$$\delta\hat{x} = \hat{\mathbf{P}}(\mathbf{H}^T W \tilde{\mathbf{b}} + \bar{\mathbf{P}}^{-1} \delta\bar{x})$$

Expanding gives us

$$\delta\hat{x} = \hat{\mathbf{P}}\mathbf{H}^T W \tilde{\mathbf{b}} + \hat{\mathbf{P}}\bar{\mathbf{P}}^{-1} \delta\bar{x}$$

Now, let's use the definition of the Kalman gain and Eq. (10-33) to produce

$$\delta\hat{x} = \mathbf{K}\tilde{\mathbf{b}} + (\mathbf{I} - \mathbf{K}\mathbf{H})\delta\bar{x}$$

Rearranging terms results in

$$\delta\hat{x} = \delta\bar{x} + \mathbf{K}(\tilde{\mathbf{b}} - \mathbf{H}\delta\bar{x})$$

Notice the term $\mathbf{H}\delta\bar{x}$ 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

* 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 “Convergence Issues” on page 816.) yields very stable algorithms, especially for onboard processing with limited computational power.

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 usually relies on batch least-squares techniques to find them. Our goal is to update the state error and covariance to the current time step-by-step, using each observation and either a linear or an “extended” approach. 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.

10.6.1 Kalman Filter (Linear System)

We've derived the basic equations for the Kalman filter (sometimes called a *discrete 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 error state transition matrix, Φ , to propagate the state and the state errors exactly because the equations of motion are linear. Although we don't show white process noise for the predicted state error because of a zero mean Gaussian assumption, we keep the second moment of the process noise, \mathbf{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. Algorithm 68 summarizes the Kalman filter for a linear system.

ALGORITHM 68: Kalman Filter—Linear System

$$\begin{aligned} & (\hat{\mathbf{X}}_k, \hat{\mathbf{P}}_k | z(\text{obs}), \mathbf{Q}, \mathbf{R} \Rightarrow \hat{\mathbf{X}}_{k+1}, \hat{\mathbf{P}}_{k+1}) \\ & \Phi_S(t_{k+1}, t_k) = \Phi(t_{k+1}, t_k) = \frac{\partial \hat{\mathbf{X}}_{k+1}}{\partial \hat{\mathbf{X}}_k} \end{aligned}$$

* Continuous systems use a variation known as the Kalman-Bucy method. These systems have a constant stream of data input for the problem. Although orbit determination often has discrete data observations, there are occasions when the data stream is continuous, or nearly continuous, such as GPS data. The process of converting radar signals into observations also uses Kalman filters to process the huge amount of data. See Gelb (1989:119–127) for additional information on this variation.

† 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.

$$\mathbf{H}_{k+1} = \frac{\partial z}{\partial \bar{\mathbf{X}}_{k+1}}$$

Prediction

$$\bar{\mathbf{X}}_{k+1} = \Phi_S \hat{\mathbf{X}}_k \quad \text{Predicted State}$$

$$\bar{\mathbf{P}}_{k+1} = \Phi \hat{\mathbf{P}}_k \Phi^T + \mathbf{Q} \quad \text{Predicted Error Covariance}$$

Update

$$\mathbf{K}_{k+1} = \bar{\mathbf{P}}_{k+1} \mathbf{H}_{k+1}^T \left[\mathbf{H}_{k+1} \bar{\mathbf{P}}_{k+1} \mathbf{H}_{k+1}^T + \mathbf{R} \right]^{-1} \quad \text{Kalman Gain}$$

$$\hat{\mathbf{X}}_{k+1} = \bar{\mathbf{X}}_{k+1} + \mathbf{K}_{k+1} [z - \mathbf{H}_{k+1} \bar{\mathbf{X}}_{k+1}] \quad \text{State Estimate}$$

$$\hat{\mathbf{P}}_{k+1} = [\mathbf{I} - \mathbf{K}_{k+1} \mathbf{H}_{k+1}] \bar{\mathbf{P}}_{k+1} \quad \text{Error Covariance Estimate}$$

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 \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 \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 10-6 illustrates a linear system.

▼ Example 10-6. 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, $\dot{\alpha}$ (see Fig. 10-8). The train engineer's job is to estimate position and velocity information from the radar's range measurements ($\sigma_{range} = 5$ m).

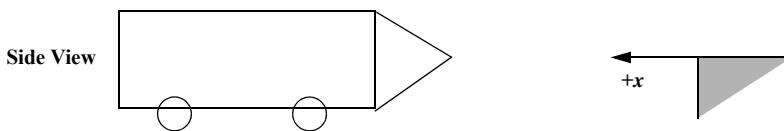


Figure 10-8. Schematic View of a High-speed Rail System for a Problem Applying a Linear Kalman Filter. This figure corresponds to Example 10-6, which looks at a high-speed rail system and tries to determine a train's position and velocity. We'll consider only one dimension.

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

$$range = 1 \text{ m}$$

$$velocity = 0.1 \text{ m/s}$$

$$acceleration = 0.1 \text{ m/s}^2$$

$Q(t)$ now becomes a constant matrix whose diagonal consists of the variances of our guessed dynamics and power-density-matrix statistics:

$$Q(t) = \begin{bmatrix} range^2 & 0 & 0 \\ 0 & velocity^2 & 0 \\ 0 & 0 & acceleration^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.01 & 0 \\ 0 & 0 & 0.01 \end{bmatrix}$$

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:

$$\begin{aligned}a(t) &= a_o \\v(t) &= a_o \Delta t + v_o \\r(t) &= \frac{1}{2} a_o \Delta t^2 + v_o \Delta t + r_o\end{aligned}$$

or in matrix form,

$$\Phi_S = \Phi = \begin{bmatrix} 1 & \Delta t & \frac{\Delta t^2}{2} \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix}$$

As the train approaches the platform, the engineer gets the following readings from his radar:

	$t_{observation}$ (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)

$$\begin{aligned}\bar{X}_1 &= \Phi_S \hat{X}_0 = \begin{bmatrix} 1 & \Delta t & \frac{\Delta t^2}{2} \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 600 \\ -45 \\ 1.5 \end{bmatrix} = \begin{bmatrix} 1 & 5 & 12.5 \\ 0 & 1 & 5 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 600 \\ -45 \\ 1.5 \end{bmatrix} = \begin{bmatrix} 393.75 \\ -37.50 \\ 1.50 \end{bmatrix} \\ \bar{P}_1 &= \Phi \hat{P}_0 \Phi^T + Q(t) = \begin{bmatrix} 1 & 5 & 12.5 \\ 0 & 1 & 5 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 225 & 0 & 0 \\ 0 & 25 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 5 & 1 & 0 \\ 12.5 & 5 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.01 & 0 \\ 0 & 0 & 0.01 \end{bmatrix} \\ \bar{P}_1 &= \begin{bmatrix} 1007.25 & 187.5 & 12.5 \\ 187.5 & 50.01 & 5 \\ 12.5 & 5 & 1.01 \end{bmatrix}\end{aligned}$$

The simplicity of the \mathbf{H} matrix allows us more simply to determine the Kalman gain. The quantity $\bar{\mathbf{P}}\mathbf{H}^T$ will always be the first column of the covariance matrix. In addition, $\mathbf{H}\bar{\mathbf{P}}\mathbf{H}^T$ 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

$$\mathbf{K}_1 = \bar{P}_1 \mathbf{H}^T [\mathbf{H}\bar{P}_1 \mathbf{H}^T + \mathbf{R}]^{-1} = \begin{bmatrix} 1007.25 \\ 187.50 \\ 12.50 \end{bmatrix} \frac{1}{1032.25} = \begin{bmatrix} 0.975 & 78 \\ 0.181 & 64 \\ 0.012 & 11 \end{bmatrix}$$

Again, the simplicity of the \mathbf{H} matrix permits some simplifications in determining the best estimates:

$$\hat{X}_1 = \bar{X}_1 + \mathbf{K}_1 [\bar{z} - \mathbf{H}\bar{X}_1] = \begin{bmatrix} 393.75 \\ -37.50 \\ 1.50 \end{bmatrix} + \begin{bmatrix} 0.975 & 78 \\ 0.181 & 64 \\ 0.012 & 11 \end{bmatrix} [385 - 393.75] = \begin{bmatrix} 385.211 & 92 \\ -39.089 & 37 \\ 1.394 & 04 \end{bmatrix}$$

$$\hat{P}_1 = [I - K_1 H] \bar{P}_1 = \begin{bmatrix} 0.024 & 22 & 0 & 0 \\ -0.181 & 64 & 1 & 0 \\ -0.012 & 11 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1007.25 & 187.5 & 12.5 \\ 187.5 & 50.01 & 5 \\ 12.5 & 5 & 1.01 \end{bmatrix} = \begin{bmatrix} 24.394 & 53 & 4.541 & 05 & 0.302 & 74 \\ 4.541 & 05 & 15.952 & 12 & 2.729 & 47 \\ 0.302 & 74 & 2.729 & 47 & 0.858 & 63 \end{bmatrix}$$

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)

$$\bar{X}_2 = \Phi_S \hat{X}_1 = \begin{bmatrix} 1 & 5 & 12.5 \\ 0 & 1 & 5 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 385.211 & 92 \\ -39.089 & 37 \\ 1.394 & 04 \end{bmatrix} = \begin{bmatrix} 207.190 & 60 \\ -32.119 & 16 \\ 1.394 & 04 \end{bmatrix}$$

$$\bar{P}_2 = \Phi \hat{P}_1 \Phi^T + Q(t) = \begin{bmatrix} 1 & 5 & 12.5 \\ 0 & 1 & 5 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 24.394 & 53 & 4.541 & 05 & 0.302 & 74 \\ 4.541 & 05 & 15.952 & 12 & 2.729 & 47 \\ 0.302 & 74 & 2.729 & 47 & 0.858 & 63 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 5 & 1 & 0 \\ 12.5 & 5 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.01 & 0 \\ 0 & 0 & 0.01 \end{bmatrix}$$

$$\bar{P}_2 = \begin{bmatrix} 952.522 & 241.835 & 24.683 \\ 241.835 & 64.723 & 7.023 \\ 24.683 & 7.023 & 0.867 \end{bmatrix}$$

$$K_2 = \bar{P}_2 H^T [\bar{H} \bar{P}_2 H^T + R]^{-1} = \begin{bmatrix} 952.522 \\ 241.835 \\ 24.638 \end{bmatrix} \frac{1}{977.522} = \begin{bmatrix} 0.974 & 42 \\ 0.247 & 39 \\ 0.025 & 25 \end{bmatrix}$$

$$\hat{X}_2 = \bar{X}_2 + K_2 [\bar{z} - H \bar{X}_2] = \begin{bmatrix} 207.190 & 60 \\ -32.119 & 16 \\ 1.394 & 04 \end{bmatrix} + \begin{bmatrix} 0.974 & 42 \\ 0.247 & 39 \\ 0.025 & 25 \end{bmatrix} [234 - 207.190 & 60] = \begin{bmatrix} 233.314 & 50 \\ -25.486 & 61 \\ 2.070 & 99 \end{bmatrix}$$

$$\hat{P}_2 = [I - K_2 H] \bar{P}_2 = \begin{bmatrix} 0.025 & 58 & 0 & 0 \\ -0.247 & 39 & 1 & 0 \\ -0.025 & 25 & 0 & 1 \end{bmatrix} \begin{bmatrix} 952.522 & 241.835 & 24.683 \\ 241.835 & 64.723 & 7.023 \\ 24.683 & 7.023 & 0.867 \end{bmatrix} = \begin{bmatrix} 24.360 & 6.185 & 0.631 \\ 6.185 & 4.894 & 0.916 \\ 0.631 & 0.916 & 0.245 \end{bmatrix}$$

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 10-3 summarizes the data for our problem.

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.

TABLE 10-3. Residuals and Standard-Deviation Values for Example 10-6: Kalman Filter (Linear System). This table shows the values for each state update of Example 10-6.

Observation	$z - \mathbf{H}\bar{\mathbf{X}}$ (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

10.6.2 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 69. For the first iteration, we assume the predicted state error, $\delta\mathbf{x}_k$, is zero. Note that I’ve omitted the process noise, \mathbf{v} , as an input. Typical practice assumes it has a Gaussian distribution with a zero mean. You can use more accurate techniques to determine Φ in the algorithm, but the following approximation [See Eq. (10-44)] will work for most linearized Kalman filters.

$$\Phi(t_{k+1}, t_k) = \mathbf{I} + \mathbf{F}(t_{k+1} - t_k) + \frac{1}{2!}\mathbf{F}^2(t_{k+1} - t_k)^2 + \frac{1}{3!}\mathbf{F}^3(t_{k+1} - t_k)^3 + \dots$$

Alternatively, we can find Φ by numerical integration [Eq. (10-27)] 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 69: Linearized Kalman Filter

$$(\hat{X}_o, \hat{P}_o, Q, R, z(\text{obs}) \Rightarrow \hat{X}_{k+1}, \hat{P}_{k+1})$$

Find one time $\dot{\hat{X}}_{t_o} = \frac{\partial \dot{\hat{X}}_{t_o}}{\partial \hat{X}_{t_o}}$, and the nominal trajectory using \hat{X}_o

Find at each observation time $H_{k+1} = \frac{\partial z}{\partial \hat{X}_{k+1}}$

Prediction

$$\Phi(t_{k+1}, t_o) \equiv I + F \Delta t + F^2 \frac{\Delta t^2}{2!} + \dots$$

PKEPLER($\hat{X}_o, t_{k+1} - t_o \Rightarrow \bar{X}_{k+1}$)

$$\text{or } \bar{X}(t_{k+1} | t_k) = \int_{t_k}^{t_{k+1}} \dot{\bar{X}}_k dt + \bar{X}_k \quad \text{Predicted State}$$

$$\delta \bar{x}_{k+1} = \Phi \delta \hat{x}_k \quad \text{Predicted State Error}$$

$$\bar{P}_{k+1} = \Phi \hat{P}_k \Phi^T + Q \quad \text{Predicted Error Covariance}$$

Update

$$\tilde{b}_{k+1} = z - H_{k+1} \bar{X}_{k+1}$$

$$K_{k+1} = \bar{P}_{k+1} H_{k+1}^T \left[H_{k+1} \bar{P}_{k+1} H_{k+1}^T + R \right]^{-1} \quad \text{Kalman Gain}$$

$$\delta \hat{x}_{k+1} = \delta \bar{x}_{k+1} + K_{k+1} [\tilde{b}_{k+1} - H_{k+1} \delta \bar{x}_{k+1}] \quad \text{State Error Estimate}$$

$$\hat{P}_{k+1} = \bar{P}_{k+1} - K_{k+1} H_{k+1} \bar{P}_{k+1} \quad \text{Error Covariance Estimate}$$

$$\hat{X}_{k+1} = \bar{X}_{k+1} + \delta \hat{x}_{k+1} \quad \text{State Estimate}$$

This is the *classic* use of a Kalman filter for nonlinear systems. Notice the addition of the state-error estimate, δx , 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 b matrix contains the residuals as in the problem using nonlinear differential correction (actual observations minus the

observations from the state $\bar{\mathbf{X}}$). Notice the use of $\mathbf{H}\delta\bar{\mathbf{x}}$ because we're now comparing it to the residual matrix, \mathbf{b} . We find the \mathbf{b} matrix as the measured observations minus the predicted value of the observations, $\mathbf{H}\bar{\mathbf{X}}$, 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 \mathbf{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. Earlier, we saw that we often propagate the state using a numerical integrator. This minimizes propagation error in the reference trajectory. However, analytical methods can also create the reference trajectory, and some of these methods can introduce large errors. For the orbit problem, because there are simplified methods to propagate the state to the new time and older computers were slow, some past solutions chose *not* to integrate the state but instead used for two-body motion (the Φ_s matrix) and let $\delta\mathbf{x}$ 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. This 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.

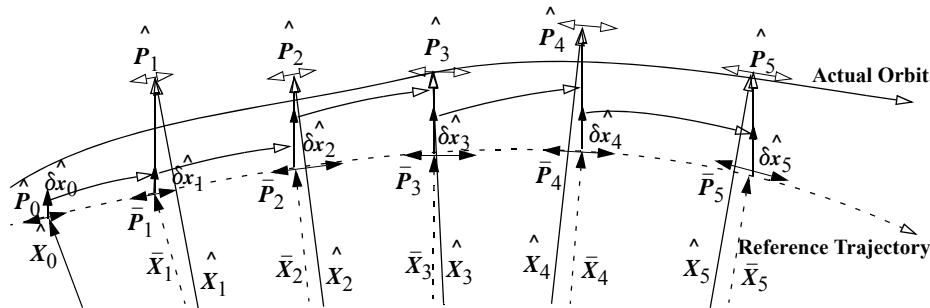


Figure 10-9. Representation of a Linearized Kalman Filter. The predicted, estimated, and state errors vary over time, so I've intentionally shown the values varying, not always converging, over time. Notice how we use only the state error estimates as inputs to determine subsequent state vectors. \mathbf{P} represents the uncertainty at each time.

10.6.3 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. 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 error is zero, $\delta\bar{x}_{k+1} = 0$, because there is no state error 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 influences covariance predictions. However, we must still include the second moment of the process noise, Q . I show these features in Algorithm 70. 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 influences computer runtime because multi-step methods have complex start-up techniques. I've shown *PKEPLER* to indicate one choice of propagator. A more 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 error state transition matrix is used only to update the state error and the covariance matrix. I've also shown the more rigorous method to find Φ using numerical integration [Eq. (10-27)]. There are other techniques discussed in Sec. 10.7. Finally, because F is no longer constant, we must recalculate it with each new observation. This adds additional processing at each step.

ALGORITHM 70: Extended Kalman Filter

$$(\hat{X}_k, \hat{P}_k, Q, R, z_{\text{obs}}) \Rightarrow (\hat{X}_{k+1}, \hat{P}_{k+1})$$

$$H_{k+1} = \frac{\partial z}{\partial \hat{X}_{k+1}}$$

Prediction

$$PKEPLER(\hat{X}_k, t_{k+1} - t_k \Rightarrow \bar{X}_{k+1})$$

$$\text{or } \bar{X}(t_{k+1} | t_k) = \int_{t_k}^{t_{k+1}} \dot{\hat{X}}_k dt + \hat{X}_k \quad \text{Predicted State}$$

$$F = \frac{\partial \dot{\hat{X}}_{t_{k+1}}}{\partial \hat{X}_{t_{k+1}}}, \quad \dot{\Phi}(t_{k+1}, t_k) = F(t) \Phi(t_{k+1}, t_k)$$

$$\delta\bar{x}_{k+1} = 0$$

Predicted State Error

$$\bar{P}_{k+1} = \Phi \hat{P}_k \Phi^T + Q$$

Predicted Error Covariance

Update

$$\tilde{\mathbf{b}}_{k+1} = \mathbf{z} - \mathbf{H}_{k+1} \bar{\mathbf{X}}_{k+1}$$

$$\mathbf{K}_{k+1} = \bar{\mathbf{P}}_{k+1} \mathbf{H}_{k+1}^T \left[\mathbf{H}_{k+1} \bar{\mathbf{P}}_{k+1} \mathbf{H}_{k+1}^T + \mathbf{R} \right]^{-1} \quad \text{Kalman Gain}$$

$$\hat{\delta \mathbf{x}}_{k+1} = \mathbf{K}_{k+1} \tilde{\mathbf{b}}_{k+1} \quad \text{State Error Estimate}$$

$$\hat{\mathbf{P}}_{k+1} = \bar{\mathbf{P}}_{k+1} - \mathbf{K}_{k+1} \mathbf{H}_{k+1} \bar{\mathbf{P}}_{k+1} \quad \text{Error Covariance Estimate}$$

$$\hat{\bar{\mathbf{X}}}_{k+1} = \bar{\mathbf{X}}_{k+1} + \hat{\delta \mathbf{x}}_{k+1} \quad \text{State Estimate}$$

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 error, $\hat{\delta \mathbf{x}}$, 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 \mathbf{b} 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 10-10 graphically depicts the state and state errors over time.

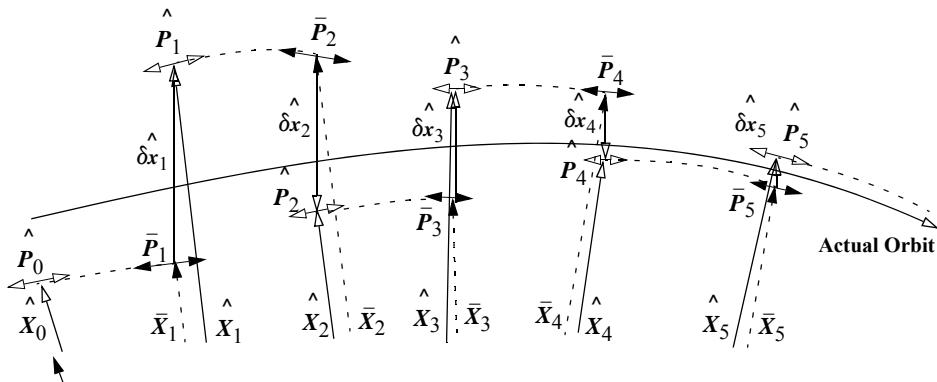


Figure 10-10. Representation of an Extended Kalman Filter. In this ideal representation, notice how all the predicted, estimated, and state updates converge over time. Normally, filter convergence is not guaranteed.

We can dynamically reject observations as the filter runs by examining the measurement matrix. Letting c be the number of standard deviations to be allowed, if

$|\tilde{b}_{k+1}| < c \sqrt{\mathbf{H}_{k+1} \bar{\mathbf{P}}_{k+1} \mathbf{H}_{k+1}^T + \mathbf{R}}$, the observations are processed, otherwise, they are rejected.

If a Kalman 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 could be an EKF using a white-noise hypothesis for process noise when the actual 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, or outside the interval of convergence (in which case iterated least squares will also diverge). A filter may not converge if the measurement parameters are 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.

Process Noise

Determining the process noise evokes considerable discussion. Successful approaches are implemented in Analytical Graphics Inc. **Orbit Determination Toolkit** (ODTK) and its predecessor, **Real Time Orbit Determination** (RTOD). Both methods were developed by Jim Wright and Richard Hujasak (Wright, 1981 and 1994). ODTK and RTOD are unique and challenging approaches that strive to define the process noise as a function of the system's physical processes. Wright recognized that process noise in the orbit problem is a real physical phenomenon that arises from uncertainty in the force models. Using gravity is an easy way to derive a generalized process-noise function. The errors can come from truncating the model (which we do to save computer time) or in 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 use a 4×4 gravity field, the process noise error is the contribution of the gravity field *above* 4×4 —called the **error of omission**. If we limit the field to say 50×50 , we can calculate the process noise exactly from the omitted coefficients (although this effectively diminishes the original reason to truncate!). We know that gravity models are found from least squares solutions. The "state" contains the coefficients, and the "covariance" represents the uncertainty of those coefficients. This uncertainty is called an **error of commission** and it permits an exact determination of the process noise (for gravity). For other force models, the process noise is assembled similarly with varying degrees of difficulty.

Wright et al. (2008) discuss the mathematical background and development, as well as introducing a computational process that is summarized here. The process relies on finding constant terms for an orbit under consideration, and dynamic terms as the propagation proceeds. Wright et al. (2008) also discuss the need to graph results and evaluate an average from those graphs. Polynomial and splining functions are commonly used.

Several static values are found first. The gravity error covariance is calculated for the various gravity degrees and stored, using an ideal circular orbital radius for the orbit under consideration. This is a function of commission and omission [first and second terms respectively in Eq. (10-35)]. The errors of commission rely on the uncertainty of the gravitational coefficients (standard deviations, $\sigma_c(\bar{C}_{t,m})$ and $\sigma_s(\bar{S}_{t,m})$ provided with the coefficients). The error of omission is found using the gravitational coefficients for the omitted terms. With very large modern gravity fields (e.g., 2159×2159), it may be desirable to limit the number of terms. Here, N represents the truncation point (e.g., 50), and N_{max} is an upper limit (e.g., 70) for very large gravity fields. Each gravity error covariance is stored individually, by degree, up to N_{max} .

$$\sigma_n^2 = \frac{\mu(l-1)^2}{R_\oplus^4} \sum_{l=0}^N (\sigma_c(\bar{C}_{t,m})^2 + \sigma_s(\bar{S}_{t,m})^2) + \sum_{l=N+1}^{N_{max}} (\bar{C}_{t,m}^2 + \bar{S}_{t,m}^2) \quad (10-35)$$

Wright (1981) shows the scalar auto covariance functions, originally developed by Kaula. Note the central Earth angle (Λ) from the development of the gravitational potential and the use of radial, along track, cross track (RSW) components.

$$\begin{aligned} \sigma_R^2(\Lambda) &= \sum_{l=0}^{\infty} \left(\frac{l+1}{l-1} \right)^2 \left(\frac{R_\oplus}{r} \right)^{2l+4} P_{l,0}[\sin(\Lambda)] \sigma_n^2 \\ \sigma_S^2(\Lambda) &= \frac{1}{2} \sum_{l=0}^{\infty} \frac{l(l+1)}{(l-1)^2} \left(\frac{R_\oplus}{r} \right)^{2l+4} \left[P_{l,0}[\cos(\Lambda)] - \frac{P_{l,2}[\cos(\Lambda)]}{l(l+1)} \right] \sigma_n^2 \\ \sigma_W^2(\Lambda) &= \frac{1}{2} \sum_{l=0}^{\infty} \frac{l(l+1)}{(l-1)^2} \left(\frac{R_\oplus}{r} \right)^{2l+4} \left[P_{l-1,0}[\cos(\Lambda)] + \frac{P_{l-1,2}[\cos(\Lambda)]}{l(l+1)} \right] \sigma_n^2 \end{aligned}$$

Each component in the equations above are evaluated at $\Lambda = 0$. The defined auto correlations are found as follows (only the radial axis is shown).

$$p_R(\Lambda) = \frac{\sigma_R^2(\Lambda)}{\sigma_R^2(0)}$$

The auto correlations are then evaluated as integrals (only the radial axis is shown).

$$I_R(\Lambda) = 2 \int_0^\Lambda p_R(M) dM$$

Convert to proper units by multiplying by the period and dividing by 360° and get T_R , T_S , T_W . These functions are graphed for later determination with polynomials or splines that can quickly represent the values.

As the filter processes data, two matrices are formed to populate the process noise matrix. $\mathbf{I}\mathbf{1}$ is found using the reconstructed T_R , T_S , T_W diagonal values (this is the middle bracketed integral below), and \mathbf{Q} is found from the following equation. Notice here that $\mathbf{H}(t_{k+1}, \tau_k) = \Phi(t_{k+1}, \tau_k)\mathbf{G}(t)$.

$$\mathbf{Q} = \int_{t_k}^{t_{k+1}} \mathbf{H}(t_{k+1}, \tau_k) R(0, t) \times \left[\int_{\gamma}^{\bar{\gamma}} \rho \gamma d\gamma \right] \mathbf{H}(t_{k+1}, \tau_k)^T d\tau$$

10.6.4 Unscented Kalman Filter (UKF)

The Kalman Filter enjoys success in many applications. However, situations arise where it is desirable to implement the propagation of the covariance in a slightly different way. This is particularly true for systems where the process noise is insufficient to properly model the dynamic nature of a problem. I will only sketch out the approach, but for the full theory and for additional information, see Tapley et al. (2004) and Simon (2006).

In its simplest form, a UKF simply takes the state and covariance and propagates the state (usually numerically), but propagates the covariance using the full force models and without the linearization present in Least Squares and Kalman filter applications. This is true for times when observations are being processed, and for prediction into the future. A potential application is the propagation of covariance and how realistically it encompasses future positions of the satellite. Instead of relying on the numerical integration of Eq. (10-27) and possibly the inclusion of process noise [Eq. (10-29)], the UKF moves the 12 critical points of the covariance forward in time individually, using all force models.

Two short routines are very useful to implement a UKF. The first is determining the 12 points to propagate from the covariance, and the second is reassembling the covariance at future times once each point has been propagated. Remember that the covariance is actually a hyper ellipse in 6 dimensions (considering position and velocity). The eigenvectors give the major and minor axes of this ellipse in 6 dimensions and the eigenvalues are the magnitude of the eigenvectors. Beware that programs like Matlab often sort the eigenvectors so there may not be a direct correspondence to the position and velocity components. In Eq. (10-36), [**covariance**] is the epoch covariance under consideration, the [**s**] array is a 6×6 array with a 6 dimensioned state of position and velocity vectors, and [**sigmapts**] is a 6×12 array, encompassing the 12 points necessary to describe the ellipse. Using Cholesky decomposition (See Tapley et al. (2004:287-288) for details), we first find the square root of the state size times the Cholesky decomposition of the covariance. Simon (2006:449) indicates the need for the square root term. Note that Matlab code is used to better show the indices and loops.

```
% compute matrix square root of state size (6) times covariance (6x6)
s = sqrt(6) * chol(covariance, 'lower');
% loop through each element of the state (position reci, velocity veci)
for i = 1:6
% increment this by 2
jj = (i-1)*2 + 1;
% transpose () to get proper vector orientation
sigmapts(1:3,jj) = reci + s(1:3,i)';
sigmapts(4:6,jj) = veci + s(4:6,i)';
% ---- find negative direction vectors
% transpose () to get proper vector orientation
sigmapts(1:3,jj+1) = reci - s(1:3,i)';
sigmapts(4:6,jj+1) = veci - s(4:6,i)';
end; % for i
```

(10-36)

To reassemble the 12 position and velocity vectors at any time into a single covariance, use the following. The last equation ensures perfect symmetry in case there are small round-off errors. (Matlab code is used to better show the indices and loops.)

```
% find the mean of 12 sigma points in a 6 dimensioned state
for i = 1 : 12
    yu = yu + pts(1 : 6, i);
end; % for i
yu = yu / (12); % (6x12)

% find the covariance (6x6)
for i = 1 : 12
    y(1:6, i) = pts(1:6, i) - yu(1:6, 1);
end; % for i
tmp = (y * y') / 12;
% ensures perfect symmetry
cov = (tmp + tmp') / 2;
```

(10-37)

10.6.5 Summary of Kalman Filters

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 continuing 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 Labora-

tory uses a filter to obtain very precise ephemerides for GPS satellites and for the TOPEX/Jason satellites.

Filters have a persistent reputation of being hard to tune, especially in setting the Q parameters for process noise. Fisher (1998) discusses the methods we can use to alleviate this problem. There are two basic methods, adaptive and non-adaptive. *Adaptive techniques* modify the structure of the filter to accommodate unknown parameters. There are two ways to implement this technique. *Structural* techniques estimate the unmodeled accelerations as a solve-for parameter. This can be computationally intensive. *Statistical* techniques include all the known errors into a mathematical model of the process noise. This technique has been successfully used in ODTK and RTOD. *Non-adaptive techniques* artificially adjust the process noise to obtain the correct performance. This approach is complex, and changes for different orbits, data, and other parameters.

The Kalman filter and Extended Kalman filters have good and bad points when compared to least squares and sequential least squares. We generically refer to Kalman filters here as having both a filter and a smoother. The following are advantages of the filter over least squares and sequential least squares:

1. It allows solving for measurement biases simultaneously in the state space, which often better 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. Insufficient data affects all orbit determination methods.
2. Filters can model time-varying coefficients of drag and solar-radiation pressure. Although a least-squares technique can solve for an average parameter during the fit span, the approach is at best only piecewise continuous. This assumes we have enough data to arrive at a solution.
3. It's not as limited by the accuracy of the force models in the fit span (depending on the data frequency) and can therefore use less accurate force models. However, subsequent propagation should use the same force models, so there are sometimes competing goals.
4. It estimates the current state and covariance matrix throughout the span of observations, which can be beneficial.
5. It may require less memory storage and processor time, but that depends on the computer software.
6. It's a very fast, recursive process that is ideally suited for real-time control of a system because it doesn't require iteration (Viking Lander, Apollo Lunar Landers, Missiles). In view of the interest and success of the Global Positioning System, these systems are being integrated into military tactics. This integration often needs orbit determination in near real-time, which requires a filter.

But 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*** which we can overcome in several ways. A quick solution is to increase \mathbf{Q} . The classical covariance recursion is numerically unstable in many applications, which only adds to this problem.

2. The filter isn't *self-starting*; to begin, it requires an initial estimate of the state, \mathbf{X} , covariance, \mathbf{P} , plus the process noise, \mathbf{v} , power-density matrix, \mathbf{Q} , and measurement error, \mathbf{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 \mathbf{Q} matrix from the physics of the problem. This can be very complex depending on the fidelity of our dynamics model.
4. ***Divergence*** is related to smugness (see #1 above), but it results from the filter "walking off" from the solution. If our data is very noisy, we may not be able to converge on a solution because noisy data has a large measurement error matrix, \mathbf{R} , which causes us to ignore new data if the filter is smug. The data can also cause the solution to degrade as more information is processed. Some centers invoke the Joseph algorithm, fading-memory filters, or square-root formulas. However, if the data is very poor, both a well-designed filter *and* a least-squares technique will fail.

We've discussed some 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.

10.7 Calculating the Matrices

The least squares equations are surprisingly simple, yet complex in their implementation. Eq. (10-16) is the basis for the majority of the effort in building a differential correction program. In general, we can use analytical and numerical techniques to calculate most of the matrices in Eq. (10-16). I'll discuss each and give examples.

There are numerous approaches to perform differential correction and a selection is ultimately tied to the calculation of the partial derivative matrix, \mathbf{A} . Remember that during the differential correction process, we can reduce the force models for the variational equations. The main effect is in the number of iterations. However, when propagating the state and covariance for a particular application, one should use rigorous force models on both the state and covariance to preserve as much accuracy from the initial solution. Table 10-4 shows several different options.

TABLE 10-4. Differential Correction Options. This table lists four options for employing least squares techniques. The complex method numerically integrates the equation for Φ . Note that we can also approximate the observation (H), partial derivative (F), and error state transition (Φ) matrices by finite differencing.

Operation	Simple	Analytical	Constant System	Complex
$\hat{X} = \hat{X}_o$		Kepler, PKepler, etc.		
H	N/A	$H = \frac{\partial \text{observations}}{\partial \hat{X}}$ (analytic)		
Φ	N/A	$\dot{\vec{r}}(\tau) = \dot{\vec{r}}_o(t) + \ddot{\vec{r}}_o(t)\tau + \frac{\dddot{\vec{r}}_o(t)\tau^2}{2!} + \frac{\dots}{3!} + \dots$ $\Phi = \frac{\partial \dot{\vec{X}}}{\partial \hat{X}_o} = \begin{bmatrix} \frac{\partial \dot{\vec{r}}}{\partial \dot{\vec{r}}_o} & \frac{\partial \dot{\vec{r}}}{\partial \dot{\vec{v}}_o} \\ \frac{\partial \dot{\vec{v}}}{\partial \dot{\vec{r}}_o} & \frac{\partial \dot{\vec{v}}}{\partial \dot{\vec{v}}_o} \end{bmatrix}$	$F = \frac{\partial \dot{\vec{X}}}{\partial \hat{X}} = \begin{bmatrix} \frac{\partial \dot{\vec{r}}}{\partial r_I} & \frac{\partial \dot{\vec{r}}}{\partial r_J} & \frac{\partial \dot{\vec{r}}}{\partial r_K} & \frac{\partial \dot{\vec{v}}}{\partial v_I} & \frac{\partial \dot{\vec{v}}}{\partial v_J} & \frac{\partial \dot{\vec{v}}}{\partial v_K} \end{bmatrix}_{ITRF}$ $\Phi = \left\{ \begin{array}{l} I + F\Delta t + \frac{1}{2!}F^2\Delta t^2 \\ \quad + \frac{1}{3!}F^3\Delta t^3 + \dots \end{array} \right\}$	$\frac{\partial v_J}{\partial r_I} \dots$ \vdots $\frac{\partial a_K}{\partial r_I} \frac{\partial a_K}{\partial r_J} \frac{\partial a_K}{\partial r_K} \frac{\partial a_K}{\partial v_I} \frac{\partial a_K}{\partial v_J} \frac{\partial a_K}{\partial v_K}$ $\Phi(t, t_o) = F(t)\Phi(t, t_o)$ $\Phi(t_o, t_o) = I$
A	Finite Differencing	$A = H\Phi$		

Calculating the Observation Partial Derivatives, H

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. We need the partial derivative of the observations with respect to the state. The solution lies in using the chain rule.

$$\mathbf{H} = \frac{\partial \text{obs}}{\partial \vec{r}_{ECEF}} = \frac{\partial \text{obs}}{\partial \vec{\rho}_{SEZ}} \frac{\partial \vec{\rho}_{SEZ}}{\partial \vec{\rho}_{ECEF}} \frac{\partial \vec{\rho}_{ECEF}}{\partial \vec{r}_{ECEF}}$$

Recalling the vector relation of the range vector, $\vec{\rho}_{ECEF} = \vec{r}_{ECEF} - \vec{r}_{siteECEF}$, the final partial derivative is simply the identity matrix.

The derivative of a coordinate system with respect to another depends on the transformation between the systems.

$$\begin{aligned}\vec{\rho}_{SEZ} &= \left[\begin{smallmatrix} SEZ \\ ECEF \end{smallmatrix} \right] \vec{\rho}_{ECEF} \\ \frac{\partial \vec{\rho}_{SEZ}}{\partial \vec{\rho}_{ECEF}} &= \frac{\partial}{\partial \vec{\rho}_{ECEF}} \left[\begin{smallmatrix} SEZ \\ ECEF \end{smallmatrix} \right] \vec{\rho}_{ECEF} + \left[\begin{smallmatrix} SEZ \\ ECEF \end{smallmatrix} \right] \frac{\partial \vec{\rho}_{ECEF}}{\partial \vec{\rho}_{ECEF}}\end{aligned}$$

Notice that the first partial is zero, and the second is 1.0. Thus, the partial derivative of a vector in one frame with respect to another frame is simply the transformation between them. In this example, we use the transpose of Eq. (3-28).

$$\frac{\partial \vec{\rho}_{SEZ}}{\partial \vec{\rho}_{ECEF}} = \left[\begin{smallmatrix} SEZ \\ ECEF \end{smallmatrix} \right] = \begin{bmatrix} \sin(\phi_{gd})\cos(\lambda) & \sin(\phi_{gd})\sin(\lambda) & -\cos(\phi_{gd}) \\ -\sin(\lambda) & \cos(\lambda) & 0 \\ \cos(\phi_{gd})\cos(\lambda) & \cos(\phi_{gd})\sin(\lambda) & \sin(\phi_{gd}) \end{bmatrix}$$

Let's consider observations that consist of range, azimuth, and elevation. In the SEZ coordinate system, each observation is found from the slant range vector and the equations in Algorithm 27 (**RAZEL**). I've used TAN and SIN to preserve the quadrant information.

$$\begin{aligned}\rho &= \sqrt{\rho_S^2 + \rho_E^2 + \rho_Z^2} \\ \beta &= \text{ATAN}\left(\frac{-\rho_S}{\rho_E}\right) \\ el &= \text{ASIN}\left(\frac{\rho_Z}{\sqrt{\rho_S^2 + \rho_E^2 + \rho_Z^2}}\right)\end{aligned}$$

The partial derivatives are found directly using standard mathematical formulae. Combining these results, we find the observation partial derivative matrix.

$$\frac{\partial \text{obs}}{\partial \dot{\vec{r}}_{ECEF}} = \begin{bmatrix} \frac{\rho_S}{\rho} & \frac{\rho_E}{\rho} & \frac{\rho_Z}{\rho} \\ \frac{\rho_E}{\left(\frac{\rho_E^2}{\rho_S^2} + 1\right)\rho_S^2} & -\frac{1}{\left(\frac{\rho_E^2}{\rho_S^2} + 1\right)\rho_S} & 0 \\ -\frac{\rho_S\rho_Z}{\rho^3 \sqrt{-\frac{\rho_Z^2}{\rho_S^2 + \rho_E^2 + \rho_Z^2} + 1}} & -\frac{\rho_E\rho_Z}{\rho^3 \sqrt{-\frac{\rho_Z^2}{\rho_S^2 + \rho_E^2 + \rho_Z^2} + 1}} & \frac{\rho - \rho_Z^2}{\rho^3 \sqrt{-\frac{\rho_Z^2}{\rho^2} + 1}} \end{bmatrix} \times \begin{bmatrix} \text{SEZ} \\ \text{ECEF} \end{bmatrix}$$

A similar method can be used for other observation types. In general, partial derivatives for other observation types can be 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), Chioma (1997), and Shaver (1980) for detailed equations.

Calculating the Partial Derivative Matrix, F

The F matrix is a critical component of the overall A matrix evaluation. F is very complex for detailed force models. Its primary use is to find the error state transition matrix and we may use analytical, numerical, or finite difference techniques. We introduced F in Eq. (10-25).*

$$F(t) = \frac{\partial f(\vec{X})}{\partial \vec{X}}$$

We can use two-body dynamics, but we sometimes wish to incorporate more dynamics into the error state transition matrix, whether it is 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. Let's assume we're given the state at the current time \hat{X} , and the epoch state \hat{X}_o , possibly augmented by dynamical parameters (c_D , $C_{t,m}$, $S_{t,m}$, for example), which define

$$\hat{X}_o = (\dot{\vec{r}}_o, \dot{\vec{v}}_o, c_D, C_{t,m}, S_{t,m})$$

The equations of motion, $\dot{\hat{X}} = f(\dot{\vec{r}}, \dot{\vec{v}}, c_D, C_{t,m}, S_{t,m})$, depend on the current state (in which the position and velocity vectors 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

* As we mentioned on page 82, there is a difference between a state transition matrix that moves the state (Φ_S) and the state errors (Φ). The F matrix in this section is exclusively for the latter.

$$\frac{\partial \dot{\hat{X}}}{\partial \hat{X}_o} = \frac{\partial}{\partial \hat{X}_o} f(\hat{r}, \hat{v}, c_D, C_{t,m}, S_{t,m}) \quad (10-38)$$

Rewriting the left-hand side using operator notation and interchanging the order of the operators yields

$$\frac{\partial \dot{\hat{X}}}{\partial \hat{X}_o} = \frac{\partial}{\partial \hat{X}_o} \left[\frac{d \hat{X}}{dt} \right] = \frac{d}{dt} \left[\frac{\partial \hat{X}}{\partial \hat{X}_o} \right] \quad (10-39)$$

Thus, Eq. (10-39) 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 $\partial \hat{X} / \partial \hat{X}_o$. Combining Eq. (10-38) and Eq. (10-39) yields a first-order system of differential equations for the *variational equations*:

$$\frac{d}{dt} \left[\frac{\partial \hat{X}}{\partial \hat{X}_o} \right] = \frac{\partial}{\partial \hat{X}_o} \left(\frac{d \hat{X}}{dt} \right) = \frac{\partial f(\hat{r}, \hat{v}, c_D, C_{t,m}, S_{t,m})}{\partial \hat{X}} \frac{\partial \hat{X}}{\partial \hat{X}_o} = \mathbf{F} \Phi \quad (10-40)$$

This is an $n \times n$ matrix of first-order differential equations where n is the number of state and solve-for parameters—an explicit form of Eq. (10-27). The equations are numerically integrated, (usually) with the same algorithms used for integrating the equations of motion, and chained with the observation partial derivatives to find the desired partial derivatives for Eq. (10-16). Eq. (10-40) implies [at least] six times the number of equations, which typically means six times the computational cost. To reduce the computational burden, the force models used to integrate the variational equations are often reduced from that used to integrate the equations of motion. Be aware that this is only valid *within* the differential correction process. As an example, precision orbit determination using a 21×21 gravity model might use a 2×2 or 4×4 field with 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. However, to propagate the converged state, you should use the same fidelity models (21×21 from above). This ensures the accuracy of the answers at the future time, but it doesn't restrict your options from a computer software standpoint. The use of identical computer code, especially for numerical operations, is almost never warranted unless undocumented empirical techniques are used in a program. You should avoid these situations!

For analytical evaluations, 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. The state vectors are taken in a body-fixed frame (ECEF) because we apply accelerations in this frame. The state and derivative of the state are

$$\hat{\dot{X}}_{\text{2-body}} = \begin{bmatrix} \dot{\vec{r}} \\ \dot{\vec{v}} \end{bmatrix}_{\text{ECEF}} \quad \text{and} \quad \dot{\hat{X}}_{\text{2-body}} = \begin{bmatrix} \dot{\vec{v}} \\ \dot{\vec{a}} \end{bmatrix} = \begin{bmatrix} \dot{\vec{v}} \\ -\frac{\mu \dot{\vec{r}}}{r^3} \end{bmatrix}_{\text{ECEF}}$$

In general, by Eq. (10-40) we have

$$\frac{\partial}{\partial \hat{\dot{X}}_o} \left(\frac{d\hat{X}}{dt} \right) = \frac{\partial \dot{\hat{X}}}{\partial \hat{X}} \frac{\partial \hat{X}}{\partial \hat{\dot{X}}_o} = \mathbf{F} \Phi$$

$$\mathbf{F} \Phi = \begin{bmatrix} \frac{\partial v_I}{\partial r_I} & \frac{\partial v_I}{\partial r_J} & \frac{\partial v_I}{\partial r_K} & \frac{\partial v_I}{\partial v_I} & \frac{\partial v_I}{\partial v_J} & \frac{\partial v_I}{\partial v_K} \\ \frac{\partial v_J}{\partial r_I} & \cdots & & & & \\ \vdots & & & & & \vdots \\ \frac{\partial a_K}{\partial r_I} & \frac{\partial a_K}{\partial r_J} & \frac{\partial a_K}{\partial r_K} & \frac{\partial a_K}{\partial v_I} & \frac{\partial a_K}{\partial v_J} & \frac{\partial a_K}{\partial v_K} \end{bmatrix}_{\text{ECEF}} \frac{\partial \dot{\hat{X}}}{\partial \hat{\dot{X}}_o}$$

The two-body contribution is obtained from the partial derivatives of the two-body acceleration, which depend only on the position, r .

$$\frac{\dot{\hat{X}}_{\text{2-body}}}{\partial \hat{X}} = \mathbf{F} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ -\frac{\mu}{r^3} + \frac{3\mu r_I^2}{r^5} & \frac{3\mu r_I r_J}{r^5} & \frac{3\mu r_I r_K}{r^5} & 0 & 0 & 0 \\ \frac{3\mu r_I r_J}{r^5} & -\frac{\mu}{r^3} + \frac{3\mu r_J^2}{r^5} & \frac{3\mu r_J r_K}{r^5} & 0 & 0 & 0 \\ \frac{3\mu r_I r_K}{r^5} & \frac{3\mu r_J r_K}{r^5} & -\frac{\mu}{r^3} + \frac{3\mu r_K^2}{r^5} & 0 & 0 & 0 \end{bmatrix}_{\text{ECEF}} \quad (10-41)$$

Notice that the equations for the \mathbf{F} matrix are exact (for a given acceleration), unlike the Taylor series representation in the analytic formulation of $\Phi_{\text{2-body}}$ on page 810.

Let's look at an example using more complex perturbations. We find the equations of motion from Algorithm 64, but for now, let's use the state notation.

$$\frac{d\hat{X}}{dt} = \dot{\hat{X}}_{\text{2-body}} + \dot{\hat{X}}_{\text{nonspherical}} + \dot{\hat{X}}_{\text{drag}} + \dot{\hat{X}}_{\text{3-body}} + \dot{\hat{X}}_{\text{SR}} + \dot{\hat{X}}_{\text{other}}$$

Then, the partial derivatives with respect to the epoch state are

$$\frac{\partial}{\partial \hat{\mathbf{X}}_o} \left(\frac{d\hat{\mathbf{X}}}{dt} \right) = \left\{ \frac{\partial \dot{\hat{\mathbf{X}}}_{\text{2-body}}}{\partial \hat{\mathbf{X}}} + \frac{\partial \dot{\hat{\mathbf{X}}}_{\text{nonspherical}}}{\partial \hat{\mathbf{X}}} + \frac{\partial \dot{\hat{\mathbf{X}}}_{\text{drag}}}{\partial \hat{\mathbf{X}}} + \dots \right\} \frac{\partial \hat{\mathbf{X}}}{\partial \hat{\mathbf{X}}_o} = \mathbf{F} \frac{\partial \hat{\mathbf{X}}}{\partial \hat{\mathbf{X}}_o}$$

The equations of motion due to the nonspherical Earth also 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,

$$\frac{\partial \dot{\hat{\mathbf{X}}}_{\text{nonspherical}}}{\partial \hat{\mathbf{X}}} = \mathbf{F} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ \frac{\partial a_I}{\partial r_I} & \frac{\partial a_I}{\partial r_J} & \frac{\partial a_I}{\partial r_K} & 0 & 0 & 0 \\ \frac{\partial a_J}{\partial r_I} & \frac{\partial a_J}{\partial r_J} & \frac{\partial a_J}{\partial r_K} & 0 & 0 & 0 \\ \frac{\partial a_K}{\partial r_I} & \frac{\partial a_K}{\partial r_J} & \frac{\partial a_K}{\partial r_K} & 0 & 0 & 0 \end{bmatrix}_{\text{ECEF}}$$

Although this equation is an exact formulation, when we use complete gravitational models, drag models, etc., the implementation can get complex. Let's examine what's involved with forming the partial derivatives of the gravitational potential. Remember that the acceleration (\vec{a}) is the gradient of the potential function, so in spherical coordinates of the satellite (r, ϕ_{gc}, λ),

$$\vec{a} = \frac{\partial U}{\partial \vec{r}} = \frac{\partial U}{\partial r} \frac{\partial r}{\partial \vec{r}} + \frac{\partial U}{\partial \lambda} \frac{\partial \lambda}{\partial \vec{r}} + \frac{\partial U}{\partial \phi_{gc}} \frac{\partial \phi_{gc}}{\partial \vec{r}}$$

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. (8-19). Remember that we evaluate the acceleration in a body-fixed (ECEF) frame (U is the aspherical potential *excluding* the two-body term):

$$\begin{aligned} \frac{\partial \vec{a}}{\partial \vec{r}} &= \frac{\partial}{\partial \vec{r}} \left(\frac{\partial U}{\partial r} \right) \frac{\partial r}{\partial \vec{r}} + \frac{\partial}{\partial \vec{r}} \left(\frac{\partial U}{\partial \phi_{gc}} \right) \frac{\partial \phi_{gc}}{\partial \vec{r}} + \frac{\partial}{\partial \vec{r}} \left(\frac{\partial U}{\partial \lambda} \right) \frac{\partial \lambda}{\partial \vec{r}} \\ &+ \frac{\partial U}{\partial r} \frac{\partial^2 r}{\partial \vec{r}^2} + \frac{\partial U}{\partial \phi_{gc}} \frac{\partial^2 \phi_{gc}}{\partial \vec{r}^2} + \frac{\partial U}{\partial \lambda} \frac{\partial^2 \lambda}{\partial \vec{r}^2} \end{aligned}$$

In matrix notation, the first set of partial derivatives are

$$\frac{\partial}{\partial \vec{r}} \begin{bmatrix} \frac{\partial U}{\partial r} \\ \frac{\partial U}{\partial \phi_{gc}} \\ \frac{\partial U}{\partial \lambda} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 U}{\partial r^2} & \frac{\partial^2 U}{\partial r \partial \phi_{gc}} & \frac{\partial^2 U}{\partial r \partial \lambda} \\ \frac{\partial^2 U}{\partial \phi_{gc} \partial r} & \frac{\partial^2 U}{\partial^2 \phi_{gc}} & \frac{\partial^2 U}{\partial \phi_{gc} \partial \lambda} \\ \frac{\partial^2 U}{\partial \lambda \partial r} & \frac{\partial^2 U}{\partial \lambda \partial \phi_{gc}} & \frac{\partial^2 U}{\partial \lambda^2} \end{bmatrix} \begin{bmatrix} \frac{\partial r}{\partial \vec{r}} \\ \frac{\partial \phi_{gc}}{\partial \vec{r}} \\ \frac{\partial \lambda}{\partial \vec{r}} \end{bmatrix} \quad (10-42)$$

Differentiate Eq. (8-25) (again excluding the two-body term) to find the partial derivatives of the matrix:

$$\begin{aligned} \frac{\partial^2 U}{\partial r^2} &= \frac{\mu}{r^3} \sum_{l=2}^{\infty} \left(\frac{R_{\oplus}}{r}\right)^l (\ell+2)(\ell+1) \sum_{m=0}^{\ell} P_{\ell,m} [\sin(\phi_{gc})] \{ C_{\ell,m} \cos(m\lambda) + S_{\ell,m} \sin(m\lambda) \} \\ \frac{\partial^2 U}{\partial \phi_{gc} \partial r} &= \frac{\partial^2 U}{\partial \phi_{gc} \partial r} \\ &= -\frac{\mu}{r^2} \sum_{l=2}^{\infty} \left(\frac{R_{\oplus}}{r}\right)^l (\ell+1) \sum_{m=0}^{\ell} \{ P_{\ell,m+1} [\sin(\phi_{gc})] - m \tan(\phi_{gc}) P_{\ell,m} [\sin(\phi_{gc})] \} \\ &\quad \times \{ C_{\ell,m} \cos(m\lambda) + S_{\ell,m} \sin(m\lambda) \} \\ \frac{\partial^2 U}{\partial r \partial \lambda} &= \frac{\partial^2 U}{\partial \lambda \partial r} \\ &= -\frac{\mu}{r^2} \sum_{l=2}^{\infty} \left(\frac{R_{\oplus}}{r}\right)^l (\ell+1) \sum_{m=0}^{\ell} m P_{\ell,m} [\sin(\phi_{gc})] \{ S_{\ell,m} \cos(m\lambda) - C_{\ell,m} \sin(m\lambda) \} \\ \frac{\partial^2 U}{\partial \phi_{gc}^2} &= \frac{\mu}{r} \sum_{l=2}^{\infty} \left(\frac{R_{\oplus}}{r}\right)^l \sum_{m=0}^{\ell} \{ C_{\ell,m} \cos(m\lambda) + S_{\ell,m} \sin(m\lambda) \} \left[\tan(\phi_{gc}) P_{\ell,m+1} [\sin(\phi_{gc})] \right. \\ &\quad \left. + \{ -m \sec^2(\phi_{gc}) + m^2 \tan^2(\phi_{gc}) - (\ell-m)(\ell+m+1) \} P_{\ell,m} [\sin(\phi_{gc})] \right] \\ \frac{\partial^2 U}{\partial \phi_{gc} \partial \lambda} &= \frac{\partial^2 U}{\partial \lambda \partial \phi_{gc}} = \frac{\mu}{r} \sum_{l=2}^{\infty} \left(\frac{R_{\oplus}}{r}\right)^l \sum_{m=0}^{\ell} m \{ S_{\ell,m} \cos(m\lambda) - C_{\ell,m} \sin(m\lambda) \} \\ &\quad \times \{ P_{\ell,m+1} [\sin(\phi_{gc})] - m \tan(\phi_{gc}) P_{\ell,m} [\sin(\phi_{gc})] \} \end{aligned}$$

$$\frac{\partial^2 U}{\partial \lambda^2} = -\frac{\mu}{r} \sum_{l=2}^{\infty} \left(\frac{R_{\oplus}}{r} \right)^l \sum_{m=0}^l m^2 P_{l,m} [\sin \phi_{gc}] \{ C_{l,m} \cos(m\lambda) + S_{l,m} \sin(m\lambda) \}$$

The partial derivatives on the far right-hand side of Eq. (10-42) are given in Eq. (8-26). Determine the second derivatives of the spherical coordinates using Eq. (8-26):

$$\begin{aligned} \frac{\partial^2 r}{\partial \dot{r}^2} &= \frac{1}{r} \left[\mathbf{I} - \frac{\dot{r} \dot{r}^T}{r^2} \right] \\ \frac{\partial^2 \phi_{gc}}{\partial \dot{r}^2} &= -\frac{1}{(r_I^2 + r_J^2)^{3/2}} \left[\left(\frac{\partial r_K}{\partial \dot{r}} \right)^T - \frac{r_K \dot{r}}{r^2} \right] \left[r_I \left(\frac{\partial r_I}{\partial \dot{r}} \right) + r_J \left(\frac{\partial r_J}{\partial \dot{r}} \right) \right] \\ &\quad - \frac{1}{r^2 \sqrt{r_I^2 + r_J^2}} \left[\dot{r} \left(\frac{\partial r_K}{\partial \dot{r}} \right) + r_K \mathbf{I} - \frac{2r_K \dot{r} \dot{r}^T}{r^2} \right] \\ \frac{\partial^2 \lambda}{\partial \dot{r}^2} &= -\frac{2}{(r_I^2 + r_J^2)^2} \begin{bmatrix} -r_J \\ r_I \left(\frac{\partial r_I}{\partial \dot{r}} \right) + r_J \left(\frac{\partial r_J}{\partial \dot{r}} \right) \\ 0 \end{bmatrix} + \frac{1}{(r_I^2 + r_J^2)} \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{aligned}$$

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 those equations here due to their length.

Finally, we can find the \mathbf{F} matrix using finite or central differences [Eq. (10-17)]. 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):

$$\mathbf{F} = \frac{\partial \dot{\bar{X}}}{\partial \bar{X}} \approx \frac{\Delta \dot{\bar{X}}_o}{\Delta \bar{X}_o}$$

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 (0.5% to 1% usually works well) 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.

$$\mathbf{F}(t) \equiv \begin{bmatrix} \frac{\dot{\bar{X}}_{1_c}(t) - \dot{\bar{X}}_{1_n}(t)}{\bar{X}_{1_c}(t) - \bar{X}_{1_n}(t)} & \dots & \frac{\dot{\bar{X}}_{1_c}(t) - \dot{\bar{X}}_{1_n}(t)}{\bar{X}_{1_c}(t) - \bar{X}_{6_n}(t)} \\ \vdots & & \vdots \\ \frac{\dot{\bar{X}}_{6_c}(t) - \dot{\bar{X}}_{6_n}(t)}{\bar{X}_{1_c}(t) - \bar{X}_{1_n}(t)} & \dots & \frac{\dot{\bar{X}}_{6_c}(t) - \dot{\bar{X}}_{6_n}(t)}{\bar{X}_{6_c}(t) - \bar{X}_{6_n}(t)} \end{bmatrix}_{t=t_k}$$

$\bar{X}_{1_c}(t)$ is the first changed state at t_k , and $\bar{X}_{1_n}(t)$ is the nominal state at the same time.

Calculating the Variational Equations and the Error State Transition Matrix, Φ

When we find the observation partial derivatives separately (\mathbf{H}), we need an equally rigorous approach to find the error state transition matrix ($\Phi = \partial\hat{\mathbf{X}}/\partial\hat{\mathbf{X}}_o$). We then combine the two matrices to find the partial-derivative matrix [Eq. (10-16)]. The most accurate technique finds the error state transition matrix by numerically integrating Eq. (10-27)—a process that depends on finding the \mathbf{F} matrix. In essence, the partial derivatives of the equations of motion are integrated with the state.

Finding the error state transition matrix often uses the variational equations. Remember, the **variational equations** are the partial derivatives of the state at time t with respect to the epoch state at t_o , and we can calculate them using analytical expressions, numerical integration, or finite differences.

Analytical methods for Φ assume analytical expressions for the state as functions of the time and epoch elements. The partial derivatives are obtained from these expressions. To simplify the process, analytical methods use limited force models. These models are usually restricted to two-body motion or two-body plus the J_2 perturbations. Consider the force model limited to two-body motion. If we use cartesian variables and consider variations only in the six epoch state parameters (\vec{r}_o, \vec{v}_o), the variational equations become

$$\Phi = \frac{\partial\hat{\mathbf{X}}}{\partial\hat{\mathbf{X}}_o} = \begin{bmatrix} \frac{\partial\vec{r}}{\partial\vec{r}_o} & \frac{\partial\vec{r}}{\partial\vec{v}_o} \\ \frac{\partial\vec{v}}{\partial\vec{r}_o} & \frac{\partial\vec{v}}{\partial\vec{v}_o} \end{bmatrix}$$

where each of the elements of the above matrix are themselves 3×3 matrices, e.g.,

$$\frac{\partial \dot{\vec{r}}}{\partial \vec{r}_o} = \begin{bmatrix} \frac{\partial r_I}{\partial r_{I_o}} & \frac{\partial r_I}{\partial r_{J_o}} & \frac{\partial r_I}{\partial r_{K_o}} \\ \frac{\partial r_J}{\partial r_{I_o}} & \frac{\partial r_J}{\partial r_{J_o}} & \frac{\partial r_J}{\partial r_{K_o}} \\ \frac{\partial r_K}{\partial r_{I_o}} & \frac{\partial r_K}{\partial r_{J_o}} & \frac{\partial r_K}{\partial r_{K_o}} \end{bmatrix}_{ITRF}$$

The analytical solution can be developed using the Taylor series as we did in developing the classical f and g series (see Sec. 2.3.1 on page 86), and differentiation. Note that the final term in the velocity expression is simply the acceleration. Thus,

$$\begin{aligned} \dot{\vec{r}}(t) &= \dot{\vec{r}}_o + \dot{\vec{v}}_o \Delta t + \frac{1}{2} \frac{d\dot{\vec{v}}}{dt} \Big|_{t=t_o} \Delta t^2 + \dots \\ &\approx \dot{\vec{r}}_o + \dot{\vec{v}}_o \Delta t - \frac{\mu \dot{\vec{r}}_o}{2r_o^3} \Delta t^2 \\ \dot{\vec{v}}(t) &= \dot{\vec{v}}_o + \frac{d\dot{\vec{v}}}{dt} \Big|_{t=t_o} \Delta t + \dots \\ &\approx \dot{\vec{v}}_o - \frac{\mu \dot{\vec{r}}_o}{r_o^3} \Delta t + \left(\frac{-\mu \dot{\vec{v}}_o}{r_o^3} + \frac{3\mu \dot{\vec{r}}_o v_o}{r_o^4} \right) \frac{\Delta t^2}{2} \\ &= \dot{\vec{v}}_o - \frac{\mu \dot{\vec{r}}_o}{r_o^3} \Delta t \end{aligned} \tag{10-43}$$

Ignoring higher order terms (Δt^3 and jerk), the error state transition matrix is given by*

* Note that this formulation will *not* move the state forward in time (as mentioned on page 82 and shown later in this chapter). It is intended only to move the state errors. Φ can also be found using Eq. (10-41) and $\Phi = I + F\Delta t + \frac{F^2\Delta t}{2}$.

$$\Phi \left[\begin{array}{c} \frac{\partial \hat{\mathbf{X}}}{\partial \hat{\mathbf{X}}_o} \\ \text{2-body} \end{array} \right] \equiv \begin{bmatrix} 1 + \frac{3\mu\Delta t^2 r_I^2}{2r_o^5} - \frac{\mu\Delta t^2}{2r_o^3} & \frac{3\mu\Delta t^2 r_I r_J}{2r_o^5} & \frac{3\mu\Delta t^2 r_I r_K}{2r_o^5} & \Delta t & 0 & 0 \\ \frac{3\mu\Delta t^2 r_I r_J}{2r_o^5} & 1 + \frac{3\mu\Delta t^2 r_J^2}{2r_o^5} - \frac{\mu\Delta t^2}{2r_o^3} & \frac{3\mu\Delta t^2 r_J r_K}{2r_o^5} & 0 & \Delta t & 0 \\ \frac{3\mu\Delta t^2 r_I r_K}{2r_o^5} & \frac{3\mu\Delta t^2 r_K r_J}{2r_o^5} & 1 + \frac{3\mu\Delta t^2 r_K^2}{2r_o^5} - \frac{\mu\Delta t^2}{2r_o^3} & 0 & 0 & \Delta t \\ \frac{3\mu\Delta t r_I^2}{r_o^5} - \frac{\mu\Delta t}{r_o^3} & \frac{3\mu\Delta t r_I r_J}{r_o^5} & \frac{3\mu\Delta t r_I r_K}{r_o^5} & 1 & 0 & 0 \\ \frac{3\mu\Delta t r_I r_J}{r_o^5} & \frac{3\mu\Delta t r_J^2}{r_o^5} - \frac{\mu\Delta t}{r_o^3} & \frac{3\mu\Delta t r_J r_K}{r_o^5} & 0 & 1 & 0 \\ \frac{3\mu\Delta t r_I r_K}{r_o^5} & \frac{3\mu\Delta t r_K r_J}{r_o^5} & \frac{3\mu\Delta t r_K^2}{r_o^5} - \frac{\mu\Delta t}{r_o^3} & 0 & 0 & 1 \end{bmatrix}$$

The analysis is much simpler using the classical orbital elements, a , e , i , Ω , ω , and $M = nt + M_o$. If \hat{c} represents orbital elements instead of cartesian vectors, the variational equations become

$$\frac{\partial \hat{c}}{\partial \hat{c}_o} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ \frac{-3n}{2a} \Delta t & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

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. You can include additional gravitational effects in the variational equations using equations like Eq. (8-56). The accelerations simply add to the two-body acceleration we used in the previous Taylor series expansion. This model is sufficient for many applications.

Under certain circumstances, we can obtain an analytical formulation for $\Phi(t_o, t)$. Junkins (1975:45-48) shows the general case for an analytical development, as well as how to derive the Peano-Baker method for determining the error 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 **sta-**

tionary 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. (10-25) 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

$$\delta\mathbf{x} = \left\{ \mathbf{I} + \mathbf{F}(t - t_o) + \frac{1}{2!}\mathbf{F}^2(t - t_o)^2 + \frac{1}{3!}\mathbf{F}^3(t - t_o)^3 + \dots \right\} \delta\mathbf{x}_o = \Phi\delta\mathbf{x}_o \quad (10-44)$$

The number of terms will vary for different problems. The expression in the brackets is the definition of a matrix exponential; thus, we can rewrite the Taylor series expansion and the error state transition matrix:

$$\begin{aligned}\delta\mathbf{x} &= \text{EXP}[\mathbf{F}(t - t_o)]\delta\mathbf{x}_o \\ \Phi(t, t_o) &= \text{EXP}[\mathbf{F}(t - t_o)]\end{aligned} \quad \left. \begin{array}{l} \mathbf{F} \text{ constant only} \\ \text{Eq. (10-45)} \end{array} \right.$$

The second equation follows by comparing the first equation with Eq. (10-26).

Eq. (10-45) 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.

When we find the error state transition matrix with numerical techniques, we often chose complete force models in the variational equations as we saw in the last section. We then numerically integrate Eq. (10-27) to find current solutions of the matrix. Finite differencing is generally not applicable to the formation of Φ because it has limited accuracy, and if it's used at all, it's used for the complete \mathbf{A} matrix.

Calculating the State Transition Matrix, Φ_S

There is the potential for confusion when finding the state transition matrix to propagate the state through time as mentioned on pg 90, and earlier in this section. To find the state transition matrix for state propagation (Φ_S), begin the same as with the derivation of the error state transition matrix. The propagation of the state is

$$\dot{\hat{X}} = \Phi_S \hat{X}_o$$

with an assumed initial condition of

$$\begin{aligned} \dot{(\Phi_S)}(t, t_o) &= \mathbf{G}(t)\Phi_S(t, t_o) \\ \Phi_S(t_o, t_o) &= \mathbf{I} \end{aligned}$$

Notice the use of \mathbf{G} to distinguish the process from the state error propagation.

To find Φ_S , begin with the Taylor series as we did in developing the classical f and g series (see Sec. 2.3.1 on page 86), and also in Eq. (10-43). Note that the final term in the velocity expression is the derivative of acceleration, or jerk, and it can be ignored for now. On pg 90, we evaluated the f and g series for orbital elements. Here, we wish to simply keep the position and velocity vectors, so we just take the Taylor series and put it in matrix form (no partials and no variational equations are required). This finds Φ_S directly. I show just the common diagonal elements of each 3×3 diagonal sub-matrix.

$$\Phi_S \equiv \begin{bmatrix} \left[1 - \frac{\mu \Delta t^2}{2 r_o^3} \right] [\Delta t] \\ \left[-\frac{\mu \Delta t}{r_o^3} \right] [1] \end{bmatrix} \quad (10-46)$$

We can obtain the same result defining a matrix \mathbf{G} as $\dot{\hat{X}} = \mathbf{G}\hat{X}_o$, and then finding the state transition matrix as $\Phi_S = \mathbf{I} + \mathbf{G}\Delta t + \frac{\mathbf{G}^2 \Delta t^2}{2}$. The equivalent expression for \mathbf{G} is found by inspection. I show the common diagonal elements of each 3×3 diagonal sub-matrix.

$$\mathbf{G} \equiv \begin{bmatrix} [0] & [1] \\ \left[-\frac{\mu}{r_o^3} \right] & [0] \end{bmatrix}$$

10.8 Implementing Differential-Correction Solutions

Most robust operational estimation systems are very complex, especially if they use detailed analytical partials. The complete architecture is important. The ability to quickly have observational data available, computational power, and rapid communications are desirable. The Navy developed a system with these features (Boers et al. 2000, and Seago et al. 2000). Determining which technique to use (See Table 10-4) and the data architecture are fundamental. Any technical changes or updates shouldn't need lengthy (and costly) periods of reprogramming.

Also important is operator knowledge. Operators and decision makers who never study the technical details may miss the nuances of a mathematical technique. This could lead to incorrect system-level decisions by people based on their knowledge of a particular software program, and not the technical details. 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 not use force models whose accuracy degrades the content of the best observations you receive (where possible). In fact, your force models should be at least an order of magnitude *more accurate* than your best observations. 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.

Mathematical Technique

The selection of using a Batch Least Squares vs a Kalman Filter evokes tremendous discussion. Ultimately, the requirements are the deciding factor. If we need continuous near real-time updates, Kalman Filter approaches are preferred. If our goal is routine position determination, the batch Least Squares techniques suffice.

Comparisons of the two techniques are limited, but Montenbruck et al. (2005) compare Least Squares and Kalman Filter approaches to orbit determination. While they note differences (accelerations at the couple nm/s² level) and conclude that the Least Squares approach is better, the positional comparisons are shown to be remarkably close. It's possible that the acceleration differences are a result of the two approaches trying to adjust the solve-for parameters to match the slightly imperfect force models. Operational decisions based on either approach would probably be identical.

Force Model Treatment

Non conservative forces (drag and solar radiation pressure) are particularly challenging for OD operations. The BC is especially difficult to model because several parameters are contained within the BC. We know that material properties drive the c_D value (Gaposchkin, 1994), and that the material properties change with exposure to the space environment over time. The area certainly changes with respect to the orientation of the satellite over time, and the mass may change if any thrusting, or venting occurs.

A ***segmented BC*** technique (also called an arc, Ullman:1992) is sometimes used with Least Squares approaches in which the BC is kept constant for a certain period of time. This could be for some interval within the fit span, or on an observation track basis. In the limiting case, a Kalman Filter considers BC changes with each observation. For batch Least Squares systems, additional overhead is required to accomplish the processing. The ability to ingest attitude profiles of a satellite to permit dynamic cross sectional area calculations is likely needed to eliminate much of the uncertainty.

SRP varies much like the BC, but there are fewer parameters, namely, the solar radiation pressure coefficient, c_R . Similar techniques to those for the BC are implemented to better model the force, but a popular approach is to implement a box-wing model to better determine the cross-sectional area of the satellite in force model calculations. As with atmospheric drag, this approach is very effective.

Update Intervals and Fit Spans

Tracking systems often use specialized methods to determine when to update a satellite's positional information. I'll define an ***update interval*** as the amount of time which may pass before we *update* the satellite state vector using new observations. Typically these values range from an update once per day for low Earth orbits, to updates once per week or more for GEO.

The length of the update interval depends on the quality and quantity of observations, the satellite orbit, whether the satellite maneuvers, the estimation and propagation schemes, the type of sensors, and the desired confidence and accuracy. Ideally this interval should occur before the propagation theories would introduce enough error to keep the next sensor from re-acquiring the satellite, or when the most stringent accuracy requirement is no longer met. Unfortunately, this is often not the case; the update-interval times are often significantly shorter than the actual requirement—something that can stress limited sensor resources. If the process is done efficiently, much longer update intervals are possible, sometimes weeks or more in length. This requires proper force models, processing, and prediction.

Closely related is the concept of a fit span, and it applies to batch Least Squares systems. A ***fit span*** is defined as the period of time over which we collect data for a differential correction. For LEO satellites, this can be 4-5 days. For GEO satellites, 30 days is common. The Russians have successfully processed fit spans of 500 days (Khutorovsky, 1998). Be careful not to confuse this with the update interval. A fit span can include many update intervals, as shown in Fig. 10-11. Schumacher and Glover (1995) describe a method below to determine the fit span for analytical theories based on the satellite's period and rate of decay (mean motion). Other methods use the eccentricity and the period. For non-maneuvering satellites,

<u>Period</u>	<u>Fit Span</u>
$\rho > 800 \text{ min}$	30 days
$600 < \rho < 800 \text{ min}$	15 days
$\dot{\rho} > -0.0005 \text{ min/day}$	10 days
$-0.001 < \dot{\rho} < -0.0005 \text{ min/day}$	7 days
$-0.01 < \dot{\rho} < -0.001 \text{ min/day}$	5 days
$\dot{\rho} < -0.01 \text{ min/day}$	3 days

where we've used

$$\dot{\rho} = \frac{-4\pi}{n^2} \left(\frac{\dot{n}}{2} \right) (1440 \text{ min/day})$$

An example where we need 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. The 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 contribution from drag. As the fit span increases, we begin to observe the separate quadratic and linear changes in the along-track motion.

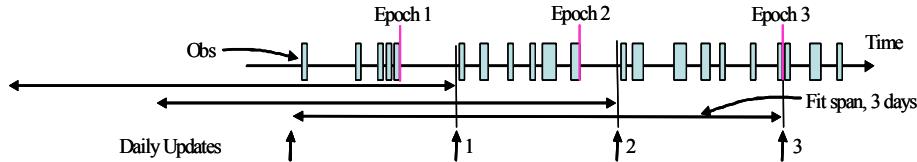


Figure 10-11. Depiction of Fit Span and Update Intervals. This figure highlights three updates in a system where the update interval is about one per day. The fit span, 3 days in this example, gathers all available data and places the epoch at the last observation so it's current. Notice that the epoch doesn't always coincide with the update time. 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.

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 move 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.” A Kalman Filter and smoother avoids this dilemma somewhat because there is no fit span and the parameters are free to vary at each observation time. However, sufficient time is needed for the filter to converge. This time *can* be less than the fit span.

10.8.1 Convergence Issues

The estimation process 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 a limited number of observations.

Data Editing

We've discussed the stopping criterion in Eq. (10-12) and Eq. (10-18) previously. But this also is dependant on rejection of outlying observations. Seago et al. (2011) provides an excellent discussion of the data editing process and portions are extracted below.

The lack of statistical independence in the elements of the residuals complicates the application of statistical tests to identify outlying measurements in batch Least Squares solutions. An assumption of the covariance of the residuals $\approx \mathbf{W}^{-1}$ would result in residual uncertainties that tend to be too small.

Seago et al. (2011) introduces a *projection matrix* ($\Lambda = \mathbf{H}(\mathbf{H}^T \mathbf{W}\mathbf{H})^{-1}\mathbf{H}^T \mathbf{W}$) as an analog of the Kalman gain, \mathbf{K} , but for Least Squares systems. However, even if Λ were computed to improve the realism of the residual covariance, its contribution is still likely to under-represent the *true* level of residual uncertainty, due to a lack of dynamical process noise within the Least Squares formulation. Therefore, the use of theoretical residual variance tends to be impractical for testing batch Least Squares residuals, and therefore for data editing.

During the first few iterations of Least Squares techniques, the residuals are affected by slowly-varying systematic errors largely induced by the errors in the state parameters and the as yet uncompensated sensor biases (Koskela, 1967:111). The root-mean-square (RMS) of the batch Least Squares residuals (Eq. (10-12)) is sometimes used after an iteration to establish the scale factor for future rejections. While RMS editing is very common, there are some disadvantages.

- Because batch Least Squares residuals are not independently and identically distributed (they are actually heteroskedastic, or time varying), the RMS is not a truly reliable estimate of residual variance. A $C \times$ RMS criterion may not reject data at the intended rate. Unfortunately, when tracking data are sparse, a criterion based on $C \times$ RMS can be completely ineffective. If the RMS is estimated from a sample size n , then outlier detection based on $C \times$ RMS is impossible whenever $C > \sqrt{n}$. For example, $10 \times$ RMS is often used to reject impossible values. Yet the sample size must be greater than 100 to detect an outlier.
- The first iteration has no RMS value to use in testing the residuals. Subsequently, it can already be contaminated by outlying observations not previously detected. This weakens the test and potentially masks the presence of outlier observations.
- The degree of convergence of non-linear batch Least Squares estimator is often gauged by changes in the (weighted) RMS. Inefficient outlier editing can substantially change the RMS and can result in unnecessarily slow convergence.

RMS calculations for data-rejection purposes are typically unweighted and may be based on observations only of a specific measurement *type*. For example, while twenty-four tracking measurements of range, azimuth, and elevation may be available for maintaining an orbit, only eight of those measurements may be ranging measurements. The previous $C \times$ RMS example applies to each observation type separately. More robust estimators of scale beyond RMS could be used, although their reliability tends to presume that the data are independently and identically distributed (IID) with known distributional characteristics, and their reliability also decreases with very small sample sizes.

Variate-difference methods look for outliers within a sequence of residual differences. Editors based on first- or second-differences can find one or two outliers, but their usage adds complexity to the estimation process. There are other limitations.

- Two or more successive outliers may mask each other when differenced, thus requiring higher-order differences, and greater complexity.
- As higher order differences are used, determining the source of the larger and smaller difference from the norm requires significant bookkeeping.

- Residual differences relative to the un-converged solution may exhibit large non-random behavior, which may not appear Gaussian. Preliminary outlier rejection based on a Gaussian assumption may prematurely eliminate good data.
- Differencing successive residuals is uninformative on short sequences typical for some types of US space-surveillance tasking (three observations per pass).

Outlier rejection schemes which are not based on the distribution of measurement residuals, can also be applied to batch or sequential estimators. A hard limit may be imposed on what is an acceptable magnitude for a measurement. Geometrical constraints may be imposed, such as rejection of very low elevation tracking, or tracking data that appear to have originated outside the field of regard of the sensor. The probability of an erroneous observation at the start of a pass is greater than subsequent measurements, and these tend to occur only in particular classes of sensors (radars). These could be weighted accordingly.

Initial Orbital Estimate

If the estimation algorithm doesn't converge, our initial estimate may be too far off (the initial estimate needs to be within the radius of convergence for the true position), or the observation quantity or quality may be inadequate. Although the observation quantity is usually the problem, insufficient calibration and poor observability can also affect the process. Initial orbit determination techniques can provide an initial estimate, but they are usually insufficient to run a successful orbit determination from. A Least Squares solution can sometimes improve the initial guess from the IOD.

Matrix Inversion

When processing 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 estimation is started incorrectly. These problems are often related to difficulties with the matrix inversion. There are several methods to mitigate this situation. Most techniques can benefit from singular value decomposition (SVD), fading-memory filters, and square root matrix inversion algorithms.

Matrix inversion of the normal equations has received some attention in the literature [Marshall, 1999 and Vallado and Crawford (2008)]. Traditionally the LU decomposition method is favored as it lends itself to efficient use in terms of computer memory and processing. However, if the matrix (A) is singular or near-singular, the classical LU method (and others) can fail completely or result in very inaccurate answers. The common use of double precision in modern software helps push up the matrix solution accuracy in near-singular cases, but this avoids the underlying problem. Typically near-singular cases yield solve-for parameters with unrealistically large values that delicately cancel each other. Marshall (1999) investigated the use of SVD in the Navy operational system and reached the same conclusions.

The use of SVD has the major advantage that (theoretically at least) it cannot fail to produce a useful solution. The disadvantage is greater processing and memory demands, but those are no longer significant because memory is large, cheap, and the processing

load is still less than the generation of the A matrix for most reasonable numbers of observations (i.e., N less than 1 – 10K points).

The fact that SVD offers a stable solution should not be treated as an excuse for ignoring the underlying problem, that of certain solve-for parameters not being well defined by the available observations or by the orbital model itself (e.g., BSTAR drag coefficient with highly eccentric orbits). This type of difficulty is far more common than it initially appears, but can often be helped by a good choice of state representation.

Computationally, SVD replaces the matrix inversion in the normal equations with the following steps. See (Press, 1992:51-62) for specific details of the SVD algorithm.

- Use SVD to decompose the weighted Jacobian matrix $[S \cdot A]$ into U , V , and w , where ($S = A^T W$).
- Inspect w_j , set any ‘very small’ w_j terms to zero rather than any $1/w_j$ term becoming nearly infinite
- Solve $d\mathbf{x} = V \cdot (\text{diag}[1/w_j]) \cdot U^T \cdot S \cdot b$
- Find the covariance matrix from $V \cdot (\text{diag}[1/w_j]) \cdot U^T$

Fading-memory filters weight older information less than the current information. We want to ensure the covariance matrix is never negative, or zero. The qualitative reason for weighting older data is to tell the estimation process that the dynamical model is imperfect, and therefore the extrapolated dynamics are suspicious. There are both additive and multiplicative forms and they apply differently to Least Squares and Kalman Filter techniques. For the Kalman Filter, Gelb (1989:286) shows how to evaluate a fading-memory variable, m_{fade} , which is

$$m_{fade} = \exp\left[\frac{\Delta t_{obs}}{\tau_{age}}\right]$$

Here, we use the time between observations, Δt_{obs} , and the constant-age weighting factor, τ_{age} , and place the values on the diagonal of a square matrix, M_{fade} . In essence, we’re increasing the measurement error for the older observations. The effect produces an exponential decay on the use of older data. For Kalman Filters then,

$$\begin{aligned} K_{k+1} &= \bar{P}_{k+1} H_{k+1}^T \left[H_{k+1} \bar{P}_{k+1} H_{k+1}^T + M_{fade} R_{k+1} \right]^{-1} \\ \hat{P}_{k+1} &= \frac{\bar{P}_{k+1}}{M_{fade}} (I - K_{k+1} H_{k+1}) \end{aligned}$$

To use a fading-memory filter for Least Squares techniques, recall the sequential-batch formulation [Eq. (10-19) and Eq. (10-20)]. The previous data could be counted less by simply multiplying by 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. For example, if $m_{fade} = 0.933^{\Delta t_{obs}}$, the fading yields a “half-life” of about 10 days for the data (Chioma, 1998). The resulting equations are

$$\begin{aligned}\hat{\delta x}(0|k+n) &= (\mathbf{A}_{new}^T \mathbf{W}_{new} \mathbf{A}_{new} + \mathbf{M}_{fade} \hat{\mathbf{P}}_k^{-1})^{-1} (\mathbf{A}_{new}^T \mathbf{W}_{new} \tilde{\mathbf{b}}_{new} + \mathbf{M}_{fade} \hat{\mathbf{P}}_k^{-1} \hat{\delta x}(0|k)) \\ \hat{\mathbf{P}}_{k+n} &= (\mathbf{A}_{new}^T \mathbf{W}_{new} \mathbf{A}_{new} + \mathbf{M}_{fade} \hat{\mathbf{P}}_k^{-1})^{-1}\end{aligned}$$

This simplistic example is just one of many methods.

Square-root filters are a more robust numerical approach to propagate poorly conditioned (near singular) covariance matrices. They use the square root of the covariance matrix, \mathbf{S} , which is derived using *Cholesky decomposition*. Notice that \mathbf{P} will always be positive, or at least non-negative.

$$\mathbf{P} = \mathbf{S}\mathbf{S}^T$$

Gelb (1989:306–308) derives an update recursion for \mathbf{S} to replace Eq. (10-33) 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), Tapley et al. (2004, Chap 5), or Maybeck (1979:368-405) for a more detailed discussion of square-root filters.

Consider Parameters

Sometimes the OD process has difficulty converging because there is significant uncertainty in the force models and satellite parameters (e.g., atmospheric drag, BC). In a traditional Least Squares approach, a single set of parameters model the BC (bias, std dev, etc). However, if the true BC value is constantly changing, error is introduced by using static parameters. Consider parameters insert additional uncertainty into the OD process to better model that uncertainty. The noise from atmospheric drag may be very large and an OD process may try to “assign” this uncertainty to individual parameters (c_D or BC) instead of attributing it to the force model. This could lead to incorrect assumptions about the parameters, and erroneous propagation results.

We know there is coupling of the BC and the gravitational model. As we solve for the BC in differential corrections, it absorbs error from poorly modeled gravitational fields, so a complete gravity field is recommended for low-Earth orbits. Atmospheric density models are often only accurate to about 10–15% (Lee and Alfreind, 2000, and Granholm et al. 2000). Simple estimation processing will allow the state parameters to “absorb” this error. With a consider parameter approach, you add a constant 15% consider parameter in the covariance matrix to introduce this known uncertainty during processing.* This option approximates the essence of the solve-for parameters, but it lacks the rigor of the solve-for approach because consider parameters are not estimated as part

* To populate the consider covariance, see Tapley et al. (2004: 389-396).

of the solution. Mathematically, the consider parameter is inserted into the 7×7 position of the covariance matrix (if the BC is in the 7th position in the state matrix). Various options exist from additive, to multiplicative, and other simple relations. Another approach uses a variable consider parameter for processing. Alternatively, a Kalman Filter can estimate the uncertainty in the atmospheric density (Wright and Woodburn, 2004). In all cases, the basic principle is to limit known errors from affecting the accuracy of the state parameters that you estimate.

Track Weighting

Batch Least Squares approaches can also be limited when parameters vary significantly within the observation span. Track weighting (also called arc-processing, Ullman:1992) acknowledges the dynamic behavior of the BC (and SRP) over the course of a fit span and is similar to the segmented BC. However, track weighting is related to the inherent dynamic behavior of the sensor and not the physical nature of the satellite. We perform track weighting during the accumulation of the normal equation matrices by simply having different weight values for each track, and summing each track separately. Ullman (1992) describes this process in greater detail. Of course, there is additional computational overhead to maintain the track (or possibly time) boundaries, but the advantages in performance improvement are usually worth it.

10.8.2 Residuals

During the OD process, the residuals often determine the success of the operation. For Least Squares, the residuals are iterated over time until the best estimate is achieved. For filters, there is a single pass through the data during which the residuals should show a reduction in uncertainty as more observations are processed. Typical plots for both approaches should show a majority of the observations being accepted and processed, with a few measurements being rejected. In Fig. 10-12, I show residuals divided by the individual standard deviations. This permits evaluation of different observation types (e.g., range and azimuth) on the same plot, while also providing a simple ± 3 metric which should capture the majority of observations.

Filters often employ smoothers that take the final filter state and re-process the observations one time, but in reverse order. We are no longer examining the residuals, and we rely on the uncertainty to ascertain the success of the operation. The result is usually a “bathtub” effect in which the uncertainty is least in the middle (where observations occur on both sides of the estimates), and greatest at the ends (where there are observations on only one side). Figure 10-13 shows a typical position uncertainty.

10.8.3 Post-processing and Prediction

We need to distinguish 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 are measured and we can better estimate the time-varying parameters (e.g., BC). Because we know the physical parameters (atmosphere and solar radiation), the number of unknowns is reduced and the estimation pro-

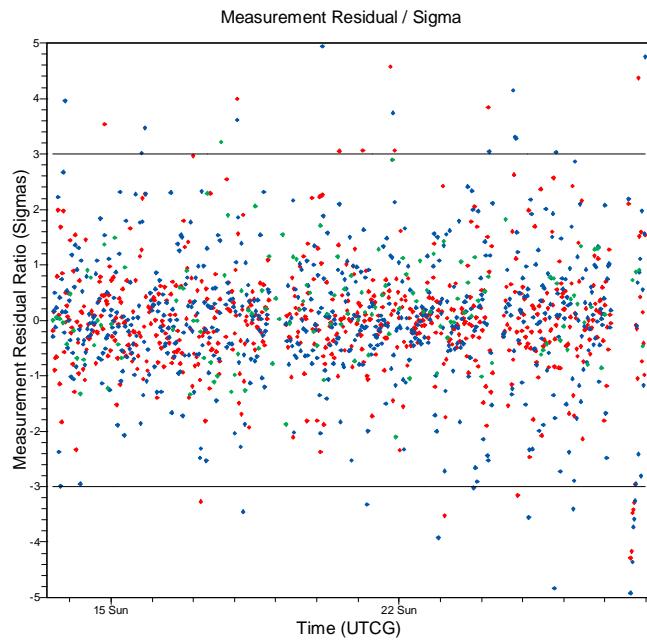


Figure 10-12. Typical Residual Ratio Plot. The range, azimuth, and elevation residual ratios are shown. Notice the majority of the observations fall within a ± 3 limit. By dividing each residual by its standard deviation, all the observations can be shown on one graph. (ODTK image courtesy of AGI)

cessing can “concentrate” on solution of the satellite parameters. In near real-time applications (processing within a few days of the current time), the predicted solar flux indices and EOP values are not known precisely, so there can be varying accuracies within post processing. Post processing can also fuse multiple OD strategies to fully reduce the observational data. This form of post processing is usually very accurate. Reference POEs from orbit determination for the TOPEX orbit routinely achieve radial position accuracies of about 2 cm (Ries and Tapley, 1999).

Prediction involves a very different problem—that of using a given state vector to determine a satellite’s future positions without knowing all the physical parameters (solar flux and EOP). This is the executing the propagation techniques discussed in Chaps. 8 and 9, but here, we are interested in the ability of the propagation to reflect the true orbital state and how best to extend the estimation results. 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 solutions! The estimation process incorporates poorly known, or constant physical parameters of the satellite with simple estimates of the atmospheric and EOP values—a process that degrades the accuracy.

For filters, an important subtlety is noted. For propagation, it may be best to propagate a state that is slightly “back” from the end time of the observations processed. This

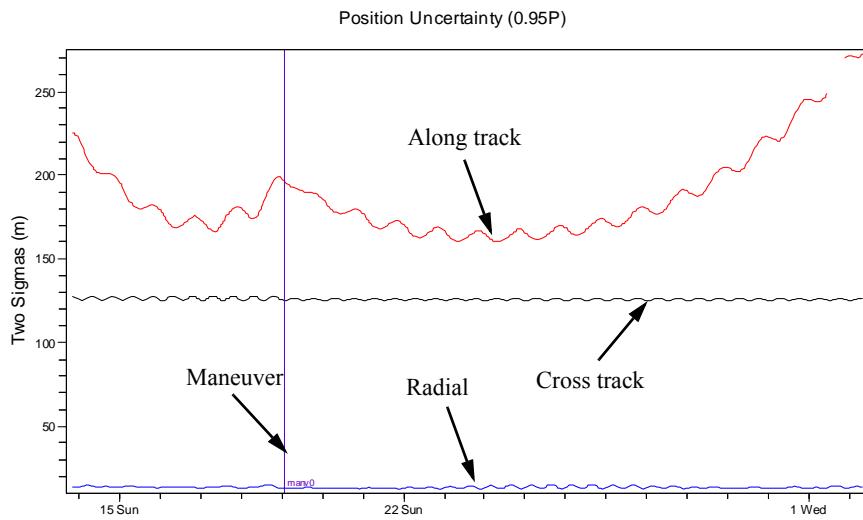


Figure 10-13. Smoother Position Uncertainty. A typical position uncertainty from a smoother. There is a known maneuver in the data and you can see the slight increase in uncertainty as the filter and smoother process near it. Cross track uncertainty is often smaller, but in this case, there is limited observability in that direction. (ODTK image courtesy of AGI)

is because the smoother has the most information there, the uncertainty is the lowest, and the sensor and satellite parameters are best known. Depending on the OD, this may be relatively close to the final state time, and therefore using the prior vector introduces little additional propagation error into the resulting ephemeris.

Getting the Initial State

Because the POE's are accurate to a few cm, they can be considered as “truth”, but are often termed *reference orbits* if the accuracy is larger. There are three possible options to find the initial state for use in propagation: interpolating a state from the POE, perturbing a state from the POE to simulate what could be achievable by an on-board processing system, and performing an OD on a portion of the POE, and then propagating into the future (the last option could also be doing an OD from observations and propagating). We assess accuracy by comparing the propagated data points with the POE. Figure 10-14 shows examples of these for a GPS satellite (NORAD 27704) in about a 20,000 km altitude circular inclined orbit.

You may be tempted to use the initial state of a reference orbit to predict (propagate) forward in time. If the reference orbit is good to a few cm, this seems logical. However, experience suggests that the force models between programs will never be identical, and different treatment of the input data (solar flux, geomagnetic indices, etc), can introduce differences into the final solution. The best approach is to perform an OD on the ephemeris data to get the solve-for parameters as close as possible.

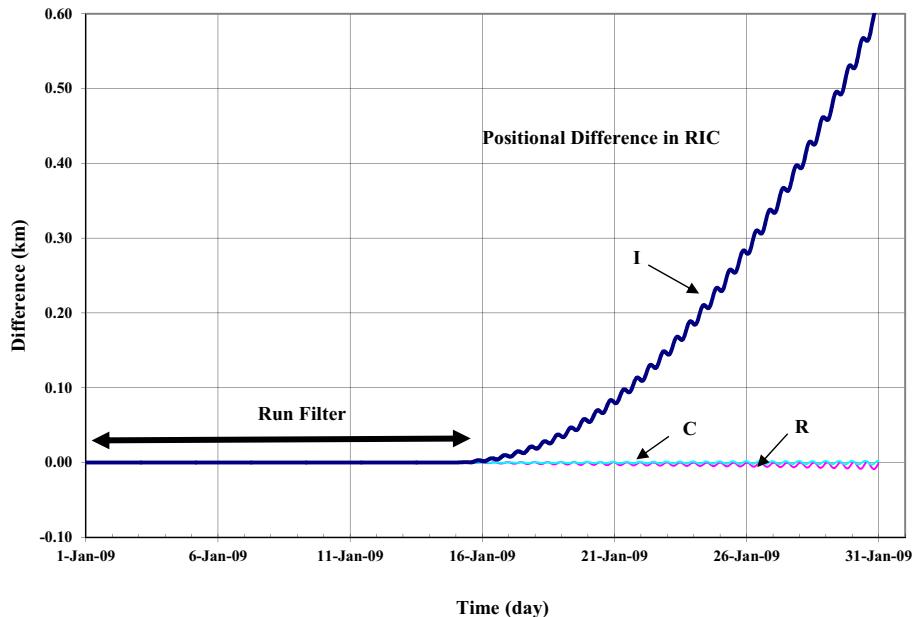


Figure 10-14. Prediction Performance. This figure shows prediction accuracies for about 14 days resulting from an Orbit Determination of a GPS POE. The POE was used as observations to find the original state vector. The comparison during the time of observations was well under a meter. The final state was then propagated and compared to the POE. Note the rapid divergence, predominantly in the in-track direction. The satellite is at about 20,000 km altitude.

The concept of performing an OD on a reference orbit applies to analytical techniques as well. Wilkins et al. (2000) show the precise conversion of elements in an operational environment. Specifically, they investigate the concept of maintaining the space catalog with numerical techniques, and reconstructing analytical element sets from propagations of the numerical vectors. Their results indicate a 2–3 fold improvement in the mean element accuracy. Cappellucci (2005) showed similar results. As more attention is focused on the operational use of numerical techniques, this could be an important benefit for older systems or for applications still requiring analytical methods.

An interesting application is processing TLE data to actually improve the accuracy for a period of time. See Muldoon et al. (2009), Flohrer et al. (2009), and Levitt and Marshall (2011) as they conclude that processing ephemerides generated by SGP4 can be more accurate than the original TLE. This may seem contradictory, but remember that the TLE's will be most accurate before their epoch (where the observations were in the fit span). An OD applying additional force models can thus better model the ephemeris states than the original SGP4 orbital ephemeris, and thus achieve greater accuracy. The improvement is directly tied to the number and accuracy of TLEs available. The technique works best with many TLEs (per day), generally available for calibration and

highly tracked satellites. Also, when forming the TLE ephemeris, it's best to splice at times between TLE epochs to take advantage of the increased accuracy of the TLE before its epoch time.

10.8.4 Propagating the Error Covariance

A common method to evaluate an estimation solution is the covariance matrix. We usually want to know how good our estimate is in the future, thus, we propagate the covariance matrix [Eq. (10-29)] along with the state propagation. A general rule of thumb is that the prediction interval should approximately equal the fit span for optimum accuracy in a Least Squares approach.

The covariance matrix gives a first indication of the solution accuracy *at epoch*. Over time, the error will grow, perhaps significantly. Some variations, like cross-track errors, are periodic and do not grow over extended time periods. The overall volume of the error ellipse remains the same over time. The specific shape depends on the observations, geometry, and the techniques used to determine the orbit. A GEO satellite observed with transponder ranging will have a thin ellipse resembling a pancake because the radial position is known to within a few meters, but the in-track and cross-track are much larger due to their being nearly unobservable. Error growth depends on propagation force model errors, and on initial errors in the covariance.

Remember that the covariance is a statistical uncertainty resulting from the processing of observations. As such, a 100 m uncertainty at some point in the future means that after many trials, the average uncertainty will be 100 m, 68.27% of the time, 1 sigma, one degree system. It would be incorrect to say that after a single trial the satellite would be within 100 m (1σ). Only after many trials and detailed comparisons with future observations, could a statistically valid statement be attached to the 100 m value.

The growth in the error covariance is best understood by examining both velocity and position errors. The initial state contains velocity errors which 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 the state space propagates through time, the growth of these errors gradually erodes 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 value only accelerates that degradation.

We can visualize the growth of the covariance over time by simply propagating points the orbit numerically, and performing the proper conversion operation (Sec. 6.8.3, See also Hill, Alfriend and Sabol 2008). A banana shape along the orbit is common as the errors get very large.

Significant discussion centers around the coordinates, elements, and form of covariance propagation. Equinoctial elements generally seem to perform better, but some UKF approaches claim to remove the uncertainties in the covariance propagation.

10.8.5 Covariance Matrix Transformations

Covariance matrices are generally expressed in cartesian coordinates, but other forms exist. Vallado (2003) examined these transformations in detail and provides specific equations. Figure 10-15 shows various combinations of transformations. Note that there are two basic types of transformations—those to different coordinate systems, and those to different sets of orbital state formats. From the figure, you may wonder why alternate orbital state formats are not shown for some of the coordinate systems, such as ECEF. This is because when the subsequent transformation is made to classical elements, the classical elements no longer represent an actual orbit of the satellite due to the velocity vector changes in the transformation from “inertial” to “fixed” coordinates. Thus, a change to these alternate orbital state formats would not yield useful information.

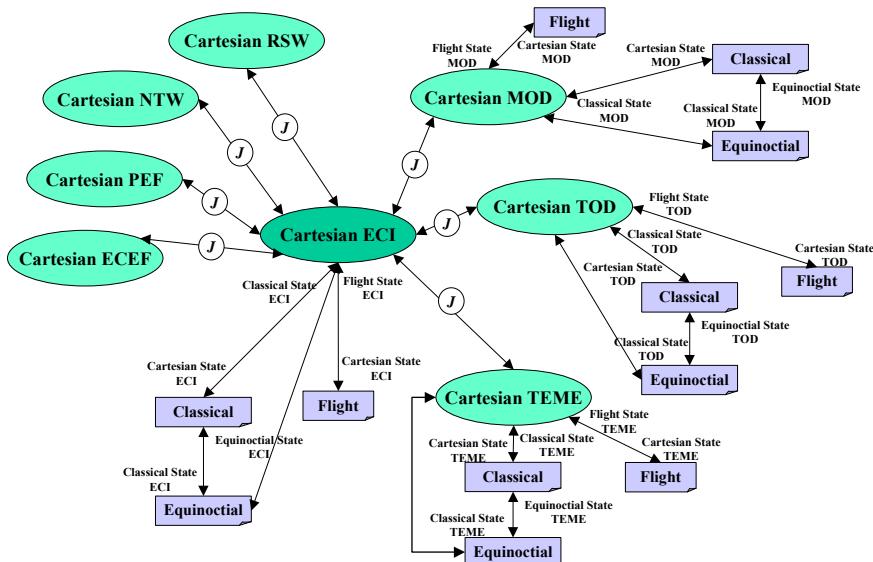


Figure 10-15. Covariance Transformations. This figure shows the types and data required to accomplish each transformation. The central point is the cartesian, ECI covariance matrix. To transform to different coordinate systems, a simple transformation matrix (J) is required. To change the format of the covariance matrix, the state is needed in the desired format (cartesian, classical, etc.).

The basic transformation relies on the similarity transformation developed in Eq. (10-8). Here, m is known as the **Jacobian** for the transformation and the subscripts indicate the ending/initial states.

$$\mathbf{P}_y = m \mathbf{P}_x m^T \quad m_{\frac{y}{x}} = \frac{\partial y}{\partial x}$$

There are several important mathematical properties that we can use for this application. First, the Jacobian matrix is orthogonal. This means that the inverse is equal to the transpose, and that its determinant is equal to unity. The covariance matrix is symmetric, implying that it can be diagonalized ($D = B^T P B$) where D is the diagonal matrix, and B is an orthogonal matrix related to the eigenvalues of P . For orbital state format conversions, the Jacobian assumes the form (where oe are any set of orbital elements),

$$J_{\frac{\dot{r}}{oe} \frac{\dot{v}}{oe}} = \begin{bmatrix} \frac{\partial \dot{r}}{\partial oe} & \frac{\partial \dot{v}}{\partial oe} \end{bmatrix}$$

The transformation for different coordinate systems uses the transformation matrix (Sec. 3.4.3 and Sec. 3.7). For example, to express ECI covariance results in MOD, using matrix multiplications and a short hand notation where [Prec] [Eq. (3-89)] and [0] are each 3×3 submatrices, the Jacobian of the system is

$$[J] = \begin{bmatrix} [Prec] & [0] \\ [0] & [Prec] \end{bmatrix} \quad \vec{P}_{MOD} = [J] \vec{P}_{ECI} [J]^T \quad (10-47)$$

To change the orientation, such as to RSW, we use

$$[J] = \begin{bmatrix} [\hat{R} \mid \hat{S} \mid \hat{W}]^T & [0] \\ [0] & [\hat{R} \mid \hat{S} \mid \hat{W}]^T \end{bmatrix} \quad \vec{P}_{RSW} = [J] \vec{P}_{ECI} [J]^T \quad (10-48)$$

We can also interpolate covariance matrices, although the process isn't quite the same as interpolating ephemerides. See Alfano (2004) for details on how to accomplish this process.

Sensor Modeling

The selection of the proper sensor bias and weight parameters is important, but difficult. Values found in Least Squares and Kalman Filters systems will likely vary, but they can be used as initial values in the other's operation.

The actual site location can strongly influence the OD results. Consider a sensor whose location is off by 500 m. Unless this bias is picked up by the OD processing, the orbit will be in error by at least this amount. In systems where the observations are sparse, the problem is accentuated. The altitude of the sensor centroid above the ground should be included.

Proper modeling of the sensor observations is crucial to forming the correct weighting matrix. Generally, a simple bias and standard deviation is used to model the uncertainty in the observations. In addition, time-varying behaviors can model more complex and longer period sensor characteristics. We usually assume that the sensor characteristics are either constant, or that they vary periodically, or randomly. Approaches to model dynamic variations in the observations are numerous. One approach is to use a Gauss-

Markov technique where there is an uncertainty and a $\frac{1}{2}$ -life which serves to set a time limit for the parameter to “move” during processing. A Vasicek approach (Seago et al. 2011) uses multiple time sequences to model the behavior.

10.8.6 Sensor Calibration

Estimating the sensor biases and noise statistics is difficult, but it strongly affects the results of differential correction because without it, these effects become aliased into the orbit solution. We could view data in Table 4-4 as *constant* values, but they’re actually only estimates of the true time-varying values calculated through many mathematical operations. In fact, the weights are derived from sample estimates and may not be representative of the true population statistics. Some systems (such as GPS and SLR) do this technically demanding process well; others do it poorly. Seago, et al. (2000) discuss the NRL process in which they use SLR data to produce definitive reference orbits. You should always try to apply external (independent and highly accurate) references, if they exist! Reference or POE orbits often allow you to determine biases, but the observability must be sufficient. Data fusion can also sometimes find unknown biases.

The fact that the noise and biases change over time and for different satellites and geometries doesn’t mean the values should be ignored. Instead, we should examine the statistics closely to detect large variations which may indicate more severe (and correctable) problems. Least-squares techniques will estimate systematic values, whereas filters can sometimes better track the time-varying values while still solving for the systematic errors. In each case, we must have enough data to permit a solution to be found. Too little data prevents the accurate determination of both the state and bias values. Fonte (1995), has shown that using incorrect biases can produce an almost 70 m error in a low-Earth satellite’s location. A satellite transponder bias of 300 nsec is nearly a 90 m difference! Miscalibration may result in a very good site appearing to provide bad data, and vice versa.

10.9 Practical Considerations

Precise orbit determination is extremely complicated. Applying orbit determination methods to real world problems presents additional difficulties which compound the existing complexity of the mathematical process. I’ll discuss several aspects.

10.9.1 Data

The accuracy we obtain from processing observational data is a function of many variables including the number, type, and quality of the observations in a pass over a sensor, the location of the data within a pass, the total number of tracks, number of sensors, the geographic location of the sensors, the processing technique, etc. To obtain the best accuracy, you must trade-off the merits of each of these aspects versus their costs. Several areas are relevant.

Availability

Sensor observations will not always be available when needed. Maintenance, downtime, and tasking priority can affect the quantity of observations, and personnel actions can even influence quality and, in some cases, prevent the observations themselves from reaching the user. Unfortunately, we can't reliably predict these effects inside a very large system (like the AFSPC SSN). One solution is gathering as many types of data as possible, so we'll have a backup whenever normal procedures fail. Recall the results from Example 10-2. When we removed the obviously incorrect point, the RMS improved and the standard deviation improved slightly. If we had had more data, RMS and standard deviation *might* have both decreased.

Quantity

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 (except possibly for satellites with GPS receivers); data is often limited for satellite orbits. As we saw in Chap. 7, 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 (solve-for) parameters. Vallado and Carter (1997) show we need more data if we want to solve for sensor biases in addition to the position and velocity vectors. This is especially true for eccentric, deep space, and drag perturbed orbits. Remember the solve-for parameters "soak up" any mismodeling in the force models, so we want to have sufficient data to accurately determine the effect of perturbations, and not just the dynamic mismodeling.

In general, more distinct tracking data should give better accuracy (Central Limit Theorem). However, this assumes that the biases are identified and removed, and that the relative accuracy (weighting) of each observation is known and used in the estimation process. If the orbit solution quality 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 most modern sensors receive 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 additional sparse sets of observations. Although that could 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 many variables, as mentioned at the beginning of this section. Usually, *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 on a satellite at about 800 km altitude. Sparse data (less than five observations per pass) even from multiple sensors can actually increase this error to over 400 m for the same satellites (Phillips Laboratory, 1995). This suggests that dense observations (perhaps on the order of 50–100 per pass) can produce precise orbits (10 m to 100 m for many satellites).

Location of Observations

We should also examine the general precision of observations we can get during a satellite's pass. Figure 10-16 shows a greatly exaggerated satellite pass in which three points of the orbit are highlighted. Although Fig. 10-16 shows only one pass, many passes of low, medium, and high elevations can eliminate some of the errors resulting from using just one pass. When the satellite is very low to the horizon, a small vertical (elevation) 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 (azimuth) 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*. Combining these results, we get an *error ellipsoid* about the satellite. In general, along-track errors are greatest because of a lack of precise timing information and the uncertain nature of the local satellite environment (non-conservative forces such as drag tend to retard the satellite's motion). Cross-track errors are usually smaller, typically resulting from a sensor's misalignment. Radial errors are usually the smallest.

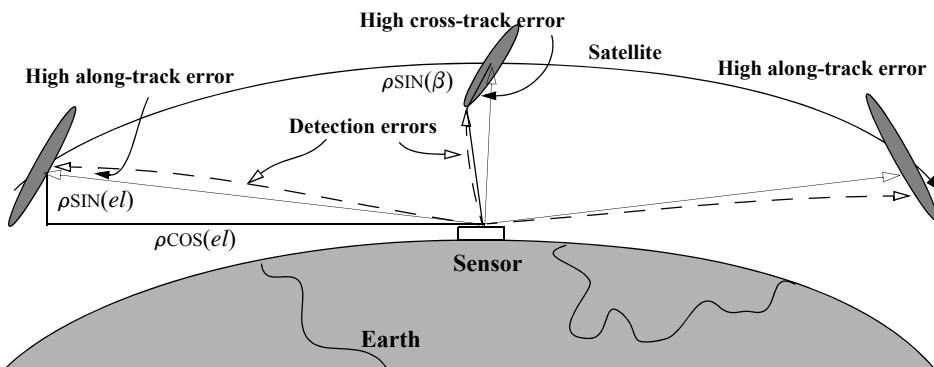


Figure 10-16. Sensor Viewing Accuracies on a Pass. Several viewing constraints affect the accuracy of the data a sensor receives. Whenever the satellite is low on the horizon, the along-track values are most difficult to resolve. Whenever the satellite is directly above the site, the cross-track errors tend to dominate.

We can see this effect mathematically by finding the vertical and horizontal components of a satellite's position in the local site's frame. Using along-track (AT) and cross-track (CT) components,

$$\text{AT} = \rho \sin(e\ell) \quad \text{CT} = \rho \cos(e\ell) \sin(\beta)$$

The partial derivatives give us the locations of the minimum and maximum values for various elevations.

$$\frac{\partial}{\partial e\ell} \text{AT} = \rho \cos(e\ell) \quad \frac{\partial}{\partial e\ell} \text{CT} = -\rho \sin(e\ell) \sin(\beta)$$

Thus, for equal angular displacements ($\Delta\alpha$ and $\Delta\beta$), the effect is largest on the along-track component at low elevation, and in the cross-track component near zenith.

Type of Data

The question often arises, “Which type of data is more accurate?” Admittedly, range data is usually preferred. As we saw in Chap. 7 (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 in the orbit estimation process, and the best solution. We can estimate a trade-off between range and angular measurements using Fig. 9-14. 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 estimate.

Observability Considerations

An ideal combination of observations is to have simultaneous range (range rate and angular measurements) data from multiple locations. The resulting triangulation is very accurate, but unfortunately, this situation rarely occurs. It’s more common to use multiple sensors to obtain observations at different times. This enhances the *observability*, and gives geometric dispersion to the observations. **Observability** indicates the ability to uniquely determine the state from the observations. It’s related to the sensitivity of the state parameters to the observations. We can evaluate observability by testing the rank of the information matrix ($\mathbf{H}^T \mathbf{H} > 0$). The partial derivatives in the \mathbf{H} matrix indicate the observability for each measurement. Partials which vanish, or have small magnitudes, provide no information to the estimation process and can impede it. We should eliminate them because they can affect the final answer and add only to the processing load. This sensitivity analysis requires no actual observation measurements. We evaluate 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 are very insensitive to any out-of-plane motion. Despite having a continuous set of very dense observations from a single site, there is often insufficient observability for accurate Orbit Determination. Filters can sometimes measure observability by noticing that processing additional measurements lowers the variances.

10.10 Applications

Specific program design for orbit determination can be very involved, and is beyond the scope of this book. However, a few applications warrant general discussion. Some of these are proven areas with working solutions, while others have the potential to improve performance by using alternate approaches.

10.10.1 Detecting Maneuvers

We've examined maneuvers in detail in Chaps. 6 and 7. However, maneuvers can present great difficulties for differential-correction techniques especially for only passive tracking. It's unlikely that observational data is available on a satellite *during* a maneuver (other than GPS or accelerometer data), but if the fit span has 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 refine the initial estimate. Post calibration is necessary to assure the maneuver went as planned. Unknown maneuvers present many additional issues to consider. There are two fundamental questions: When did the maneuver occur (***maneuver detection***), and what was the size of the maneuver (***maneuver recovery***)? Several factors influence answers to these questions.

To determine when a maneuver occurs, both LS and filter approaches can determine the times, but with differing results. For LS approaches, the uncertainty from the processing will be increased due to the maneuver within the fit span. The residuals may begin to show increased values. However, the length of the fit span will make subsequent refinement more difficult because the fit span masks individual observation contributions. The usual approach is to move the fit span back until the uncertainty increase diminishes. Filters have an advantage in processing each observation sequentially. However, the noise of the observations may be such that a bad observation could be mistaken for a post maneuver observation. Complicating both approaches is the observation density. If a 3 minute maneuver occurs in a period where there are no observations for 2 days, precise location will be very difficult!

The covariance from an OD process may also be misleading. The weighting matrix \mathbf{W} and filter (\mathbf{R}) are assumed to be constant throughout the OD operation. Thus, observations that are rejected will have no affect on the covariance (uncertainty), and even if a few observations are accepted while most are being rejected, the covariance can actually shrink!

Other factors influence the determination of when a maneuver occurs, but they also affect the ability to recover a maneuver. The observation type and sensor/satellite geometry affect the observability. A single sensor observing a GEO satellite from approximately the same longitude ensures ambiguous results. If you have very dense observations, but poor observability, a maneuver can be missed.

The cadence of a satellites' maneuvers is important as well. Suppose a satellite performs 3 maneuvers per day, but observations exist only every other day. Precisely modeling each maneuver is very difficult. The length and magnitude of the maneuver (impulsive, finite, continuous) have a similar effect. Continuous thrusting is usually handled by inserting process noise, but precise mathematical modeling is better.

Most solutions address a subset of maneuver types and situations. Some perform well on these cases. However, a single universal approach may not be feasible given the many factors influencing the process.

10.10.2 Predicting Reentry

Reentering objects like Skylab (July 11, 1979), MIR (March 23, 2001), and ISS(?) cause great concern as large fragments can often survive reentry. Two issues are relevant—when and where will it reenter? When is a function of satellite lifetime, which seeks to determine when the orbital elements will be perturbed enough that the perigee dips low enough into the atmosphere that it reenters. This process is heavily dependant on the atmospheric model chosen, and the predicted solar flux. Klinkrad (2009) provides an excellent background into the problem and Oltrogge and Chao (2007) provide a unique method of using actual solar flux data, but in a realistic simulation sense.

The second problem of where the reentry will occur is related to the first. When the perigee drops sufficiently, atmospheric drag imparts enough slowing on the satellites orbit the it will reenter. However, the size, orientation, and shape of the satellite can cause lift which makes reentry location less precise. Satellite break-up can also occur, and individual pieces can move in differing directions. Finally, any fuel remaining on the satellite can explode, further increasing the number of particles reentering. The Chinese ASAT test demonstrated an intentional collision with high momentum—resulting in 1000's of pieces of debris. While many particles were created during the event, they mostly decayed within a few years due to atmospheric drag effects.

To adequately solve the reentry problem, force models will generally need a lift state because the satellite may produce lift depending on its altitude, orientation, and velocity.

Hujšak and Dicello (1999) show a Kalman filter approach to a stressing reentry problem for a Molnyia satellite (SSC# 13390) in March 1995. They used available Space Surveillance Network tracking data to recreate the scenario. The Molnyia satellite provides additional difficulties because of the high perigee velocities during reentry. The filter automatically processed all the SSN data (normal sparse tasking and increased tasking at end of life) from November 16, 1994, to March 13, 1995. Note that calibration data (noise and bias statistics) were not provided for the study and additional tracking data of DMSP satellites was used to establish basic calibration statistics. Also, the original direction cosines from NAVSPASUR were no longer available and the resulting computation of range introduces enough additional error to render them useless. Thus, only the angular NAVSPASUR observations were available for the study. Because NAVSPASUR data was the primary data source before reentry, the reduced accuracy compounded an already difficult problem. As the satellite approached reentry, batch least squares approaches required fit spans that were only about 3 revolutions because the orbit was changing so fast. Unfortunately, the data in 3 revolutions is often not adequate for solution convergence. The filter successfully processed through several data gaps of over 7 days during which the satellite's apogee dropped 3000 km.

The satellite broke up during the reentry. The first processing predicted reentry on March 14, 1995. Subsequent tracking showed a piece still orbiting. After re-initialization, the filter predicted the final reentry on March 28, 1995. The study concluded that the Jacchia models were unsatisfactory for the problem, while the MSIS-90 model was acceptable, and actually required for successful completion of the analysis. Lift states were included in the simulation to achieve these specialized results.

Problems

1. Using the Gaussian probability density function [Eq. (10-6)], derive

$$E\{x\} = \bar{x} \quad E\{(x - \bar{x})^2\} = \sigma^2$$

2. What are the terms of the \mathbf{F} matrix (analytical technique) for a state space of position and velocity vectors including solar-radiation pressure? (Hint: see Chap. 8 for a discussion of the equations of motion)
3. Derive analytical variational equations for two-body plus J_2 perturbed motion.
4. 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. 9.6.2 and list the trade-offs for each factor that affect atmospheric drag.
5. Design a least-squares technique for the angles-only method of orbit determination (Sec. 7.3.2). 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

6. Sec. 10.4.1 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?
7. 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 propaga-

tion 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/semanalytical)
 - 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, the fit spans
 - g) Use covariance and schedule optimization to improve tasking
 - h) Improve the covariance by scaling it, consider parameters, track weighting, etc.
 - i) Improve the star catalog
 - j) Build new sensors or close older ones
 - k) Take more or fewer observations per pass
 - l) Use a Kalman-filter technique with a mathematically derived process noise
 - m) 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
 - n) Do something else
8. Use the observations from Kaena Point on January 29, 1995 for GEOS-III in Table 10-2 to find how the state vector and covariance changes as eight more observations arrive, after the answers calculated in Example 10-4. 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.
9. Given the general nature of the position uncertainty resulting from a smoother in a Kalman Filter solution (Fig. 10-13), discuss which state vectors (in time) should perform the best in a prediction sense, and why. How does this compare to a Batch Least Squares solution?

836

ORBIT DETERMINATION AND ESTIMATION

10.10

CHAPTER 11 MISSION ANALYSIS

- 11.1 Introduction
- 11.2 Mission Orbits
- 11.3 Geometries for Surveillance and Reconnaissance
- 11.4 Designing and Maintaining Mission Orbits
- 11.5 Navigation—the Global Positioning System
- 11.6 Predicting Satellite Look Angles
- 11.7 Performing Close Approach Analyses

11.1 Introduction

Military applications in space began when the former 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 of balloons and the 20th century’s development of aircraft. But applying such platforms to space poses interesting and new challenges to military planners. The former Soviet Union’s attempt to transmit pictures from the far side of the Moon was to be a statement of military capability. Although Luna 1 was launched on January 2, 1959, it crashed into the Moon and pictures weren’t sent back from the far side of the Moon until the Luna 3 transmission on October 7, 1959.

This chapter provides insight into some practical applications which may be useful when putting together “back-of-the-envelope” programs. We’ll examine *specialized orbit design* (Secs. 11.4 and 11.6) and *applications* (Sec. 11.7) and interplanetary fundamentals. The section on specialized orbits focuses on the analysis required to craft and maintain 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. The mission-applications algorithms are useful for analyzing orbital motion and characteristics related to 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 mission that uses these techniques is surveillance of the Earth, using satellites such as Landsat, RADARSAT, and SPOT.

Satellite mission design and analysis tie together many of the concepts presented in this book. Figure 11-1 shows the interactions.

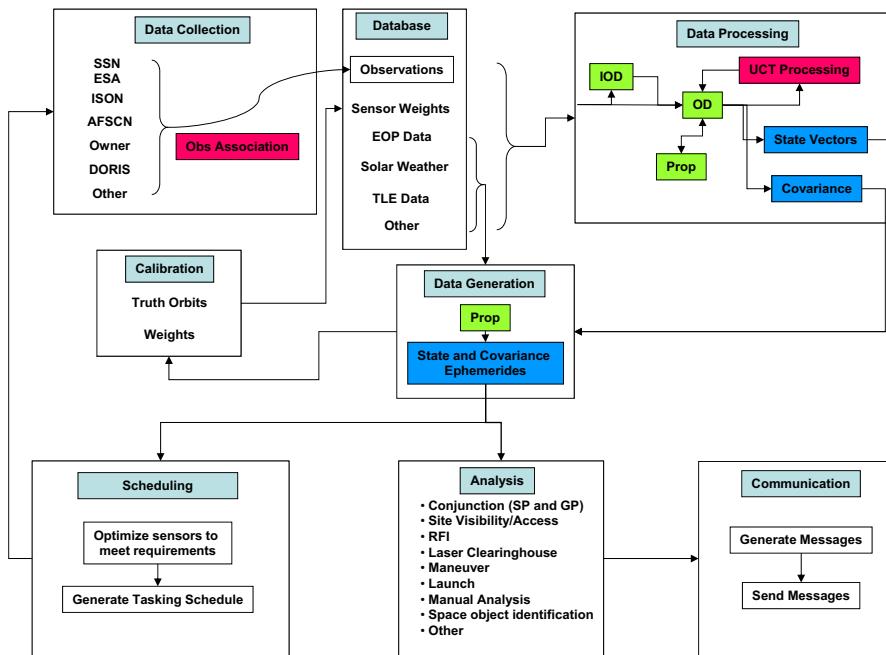


Figure 11-1. Operational Flight Mechanics Processing. This figure shows the interrelations of the components needed to maintain and produce a space object catalog, and to perform the associated surveillance and analysis tasks. Note the relationships between initial orbit determination (IOD, Chap. 7), propagation (Chaps. 8 and 9), and orbit determination (OD, Chap. 10). Time and coordinate systems (Chap. 3) ensure consistency among all the operations. The tasks and subsequent communication are certainly more complicated than shown because the focus is on the initial steps, rather than the final result.

Figure 11-1 also shows the flow of operations for a flight dynamics system. The processing begins by collecting observations and basic data. Initial orbit determination is shown as an option before orbit determination depending on the availability of *a priori* satellite data. Because we consider orbit determination to be the routine processing of observations, we need an initial estimate of the orbit (either from IOD or a previous OD). Once we determine the state vector (or orbital elements), we can use propagation techniques to create an ephemeris for each satellite. To complete the cycle, we create look angles for a sensor to reacquire the satellite at some future time. This requires finding the precise times when a satellite will rise and set from a particular sensor's field of

view, so we can take observations. This problem is the *rise/set* or *prediction* problem and it's useful in determining where to look for future observations. Many other analyses are possible. We may want to determine if any close approaches between two satellites will occur. As the number of satellites increases and their associated hardware (debris) accumulates in space, precise methods of determining close approaches are necessary to safeguard payloads and astronauts. The problem is generally known as ***close-approach determination***. Although some basic theories have existed for over 20 years, increased computational power and new techniques are changing the way we address this problem.

11.1.1 Operational Processing

Implementing an operational program for space mission operations is very challenging. Some common terminology in use today references ***Space Situational Awareness*** (SSA) to imply the knowledge of how, what, where, when and why things are in orbit. I prefer to refine that definition so that SSA is the process by which an organization maintains orbital state and mission information on all objects in space to some level of precision, now, and at a future time. The challenges imposed by any comprehensive solution to provide SSA (or whatever name is used in the future) have been unresolved for decades. Vallado and Alfano (1999) capped several years of study on the operational space picture and methods to improve and initiate a system that could support the requirements listed above. I summarize our findings below because our analysis and observations are still valid today.

We'll focus much of this discussion on the US military space surveillance system—the Joint Space Operations Center (JSPOC), formerly known as the Cheyenne Mountain Operations Center (CMOC). First, we introduce the organization and nomenclature used by AFSPC as it can be confusing. Vallado et al. (2006) extracted some history of AFSPC (See also Weeden and Cefola 2010 for additional information)

"The original North American Air Defense Command (NORAD) Combat Operations Center has evolved into the Cheyenne Mountain Operations Center (CMOC). The original requirement for an operations center in Cheyenne Mountain was to provide command and control in support of the air defense mission against the Soviet manned bomber threat ... In the early 1960s, the advent of an Intercontinental Ballistic Missile (ICBM) attack against North America became a top priority. Missile warning and air sovereignty were the primary missions in the Mountain throughout the 1960s and 70s. During a brief period in the mid 1970s, the Ballistic Missile Defense Center was installed within the Mountain. In 1979, the Air Force established a Space Defense Operations Center [SPADOC] to counter the emerging Soviet's anti-satellite threat. The evolution continued into the 1980s when Air Force Space Command [AFSPC] was created and tasked with the Air Force Space mission In April 1981, Space Defense Operations Center crews and their worldwide sensors, under the direction of Air Defense Command [ADC], supported the first flight of the space shuttle Oct. 1, 2002 marked the welcoming of two new commands, U.S. Northern Command and U.S. Strategic Command, to Cheyenne Mountain. CMOC is responsible for providing support to USNORTHCOM's mission of homeland defense and USSTRATCOM's mission of space and missile warning, formerly associated with U.S. Space Command.

Today, the Cheyenne Mountain Complex is known as Cheyenne Mountain Air Force Station (CMAFS). CMAFS is host to four commands: North American Aerospace Defense Command (NORAD), United States Northern Command (USNORTHCOM), United States Strategic Command (USSTRATCOM), and Air Force Space Command (AFSPC). CMOC serves as the command center for both NORAD and USNORTHCOM. It is the central collection and coordination center for a worldwide system of satellites, radars, and sensors that provide early warning of any missile, air, or space threat to North America. Supporting the NORAD mission, CMOC provides warning of ballistic missile or air attacks against North America, assists the air sovereignty mission for the U.S. and Canada, and if necessary, serves as the focal point for air defense operations to counter enemy bombers or cruise missiles. In addition, CMOC also provides theater ballistic missile warning for U.S. and allied forces. In support of USSTRATCOM, CMOC provides a day-to-day picture of precisely what is in space and where it is located. Space control operations include protection, prevention, and negation functions supported by the surveillance of space. Operations are conducted in seven centers manned 24 hours a day, 365 days a year. The centers are the Air Warning Center, Missile Correlation Center, Space Control Center [JSPOC], Operational Intelligence Watch, Systems Center, Weather Center, and the Command Center ... The Joint Space Operations Center (JSPOC) supports United States Strategic Command (USSTRATCOM) missions of surveillance and protection of U.S. assets in space. The JSPOC's primary objective in performing the surveillance mission is to detect, track, identify, and catalog all man-made objects orbiting earth. The JSPOC maintains a current computerized catalog of all orbiting man-made objects, charts preset positions, plots future orbital paths, and forecasts times and general location for significant man-made objects reentering the Earth's atmosphere."

The organizational names identify whether they are joint or not. For generic applications, we use DoD because this is the broadest identification for all the organizations. We generally use NORAD when referring to the TLE data because their formation was begun under NORAD. Today, the JSPOC [a.k.a. Space Control Center and formerly CMOC] produces the TLE data using a batch least squares approach, but we sometimes retain the historical name due to its familiarity in the community. Regulations and documentation have generally come from AFSPC and are identified as such.

We've all noticed an increase in the commercial use of space. The loss of Galaxy 4 (SSC 22694) on May 19, 1998 and the ensuing credit card, television, and cellular communication difficulties made national news. The International Space Station (ISS) requires precise knowledge of all objects to avoid costly (and perhaps unnecessary) maneuvers. These trends call for accurate and timely positional information for all satellites, as well as debris.

After reviewing the concepts of Chapter 10, you should realize that a batch Least Squares mathematical approach constrains the solution of certain operations. The fundamental dependence on fit spans is a limitation, and it may negatively affect the results. Batch Least Squares techniques are sometimes compensated with add-on features like consider parameters, track weighting, fading memories, etc, but these are all approximate solutions to solve the truly dynamic physical principles at work. The relative sparsity of observations and reduced force model availability in processing creates a compounding effect on the accuracy that is achievable. True statistical measures require

several orders of magnitude larger sample populations—something that just isn't available with routine AFSPC/SSN observational data.

A Kalman filter can use mathematically derived, rigorous principles to avoid some of these shortfalls, but predictions may not be as accurate as Least Squares approaches. In general, filters work better with higher sampled data rates, and batch Least Squares techniques are often optimized for sparser data. The combination of both techniques enables optimal performance for all orbit types.

Figure 11-1 shows the general flowchart of the major activities associated with designing and operating an operational space catalog. The following characteristics are needed to meet any operational system supporting SSA—complete and robust, timely and efficient, standardized and maintainable, accurate, and trusted. Trust may be the hardest to achieve as it cannot be guaranteed through simple regulations or enforcement of technical standards.

Although space surveillance lies at the heart of many operational activities in space (planning, routine and crisis operations, and engagements), use of academic advancements have traditionally lagged within operational systems. For instance, Telstar 401 (SSC 22927) failed in January 1997 and is adrift in the GEO belt threatening many other satellites. Likewise, Galaxy 15 (SSC 28884) failed in April 2010 and was adrift in the GEO belt until December 2010. Satellite owners whose satellites are projected to come close or collide with these satellites face a decision about maneuvering. If they maneuver because the error box is 10 km (for instance), but the actual encounter is 350 m, precious maneuver fuel is wasted because of the larger error box. Research has shown significant over-simplification in the existing operational planning (Alfano, 2006), but these advances have not yet made it into some operational systems. In many cases, it's likely that the attitude control engines could be enough to minimize the cross-sectional area and avoid a collision, provided that sufficient warning time and data accuracy are available. Unfortunately, this level of accuracy is not generally available.

Complete / Robust

Operations such as conjunction processing must use a complete catalog. Consider doing conjunction processing with only 80% of the space catalog. The results would not be very useful if a satellite in the remaining 20% had a conjunction in the near future! We can categorize the complete catalog as follows. Satellites are known, mistagged, uncorrelated, classified, or lost.

Mistagged satellites (assigning one satellite name to another satellite) occur routinely in systems that try to passively monitor satellites that maneuver often, such as JSPOC trying to maintain the GEO catalog. Satellite clusters are often mistagged, and warnings sent out for Astra 1H (for instance), could actually be for Astra 1KR when in actuality, the two satellites are over 50 km apart!

If we lose track of a satellite (from either an unknown maneuver, faulty predictions, or a new launch), but then re-observe the satellite at a later time, it's called an ***Uncorrelated Target*** (UCT, track is also sometimes used for observation tracks that cannot be attributed to any known satellite) because the new satellite doesn't match any known satellite and we can't associate new observations with existing orbits.* Most sensors are

tasked to observe a particular satellite. When a sensor tracks an object, but it's not the intended satellite, we have an uncorrelated track and initial orbit determination techniques and manual intervention are usually required to determine the new orbit. Khutorovsky (1996) remarks that "the errors in predicted orbits are the only cause of miscues in the assignment of sensor taskings," underscoring the need for accurate tasking techniques. A large number of UCTs are an indication of a larger problem though—the accuracy, architecture, and concept of operations in a space surveillance/flight dynamics system. The astrodynamics community recognized that we can perform significantly better (at least several orders of magnitude improvement in accuracy) using existing resources (parallel processing, numerical propagation and dense observations) at the former Phillips Laboratory (PL) Astrodynamics Division workshop in December 1993.

Note that UCT satellites may quickly become *lost* if the sensors are no longer able to obtain tracks on the satellite. Because a majority of observations are taken in a tracking sense (as opposed to surveillance), this can happen quite easily.

Classified satellites present another problem when trying to establish and maintain a complete catalog. Satellites that one country considers classified are not necessarily considered classified by another. In the past, some countries operated on the notion that satellites could be "hidden". In May 2010, the media reported that "General James Cartwright, vice chairman of the Joint Chiefs of Staff, said it was no longer possible for the United States and other countries to keep vast numbers of orbiting satellites a secret." (Space-travel, 2010)

A robust system must be able to use all forms of data for processing. For instance, the system should be able to accept SSN, Russian Space Surveillance System (RSSS), Air Force Satellite Control Network (AFSCN), ESSS, ISON, SLR, Global Positioning System (GPS), University, space-based (MSX, SBSS), bi-static, transponder, DORIS, and other observational data. It should also be able to process all types of data (radar, optical, laser, interferometer, transponder, etc.)

There needs to be true surge capability in the system, including some redundancy needed for weather, mechanical, and other outages. Additional capacity can be gained by examining the tasking protocols. Task only important objects (don't spend valuable sensor time tracking a piece of debris in a relatively "stable" orbit and don't track a satellite which is independently maintained). This requires collaborative arrangements with the owner operators—and this requires trust!

Timely / Efficient

Data must be available in a timely manner. In the past, communication links typically constrained the volume of data that could be sent to users. While this is largely gone today, some limitations still exist. The Internet moves vast quantities of data between worldwide locations, however, its use in some systems is often very limited.

* Don't confuse UCTs [targets] with ***observation association***, or the process of assigning observations to a known satellite. Unassociated observations usually result in a UCT that is processed as future observations are gathered on it.

Efficiency is difficult to design into a system, especially when the system has evolved over many years from disparate original functions. The power of the modern computer makes decentralized execution a firm reality, although one that has seen limited acceptance in operational systems to date.

Tasking involves determining which sensors will be assigned to track which satellites. Although there have been numerous studies examining the tasking algorithms to look at how to effectively allocate sensor resources to a large population of space objects, few have examined the problem separated from existing system architectures. As a result, progress has been limited. Currently, tasking at JSPOC is handled by a table of priority values. If additional accuracy is needed, the priority is moved to the next highest level (Wilson, 2004). While this works for most situations, it is not the most efficient nor flexible process.

Another tasking approach would propagate the covariance into the future (assuming of course that the covariance is realistic). There are studies on the positive aspects of batch Least Squares and Kalman filter approaches and the use of an accurate covariance matrix seems to be an obvious choice for efficiently addressing the tasking problem (Hill et al. 2010). For each satellite, the point at which the covariance error exceeds the accuracy requirement would trigger a tasking request. The actual time would need to be before this point to ensure no mission degradation beyond the requirement. For satellites having precise owner operator control, little to no tracking would be required as cooperative arrangements could establish the exchange of data.

Any tasking process needs coupling with an optimization routine for maximum efficiency. Several constraints are required for the optimization to work effectively. In particular, the type of sensor, the observability, probability of the tasking actually being executed, etc. While the specific optimization details are beyond our current scope, optimizing several dozen sensors against a catalog of about several ten-thousand satellites is not particularly challenging for modern techniques.

Standardized / Maintainable

Ultimately, the issue of maintainability rests with the specific software implementation and the computing hardware. This suggests we be very cautious when implementing new systems and ensure we can port the existing code to a new system (computer hardware) as requirements dictate.

For almost all modern systems, the importance and availability of computer code, test cases, and documentation is critical. Software must be updated and thoroughly evaluated, tested, documented and researched by a truly independent analysis and possibly by other organizations wishing to use the code. Organizations that declare certain codes to be a standard without a truly independent evaluation miss the opportunity to improve their code by testing, and incorporating other experience and design lessons. To assume the code is in no need of peer review is questionable. Sharing computer code is not very popular because no computer code is perfect and there may be proprietary issues.

Open source software initiatives have the potential to break down some of these barriers.

Upgrades should be designed for future growth. The dependence on one type of processing (batch Least Squares) is extremely limited. Including a Kalman Filter approach may be beneficial for well-tracked, or high interest satellites, thereby augmenting the capabilities of an existing system.

Finally, international standards should be adhered to even though compliance is voluntary. This will ensure compatibility with users. There are several efforts underway dealing with astrodynamics standards in the international arena (Finkelman and Oltrogge 2006). The AIAA recommended practice (ANSI/AIAA 2010) contains useful information about implementing propagation techniques.

Accurate

The accuracy of a space catalog stimulates significant debate. A simple solution is to be flexible to provide whatever accuracy the user requires. A fundamental question is—*what observations and processing are needed to achieve a certain level of accuracy on a particular satellite, now, and at a future time?* Note the use of a “particular” satellite. Broad accuracy assignments and values are simply not useful for precise operations such as conjunction processing.

Measuring the accuracy ultimately involves post-processing analysis of the results and examination of the covariance resulting from the estimation. The issue of a realistic covariance matrix cannot be overemphasized, and in fact, it has received considerable discussion over the last few years. In reality, only two alternatives exist and have been proven to be effective. Batch Least Squares techniques produce a covariance matrix as a result of the estimation process and the additional state parameters included—solve-for parameters, track weighting, etc. This technique has been used for many years, but may experience difficulties when higher accuracy is required, and non-conservative forces introduce significant error into the solutions. Using a Kalman Filter that incorporates mathematically derived process noise has the advantage of not being constrained by the fit span, and the limitations that imposes on dynamic variables in the solution. However, the Kalman filter solution may not model the mean long term behavior in prediction as well as batch Least Squares methods, although various fading memory approaches can be used to adjust bias and drift uncertainty.

Trusted

Trust is a key to establishing any operation, and this is especially important in space. However, it is not generally associated with astrodynamical operations. Regulations and rigid procedures can dictate and direct, but trust is earned over a period of time, usually coinciding with openness and independent verification of results and techniques. This is absent in some large scale operations today. Baron (2005) remarks

The ability to trust is often thought of as a subjective trait subject to irrational assessments such as familiarity and name recognition. Trust, however is not completely subjective, nor completely bipolar; that is we do not necessarily either trust or distrust. Sometimes, we can trust a little, or trust a lot, or trust individuals or organizations in some aspects but not others. Trust, therefore, can more rationally and objectively be treated as a risk assessment, unique for each case. Software is one of the most important products used in aerospace, yet

one of the most difficult to assess. Managers can be prone to trust organizations with whom they have had previous dealings or are familiar without a complete assessment of risk and other products.

For astrodynamics, sharing observations and test data is not yet a common occurrence, and trust may be lacking in many areas. The SLR and GPS communities have provided much useful data for researchers. High quality reference ephemerides exist at the Center for Space website (www.centerforspace.com/downloads). These ephemerides provide a comprehensive set of data for general comparison and use.

11.1.2 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. Recognize that the distributions in the following figures change over time.

Figure 11-2 shows the distribution of existing satellite orbits by inclination (in December 2012). The large concentration of orbital inclinations in the 60° to 110° range is because this region allows satellites to observe and communicate with large regions of the Earth. The band occurring near 100° is mainly Sun-synchronous orbits used by Earth observing satellites. A smaller group represents the equatorial geosynchronous satellites.

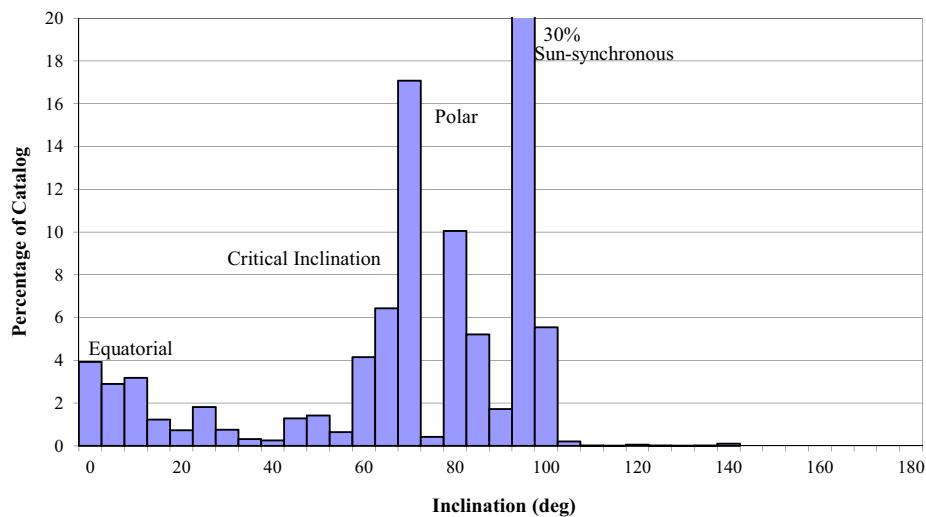


Figure 11-2. Inclination of Earth Satellite Orbits. This figure shows how satellite orbits are distributed based on inclination. Notice the large concentration of satellites near the critical inclination, i_ω , as well as the polar and Sun-synchronous regions.

The eccentricity distribution is less varied 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 Fig. 11-3 shows that more than 9% of satellites are in moderately to highly eccentric orbits (e greater than 0.1).

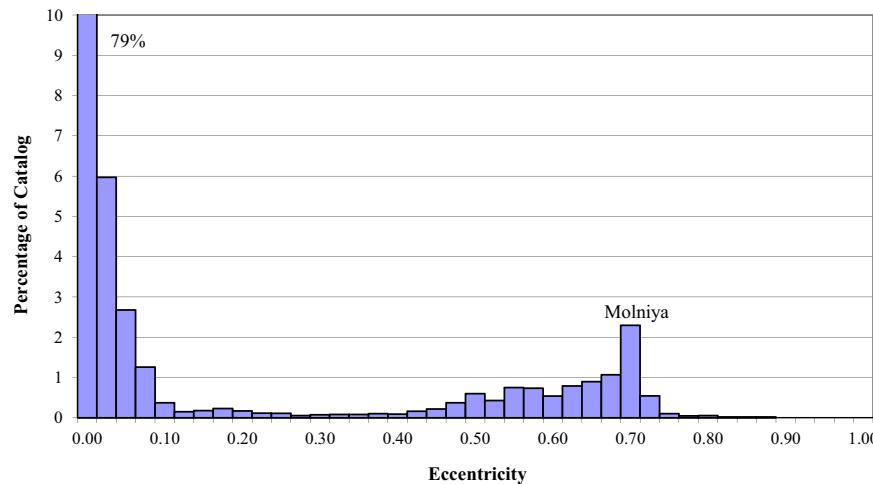


Figure 11-3. Eccentricity of Earth Satellite Orbits. This figure shows the satellite population distributed by eccentricity. Although many satellites are still in roughly circular orbits, about 9% of the catalog has $e > 0.1$. I limit the plot to 10% for clarity in the remaining groups.

We can obtain information about the period and altitude by examining the values and distribution of satellite mean motions. Figure 11-4 shows this distribution. Notice that the plot shows equivalent values for mean motion, altitude, and period.

Finally, we can group satellites into classes determined by the apogee and perigee altitudes. These groups are sometimes called ***Gabbard classes*** (Kaya and Snow, 1999). Figure 11-5 shows such a snapshot of the space catalog in December 2012.

11.2 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 Table 11-1 and representative orbits that satisfy each one. Many of these specialized orbits are required in later sections.

Communication satellites are an integral part of our society. Few imagined the explosion of television, telephone, the Internet, and other 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 storm and temperature pre-

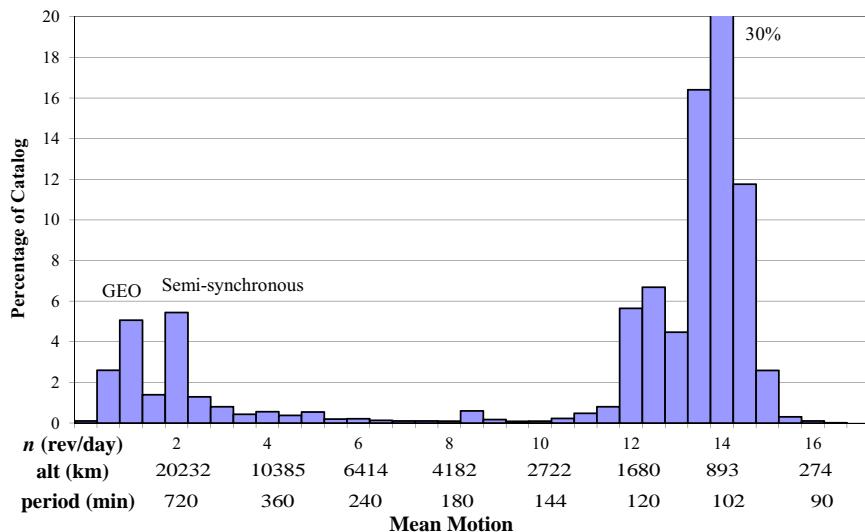


Figure 11-4. Mean Motion of Earth-Orbiting Satellites. We determine two important quantities from mean motion (n): the orbital period and semimajor axis. The circular orbital altitude for each mean motion is also shown.

dictions. TIROS 1 was the first meteorological satellite—launching on April 1, 1960. **Navigation** satellites are helping redefine how we determine any location on, or above 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, missile and launch 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, our environment, and outer space.

Each of these missions uses different orbits, as shown in Table 11-1. The specialized orbits contain **low-Earth** (LEO) to **geosynchronous** (GEO) orbits. The types of low-Earth satellites are varied, whereas geosynchronous orbits are nearly circular and nearly equatorial, with periods of one sidereal day. **Semi-synchronous** orbits have orbital periods of 12 hours and are especially useful in navigation satellites.

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 on the solar panels, or certain ground lighting conditions. **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

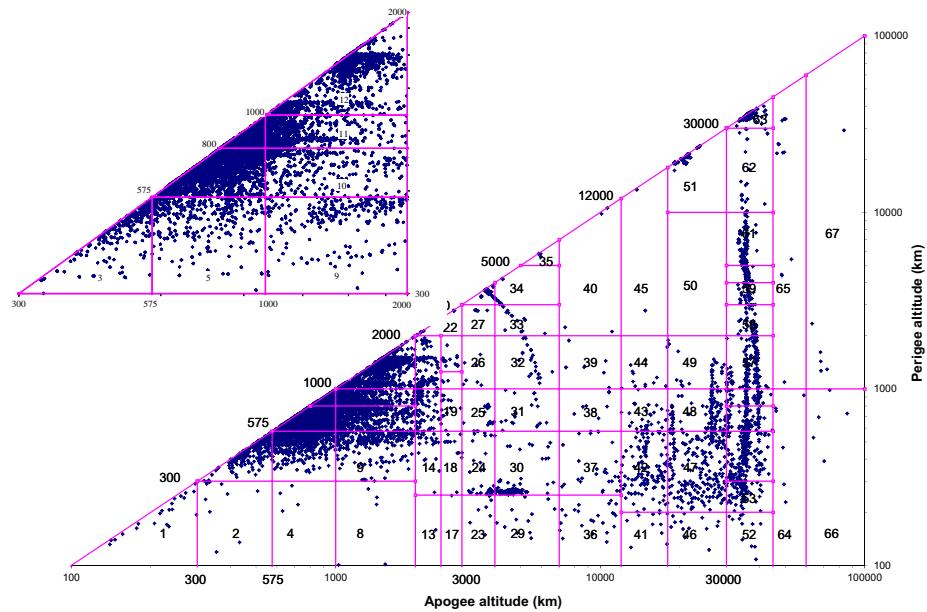


Figure 11-5. Categorizing Satellite Orbits. This figure shows the distribution of satellite orbits. The categories are numbered sequentially from the left (1), proceeding to the right, and then vertically up. Apogee and perigee altitudes values (km) are also shown. The inset shows additional details for the LEO region.

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.

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 specialized 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, we must understand the present and future distribution of satellite orbits when planning operational software. As shown in Table 11-2 and Fig. 11-6, objects in space are either payloads (active and inactive) or debris. **Debris** includes rocket bodies, platforms, cables and devices used to secure and release satellites for deployment, pieces of satellites, and dead satellites. The disparity between the orbiting and reentered

TABLE 11-1. 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, C = Constellations, FF = Formation Flying.)

Mission	Satellite Orbit			Specialized Orbits / "Frozen" Parameter					Grouping		Examples			
	LEO	MEO	GEO	SS	RG*	MA	O	i	ρ_Ω	$e + \omega$	varies	C	FF	
Communication	X	X	X				X					X		Intelsat, Inmarsat, SES, Comstar, Molniya
Navigation		X	X									X		GPS, GLONASS, Galileo
Military †	X	X	X	X	X	X						X	X	DSCS, DMSP, TDRS, Milstar, DSP
Weather		X		X	X									GOES, NOAA, TIROS
Earth Research	X		X	X	X							X		TOPEX, GEOSAT, GRACE, LAGEOS
Earth imaging	X	X		X								X	X	Landsat, SPOT, RADARSAT
Scientific	X		X	X		X	X					X		Explorer, Hubble, GEOS

* Periodic maneuvers are required to maintain repeat groundtrack orbits.

† Military missions include communication, reconnaissance, weather, navigation, detection, and others.

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 and end-of-life disposal has become an important issue. There is also considerable debate about the actual size of the space object catalog (50-100 thousand?) because sensors can detect only certain sized objects. Thus, debris models are only estimates. A popular model is the European Space Agency's Meteoroid and Space Debris Terrestrial Environment Reference (MASTER) model which periodically estimates what the future debris population might look like under different scenarios (e.g., Release 1.0 was in April 2006).

TABLE 11-2. Orbiting and Decayed Satellites. The table shows several countries (as of December 2012) who operate in space and their contribution to its population.

Country	Orbiting			Reentered		
	Payloads	Debris	Total	Payloads	Debris	Total
Former Soviet Union	1464	3868	6328	2483	9094	14322
United States	1168	3186	5002	901	4016	5538
People's Republic of China	146	3565	3784	59	665	821
France	56	308	496	8	612	685
Japan	135	47	224	30	141	230
European Space Agency	50	38	95	10	15	32
Other	733	191	972	72	392	487

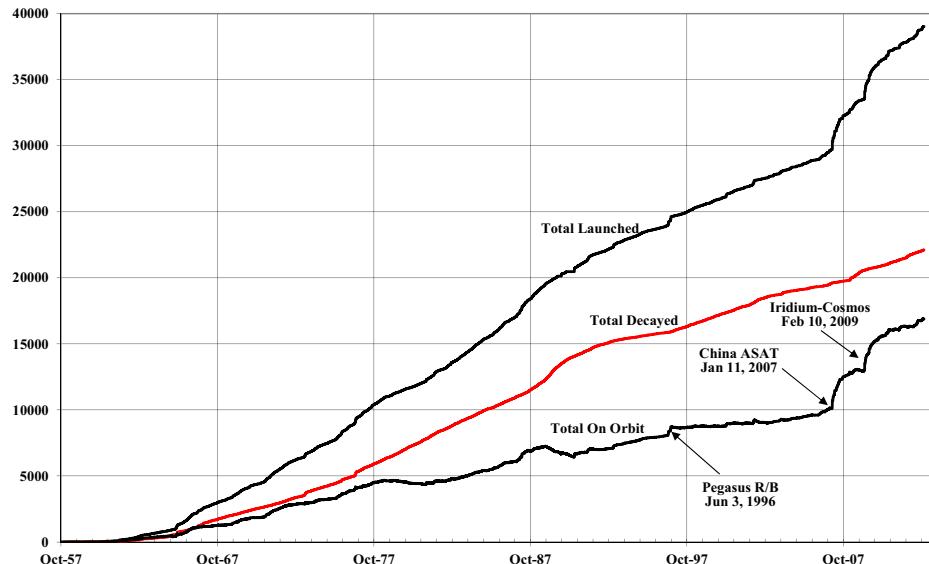


Figure 11-6. Satellite Population Growth. This figure shows the satellite population growth since the first launch. Note the amount of decayed objects. Although accounting for drag is a challenge to model accurately, drag reduces the satellite population through decay and eventual atmospheric reentry. (as of December 2012)

11.2.1 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 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. In addition, Chao (2005) presents a more in depth look at operational maneuvers and planning.

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 (May 14, 1973 to July 11, 1979), Mir (February 19, 1986 to March 23, 2001), and ISS (November 20, 1998 to 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 would reenter and be destroyed.

There are many options for stationkeeping on low- and mid-Earth-orbit satellites. First, resupply often isn't practical, so considerable effort goes into designing orbits which require little adjustment due to the precise balancing of perturbations. One option 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.

Maintaining stable equilibrium locations for geosynchronous satellites is important because many satellites are "co-located" in these orbits. The operation is often called ***east-west stationkeeping*** because we try to maintain the longitude for the satellite. This is particularly useful for communication satellites. We want to design the orbit (if possible) to keep the stationkeeping velocity as low as possible. The qualitative analysis on page 665 (resonance) has shown that under the influence of the $C_{2,2}$ 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. 9-8 or Fig. 9-9. 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. 9, the effects of third-body perturbations cause a long-periodic motion in the inclination for geosynchronous satellites. The compensation process is usually called ***north-south stationkeeping*** because the perturbation results in changes that are normal to the equator. For north-south stationkeeping, Chao (2005:62) indicates that the period is about 26 years to complete a half period in the inclination of about 15° . The satellite gains about 1° of inclination per year for ten years, and then it slows to reach 15° after about 16 additional 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 out of the equatorial plane several years ago are now coming back towards $i = 0^\circ$, increasing the possibility of collision with newer satellites. This process points to the importance of accurately tracking satellites and developing end-of-life strategies.

11.2.2 Mission Planning

A significant amount of time and money is spent analyzing a satellite orbit for a particular mission. I'll briefly introduce three primary phases, concept exploration, research and development, and operation.

The concept exploration phase attempts to match mission objectives with available technology and expected funding. A first step is the evaluation of potential orbits for the particular mission. Entire texts are devoted to these trade spaces [e.g., Larson and Wertz (1999), Wertz et al. (2011), or Griffen and French (1991)], so I won't cover those topics. However, once the basic orbital parameters are selected, simulations are often performed to assist the planning and sizing for the spacecraft. These activities aid in determining how much fuel to carry, how much solar power will be needed, etc. This material is covered in the texts mentioned previously, but there often exists a need for a reasonably accurate simulation of the long term behavior of the orbits. This is especially true for formation flying satellites and constellations. Some analytical methods, and the semianalytical techniques mentioned in Chap. 9 are well-suited for this task. Several studies have employed techniques like the DSST to accomplish these goals [Smith (1999), Shah (1997), Wallace (1995), Sabol (1994), and others]. In addition, Brumberg and Brumberg (1999) present compelling arguments for continued use of analytical and semianalytical techniques to gain insight into the behavior of orbits over time.

Research and development is the process of actually building the spacecraft. From an astrodynamics perspective, the focus is usually on the computer codes used in the ground processing system, and with the satellite itself. The discussions throughout this text apply (commonality, well-defined coordinate systems, etc.) and the importance of computer programming becomes evident. Computer programming practices are mentioned in Appendix D.

Finally, operations play an important part in ensuring that the mission is a success. An obvious objective is to operate the vehicle, but this can also include incorporating modified or new algorithms. Calibration and periodic independent checks are also necessary to ensure optimum performance. In addition, anomaly resolution is required to have a plan of action in case of an unexpected event, such as a power loss, data glitch in the computer, component failure, etc.

11.3 Geometries for Surveillance and Reconnaissance

We can view surveillance and reconnaissance topics when the observing takes place *from space*. The distinction between surveillance and reconnaissance is subtle, but distinct. ***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. 11-7 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.

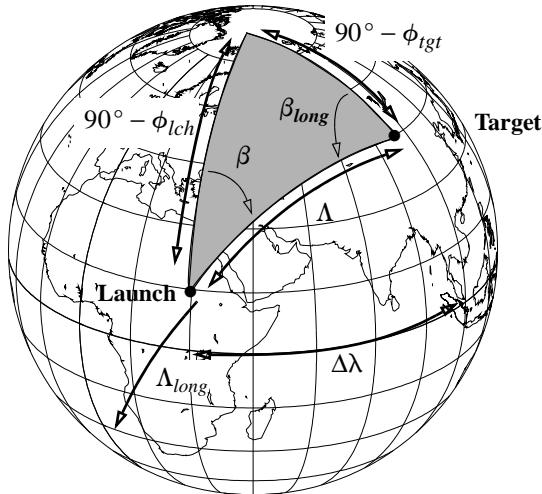


Figure 11-7. Earth Geometry. We can develop several formulas from the spherical triangles formed by two locations on the Earth. Of primary interest are the azimuth, β , and the ground range to the target, Λ . These values are unique for both the short and long way to the target. The launch and target latitudes are ϕ_{lch} and ϕ_{tgt} .

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-24) and written using the ground range angle,

$$\Lambda_{km} = R_{\oplus} \Lambda_{radians} \quad (11-1)$$

applies to this problem. Remember, this relation is for a spherical Earth. Although I've shown kilometers in Eq. (11-1), 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. (8-14):

$$\cos(\Lambda) = \sin(\phi_{tgt}) \sin(\phi_{lch}) + \cos(\phi_{tgt}) \cos(\phi_{lch}) \cos(\Delta\lambda) \quad (11-2)$$

This formula always gives the answer for the *short way* to the target or the acute angle (less than 180°). From the law of sines for spherical triangles,

$$\frac{\sin(\beta)}{\sin(90^\circ - \phi_{tgt})} = \frac{\sin(\Delta\lambda)}{\sin(\Lambda)}$$

Solving for Λ results in an equation we can use with Eq. (11-2) in ATAN2 to resolve the quadrant.

$$\sin(\Lambda) = \frac{\sin(\Delta\lambda) \cos(\phi_{tgt})}{\sin(\beta)} \quad (11-3)$$

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

$$\cos(\beta) = \frac{\sin(\phi_{tgt}) - \sin(\phi_{lch}) \cos(\Lambda)}{\cos(\phi_{lch}) \sin(\Lambda)} \quad (11-4)$$

Then, let's find a sine expression using the sine law, as in the development of Eq. (11-3):

$$\sin(\beta) = \frac{\sin(\Delta\lambda) \cos(\phi_{tgt})}{\sin(\Lambda)} \quad (11-5)$$

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:

$$\sin(\phi_{tgt}) = \cos(\beta) \cos(\phi_{lch}) \sin(\Lambda) + \sin(\phi_{lch}) \cos(\Lambda) \quad (11-6)$$

The longitude relies on solving Eq. (11-3) for the change in longitude. We can resolve angles without quadrant checks if we find the cosine expression by solving Eq. (11-2):

$$\begin{aligned} \sin(\Delta\lambda) &= \frac{\sin(\beta) \sin(\Lambda)}{\cos(\phi_{tgt})} \\ \cos(\Delta\lambda) &= \frac{\cos(\Lambda) - \sin(\phi_{lch}) \sin(\phi_{tgt})}{\cos(\phi_{lch}) \cos(\phi_{tgt})} \end{aligned} \quad (11-7)$$

Otherwise, we must specify the direction (long or short) to determine longitude. 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. 11-8). 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.

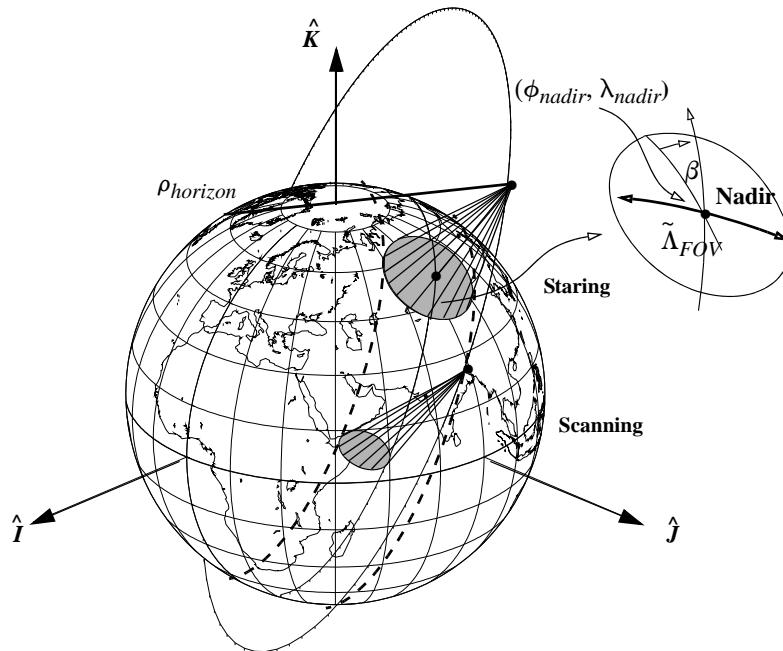


Figure 11-8. Geometry for Field of View (FOV). The FOV problem consists of achieving mission objectives within a satellite sensor's limit. An arbitrary 1500 km radius circle about the staring-satellite's subpoint depicts this concept. Over an interval of time, the scanning sensor can view the same area as the staring sensor, but it can't continually observe the entire region of the staring sensor.

The general problem associated with calculating the field of view is determining how much of the Earth we can see, given certain parameters about the viewing satellite. Sev-

eral specific cases arise, and we'll explore how to describe the geometries. Consider the basic geometry shown in Fig. 11-9 and remember we're still using a spherical Earth.

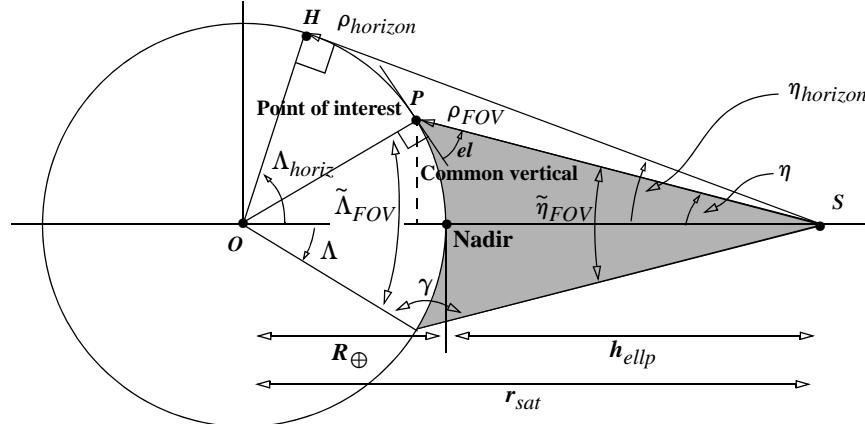


Figure 11-9. Basic Geometry for Field of View—Nadir Pointing. We require several parameters to determine the location of the Earth's horizon and the ground-range angle for a point of interest, P . From the nadir point defined directly beneath a satellite, S , at an altitude of h_{ellp} , measure the 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 12 in Chap. 3. Next, we determine the various boresight angles: specifically, the horizon from the satellite, $\eta_{horizon}$, the maximum displacement the sensor can view, ($\eta \approx \tilde{\eta}_{FOV}/2$), or the location of another satellite, η . We find the horizon-boresight angle using Fig. 11-9. From the right triangle shown, we can directly determine

$$\sin(\eta_{horizon}) = \frac{R_\oplus}{r_{sat}} \quad (11-8)$$

The sine expression is sufficient because the angle will never exceed $\pm 90^\circ$.*

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:

$$\rho_{horizon} = r_{sat} \cos(\eta_{horizon}) \quad (11-9)$$

To find another expression, let's use the Pythagorean theorem on the right triangle (*SHO*) in Fig. 11-9 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. 11-9. Now, find the intermediate angle, γ , using the sine law for oblique triangles:

$$\sin(\gamma) = \frac{r_{sat} \sin(\eta)}{R_\oplus} \quad (11-10)$$

Remember to use 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-23).

$$\rho = R_\oplus \cos(\gamma) + r_{sat} \cos(\eta) \quad (11-11)$$

As a check, the slant-range value is always between h_{ellp} and $\rho_{horizon}$, unless we're observing another satellite. Solve for the *general boresight angle*, η , using Eq. (C-21) (See Fig. 11-10).† This gives

$$\cos(\eta) = \frac{\rho^2 + r_{sat}^2 - R_\oplus^2}{2\rho r_{sat}} \quad (11-12)$$

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, $R_\oplus + h_{ellp}$.

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

* 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 often because of mechanical and electrical design limits.

† Wertz (1978:87) shows an alternate formula.

$$\tan(\eta) = \frac{R_\oplus \sin(\Lambda)}{R_\oplus + h_{ellp} - R_\oplus \cos(\Lambda)} = \frac{\sin(\eta_{horizon}) \sin(\Lambda)}{1 - \sin(\eta_{horizon}) \cos(\Lambda)}$$

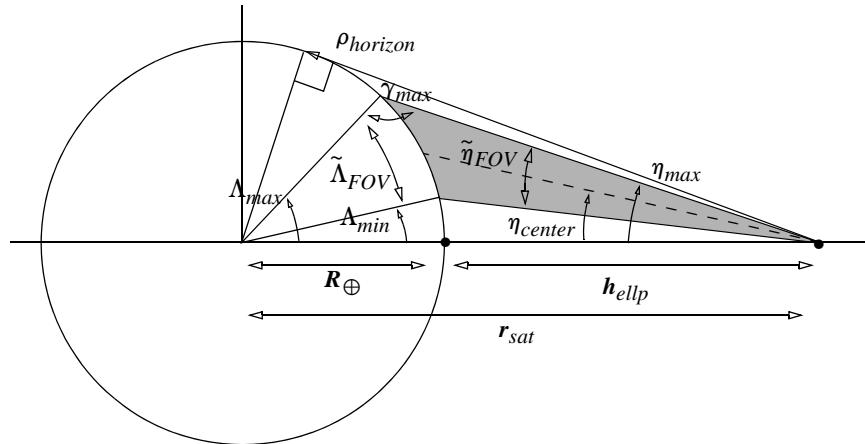


Figure 11-10. Geometry for Field of View—Oblique Pointing. This shows a scanning sensor, as well as another angle, η_{center} , we need to calculate ground ranges for oblique pointing. When η_{center} is zero, the diagram reverts to the geometry of Fig. 11-9.

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. (11-13). Then use β to find the particular latitude and longitude from Eq. (11-6) and Eq. (11-7), using the spherical triangle in Fig. 11-7. 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, $\tilde{\Lambda}_{FOV}$, that corresponds to a given boresight angle, $\tilde{\eta}_{FOV}$. Let's begin by finding the intermediate angle from Eq. (11-10) for the given boresight angle. Next, find the slant range to the satellite using Eq. (11-11). The common vertical point of the field-of-view angles in Fig. 11-9 permits you to use the sine law and find the ground range from the nadir point:

$$\sin(\Lambda) = \frac{\rho \sin(\eta)}{R_\oplus} \quad (11-13)$$

The more general case occurs for a scanning sensor, as shown in Fig. 11-10. 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. 11-10, we can do two calculations using the center boresight angle, η_{center} , to determine maximum and minimum boresight angles, and ultimately, maximum and mini-

mum ground-range angles. The difference in ground-range angles ($\Delta_{max} - \Delta_{min}$) is the value we seek.

Let's look at an example to show how we can use these relations.

▼ Example 11-1: Calculating Field of View.

GIVEN: Satellite in a circular orbit ($i = 60^\circ$) at an altitude of 800 km. $\tilde{\eta}_{FOV} = 25.0^\circ$. The current nadir location is 50°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

$$r_{sat} = 6378.1363 + 800 = 7178.1363 \text{ km}$$

First, determine initial values to bound the problem. Begin with the horizon-boresight angle:

$$\sin(\eta_{horizon}) = \frac{R_\oplus}{r_{sat}} = \frac{6378.1363}{7178.1363}, \eta_{horizon} = 62.6917^\circ$$

We can also determine the ground range that is visible to the sensor by

$$\cos(\Lambda_{horiz}) = \frac{R_\oplus}{R_\oplus + h_{ellp}} = \frac{6378.1363}{7178.1363} \Rightarrow \Lambda_{horiz} = 27.3083^\circ$$

Therefore, the horizon is $(27.3083)(\pi/180)(R_\oplus) = 3040$ km away from the nadir point. It's also useful to find the maximum slant range so you can check later calculations.

$$\rho_{horizon} = r_{sat} \cos(\eta_{horizon}) = 7178.1363 \cos(62.6917) = 3293.178 \text{ km}$$

Now, determine the maximum locations that the sensor can see. Find the ground-range angle by first finding the intermediate angle, γ . Because you're looking directly down ($\eta_{center} = 0^\circ$), use half the complete field of view angle:

$$\sin(\gamma) = \frac{r_{sat} \sin(\eta)}{R_\oplus} = \frac{7178.1363 \sin(12.5)}{6378.1363}, \gamma = 14.09836^\circ, 165.90164^\circ$$

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 the answer is just slightly larger than the orbital altitude of 800 km.

$$\begin{aligned} \rho &= R_\oplus \cos(\gamma) + r_{sat} \cos(\eta) = 6378.137 \cos(165.90164) + 7178.1363 \cos(12.5) \\ &= 821.97 \text{ km} \end{aligned}$$

Now find the ground range from the satellite subpoint:

$$\sin(\Lambda) = \frac{\rho \sin(\eta)}{R_\oplus} = \frac{800.0 \sin(12.5)}{6378.1363}, \Lambda = 1.5984^\circ \Rightarrow 177.9290 \text{ km}$$

You'll always use the smaller value because a satellite can't observe more than a 90° arc of the Earth's surface. Simply subtract or add the ground-range angle to the satellite nadir's location to find the locations of minimum and maximum latitudes:

$$\phi_{tgtmax} = 51.5984^\circ, \phi_{tgtmin} = 48.4016^\circ, \lambda = 40^\circ\text{W}$$

If you need the east-west variations, you simply 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. The maximum and minimum boresight angles are:

$$\eta_{max} = \eta_{center} + \frac{\tilde{\eta}_{FOV}}{2} = 52.5^\circ, \text{ and } \eta_{min} = \eta_{center} - \frac{\tilde{\eta}_{FOV}}{2} = 27.5^\circ$$

Use these values ($\eta = \eta_{max}, \eta_{min}$) and calculate the following parameters. Be careful to check quadrants.

$$\sin(\gamma) = \frac{r_{sat} \sin(\eta)}{R_{\oplus}} : \text{choose both angles } > 90^\circ, \gamma_{max} = 116.764^\circ 82', \text{ and } \gamma_{min} = 148.690^\circ 21'$$

$$\rho = R_{\oplus} \cos(\gamma) + r_{sat} \cos(\eta) : \rho_{max} = 1497.51 \text{ km, and } \rho_{min} = 917.796 \text{ km}$$

$$\sin(\Lambda) = \frac{\rho \sin(\eta)}{R_{\oplus}} : \Lambda_{max} = 10.735^\circ 17' (1195 \text{ km}), \text{ and } \Lambda_{min} = 3.809^\circ 77' (424 \text{ km})$$

The total ground range is the difference between the maximum and minimum ranges:

$$\Lambda_{max} - \Lambda_{min} = 6.9254^\circ = 770.9307 \text{ km}$$

Notice that the total ground range is larger than 356 km (twice 177.9 km in the $\eta_{center} = 0^\circ$ case) for the observation beneath the satellite. In fact, the 770 km groundrange is really an instantaneous value. If the sensor can move $\pm 40^\circ$, the total potential ground range is $\Delta_{FOV} = \Lambda_{max}(2) = 1195(2) = 2390$ km. You determine the ground locations the same as before, but because the center of the field-of-view area 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 [azimuth, Eq. (6-12)] the sensor is looking.

$$\sin(\beta) = \frac{\cos(i)}{\cos(\phi_{gc})} = \frac{\cos(60)}{\cos(50)}, \beta = 51.0652 \text{ or } 128.9348^\circ$$

You can determine the correct value by testing the cosine of the argument of latitude [Eq. (6-13)]. Use the larger value when $\cos(u)$ is less than zero. In this case, $u = 62.1964^\circ$, so $\beta = 51.0652^\circ$. Remember that this is the azimuth for the direction of the satellite. The sensor can be aimed in many directions, depending on its specific design. Let's use 140° . Now use Eq. (11-6) to find the center that the sensor is observing. Be sure to find Λ using the process above because Λ and η are not simply additive. ($\eta = 40^\circ$, $\gamma = 133.6632^\circ$, $\rho = 1095.197$ km, $\Lambda = 6.3368^\circ$).

$$\sin(\phi_{ctr}) = \cos(\beta) \cos(\phi_{nadir}) \sin(\Lambda) + \sin(\phi_{nadir}) \cos(\Lambda)$$

$$\sin(\phi_{ctr}) = \cos(140^\circ) \cos(50^\circ) \sin(6.3368^\circ) + \sin(50^\circ) \cos(6.3368^\circ)$$

$$\phi_{ctr} = 44.9926^\circ$$

From Eq. (11-7),

$$\sin(\Delta\lambda) = \frac{\sin(\beta) \sin(\Lambda)}{\cos(\phi_{ctr})} = \frac{\sin(140^\circ) \sin(6.3368^\circ)}{\cos(44.9926^\circ)} \quad \Delta\lambda = 5.7577^\circ$$

$$\lambda_{ctr} = 45.7577^\circ$$

The locus of points from the center is found using half the distance of the minimum and maximum ranges ($\Lambda = 6.9254 / 2$). The maximum and minimum latitudes occur at $\beta = 0$ and 180° .

$$\phi_{tgtmax} = 48.4553^\circ, \text{ and } \phi_{tgtmin} = 41.5299^\circ$$

For an azimuth within the FOV of $\beta = 60^\circ$

▲ $\phi_{tgt} = 46.6423^\circ, \text{ and } \lambda_{tgt} = 50.1721^\circ$

11.3.1 Ground Distance on the Ellipsoid

The previous section used a spherical Earth to model distances on the ground. However, we know that the Earth has a significant bulge and is not spherical. Nguyen and Dixson (2003) present an iterative method that solves for the ground range on an ellipsoidal Earth. Vincenty, (1966, 1975) also presents a method, reproduced here. For the direct approach to find the ending location (ϕ_{gd2}, λ_2) given a starting location (ϕ_{gd1}, λ_1) , azi-

muth (β) and distance (Λ_{ell}), begin by finding temporary variables and let $u = f^2 \cos(\beta)$ where f is the flattening of the Earth.

$$A = 1 + \frac{u^2}{16384} \{ 4096 + u^2(-768 + u^2(320 - 175u^2)) \}$$

$$B = \frac{u^2}{1024} \{ 256 + u^2(-128 + u^2(74 - 47u^2)) \}$$

Iterate the following equations until the difference in Λ is small (e.g., 1×10^{-6}). Vicenty uses three angular distances on a sphere in the calculations. Λ_m is the angular distance from the equator to the midpoint of the line, Λ_1 is the angular distance from the equator to the starting point, and Λ is the angular distance between the two points. Use $\Lambda = \Lambda_{ell}/b \oplus A$ as a first guess.

$$2\Lambda_m = 2\Lambda_1 + \Lambda$$

$$k = \cos(2\Lambda_m) + \frac{B}{4} \left(\cos(\Lambda)(-1 + 2\cos^2(2\Lambda_m)) - \frac{B}{6} \cos(2\Lambda_m) \right. \\ \left. (-3 + 4\sin^2(\Lambda))(-3 + 4\cos^2(2\Lambda_m)) \right)$$

$$\Delta\Lambda = B \sin(\Lambda) k$$

$$\Lambda = \frac{\Lambda_{ell}}{b \oplus A} + \Delta\Lambda$$

Now find the final result for the ending location. Note the use of reduced latitude (see page 140).

$$\tan(\phi_{gd2}) = \frac{\sin(\phi_{rd1}) \cos(\Lambda) + \cos(\phi_{rd1}) \sin(\Lambda) \cos(\beta)}{(1-f) \sqrt{\sin^2(\beta) + (\sin(\phi_{rd1}) \sin(\Lambda) - \cos(\phi_{rd1}) \cos(\Lambda) \cos(\beta))^2}}$$

$$\tan(d\lambda) = \frac{\sin(\Lambda) \sin(\beta)}{\cos(\phi_{rd1}) \cos(\Lambda) - \sin(\phi_{rd1}) \sin(\Lambda) \cos(\beta)}$$

$$C = \frac{f}{16} \cos^2(\beta) (4 + f(4 - 3\cos^2(\beta)))$$

$$\Delta\lambda = d\lambda - (1 - C)f \sin(\beta)(\Lambda + C \sin(\Lambda)[\cos(2\Lambda_m) + C \cos(\Lambda)(-1 + 2\cos^2(2\Lambda_m))])$$

$$\lambda_2 = \lambda_1 + \Delta\lambda$$

Vincenty also shows the reverse formulation, finding the azimuth and distance given the starting and ending locations. Begin with an initial guess of $d\lambda = \Delta\lambda$. The longitude is recovered at the end of the following equations using the C and $\Delta\lambda$ expressions from the first transformation. These equations are iterated until the longitude stops changing.

$$\begin{aligned}\sin^2(\Lambda) &= \{\cos(\phi_{rd2})\sin(\lambda)\}^2 + \{\cos(\phi_{rd1})\sin(\phi_{rd2}) - \sin(\phi_{rd1})\cos(\phi_{rd2})\cos(\Delta\lambda)\}^2 \\ \cos(\Lambda) &= \sin(\phi_{rd1})\sin(\phi_{rd2}) + \cos(\phi_{rd1})\cos(\phi_{rd2})\cos(\Delta\lambda) \\ \sin(\beta) &= \frac{\cos(\phi_{rd1})\cos(\phi_{rd2})\sin(\Delta\lambda)}{\sin(\Lambda)} \\ \cos(2\Lambda_m) &= \cos(\Lambda) - \frac{2\sin(\phi_{rd1})\sin(\phi_{rd2})}{\cos^2(\beta)}\end{aligned}$$

Using $\Delta\Lambda$ from the first transformation,

$$\begin{aligned}\Lambda_{ell} &= b_{\oplus}A(\Lambda - \Delta\Lambda) \\ \tan(\beta_1) &= \frac{\cos(\phi_{rd2})\sin(\Delta\lambda)}{\cos(\phi_{rd1})\sin(\phi_{rd2}) - \sin(\phi_{rd1})\cos(\phi_{rd2})\cos(\Delta\lambda)}\end{aligned}$$

11.4 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. 9 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. 9.8.1) because it often requires accurate predictions for the distant future.

As discussed in Table 11-1, 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.

11.4.1 Sun-Synchronous Orbit

The first specialized orbit we'll consider is the ***Sun-synchronous*** orbit, which maintains a constant orientation towards the Sun over time. 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. 11-11.

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. (9-37) which gave the secular perturbations for the right ascension of the ascending node as a result of J_2 perturbation. In one year (365.242 189 7 days, see page 4), the Earth completes 360° around the Sun.

$$\frac{d\Omega}{dt_{desired}} = \dot{\Omega}_{SunSyn} = \frac{360^\circ}{365.242 189 7} = 0.985 647 36 \frac{\text{deg}}{\text{day}} = 0.017 202 79 \frac{\text{rad}}{\text{day}}$$

$$\dot{\Omega}_{SunSyn} = 1.991 063 853 \times 10^{-7} \frac{\text{rad}}{\text{sec}} = 1.606 378 006 \times 10^{-4} \frac{\text{rad}}{\text{TU}}$$

Substituting this value into Eq. (9-37) results in

$$\frac{\dot{\Omega}_{SunSyn}}{n} = \dot{\Omega}_{SunSyn} \sqrt{\frac{a^3}{\mu}} = -\frac{3R_\oplus^2 J_2}{2a^2(1-e^2)^2} \cos(i)$$

$$a^{7/2} = -\frac{3R_\oplus^2 J_2 \sqrt{\mu}}{2\dot{\Omega}_{SunSyn}(1-e^2)^2} \cos(i) \quad (11-14)$$

Be careful of the units in this relation! Also notice the relation between the semimajor axis and the inclination.

To complete the analysis of Sun-synchronous orbits, we solve Eq. (11-14) for the other parameters. Thus, if we are given semimajor axis and eccentricity, or semimajor axis and inclination, we can solve for inclination, or eccentricity, respectively.

$$i = \cos^{-1} \left(\frac{-2a^{7/2}\dot{\Omega}_{SunSyn}(1-e^2)^2}{3R_\oplus^2 J_2 \sqrt{\mu}} \right) \quad (11-15)$$

$$e = \sqrt{1 - \sqrt{\frac{-3R_\oplus^2 J_2 \sqrt{\mu} \cos(i)}{2a^{7/2}\dot{\Omega}_{SunSyn}}}} \quad (11-16)$$

We can examine the various factors to select a Sun-synchronous orbit. Figure 11-12 shows the combinations of semimajor axis, eccentricity, and inclination. The solution of Eq. (11-14) shows the semimajor axis (altitude), eccentricity, and inclination for a Sun-synchronous orbit

In practice, we solve Eq. (11-14) 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

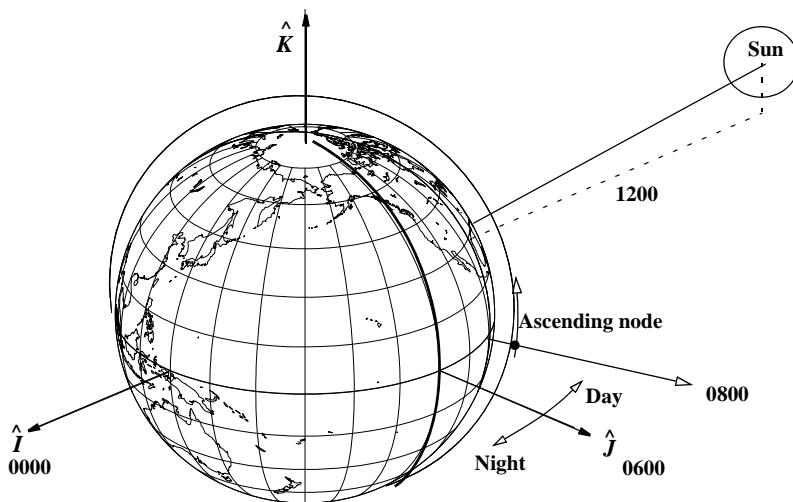


Figure 11-11. Geometry for Sun-Synchronous Orbits. We usually use a nomenclature based on the time of the ascending node to distinguish Sun-synchronous orbits. The orbit shown is an 0800 orbit because the satellite reaches the ascending node at 8:00 A.M. local time. I show the Sun offset from the equatorial plane for generality.

value, thus minimizing altitude variations from perigee to apogee without the need for special stationkeeping burns to maintain the eccentricity. The eccentricity is often chosen to be near-zero. Therefore, the problem is one of determining the inclination given other orbital parameters.*

Now consider an example calculation based on RADARSAT [Jane's (1994:385)].

▼ **Example 11-2. Calculating a Sun-synchronous Orbit.**

GIVEN: The Canadian government launched the RADARSAT satellite on November 4, 1995 to monitor the polar icecap's motion. A Sun-synchronous orbit ($e = 0$) 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.

$$a = 800.0 + 6378.1363 \text{ km} = 7178.1363 \text{ km}$$

* The other orbital element to choose is the right ascension of the ascending node. The node lets us choose an orientation such as maintaining the solar panels perpendicular to the Sun (0600, 1800 orbits). Note that these times are local times for the ascending node [usually], and they are often mean times (α_{FMS} from pg 183). From Fig. 11-11, if the Sun is over the J -axis and we desire a noon ascending node, Ω would simply be 90° , or α_{\odot} . Letting τ_{node} be the number of hours from noon (\pm) for the desired node, and using a general solar position, use $\Omega = \alpha_{\odot} + \tau_{\text{node}} + 12^\circ$. Thus for the 0600 orbit, the offset $\tau_{\text{node}} = -6^\circ$.

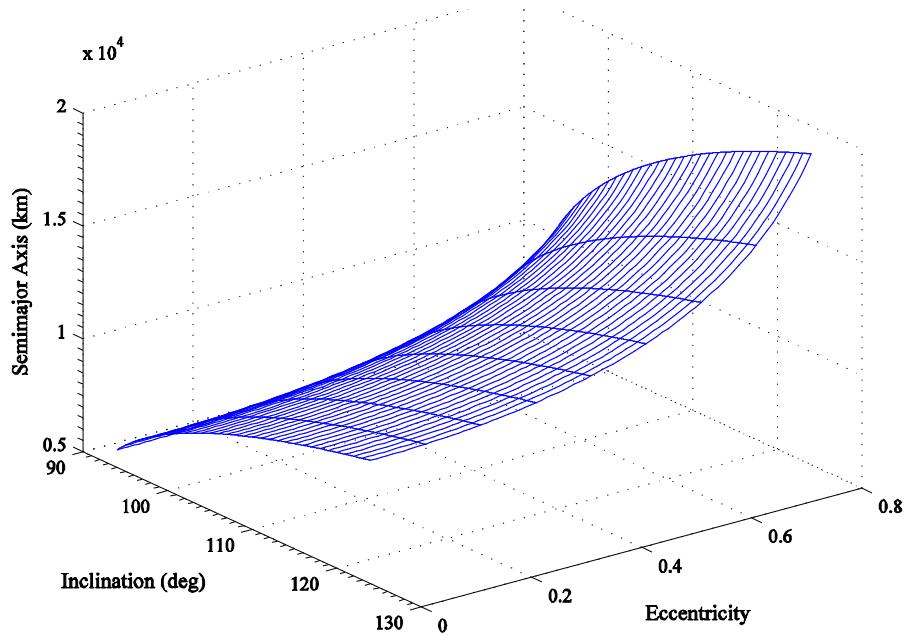


Figure 11-12. Elements for Sun-Synchronous Orbits. This figure shows that choosing two orbital elements from i , a , and e determines the third element needed for a Sun-synchronous orbit. The semimajor axis and radius of perigee must be greater than the Earth's radius to be realistic.

Remember, the rate must be in rad/s.

$$\cos(i) = -\frac{a^{7/2} \dot{\Omega}_{SunSyn} (1-e^2)^2}{3R_\oplus^2 J_2 \sqrt{\mu}} = -\frac{(1.125\ 32)^{7/2} (1.991\ 063\ 853 \times 10^{-7})}{3(6378.137^2)(0.001\ 082\ 63)\sqrt{398600.4418}}$$

$$i = 98.6^\circ$$

If you keep $i = 98.6^\circ$ and change the eccentricity to 0.2,

$$a^{7/2} = \frac{-3(6378.137^2)(0.001\ 082\ 63)\sqrt{398600.4418}}{2(1.991\ 063\ 853 \times 10^{-7})(1-0.2^2)^2} \cos(98.6^\circ)$$

$$a = 7346.846 \text{ km, or apogee and perigee altitudes of } 674.8 \text{ km} \times 1262.6 \text{ km}$$

▲ 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 on Sun-synchronous Orbits

The secular rate of RAAN (Ω_{sec}) is caused mainly by the J_2 perturbation (Fig. 9-3) 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 . Ignoring effects of a and e , the effect is given by

$$\Delta\dot{\Omega}_{sec} \equiv \frac{\partial\dot{\Omega}}{\partial i}\Delta i + O(\Delta i)^2$$

where using Eq. (9-37) yields

$$\begin{aligned}\Delta\dot{\Omega}_{sec} &\equiv \frac{3}{2}n\left(\frac{R_{\oplus}}{p}\right)^2 J_2 \sin(i)\Delta i + O(\Delta i)^2 \\ &\equiv -\dot{\Omega}_{sec} \tan(i)\Delta i + O(\Delta i)^2\end{aligned}\quad (11-17)$$

The primary drift in the inclination is not caused by J_2 , but by the Sun because the Sun-synchronous constraint causes a resonance effect due to the critical argument $\dot{\Omega} - \dot{\alpha}_{\odot} \approx 0$ (Sec. 9.5.2), or equivalently, $\dot{\Omega} - (\omega_{\odot} + \dot{M}_{\odot}) \approx 0$. 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 40. 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

$$\Delta i = \frac{1}{na} \cos(u) \Delta v_W + O(e^2) \quad (11-18)$$

Inspection of Eq. (11-18) shows that the efficiency of the Δv is maximized by thrusting at the equator ($u = 0$). This is a general form of the inclination-only maneuvers in Algorithm 39. 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. (11-17) 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

$$\Delta i(t) = \Delta i_o + \frac{di}{dt}(t - t_o) \quad (11-19)$$

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. (9-34)] 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. (11-17):

$$\Delta\Omega_{sec} = -\int_{t_o}^t \dot{\Omega}_{sec} \tan(i) \Delta i dt \quad (11-20)$$

Consistent with other approximations already made, we assume nominal (constant) values for $\dot{\Omega}_{sec}$ and $\tan(i)$. Substituting Eq. (11-19) into Eq. (11-20) and integrating,

$$\begin{aligned} \Delta\Omega &= A_1(t - t_o) + \frac{A_2}{2}(t - t_o)^2 \\ A_1 &= -\dot{\Omega}_{sec} \tan(i) \Delta i_o \\ A_2 &= -\dot{\Omega}_{sec} \tan(i) \frac{di}{dt} \end{aligned} \quad (11-21)$$

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. (11-21) and Fig. 11-13 also show that the drift is quadratic in the time.

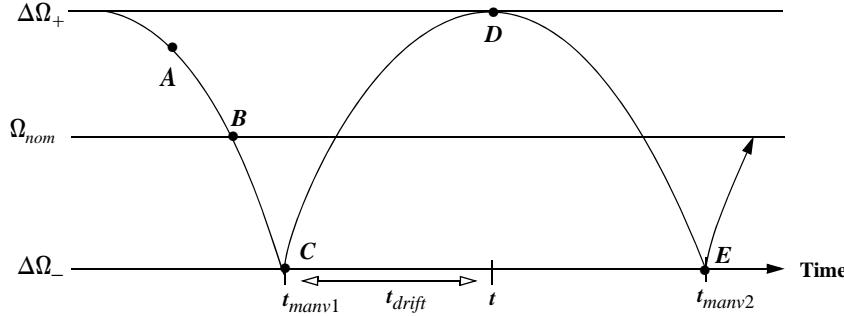


Figure 11-13. Profile of Nodal Drift. This figure shows the change in the nodal drift, $\Delta\Omega$, over time. There are usually limits placed about the nominal value (\pm) within which the mission is successful. The satellite mission begins at point A with orbit insertion. The first burn occurs at 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 $\Omega_o - \Delta\Omega$. Taking the derivative of Eq. (11-21), 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 $-\Delta\Omega$ over the interval $[t_{manv1}, t]$, so

$$-\Delta\Omega = A_1 t_{drift} + \frac{A_2}{2} t_{drift}^2 = A_1 \left(\frac{-A_1}{A_2} \right) + \frac{A_2}{2} \left(\frac{-A_1}{A_2} \right)^2 = -\frac{1}{2} \frac{A_1^2}{A_2}$$

Solving for A_1 yields

$$A_1 = \pm \sqrt{2A_2\Delta\Omega}$$

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 in Eq. (11-21):

$$\Delta i_a = \frac{-A_1}{\dot{\Omega}_{sec} \tan(i)}$$

The approximate time interval between inclination adjustments is $2t_{drift}$.

The inclination rate in Eq. (11-19) [and Eq. (11-21)] is given by the Lagrange VOP equations [Eq. (9-12)]:

$$\frac{di}{dt} = \frac{1}{na^2\sqrt{1-e^2}\sin(i)} \left\{ \cos(i) \frac{\partial R}{\partial \omega} - \frac{\partial R}{\partial \Omega} \right\}$$

Taking the dominant term in Kaula's third-body disturbing function [Eq. (9-34)] for the Sun and further restricting the indices to the resonant term of interest $(l,m,p,h,q,q') = (2,2,1,0,0,0)$ 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, $\Omega \approx \dot{M}_{\odot}$. So the rate of the cosine 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.

11.4.2 Repeat-Groundtrack Orbits

Repeat-groundtrack orbits are those which retrace their groundtrack over a certain time interval. They're used by 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 apply 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. (1-26)]. Thus we can view the anomalistic period as a perturbed or osculating Keplerian period.

$$\rho_K = \frac{2\pi}{n} = 2\pi \sqrt{\frac{a^3}{\mu}} \quad (11-22)$$

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 (ascending or descending) 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

$$\rho_{\theta_G} = \frac{2\pi}{\omega_{\oplus} - \dot{\Omega}} \quad (11-23)$$

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. (9-41)] and $\dot{\omega}$ [Eq. (9-39)] 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:

$$\rho_{\Omega} \approx \rho_K \left[1 - \frac{3J_2}{2} \left(\frac{R_{\oplus}}{a} \right)^2 (3 - 4\sin^2(i)) \right] \quad (11-24)$$

We can also maintain the full accuracy and not simplify to obtain a direct expression.

$$\rho_{\Omega} = \rho_K \left[\frac{1}{1 + \frac{3J_2}{4} \left(\frac{R_{\oplus}}{p} \right)^2 \left\{ \sqrt{1 - e^2} (2 - 3\sin^2(i)) + (4 - 5\sin^2(i)) \right\}} \right] \quad (11-25)$$

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.

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 *revolutions to repeat* (or the equator crossing points), $k_{rev2rep}$, and the *days to repeat* (or the frequency), $k_{day2rep}$, with which they occur (see Fig. 2-20 for example groundtracks). Usually, equator crossing points aren’t overflowed in contiguous order (the distinction between passes and revolutions in Chap. 2) 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 ($k_{day2rep} = 16$ d) at a spacing of roughly 172 km along the equator per pass. This required 233 crossing points because $k_{rev2rep} = 233 \approx 2\pi R_{\oplus} / 172$ 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 11-3 demonstrates this concept. The time interval between two successive arrivals at the ascending node is called the nodal period, ρ_{Ω} , in Eq. (11-24).

The spacing, or the **groundtrack shift**, $\Delta\lambda_{rev}$, 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, $\dot{\Omega}$, are measured in radians per unit time, the angular spacing (in radians) is

$$\Delta\lambda_{rev} = (\omega_{\oplus} - \dot{\Omega})\rho_{\Omega} = \frac{2\pi R_{\oplus} k_{day2rep}}{k_{rev2rep}} \quad (11-26)$$

We convert to kilometers by multiplying by the Earth’s radius. For Landsat 6, $\Delta\lambda_{rev} = 24.721\ 03^\circ$, or 2751.93 km [$2\pi R_{\oplus}(16) / 233$]. 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 $\Delta\lambda_{rev}$ gives the magnitude, the direction of the shift is westward.

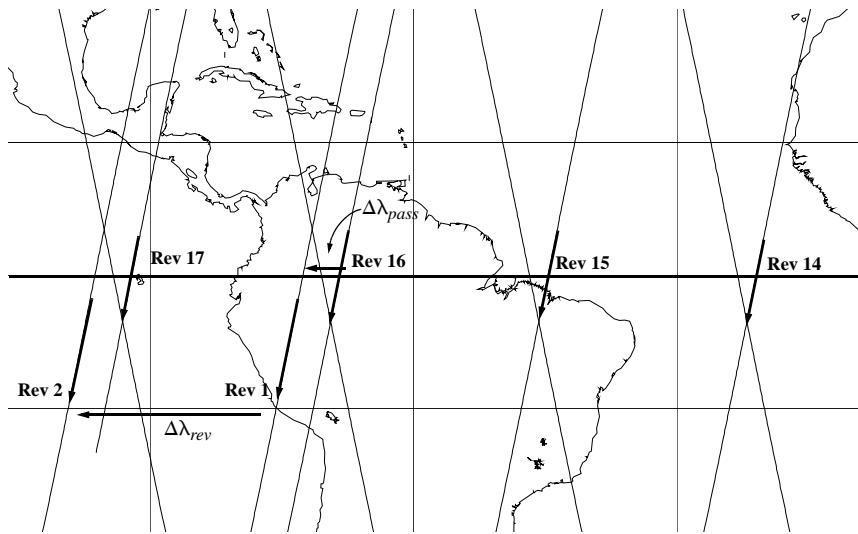


Figure 11-14. Equator Crossings for Repeat Ground Track Orbit. This shows several repeat groundtrack parameters. Be careful to distinguish between passes ($\Delta\lambda_{pass}$) and revolutions ($\Delta\lambda_{rev}$). An exact repeat groundtrack orbit will have $\Delta\lambda_{pass} = 0$ while $\Delta\lambda_{rev}$ can be thousands of kilometers.

Note that we defined the $\Delta\lambda_{rev}$ 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_Q , 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 (Fig. 11-14). This unwanted westward (or eastward, Eq. (9-37)) groundtrack shift is due to the nodal regression—a function of the inclination. That's why Eq. (11-26) 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.

We can also couple a repeat groundtrack orbit with a Sun-synchronous orbit. The process treats each individually. First, you find the semimajor axis for the repeat groundtrack orbit. The solution is then iterated until the Sun-synchronous parameters are satisfied. It's possible that the semimajor axis may change slightly during this process, but it is generally constant.

Finding a Repeat Groundtrack Orbit

Algorithm 71 shows two approaches for solution including differing perturbations. The first iterates on a to match the desired change in longitude. The second iterates on a to maintain the input revolutions per day. There are many variations for possible input values. I've shown one, but others may be derived based on available information. Note that the calculation of eccentricity is optional, as shown, in the second approach.

ALGORITHM 71: Repeat Ground Track ($k_{rev2rep}, k_{day2rep}, e, i \Rightarrow a$)

$k_{revpday} = \frac{k_{rev2rep}}{k_{day2rep}}$ $n = k_{revpday} \omega_{\oplus}$ $a_{new} = \left(\mu \left(\frac{1}{n} \right)^2 \right)^{1/3}$ LOOP $\Delta\lambda_{rev} = \frac{2\pi k_{day2rep}}{k_{rev2rep}}$ LOOP $a = a_{new}$ $p = a(1 - e^2)$ $\dot{\Omega} = -\frac{3nJ_2}{2} \left(\frac{R_{\oplus}}{p} \right)^2 \cos(i)$ $\dot{\omega} = \frac{3nJ_2}{4} \left(\frac{R_{\oplus}}{p} \right)^2 (4 - 5\sin^2(i))$ $\Delta\lambda_{period} = \frac{2\pi \dot{\Omega}}{n}$ $\Delta lon = \Delta\lambda_{rev} + \Delta\lambda_{period}$ $n = \frac{2\pi \omega_{\oplus}}{\Delta lon}$ $a_{new} = \left(\mu \left(\frac{1}{n} \right)^2 \right)^{1/3}$ END LOOP	LOOP $a = a_{new}$ $(e = e_{new}) optional$ $p = a(1 - e^2)$ $\dot{\Omega} = -\frac{3nJ_2}{2} \left(\frac{R_{\oplus}}{p} \right)^2 \cos(i)$ $\dot{\omega} = \frac{3nJ_2}{4} \left(\frac{R_{\oplus}}{p} \right)^2 (4 - 5\sin^2(i))$ $\dot{M}_o = \frac{3nJ_2}{4} \left(\frac{R_{\oplus}}{p} \right)^2 \sqrt{1 - e^2} (2 - 3\sin^2(i))$ $n = k_{revpday} (\omega_{\oplus} - \dot{\Omega}) - (\dot{M}_o + \dot{\omega})$ $a_{new} = \left(\mu \left(\frac{1}{n} \right)^2 \right)^{1/3}$ $(e_{new} = \frac{a_{new} - r_p}{a_{new}}) optional$ END LOOP
--	--

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 on 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, $\Delta\varphi_\Omega$, will cause a change in the groundtrack-shift parameter ($\Delta\lambda_{rev}$) similar to Eq. (11-26)

$$\Delta\lambda_{rev} = (\omega_\oplus - \dot{\Omega})\Delta\varphi_\Omega - \Delta\dot{\Omega}\varphi_\Omega$$

The final term is really a higher-order effect, thus we will ignore it for this section. The change in φ_Ω due to a decaying semimajor axis

$$\Delta\varphi_\Omega = \frac{\partial\varphi_\Omega}{\partial a}\Delta a + O(\Delta a)^2 \approx \frac{3\pi}{na}\Delta a + O(J_2)$$

caused mainly by the change in the Keplerian period, φ_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 ($\Delta\varphi_\Omega < 0$), 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.

For convenience, let's assume that the semimajor axis's rate of decay is constant. Also for convenience, I'll drop the subscript on φ_Ω and use φ_j to denote the nodal period at the end of each revolution. Consider the decay of the nodal period starting from the nominal value φ_o . After one revolution the value is $\varphi_1 = \varphi_o - \Delta\varphi$, where $\Delta\varphi$ 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 φ over one revolution because $\Delta\varphi = 0$ at the beginning. Thus, the accumulated change in the time of arrival (Δt) relative to the desired time is $\Delta t = \Delta\varphi/2$. With no further decay, because $\varphi_1 = \varphi_o - \Delta\varphi$, the change in the time of arrival after the next revolution would be $\Delta\varphi + \Delta\varphi/2$: the average period decay over the first revolution summed with the $\Delta\varphi$ 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, $\Delta t = \Delta\varphi + \Delta\varphi/2 + \Delta\varphi/2$. The resulting nodal period is now $\varphi_2 = \varphi_o - 2\Delta\varphi$. 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 φ_o is $2\Delta\varphi + 2\Delta\varphi = 4\Delta\varphi$. If decay takes place over this third revolution, the accumulated Δt over this revolution is $\Delta\varphi_n/2$, which we add to the $4\Delta\varphi$ to obtain $4\Delta\varphi + \Delta\varphi/2$. The relationship becomes apparent in Table 11-3. Keep in mind that the change in the satellite's time of arrival over the equator, $\Delta t_{rev} = \Delta\varphi/2$, 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 φ_o and φ is their difference, $\Delta t_{diff} = \varphi - \varphi_o$.

Thus we deduce that, if $\Delta\varphi_\Omega$ is the change in nodal period per revolution, the accumulated effect on the groundtrack-shift parameter after j revolutions is

TABLE 11-3. 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 ρ_i	$\Delta t = \Delta t_{diff} + \Delta t_{rev}$ over 1 Revolution	Accumulated Δt_{acc} from Revolution 0
0	ρ_o	0	0
1	$\rho_1 = \rho_o - \Delta\rho$	$\Delta\rho/2$	$\frac{\Delta\rho}{2}$
2	$\begin{aligned}\rho_2 &= \rho_1 - \Delta\rho \\ &= \rho_o - 2\Delta\rho\end{aligned}$	$\Delta\rho + \Delta\rho/2$	$\frac{3\Delta\rho}{2} + \frac{\Delta\rho}{2} = 2\Delta\rho$
3	$\begin{aligned}\rho_3 &= \rho_2 - \Delta\rho \\ &= \rho_o - 3\Delta\rho\end{aligned}$	$2\Delta\rho + \Delta\rho/2$	$\frac{9}{2}\Delta\rho$
j	$\rho_j = \rho_o - j\Delta\rho$	$(j-1)\Delta\rho + \Delta\rho/2$	$\sum_{k=1}^j \left((k-1)\Delta\rho + \frac{\Delta\rho}{2} \right) = \frac{1}{2}j^2\Delta\rho$

$$\sum_{k=1}^j \Delta\lambda_{rev_k} = \frac{1}{2}j^2(\omega_{\oplus} - \dot{\Omega})\Delta\rho_{\Omega}$$

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 Repeat Groundtrack 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. (11-26)]

$$\begin{aligned}
\Delta\lambda_{rev} &= \frac{\partial\Delta\lambda_{rev}}{\partial a}\Delta a + \frac{\partial\Delta\lambda_{rev}}{\partial i}\Delta i \\
\frac{\partial\Delta\lambda_{rev}}{\partial a} &= R_{\oplus}(\omega_{\oplus} - \dot{\Omega})\frac{\partial\rho_{\Omega}}{\partial a} - \frac{\partial\dot{\Omega}}{\partial a}R_{\oplus}\rho_{\Omega} \\
\frac{\partial\Delta\lambda_{rev}}{\partial i} &= R_{\oplus}(\omega_{\oplus} - \dot{\Omega})\frac{\partial\rho_{\Omega}}{\partial i} - \frac{\partial\dot{\Omega}}{\partial i}R_{\oplus}\rho_{\Omega} \\
\Delta a &= \Delta a_o + \frac{da}{dt}(t - t_o) \quad \Delta i = \Delta i_o + \frac{di}{dt}(t - t_o)
\end{aligned} \tag{11-27}$$

The average rate of the drift over one revolution is

$$\frac{\Delta(\Delta\lambda_{rev})}{\rho_{\Omega}} = \frac{1}{\rho_{\Omega}} \left[\frac{\partial\Delta\lambda_{rev}}{\partial a}\Delta a_o + \frac{\partial\Delta\lambda_{rev}}{\partial i}\Delta i_o + \left(\frac{\partial\Delta\lambda_{rev}}{\partial a}\frac{da}{dt} + \frac{\partial\Delta\lambda_{rev}}{\partial i}\frac{di}{dt} \right)(t - t_o) \right]$$

Substituting a continuous representation for the discrete representation above yields

$$\begin{aligned}
\frac{d\Delta\lambda_{rev}}{dt} &= k_1 + k_2(t - t_o) \\
\frac{d\Delta\lambda_{rev}}{dt} &= \frac{d}{dt}[\Delta(\Delta\lambda_{rev})] = \frac{d}{dt}(\Delta\lambda_{rev} - \lambda_{revnom}) \\
&= \frac{d\Delta\lambda_{rev}}{dt} - \frac{d\lambda_{revnom}}{dt} = \frac{d\Delta\lambda_{rev}}{dt} - 0 \\
k_1 &= \frac{1}{\rho_{\Omega}} \left(\frac{\partial\Delta\lambda_{rev}}{\partial a}\Delta a_o + \frac{\partial\Delta\lambda_{rev}}{\partial i}\Delta i_o \right) \quad k_2 = \frac{1}{\rho_{\Omega}} \left(\frac{\partial\Delta\lambda_{rev}}{\partial a}\frac{da}{dt} + \frac{\partial\Delta\lambda_{rev}}{\partial i}\frac{di}{dt} \right)
\end{aligned} \tag{11-28}$$

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,

$$\Delta(\Delta\lambda_{rev}) = \Delta\lambda_{rev} - \lambda_{rev0} = k_1(t - t_o) + \frac{1}{2}k_2(t - t_o)^2 \tag{11-29}$$

where $\Delta\lambda_{rev}$ is the groundtrack drift from the initial value λ_{rev0} over the time interval $t - t_o$. Usually, these values are offsets from the nominal groundtrack value, λ_{revnom} in Eq. (11-26), 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. 11-15. 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 .

Note that this quadratic profile was derived from a first-order Taylor series expansion given in Eq. (11-27); thus, the coefficients are evaluated at time t_o , usually the burn time

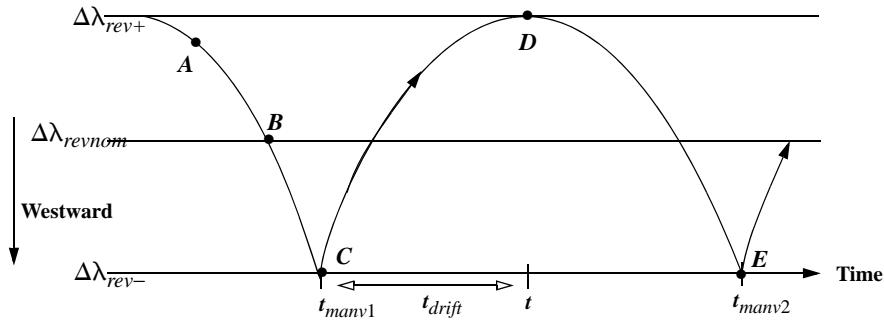


Figure 11-15. Profile of Groundtrack Drift. This figure shows the change in the groundtrack-shift parameter, $\Delta\lambda_{rev}$, over time as corrected by periodic burns for a direct orbit. There are usually limits placed about the nominal value (\pm) within which the satellite mission is successful. 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 11-3 begins at point B .

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, depending only on the natural perturbing forces acting on the spacecraft's motion. 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. (11-29)] provides the basis for finding the Δv necessary to cause the groundtrack to drift from the current offset value $\Delta\lambda_{rev-}$ at the burn time t_{manv1} , to the maximum allowed westward drift $\Delta\lambda_{rev+}$ 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 westward drift. The maximum groundtrack drift occurs at time $t - t_{manv1} = t_{drift}$ and satisfies the condition $d\Delta\lambda_{rev} / dt = 0$ which, in view of Eq. (11-28), 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. (11-29)] at this time obtains the maximum westward drift value:

$$\Delta\lambda_{rev+} - \Delta\lambda_{rev-} = -k_1 \frac{k_1}{k_2} + \frac{1}{2} k_2 \left(-\frac{k_1}{k_2} \right)^2 = -\frac{1}{2} \frac{1}{k_2} k_1^2$$

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

$$k_1 = \sqrt{2k_2(\Delta\lambda_{rev-} - \Delta\lambda_{rev+})}$$

The desired offset in semimajor axis is obtained by inverting the expression for k_1 in Eq. (11-28). Adding this offset to the original value gives us the corrected semimajor axis to maintain the repeat-groundtrack criteria (remember that $\Delta i_o \equiv 0$)

$$\Delta a_o = k_1 \rho_\Omega \left(\frac{\partial \Delta\lambda_{rev}}{\partial a} \right)^{-1}$$

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 [Eq. (9-52)]

$$\Delta v = \frac{n^2 a}{2v} \Delta a$$

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, $v = na$), and the expression is

$$\Delta v = \frac{n}{2} \Delta a + O(e)$$

Notice 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 undesirable 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 72. 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 72: Maintain Repeat Groundtrack

$$(a, e, i, \Delta\lambda_{rev+}, \Delta\lambda_{rev-}, da/dt, di/dt \Rightarrow a_{new})$$

$$p = a(1 - e^2)$$

$$\dot{\Omega} = -\frac{3nR_\oplus^2 J_2}{2p^2} \cos(i) \quad n = \sqrt{\frac{\mu}{a^3}}$$

$$\frac{\partial \rho_\Omega}{\partial a} = \frac{3\pi}{na} \left[1 + \frac{1}{2} J_2 \left(\frac{R_\oplus}{a} \right)^2 (4 \cos^2(i) - 1) \right] + O(e)$$

$$\frac{\partial \rho_\Omega}{\partial i} = \frac{12\pi}{n} J_2 \left(\frac{R_\oplus}{a} \right)^2 \sin(2i) + O(e)$$

$$\frac{\partial \dot{\Omega}}{\partial a} = -\frac{7\dot{\Omega}}{2a} \quad \frac{\partial \dot{\Omega}}{\partial i} = -\dot{\Omega} \tan(i)$$

$$\rho_\Omega \equiv \frac{2\pi}{n} \left[1 - \frac{3J_2}{2} \left(\frac{R_\oplus}{a} \right)^2 (3 - 4 \sin^2(i)) \right]$$

$$\Delta\lambda_{revnom} = (\omega_\oplus - \dot{\Omega}) \rho_\Omega$$

$$\frac{\partial \Delta\lambda_{rev}}{\partial a} = R_\oplus (\omega_\oplus - \dot{\Omega}) \frac{\partial \rho_\Omega}{\partial a} - \frac{\partial \dot{\Omega}}{\partial a} R_\oplus \rho_\Omega$$

$$\frac{\partial \Delta\lambda_{rev}}{\partial i} = R_\oplus (\omega_\oplus - \dot{\Omega}) \frac{\partial \rho_\Omega}{\partial i} - \frac{\partial \dot{\Omega}}{\partial i} R_\oplus \rho_\Omega$$

$$k_2 = \frac{1}{\rho_\Omega} \left(\frac{\partial \Delta\lambda_{rev}}{\partial a} \frac{da}{dt} + \frac{\partial \Delta\lambda_{rev}}{\partial i} \frac{di}{dt} \right) \quad k_1 = \sqrt{2k_2(\Delta\lambda_{rev-} - \Delta\lambda_{rev+})}$$

$$t_{drift} = -\frac{k_1}{k_2}$$

$$\Delta a = k_1 \rho_\Omega \left(\frac{\partial \Delta\lambda_{rev}}{\partial a} \right)^{-1}$$

$$\Delta v = \frac{n}{2} \Delta a \text{ for small } e$$

11.4.3 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 $h_{\text{ellp}} = r - r_{\text{site}}$. From Eq. (1-24) 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 12). 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

$$\Delta h_{\text{ellp}} = \Delta r - \Delta r_{\text{ellipsoid}}$$

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: $\Delta r_{\text{max}} = 2ae$.

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 v from the trajectory equation [Eq. (1-24)] in favor of u yields

$$r = \frac{a(1 - e^2)}{1 + e \cos(u - \omega)} \quad (11-30)$$

This expression yields r as a function of the argument of latitude, u and the argument of perigee, in addition to a and 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 ($\Delta r_{\text{max}} = 2ae$). 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

$$\Delta r = \frac{\partial r}{\partial a} \Delta a + \frac{\partial r}{\partial e} \Delta e + \frac{\partial r}{\partial \omega} \Delta \omega + \text{higher-order terms} \quad (11-31)$$

Let's consider each of these contributions separately.

The semimajor axis's decay due to drag maps directly into altitude decay. Using Eq. (11-30),

$$\Delta r = \frac{\partial r}{\partial a} \Delta a + O(\Delta a)^2 = \frac{r}{a} \Delta a + O(\Delta a)^2 = \frac{(1 - e^2) \Delta a}{1 + e \cos(u - \omega)} + O(\Delta a)^2 \quad (11-32)$$

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. (11-32) 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,

$$\Delta r = -\frac{r}{a(1 - e^2)} [r \cos(u - \omega) + 2ae] \Delta e \quad (11-33)$$

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 337,

the most efficient eccentricity-adjustment burn uses thrust in the transverse (normal to the radius vector) direction while at perigee or apogee [Eq. (9-24)]:

$$\Delta e = \frac{\sqrt{1-e^2}}{na} (\cos(E) + \cos(\nu)) \Delta v_T$$

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 (which is always defined for circular orbits), and the Δv is applied along the 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 11-16 shows the general geometry depicting variations in perigee location.

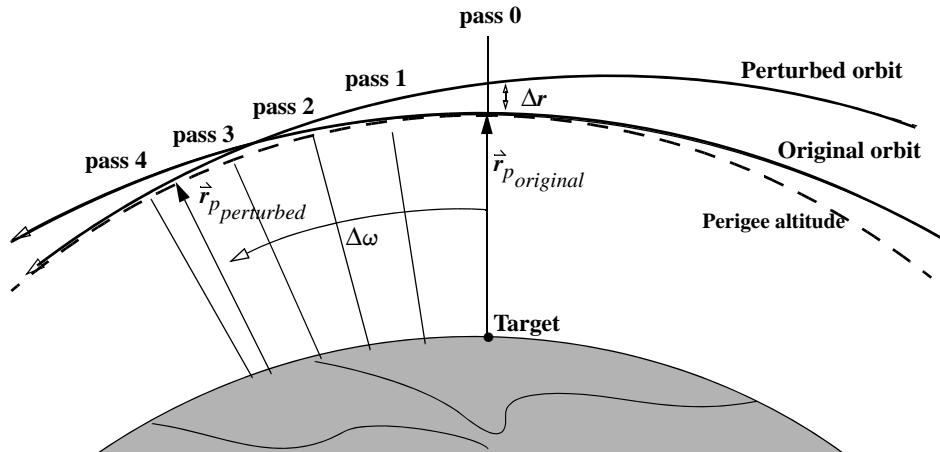


Figure 11-16. Perigee Drift. This figure is an exaggerated view of how J_2 secular perturbations cause perigee to drift. On each pass, perigee drift causes the satellite's altitude over the target to increase. We must correct the orbit between passes three and four to keep the satellite within mission limits. The perturbed orbit is shown at the exact point where the target will exit the satellite's field of view.

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,

$$|\Delta r_{max}| = ae(1-e^2)|\Delta \omega| \quad (11-34)$$

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. (11-34) 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 $|\dot{\omega}|$.

$$\tau_{max} = \left| \frac{\Delta \omega}{\dot{\omega}} \right| \quad (11-35)$$

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 $u-\omega = 0, 180^\circ$. 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 $\left. \partial^3 r / \partial \omega^3 \right|_{u-\omega=0}$ 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

$$\Delta \omega = \pm \sqrt{\frac{2}{ae} \left(\frac{1+e}{1-e} \right) \Delta r_{max}} \quad (11-36)$$

This is the maximum drift from perigee. To get the maximum drift time, we use Eq. (11-35) 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 from Eq. (9-24),

$$\Delta \omega = \frac{\sqrt{1-e^2}}{nae} \left(1 + \frac{r}{p} \right) \sin(\nu) \Delta v_S$$

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 73 shows the required equations to determine the time before burning and the magnitude of the burn.

ALGORITHM 73: Minimum Altitude Variation ($\Delta r_{max}, a, e, i \Rightarrow \tau_{max}, \Delta v_s$)

$$\begin{aligned} n &= \sqrt{\frac{\mu}{a^3}} & \dot{\omega} &= \frac{3nR_{\oplus}^2 J_2}{4a^2(1-e^2)^2} (4 - 5\sin^2(i)) \\ \Delta\omega_{max} &= \sqrt{2 \frac{\Delta r_{max}}{ae} \left(\frac{1+e}{1-e} \right)} \\ \tau_{max} &= \left| \frac{\Delta\omega_{max}}{\dot{\omega}} \right| \\ \Delta v_s &= \frac{nae}{2} |\Delta\omega_{max}| \end{aligned}$$

11.4.4 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

$$\begin{aligned} \frac{d\omega}{dt} &= \frac{3n}{(1-e^2)^2} J_2 \left(\frac{R_{\oplus}}{a} \right)^2 \left(1 - \frac{5}{4} \sin^2(i) \right) \theta \\ \theta &= 1 + \frac{J_3}{2J_2} \left(\frac{R_{\oplus}}{a} \right) \left(\frac{1}{(1-e^2)} \right) \left(\frac{\sin^2(i) - e^2 \cos^2(i)}{\sin(i)} \right) \frac{\sin(\omega)}{e} \end{aligned} \quad (11-37)$$

Combining terms gives us

$$\frac{d\omega}{dt} = \frac{3n}{4} J_2 \left(\frac{R_{\oplus}}{a} \right)^2 (4 - 5\sin^2(i)) + \frac{3n}{8e} J_3 \left(\frac{R_{\oplus}}{a} \right)^3 \sin(\omega) (4 - 5\sin^2(i)) \left(\frac{\sin^2(i) - e^2 \cos^2(i)}{\sin(i)} \right)$$

Alfriend (2006) suggests that an additional term is needed in the relation although we note that averaging long periodic terms seems contradictory.

$$\frac{d\omega}{dt} = \frac{3n}{4} J_2 \left(\frac{R_{\oplus}}{p} \right)^2 (4 - 5 \sin^2(i)) + \frac{3n}{8e} J_3 \left(\frac{R_{\oplus}}{p} \right)^3 \sin(\omega) \left[(4 - 5 \sin^2(i)) \left(\frac{\sin^2(i) - e^2 \cos^2(i)}{e \sin(i)} \right) + 2e^2 \sin(i) (15 \sin^2(i) - 13) \right]$$

The long-periodic eccentricity rate is

$$\frac{de}{dt} = -\frac{3}{2} \frac{n}{(1-e^2)^2} J_3 \left(\frac{R_{\oplus}}{a} \right)^3 \sin(i) \left(1 - \frac{5}{4} \sin^2(i) \right) \cos(\omega) \quad (11-38)$$

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_ω , 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 $O(e^2)$,

$$e_o \approx -\frac{1}{2} \frac{J_3}{J_2} \left(\frac{R_{\oplus}}{a} \right) \sin(i)$$

For real-world applications, we extend the analysis to include other zonal harmonics. Continuing the Landsat 6 example (page 871), $J_2 = 0.108\ 265 \times 10^{-2}$ 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.

Using the example of the Navy Remote Ocean Sensing System (NROSS, McClain 1987), we can plot the phase space near the frozen orbit values. NROSS used a semimajor axis of 7198.65 km, and an inclination of 98.697°. The frozen orbit parameters are found from Eq. (11-37) and Eq. (11-38) as $e = 0.001\ 029\ 4$ and $\omega = 90^\circ$. Figure 11-17 shows the results generated by making several runs with a numerical propagation (STK/HPOP) with a 21×21 gravity field and combinations of initial e/ω combinations.

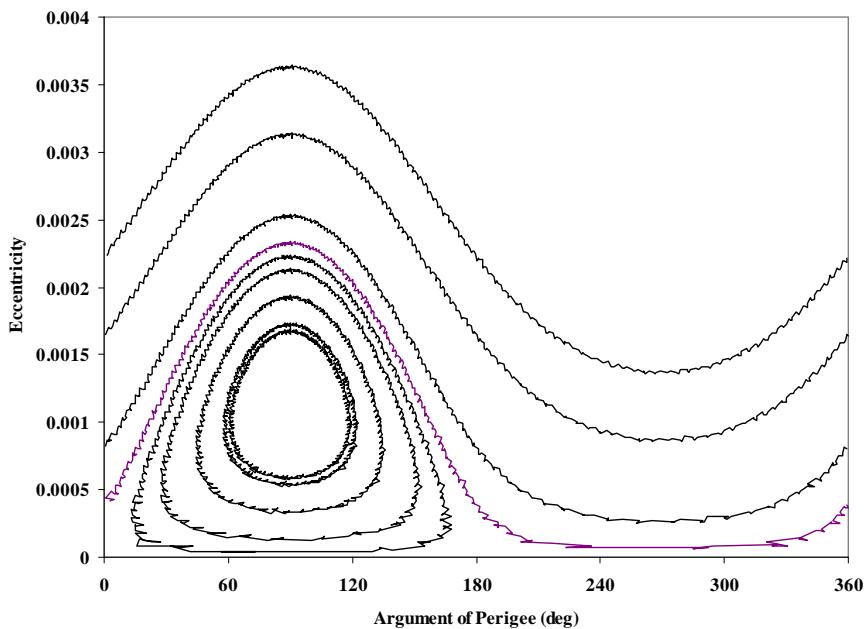


Figure 11-17. Frozen Orbit Phase Space. This plot shows various combinations of mean eccentricity and argument of perigee near the frozen orbit values for the NROSS satellite. The small variations are an artifact of using sectoral and tesseral harmonics in the trajectory simulations, and ambiguity in ω from nearly circular orbits. Note that these values are Kozai (1959) mean values (See “Mean to Osculating Element Conversions” on page 708.).

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 phenomenon of a frozen-orbit eccentricity/perigee depends only on the perturbing force from the zonal harmonics. Using the element his-

tories 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. Earth-observing missions that employ repeat-groundtrack orbits and Sun-synchronous orbits often use frozen-orbit conditions for eccentricity/perigee. 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.

11.4.5 Application: Designing a Specialized Orbit

Let's examine a photo-reconnaissance satellite over a hypothetical target at the Colosseum in Rome, Italy, to summarize the design process of the previous sections. 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 $\eta_{FOV} = 20^\circ$, 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 site's radius at the target latitude. Figure 11-18 shows the general geometry.

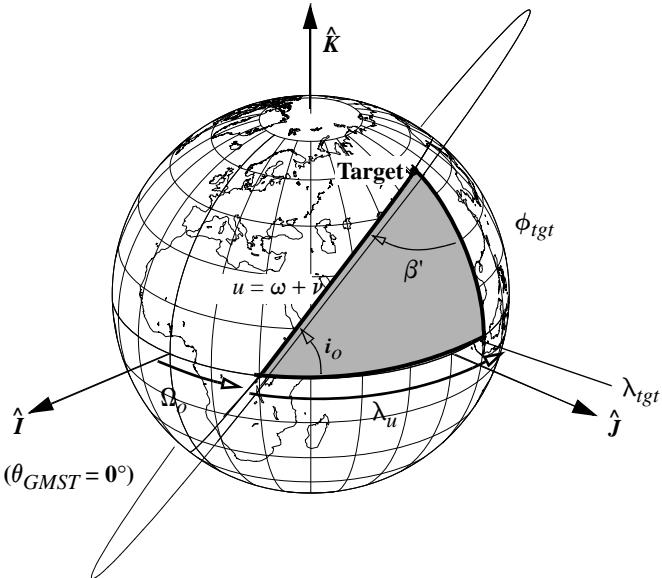


Figure 11-18. Geometry for Injecting a Satellite into Orbit. The inclination is a critical parameter in determining a strategy for satellite injection. Ω depends on the time (GMST) and longitude (λ_u) from the initial placement (target), whereas the remaining orbital elements arise from mission constraints. 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 (β').

We find the magnitude of the site radius using Eq. (3-7). 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.

$$r_p = r_{site} + alt_{perigee} = a_o(1.0 - e_o) \quad (11-39)$$

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 ($\Delta\lambda_{pass}$). 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 ($k_{day2rep}$).

The mean motion is found by dividing the number of crossings by the repeat interval, $n = k_{rev2rep}/k_{day2rep}$. 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. (1-29).

In some cases (Example 11-3), 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.

$$e_o = \frac{a_o - r_p}{a_o} \quad (11-40)$$

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

$$\phi_{tgt} < i_o < 180.0^\circ - \phi_{tgt} \quad (11-41)$$

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. 6.4.1 in Chap. 6) 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.

For this discussion, we'll ignore the ECI/ECEF distinctions and find the right ascension of the ascending node, Ω , using λ_u , λ_{tgt} , and Greenwich Sidereal Time, θ_{GMST} , as shown in Fig. 11-18. The first step in solving for Ω is to use equations from the launch-window problem (Sec. 6.4.1) to solve for the satellite's azimuth [β , Eq. (6-12)] as it passes over the target and to find the angle between the node and the target's longitude meridian [λ_u , Eq. (6-16)].

The next step is to solve for θ_{GMST} , given the start time of the mission. Algorithm 15 determines the θ_{GMST} and LST, but be sure to use Algorithm 14 to determine the correct JD input for the routine. From Fig. 11-18, $LST = \Omega + \lambda_u$. Finally, solve for Ω :

$$\Omega_o = \theta_{GMST} + \lambda_{tgt} - \lambda_u \quad (11-42)$$

The argument of perigee is defined by the target latitude and orbit inclination, if we assume perigee is above the target. Figure 11-18 depicts the geometry necessary to calculate the argument of perigee. Using the sine law from spherical trigonometry [Eq. (C-27)] gives us $\text{SIN}(\omega_o) / \text{SIN}(90^\circ) = \text{SIN}(\phi_{tgt}) / \text{SIN}(i_o)$. Simplifying,

$$\text{SIN}(\omega_o) = \frac{\text{SIN}(\phi_{tgt})}{\text{SIN}(i_o)} \quad (11-43)$$

The last parameter, mean anomaly, is chosen so the satellite starts at perigee over the target at the epoch time, thus $M_o = 0.0^\circ$.

▼ Example 11-3. 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 51,

$$r_\delta = 4782.104 \text{ km}, r_K = 4207.153 \text{ km} \text{ and } |r_{site}| = \sqrt{r_\delta^2 + r_K^2} = 6368.936 \text{ km}$$

This allows you to find the radius of perigee:

$$r_p = 6368.936 \text{ km} + 160 \text{ km} = 6528.936 \text{ km}$$

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. (11-10), Eq. (11-11), and Eq. (11-13) 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; $r_{sat} = 6538.14 \text{ km}$.

$$\sin(\gamma) = \frac{r_{sat} \sin(\eta)}{R_\oplus} = \frac{(160 + 6378.1363) \sin(25^\circ)}{6378.1363}, \gamma = 25.672\ 079^\circ, \mathbf{154.327\ 920^\circ}$$

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 ($160 \pm 30 \text{ km}$).

$$\rho = R_\oplus \cos(\gamma) + r_{sat} \cos(\eta) = \cos(154.327\ 920) + 6538.14 \cos(25^\circ)$$

$$\rho = 177.024\ 616 \text{ km}$$

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. (11-1) or Eq. (C-24).

$$\sin(\Lambda_{max}) = \frac{\rho \sin(\eta)}{R_\oplus} = \frac{177.024\ 616 \sin(25^\circ)}{6378.1363}, \Lambda_{max} = 0.672\ 066^\circ = 74.814 \text{ km}$$

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 revs to repeat, or the number of crossing points— $k_{rev2rep} = 2\pi R_\oplus / 243 = 164.918$ 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 ($k_{day2rep}$) that gives us $n_o = 164 / 10 = 16.4$ 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. $n_o = 0.001\ 192\ 642 \text{ rad/s}$.

$$a_o = \left(\mu \left(\frac{1}{n_o} \right)^2 \right)^{1/3} = \left(398600.4418 \left(\frac{86400}{(16.4)(2\pi)} \right)^2 \right)^{1/3} = 6543.364 \text{ km}$$

Now find the eccentricity to make our perigee height exactly 160 km—it should be small.

$$e_o = (a_o - r_p) / a_o = (6543.364 - 6528.936) / 6543.364 = 0.002\ 205$$

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. Choosing $e_o = 0.001$ lets us find a_o and n_o .

$$a_o = r_p / (1 - e_o) = 6528.936 / (1 - 0.001) = 6535.472 \text{ km}$$

$$n_o = \sqrt{\frac{\mu}{a_o^3}} = \sqrt{\frac{398600.4418}{6535.472^3}} = 0.001195 \text{ rad/s}$$

To continue, you need to select an inclination. Eq. (11-43) 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 (you'll recognize some concepts from analyzing resonance in Chap. 9). 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 666. Using $\dot{M} = n + M_o$,

$$\frac{n + \dot{\omega} + \dot{M}_o}{\omega_{\oplus} - \dot{\Omega}} \approx \frac{P}{Q} = \frac{k_{rev2rep}}{k_{day2rep}}$$

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 $n/\omega_{\oplus} = 0.001195 / 0.00007292 = 16.387017$ rev/side-real 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. (11-22), Eq. (11-23), and Eq. (11-24).

$$\frac{Q_i}{P_i} = \frac{\theta_{\theta_G}}{\theta_{\Omega}} \text{ and } n = \frac{P_i}{Q_i} \left\{ \frac{2\pi}{\theta_{\theta_G}} \right\} \frac{\theta_{\Omega}}{\theta_K} = \frac{P_i}{Q_i} (\omega_{\oplus} - \dot{\Omega}) \left\{ 1 - \frac{3J_2}{2} \left(\frac{R_{\oplus}}{a_o} \right)^2 (3 - 4\sin^2(i_o)) \right\}$$

Now evaluate this expression to find the semimajor axis and eccentricity. Using the equations on page 668 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	104.915552	6.507022	6.517102	0.956726	6569.0409	0.006105
		1	104.948259	6.557325	6.567380	0.956443	6570.3375	0.006301
33	2	0	104.915552	6.507022	6.517102	0.986624	6435.6536	invalid
49	3	0	104.915552	6.507022	6.517102	0.976658	6479.3595	invalid
82	5	0	104.915552	6.507022	6.517102	0.980645	6461.7883	invalid
213	13	0	104.915552	6.507022	6.517102	0.979725	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:

$$\text{SIN}(\beta) = \frac{\cos(i_o)}{\cos(\phi_{tgt})} = \frac{\cos(45^\circ)}{\cos(41.52^\circ)} , \beta = 70.8068^\circ$$

$$\cos(\lambda_u) = \frac{\cos(\beta)}{\text{SIN}(i_o)} = \frac{\cos(70.8068)}{\text{SIN}(45.0^\circ)} , \lambda_u = 62.2943^\circ$$

Next, find the Greenwich Sidereal Time from Algorithm 15. The $JD = 2,451,637.0$. Remember that the rigorous approach would use ECI and ECEF vectors.

$$\text{LSTIME}(JD, UT, \lambda \Rightarrow \theta_{LST}, \theta_{GMST}), \theta_{GMST} = 11.140\ 176^\circ$$

$$\Omega_o = \theta_{GMST} - \lambda_u + \lambda_{tgt} = -38.784\ 170^\circ$$

$$\text{SIN}(\omega_o) = \frac{\text{SIN}(\phi_{tgt})}{\text{SIN}(i_o)} = \frac{\text{SIN}(41.52^\circ)}{\text{SIN}(45.0^\circ)} , \omega_o = 69.628\ 610^\circ$$

Finally, set $M_o = 0^\circ$ so the satellite starts over the target. The preliminary values are

$$a_o = 6570.3375 \text{ km}, e_o = 0.006\ 301, i_o = 45.0^\circ$$

$$\Omega_o = -38.784\ 170^\circ, \omega_o = 69.628\ 610^\circ, M_o = 0.000^\circ$$

 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 still have to perform periodic maneuvers to adjust the orbit.

▼ Example 11-4. Maintaining the Repeat-Groundtrack Orbit.

GIVEN: The results from Example 11-3

FIND: The rate of nodal regression from J_2 , t_{drift} , and Δv

$$p = a(1 - e^2) = 6570.3358(1 - 0.006\ 301^2) = 6570.0749 \text{ km}$$

$$\dot{\Omega} = -\frac{3nR_\oplus^2 J_2}{2p^2} \cos(i) = -1.2829 \times 10^{-6} \text{ rad/s} = -6.350\ 78^\circ/\text{day}$$

You will require several partial derivatives to find the solution.

$$\frac{\partial \dot{\Omega}}{\partial a} = \frac{3\pi}{na} \left[1 + \frac{1}{2} J_2 \left(\frac{R_\oplus}{a} \right) (4 \cos^2(i) - 1) \right] = 1.2106 \text{ s/km}$$

$$\frac{\partial \dot{\Omega}}{\partial i} = -\frac{7}{2} \frac{\dot{\Omega}}{a} = 6.8339 \times 10^{-10} \text{ rad/(s·km)}$$

$$\frac{\partial \dot{\Omega}}{\partial i} = \frac{12\pi}{n} J_2 \left(\frac{R_\oplus}{a} \right) \sin(2i) = 32.4441 \text{ /s}$$

$$\frac{\partial \dot{\Omega}}{\partial i} = -\dot{\Omega} \tan(i) = 1.2829 \times 10^{-6} \text{ rad/s}$$

Find the anomalistic period.

$$\rho_{\Omega} \approx \frac{2\pi}{n} \left[1 - \frac{3J_2}{2} \left(\frac{R_{\oplus}}{a} \right)^2 (3 - 4\sin^2(i)) \right] = 5242.1016 \text{ s} = 88.20169 \text{ min}$$

The groundtrack shift should be approximately equal to $2\pi R_{\oplus} k_{days2rep} / k_{rev2rep}$.

$$\Delta\lambda_{rev} = (\omega_{\oplus} - \dot{\Omega})\rho_{\Omega} = (7.2921 \times 10^{-5} - 1.2829 \times 10^{-6}) 5242.1016 = 0.392695 \Rightarrow$$

$$2504.6604 \text{ km}$$

The partial derivatives give you the sensitivity of λ_S to changes in a and i .

$$\frac{\partial \Delta\lambda_{rev}}{\partial a} = R_{\oplus}(\omega_{\oplus} - \dot{\Omega}) \frac{\partial \rho_{\Omega}}{\partial a} - \frac{\partial \dot{\Omega}}{\partial a} R_{\oplus} \rho_{\Omega} = 0.549921$$

$$\frac{\partial \Delta\lambda_{rev}}{\partial i} = R_{\oplus}(\omega_{\oplus} - \dot{\Omega}) \frac{\partial \rho_{\Omega}}{\partial i} - \frac{\partial \dot{\Omega}}{\partial i} R_{\oplus} \rho_{\Omega} = -27.947 \text{ km}$$

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. 8 and Chap. 9). Using the simplified numerical technique in Chap. 8.

$$da/dt = -0.174562 \text{ km/rev} = -3.2985 \times 10^{-5} \text{ rad/s}$$

$$di/dt = 0.000132 \text{ /rev} = 4.3533 \times 10^{-10} \text{ rad/s}$$

The two k -parameters are found using tolerances that are chosen for the maximum allowable groundtrack drift. Let $\Delta\lambda_{revnom}$ vary by about ± 50 km. This puts a 100 km tolerance band about the nominal groundtrack shift parameter, $\Delta\lambda_{revnom}$. Even if the satellite drifts this far, the sensor will still be able to see the Colosseum ($\Lambda_{max} = \pm 74$ km).

$$k_2 = \frac{1}{\rho_{\Omega}} \left(\frac{\partial \Delta\lambda_{rev}}{\partial a} da + \frac{\partial \Delta\lambda_{rev}}{\partial i} di \right) = -3.4299 \times 10^{-9} \text{ km/s}^2$$

$$k_1 = \sqrt{2k_2(\Delta\lambda_{rev-} - \Delta\lambda_{rev+})} = \sqrt{2(-3.4299 \times 10^{-9})(\frac{-50 - 50}{6378.1363})} = 8.2824 \times 10^{-4} \text{ km/s}$$

This allows you to find the correction to the semimajor axis.

$$\Delta a = k_1 \rho_{\Omega} \left(\frac{\partial \Delta\lambda_{rev}}{\partial a} \right)^{-1} = 8.2824 \times 10^{-4} (5242.1016)(1.0 / 0.549921) = 7.9706 \text{ km}$$

Finally, find out how often to maneuver, and the magnitude of each burn.

$$t_{drift} = -\frac{k_1}{k_2} = -\frac{8.2824 \times 10^{-4}}{-3.4299 \times 10^{-9}} = 241477 \text{ s} = 4024.6 \text{ min} = 2.79 \text{ day}$$

$$\Delta v = \frac{n}{2} \Delta a = \frac{0.00118}{2} (-3.2985 \times 10^{-5}) = 4.724 \text{ m/s}$$

Remember that you let the satellite drift for twice the drift time and you over correct to get the most drift, so you'll have to perform a maneuver of about 9.45 m/s every 5.58 days. This is a moderate maneuver, and you must correct the orbit about every few days—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 will also cause the perigee location to change. Perigee drift soon violates the requirement for the camera to be 160 km ± 30 km

over the target unless the orbit is nearly circular. Even for nearly circular orbits, the semimajor axis and eccentricity variations will nullify the initial analysis over time.

Algorithm 73 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 11-5. Maintaining Minimum Altitude Variations.

GIVEN: The results of Example 11-4.

FIND: Corrections to minimize altitude variations.

The first step is to find the secular perturbations resulting from J_2 : $\dot{\omega} = 6.736\ 023^\circ/\text{day} = 1.3607 \times 10^{-6}$ rad/s. The maximum change in distance is ± 30 km. Thus the maximum allowable apsidal drift is

$$\Delta\omega_{max} = \sqrt{2 \frac{\Delta r}{ae} \left(\frac{1+e}{1-e} \right)} = \sqrt{2 \frac{30}{(6570)(0.006\ 301)} \left(\frac{1+0.006\ 301}{1-0.006\ 301} \right)} = 1.211\ 473 = 69.4123^\circ$$

The amount of time until this condition occurs is

$$\tau_{max} = \left| \frac{\Delta\omega_{max}}{\dot{\omega}} \right| = \left| \frac{1.211\ 473}{1.36 \times 10^{-6}} \right| \Rightarrow 14,838.68 \text{ min} = 10.3046 \text{ days}$$

Now find the change in velocity.

$$\Delta v = \frac{nae}{2} \Delta\omega_{max} = \frac{0.001\ 186(6570)0.006\ 301}{2} 1.211\ 473 = 29.728 \text{ m/s}$$

Notice the long time delay between maneuvers. In addition, each burn is rather large. We could further reduce these maneuvers by incorporating frozen-eccentricity designs. We could also do the corrections at ± 10 km, resulting in $\Delta\omega_{max} = 40.076^\circ$, $\tau_{max} = 5.948$ days, and $\Delta v = 17.164$ m/s.

▲ 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.

11.5 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 event occurred when the British Parliament offered £20,000 for anyone who could determine longitude within about $1/4^\circ$ 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, but he wasn't paid in full until 1773 (the Board of

Longitude is now synonymous with stagnant bureaucracies). His clock worked superbly (determining longitude to about $1\frac{1}{4}$). This instrument permitted vastly improved astrometric observations, and modern versions are still used today. See also Sobel (1998).

Modern clocks and navigation satellites mark tremendous technological advances that are revolutionizing 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. New applications for GPS are continuing to emerge. 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).

11.5.1 Historical Background

Although the history of navigation reaches back before the Christian era, the first *practical* radio-navigation systems emerged only in the last century. Navigators 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 two observations to create a two-dimensional fix with 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 constellation, 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 obvi-

ous, 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 11-19 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.

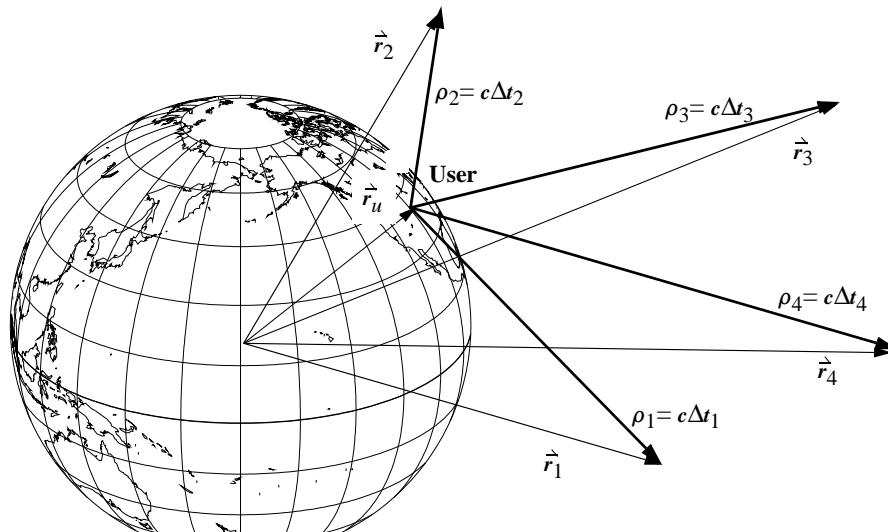


Figure 11-19. Navigating with the GPS—the 4-Dimensional Simplex. A GPS receiver starts its navigation solution by measuring four pseudoranges to four widely dispersed GPS satellites (ρ_i , $i = 1 \dots 4$). Then it substitutes these four measured pseudoranges into a system of four equations in four unknowns. The first three unknowns are the receiver's three mutually orthogonal position coordinates. The fourth unknown is the clock offset error.

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 with

the precise tracking of time—a problem we've seen throughout this book. Remember that time delays are a major consequence of positional errors for a satellite.

Precise global synchronization of clocks is integral to GPS operations. The GPS satellites' atomic clocks are moving fast enough in space 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 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,

$$\Delta\tau_{GPS} = \left\{ 1 - \left(\frac{\mu_{\oplus}}{c^2 r_{GPS}} - \frac{v_{GPS}^2}{2c^2} \right) \right\} \Delta\tau \quad (11-44)$$

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

(See Danielson, 1997:216-222 for an example calculation). 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. The user receiver solves the second-order Doppler correction. Consult Wheeler (1990), Logsdon (1992), or Hofmann-Wellenhof et al. (1994) for further information.

11.5.2 System Introduction

The Global Positioning System has three segments: the *space segment*, the *control segment*, and the *user segment*. The space segment consists of a constellation 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.

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 several designs: Block I, Block II, Block IIA, Block IIR, Block IIR-M, Block IIF, and Block III (which is now in design). Figure 11-20 represents a snapshot in time to show how the GPS satellites are positioned in the constellation.

Block I satellites were first launched in 1978 as a proof of concept. 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 signals. The Block IIA (advanced) satellites talk to each other, so the constellation is more autonomous. Two Block IIA satellites (GPS-35 and GPS-36) 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. Final Operational Capability occurred in April 1995.

Block IIR (replenishment/replacement) satellites feature hydrogen-maser atomic clocks instead of the earlier 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 augment the constellation's autonomy from ground control. The Block III satellites further improve autonomy over the Block IIR satellites and will have an additional signal for civilian use.

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 Schriever 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

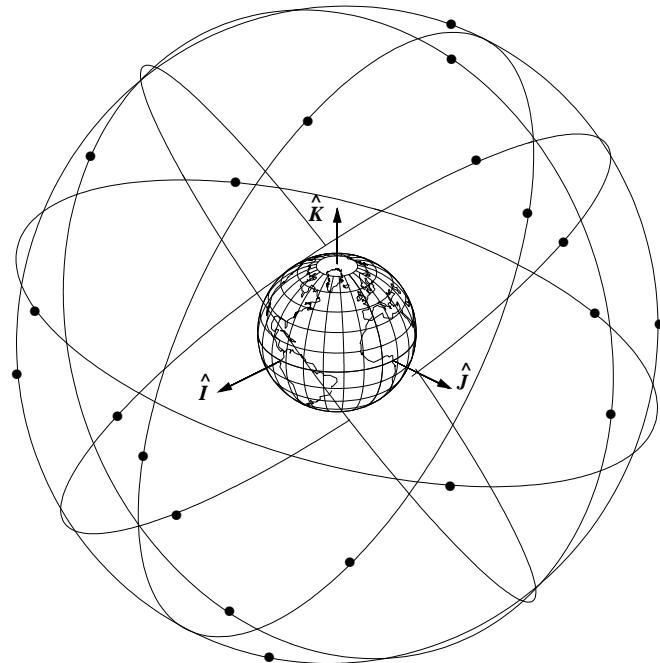


Figure 11-20. The GPS Constellation of 24 Satellites. The 24 satellites in the GPS constellation travel around the Earth in 12-hour circular orbits with $i = 55^\circ$.

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 positioning 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* (IGS), which includes 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.

If you are using GPS measurements, then satellite attitude becomes very important because the measurements are relative to the antenna phase center. The location of the antenna is in the satellite body frame. The GPS community produces precise ephemerides that can be very useful for precision applications (OD, propagation, etc). You'll need these ephemerides (SP3 format) for the applicable period of time. You can obtain the antenna phase center ephemerides (best for using when you have raw GPS data) at

<ftp://ftp.nga.mil/pub2/gps/apcpe/2012apc/>

or the center of mass ephemerides (best for propagating the satellite) at

<ftp://ftp.nga.mil/pub2/gps/pedata/2012pe/>

11.5.3 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 the *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. Multiplying the fundamental frequency by 154 and 120, respectively, generates the L_1 and L_2 carrier frequencies. Upgrades to the L_1 and L_2 bands are in progress, and a new signal called L_5 (1176.45 MHz) is being broadcast from a few satellites. Final completion is expected in about 2026.

The *Coarse-Acquisition code*, C/A, sometimes called the *Standard Positioning Service*, is a pseudorandom noise code that is modulated on both carriers. 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* (S/A), to deny full system accuracy to unauthorized users. This technique intentionally corrupted the GPS satellites' clocks and the broadcast ephemerides. Errors were introduced into the fundamental frequency of the GPS clocks. This clock "dithering" affected corrections to the satellites' clocks, as well as the pseudorange observables. Errors were introduced into the broadcast ephemerides by truncating the orbital information in the navigation message. Clock dithering contributed an error signal to the pseudoranges with 50 m amplitude and periods of several minutes. Truncating the broadcast ephemerides contributed an error signal with 50–150 m amplitude and periods of some hours (Hofmann-Wellenhoff, 1994). S/A was discontinued in May 2000 by Presidential order.

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. The Y-code actually combines the P-code with a W-encryption code and requires a Department of Defense-authorized receiver with the key 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 from GPS is the *pseudorange*—the distance between the GPS satellite at some transmit time, t_{trans} , and the receiver at some receive time, t_{rec} , in a particular coordinate system (WGS-84). 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 can 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 ($\delta t_{rec} = \alpha + \beta\delta T + \gamma\delta T^2$ and δt_{trans} likewise), the pseudorange becomes

$$\rho = c(T_{rec} - T_{trans}) + c(\delta t_{rec} - \delta t_{trans})$$

We now see the pseudorange 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, respectively) and the clock correction terms.

Let's substitute for the magnitude of the true pseudorange with position vectors using an orthogonal coordinate system with the origin at the center of the Earth (typically Earth-Fixed for surveying and point positioning). Now, the model becomes

$$\rho = |\vec{r}_{GPS}(T_{trans}) - \vec{r}_{rec}(T_{rec})| + c(\delta t_{rec} - \delta t_{trans}) \quad (11-45)$$

The satellite coordinates (first term) and clock corrections (last term) are available from the broadcast ephemeris as well as several external sources. Thus, four unknowns remain in Eq. (11-45). 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 $3 + k \leq i_{GPS}$ for a stationary receiver. 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, although in practice, more information is often needed. This is called **time determination**.

* The actual process of working with observables is complex. Gabor (1999) shows more extensive development of the equations and the necessary assumptions.

Carrier-Phase Observables

Another observable is based on the *carrier phase* of the signal. Hofmann-Wellenhof et al. (1994) show the development. The receiver observes the signal frequency of the carrier wave and doesn't require any information contained in 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 and substitute the clock errors as $\phi_o = c(\delta t_{rec} - \delta t_{trans})$, the phase observable equation becomes

$$N + \Delta\phi = (f_s - f_r)T - \frac{f_s}{c}\rho_{true} + f(\delta t_{rec} - \delta t_{trans})$$

If we assume that $f_s = f_r$ (by design) and recall that $c = \lambda f$, we find that

$$\begin{aligned} \rho_\phi &= \rho_{true} + c(\delta t_{rec} - \delta t_{trans}) + \lambda N \\ \rho_\phi &= -\lambda\Delta\phi \end{aligned} \tag{11-46}$$

As a shorthand notation, we'll define $\rho_\phi = -\lambda\Delta\phi$. Note that by comparing ρ_ϕ with Eq. (11-45), we can form $\rho \equiv \rho_\phi - \lambda N$ where we have a distance equal to N cycles of wavelength λ , and a fraction of the wavelength, $\Delta\phi$.

The integer number of cycles, N , is typically not known and varies for every phase connected pass of 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 N 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, and there is

an integer jump in the cycles of N . 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 (for a stationary receiver) 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, we must 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. The ionosphere is a dispersive region and its effect is dependent on the signal wavelength. Thus, 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 that accounts for 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 linear combination of the L₁ and L₂ frequencies. Manufacturers of receivers typ-

ically supply the location of the offset of the antenna phase center 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, wide-band 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 to the Earth. The Earth's gravitational field also 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 L₁ and L₂ 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 pseudorange and carrier phase observables to eliminate an undesirable effect, or aid resolution of other ambiguities. 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 because the combination has a much larger wavelength than either frequency. 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. There are several ambiguity resolution techniques.

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 three.

Figure 11-19 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 a system of four equations and four unknowns. We can regard each of these four equations as the algebraic characterization of the diagonal of a parallelepiped. From Fig. 11-19 and Eq. (11-45), 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, $(\delta t_{rec} - \delta t_{trans})$, each component ($i = 1\dots4$) is

$$(\rho_i - c \cdot (\delta t_{rec} - \delta t_{trans}))^2 = (r_{iI} - r_{uI})^2 + (r_{iJ} - r_{uJ})^2 + (r_{iK} - r_{uK})^2 \quad (11-47)$$

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. (11-47)]. Ordinary algebraic 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:

$$\rho_1 = \sqrt{(r_{1I} - r_{uI})^2 + (r_{1J} - r_{uJ})^2 + (r_{1K} - r_{uK})^2} + c \cdot (\delta t_{rec} - \delta t_{trans}) \quad (11-48)$$

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.

$$\begin{aligned} \Delta\rho_1 &= \rho_1 - \rho_1(r_{uIest}, r_{uJest}, r_{uKest}, \delta t_{rec_{est}}) \\ &= \frac{\partial\rho_1}{\partial r_{uI}}\Delta r_I + \frac{\partial\rho_1}{\partial r_{uJ}}\Delta r_J + \frac{\partial\rho_1}{\partial r_{uK}}\Delta r_K + \frac{\partial\rho_1}{\partial\delta t_{rec_{est}}}\Delta\delta t_{rec_{est}} + u_1 \end{aligned} \quad (11-49)$$

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. 10.

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 are from the receiver to the satellites, the better the observability, and thus the estimation 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 (See Langley, 1999 for more details).

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.

Although DGPS is no longer needed to counter the effects of selective availability, it can be useful in improving real-time accuracy. 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

Differencing the signals from two satellites to a receiver forms single differences. The single differences from two receivers are then "differenced" again to arrive at the double difference combination. This technique is similar to DGPS. It uses double-differenced observables to eliminate the effects of biases in solving for baseline distances and relative positions, and for ambiguity resolution. 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 double differencing. Some approaches form the double differences from the phase-derived pseudoranges instead of the phases themselves.*

GPS Summary

At this point, the GPS system may seem to have an incredible number of variables. It does! The system is complex because it must produce accurate timing and positional data including the effects of special and general relativity. The ephemerides provide a unique, highly accurate data source with which to perform experiments. Be sure to properly account for solar radiation pressure (ROCK models). It also provides highly accurate measurements to derive orbits with. There are numerous types of measurements, but in short, single differencing (SD of pseudo-range and phase) allows you to remove receiver clock errors, and dual frequency (DF pseudo-range and phase) solutions allow you to remove Ionosphere errors. Combinations (SD and DF) remove both errors, but require significant data.

I've only sketched the system. I encourage you to study it further and recommend Hofmann-Wellenhof et al. (1994), Wheeler (1990), Ashby and Spilker (1996), or Logsdon (1992) for more information.

11.6 Predicting Satellite Look Angles

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

* Single differencing can remove satellite clock corrections and other effects like the ionosphere.

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 or the *prediction 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 [ECEF] position vectors of each. It advances vectors by some 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. We'll introduce this method shortly along with 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 visibility restrictions.

Escobal ([1965] 1985:162–174) also proposed a 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.

11.6.1 PREDICT Formulation

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 the corresponding 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 geodetic 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 11-21 shows the geometry. The proximity of the site to the terminator determines how the satellite will appear (visually) from the site. Remember there are several conditions for twilight (Sec. 5.1.2). The **terminator** is the line on the Earth that separates night and day. As Fig. 11-21 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 time it's illuminated increases, but the satellite becomes dimmer (its *magnitude* increases) and we'll eventually not be able to see it.

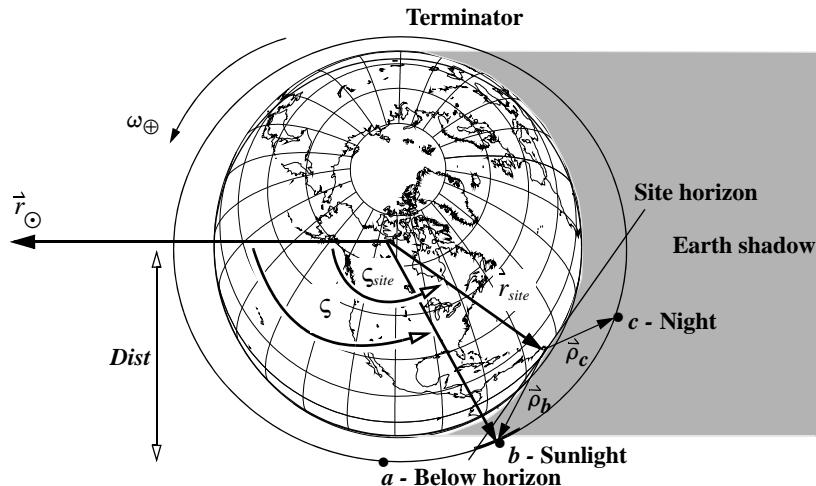


Figure 11-21. Determining Visibility. The visibility of a low-Earth satellite from a ground station, r_{site} , has several cases: the Earth obstructs the view, *a*; the satellite is in the sunlight, *b*; and the satellite is in the Earth's shadow, *c*. We often assume the Earth's shadow is cylindrical. I've shown only one value of ζ although there are actually three distinct values of ζ corresponding to the three points (*a*, *b*, and *c*).

Three conditions are necessary to determine visibility from a site. To establish these conditions, we must first find the slant-range vector from the site to the satellite ($\hat{\rho}_{sez} = \vec{r}_{sat_{sez}} - \vec{r}_{siteSEZ}$). 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. We must find the actual angle (ζ_{site}) if a twilight condition is used. The final condition

consists of determining if the satellite is in the Sun. We must find 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. 5.1.2, page 281. If the perpendicular distance is larger than the Earth's radius, the satellite is in the Sun and is visible to the sensor site. The analysis is more complex if we use the shadow analysis on page 302. 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 these 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 either in the light or 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 (See also Sec. 4.3). 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 limitations is the orbital geometry, 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, sometimes 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 limits the accuracy because of atmospheric refraction. High elevations improve the observation quality and occur with the closest approaches.

Taking these constraints into account, we can write Algorithm 74 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. 8 or Chap. 9. Be careful with the units of the Sun's position vector. The first propagation is included to move the vectors from the satellite epoch to the start time, if they are different.

ALGORITHM 74: PREDICT

$$\begin{aligned}
 & (\hat{\vec{r}}_o, \hat{\vec{v}}_o, JD_o, \phi_{gd}, \lambda, h_{ellp}, \Delta t, JD_{start}, JD_{end} \Rightarrow JD_i, \rho_i, \beta_i, el_i, Vis_i) \\
 & \textbf{PKEPLER} (\hat{\vec{r}}_o, \hat{\vec{v}}_o, JD_{start} - JD_o \Rightarrow \hat{\vec{r}}, \hat{\vec{v}}) \\
 & \text{LOOP} \\
 & JD = JD + \Delta t \\
 & \textbf{PKEPLER} (\hat{\vec{r}}_o, \hat{\vec{v}}_o, \Delta t \Rightarrow \hat{\vec{r}}, \hat{\vec{v}}) \text{ (ECI)} \\
 & \textbf{FK5}(\hat{\vec{r}}_{ECI}, \hat{\vec{v}}_{ECP} \text{ yr, mo, day, UTC, } \Delta \text{UT1})
 \end{aligned}$$

```

 $\Delta AT, x_p, y_p \Rightarrow \dot{\vec{r}}_{ECEF}, \dot{\vec{v}}_{ECEF})$ 
 $SITE(\phi_{gd}, \lambda, h_{ellp} \Rightarrow \dot{\vec{r}}_{siteECEF})$ 
 $\dot{\vec{r}}_{ECEF} = \dot{\vec{r}}_{ECEF} - \dot{\vec{r}}_{siteECEF}$ 
 $\dot{\vec{r}}_{SEZ} = [\text{ROT2}((90^\circ - \phi_{gd}))][\text{ROT3}(\lambda)]\dot{\vec{r}}_{ECEF}$ 
IF  $\dot{\rho}_Z > 0$ 
   $SUN(JD \Rightarrow \dot{\vec{r}}_\odot)$  (make sure units are km)
    IF  $\dot{\vec{r}}_\odot \cdot \dot{\vec{r}}_{siteECEF} > 0$ 
      Vis = 'Radar Sun'
    ELSE
       $\sin(\zeta) = \frac{|\dot{\vec{r}}_\odot \times \dot{\vec{r}}_{ECEF}|}{|\dot{\vec{r}}_\odot| |\dot{\vec{r}}_{ECEF}|}$ 
       $Dist = |\dot{\vec{r}}_{ECEF}| \cos(\zeta - 90^\circ)$ 
      IF  $Dist > R_\oplus$ 
        Vis = 'Visible'
        Find  $\alpha, \delta, \alpha_t, \delta_t, \beta, el$ , as appropriate
      ELSE
        Vis = 'Radar night'
    ELSE
      Vis = 'not visible'
UNTIL  $JD > JD_{end}$ 
-----
```

▼ **Example 11-6. 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 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 10, find

$$\begin{aligned}\dot{\vec{r}} &= 6585.038\ 266\hat{I} + 1568.184\ 321\hat{J} + 9.116\ 355\hat{K} \text{ km} \\ \dot{\vec{v}} &= -1.115\ 776\ 6\hat{I} + 4.631\ 681\ 6\hat{J} + 6.014\ 957\ 6\hat{K} \text{ km/s}\end{aligned}$$

Don't forget to move the satellite to the initial start time. The epoch time is found (UTC = 21:36) giving $JD_{epoch} = 2,450,540,400$. Now determine an ephemeris for the satellite during the time of interest using the following formulas:

$$\begin{aligned}
 & \textbf{PKEPER}(\hat{\vec{r}}_o, \hat{\vec{v}}_o, JD_{start} - JD_o \Rightarrow \hat{\vec{r}}, \hat{\vec{v}}) \\
 & \textbf{FK5}(\hat{\vec{r}}_{ECI}, \hat{\vec{v}}_{ECI}, yr, mo, day, UTC, \Delta UT1, \Delta AT, x_p, y_p \Rightarrow \hat{\vec{r}}_{ECEF}, \hat{\vec{v}}_{ECEF}) \\
 & \textbf{SITE}(\phi_{gd}, \lambda, h_{ellp} \Rightarrow \hat{\vec{r}}_{siteECEF}) \\
 & \hat{\rho}_{ECEF} = \hat{\vec{r}}_{ECEF} - \hat{\vec{r}}_{siteECEF}
 \end{aligned}$$

Because we'll get many vectors throughout the time span, let's look at the process for one set of vectors on April 2, 1997 at 1:08:00 UTC (JD = 2,450,540.5472). Determine visibility as follows. The position and velocity vectors are (ECI)

$$\begin{aligned}
 \hat{\vec{r}} &= -2811.2769\hat{I} + 3486.2632\hat{J} + 5069.5763\hat{K} \text{ km} \\
 \hat{\vec{v}} &= -6.859\ 691\hat{I} - 2.964\ 792\hat{J} - 1.764\ 721\hat{K} \text{ km/s}
 \end{aligned}$$

In ECEF coordinates,

$$\begin{aligned}
 \hat{\vec{r}} &= 885.7296\hat{I} - 4389.3856\hat{J} + 5070.1765\hat{K} \text{ km} \\
 \hat{\vec{v}} &= 7.134\ 377\hat{I} - 0.596\ 976\hat{J} - 1.762\ 745\hat{K} \text{ km/s}
 \end{aligned}$$

Forming (in ECEF) and rotating the slant range vector ($\hat{\vec{r}} - \hat{\vec{r}}_{site}$) to the SEZ coordinate system

$$\hat{\rho}_{SEZ} = -773.8654\hat{S} + -581.4980\hat{E} + 328.8145\hat{Z} \text{ km}$$

Because $\hat{\rho}_Z > 0$, we must find coordinates for the Sun to determine the visibility.

$$SUN(JD \Rightarrow \hat{\vec{r}}_{\odot}), \hat{\vec{r}}_{\odot} = 0.9765\hat{I} + 0.1960\hat{J} + 0.0850\hat{K} \text{ AU}$$

Be sure to convert this vector to km, and find the site vector in ECEF coordinates before finding the dot product, $\hat{\vec{r}}_{\odot} \cdot \hat{\vec{r}}_{site}$. We can find the Sun-satellite angle because the dot product is less than zero.

$$\hat{\vec{r}}_{siteECI} = -3414.0283\hat{I} + 3258.1636\hat{J} + 4276.1212\hat{K} \text{ km}$$

$$\sin(\zeta) = \frac{|\hat{\vec{r}}_{\odot} \times \hat{\vec{r}}|}{|\hat{\vec{r}}_{\odot}| |\hat{\vec{r}}|}, \zeta = 76.0407^\circ$$

The distance quantity is $Dist = |\hat{\vec{r}}| \cos(\zeta - 90^\circ) = 6564.6870$ km. Because $Dist$ is greater than R_{\oplus} , the satellite is visible. We can find α , δ , α_p , δ_p , β , etc., as appropriate. We continue this process for the remaining time interval to arrive at the answers in Table 11-4.

▲ Implementations not using rigorous ECI-ECEF conversions can use $\theta_{LST} = 136.2944^\circ$. The answers are very similar.

11.6.2 Rise / Set

The analytical method from Alfano (1993) and Alfano, Negron, and Moore (1992) uses pairs of fourth-degree polynomials (quartic blending, Sec. C.5.3) 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.

TABLE 11-4. Prediction Values for Example 11-6. This table shows sample visibility parameters for two passes in the example problem.

Visibility	Range km	β °	el °	Date	UTC (h/min/s)
Not Visible	2449.4937	293.7015	-1.3611	Apr 1, 97	23:28:0.00
Visible	1741.7981	313.8199	5.7429	Apr 1, 97	23:30:0.00
Visible	1210.5453	341.1353	14.2021	Apr 1, 97	23:32:0.00
Visible	1149.6434	26.6257	15.5305	Apr 1, 97	23:34:0.00
Visible	1614.2258	58.2987	7.3505	Apr 1, 97	23:36:0.00
Visible	1791.0127	310.1286	5.1344	Apr 2, 97	01:06:0.00
Visible	1022.3143	323.0780	18.7619	Apr 2, 97	01:08:0.00
Radar Night	574.2170	28.7628	41.9326	Apr 2, 97	01:10:0.00
Radar Night	1036.7458	92.3907	18.2101	Apr 2, 97	01:12:0.00
Radar Night	1808.0172	109.8418	4.7758	Apr 2, 97	01:14:0.00

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 ECEF 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, $\rho_{ECEF}(t)$, as

$$\hat{\rho}_{ECEF}(t) = \hat{r}_{sat}(t) - \hat{r}_{site}(t) \quad (11-50)$$

This vector is easily transformed to the topocentric horizon coordinate system [Eq. (3-28)], yielding $\rho_{SEZ}(t)$.

At first glance, one might think the range, azimuth, and elevation solutions from Algorithm 27 would be sufficient, but they're *not* for this problem because they're not "smooth" (See also Fig. 11-22). We'll define a "smooth" function as one which allows us to recover the rise/set parameters with certain processing. Alfano (1993) found that quartic blending works best with these rise/set parameters. 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. (4-10)] has a step discontinuity when the function changes from 359° to 0° , so use the tangent expression to capture the time-varying components that define β . Finally, for el , use a trigonometric equation that compares the limiting value with the time-varying values. Alfano (1993) uses

$$\begin{aligned} f_{range}(t) &= \rho(t) - \rho_{\text{LIM}} \\ f_{\beta_{lim}}(t) &= \rho_E(t) + \rho_S(t) \tan(\beta_{lim}) \Big|_{\text{High}, \text{Low}} \\ f_{el_{lim}}(t) &= \cos^{-1} \left[\frac{\cos(el_{lim}(t))}{|\dot{r}_{sat}(t)|} \right] - el_{lim}(t) - \cos^{-1} \left[\frac{\cos(el)}{|\dot{r}_{sat}(t)|} \right] + el \Big|_{\text{High}, \text{Low}} \end{aligned} \quad (11-51)$$

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. 11-22 shows representative examples.

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 75 shows the method. Generating site and satellite ephemerides is identical to the simple numerical approach but permits larger time steps. Eq. (C-47) (see Algorithm 75) 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

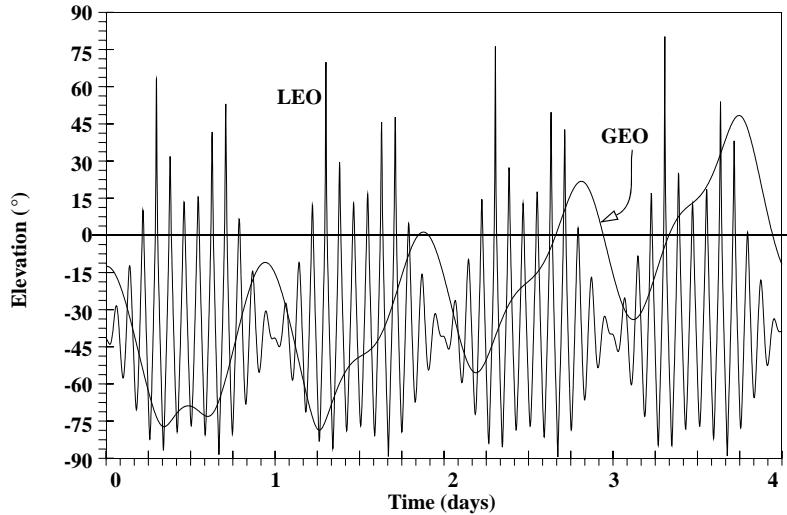


Figure 11-22. Elevation Versus Time. The standard functions of azimuth and elevation vary too much to permit an accurate solution with only a fourth-degree polynomial fit. The curves show elevation values for low-Earth and GEO orbits. Notice the slight long-periodic variation evident in the GEO orbit. That's because it's a 55° inclined orbit.

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 $F_{qrt}(\tau)$ 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. (11-51). 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 11-23 shows the overall situation, as it applies to the rise/set problem.

Once we evaluate each of the constraints, we must “add” the results to determine actual rise and set times. In Fig. 11-23, there are only small visibility opportunities due to the many constraints.

Algorithm 75 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 75: RISE/SET

$$(\dot{\vec{r}}_{sat}, \dot{\vec{v}}_{sat}, \phi_{gd}, \lambda, h_{ellp}, JD, \rho_{lim}, \beta_{lim}, el_{lim} \Rightarrow t_{rise/set}, \beta_{rise/set}, el_{rise/set})$$

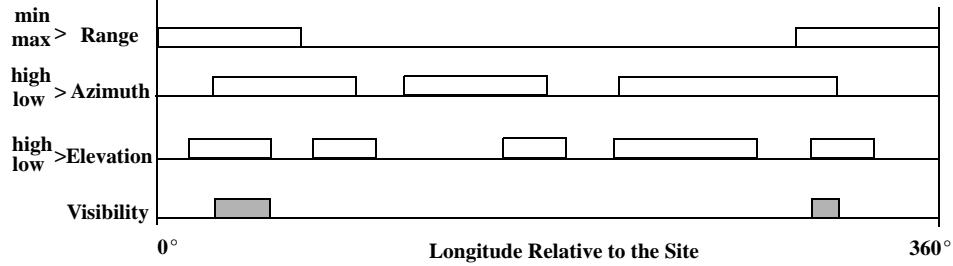


Figure 11-23. How Restrictions Affect Site Visibility. This figure shows how multiple restrictions add to yield the final answer. Rectangles show when a respective constraint is met. The satellite is visible only when all three restrictions are met. We formulate the solution for longitude values from 0° to 360°. The range function will always be symmetrical because the longitude is relative to the site. Azimuth and elevation aren't symmetrical.

Find the site vector (ECEF) and the transformation from ECEF to SEZ:

$$\text{SITE}(\phi_{gd}, \lambda, h_{ellp} \Rightarrow \vec{r}_{siteECEF})$$

$$\begin{bmatrix} \frac{SEZ}{IJK} \end{bmatrix} = \begin{bmatrix} \sin(\phi_{gd})\cos(\lambda) & \sin(\phi_{gd})\sin(\lambda) & -\cos(\phi_{gd}) \\ -\sin(\lambda) & \cos(\lambda) & 0 \\ \cos(\phi_{gd})\cos(\lambda) & \cos(\phi_{gd})\sin(\lambda) & \sin(\phi_{gd}) \end{bmatrix}$$

Generate a file of position vectors for the satellite:

$$\text{PKEPLER}(\vec{r}_{sat}, \vec{v}_{sat}, \Delta t \Rightarrow \vec{r}_{sat}(t), \vec{v}_{sat}(t)) \text{ (ECI)}$$

Evaluate the functions at each time:

LOOP:

Convert the position vector to ECEF and find range and elevation.

$$\text{FK5}(\vec{r}_{ECI}, \vec{v}_{ECP}, yr, mo, day, UTC, \Delta UT1, \Delta AT, x_p, y_p \Rightarrow \vec{r}_{ECEF}, \vec{v}_{ECEF})$$

$$\vec{\rho}_{ECEF}(t) = \vec{r}_{ECEF}(t) - \vec{r}_{siteECEF}(t)$$

$$\vec{\rho}_{SEZ}(t) = \begin{bmatrix} \frac{SEZ}{IJK} \end{bmatrix} \vec{\rho}_{ECEF}(t)$$

$$\rho(t) = |\vec{\rho}_{SEZ}|$$

$$\sin(el(t)) = \frac{\rho_z(t)}{\rho(t)}$$

Evaluate the functions at each time:

$$f_{range}(t) = \rho(t) - \rho_{lim}$$

$$f_{\beta_{lim}}(t) = \rho_E(t) + \rho_S(t) \text{TAN}(\beta_{lim}) \Big|_{High, Low}$$

$$f_{el_{lim}}(t) = \cos^{-1} \left[\frac{\cos(el_{lim}(t))}{|\vec{r}_{sat}(t)|} \right] - el_{lim}(t) - \cos^{-1} \left[\frac{\cos(el)}{|\vec{r}_{sat}(t)|} \right] + el \Big|_{High, Low}$$

Search for rise/set (or any other function) times.

- Find the coefficients (α_{qi}) from Eq. (C-47) using each of the five functions.
- Use these coefficients to find the roots of f_{qi} in the first equation of Eq. (C-47). This will find each τ_{root} of the high/low constraints.

$$f_{qi}(\tau) = \alpha_{qi5}\tau^5 + \alpha_{qi4}\tau^4 + \alpha_{qi3}\tau^3 + \alpha_{qi2}\tau^2 + \alpha_{qi1}\tau + \alpha_{qi0}$$

- Combine all five results and determine a single τ_{root} for each time, satisfying all constraints
- Recover the original values for range, azimuth, and elevation. For instance: Find new coefficients from Eq. (C-47) using the time values, and then use the real root between 0.0 and 1.0 and substitute into the first equation of Eq. (C-47). The solution is the time answer, τ .
- Any other parameters at the same time (range, azimuth, etc.) are found using coefficients from Eq. (C-47) formed with the function values, and the single root found above.

▼ Example 11-7. Predicting Satellite Rise and Set Times.

GIVEN: Satellites with data shown in Table 11-5.

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 $h_{ellp} = 2.918$ km.

TABLE 11-5. Orbital Elements for Example 11-7. For each satellite, $\omega = \Omega = M = 0.0^\circ$.

Object	n (rev/solar day)	e	i (°)
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

For quartic blending, advance the site and satellite in three-minute intervals and check the limit functions for crossing times.

Table 11-6 shows the viewing times for a one-day simulation using the parameters of Table 11-5. 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 time. Notice this method's results differ by less than one second from that of a truth model using a 0.1-second step.

TABLE 11-6. One-day Timetable Solutions for Rise/Set of Satellites 6 and 8, Respectively (for Example 11-7). 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

Despite the positive results in this example, it's possible to show another satellite orbit for which the parabolic and quartic methods are remarkably close. The reason is the shape of the function curve, which will be different for every satellite (recall Fig. 11-22). In all cases, the form of the function should be closely examined before implementing any operational design. In general, the quartic method works better than the parabolic, and the shorter the ephemeris time step, the better the estimate.



11.7 Performing Close Approach Analyses

Close Approach refers generically to the time when two satellites are closest to each other, but can also be defined as the interval of time when the satellites are within some threshold distance. The analysis process is sometimes called **Conjunction Assessment**. A related term you may encounter is **Collision Avoidance** (sometimes abbreviated COLA)—evaluating maneuvers to prevent a predicted close approach. Solution techniques apply equally for operations determining radio frequency interference (RFI) and laser clearinghouse because they use the same principles.

Several events highlight the need for accurate close approach calculations. The Iridium-33 (24946) - Cosmos-2251 (22675) conjunction on February 10, 2009 generated more than 1875 pieces of debris. On January 11, 2007, the Chinese conducted an Anti-

Satellite Test (ASAT) and hit an aging Chinese weather satellite (25730) with a missile launch from Xichang. The resulting debris spread quickly and entered many orbital regimes. Figure 11-24 shows snapshots of the debris cloud. Over 3500 pieces of debris were generated and by March 2012, over 3100 pieces still remained in orbit.

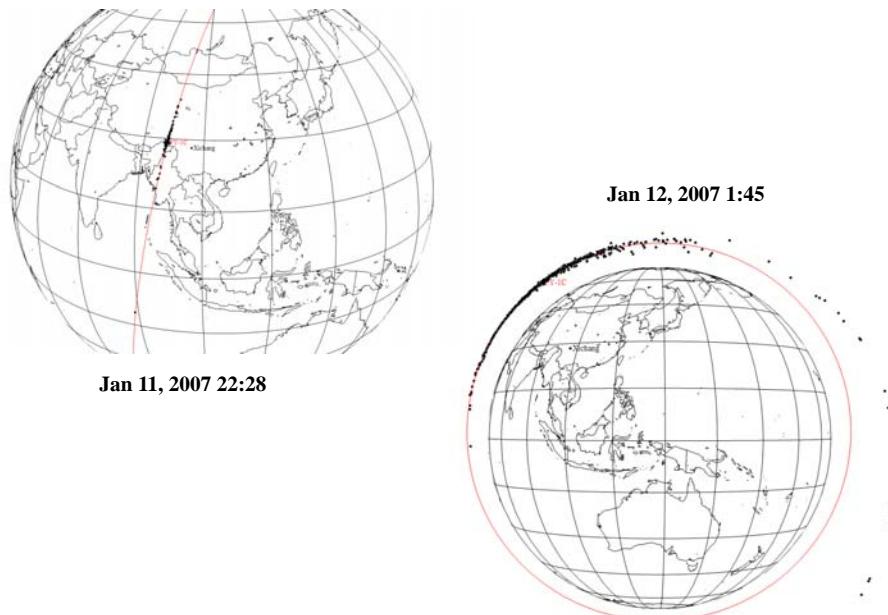


Figure 11-24. Initial Debris Spread of debris. The figure shows the initial spread of the debris from the January 2007 Chinese ASAT test. The original Feng Yun 1C orbit is shown with dots representing the debris pieces. Note the large range of eccentric orbits produced about 4 hours past the event. (STK images courtesy of AGI).

On July 24, 1996, a French satellite (Cerise 23606, 700 km altitude polar orbit) collided with a piece of debris from an Ariane rocket-body that was launched in 1986. Analysis has suggested that Cosmos 1394 (13381) and Cosmos 926 (10137) had a conjunction in December 1991 (980 km altitude, 83° orbit), and the CZ-4 3rd stage (19468) and a DMSP rocket body collided in January 2005 (885 km altitude over South Pole).

Satellites have moved to avoid a predicted conjunction. On Shuttle Missions 44 and 48, debris was predicted to enter a **safety zone** resulting in NASA's maneuvering the Space Shuttle away from the approaching objects.* Analysis of the shuttle on Mission

* The safety zone is a pre-determined region used to screen objects that may come near the satellite of interest. Don't confuse this with the covariance (also called an **error ellipsoid** or triaxial spheroid) that represents the uncertainty in the location of a satellite as determined through Orbit Determination. The safety zone is generally larger than the covariance, but it may use the covariance as a guide. They are all generally largest in the along-track direction.

57 predicted a close encounter with a spent Cosmos booster (Kennedy, 1993), so NASA again maneuvered the Shuttle to avoid it. Numerous Space Shuttle flights have been hit by small pieces of debris. One of the missions for the ***International Space Station*** (ISS) is to conduct micro-gravity experiments. Each time the ISS executes a maneuver, the thrust forces upset the delicate micro-gravity environment and place the experiments at risk.

Avoiding satellite conjunctions is one way to prevent producing additional debris. The GEO satellite population is increasing rapidly for numerous missions, predominantly communication. Some satellites have been “stacked” in the “same” slot—that is, their station keeping boxes may overlap each other. As we deploy more satellites, the ability to determine periods when they are close to one another is a much more significant challenge for satellite operators.

Since May 2004, the ***Center for Space Standards and Innovation*** (CSSI) has been providing daily reports of likely conjunctions within a distance threshold of 5 km for the upcoming week for all objects in Earth orbit using the full catalog of unclassified NORAD TLEs available to the public. The program is called Satellite Orbital Conjunction Reports Assessing Threatening Encounters in Space (SOCRATES, Kelso and Alfano 2005). The information is publicly posted twice daily at <http://www.celesttrak.com/SOCRATES/>.

SOCRATES-GEO grew from the initial SOCRATES effort and began operations in December 2007 focusing on geosynchronous satellites. In GEO (actually ± 300 km from GEO altitude), over 25% of the total known population is comprised of operational satellites, where precise orbital data is available directly from the operators (in December 2012, there were 1540 total objects of which 641 were debris (259 GEO Transfer Orbits and 258 HEO), about 390 were dead satellites, and 509 were operational. These satellites represent an important region because it’s a limited resource, close locations are desired for many satellites, existing space surveillance is not very good, and any debris created from a collision would impact hundreds of additional satellites for many centuries. Owner ephemerides include all maneuvers whether planned or already executed. The maneuver information is possibly the most important and unique aspect of the owner data.

Further expansion of SOCRATES-GEO occurred on April 12, 2010, when the Space Data Association Ltd. (SDA) selected Analytical Graphics, Inc. (AGI) to develop and operate its Space Data Center (SDC). The SDA is a non-profit organization that brings together satellite operators who value controlled, reliable and efficient data-sharing critical to the safety and integrity of the space environment and the Radio Frequency (RF) spectrum. The SDA was founded in 2009 by Inmarsat, Intelsat, and SES—three of the leading global satellite communications companies. By collecting and analyzing its authoritative radio frequency, close approach, ephemeris and points-of-contact data, the SDA’s Space Data Center performs Space Situational Awareness and threat mitigation analyses with previously unachievable accuracy and expedience. By 2013, the SDA membership spanned both GEO and LEO regimes, and provided CA processing for more than 70% of all operational satellites in GEO.

For this discussion, we'll call the satellite of interest the *primary* and the other satellite the *secondary*. The analysis seeks to determine several quantities: the time of closest approach, the relative distance between the two satellites, and the probability that a collision will occur.*

11.7.1 Close Approach Solutions

The process of evaluating close approaches involves several steps. First, we screen the satellites to determine if there is any possibility of an encounter. Many satellite geometries simply never approach each other. Next, satellites are propagated through a time interval to form an ephemeris. The overall accuracy is very dependent on the type and the fidelity of the propagation. The ephemerides are differenced to determine the minimum ranges. If a close approach is found, the probability of collision may be evaluated if covariance information is available. There are several techniques for the probability calculations (Alfano 2005, Chan 2003, and Patera 2001 and 2005). Finally, the actual processing may be optimized for speed (Healy, 1995). Using his Conjunction Sieve (CSIEVE) technique may provide significant speed to the overall conjunction problem.

An important operational consideration is noted. The USSTRATCOM www.space-track.org website has a disclaimer stating that "TLEs should not be used for conjunction assessment predictions." Analysis by the SDA (Morring, 2012) has suggested that the AFSPC (JSPOC) numerical (SP) vectors also contain significant errors, sometimes exceeding those of the TLEs! This is because JSPOC processes observations for satellites in an automated fashion. Batches of data arrive and are processed to form TLEs (along with a parallel operation to produce SP vectors). There is no knowledge of maneuvers before or after the current time of OD generation. During the processing, the presence of maneuvers inflates the uncertainty of any estimate, but because the TLEs are delivered without a covariance, there is little way of knowing if there is increased uncertainty in any particular TLE. Likewise, there is no way to know if future maneuvers are planned by an owner operator, and therefore not included in the TLE (or SP) prediction. The solution is to have cooperation between organizations, and data fusion to improve accuracy.

Screening Filters

Hoots, Crawford and Roehrich (1984) published a detailed look at close approach processing, detailing several filters to reduce the computational complexity. Filters are often proposed to alleviate some unnecessary calculations in the problem. However, realize that filters will *always* introduce some degree of uncertainty (missed conjunc-

* Besides predicting times, distances, and probabilities, close-approach techniques have other uses. One example 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 (an uncorrelated target, UCT). By running the close-approach technique backwards, we can usually find the maneuver time and separate the pre- and post maneuver orbits.

tions*) into the solution depending on the type of filter, type of orbital element used (mean or osculating), and the time span for the conjunction calculations.

A first filter 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 ($r_{p_{max}}$) and the smallest apogee ($r_{a_{min}}$) is greater than the desired distance tolerance, the two satellites will never come close enough to consider (unless a maneuver is applied).

$$\text{No conjunction if } r_{p_{max}} - r_{a_{min}} > \text{Dist} \quad (11-52)$$

This test is good for all closed orbits, including those in different planes. Orbits that pass this initial test need no further processing. Figure 11-25 illustrates the geometry.

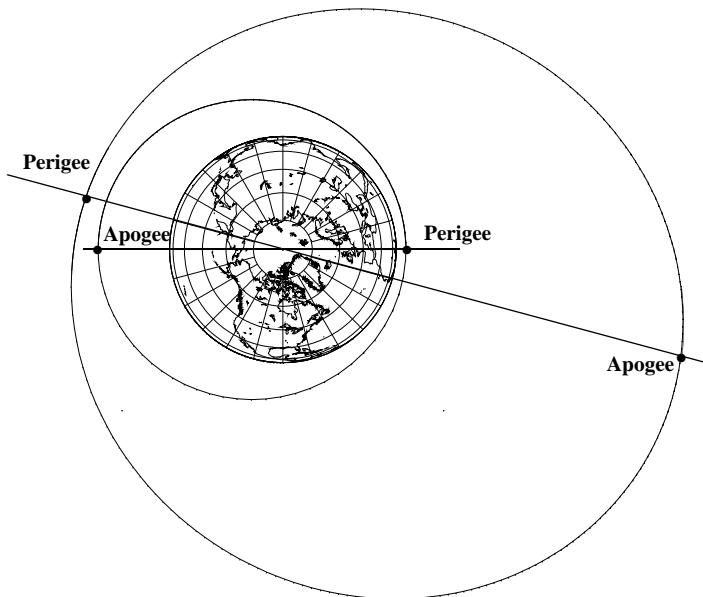


Figure 11-25. Apogee-Perigee Filter. The example shows two orbits, one with $e = 0.2$ and $a = 8637.018$ km and the other with $e = 0.35$ and $a = 17,444.459$ km. Even with perturbations, the apogee of the smaller orbit ($r_a = 10,364$ km) will never approach the perigee of the larger ($r_p = 11,339$ km). These orbits lie in the same plane, but this concept works for orbits in different planes.

* There are two common errors associated with missed conjunctions. **Type I** errors occur when a maneuver does not occur, but is predicted to take place, a false positive. **Type II** errors happen when the conjunction actually occurs, but is not detected—a false negative.

Woodburn et al. (2009) note some difficulties with this filter as the orbits get closer. Nearly circular orbits often have apogee (and perigee) values that are nearly the same in magnitude, yet perturbations can move them around substantially in the orbit. If the screening distance is close to the difference in the orbital parameters caused by perturbations, a simple distance test may be exceeded at a future time when the orbit has been perturbed. Another example arises for an apogee at the beginning of the period for one satellite, and a perigee for the other satellite at the end of the period that are closer than the screening distance. Here, the timing is off. For the apogee-perigee filter, it's rather simple to pad (increase) the screening distance to account for the variability of the orbital elements, but as the time span increases, the required pad to account for perturbations can become very large. This is an unfortunate reality of using filters.

Woodburn et al. (2009) discuss two additional filters. The first is an orbital path filter which seeks to find the shortest distance between the two elliptical orbits. Unfortunately, even before considering difficulties of mean and osculating elements, Woodburn notes that the approach does not seem to perform very well. The other filter is a time filter that determines if the satellites are close while at the intersecting nodes of the two orbits. Although this filter has promise in reducing the overall numbers of computations, it is still plagued by orbital element types. It seems that if accuracy is the goal, the first (apogee-perigee) filter may be the only one that has a reasonable chance to save calculations as well as maintaining accuracy.

Miss Distance Determination

An obvious solution to the time and distance portion of the close-approach problem is to step sequentially along the orbits of the 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 for analyses. We can easily estimate the minimum distance using Eq. (C-38). Although straightforward, this approach requires 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, non maneuvering trajectories (Note that if there is a maneuver, the process must be restarted because the derivatives will not be continuous across the maneuver). 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 (the resulting curve is ‘smooth’ to first order). The main goal is determining the distance between the two satellites. The method may use any propagation scheme, permitting us to select a desired speed and accuracy. The answers result from closed-form solutions and eliminate the need for iteration.

Begin with the geocentric equatorial (ECI) position vectors* of the *primary*, \hat{r}_p , and *secondary*, \hat{r}_s , satellites at time t . The *relative-distance vector*, \hat{r}_d , and its time derivatives become

* We generally use inertial vectors, although Earth-fixed vectors are feasible as long as you're consistent. Remember that you can't evaluate orbital elements from Earth fixed vectors.

$$\begin{aligned}\hat{\vec{r}}_d &= \hat{\vec{r}}_s - \hat{\vec{r}}_p \\ \dot{\hat{\vec{r}}}_d &= \dot{\hat{\vec{r}}}_s - \dot{\hat{\vec{r}}}_p \\ \ddot{\hat{\vec{r}}}_d &= \ddot{\hat{\vec{r}}}_s - \ddot{\hat{\vec{r}}}_p\end{aligned}\tag{11-53}$$

where *any* orbit propagator can produce the ephemerides for the primary and secondary satellites. Large time steps are possible with the propagator because of the splining techniques we use. If the satellite's acceleration vectors aren't available, we can approximate the vector components using Algorithm 64, but this simplification includes only J_2 perturbative forces, making it only an approximation. The *distance function*, $f_d(t)$, and its time derivatives are found using dot products. The distance function is the square of the distance so we don't need to evaluate a square root.

$$\begin{aligned}f_d(t) &= \hat{\vec{r}}_d \cdot \hat{\vec{r}}_d \\ \dot{f}_d(t) &= 2(\dot{\hat{\vec{r}}}_d \cdot \hat{\vec{r}}_d) \\ \ddot{f}_d(t) &= 2(\ddot{\hat{\vec{r}}}_d \cdot \hat{\vec{r}}_d + \dot{\hat{\vec{r}}}_d \cdot \dot{\hat{\vec{r}}}_d)\end{aligned}$$

We evaluate these equations for a sequence of times until two adjoining times (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 that 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 cubic polynomial equation $P_{cd}(\tau)$ [Eq. (C-42), see Algorithm 76] that corresponds to $\dot{f}_d(t)$. Use the additional “*d*” subscript to indicate the distance function. If you use the derivative of the function, the cubic splining in Eq. (C-42) still applies—simply use the first and second derivatives instead of the function and first derivative. Remember that τ is a normalized time variable that 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. This is necessary to distinguish the results from the ellipsoidal function we'll discuss shortly. If

$$\left. \frac{dP_{cd}(\tau)}{d\tau} \right|_{\tau=\tau_{droot}} > 0$$

a local minimum range exists. You must still determine the time and range corresponding to τ_{droot} . Use quintic splines to capture the contribution of acceleration and to keep the solution accurate. Form the quintic spline, P_{qi_I} , using Eq. (C-48) (see Algorithm 76) and the *I*-components of Eq. (11-53). Form P_{qi_J} and P_{qi_K} similarly. The quintic polynomial uses τ_{droot} . 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 $\Delta t = t_{n+1} - t_n$.

Projections

Finding the close approach time and distance is just half the problem. We know from Chap. 10 that orbit determination produces uncertainty for each object. The uncertainties (covariance) are usually aligned differently and the computations can be complex. Finding the probability in 3 dimensions requires integrating over each axis of the position. Alfano (2007) shows the equation (with xyz as position components).

$$Prob_{conj} = \frac{1}{\sqrt{8\pi^3}\sigma_x\sigma_y\sigma_z} \int \int \int_{Vol} \text{EXP} \left[\frac{-x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2} - \frac{z^2}{2\sigma_z^2} \right] dx(dy)dz$$

Projecting the error ellipses onto a plane considerably reduces the complexity, but it relies on several assumptions. The physical objects are often treated as spheres, thus eliminating the need for attitude information. The positional errors are assumed to be zero-mean, Gaussian, uncorrelated, and constant for the encounter. The satellites relative motion is considered linear for the encounter by assuming the effect of the relative acceleration is dwarfed by that of the velocity. The relative velocity at the point of closest approach is deemed sufficiently large to ensure a brief encounter time and static covariance. The **encounter region** is defined when one object is within a standard deviation (σ) combined covariance ellipsoid shell scaled by a factor of n . This user-defined, three-dimensional, $n\sigma$ shell is centered on the primary object; n is typically in the range of 3 to 8 to accommodate conjunction probabilities that range from 97.070 911% to 99.999 999%.

Because the covariances are assumed to be uncorrelated, they are simply summed to form one large, combined, covariance ellipsoid that is centered at the primary object (Fig. 11-26).*

The physical dimensions of each satellite are also summed. The secondary object (defined here as the combined satellite size) passes quickly through this ellipsoid creating a tube-shaped path. A conjunction occurs if the secondary sphere touches the primary sphere, i.e., when the distance between the two projected object centers is less than the sum of their radii. The radius of this collision tube is enlarged to accommodate all possibilities of the secondary touching the primary by combining the radii of both objects.

A plane perpendicular to the relative velocity vector is formed and the combined object and covariance ellipsoid are projected onto it. This is the *encounter plane*. (Fig. 11-27).

As stated previously, the encounter region is defined by an $n\sigma$ shell determined by the user to sufficiently account for conjunction possibilities. Within that shell, the tube is

* No special operation is needed to sum the matrices (Chan 2008:14, Eq 2.3) so that simple matrix addition is sufficient.

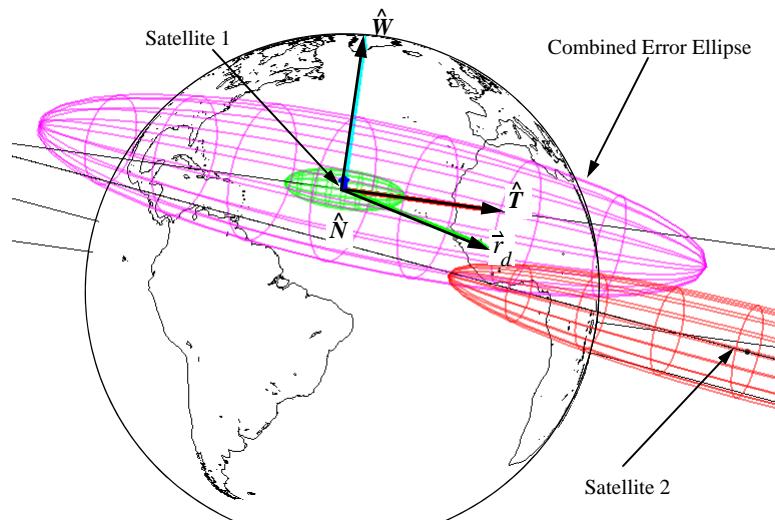


Figure 11-26. Conjunction Encounter Geometry. The figure shows a generic close approach scenario. Each satellite has an associated velocity and covariance. The combined error ellipse is simply the summation of each individual covariance. The relative velocity vector is used to define the encounter plane. The standard deviations, (σ s), should be the same for each error ellipse. (STK image courtesy of AGI)

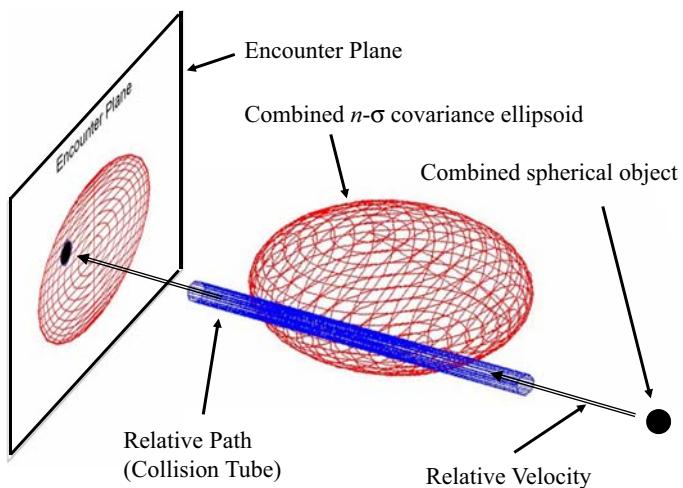


Figure 11-27. Conjunction Encounter Visualization and Reduction. As the combined spherical object enters the secondary covariance, the tube can be a straight path (shown). Note that the combined object size is often considerably smaller than the covariance matrix.

straight and rapidly traversed, allowing a decoupling of the dimension associated with the tube path (i.e., relative velocity). The tube becomes a circle in the encounter plane. Likewise, the covariance ellipsoid becomes an ellipse.

The relative velocity vector is associated with the time of closest approach and contributes no uncertainty to the closest approach vector. Thus, we consider just the two-dimensional projection (Akella and Alfriend 2000). The conjunction assessment here is concerned with cumulative probability over the time it takes to span the $n\sigma$ shell, not an instantaneous probability at a specific time within the shell. Along this dimension, integration of the probability density across the shell produces a number very near unity, meaning the close approach will occur at some time within the shell with near absolute certainty. The new equation is

$$Prob_{conj} = \frac{1}{2\pi\sigma_x\sigma_y} \int_{-obj}^{obj} \int_{-\sqrt{obj^2 - x^2}}^{\sqrt{obj^2 - x^2}} \text{EXP}\left[-\frac{1}{2}\left(\left(\frac{x + x_m}{\sigma_x}\right)^2 + \left(\frac{y + y_m}{\sigma_y}\right)^2\right)\right] dy dx \quad (11-54)$$

Error Ellipse Entry/Exit - The Ellipsoidal Function

Once a candidate close approach is found, we examine the safety zone around the primary satellite. Remember that the covariance is the uncertainty in the predicted state while the safety zone is [arbitrarily] created to balance program risk and safety. The two usually differ. We need this zone because propagation and differential correction techniques (Chaps. 8, 9, and 10) introduce errors into the orbit solution—the covariance matrix. Although a single number indicates the satellite's mean location at the point in time where the satellites are closest, the propagation of the covariance describes the satellite's probability density for a certain region about the propagated state. The size of the covariance matrix is dependent on the original formation in the differential correction process, and the techniques used to move it through time.

We need to obtain a function that determines the entry and exit conditions for the secondary satellite from the ellipsoidal region (combined covariance) about the primary satellite. We're only interested in a two-dimensional result because the direction associated with relative velocity does not contribute to the uncertainty of the close approach vector, so by projecting \vec{r}_d along the primary axis (T) and using the Pythagorean theorem, we can use the general form of an ellipse (with a and b as the major and minor axes) to achieve the desired result. Notice that this assumes equal minor axes.

$$\frac{N^2}{a^2} + \frac{T^2}{b^2} - 1 = 0 \quad N = \frac{\vec{r}_d \cdot \vec{v}_p}{\vec{v}_p} \quad T^2 = \vec{r}_d \cdot \vec{r}_d - N^2$$

Substituting N and T into the ellipse equation gives us the **ellipsoidal function**,

$$f_e(\tau) = \frac{\frac{(\vec{r}_d \cdot \vec{v}_p)^2}{(\vec{v}_p \cdot \vec{v}_p)}}{a^2} + \frac{(\vec{r}_d \cdot \vec{r}_d) - \frac{(\vec{r}_d \cdot \vec{v}_p)^2}{(\vec{v}_p \cdot \vec{v}_p)}}{b^2} - 1$$

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. The primary satellite's velocity is needed 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

$$f_e(\tau) = \frac{|\vec{r}_d \cdot \vec{r}_d|}{a^2} - 1$$

To determine the times of entry and exit, find the coefficients α_{ce} from Eq. (C-43) (see Algorithm 76) 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

$$p_1 = f_e(0), (t = t_n) \quad p_4 = f_e(1), (t = t_{n+1})$$

But you'll need to reconstruct the intermediate points from the distance functions of Eq. (11-53), 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

$$p_2 = f_e(\tau_1), \tau_1 = 1/3 \quad p_3 = f_e(\tau_2), \tau_2 = 2/3$$

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.

```

 $\tau_2 = \tau_{droot}$ 
IF  $\tau_{droot} < 0.5$ 
     $\tau_1 = (1 + \tau_{droot})/2$ 
ELSE
     $\tau_1 = \tau_{droot}/2$ 

```

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-42). 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 ($\Delta t = t_{n+1} - t_n$)

$$\text{An exit exists if } \left. \frac{dP_{ce}(\tau)}{d\tau} \right|_{\tau=\tau_{eroot}} > 0, \text{ otherwise, it's an entry.}$$

$$t(\tau_{eroot}) = t_n + \tau_{eroot}\Delta t$$

Determining when Two Ellipsoids touch

Analytical methods for determining if two ellipsoids share the same volume were investigated by Alfano and Greer (2003). There are no dimensional distortions, coordinate rotations, transformations, or eigenvector computations needed. The formulation involves adding an extra dimension in the solution space and examining eigenvalues that are associated with degenerate quadric surfaces. This process is involved, but also part of the subsequent probability calculations, thus, we do not explore it here. The reader is referred to the original article for implementation.

Maximum Probability and Dilution Regions

Once the entry/exit points have been found, determining the probability of collision provides further data for a decision maker to consider. *ANCAS* tells you when the orbits will approach each other, but the satellite positions are uncertain. 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 300 m of the Space Shuttle. Should the Shuttle be commanded to move? 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? Are there legal considerations? Questions like these are the basis for lengthy analyses to determine the correct option—and the actual probability calculation is just one step.

Probability calculations for conjunction assessment should ensure sufficient accuracy for meaningful results. Because all operational decisions are ultimately made with respect to the amount of acceptable risk, the action threshold should not be based solely on an unacceptable miss distance but rather on an unacceptable collision probability. This is already done with the ISS where avoidance maneuvers are initiated if the collision probability exceeds an acceptable risk threshold (e.g., 1×10^{-5}).

The accuracy of positional covariance matrices resulting from Orbit Determination of sparse data is questionable. Covariance matrices formed in this manner often provide overly optimistic results. Foster and Frisbee (1998) noted that “The primary problem with state error covariances determined from observations of objects in Earth orbit is that they are not truly reflective of the uncertainties in the dynamic environment.” To address this concern, they devised a method to scale covariances provided to NASA by AFSPC (JSPOC). The covariance values are simply multiplied at the propagated times.*

* Scaling the initial covariance could also work but the scaling would need to properly account for force models, process noise, etc, for the covariance propagation. If the scaling is accomplished after propagation, it would likely need to be a function of time. The exact process used by AFSPC (JSPOC) is unknown, but may include both scaling options.

It is possible to find the absolute worst-possible covariance size and orientation that maximizes the probability for a given encounter where the object sizes are known and the closest-approach distance is fixed—the maximum probability. **Maximum probability** explains the effects of positional uncertainty on the Gaussian probability computation for the orbit conjunction (Alfano 2005, 2005b). It's also possible to find covariance parameters that maximize the probability for various covariance shapes as determined by the **aspect ratio** (i.e., the ratio of major-to-minor axes of the projected combined covariance ellipse). Regions of maximum probability are mapped for various satellite sizes, encounter geometries, and covariance sizes and shapes. These regions are examined to assess probability dilution which should be used to determine if the data is sufficiently accurate to support a probability calculation.

If the maximum probability is below a predefined user threshold, then no further calculations are needed. Even in the absence of known covariances, the numerical approximations that follow will provide the user with the worst-case collision potential for aspect ratios ranging from 1 to 50, capturing 99% of all conjunction possibilities. Beyond those bounds, an iterative search method is recommended. Such analysis can be insightful when one only has knowledge of the miss distance and physical object sizes.

Two situations are possible. If the object is non-spherical, we reorient for various configurations to determine the maximum probability. We can also rescale the covariance to reflect uncertainty in the covariance estimates. Reorienting finds a maximum probability (highest point on the curve in Fig. 11-28) while doing both finds the absolute maximum probability (a point above the highest probability in Fig. 11-28).

Mathematically then, after combining and projecting the covariances, we find the two-dimensional result in Eq. (11-54). If we include the aspect ratio, or the ratio of σ_y to σ_x , the probability expression becomes

$$Prob_{conj} = \frac{1}{2\pi\sigma_x^2 AR} \int_{-r_{obj-comb}}^{r_{obj-comb}} \int_{-\sqrt{r_{obj-comb}^2 - x^2}}^{\sqrt{r_{obj-comb}^2 - x^2}} \exp\left[-\frac{1}{2}\left(\left(\frac{x + x_m}{\sigma_x}\right)^2 + \left(\frac{y + y_m}{\sigma_x AR}\right)^2\right)\right] dy dx \quad (11-55)$$

In the encounter plane, the x_m and y_m components are varied as a function of the fixed relative distance (*dist*) and the angle representing the orientation of each satellite. A case of interest involves an infinite aspect ratio ($AR = \infty$). For this case, the combined object radius is normalized with respect to the miss distance

$$r = \frac{r_{obj-comb}}{dist}$$

Using Eq. (11-55) for maximum probability determines the worst-case conjunction scenario by finding the combined Gaussian probability density that maximizes collision probability. The only parameters required are the distance (*dist*) of closest approach, the radius of the combined object ($r_{obj-comb}$), and the aspect ratio of major-to-minor projected covariance ellipse axes (*aspect ratio AR*). The major axis of the combined cova-

riance ellipse is aligned with the relative position vector (at the point of closest approach) such that it passes through the center of the combined object. The projected, combined object is assumed circular with its probability mass distributed symmetrically about the major axis. This means that only a single axis length needs to be examined to maximize the probability, the other being determined from the aspect ratio.

Clearly, if the combined object footprint contains the covariance ellipsoid center, the minor axis' standard deviation can be chosen to drive the maximum probability to one. For spherical objects this occurs when the predicted miss distance is less than the combined object size ($dist < r_{obj-comb}$). This is the limiting case and need not be addressed; it's inferred that a decision maker faced with such a predicted "direct hit" would not need a probability calculation. The method described here only applies when the combined object does not encompass the covariance center ($dist \geq r_{obj-comb}$). Given the combined object radius and distance from center, the minor axis size can be determined by maximizing a two-dimensional probability expression. Once determined, the worst-case collision probability is calculated.

Alfano (2005) shows the derivation of the maximum probability. Using error functions instead of integrals in Eq. (11-54), and finding σ in terms of r ,

$$\sigma = \sqrt{\frac{2r}{\ln\left(\frac{1+r}{1-r}\right)}}$$

the final result for an infinite AR is

$$Prob_{conjmax} = \frac{1}{2} \left\{ erf\left[\frac{r+1}{2\sqrt{r}} \sqrt{-\ln\left(\frac{1-r}{1+r}\right)}\right] + erf\left[\frac{r-1}{2\sqrt{r}} \sqrt{-\ln\left(\frac{1-r}{1+r}\right)}\right] \right\} \quad (11-56)$$

The length of the single axis has been optimized to produce the greatest overall probability. The reader is reminded that σ , is a normalized value and should be multiplied by $dist$ to produce σ_x .

Due to the near-linear nature of the previous equation, $Prob_{conjmax}$ can be approximated to at least three significant figures with the equations (and an infinite AR)

$$Prob_{conjmax} = 0.48394(r) \text{ if } r < 0.8 \quad (11-57)$$

$$Prob_{conjmax} = 0.21329 \exp(1.01511 r) - 0.09025 \text{ if } 0.8 \leq r < 1.0$$

For the limiting case where $r = 1$, $Prob_{conjmax}$ is 0.5. Eq. (11-57) is used for computational prescreening. As the aspect ratio grows, so does $Prob_{conjmax}$. To properly assess the maximum probability, it is recommended that the reader not use Eq. (11-57) for anything other than prescreening unless it is certain that the encounter's aspect ratio is near infinity ($AR > 1000$).

If you know the AR (and trust your covariance), the maximum probability is

$$\alpha = AR \frac{(r_{obj-comb})^2}{(dist)^2} \quad Prob_{conjmax} = \left(\frac{\alpha}{1 + \alpha} \right) \left(\frac{1}{1 + \alpha} \right)^{1/\alpha} \quad (11-58)$$

The **probability dilution** region is that region where the standard deviation of the combined covariance minor axis (σ_x) exceeds that which yields $Prob_{conjmax}$. For fixed object sizes and miss distance, the σ_x that produces $Prob_{conjmax}$ defines the dilution region boundary as shown in Fig. 11-28. To left of the vertical line, greater positional accuracy (smaller σ_x) decreases collision probability. To the right of the vertical line, lesser positional accuracy (greater σ_x) also decreases collision probability. Therefore,

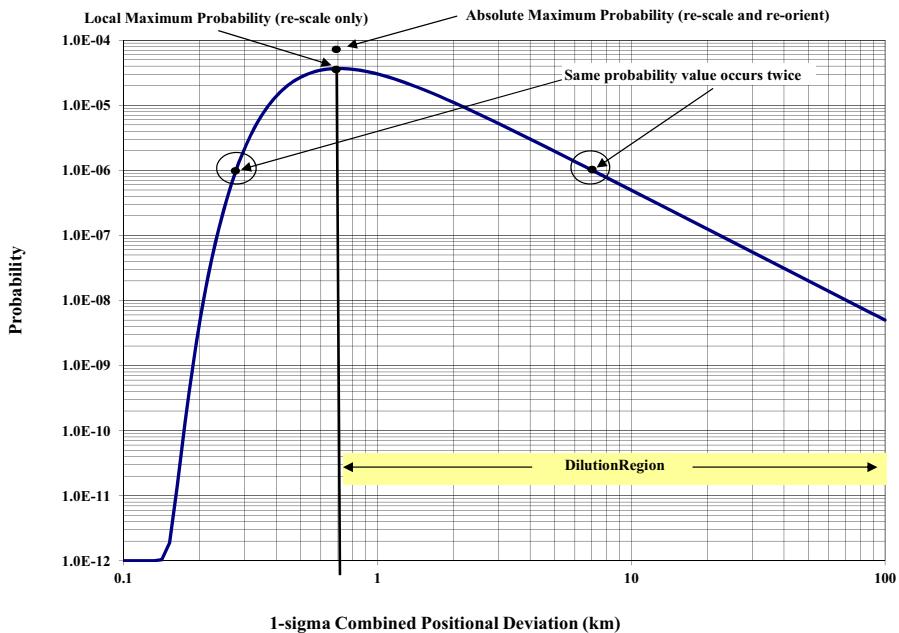


Figure 11-28. Dilution Region for a Nominal Encounter. The probability is often the single value used in conjunction decision making. Because there can be multiple probabilities associated with good and bad input data, the maximum is often used to bound the upper limit of an event, without requiring the accuracy of the initial data uncertainty. The dilution region occurs where the input data is not well calibrated, or has large uncertainties. Maximum probability is found by reorienting non-spherical objects. The absolute maximum is found by re-scaling the covariance matrix as well.

large measurement noise values or improperly known measurement noise can produce the same probability (10^{-6} is given as an example in Fig. 11-28). Although both calculations are mathematically correct, we may find greater confidence in calculations where the measurement statistics are better known, or smaller.

If the calculated (true) probability is below an action threshold, the user could be led to conclude the encounter does not pose a threat. With great surety one can infer this when outside the dilution region, provided that the positional uncertainty is correct. This is because the low probability is coupled with great confidence in the predicted miss distance. This is not the case in the dilution region where there is little confidence in the predicted miss distance.

Faced with a low probability below an action threshold, you could use the dilution region boundary as a discriminator. If the positional uncertainty is not small enough to avoid the dilution region, then it would be prudent to attempt to reduce the uncertainty and reassess the true probability. If the uncertainty still yields a low probability below action threshold in the dilution region, you should also consider the maximum probability (worst-case scenario for predicted miss distance) as well as the calculated one. This will help to ensure that a decision maker is not lulled into a false sense of security by a low probability calculation that may be misleading.

A final note on the physical size of the satellites. Alfano (2006) and Chan (2009) summarize the current status of conjunction calculations, including the physical dimensions of the objects. The hard-body radii are summed to form a combined object radius, and the positional covariances are also summed. Peterson (2005) also handles spherical satellite shapes, requiring a re-evaluation of the integrals. A width-factor histogram for a single run of SOCRATES can give the reader a snapshot of variability encountered when considering the size of the satellites. These methods can be used in cascading filters to determine if the risk of collision is sufficiently low to dismiss further computations.

Implementing Close Approach Processing

A combined close approach solution couples the geometrical and statistical operations discussed thus far. 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.5.1 to provide the best performance by ignoring all complex roots. Higher-degree 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.

Coppola et al. (2009) discuss some practical aspects to consider when parsing out the computational resources in performing the entire catalog calculations on the entire catalog. By judicious assignment of portions of the catalog, you can significantly enhance performance. As mentioned before, the CSIEVE technique (Healy, 1995) may provide significant speed as well.

Algorithm 76 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, and are only interested in the minimum distance, 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 76: CLOSE APPROACH

$$(\dot{\vec{r}}_s, \dot{\vec{v}}_s, \dot{\vec{r}}_p, \dot{\vec{v}}_p, JD_{start}, JD_{stop}, \Delta t \Rightarrow t_{close}, d_{close}, \tau_{entry}, \tau_{exit}, P_{max})$$

Run initial filter using orbital elements to eliminate obvious candidates:

$$r_{p_{max}} - r_{a_{min}} > Dist$$

Generate ephemerides for the primary and secondary:

$$\textbf{PKEPLER}(\dot{\vec{r}}_s, \dot{\vec{v}}_s, \Delta t \Rightarrow \dot{\vec{r}}_s, \dot{\vec{v}}_s)$$

$$\textbf{PKEPLER}(\dot{\vec{r}}_p, \dot{\vec{v}}_p, \Delta t \Rightarrow \dot{\vec{r}}_p, \dot{\vec{v}}_p)$$

$$\dot{\vec{r}}_d = \dot{\vec{r}}_s - \dot{\vec{r}}_p$$

$$\dot{\vec{r}}_d = \dot{\vec{v}}_s - \dot{\vec{v}}_p$$

$$\ddot{\vec{r}}_d = \ddot{\vec{r}}_s - \ddot{\vec{r}}_p$$

$$g_d(t) = \dot{\vec{r}}_d \cdot \dot{\vec{r}}_d$$

$$\dot{g}_d(t) = 2(\dot{\vec{r}}_d \cdot \ddot{\vec{r}}_d)$$

$$\ddot{g}_d(t) = 2(\ddot{\vec{r}}_d \cdot \dot{\vec{r}}_d + \dot{\vec{r}}_d \cdot \dot{\vec{r}}_d)$$

Search for the root for the time of minimum range, τ_{droot} . Roots occur where

$$\dot{g}_d = 0, \text{ and } \ddot{g}_d > 0$$

Using Eq. (C-42) with $f_d = g_d$ and $\dot{f}_d = \dot{g}_d$, form coefficients α_{ci} and

$$P_{cd}(\tau) = \alpha_{c3}\tau^3 + \alpha_{c2}\tau^2 + \alpha_{c1}\tau + \alpha_{c0}$$

Solve $P_{cd}(\tau) = 0$ to find τ_{droot}

Find the minimum distance at τ_{droot} by solving the three quintics (each IJK axis)

Using Eq. (C-48) with $f_d = \dot{\vec{r}}_d$ $\dot{f}_d = \ddot{\vec{r}}_d$ $\ddot{f}_d = \dddot{\vec{r}}_d$, form α_{qii}

$$P_{qi}(\tau) = \alpha_{qi5}\tau^5 + \alpha_{qi4}\tau^4 + \alpha_{qi3}\tau^3 + \alpha_{qi2}\tau^2 + \alpha_{qi1}\tau + \alpha_{qi0}$$

Solve $P_{qi}(\tau_{droot})$ to find the minimum distance and time

$$d_{close} = \sqrt{P_{qi_I}^2(\tau_{droot}) + P_{qi_J}^2(\tau_{droot}) + P_{qi_K}^2(\tau_{droot})}$$

$$t_c(\tau_{droot}) = t_n + \tau_{droot}\Delta t$$

Now check for the error-ellipsoid entry/exit times.

$$f_e(\tau) = \frac{\frac{(\dot{\vec{r}}_d \cdot \dot{\vec{v}}_p)^2}{a^2} - \frac{(\dot{\vec{r}}_d \cdot \dot{\vec{r}}_d)}{(\dot{\vec{v}}_p \cdot \dot{\vec{v}}_p)} - 1}{b^2}$$

$$p_1 = f_e(0)$$

$$p_2 = f_e(\tau_1), \tau_1 = 1/3$$

$$p_3 = f_e(\tau_2), \tau_2 = 2/3$$

$$p_4 = f_e(1)$$

$$P_{ce}(\tau) = \alpha_{c3}\tau^3 + \alpha_{c2}\tau^2 + \alpha_{c1}\tau + \alpha_{c0}$$

Using Eq. (C-43), solve $P_{ce}(\tau) = 0$ to find real $\tau_{eroot(s)}$.

IF τ_{eroot} exists, find the close approach entry/exit time

$$\tau_2 = \tau_{eroot}$$

$$\text{IF } \tau_{eroot} < 0.5$$

$$\tau_1 = (1 + \tau_{eroot})/2$$

ELSE

$$\tau_1 = \tau_{droot}/2$$

Solve $P_{qi}(\tau_{eroot})$ using Eq. (C-48) to find time (s)

Check derivatives for entry or exit.

$$t_{entry/exit}(\tau_{eroot}) = t_n + \tau_{eroot}\Delta t$$

Evaluate instantaneous probability

$$Prob_{conj} = \frac{1}{2\pi\sigma_x^2 AR} \int_{-\infty}^{obj} \int_{-\sqrt{obj^2 - x^2}}^{\sqrt{obj^2 - x^2}} \exp\left[\frac{-1}{2} \left\{ \left(\frac{x + x_m}{\sigma_x}\right)^2 + \left(\frac{y + y_m}{\sigma_x AR}\right)^2 \right\}\right] dy dx$$

or find the maximum probability

$$Prob_{conjmax} = \frac{1}{2} \left\{ erf\left[\frac{r+1}{2\sqrt{r}} \sqrt{-LN\left(\frac{1-r}{1+r}\right)}\right] + erf\left[\frac{r-1}{2\sqrt{r}} \sqrt{-LN\left(\frac{1-r}{1+r}\right)}\right] \right\}$$

Problems

1. Can η_{FOV} equal Λ_{FOV} ? If so, how?
2. In Example 11-1, we found several center locations for a satellite to observe using an off-axis sensor. We then found selected maximum and minimum positions from these central locations. What is the total locus of points representing the maximum locations visible to a satellite using all azimuths? Use a spacing of $\Delta\beta = 5^\circ$ to find each possible center location.
3. For Example 11-3, suppose $i = 130^\circ$. How do your answers change? Consider nodal, perigee, and drag corrections.
4. For Example 10-4, 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. 8 and 9 and see Sec. 10.7.)
5. You're tracking a satellite with orbital elements equal to

$$\begin{aligned}\hat{r}_{IJK} &= 191.955\ 268 \hat{I} + 1069.525\ 308 \hat{J} + 7098.497\ 371 \hat{K} \text{ km} \\ \hat{v}_{IJK} &= 7.206\ 807\ 1 \hat{I} - 1.780\ 491\ 4 \hat{J} + 0.060\ 407\ 9 \hat{K} \text{ km/s}\end{aligned}$$

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,

$$\begin{aligned}\hat{r}_{IJK} &= 6078.803\ 133 \hat{I} - 754.062\ 501 \hat{J} + 3748.213\ 782 \hat{K} \text{ km} \\ \hat{v}_{IJK} &= 3.641\ 897\ 3 \hat{I} - 1.770\ 935\ 1 \hat{J} - 6.256\ 019\ 3 \hat{K} \text{ km/s}\end{aligned}$$

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.

6. 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 an answer by 09:00 Monday. How do you proceed?
7. You're asked to design a new satellite-reconnaissance orbit for the satellite in Example 11-3 through Example 11-5. 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 11-3), but calculate the answers for a Sun-synchronous orbit (06:50), and a frozen eccentricity orbit ($e = 0.01$). Be sure to support your answers with numerical values.
8. 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.

9. 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. 11-2 to Fig. 11-4 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.
10. Continuing with the concept in Problem 9, 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. 4, and the information from Fig. 11-5 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.
11. Why do navigation satellite constellations generally use semisynchronous orbits?
12. Following the discussion from Problem 9, 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 a particular propagator, reduce computer resources, and speed execution. In particular, pay close attention to the function you derive for the blending and splining techniques.

13. You're a defense attorney representing a company whose satellite is accused of colliding with GALAXY 4 and knocking out television coverage just as the Denver Broncos were about to win the Super Bowl on January 25, 1998 at 16:00 Mountain Standard Time (Incidentally, the satellite actually failed on May 19, 1998). 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 radius equal to 25 m, RMS.

Defense Satellite: January 25, 1998, 12:47 UTC

$$\begin{aligned}\hat{\vec{r}}_{IJK} &= 31,497.390\,167 \hat{I} + 28,026.945\,431 \hat{J} + 276.112\,685 \hat{K} \text{ km} \\ \hat{\vec{v}}_{IJK} &= -2.044\,055 \hat{I} + 2.296\,802 \hat{J} - 0.017\,964 \hat{K} \text{ km/s}\end{aligned}$$

GALAXY 4: January 18, 1998, 13:43:45.402 UTC

$$\begin{aligned}\hat{\vec{r}}_{IJK} &= 28,478.560\,600 \hat{I} + 31,090.200\,400 \hat{J} + 4.008\,900 \hat{K} \text{ km} \\ \hat{\vec{v}}_{IJK} &= -2.267\,700 \hat{I} + 2.076\,500 \hat{J} + 0.000\,300 \hat{K} \text{ km/s}\end{aligned}$$

Assume the jury will comprehend and act on factual information—not courtroom drama. If you can win the case, how do you proceed? If not, how will the prosecution present the true facts of the case?

14. 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?

$a = 7047.9490 \text{ km}$	$a = 7048.0192 \text{ km}$
$e = 0.000\,899$	$e = 0.001\,470\,2$
$i = 98.1025^\circ$	$i = 98.4535^\circ$
$\Omega = 141.7519^\circ$	$\Omega = 334.7433^\circ$
$\omega = 67.4104^\circ$	$\omega = 119.3840^\circ$
$M = 292.8048^\circ$	$M = 240.8797^\circ$
Epoch = July 23, 1997 09:25:32.358 UTC	Epoch = July 23, 1997 08:15:32.965 UTC

15. Calculate the orbital parameters needed to repeat every 6.875 days. The satellite should complete 107 revolutions during this period. Is a Sun-synchronous orbit possible at this altitude? Finally, design a repeat groundtrack-Sun synchronous orbit to accomplish these requirements. Assume a circular orbit, but consider different inclination values. How does the initial choice of inclination affect the results?

940

MISSION ANALYSIS

11.7

CHAPTER 12 INTERPLANETARY MISSION ANALYSIS

- 12.1 Historical Background
- 12.2 Patched Conic Trajectories
- 12.3 Porkchop Plots
- 12.4 Gravity Assist Trajectories
- 12.5 B-Plane Targeting
- 12.6 Planetary Capture
- 12.7 Three-body Mission Design
- 12.8 Low-energy Transfers

12.1 Historical Background*

Unlike the previous chapters with hundreds to thousands of years of history, space exploration generally began on October 4, 1957 with the launch of Sputnik. Initially the focus was more on the development of Intercontinental Ballistic Missiles (ICBMs) to support the Cold War. We saw in Chap. 4 how the sensors built in the 1960's still affect many of the space surveillance systems today! However, the focus on ICBMs lost some luster as technology improved, and more importantly, as the race to put a man on the Moon took hold—a symbol of economic prowess and national pride. Thus began a vigorous expansion of space exploration.

The first *winner* in the space race was the former Soviet Union and the successful launch of Yuri Gagarin on April 12, 1961. They had also been first to orbit (Sputnik 1), and the first to send instruments to the Moon (Luna 1, January 4, 1959).

In October 1958, the National Aeronautics and Space Administration (NASA) was formed. It quickly combined other agencies and programs performing related work—Redstone Arsenal, Vanguard, and the Jet Propulsion Laboratory. Their immediate focus became landing a man on the Moon, and bringing him back!

The logical first choice for flight was Earth orbit. Six manned Mercury and ten manned Gemini missions quickly explored this region. Although much remained for investigation, the perceived urgency of the space race guaranteed that additional exploration would follow.

The Moon was the next focus. Several projects began exploration of the Moon: Pioneer, Ranger, Surveyor, Lunar Orbiter, and finally, Apollo. It was the Apollo program that eventually culminated in man walking on the Moon. But not without loss. The tragic fire on January 27, 1967 took astronauts Grissom, White, and Chaffee in a

* This chapter was written by Jeff Parker and combined with sections previously in Chap. 1.

moment. Despite this loss, the program quickly recovered, redesigned, and by the end of 1968, a totally new vehicle blasted off from Earth. The mission took it on to lunar orbit. It was Christmas Eve, 1968 ...

“For all the people of the Earth,” said Anders, ‘the crew of Apollo 8 has a message we would like to send you.’ He paused a moment and then began reading: ‘In the beginning God created the Heaven and the Earth.’

After four verses of Genesis, Lovell took up the reading: ‘And God called the light Day, and the darkness he called Night.’

At the end of the eighth verse Borman picked up the familiar words: ‘And God said, Let the waters under the Heavens be gathered together unto one place, and let the dry land appear; and it was so. And God called the dry land Earth; and the gathering together of the waters He called seas; and God saw that it was good.’

The commander added: ‘And from the crew of Apollo 8, we close with good night, good luck, a Merry Christmas, and God bless all of you - all of you on the good Earth.’ It was a rare emotion.” Cortright [1975, pg 183]

After a docking test in Earth orbit (Apollo 9) and another in lunar orbit (Apollo 10), Apollo 11 lifted off on July 16, 1969. The culmination of the early phases of exploration surely was the landing of man on the Moon on July 20, 1969, 2256 EDT.



Figure 12-1. Apollo 11 Footprint on the Moon. Perhaps one of the best recognized photographs from the Apollo program. In actuality, the footprint in this photo was made and then taken by Buzz Aldrin to examine the soil mechanics. (NASA photo ID AS11-40-5877 from <http://www.hq.nasa.gov/alsj/a11/a11.html>).

Nine Apollo missions would eventually go to the Moon, with six landing on the surface, in what became almost commonplace. Apollo 13 suffered a catastrophic explosion on the way to the Moon on April 14, 1970, but thanks in large part to the redesign after the Apollo fire, the condensed humidity in the command module resisted sparks and performed just long enough to permit a safe, but cold, return on April 17, 1970.

Interplanetary travel was interwoven in the mix as well. Requiring additional escape velocity, the craft were small, but built well. The Voyager spacecraft are still operating in 2013 using their **radioisotope thermal generators** (RTG), or nuclear power. Several other programs took place.

Mars was explored by the Soviet Mars 3 spacecraft which had a successful landing in December 2, 1971, but only provided 20 sec of data. The US Mariner program explored Mercury, Venus, and Mars. Viking 1 later explored Mars and landed successfully on July 20, 1976.

Pioneer 10 and 11 launched in April 1973, Voyager 1 and 2 launched in September and August 1977 (respectively), and they explored Jupiter, Saturn, Uranus, and Neptune. Each had multiple RTGs to power them as their trajectories were away from the Sun. The Pioneer spacecraft lasted until about 1996 and the Voyager spacecraft are projected to last until about 2020.

A detailed discussion of all the missions would take considerable space. A concise graphical representations may be found at (as of Dec 2012):

<http://books.nationalgeographic.com/map/map-day/index>

For this chapter, we adopt a slight variation in nomenclature. Vehicles launched into orbit are called spacecraft to distinguish them from satellites, herein used only for moons of larger objects.

With enough energy, a spacecraft's orbit may be boosted high enough such that it departs the influence of the Earth and enters the region of space dominated by the Sun's gravity. At this moment, we call it an interplanetary mission. Interplanetary missions all depart the Earth and, with the exception of lunar missions, enter heliocentric orbits. Some encounter other planets, others encounter asteroids, comets, and others remain in a heliocentric orbit. This section provides an introduction to some basic, yet important concepts in interplanetary mission design. Numerous interplanetary missions have been launched to date, and it's reasonable to expect that the rate of exploration will increase in the future.

The Earth provides a huge boost to spacecraft destined for other planets in the solar system because the Earth travels nearly 30 km/s about the Sun. Thus, a spacecraft that barely escapes the gravity of the Earth arrives in a heliocentric orbit with a velocity of almost 30 km/s. A spacecraft must obtain a velocity of at least 32.7 km/s at the Earth's orbit to reach Mars. Table 12-1 summarizes the approximate change in velocity needed at the Earth's orbit to reach other destinations in the solar system, without using any further maneuvers or gravity assist trajectories. Of course, higher relative velocities will reduce the transfer durations.

The New Horizons spacecraft was launched in 2006 on an Atlas V 551 launch vehicle on a mission to fly by Pluto. At the time, this was one of the largest boosters available. The New Horizons spacecraft was small—only about 478 kg at launch, including onboard fuel, compared to the fully fueled Atlas V with a ground mass of about 587,000 kg. This mission pushed the limits of conventional rockets.

Fortunately, interplanetary spacecraft can use gravity assist techniques to reach very high speeds with less fuel. Gravity assists will be described in detail later, but they have enabled missions such as Pioneer, Voyager, Galileo, Cassini, and many others.

TABLE 12-1. Interplanetary Velocities. This shows the change in velocity and transfer duration of a Hohmann transfer between Earth's orbit and other destinations in the Solar System, assuming circular coplanar orbits.

Destination	Change in Velocity (km/s)	Relative to Earth velocity (km/s)	Transfer duration (years)
Mercury	22.252	-7.533	0.289
Venus	27.289	-2.495	0.400
Earth	29.785		
Mars	32.729	2.945	0.709
Ceres	36.099	6.315	1.292
Jupiter	38.577	8.793	2.731
Saturn	40.077	10.292	6.062
Uranus	41.067	11.282	16.071
Neptune	41.439	11.655	30.675
Pluto	41.599	11.815	45.639

This chapter provides an overview of the process of building an interplanetary spacecraft trajectory. We describe the patched conic process that is used to build direct transfers to any orbit in the solar system. Next, we show how to extend the patched conic process to include gravity assists. Methods to insert the spacecraft into a desirable orbit at a planet or moon are then found. In addition, we extend the discussions to three-body trajectories, halo orbits, and low-energy transfers. We discuss the underlying equations of motions, trajectories, and practical applications of interplanetary missions.

12.2 Patched Conic Trajectories

The *patched conic* method is a quick way to estimate ballistic interplanetary trajectories—trajectories that don't require any propulsion between planetary encounters. Figure 12-2 shows an illustration of a simple transfer to Mars using the patched conic method. The transfer consists of three phases. Notice that the central body is different in each phase.

1. Earth Departure, a hyperbola relative to the Earth
2. Interplanetary Transfer, modeled as an ellipse relative to the Sun
3. Mars Arrival, a hyperbola relative to Mars

The transfer uses a series of conic sections that are patched together based on which central body is primarily affecting the motion of the spacecraft. The model is an approximation, of course, because the spacecraft is always under the influence of the gravitational force of every massive body in the solar system during each phase of its transfer. The influence of the Earth's gravity on the spacecraft diminishes rapidly as the spacecraft departs the Earth. Soon after departure, the Sun becomes the dominant force and it remains so until the spacecraft approaches very near Mars. Figure 12-3 illustrates this effect by plotting the gravitational force of the Earth, the Sun, and Mars during the transfer. Table 12-2 explores this further by summarizing the effects of different forces during each phase of the transfer. You can see in each phase, one body's gravity dominates the motion. If the approximation switches bodies appropriately, then it's apparent

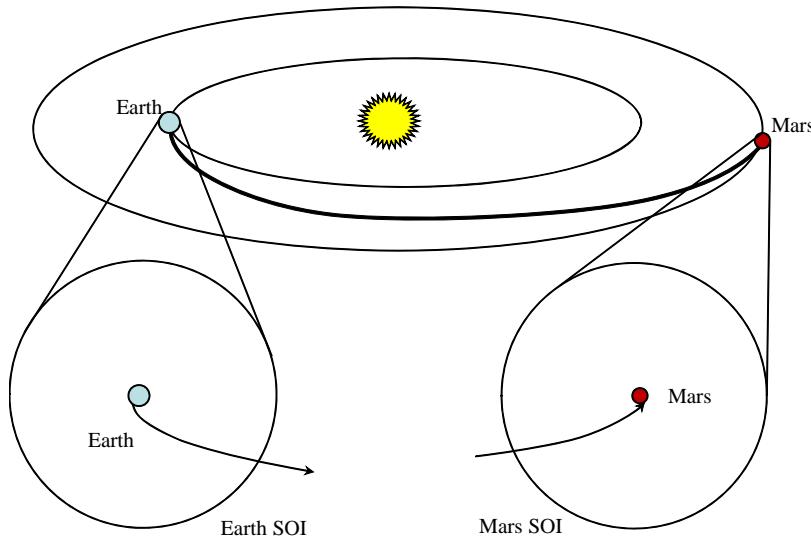


Figure 12-2. Example patched-conic Earth-Mars transfer. The transfer includes a hyperbolic Earth departure, an elliptical heliocentric transfer, and a hyperbolic Mars arrival. The figure is representative but not to scale. The Sphere of Influence, SOI, is the region that depicts the extent of a central body gravitational influence.

that two-body equations of motion are sufficient to approximate each phase of this transfer.

A mission designer using a patched conic interplanetary transfer begins by constructing the interplanetary cruise. The cruise defines the velocity vector the spacecraft must obtain during its Earth departure, as well as the velocity vector that the spacecraft will have upon arriving at Mars. The hyperbolic Earth departure and Mars arrival are then patched to fit that interplanetary cruise.

12.2.1 Sphere of Influence

A central body's *Sphere of Influence* (SOI) (see page 25) is described as the region of space within which the motion of a particle is influenced by that object's gravity more than any other perturbing body. For instance, the Earth's SOI is found by identifying the region of space where the Earth's gravity is more dominant than the Sun's gravity (or any other planet).

Eq. (12-1) describes the acceleration of a spacecraft in the presence of the Earth and the Sun, Eq. (8-34).

$$\ddot{\vec{r}}_{\oplus sat} = -\frac{G(m_{\oplus} + m_{sat})\dot{\vec{r}}_{\oplus sat}}{r_{\oplus sat}^3} + Gm_{\odot} \left(\frac{\dot{\vec{r}}_{sat\odot}}{r_{sat\odot}^3} - \frac{\dot{\vec{r}}_{\oplus\odot}}{r_{\oplus\odot}^3} \right) \quad (12-1)$$

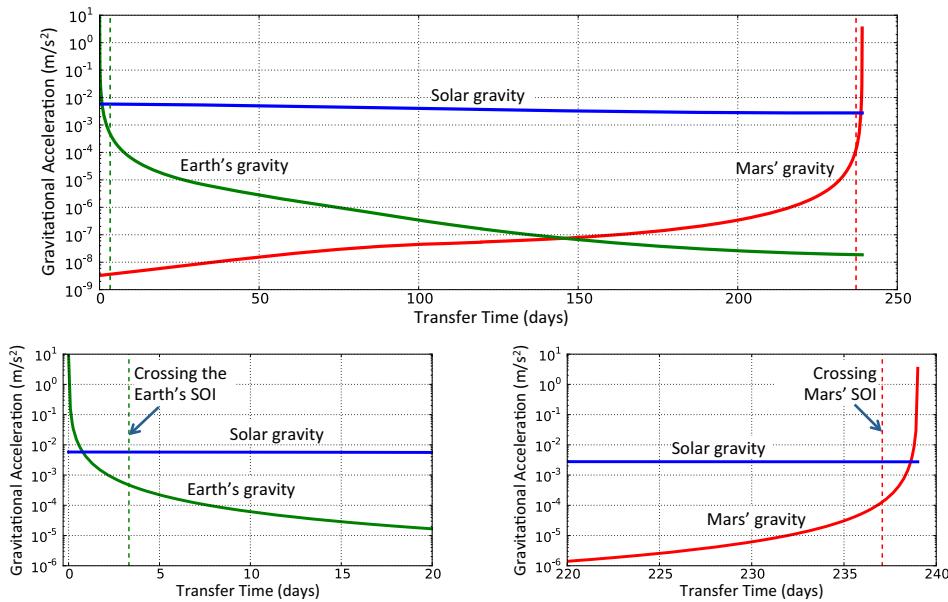


Figure 12-3. Example Accelerations. The gravitational acceleration experienced by a spacecraft due to the Sun, Earth, and Mars on a typical Earth-Mars transfer

TABLE 12-2. Perturbation Acceleration Magnitudes. The magnitude of the dominant forces acting upon a spacecraft during a hypothetical Mars transfer. The spacecraft is modeled as a 500 kg sphere with a radius of 3 meters.

Force	Earth Departure 185 km Altitude		Mid point of Heliocentric Transfer		Mars Arrival 150 km Altitude	
	Magnitude N/kg	Relative to Earth Gravity	Magnitude N/kg	Relative to Earth Gravity	Magnitude N/kg	Relative to Earth Gravity
Earth gravity	9.25	1.0	1.66×10^{-6}	4.58×10^{-5}	1.87×10^{-8}	5.48×10^{-9}
Solar gravity	5.79×10^{-3}	6.26×10^{-4}	3.62×10^{-3}	1.0	2.75×10^{-3}	8.06×10^{-4}
Mars gravity	3.29×10^{-9}	3.55×10^{-10}	5.34×10^{-8}	1.47×10^{-5}	3.41	1.0
Spherical Harmonics	1.43×10^{-2}	1.55×10^{-3}			9.56×10^{-3}	2.81×10^{-3}
Atmospheric Drag	3.42×10^{-3}	3.69×10^{-4}			2.77×10^{-6}	8.11×10^{-7}
Jupiter gravity	9.75×10^{-8}	1.05×10^{-8}	4.84×10^{-8}	1.34×10^{-5}	7.48×10^{-8}	2.20×10^{-8}
Saturn gravity	3.24×10^{-9}	3.50×10^{-10}	2.72×10^{-7}	7.51×10^{-7}	3.21×10^{-9}	9.43×10^{-10}
Venus gravity	2.35×10^{-8}	2.54×10^{-9}	6.54×10^{-9}	1.80×10^{-5}	2.54×10^{-8}	7.46×10^{-9}
Solar Radiation Pressure	2.52×10^{-7}	2.72×10^{-8}	1.57×10^{-8}	4.35×10^{-5}	1.19×10^{-7}	3.50×10^{-8}
General Relativity	6.93×10^{-7}	7.48×10^{-8}	9.05×10^{-11}	2.50×10^{-8}	1.56×10^{-7}	4.57×10^{-8}

Recall the notation where \mathbf{r}_{AB} is the radius vector from body A to body B. The first term in Eq. (12-1) is the dominant two-body equation of motion for a spacecraft in orbit about the Earth and the second term is the third-body gravitational perturbation due to the Sun. Laplace constructed a derivation for the sphere of influence, and we trace it out here. He reasoned that the boundary of the SOI should be the surface where the perturbation of the Earth in a Sun centered analysis would equal the perturbation of the Sun in an Earth centered analysis. Laplace's approach finds the radius where the ratios of the perturbation acceleration to the primary acceleration are equal between the Earth and Sun centered accelerations. Writing Eq. (12-1) for a Sun centered analysis, we find

$$\ddot{\mathbf{r}}_{\oplus sat} = -\frac{G(m_\odot + m_{sat})\dot{\mathbf{r}}_{\odot sat}}{r_{\odot sat}^3} + Gm_\oplus \left(\frac{\dot{\mathbf{r}}_{sat\oplus}}{r_{sat\oplus}^3} - \frac{\dot{\mathbf{r}}_{\odot\oplus}}{r_{\odot\oplus}^3} \right)$$

We can simplify this by ignoring the mass of the satellite and by assuming that $\dot{\mathbf{r}}_{\odot\oplus}$ is equal to the average semimajor axis of the Earth's orbit (a_\oplus). Further, we'll set $\dot{\mathbf{r}}_\odot = \dot{\mathbf{r}}_{sat\odot}$ and the radius of the SOI equal to the distance between the Earth and the satellite ($\dot{\mathbf{r}}_{SOI} = \dot{\mathbf{r}}_{\oplus sat}$). Distributing the negative sign from the primary term to the numerator and keeping the vector directions as above, the ratios of $a_{third\text{-body}} / a_{primary}$ for the Earth and Sun, respectively, are now

$$\frac{m_\odot \left| \frac{\dot{\mathbf{a}}_\oplus}{a_\oplus^3} - \frac{\dot{\mathbf{r}}_\odot}{r_\odot^3} \right|}{m_\oplus \left| \frac{\dot{\mathbf{r}}_{SOI}}{r_{SOI}^3} \right|} \quad \frac{m_\oplus \left| \frac{\dot{\mathbf{a}}_\oplus}{a_\oplus^3} - \frac{\dot{\mathbf{r}}_{SOI}}{r_{SOI}^3} \right|}{m_\odot \left| \frac{\dot{\mathbf{r}}_\odot}{r_\odot^3} \right|}$$

Notice that $\dot{\mathbf{r}}_{SOI} = \dot{\mathbf{a}}_\oplus - \dot{\mathbf{r}}_\odot$. If we assume that $r_{SOI} \ll a_\oplus$, $r_{SOI} \ll r_\odot$, and $r_\odot \approx a_\oplus$, then we can simplify the equations. The first step requires that we evaluate and simplify the vector magnitudes.

$$\begin{aligned} \frac{m_\odot \left| \frac{\dot{\mathbf{a}}_\oplus - \dot{\mathbf{r}}_\odot}{a_\oplus^3} \right|}{\frac{m_\oplus}{r_{SOI}^2}} &= \frac{m_\odot \frac{r_{SOI}}{a_\oplus^3}}{\frac{m_\oplus}{r_{SOI}^2}} = \frac{m_\odot r_{SOI}^3}{m_\oplus a_\oplus^3} \\ \frac{m_\oplus \left| \frac{-a_\oplus^3 \dot{\mathbf{r}}_{SOI} + r_{SOI}^3 \dot{\mathbf{a}}_\oplus}{a_\oplus^3 r_{SOI}^3} \right|}{\frac{m_\odot}{r_\odot^2}} &= \frac{m_\oplus \frac{1}{r_{SOI}^2}}{\frac{m_\odot}{r_\odot^2}} \underset{m_\oplus a_\oplus^2}{\asymp} \frac{m_\oplus a_\oplus^2}{m_\odot r_{SOI}^2} \end{aligned}$$

Setting the two ratios equal,

$$\frac{m_{\odot} r_{SOI}^3}{m_{\oplus} a_{\oplus}^3} = \frac{m_{\oplus} a_{\oplus}^2}{m_{\odot} r_{SOI}^2}$$

The final relation is (the Earth is shown, but the relation holds for any planet)

$$r_{SOI} = \left(\frac{m_{\oplus}}{m_{\odot}} \right)^{2/5} a_{\oplus} \quad (12-2)$$

where a_{\oplus} is equal to the semimajor axis of the Earth's orbit about the Sun. For the Earth, the SOI is about 924,645 km and for the Moon, it's about 66,183 km.

The moons of a planet are typically well inside of the planet's SOI. They in turn have their own SOIs, where the motion of a particle is governed more by the Moon's gravity than by the Sun's or by the planet's gravity. A patched conic analysis of a trajectory from the Earth to Europa would model the spacecraft entering the SOI of Jupiter and then while inside of Jupiter's SOI, it would enter Europa's SOI. The motion of the spacecraft within Europa's SOI would be modeled using conic sections, ignoring the effect of Jupiter as well as the Sun. Accurate bookkeeping is essential!

12.2.2 Interplanetary Transfers

The interplanetary trajectory for a patched conic approximation from Earth to Mars is found by first building an ellipse between the Earth's position in space at the departure time to Mars' position in space at the arrival time, with the Sun as the only gravitational force. This is a good first approximation because the motion takes place while the spacecraft is outside any planetary SOI (See Fig. 12-3 which shows when a spacecraft passes through Earth's and Mars' SOI).

Constructing a transfer from one location to another could be approximated with a Hohmann transfer (Sec. 6.3.1), but it's better formulated as Lambert's Problem (Sec. 7.6). This is because you use 3-dimensional vectors (thereby considering the various inclination and orbital parameters involved during each phase), and the departure and arrival are not constrained to be 180° apart. The solution to Lambert's Problem works just as well for heliocentric transfers as it does for geocentric transfers. You use the position of the Earth and Mars relative to the Sun and the transfer duration as inputs, and you obtain the minimum-energy transfer orbit which defines the velocity of the spacecraft on that transfer at the Earth and Mars.

Using a departure time (t_d), an arrival time (t_a), and a transfer duration (Δt), we write Lambert's problem as follows.

$$LAMBERT(\vec{r}_{sat}(t_d), \vec{r}_{sat}(t_a), \Delta t, t_m, \Rightarrow \vec{v}_{sat}(t_d), \vec{v}_{sat}(t_a))_{\odot} \quad (12-3)$$

Note that we assume the satellite's position at the departure and arrival times relative to the Sun are the same as the departure planet's and arrival planet's positions relative to the Sun at those times. This is reasonable given that the planetary SOI radii are typically much smaller than the distance traveled during the transfer. Be sure to use the gravitational parameter of the Sun instead of the Earth for any heliocentric transfers. There is

also a short way and long way solution (t_m) just as in Chap. 7. For interplanetary transfers, these directions are commonly referred to as Type I ($\Delta\nu < 180^\circ$) and Type II ($180^\circ > \Delta\nu < 360^\circ$) transfers respectively. Multi-revolution transfers continue so Type III transfers are $360^\circ > \Delta\nu < 540^\circ$.

Once the Lambert solution is found, we need to find the velocity of the spacecraft relative to the Earth at the SOI. This is found by subtracting the velocity of Earth relative to the Sun, from the departure velocity vector from Lambert's solution.

$$\hat{\vec{v}}_{\infty\oplus} = \hat{\vec{v}}_{sat}(t_d) - \hat{\vec{v}}_{\oplus}(t_d) \quad (12-4)$$

This velocity vector is known as the spacecraft's hyperbolic excess velocity (page 26), or v -infinity ($v_{\infty\oplus}$). It is the velocity that the spacecraft has beyond the velocity it would have if it just barely escaped Earth's gravity. In other words, it's the velocity a spacecraft would have relative to the Earth at an infinite distance from Earth under the influence of no other forces.

Upon arrival at Mars, we find the velocity relative to Mars at the SOI.

$$\hat{\vec{v}}_{\infty\mathcal{M}} = \hat{\vec{v}}_{sat}(t_a) - \hat{\vec{v}}_{\mathcal{M}}(t_a) \quad (12-5)$$

12.2.3 Departure from Earth

Within the Earth's SOI, the departure is modeled as a hyperbola, where the hyperbola is designed to give the spacecraft a velocity relative to the Earth of $\hat{\vec{v}}_{\infty\oplus}$ at infinite distance. Of course, the spacecraft never reaches an infinite distance from the Earth, but as the spacecraft transitions to a heliocentric orbit at the Earth's SOI, its velocity approaches $v_{\infty\oplus}$.

We find the $v_{\infty\oplus}$ starting at infinite distance with the specific mechanical energy of the hyperbola (note that “specific energy” is in units of Joules per kilogram and it's a measurement of the energy per mass of a spacecraft on a trajectory). Using Eq. (1-20) at an infinite radius,

$$\xi = \frac{v^2}{2} - \frac{\mu}{r} = \frac{v_{\infty\oplus}^2}{2} - \frac{\mu_{\oplus}}{\infty} = \frac{v_{\infty\oplus}^2}{2} \quad (12-6)$$

Given the energy of the hyperbola, you can determine the velocity a spacecraft must have at perigee of the hyperbola. Using Eq. (1-30) and Eq. (12-6),

$$v_p = \sqrt{2\left(\frac{\mu_{\oplus}}{r} + \xi\right)} = \sqrt{\frac{2\mu_{\oplus}}{r_{p\oplus}} + v_{\infty\oplus}^2} \quad (12-7)$$

If the launch vehicle uses a low circular parking orbit prior to the interplanetary injection, we use the velocity of a vehicle in a circular orbit, Eq. (1-33).

$$v_{circle} = \sqrt{\frac{\mu_{\oplus}}{r_{p\oplus}}}$$

where $r_{p\oplus}$ is the radius of the parking orbit. A vehicle in a circular orbit at an altitude of 185 km has a velocity of approximately 7.793 km/s.

The impulsive change in velocity needed to inject the spacecraft onto the interplanetary transfer is equal to the difference of the hyperbolic perigee velocity (v_p) and the [circular] initial transfer orbit.

$$\Delta v(t_{inj}) = \sqrt{\frac{2\mu_\oplus}{r_{p\oplus}} + v_\infty^2} - \sqrt{\frac{\mu_\oplus}{r_{p\oplus}}} \quad (12-8)$$

The Δv (t_{inj}) required for a departure with a $v_\infty\oplus$ of 0.0 km/s is 3.23 km/s, so virtually all interplanetary missions must gain at least that amount of velocity from their parking orbits, and some must gain 4.5 km/s or more in velocity. This places significant limits on the spacecraft design.

If the value of μ is very small (missions that land or depart from an asteroid), then $v_p \sim v_\infty$. The value of v_∞ is typically on the order of 2 – 6 km/s for interplanetary missions departing Earth and the value of $\sqrt{2\mu_\oplus/r_{p\oplus}}$ is approximately 11.0 km/s for standard Earth departures that have a perigee altitude of 185 km. The interplanetary departure perigee velocity v_p is thus between 11 and 14 km/s for most missions.

It's worth noticing that any increase in a spacecraft's perigee velocity raises the spacecraft's $v_\infty\oplus$ by a larger amount, and the effect is greater for more massive bodies and when the spacecraft's distance to the planet is smaller. Thus, it's better to use a lower altitude parking orbit when launching an interplanetary mission. We saw this in Chap. 6 (page 336) with the placement of energy-changing maneuvers in places where the velocity is greatest—known as the ***Oberth effect***. We can quickly prove that this is the case by starting with the energy equation (Eq. (1-20)) and taking the partial derivative of the energy with respect to v .

$$\xi = \frac{v^2}{2} - \frac{\mu}{r} \Rightarrow \frac{\partial \xi}{\partial v} = v \Rightarrow \partial \xi = v \partial v$$

Energy changes are larger when the initial velocity is larger.

A launch vehicle for an interplanetary mission normally targets a parameter known as c_3 which is proportional to the target energy, $c_3 = 2\xi$. This is equivalent to $c_3 = v_\infty^2$ or $c_3 = v^2 - 2\mu/r$. Figure 12-4 illustrates the performance of several example launch vehicles as functions of c_3 . These performance curves are part of the NASA Launch Services Program NLS-2 contract (<http://elvperf.ksc.nasa.gov/>).

In some cases the launch vehicle may inject the spacecraft directly onto the interplanetary transfer. In other cases, for instance when the launch site is on the opposite side of the Earth, the launch vehicle first targets a low, near-circular parking orbit, coasts for a fraction of an orbit (or many orbits), and then performs an additional burn to inject the spacecraft onto the interplanetary transfer.

The direction that a spacecraft travels to achieve a particular interplanetary transfer is determined using the components of the v_∞ vector. Typical launch vehicle targets include c_3 and two angular measurements: the right ascension and declination (generally

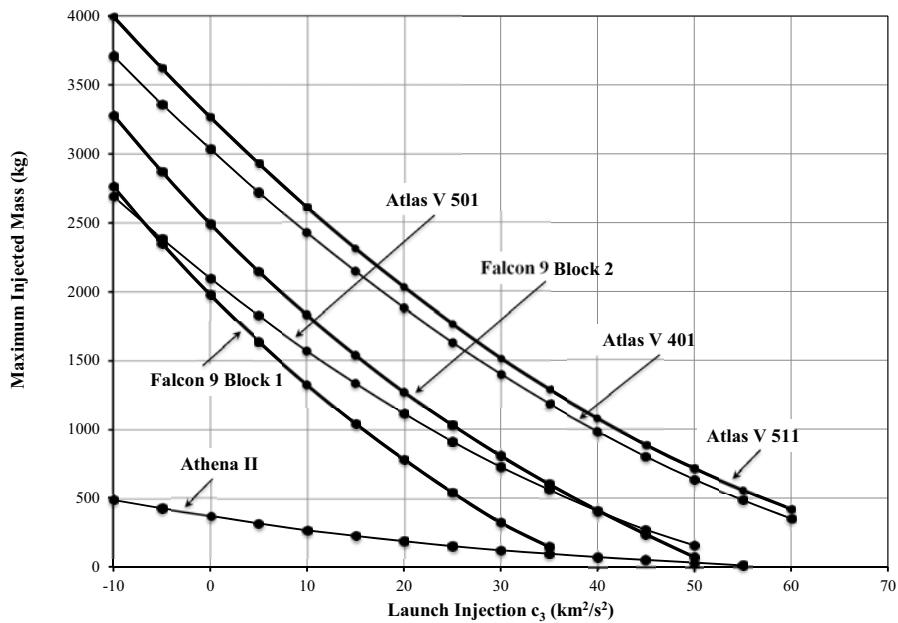


Figure 12-4. Payload Mass Performance. Several example curves that illustrate the performance of launch vehicles as functions of the target parameter c_3 .

in IAU-76/FK5—also called EME2000) of the launch asymptote, RLA and DLA respectively. These are found using Eq. (12-9).

$$\begin{aligned} RLA &= \tan^{-1}\left(\frac{v_{\infty y}}{v_{\infty x}}\right) \\ DLA &= \sin^{-1}\left(\frac{v_{\infty z}}{v_{\infty}}\right) \end{aligned} \quad (12-9)$$

The spacecraft must achieve an outgoing trajectory with a particular asymptote and energy, but there are many hyperbolae that share those characteristics, as shown in Fig. 12-5.

Figure 12-5 shows there is an unconstrained degree of freedom when designing the outgoing hyperbola. One constraint that limits which hyperbolae may be used is the latitude of the launch site. Launch vehicles may target departure hyperbolae with inclinations less than or equal to the launch site latitude without suffering performance penalties. The minimum inclination of the departure hyperbolae is equal to the declination of the launch asymptote, (in EME2000). Thus, if the DLA value for a particular mission is 26° , then the minimum inclination of the set of departure hyperbolae that may be used is 26° , which means that a launch vehicle may launch from Cape Canaveral (latitude of $\sim 28.5^\circ$) without suffering any performance penalty.

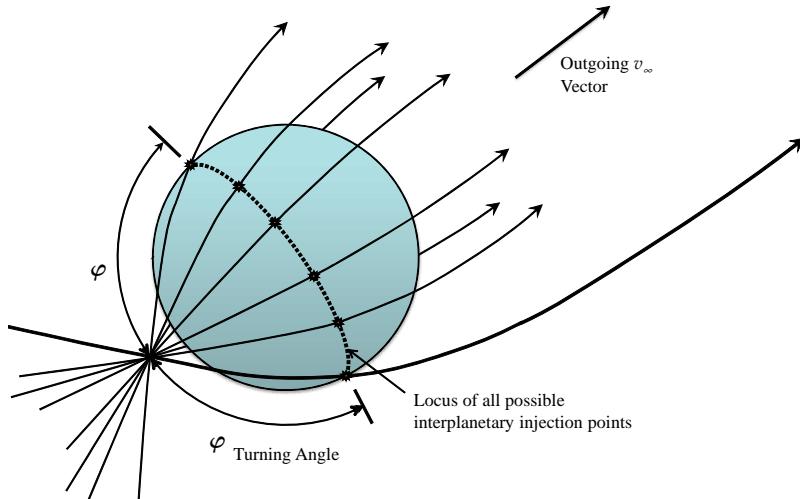


Figure 12-5. Outgoing Trajectory. Example hyperbolae that depart the Earth and achieve a velocity vector parallel to the target departure v_∞ vector.

12.2.4 Arrival

A spacecraft arriving at another planet does so in a similar way to the Earth departure. The interplanetary cruise defines its arrival velocity from the Lambert solution, which in turn specifies the specific energy of the spacecraft relative to the target planet. The spacecraft may target a particular orbit for insertion, or it may target a direct entry into the target body's atmosphere. Both problems have similar geometries, but very different design specifications.

For a given radius (r_p) about the target planet, the velocity of a spacecraft is determined using Eq. (12-7). Note that r_p is now about Mars

$$v(t_{ins}) = \sqrt{\frac{2\mu_\delta}{r_p\delta} + v_\infty^2}$$

If r_p is the planet's surface, this relationship describes how fast a probe might impact the surface of an airless body such as Vesta. Notice there is no change in velocity required.

If the mission calls for an orbit at a particular altitude we find the change in velocity similar to Eq. (12-8). This is the total change in velocity from the hyperbolic entry orbit into a target circular orbit. It's generally accomplished with multiple burns.

$$\Delta v(t_{ins}) = \sqrt{\frac{2\mu_\delta}{r_p\delta} + v_\infty^2} - \sqrt{\frac{\mu_\delta}{r_p\delta}} \quad (12-10)$$

Most modern Martian missions don't use fuel to achieve a circular orbit, but rather use aerobraking. Here the change in velocity is equal to the difference between the

hyperbolic periapsis velocity and the periapsis velocity of the elliptical capture orbit (the second term in Eq. (12-10) would be different).

If the spacecraft targets a direct atmospheric entry, the mission designer needs to characterize the flight path angle of the arrival as a function of the altitude. To do this, we need the arrival parameters. The eccentricity of a hyperbola may be determined using the hyperbola's periapsis range, r_p , and its energy.

$$e = 1 - \frac{r_p}{a} = 1 + \frac{2r_p \xi}{\mu} = 1 + \frac{r_p v_\infty^2}{\mu}$$

The semimajor axis may be similarly solved from these equations.

The position of a spacecraft on a hyperbola using the true anomaly is from Eq. (1-24). The equation to determine the flight path angle, ϕ_{fpa} , of any point along a hyperbola knowing the true anomaly is found using Eq. (2-95) and Eq. (2-96). More information about planetary arrival will be given in Sec. 12.6.

$$\tan(\phi_{fpa}) = \frac{e \sin(\nu)}{1 + e \cos(\nu)}$$

Algorithm 77 shows the combined equations for a departure and arrival between interplanetary transfers. Keeping the proper gravitational constants and other parameters is very important. In the algorithm, “1” represents the departure planet and “2” is the arrival planet. Note that we assume departure and arrival from circular orbits.

ALGORITHM 77: Patched Conic

$$(\dot{\vec{r}}_{sat}(t_d), \dot{\vec{r}}_{sat}(t_a), \dot{\vec{v}}_1(t_d), \dot{\vec{v}}_2(t_a), \Delta t, t_m, r_{p1}, r_{p2} \Rightarrow \Delta v(t_{inj}), \Delta v(t_{ins}))$$

$$LAMBERT(\dot{\vec{r}}_{sat}(t_d), \dot{\vec{r}}_{sat}(t_a), \Delta t, t_m \Rightarrow \dot{\vec{v}}_{sat}(t_d), \dot{\vec{v}}_{sat}(t_a))_{\odot}$$

$$\dot{\vec{v}}_{\infty 1} = \dot{\vec{v}}_{sat}(t_d) - \dot{\vec{v}}_1(t_d)$$

$$v_{p1} = \sqrt{\frac{2\mu_1}{r_{p1}} + v_{\infty 1}^2}$$

$$\Delta v(t_{inj}) = v_{p1} - \sqrt{\frac{\mu_1}{r_{p1}}}$$

$$\dot{\vec{v}}_{\infty 2} = \dot{\vec{v}}_{sat}(t_a) - \dot{\vec{v}}_2(t_a)$$

$$v_{p2} = \sqrt{\frac{2\mu_2}{r_{p2}} + v_{\infty 2}^2}$$

$$\Delta v(t_{ins}) = v_{p2} - \sqrt{\frac{\mu_2}{r_{p2}}}$$

12.2.5 Realistic Interplanetary Transfers

We have examined a patched conic method to build an interplanetary transfer from Earth to Mars as shown in Fig. 12-2, and discussed in Sec. 12.2. Here we briefly look at how accurate this approximation is.

Spacecraft are perturbed by the gravity of all planets and the Sun at all times, even though these perturbations may be very small. We've generated a more realistic trajectory using the patched conic approximation as an initial guess and then performing a numerical integration including the gravitational force of the Sun, all planets, and the Moon at all times. This trajectory was designed to depart from a point 185 km above the Earth at an inclination of 28.5° , to arrive at a point 125 km above Mars, and to have precisely the same transfer duration as the patched conic approximation. The geometry of the transfer was permitted to change. Figure 12-6 presents a plot of the difference in position of the realistic transfer compared with the patched conic transfer. A difference of 200,000 km *could* be viewed as significant in some situations, but it is usually not.

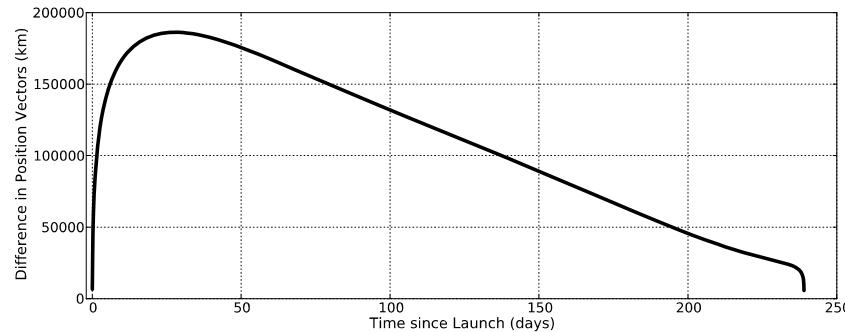


Figure 12-6. Realistic Transfer to Mars. This figure shows the difference in position of a realistic transfer compared with a patched conic transfer to Mars. The transfer durations are the same, but the positions vary.

12.3 Porkchop Plots

We saw in Sec. 12.2.2 that the solution to Lambert's Problem could be used to compute the minimum-energy solution between Earth and Mars for a particular departure and arrival date. This solution dictates a particular departure c_3 and arrival v_∞ . If you change the departure date and/or the arrival date, the result is a new trajectory with new values for the departure c_3 and arrival v_∞ . Plotting the departure date vs the arrival date, several transfer performance characteristics can be shown. The first such maps produced explored transfers from Earth to Mars (Fig. 12-7). These maps were quickly known as porkchop plots, given their resemblance to a porkchop.

Figure 12-8 gives some insight into the trajectory characteristics of the porkchop plot by plotting several Type I and Type II transfers, where the Mars arrival date is held fixed from one transfer to the next.

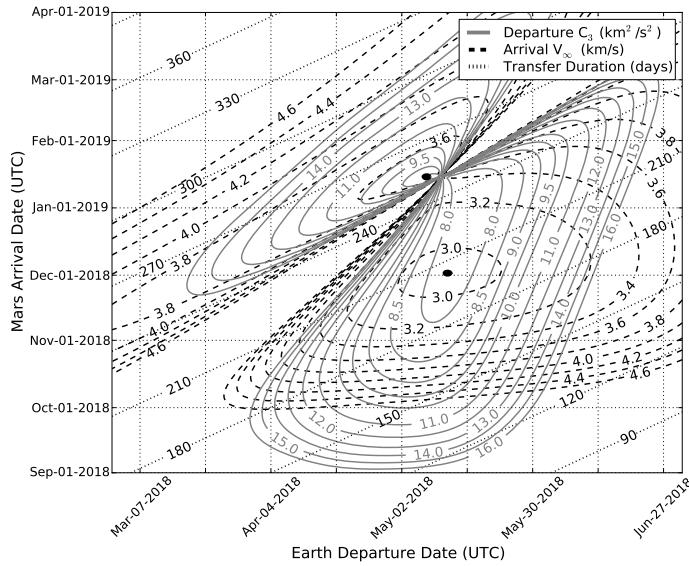


Figure 12-7. Basic Porkchop Plot Earth-Mars 2018: An example porkchop plot, summarizing the design space for transfers between Earth and Mars during the 2018 opportunity. The straight lines depict the transfer duration in days. The solid curved contours are the departure c_3 , and the dashed linear contours are the arrival v_∞ . The minimum arrival v_∞ long way (Type II) and short way (Type I) positions are highlighted (black ovals). The minimum c_3 location is close, up and to the right.

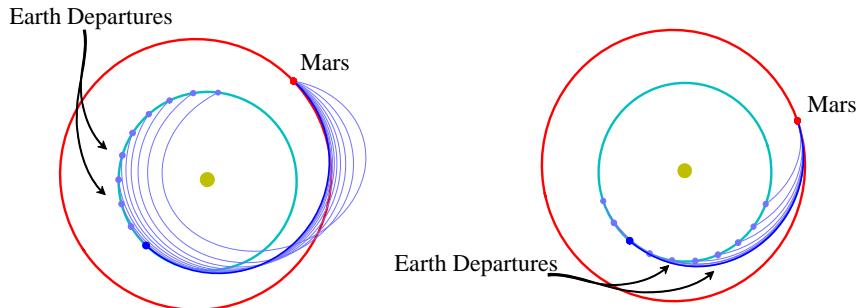


Figure 12-8. Type I and Type II Transfers: Example Type II transfers (left) and Type I transfers (right) for the Earth-Mars 2018 opportunity are shown (long way and short way in Chap. 7). The transfer with the smallest departure c_3 is emphasized in each plot. Note that the arrival date is held constant.

Many aspects of an interplanetary transfer may be plotted on a porkchop plot, including departure c_3 , arrival v_∞ , departure DLA, transfer duration, maximum distance to the Sun, etc. The porkchop plot quickly identifies favorable transfers for a particular mis-

sion. Figure 12-7 illustrates a porkchop plot for the 2018 opportunity to Mars. There are two local minima for the c_3 contours, as well as two local minima for the arrival v_∞ contours. The solutions in the lower-right correspond to Type I (short way) transfers. The solutions in the upper-left correspond to Type II (long way) transfers. The best Type I transfers require a departure c_3 of $7.7 \text{ km}^2/\text{s}^2$, whereas the best Type II transfers require c_3 values only as low as $8.1 \text{ km}^2/\text{s}^2$. Figure 12-9 illustrates other parameters that may be plotted in the chart.

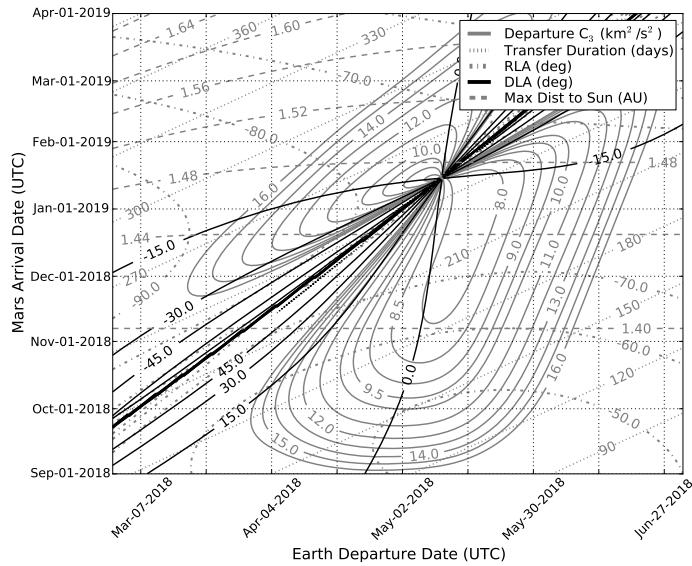


Figure 12-9. Additional Parameters on a Porkchop Plot, Earth-Mars 2018: Alternate parameters may be included in a porkchop plot. Here, the RLA, DLA, and maximum distance to the Sun are added as additional lines from Fig. 12-7.

There can be significant variability in the features of a porkchop plot from one mission opportunity to the next. For instance, Fig. 12-10 shows the 2020 Earth-Mars opportunities while Figs. 12-7 and 12-9 show the opportunities for 2018. In 2018, a mission could launch from Earth with a c_3 about $7.7 \text{ km}^2/\text{s}^2$ to reach Mars, implementing a Type I transfer, and could encounter Mars with a v_∞ near 3 km/s . In 2020, all missions departing Earth on a direct Earth-Mars transfer require c_3 values greater than $13 \text{ km}^2/\text{s}^2$ —and such transfers arrive at Mars with far higher velocities! The locations in the porkchop plot where the minimum c_3 and minimum arrival velocity occur do not line up. This places more demands on the interplanetary mission planners.

The porkchop plot included in Fig. 12-10 illustrates a Type I design space and a Type II design space, separated by a steep ridge of undesirable trajectories. That ridge is ubiquitous in porkchop plots between planets in the solar system and results because the planets are not perfectly coplanar. If they were, then all minimum-energy transfers would also be coplanar and the least-energy transfers would be 180 deg Hohmann trans-

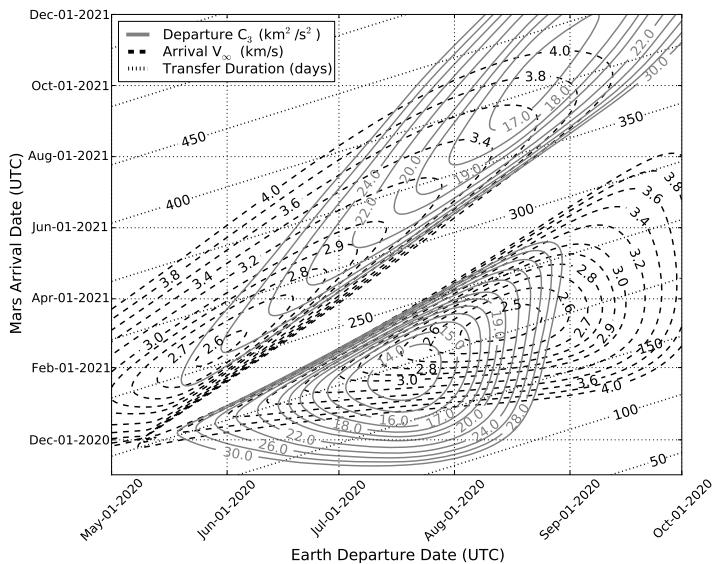


Figure 12-10. Earth-Mars Transfer 2020: When the 2020 Earth-Mars opportunities are compared to those from 2018, you can see some large differences in the available transfers. (See Figure 12-7)

fers. However, even the minimum-energy Type I and Type II solutions are not coplanar transfers, but exist in a plane that is defined by the three-dimensional positions of the Sun, Earth, and Mars. The orbital planes of optimal Type I and Type II transfers are very close to the planes of Earth and Mars. But as the transfers approach 180 deg, the non-coplanar nature of the planets forces the transfer plane to go far out of the ecliptic. One way to improve such missions is to insert a large deep space maneuver into the transfer. This type of transfer is known as a “broken plane transfer” and it permits the spacecraft to depart the Earth near the Earth’s orbital plane, and to arrive at Mars near its plane, even though the maneuver may be large.

The porkchop plot has many uses. One is to identify extended launch periods. For instance, many missions to Mars have implemented 21-day launch periods, to ensure a high probability that the project, weather, and all other requirements are satisfied on at least one day during that time span to launch the mission. Mission designers use porkchop plots to find acceptable launch/arrival dates for each of the 21 launch dates. Engineers build the spacecraft to accommodate the worst-case conditions found in each of those 21 opportunities.

Porkchop plots may also be used in the design of a gravity assist as we’ll see next.

12.4 Gravity Assist Trajectories

The patched conic approximation assumes the spacecraft is departing or arriving at a single body. If the spacecraft passes near a body, with proper geometry, its velocity can

be increased or decreased. We call this a ***gravity assist trajectory*** because the object's gravity is used to alter the spacecraft's velocity.

Early missions to the Moon, Venus, and Mars passed by the target body, flying by within several thousand kilometers of the surface of the body. Some of those included Luna 1 passing near the Moon at an altitude of 5995 km on January 1, 1959. On December 14, 1962, Mariner 2 passed about 28,710 km above Venus.

More recently, Dawn performed a flyby past Mars en route to Vesta, and Cassini flies by Titan frequently during its Saturn tour. Mariner 10, Pioneer 11, Voyager, and Messenger are other examples. In all cases, it's very important to design the flyby such that the spacecraft trajectory lines up with the new interplanetary cruise segment. The hyperbolic passage must be well known to determine how to point communication antennae and/or scientific instruments during the flyby.

The important element of a gravity assist is that the spacecraft's energy with respect to the assisting planet does not change, but the energy relative to the Sun does. The change in the spacecraft's energy comes from the energy of the assisting planet. It changes by an equal and opposite amount due to the flyby as per Newton's Laws.

Figure 12-11 illustrates the geometry of two gravity assists. The flyby on the left is designed to reduce the energy of a spacecraft relative to the planet while the flyby on the right increases the energy relative to the planet. Notice that the v_∞ magnitude remains constant and the planet's velocity remains constant (except for the minuscule change mentioned above). The heliocentric velocity vector changes a significant amount due to the rotation of the v_∞ vector.

There are several conventions for the turning angle, φ [Eq. (2-28)]. JPL (1983:19) and Kaplan (1976:93) depict a turning angle that shows the amount that the spacecraft trajectory turns, with respect to the hyperbolic asymptotes. The supplementary angle φ_s is useful for some applications so both angles are found as follows.

$$\begin{aligned} \cos(\varphi_s) &= \frac{1}{e} = \frac{1}{1 + \frac{v_\infty^2 r_p}{\mu_\oplus}} & \tan(\varphi_s) &= \frac{b}{a} = \sqrt{e^2 - 1} \\ \sin\left(\frac{\varphi}{2}\right) &= \frac{1}{e} & \varphi &= 180^\circ - 2\varphi_s \end{aligned} \quad (12-11)$$

When constructing a gravity assist, a mission designer must ensure that the spacecraft does not hit the surface of the planet, or impact a natural satellite or ring. It's straightforward to determine the radius of closest approach of the flyby. Let's say we have a trajectory that departs the Earth and performs a gravity assist by Jupiter en route to Saturn. In this case, we can build a trajectory between Earth and Jupiter using Lambert's solution. We build a second trajectory between Jupiter and Saturn. The two trajectories should have the same v_∞ magnitude at Jupiter, neglecting any satellite encounters at Jupiter. Given the v_∞ vectors inbound and outbound, $v_\infty(t_{in})$ and $v_\infty(t_{out})$, respectively, we find the turn angle as

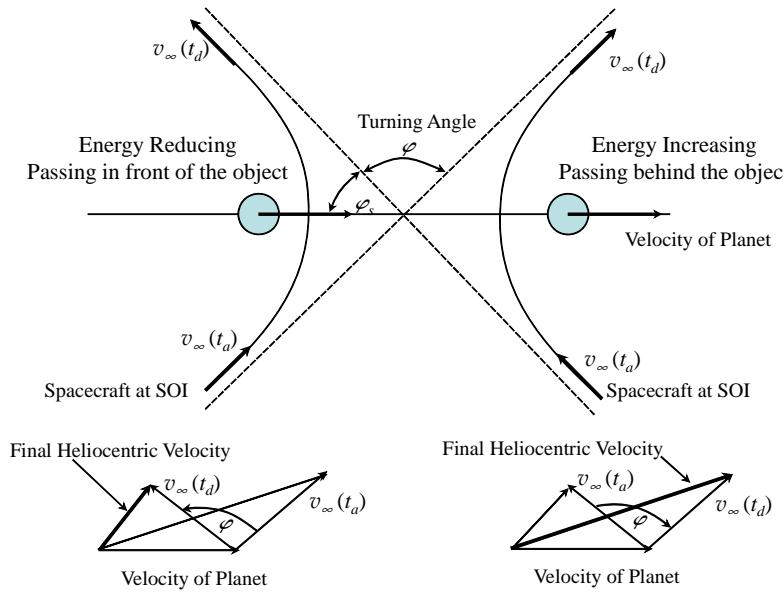


Figure 12-11. Gravity Assist Trajectories. An illustration of two gravity assists: one that increases a spacecraft's energy relative to the Sun (right) and one that decreases it (left). The spacecraft velocities are the same relative to the planet at both the arrival and departure from the SOI of the planet. The turning angle (φ) shows the change in direction after the flyby, relative to the assisting body.

$$\cos(\varphi) = \frac{\hat{v}_\infty(t_{in}) \cdot \hat{v}_\infty(t_{out})}{|\hat{v}_\infty(t_{in})| |\hat{v}_\infty(t_{out})|}$$

The radius of closest approach is then a function of the turn angle, the gravitational parameter of the gravity assist body (Jupiter, in this case), and the spacecraft velocity.

$$r_p = \frac{\mu_J}{|\hat{v}_\infty(t_{in})|^2} \left(\frac{1}{\cos\left(\frac{\pi - \varphi}{2}\right)} - 1 \right) \quad (12-12)$$

We then check to ensure that r_p is not less than the radius of the planet or in any forbidden regions, such as rings.

Consider building a mission that departs Earth, flies by Jupiter, and arrives at Saturn. We know that if we don't perform any maneuvers at Jupiter, or fly by any of its moons, the spacecraft's v_∞ magnitude will not change during the flyby. One way to design a good approximation of such an interplanetary transfer is to construct a porkchop plot of the Earth-Jupiter transfer and a second porkchop plot for the Jupiter-Saturn transfer. We search both maps for conditions where the arrival v_∞ magnitude and departure v_∞ mag-

nitude are equal on the same Jupiter encounter date. If they are equal, then we check to be sure that the spacecraft would not approach too close to the planet (Eq. (12-12)). Finally, we take the set of all good transfers and select the one with the best Earth departure c_3 and Saturn arrival v_∞ , or other parameter.

An example will illustrate the complete process.

▼ Example 12-8. Solving a Gravity Assist.

Consider Voyager 1's Jupiter flyby. This flyby was designed to propel the spacecraft to an encounter with Saturn. Mission designers didn't intend to strictly raise the spacecraft's energy, but to raise its energy such that the resulting trajectory intersected Saturn at a designated time.

GIVEN: The time of Earth at the departure of Voyager was September 8, 1977 09:08:17 UTC, the time of Jupiter flyby was March 5, 1979 12:05:26 UTC, and the time of the Saturn flyby was November 12, 1980 23:46:30 UTC.

FIND: The change in heliocentric velocity vector produced by the Jupiter flyby, the turning angle, and the radius of closest approach to Jupiter.

We can use our ephemerides or analytical approximations to determine the position of the Earth at the Earth departure time and the position and velocity of Jupiter at the Jupiter encounter time, all in the EME2000 (IAU-76/FK5) Sun centered coordinate frame.

$$\hat{r}_{\oplus}(t_{d\oplus}) = 146,169,549 \hat{X} - 36,666,991 \hat{Y} - 644 \hat{Z} \text{ km}$$

$$\hat{v}_{\oplus}(t_{d\oplus}) = 6.774\,093 \hat{X} + 28.786\,741 \hat{Y} + 0.001\,400 \hat{Z} \text{ km/s}$$

$$\hat{r}_{\mathcal{J}}(t_{a\mathcal{J}}) = -482,178,605 \hat{X} + 627,481,695 \hat{Y} + 8,221,250 \hat{Z} \text{ km}$$

$$\hat{v}_{\mathcal{J}}(t_{a\mathcal{J}}) = -10.524\,225 \hat{X} - 7.363\,294 \hat{Y} + 0.266\,050 \hat{Z} \text{ km/s}$$

We can use Lambert (Eq. (12-3) or Sec. 7.6) to generate the minimum-energy trajectory between Earth and Jupiter for those dates ($\Delta t = 1.487$ years). Note that different Lambert approaches may produce slightly different results. The spacecraft's arrival at Jupiter has a heliocentric velocity of approximately:

$$\hat{v}_{vgr}(t_{a\mathcal{J}}) = -12.234\,317 \hat{X} + 3.473\,749 \hat{Y} + 0.006\,621 \hat{Z} \text{ km/s}$$

We then subtract Jupiter's heliocentric velocity to determine the spacecraft's hyperbolic excess velocity, $v_\infty(t_{in})$.

$$\hat{v}_\infty(t_{in}) = \hat{v}_{vgr}(t_{a\mathcal{J}}) - \hat{v}_{\mathcal{J}}(t_{a\mathcal{J}}) = -1.710\,092 \hat{X} + 10.837\,042 \hat{Y} - 0.259\,429 \hat{Z} \text{ km/s}$$

The magnitude of the $v_\infty(t_{in})$ vector is approximately equal to 10.974 km/s. Conservation of energy dictates that the spacecraft must depart Jupiter with the same Jovian-centered velocity that it had at arrival, assuming no satellite flybys or maneuvers during the Jovian encounter. The flyby was designed to have a turning angle of about 98.3° , to target an outgoing trajectory that would encounter Saturn.

The position and velocity of Saturn at the time of the spacecraft's arrival was:

$$\hat{r}_{\mathcal{S}}(t_{a\mathcal{S}}) = -1,421,327,810 \hat{X} - 54,598,217 \hat{Y} + 57,422,412 \hat{Z} \text{ km}$$

$$\hat{v}_{\mathcal{S}}(t_{a\mathcal{S}}) = -0.148\,850 \hat{X} - 9.684\,595 \hat{Y} + 0.175\,045 \hat{Z} \text{ km/s}$$

The Lambert solution is used again to generate the minimum-energy trajectory between Jupiter and Saturn for the dates of that encounter ($\Delta t = 1.693$ years). We find that the spacecraft's heliocentric velocity at the Jupiter departure time is approximately

$$\hat{v}_{vgr}(t_{d\gamma}) = -20.979\,251 \hat{X} - 10.604\,676 \hat{Y} + 1.018\,354 \hat{Z} \text{ km/s}$$

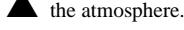
Subtract Jupiter's heliocentric velocity to determine the spacecraft's hyperbolic excess velocity $v_\infty(t_{out})$

$$v_\infty(t_{out}) = \hat{v}_{vgr}(t_{d\gamma}) - \hat{v}_\gamma(t_a) = -10.455\,026 \hat{X} - 3.241\,382 \hat{Y} + 0.752\,304 \hat{Z} \text{ km/s}$$

The magnitude of $v_\infty(t_{out})$ is 10.972 km/s—very close to that of the incoming v_∞ vector. Voyager 1 encountered all four Galilean satellites during its Jupiter flyby. The encounter with Io (within 21,000 km) certainly affected Voyager, but even with these effects the v_∞ vector changed less than 3 m/s in magnitude.

The Jupiter encounter did change Voyager 1's heliocentric velocity and orbital plane. Voyager 1's heliocentric velocity prior to the Jupiter encounter had a magnitude of approximately 12.718 km/s. Its velocity after the encounter had increased to 23.529 km/s! Furthermore, the inclination of the heliocentric orbit changed from 0.96° to 2.48°. This change in energy and orbit plane would certainly have required a sizable amount of fuel if performed with a conventional propulsion system.

Lastly, we can compute how close Voyager I came to Jupiter using Eq. (12-12). Doing so we find that Voyager I came within 792,000 km of Jupiter—certainly far enough out to avoid encountering



the atmosphere.

12.5 B-Plane Targeting

A useful mechanism for targeting a particular hyperbolic passage is the B-Plane. The **B-Plane**, also called the *Body Plane*, is an imaginary plane in space that's orthogonal to the incoming asymptote of the spacecraft's approach trajectory and includes the target body itself. Figure 12-12 shows an illustration of the B-Plane for a spacecraft approaching a body for a flyby.

The advantage to using the B-Plane is apparent when a mission requires a spacecraft to target precise conditions upon arriving at a target body—for a flyby, an orbit insertion, or an entry state to a planetary lander. Let's say the spacecraft is targeting an orbit insertion at Mars, where the mission requires the spacecraft to arrive at a particular time at an altitude of 200 km and at an inclination of 93° with respect to the Martian equator. These parameters vary in highly nonlinear fashions as the spacecraft's trajectory is adjusted with *Trajectory Correction Maneuvers* (TCMs). However, if you convert these arrival parameters relative to the B-Plane, they vary in a linear fashion and it eases the design of TCMs because the gravity of the target planet does not affect them.

The vector \hat{B} points from the center of the target body to the point where the spacecraft would pierce the B-Plane if the target body's gravity did not influence the spacecraft's trajectory, as shown in Fig. 12-12. \hat{S} lies in the trajectory plane in the direction of motion of the satellite. \hat{T} is (typically) in the ecliptic plane, and \hat{R} is normal to \hat{S} and \hat{T} . The linear time of flight (LTOF) is the time interval required for the spacecraft to hit that point, again assuming no gravitational attraction from the target body. The most common B-Plane parameters are the set $[B, \theta_B, \text{LTOF}]$, where θ_B is the angle from \hat{T} to \hat{B} , or alternatively $[B_T, B_R, \text{LTOF}]$ with components of \hat{B} .

There are several formulae that are useful when converting trajectory coordinates to B-Plane parameters. We'll discuss two. The first describes how to convert from a position and velocity vector to B-Plane parameters for a spacecraft approaching a target body. Algorithm 78 shows the equations. Note the use of the angular momentum vector,

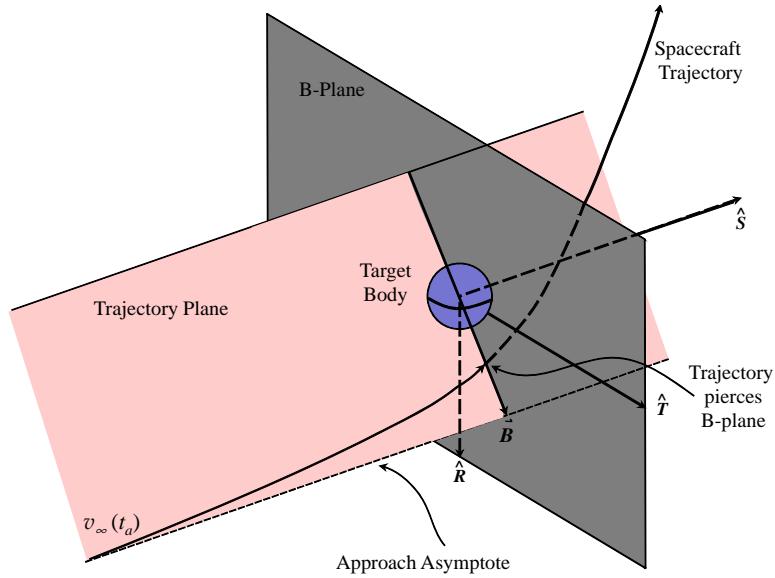


Figure 12-12. B-Plane Targeting. An illustration of the Body Plane, or B-Plane. The B-Plane is orthogonal to the spacecraft's approach asymptote, and intersects the target body center.

and \hat{K} is defined to be in the Z-axis direction of the coordinate system in use (usually ecliptic, but sometimes body-fixed). Also, the position and velocity vectors are relative to the flyby body and not the Sun.

ALGORITHM 78: Algorithm B-Plane I ($\hat{r}, \hat{v}, v_\infty^2 \Rightarrow B_R, B_T$)

$$\hat{h} = \frac{\hat{r} \times \hat{v}}{|\hat{r} \times \hat{v}|}$$

$$\hat{e} = \frac{\left(v^2 - \frac{\mu}{r}\right)\hat{r} - (\hat{r} \cdot \hat{v})\hat{v}}{\mu} \quad a = -\frac{\mu}{v_\infty^2}$$

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

$$\varphi_s = \cos^{-1}\left(\frac{1}{e}\right) \quad \hat{K} = [001]$$

$$\hat{S} = \frac{\hat{e}}{|\hat{e}|} \cos(\varphi_s) + \frac{\hat{h} \times \hat{e}}{|\hat{h} \times \hat{e}|} \sin(\varphi_s) \quad \hat{T} = \frac{\hat{S} \times \hat{K}}{|\hat{S} \times \hat{K}|}$$

$$\begin{aligned}\hat{R} &= \hat{S} \times \hat{T} \\ \vec{\hat{B}} &= b(\hat{S} \times \hat{h}) \\ B_T &= \vec{\hat{B}} \cdot \hat{T} \quad B_R = \vec{\hat{B}} \cdot \hat{R}\end{aligned}$$

The second formulation describes how to convert from v_∞ vectors to B-Plane parameters for a spacecraft approaching a flyby. This is particularly useful for gravity assists, when both $v_\infty(t_{in})$ and $v_\infty(t_{out})$ are defined easily.

ALGORITHM 79: *Algorithm B-Plane II* ($\vec{\hat{v}}_\infty(t_{in}), \vec{\hat{v}}_\infty(t_{out}) \Rightarrow \vec{\hat{B}}, r_p$)

$$\begin{aligned}\hat{S} &= \frac{\vec{\hat{v}}_\infty(t_{in})}{\|\vec{\hat{v}}_\infty(t_{in})\|} \\ \hat{h} &= \frac{\vec{\hat{v}}_\infty(t_{in}) \times \vec{\hat{v}}_\infty(t_{out})}{\|\vec{\hat{v}}_\infty(t_{in}) \times \vec{\hat{v}}_\infty(t_{out})\|} \\ \hat{B} &= \hat{S} \times \hat{h} \quad \hat{K} = [001] \\ \hat{T} &= \hat{S} \times \hat{K} \quad \hat{R} = \hat{S} \times \hat{T}\end{aligned}$$

$$\cos(\varphi) = \frac{\vec{\hat{v}}_\infty(t_{in}) \cdot \vec{\hat{v}}_\infty(t_{out})}{\|\vec{\hat{v}}_\infty(t_{in})\| \|\vec{\hat{v}}_\infty(t_{out})\|}$$

$$r_p = \frac{\mu}{\|\vec{\hat{v}}_\infty(t_{in})\|^2} \left(\left(\frac{1}{\cos\left(\frac{\pi - \varphi}{2}\right)} \right) - 1 \right)$$

$$B = \frac{\mu}{\|\vec{\hat{v}}_\infty(t_{in})\|^2} \sqrt{\left(1 + \frac{\|\vec{\hat{v}}_\infty(t_{in})\|^2 r_p}{\mu} \right)^2 - 1}$$

$$\vec{\hat{B}} = B \cdot \hat{B}$$

When performing B-Plane targeting, we convert the target parameters into B-Plane parameters and then use those to construct TCMs. The basic linear targeting scheme uses the relationships:

$$\begin{bmatrix} \partial B_T \\ \partial B_R \\ \partial TOF \end{bmatrix} = \begin{bmatrix} \frac{\partial B_T}{\partial \Delta v_x} & \frac{\partial B_T}{\partial \Delta v_y} & \frac{\partial B_T}{\partial \Delta v_z} \\ \frac{\partial B_R}{\partial \Delta v_x} & \frac{\partial B_R}{\partial \Delta v_y} & \frac{\partial B_R}{\partial \Delta v_z} \\ \frac{\partial T}{\partial \Delta v_x} & \frac{\partial T}{\partial \Delta v_y} & \frac{\partial T}{\partial \Delta v_z} \end{bmatrix} \begin{bmatrix} \Delta v_x \\ \Delta v_y \\ \Delta v_z \end{bmatrix}$$

Each of the partials is typically constructed numerically, although there are analytical equations for a number of relationships. The B-Plane formulation is useful because these partials have a very wide linear regime.

12.6 Planetary Capture

A spacecraft arriving at a planet typically enters a capture orbit, encounters that planet's atmosphere, or lands on the surface. The spacecraft may also land on a satellite of the planet, such as Titan.

Many considerations are made when designing a spacecraft's approach. If the spacecraft is going to enter the atmosphere, then it's typically desirable to arrive with a relatively low velocity and a shallow flight path angle, although not so low as to skip out of the atmosphere. If the spacecraft is going to directly land on the surface, then the arrival velocity is minimized and a desired flight path angle range is targeted. In both situations, the spacecraft must arrive at the body within specified latitude and longitude values. In some cases (LCROSS), the spacecraft intends to impact the surface from a near-normal direction with a high velocity.

If a mission objective is to place the spacecraft into orbit, then mission designers typically target an initial capture orbit before transitioning the spacecraft into its operational orbit. The capture orbit is usually very large—one that has a low periapsis and a high apoapsis. The spacecraft performs its orbit insertion at periapsis because the higher velocities yield the most efficient transfers. An orbit insertion's primary goal is to get the spacecraft safely captured and to then transition to an operational orbit. For Mars and Venus, spacecraft can use aerobraking to reduce the size of their orbits. At the large gas planets, a spacecraft may use flybys of other satellites to change the characteristics of its orbit.

One approach is to apply the Δv when at the proper altitude. From Eq. (12-10) the total Δv needed to enter a circular orbit is:

$$\Delta v = \sqrt{\frac{2\mu}{r_p} + v_\infty^2} - \sqrt{\frac{\mu}{r_p}}$$

When inserting directly into a circular orbit using one or more maneuvers only performed at periapsis, Kaplan (102-103) shows the minimum Δv for capture as

$$r_p = \frac{2\mu}{v_\infty^2} \quad \Delta v_{min} = \frac{v_\infty}{\sqrt{2}}$$

12.6.1 Aerobraking

A spacecraft in orbit about any body with an atmosphere may conduct aerobraking maneuvers. The spacecraft first performs a maneuver to drop its periapsis altitude into the atmosphere. As the spacecraft skims through the atmosphere, it experiences drag, slowing it down slightly. The more the spacecraft descends, the more drag the spacecraft experiences. Eventually, after dozens of orbits, the spacecraft's apoapsis will have dropped significantly. Magellan, in orbit about Venus, and several Mars orbiters have used aerobraking maneuvers to change their orbits in just a few months using far less fuel than would have been needed to do so with chemical propulsion.

Aerobraking maneuvers involve 3 primary phases. The *walk-in phase* consists of the spacecraft gradually dropping its periapsis until it reaches a good ***aerobraking corridor***, defined as the region where the spacecraft will experience the most atmospheric drag while not exceeding its tolerable heat rates. The *main phase* consists of multiple orbits with aerobraking maneuvers to reduce the spacecraft's apoapsis altitude. The *walk-out phase* is performed as the spacecraft approaches its final orbit to slow aerobraking down and finally to exit the procedure.

The *walk-in phase* consists of several orbits, where the spacecraft performs small maneuvers at each apoapsis to reduce the periapsis altitude. The navigators of Magellan and Mars Odyssey, for example, did not have good knowledge of the density profile of the planet's atmosphere. Thus, they gradually dropped the spacecraft's periapsis altitude, exploring the amount of drag that the spacecraft experienced.

The spacecraft typically rotates into a particular attitude as it approaches an atmospheric pass. The attitude is usually chosen to keep it dynamically stable, such as a badminton shuttlecock, where the atmosphere's force acts as a restoring force. In this way, the spacecraft maintains attitude control, and often can use the atmosphere to its benefit to desaturate its momentum wheels.* The attitude is also typically chosen to expose only certain surfaces to the atmosphere, if possible. These exposed surfaces experience the highest heating rates due to the atmospheric drag. The heating rates determine the maximum possible descent as well. Mars Odyssey flew through the atmosphere with its solar panels trailing, stabilizing the spacecraft. The solar panels could only experience a maximum heating rate that would keep the temperature below 175° Celsius.

The *main phase* of aerobraking consists of numerous orbits when the spacecraft is performing nominal aerobraking passes through the atmosphere. The operational aerobraking corridor, namely, the altitude that the spacecraft could fly to achieve the most drag braking without exceeding a maximum heating rate, is a dynamic corridor—the atmospheric density profile changes due to the level of solar activity, the season, and

* Momentum wheels are typically used to precisely maintain and adjust a spacecraft's attitude. Over time they speed up to absorb angular momentum; as they approach their maximum speed, they must be spun down to permit future operation—a process known as ***desaturation***.

other reasons. Hence, the spacecraft's orbit is continuously monitored to ensure that it remains within that corridor, to track how deep it dips into the atmosphere, and how high the heating rate climbs during each pass. If the periapsis dips too far, or if the atmospheric density rises too much, then the spacecraft performs a small maneuver at apoapsis to raise periapsis. If the atmospheric density has dropped due to solar activity, then the spacecraft performs a small maneuver to reduce its periapsis altitude to stay within the corridor. Historical missions have required small maneuvers, on the order of 0.1 m/s, about every 10 orbits to maintain the corridor.

Because our knowledge of planetary atmospheres is often quite limited, aerobraking maneuvers are prone to contingency situations. In such circumstances, designers have escape maneuvers to abort the operation. Given the uncertainty of the atmospheres and the solar activity, this is an important step. The usual action is to perform a maneuver at the next apoapsis to raise periapsis out of the atmosphere. The walk-in and main phases can be resumed at a later time.

The *walk-out phase* of aerobraking is performed when the spacecraft's apoapsis altitude has fallen to a desirable altitude. If a precise altitude is desired, the spacecraft may raise its periapsis altitude in a series of burns, experiencing less drag in each subsequent aerobraking pass until the desired orbit is attained. The final maneuver lifts the spacecraft's periapsis out of the atmosphere and ends the aerobraking maneuvers, much like the escape maneuver described earlier.

Spacecraft operators closely monitor a spacecraft during aerobraking operations to predict the timing of the next periapsis passage. Typically, the spacecraft's first few orbits are so large that navigators can determine its orbit each orbit, predicting the periapsis timing to within a few seconds. Later on though, operators do not usually get enough tracking data to determine the spacecraft's orbit each orbit, but the periapsis timing does not change as rapidly.

Magellan was the first spacecraft mission to use aerobraking to purposefully reduce the spacecraft's orbit from one operational orbit to another operational orbit. Magellan's aerobraking campaign budgeted a total of about 80 m/s of Δv to reduce its orbit from a 171×8470 km to a 197×541 km orbit using 730 passes through the Venusian atmosphere, over a course of 70 days (Giorgini et al. 1995). This total, which includes contingencies that were not needed, was far lower than the 1226 m/s required to perform the orbit transfer using chemical means.

The data for Magellan's Venus aerobraking as well as the aerobraking maneuvers for Mars Odyssey and for Mars Reconnaissance Orbiter are summarized in Table 12-3. This summary doesn't include inclination changes or specific margin allocations (Smith and Bell, 2005 and Halsell et al. 2003). Note that Mars Odyssey used aerobraking to descend from a $292 \times 26,800$ km (18.6-hour) orbit to a 387×451 km orbit and Mars Reconnaissance Orbiter (MRO) used aerobraking to descend from a $300 \times 44,600$ km (35.0-hour) orbit to a 274×450 km orbit, which was the target post-aerobraking orbit.

TABLE 12-3. Example Aerobraking Maneuvers. A summary of the aerobraking maneuvers performed by three example missions is shown. The comparison with chemical propulsion is given to illustrate the efficiency in aerobraking. The percentage of the aerobraking vs the chemical is shown on the total line. All units are m/s.

Phase	Magellan	Mars Odyssey	MRO
Walk-in	6.7	12.8	8.9
Main phase	~30	~30	~30
Contingencies	~20	~20	~20
Walk-out	16.4	65.5	65.5
Total Aerobraking, %	73.1, 6.0%	128.3, 11.5%	124.4, 10.3%
Chemical	1225.3	1113.1	1206.1

12.7 Three-body Mission Design

Traditional spacecraft trajectories are shaped by the gravity of one body (Sun, Earth, or another planet) and perturbed by other third bodies. The patched conic analyses demonstrate how to shift the design from one central body to another as the spacecraft passes through a body's sphere of influence. In some cases, the spacecraft's mission places it in a region of space that is strongly influenced by two massive bodies simultaneously. These trajectories cannot use a two-body analysis with third body perturbations, but must consider the effects of both bodies simultaneously. To understand the problem, we'll begin by studying the general three-body problem and how to apply it to real systems. We'll then present some particularly interesting trajectories and analyses that are being used by modern spacecraft.

In Chap. 1 we introduced general equations for the motion of many bodies. 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. Sundman 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.

Scientists have studied the three-body problem for hundreds of years, though the majority of the research has been accomplished since the 1960s, using modern computing technology. Their analyses have revealed a plethora of three-body orbits, many of which are very useful for spacecraft. We'll investigate several three-body orbit options in this section.

12.7.1 Circular Restricted Three-body Problem (CRTBP)

The general three-body problem considers the trajectories of three arbitrary massive bodies, traveling under mutual gravitation. However, this problem is much more general than is needed by practical mission designers. There are two simplifications that are

commonly made to the general three-body problem to make it more tractable and more applicable to trajectories in the real world. These include:

1. The mass of the third body (satellite) is negligible compared to that of the major bodies.
2. The primary and secondary bodies move in circular orbits about the center of mass, which lies between the two objects.

The resulting problem is usually called the *circular restricted three-body problem* (CRTBP). 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 simple equations to express solutions in the three-body problem 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. 12-13.* I'll introduce equations for a barycentric (fixed) frame shortly. This frame uses lowercase 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 .

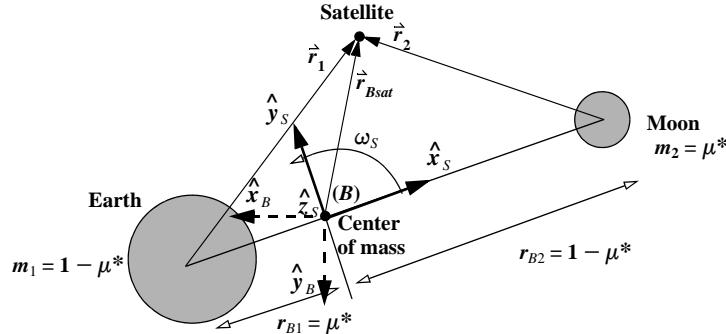


Figure 12-13. Geometry for the Restricted Three-body Problem. Notice the x_S -axis is in the direction of the secondary object (Moon), and the y_S -axis lies in the (Earth-Moon) orbital plane. Remember, this synodic coordinate system is rotating with respect to the fixed barycentric frame (x_B , y_B , z_B). The z_B and z_S axes are aligned. The distances from the barycenter to each object (r_{B1} , r_{B2}) are used to develop 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. (1-38), with $n = 3$, the inertial acceleration is

* Alternate notations exist for the primary direction. Szebehely uses a primary direction towards the larger body. Much of the literature and many operational missions use a primary axis towards the smaller body. The equations differ by only a few signs. We use the latter convention.

$$\ddot{\vec{r}}_{Bsat} = -Gm_1 \frac{\vec{r}_{1sat}}{r_{1sat}^3} - Gm_2 \frac{\vec{r}_{2sat}}{r_{2sat}^3}$$

The inherent rotation of the coordinate system requires us to examine the additional terms which arise in the acceleration. Eq. (3-26) delineates inertial and rotating accelerations (I've used symbols from this section and the rotating synodic frame, $\dot{\vec{r}}_S, \dot{\vec{v}}_S, \dot{\vec{a}}_S$):

$$\ddot{\vec{r}}_{Bsat} = \dot{\vec{a}}_S + \dot{\vec{\omega}}_S \times \dot{\vec{r}}_S + \dot{\vec{\omega}}_S \times (\dot{\vec{\omega}}_S \times \dot{\vec{r}}_S) + 2\dot{\vec{\omega}}_S \times \dot{\vec{v}}_S + \dot{\vec{a}}_{org}$$

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 the common practice of using cartesian coordinates in the synodic frame for the satellite and performing the cross product, we have

$$\ddot{\vec{r}}_{Bsat} = \ddot{\vec{r}}_S - \omega_S^2 (x\hat{x}_S + y\hat{y}_S) - 2\omega_S(\dot{y}\hat{x}_S - \dot{x}\hat{y}_S)$$

Now we must find simplified expressions for the inertial acceleration. Eq. (1-38) gives the general expression. At this point it's convenient to introduce gradients. We used gradients often in Chap. 8. 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. (3-26), 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 $\ddot{x}, \ddot{y}, \ddot{z}$ to represent the components of the rotating acceleration $\ddot{\vec{r}}_S$, in the synodic frame.

$$\begin{aligned} \ddot{x} - 2\omega_S \dot{y} - \omega_S^2 x &= \frac{\partial}{\partial x} \left(\frac{\mu_1}{r_1} + \frac{\mu_2}{r_2} \right) = -\frac{\mu_1(x+r_{B1})}{r_1^3} - \frac{\mu_2(x-r_{B2})}{r_2^3} \\ \ddot{y} + 2\omega_S \dot{x} - \omega_S^2 y &= \frac{\partial}{\partial y} \left(\frac{\mu_1}{r_1} + \frac{\mu_2}{r_2} \right) = -\frac{\mu_1 y}{r_1^3} - \frac{\mu_2 y}{r_2^3} \\ \ddot{z} &= \frac{\partial}{\partial z} \left(\frac{\mu_1}{r_1} + \frac{\mu_2}{r_2} \right) = -\frac{\mu_1 z}{r_1^3} - \frac{\mu_2 z}{r_2^3} \end{aligned} \quad (12-13)$$

$$r_1 = \sqrt{(x+r_{B1})^2 + y^2 + z^2} \quad r_2 = \sqrt{(x-r_{B2})^2 + y^2 + z^2}$$

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 *Jacobi integral* equation:

$$\dot{x}^2 + \dot{y}^2 + \dot{z}^2 - \omega_s^2(x^2 + y^2) = \frac{2\mu_1}{r_1} + \frac{2\mu_2}{r_2} - C \quad (12-14)$$

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* or the *three-body constant* of the restricted three-body problem^{*}—it's always less than $1/2$. The asterisk distinguishes μ^* 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

$$\begin{aligned} r_1 &= \sqrt{(x + \mu^*)^2 + y^2 + z^2} \\ r_2 &= \sqrt{(x + \mu^* - 1)^2 + y^2 + z^2} \end{aligned}$$

With the normalized angular velocity ($\omega_s = 1$), Jacobi's integral is now

$$v^2 = x^2 + y^2 + \frac{2(1 - \mu^*)}{r_1} + \frac{2\mu^*}{r_2} - C \quad (12-15)$$

Eq. (12-14) or Eq. (12-15) is called the *Jacobi integral* for the restricted three-body problem. Note that the Jacobi constant (C) is analogous to the two-body energy, but inverted. Thus, a spacecraft can travel further away from the Earth with higher two-body energies and lower Jacobi constant values. Most analyses use Eq. (12-15) as their starting point. Next we solve the equations of motion. We begin with Eq. (12-13) and simply do the partial differentiation. Using the mass ratio instead of μ_1 and μ_2 gives us the rotating, non-dimensional equations.

* The actual mass ratio is $\frac{\mu^*}{1 - \mu^*}$.

$$\begin{aligned}
 \ddot{x} - 2\dot{y} - x &= -\frac{(1-\mu^*)(x+\mu^*)}{r_1^3} - \frac{\mu^*(x-1+\mu^*)}{r_2^3} \\
 \ddot{y} + 2\dot{x} - y &= -\frac{(1-\mu^*)y}{r_1^3} - \frac{\mu^*y}{r_2^3} \\
 \ddot{z} &= -\frac{(1-\mu^*)z}{r_1^3} - \frac{\mu^*z}{r_2^3}
 \end{aligned} \tag{12-16}$$

Some applications require the barycentric equations of motion—those in a nonrotating frame. Szebehely (1967:21) shows the formulas.

$$\begin{aligned}
 \ddot{x} &= -\frac{\mu_1(x+r_{B1}\cos(\omega_S t))}{r_1^3} - \frac{\mu_2(x-r_{B2}\cos(\omega_S t))}{r_2^3} \\
 \ddot{y} &= -\frac{\mu_1(y+r_{B1}\sin(\omega_S t))}{r_1^3} - \frac{\mu_2(y-r_{B2}\sin(\omega_S t))}{r_2^3} \\
 \ddot{z} &= -\frac{\mu_1 z}{r_1^3} - \frac{\mu_2 z}{r_2^3} \\
 r_1 &= \sqrt{(x+r_{B1}\cos(\omega_S t))^2 + (y+r_{B1}\sin(\omega_S t))^2 + z^2} \\
 r_2 &= \sqrt{(x-r_{B2}\cos(\omega_S t))^2 + (y-r_{B2}\sin(\omega_S t))^2 + z^2}
 \end{aligned} \tag{12-17}$$

Three general areas are of interest: special cases called *equilibrium solutions* (or the *Lagrange points*), regions of motion which are related to the Jacobian constants (*periodic orbits*), and stability of motions from analyzing the *phase space* (or *Manifolds*). We'll examine each.

12.7.2 Lagrange Points

To examine the equilibrium positions, we first set the velocity and acceleration equal to zero in Eq. (12-16). There are 5 equilibrium positions, or *Lagrange points*. 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. The equilibrium positions for x and y are

$$x - \frac{(1-\mu^*)(x+\mu^*)}{r_1^3} - \frac{\mu^*(x-1+\mu^*)}{r_2^3} = 0 \quad y \left(1 - \frac{1-\mu^*}{r_1^3} - \frac{\mu^*}{r_2^3} \right) = 0 \tag{12-18}$$

If we set $y = 0$, we can solve x using a quintic equation to yield the three *collinear* 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 to find the distance (ξ) from each Lagrange point to the nearest primary (Szebehely, 1967:135–138):

$$\begin{aligned}\xi^5 + (3 - \mu^*)\xi^4 + (3 - 2\mu^*)\xi^3 - \mu^*\xi^2 - 2\mu^*\xi - \mu^* &= 0 \\ \xi^5 - (3 - \mu^*)\xi^4 + (3 - 2\mu^*)\xi^3 - \mu^*\xi^2 + 2\mu^*\xi - \mu^* &= 0 \\ \xi^5 + (2 + \mu^*)\xi^4 + (1 + 2\mu^*)\xi^3 - (1 - \mu^*)\xi^2 - 2(1 - \mu^*)\xi - (1 - \mu^*) &= 0\end{aligned}$$

There are two notations for where L_1 lies, either between the bodies, or outside. Szebehely (1967) and Kaplan (1976) place L_2 between the two bodies, however common practice places L_1 between the Earth and Moon. This is the convention adopted here.

Eq. (12-18) 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 points (with x, y coordinates):

$$L_4 = \mu^* - \frac{1}{2}, \frac{\sqrt{3}}{2} \quad L_5 = \mu^* - \frac{1}{2}, -\frac{\sqrt{3}}{2}$$

Figure 12-14 shows the Lagrange points for the Earth-Moon system, $\mu^* = 0.012\ 14$. The Sun-Jupiter system is also common with $\mu^* = 0.000\ 95$. This value is 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 can determine the boundary of permissible motion for the spacecraft by observing the spacecraft's zero-velocity surface, defined by its Jacobi constant. A spacecraft's zero-velocity curve is found by considering its Jacobi constant and setting its velocity to zero in Eq. (12-15). Example zero-velocity curves are illustrated in Fig. 12-14 for a variety of Jacobi constant values in the Earth-Moon system. For instance, a spacecraft with $C = 3.01$ can reach any position in the system except for the forbidden regions near L_4 and L_5 .

Figure 12-15 shows plots of zero-velocity curves for a system in which $\mu^* = 0.3$. Notice how different the shapes of these curves are than for the Earth-Moon's μ^* of 0.01214 in Fig. 12-14. 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.

The motion of a spacecraft with a large Jacobi constant is restricted to a small region of space around either primary (whichever it happens to be at) or the region of space far from either primary. Think of those spacecraft as not having enough energy to match the rotational motion of the primaries. If the spacecraft's Jacobi constant were lower, it could reach further. The Lagrange points define the limiting cases: a spacecraft with the Jacobi constant equal to that of the L_1 point can just cross over from one primary to the other. A spacecraft with the Jacobi constant of the L_2 point can cross the boundary to the exterior. A spacecraft with the Jacobi constant of the L_4/L_5 point can reach any point in the rotating system.

Figure 12-16 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 the restricted nature of regions of motion outside the horizontal plane. Three-

* These triangular points are also sometimes called *equilateral* points.

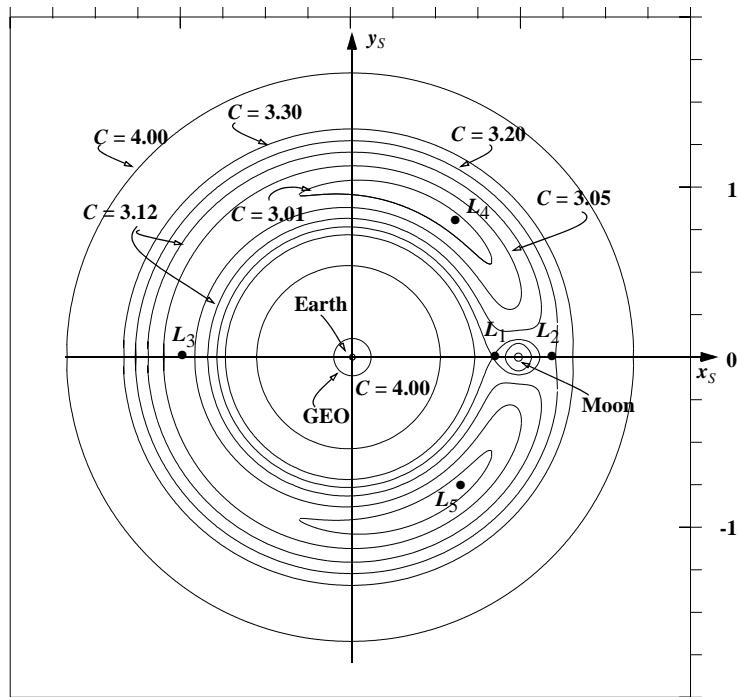


Figure 12-14. Regions of x - y Motion for the Earth-Moon System, $\mu^* = 0.012\ 14$. This representation of motion in the Earth-Moon system shows the Lagrange points, as well as the positions of the Earth and Moon. A geosynchronous orbit is shown about the Earth to give a sense of scale. Motion across curves of lower C values is possible only with additional force. The axis center is actually at the barycenter, although not visible at this scale.

dimensional figures are possible by combining Fig. 12-14 and Fig. 12-16. Consult Lundberg et al. (1985) for additional information.

12.7.3 Periodic Orbits

Apart from the five equilibrium points—the Lagrange points—the three-body problem contains a wide variety of periodic and quasi-periodic orbits. Some orbits resemble two-body orbits, but are perturbed by the other primary. The orbits that revolve about the Lagrange points are commonly called *libration orbits*. *Periodic orbits* are trajectories that retrace their path in the synodic reference frame over time while *quasi-periodic orbits* trace a nearby path that is confined to a particular region in the synodic frame. There are several types of libration orbits that further qualify the motion. *Lyapunov orbits* are two-dimensional orbits that remain in the orbital plane of the primaries. *Halo orbits* are perfectly periodic three-dimensional orbits.* *Lissajous orbits* include quasi-periodic as well as periodic three-dimensional orbits similar to halo orbits. *Vertical*

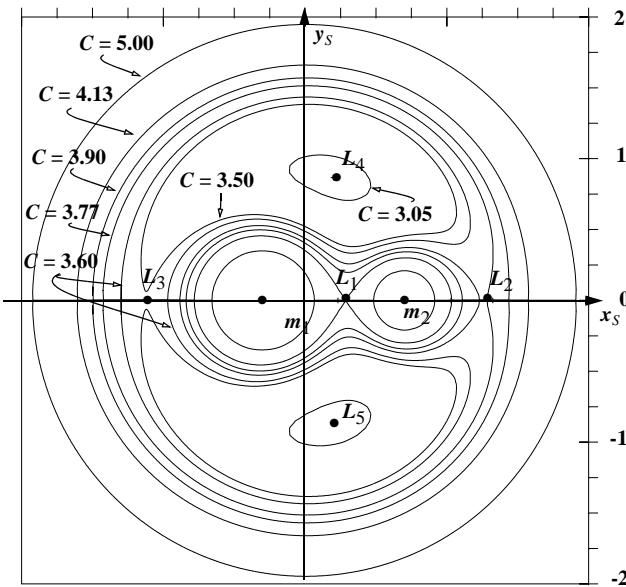


Figure 12-15. Regions of x-y Motion for a System with $\mu^* = 0.3$. This view of the motion in the x - y plane shows the five Lagrange points. This large mass ratio applies best to binary star systems.

Lyapunov orbits are perfectly periodic three-dimensional orbits that cross the xy plane through the Lagrange point. The first periodic orbits were discovered in the late 19th century and the catalog of known orbits has since exploded.

To develop the equations, think about periodic three-body orbits using two-body intuition and consider periodic orbits about either of the two massive bodies. The periodic orbit is essentially a two-body orbit in resonance with the perturbing third body. The perturbing body shifts the orbit away from a perfect conic, but it's close. We will demonstrate why periodic orbits can exist about a Lagrange point using an analytical construction [Moulton (1914:296), Richardson (1980), and Howell and Pernicka (1988)]. We closely follow Howell and Pernicka in the following paragraphs.

Begin by translating the origin of the synodic frame to one of the collinear libration points L_i . The parameter γ is defined to be equal to the distance from L_i to the smaller primary. The value of γ is positive when referring to L_2 and negative when referring to L_1 and L_3 . The new position coordinates x' , y' , and z' are defined as

$$\begin{aligned}x' &= x - (1 - \mu^* + \gamma) \\y' &= y\end{aligned}$$

* The name “halo” originated with Robert Farquhar. He noted that a perfectly periodic orbit about the Moon’s L_1 or L_2 point would appear to make a halo around the Moon, as viewed from Earth.

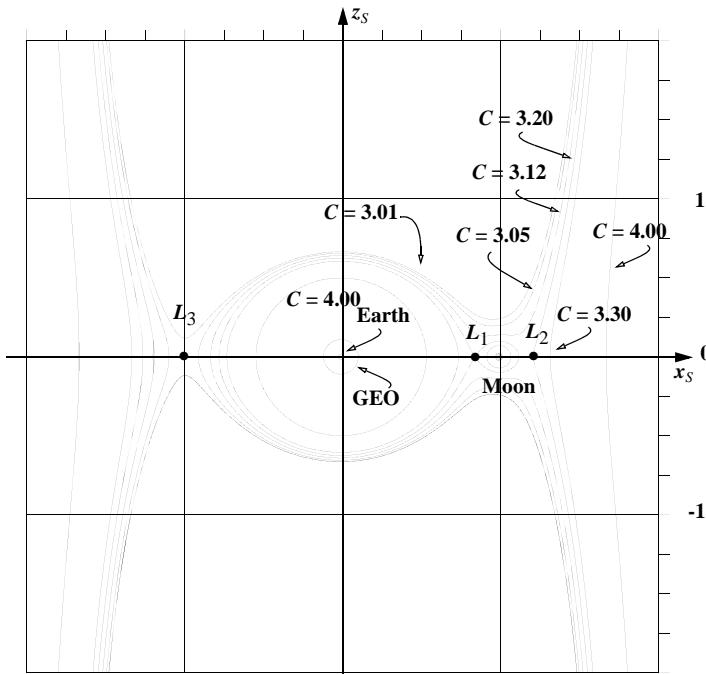


Figure 12-16. Regions of x - z Motion for the Earth-Moon System, $\mu^* = 0.012\ 14$. This equatorial view shows the zero-velocity curves for the Earth-Moon system. The Earth and geosynchronous orbit are shown for scaling purposes. Notice how restricted out-of-plane motion becomes near the Earth.

$$z' = z$$

If we now linearize the equations of motion of the CRTBP given in Eq. (12-16) under this transformation, we find the following,

$$\ddot{x}' - 2\dot{y}' - (1 + 2c)x' = 0$$

$$\ddot{y}' + 2\dot{x}' + (c - 1)y' = 0$$

$$\ddot{z}' + cz' = 0$$

where c is a constant. The analytical solution to the out-of-plane z motion describes simple harmonic motion. The solution for the in-plane x - y motion involves a characteristic equation that has two real roots and two imaginary roots. The roots represent modes of motion—one divergent and one nondivergent. If the nondivergent mode is excited, then the solution is bounded and is written as follows. Note that the frequency and phase angle in the z -equation are independent of the frequency and phase angle for the x - and y -equations.

$$x' = -kA_y \cos(\lambda t + \varphi)$$

$$y' = A_y \sin(\lambda t + \varphi)$$

$$z' = A_z \sin(\nu t + \psi)$$

This motion is described by six variables: the amplitudes of the in-plane and out-of-plane motion A_y and A_z , the frequency of oscillation in the in-plane and out-of-plane motion λ and ν , and the phase angles for the in-plane and out-of-plane motion φ and ψ . The linearized approximation to the analytical solution for periodic motion about a Lagrange point is characterized by oscillatory motion. If the two frequencies λ and ν are equal or otherwise commensurate, the resulting motion is periodic. If the frequencies are incommensurate, the resulting motion will be quasiperiodic. Figure 12-17 illustrates this periodic and quasiperiodic motion.

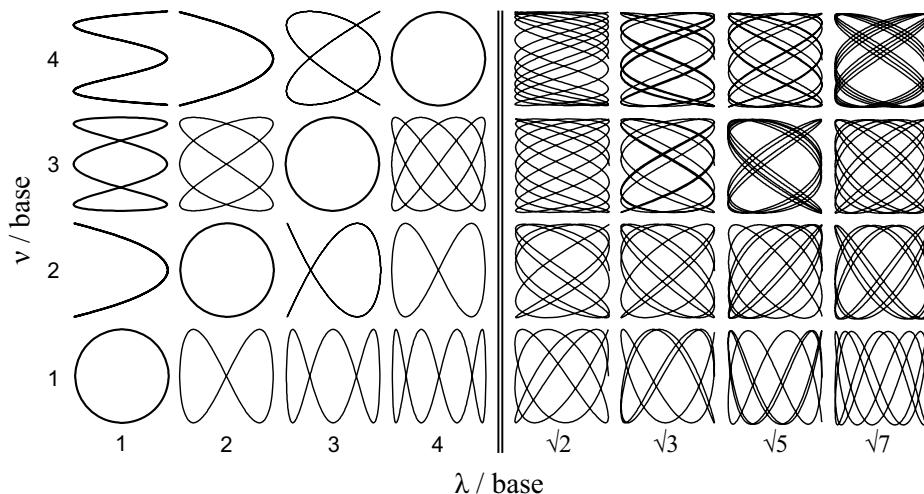


Figure 12-17. Variations of Perfectly-Periodic and Quasi-Periodic Orbits. An illustration of the yz motion of a spacecraft traversing Lissajous* orbits for different ratios of λ to ν . The orbit is periodic if the ratio is commensurate and quasi-periodic otherwise. Recognize that these are actually idealized Lissajous curves with the Lagrange point in the center. Perturbations would change the curves slightly.

The motion of a spacecraft about a Lagrange point may be illustrated using two-body intuition as follows. First, we start with a spacecraft that is closer to the Earth than the L_1 point, and traveling faster in its orbit than the rotation speed of the primaries. The spacecraft gradually moves ahead of the Moon in its orbit about the barycenter. As it moves ahead, the Moon's gravity pulls on it, slowing it down. This causes the spacecraft to move outward, away from the Earth. As it does, its velocity relative to the barycenter slows down and it begins to fall back toward the centerline. After it crosses the Earth-Moon line, the Moon again pulls on it, causing it to accelerate. This acceleration pulls the spacecraft toward the Earth again, and back to where it started. This cycle may be carefully designed to repeat "indefinitely".

Halo Orbits

The dynamics of the three-body problem permit a spacecraft to revolve about a Lagrange point. The periodic orbits whose in-plane and out-of-plane frequencies are equal are known as **halo orbits**. A portion of the family of halo orbits about the Earth-Moon L_2 is shown in Fig. 12-18. Notice how the orbits are asymmetric about the x - y plane. Figure 12-18 shows orbits in the family of southern halo orbits. The northern halo orbits are simply reflected about the x - y plane.

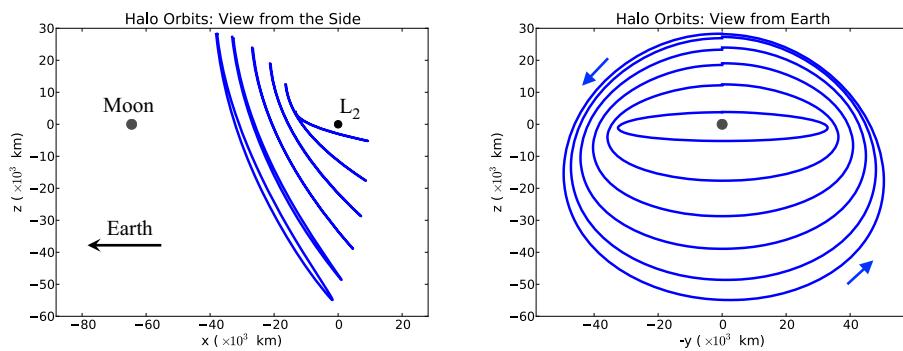


Figure 12-18. Earth-Moon Southern Orbits: Example southern halo orbits that exist about the Earth-Moon L_2 point, viewed from several angles in the Earth-Moon synodic reference frame. (Figure adapted from Parker 2007)

12.7.4 Manifolds

A spacecraft placed precisely at a Lagrange point will remain there indefinitely and remain fixed relative to the motion of the two massive bodies. However, any perturbation will cause the spacecraft to drift away. The spacecraft will exponentially depart L_1 , L_2 , or L_3 , and it will remain in the vicinity of L_4 or L_5 . The two triangular points (L_4 and L_5), are dynamically stable points, so perturbed particles remain in their vicinity over time. The collection of trajectories that travel away from a Lagrange point over time define its **unstable manifold**. Likewise, trajectories that approach a Lagrange point define its **stable manifold**.

Stability of the motion about the Lagrange points always sparks considerable interest. Orbits near one of the primaries are stable as discussed previously. Orbits at the collinear Lagrange points are unstable. Thus, there is a transition near the Lagrange points where trajectories become unstable. A trajectory is unstable if it's sensitive to the initial conditions. This is clearly the case for a spacecraft that has a value of C (~ 3.0) very near one of the Lagrange points (see Fig. 12-14 and Fig. 12-16) that has the potential to “cross over” to another part of the three-body system. The trajectories used during the Apollo program had values of C that crossed between the Earth and the Moon ($C \sim 1.756$). Many of the trips to the Moon used so-called *free-return trajectories*, in which the spacecraft could pass to the Moon and back with only a small thrusting force (not all missions used these trajectories). Much of the analysis in this area deals with nonlinear

dynamics and *phase space* or *manifold* techniques. Consult Szebehely (1967) for additional details.

The trajectories that describe the most stable and most unstable pathways from each Lagrange point define that point's stable (W^S) and unstable manifolds (W^U). This is illustrated in Fig. 12-19. Any random perturbation in the state of a spacecraft at the Lagrange point has stable and unstable components. The stable components shrink asymptotically to zero while the unstable components grow exponentially. Thus, any random perturbation drives the spacecraft toward the unstable manifold.

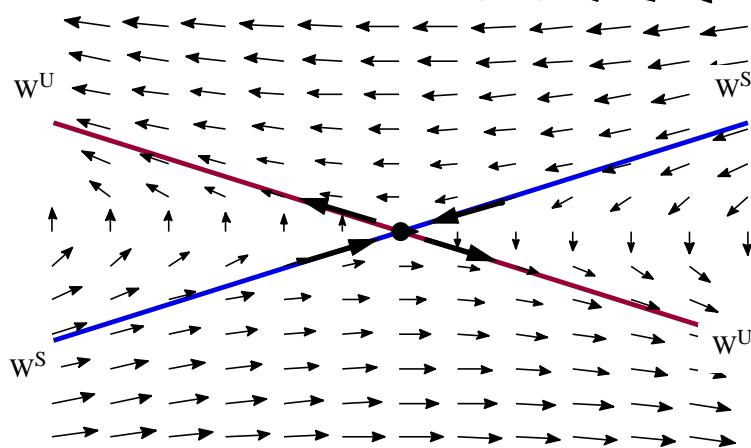


Figure 12-19. Stable and Unstable Points: A dynamical saddle point, such as between the three collinear Lagrange points in the CRTBP, with a vector field that indicates the motion of a particle near the point. Note how the magnitude (arrow length) changes along the various paths from the stable to unstable points. The notional Lagrange point is the black dot in the center.

A mission requiring the spacecraft to remain at the Lagrange point may do so using station-keeping maneuvers, which may be very small if performed frequently (about weekly for the Earth-Moon system, and about monthly for the Earth Sun system) and/or accurately enough.

Figure 12-19 shows that the component of the perturbation of a spacecraft's state along the stable manifold shrinks over time. By extension, if a spacecraft's state is precisely on the stable manifold, it will be driven to the Lagrange point. A spacecraft may be placed at the Lagrange point using virtually no fuel by placing it anywhere on the Lagrange point's stable manifold—even if it starts thousands of kilometers away. The two trajectories in the stable manifold may be generated by propagating them away from the Lagrange point backward in time.

We determine the stable and unstable directions, and corresponding manifolds by analyzing the eigenvalues and eigenvectors of the Jacobian of the Lagrange point in the CRTBP. If \mathbf{X} is the state of one of the collinear Lagrange points $[x, y, z, \dot{x}, \dot{y}, \dot{z}]^T$, then its Jacobian is

$$J = \frac{d\dot{\hat{X}}}{d\hat{X}} = \begin{bmatrix} \frac{\partial \dot{x}}{\partial x} & \frac{\partial \dot{x}}{\partial y} & \frac{\partial \dot{x}}{\partial z} & \frac{\partial \dot{x}}{\partial \dot{x}} & \frac{\partial \dot{x}}{\partial \dot{y}} & \frac{\partial \dot{x}}{\partial \dot{z}} \\ \frac{\partial \dot{y}}{\partial x} & \dots & & & & \\ \vdots & & & & & \vdots \\ \frac{\partial \ddot{z}}{\partial x} & \frac{\partial \ddot{z}}{\partial y} & \frac{\partial \ddot{z}}{\partial z} & \frac{\partial \ddot{z}}{\partial \dot{x}} & \frac{\partial \ddot{z}}{\partial \dot{y}} & \frac{\partial \ddot{z}}{\partial \dot{z}} \end{bmatrix}$$

Inserting Eq. (12-16), the Jacobian is simplified.

$$J = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ \frac{\partial \ddot{x}}{\partial x} & \frac{\partial \ddot{x}}{\partial y} & \frac{\partial \ddot{x}}{\partial z} & 0 & 2 & 0 \\ \frac{\partial \ddot{y}}{\partial x} & \frac{\partial \ddot{y}}{\partial y} & \frac{\partial \ddot{y}}{\partial z} & -2 & 0 & 0 \\ \frac{\partial \ddot{z}}{\partial x} & \frac{\partial \ddot{z}}{\partial y} & \frac{\partial \ddot{z}}{\partial z} & 0 & 0 & 0 \end{bmatrix}$$

The eigenvalues of the Jacobian for each of the three collinear Lagrange points include two pairs of imaginary numbers and one pair of real numbers. The eigenvector corresponding to the larger real eigenvalue indicates the unstable direction, v^U , and the eigenvector corresponding to the other real eigenvalue indicates the stable direction, v^S . The unstable manifold of the Lagrange point, W^U , may be mapped by propagating the state X^U forward in time, where $X^U = X \pm \epsilon v^U$ and ϵ is a small perturbation. Similarly, the stable manifold, W^S , may be mapped by propagating the state X^S backward in time, where $X^S = X \pm \epsilon v^S$.

Figures 12-20 to 12-22 illustrate the stable and unstable manifolds of the three collinear Lagrange points of the Earth-Moon system. You can see that a spacecraft with just enough energy to reach the inner Lagrange point does not have enough energy to escape the Earth altogether. A spacecraft with enough energy to reach the outer Lagrange point can almost escape.

Many three-body orbits are unstable, including libration orbits about the collinear Lagrange points of the Earth-Moon and Sun-Earth systems, just as the Lagrange points are. A spacecraft attempting to traverse an unstable orbit diverges away from that trajectory after any random perturbation in its motion, much like the motion of a spacecraft at one of the collinear Lagrange points. Unstable periodic orbits also have stable and unstable manifolds. Their manifolds aren't one-dimensional like the Lagrange point manifolds. Instead, they define a tube-like surface in three-dimensional space. A spacecraft may depart a periodic orbit at any point along that orbit, thereafter tracing out a

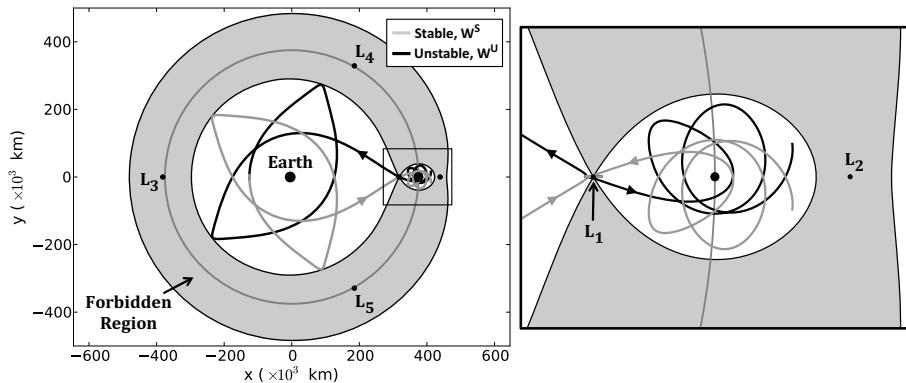


Figure 12-20. Earth-Moon L_1 Manifolds: The stable and unstable manifolds of the Earth-Moon L_1 point, viewed from above in the Earth-Moon synodic frame. The Forbidden Region is shown for reference, indicating the region of space that is inaccessible by a spacecraft with the energy of the L_1 point.

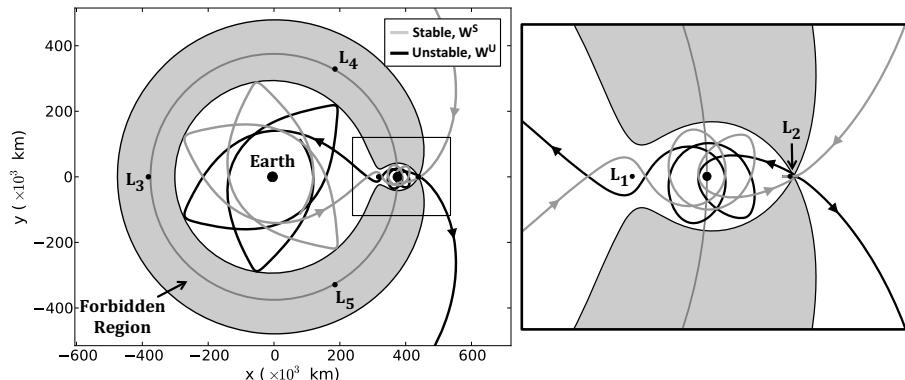


Figure 12-21. Earth-Moon L_2 Manifolds: The stable and unstable manifolds of the Earth-Moon L_2 point, viewed from above in the Earth-Moon synodic frame.

single trajectory within the orbit's unstable manifold. The unstable manifold includes the full set of all possible trajectories a spacecraft can take when departing the orbit. Similarly, the set of all possible trajectories that asymptotically approach an orbit forms that orbit's stable manifold.

To produce the invariant manifolds of an unstable periodic orbit, you need information about the local stability characteristics of each point along the orbit. In theory, you can evaluate the eigenvalues and eigenvectors of the Jacobian at each and every state along the orbit, and use that information to produce the orbit's invariant manifold. However, evaluating so many eigenvalues requires significant computational resources. A more efficient manner of producing the invariant manifolds uses the eigenvalues and

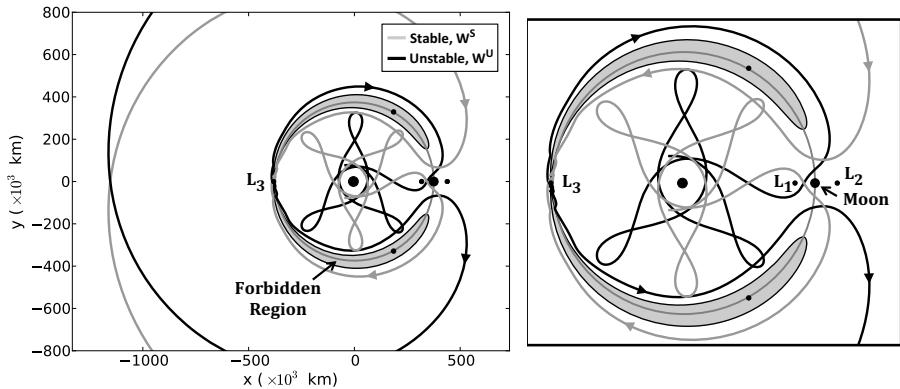


Figure 12-22. Earth-Moon L_3 Manifolds: The stable and unstable manifolds of the Earth-Moon L_3 point, viewed from above in the Earth-Moon synodic frame.

eigenvectors of the monodromy matrix (Gómez et al. 2004 or Ross, 2004). The **monodromy matrix** is constructed by propagating the state transition matrix, Φ , around the periodic orbit, and forming $M = \Phi(t_0+T, t_0)$. The monodromy matrix contains information about the stability of the entire orbit. To determine the stable and unstable directions at each point along the orbit, we only need to propagate the stable and unstable eigenvectors of the monodromy matrix about the orbit using the state transition matrix. For instance, the stable and unstable vectors at time t_i about the orbit, v_i^S and v_i^U , respectively, may be determined using the stable and unstable eigenvectors of the monodromy matrix, v^S and v^U , respectively, using the following equations.

$$v_i^S = \Phi(t_i, t_0)v^S \quad v_i^U = \Phi(t_i, t_0)v^U$$

A small perturbation ϵ is then applied to the state of the orbit at that time, X_i , and the result is propagated in time. Because the state transition matrix grows exponentially along an unstable orbit, the magnitudes of the vectors v_i^S and v_i^U grow along the orbit. It's important to normalize the vectors so that a consistent perturbation is applied to each orbit state. The final equations to produce the initial conditions for the stable and unstable manifolds at time t_i about the orbits X_i^S and X_i^U respectively, are then equal to

$$X_i^S = X_i \pm \epsilon \frac{v_i^S}{|v_i^S|} \quad X_i^U = X_i \pm \epsilon \frac{v_i^U}{|v_i^U|}$$

Mathematicians may use a very small value (ϵ) to accurately trace the manifolds. Mission designers often use large perturbations when building practical missions, so the spacecraft arrives/departs more quickly. Typical perturbations are on the order of 100 km in the Earth-Moon system and 1000 km in the Sun-Earth system. The perturbations are scaled proportionally for their velocity components.

To plot these, take an orbit and find the states of that orbit at evenly-spaced intervals in time. Then perturb each of those states in the direction of the local unstable or stable eigenvector. If you’re computing the unstable manifold, you propagate the perturbed state forward in time. For the stable manifold, you propagate the perturbed state backward in time.

Figure 12-23 shows the stable and unstable manifolds of a libration orbit about the Earth-Moon L_2 point. Notice that the manifolds are smooth and form a tube-like structure. They remain well-defined until they encounter the Moon, when they spread out very rapidly and the tube-like structure becomes less obvious. The stable and unstable manifolds are symmetric about the x -axis of the rotating frame, due to the symmetry in the CRTBP.

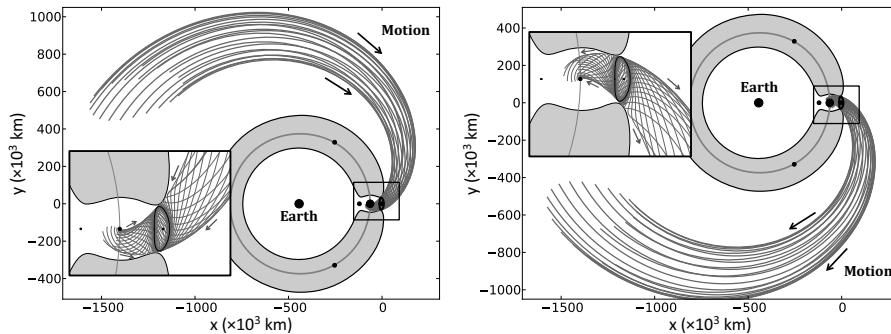


Figure 12-23. Earth-Moon L_2 Manifolds: The stable (left) and unstable (right) manifolds of a libration orbit about the Earth-Moon L_2 point, viewed from above in the Earth-Moon synodic frame.

In reality, a spacecraft can follow an unstable orbit if it performs periodic maneuvers. Genesis, SOHO, WMAP, (etc) have remained in Sun-Earth libration orbits using very little fuel. They typically require on the order of 1-10 m/s per year of Δv . The two ARTEMIS spacecraft used maneuvers to traverse libration orbits in the Earth-Moon system, also using very little fuel.

12.8 Low-energy Transfers

Several spacecraft missions have taken advantage of low-energy trajectories, mostly near the Earth or Moon, including ISEE-3, WMAP, SOHO, Genesis, WIND, GRAIL, and several others. Trajectories are labeled “low-energy” if they involve significant orbit changes that require very little fuel (admittedly an imprecise definition!). This scenario is usually possible with gravity assist trajectories as well as three-body trajectories.

The GRAIL mission is an example of a low-energy trajectory. The two GRAIL spacecraft were launched onto a low-energy lunar transfer—a very large orbit about the Earth that was significantly influenced by the Sun’s gravity. The Sun’s gravity increased each spacecraft’s energy during the lunar transfer, so that when the two spacecraft

arrived at the Moon, each was able to enter its lunar orbit for about 200 m/s less Δv than a conventional transfer.

The Genesis spacecraft is another example. Genesis was launched into an orbit that was essentially an Earth-perturbed heliocentric orbit. Its orbit drifted between the Sun and the Earth, about 1.5 million kilometers from the Earth, in such a way that wouldn't be possible with two-body motion. Toward the end of its mission, it traveled away from the Sun to the far side of the Earth—about 1.5 million kilometers from Earth—turned around and then targeted an Earth-return trajectory to deposit a capsule in Utah. All of these orbit transfers used virtually no fuel.

Several other spacecraft have been placed in orbits about either the Sun-Earth L_1 or L_2 points for scientific missions. The SOHO spacecraft has taken advantage of its position at L_1 to have a constant, unimpeded view of the Sun. The Herschel and Planck spacecraft reside at the Sun-Earth L_2 point, where it's possible to hide a heat-sensitive infrared telescope behind a sunshade. The sunshade blocks the Sun, Earth, and Moon from view, making it possible for the infrared telescope to remain very cold while it observes the cosmos.

It is possible to transfer a spacecraft to an unstable libration orbit at either the Sun-Earth L_1 and L_2 points using the orbit's stable manifold. You need only find a libration orbit whose stable manifold intersects the Earth at a desirable time, inclination, declination, etc. For instance, Fig. 12-24 illustrates a halo orbit about the Sun-Earth L_2 point and its stable manifold. The highlighted trajectories are those that come within 500 km of the Earth. If a spacecraft launches from the Earth onto any of those trajectories, the spacecraft will ballistically approach that host orbit.

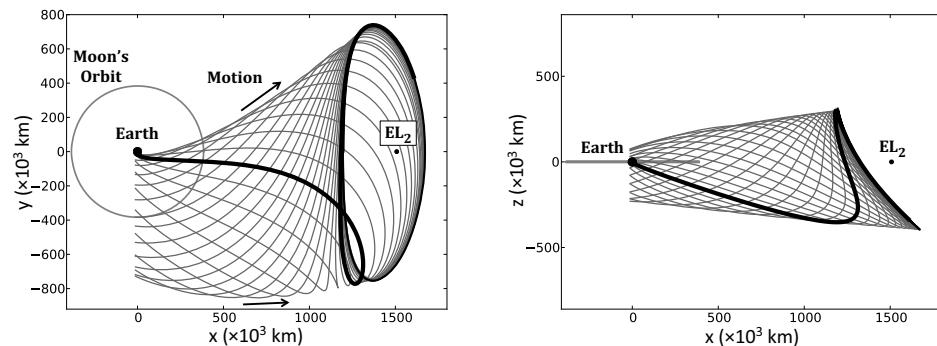


Figure 12-24. Low-energy Transfer: An example unstable halo orbit about the Sun-Earth L_2 point (EL_2) and its stable manifold. The highlighted trajectories come within 500 km of the Earth's surface.

Mission designers can also use the manifolds of unstable orbits to transfer a spacecraft from one unstable orbit to another with very little fuel. Figure 12-25 illustrates this by showing two example low-energy transfers from one orbit about the Earth-Moon L_1 point to another orbit about the Earth-Moon L_2 point. Transfers like these require virtu-

ally no fuel as long as the orbits' manifolds connect in phase space. At a minimum, this requires that the orbits have the same Jacobi constant.

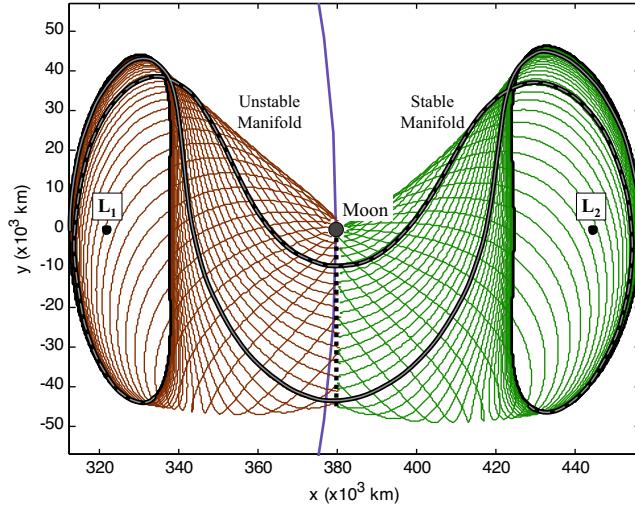


Figure 12-25. Low-Energy Orbit Transfers: Two possible trajectories that a spacecraft can take from an orbit about L_1 to an orbit about L_2 following their manifolds. These transfers are viewed in the Earth-Moon synodic frame.

Low-energy transfers may also be constructed between orbits in the Earth-Moon system and orbits in the Sun-Earth system. Figure 12-26 illustrates the relative proximity of the Sun-Earth L_1 (EL_1) and L_2 (EL_2) points relative to the five Earth-Moon Lagrange points (LL_1 - LL_5).

Figure 12-24 shows how you can launch from the Earth onto a trajectory within the stable manifold of an EL_2 libration orbit. From that orbit, a satellite can follow its unstable manifold back toward the Earth, but rather than orbiting the Earth, it can enter the stable manifold of an orbit about the Earth-Moon L_2 point. Mission designers have studied such transfers since the 1960s, though the trade space is only recently being well mapped out. Charles Conley was among the first to consider trajectories such as this, that arrive at the Moon via the LL_2 point (1968). Belbruno and Miller constructed a similar low-energy transfer to save the Japanese Hiten mission (1991). A total of five spacecraft have used low-energy transfers to the Moon: Hiten in 1991, the two ARTEMIS spacecraft in 2010 and the two GRAIL spacecraft in 2011.

The low-energy transfer does not have to enter a three-body orbit in the Sun-Earth system, as illustrated in Fig. 12-24. It can approach the orbit and transfer to the Moon instead. Parker (2007) describes how to model a low-energy transfer as a series of three-body orbit transfers, but never entering the three-body orbit, as illustrated in Fig. 12-27.

The low-energy transfer has several advantages over a direct transfer to the Moon. First, the low-energy transfer typically requires significantly less Δv than a direct trans-

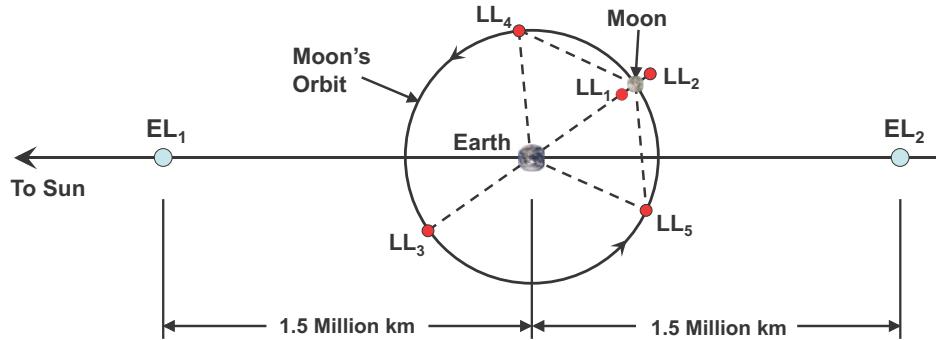


Figure 12-26. Earth-Sun Lagrange points: This figure (not to scale) shows the Sun Lagrange points (EL_1 and EL_2) and the Earth-Moon Lagrange points (LL_1 through LL_5).

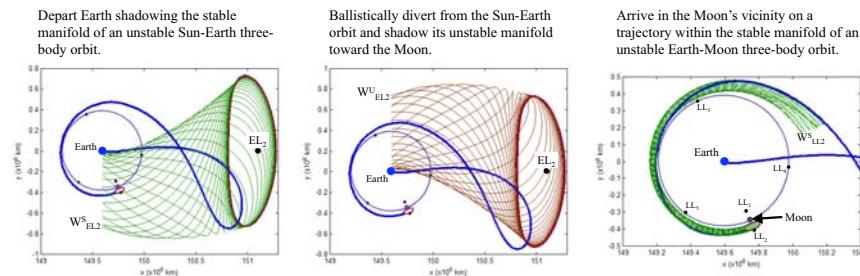


Figure 12-27. Series of Low Energy Orbit Transfers: Several low-energy transfers are shown from the Earth to the LL_2 Lagrange point.

fer. The Hiten and ARTEMIS missions were only possible because of the lower Δv requirements, because neither was launched with the intention of traveling to the Moon, and neither had enough fuel to follow a direct transfer. The 3-4 month flight time is usually undesirable for crewed missions to the Moon, but it is not usually a problem for robotic spacecraft.

Figure 12-28 shows the difference between a 4-day direct lunar transfer and an 84-day low-energy transfer between the Earth and a 100-km low lunar, polar orbit. The direct transfer requires a c_3 around $-2.0 \text{ km}^2/\text{s}^2$, whereas the low-energy transfer requires a c_3 around $-0.6 \text{ km}^2/\text{s}^2$. This means that the launch vehicle must achieve a velocity about 64 m/s higher, when using an altitude of 185 km above the surface. This difference can often be filled using leftover fuel in the rocket. When the spacecraft arrives at the Moon, it arrives with a velocity typically around 170 m/s lower following a low-energy transfer than a direct transfer. The difference can vary significantly.

Missions to lunar libration orbits, such as ARTEMIS, gain a much more significant benefit by following low-energy transfers. As Fig. 12-27 shows, a spacecraft can follow

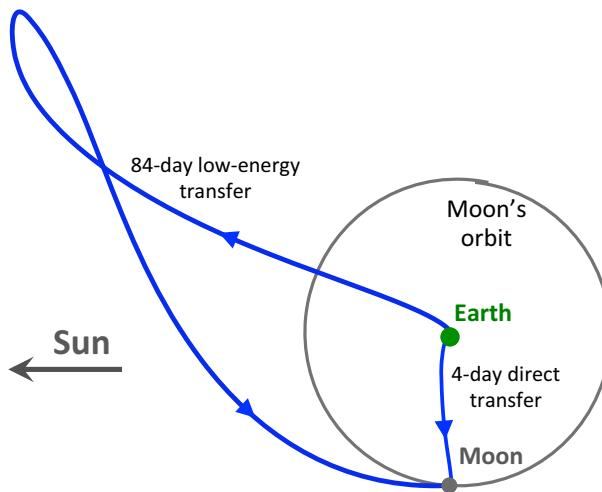


Figure 12-28. Low Energy 4-day: An illustration of the difference between a direct 4-day transfer to the Moon and an 84-day low-energy transfer, each viewed from above in the Sun-Earth rotating frame.

a low-energy lunar transfer and arrive at a target lunar libration orbit without requiring any orbit insertion. The trajectory arrives on the libration orbit's stable manifold, which carries it into orbit ballistically. Direct transfers, on the other hand, require 400–500 m/s or more to enter a libration orbit. This insertion maneuver can be reduced for missions to LL_2 orbits using a lunar flyby, though the transfer cost is still typically 300 m/s or higher.

Problems

1. Compute the sphere of influence of Jupiter. The natural satellite furthest from Jupiter (S/2003 J 2) is in an orbit with an average distance to Jupiter of approximately 29,540,000 km (as of 2013). It's commonly thought that satellite's orbits are stable within about 67% of the planet's SOI. Is this moon's orbit stable? How does this relate with Saturn and its furthest natural satellite (Fornjot), which has a semi-major axis of 24,504,879 km relative to Saturn? Could additional satellites be discovered based on these numbers?
2. 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?

3. Compute the departure c_3 and arrival v_∞ for a Mars transfer that departs Earth on May 15, 2018 and arrives at Mars on January 1, 2019. If the transfer begins in a circular orbit about Earth with an altitude of 185 km, what impulsive Δv is required to obtain the necessary c_3 to depart Earth onto this transfer? What velocity will the spacecraft have at an altitude of 100 km above Mars' surface? What is the declination of the departure asymptote from Earth relative to Earth's equator (J2000)? Can this transfer depart from Cape Canaveral (28.5° latitude) without a plane change?
4. 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.
5. 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)
 - a. How do you proceed?
 - b. If you need reduction (See Sec. 3.7), what epoch might you choose?
 - c. Does the n -body formula apply?

On May 5, 2000, the seven inner planets were *somewhat* aligned. Does this present another opportunity to “tour” the planets? Explain your answer. (Consider patched conic approximations)

6. Find the colinear Lagrange points for Jupiter (Use data from Table D-4).
7. 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. 6.3.1)
8. Determine the theoretical limits to the capacity of the gravity assist to change the spacecraft's heliocentric velocity. The maximum change in the spacecraft's heliocentric speed occurs for gravity assists with a 180° turn angle, changing the Δv vector from a vector that pointed precisely in the same direction as the heliocentric velocity vector to a vector pointed in the opposite direction (or vice versa). The maximum change in inclination may be determined similarly. These gravity assists may be theoretically valid, however, they may require a flyby with a radius of closest approach within the planet's surface. It's prudent to verify that any flyby doesn't approach too close to the planet's surface. Find the following:
 Maximum turn angle that a spacecraft could achieve flying by the Earth with a v_∞ of 6 km/s and with a minimum flyby altitude of 300 km?
 Theoretical maximum heliocentric velocity a spacecraft could have departing the Earth with a v_∞ of 6 km/s?

Minimum velocity?

Maximum turn angle that may be achieved during a Jupiter flyby if the minimum radius of the flyby coincides with Io's orbit?

Maximum heliocentric inclination that may be achieved during a Jupiter flyby with a v_∞ of 11 km/s?

9. Galileo flew to Jupiter using a VEEGA trajectory (Venus-Earth-Earth Gravity Assist). It performed a few small maneuvers to refine the trajectory, but otherwise visited the planets using ballistic trajectories from one encounter to the next, encountering each body on the following schedule:

Body	Date
Earth Departure	10/8/1989
Venus Flyby	2/10/1990
Earth Flyby	12/10/1990
Earth Flyby	12/9/1992
Jupiter Arrival	3/21/1996

What is the v_∞ value that Galileo had as it departed Earth en route to Venus, and en route to Jupiter?

How close did Galileo get to Venus during its flyby?

10. Set up a normalized circular restricted three-body system for the Jupiter-Europa system, using the masses and orbits given in Table D-4. Use a Europa mass = 4.7998×10^{22} kg. What is the value of μ^* ? What is the value of the Jacobi constant for a spacecraft that has zero velocity in the Jupiter-Europa rotating frame and is located at a position of $(x, y, z) = (0.7, 0.2, 0.1)$? Solve the quintics in Sec. 12.7.2 to identify the locations of the 5 Lagrange points in this three-body system.
11. Compute the B-Plane parameters (B_R and B_T) for a spacecraft approaching Venus with a position vector of $[-1.5 \times 10^7, 2.2 \times 10^6, -4.2 \times 10^6]$ km and a velocity vector of $[4.1, 0.2, 0.3]$ km/s, relative to Venus in J2000 coordinates. Note that the v_∞ of this incoming trajectory may be determined by computing the specific energy of the spacecraft because the B-Plane equations (Algorithm 78) requires v_∞ as an input, but in reality, v_∞ may be computed from r and v . Compute the partials of B_R and B_T with respect to the incoming velocity. What is the estimated Δv required to increase B_R by 50 km? What is the estimated Δv to decrease B_T by 10 km? Will the time of flight change for these adjustments? By how much?
12. Assume that an aerobraking campaign at Mars uses the same Δv in the main phase and contingencies as illustrated in the examples shown in Table 12-3. Thus, assume that the main phase Δv is 30 m/s and an additional 20 m/s is allocated to contingencies. What is the Δv savings by using aerobraking compared to chemically reducing an orbit when a spacecraft's initial orbit is 300 km \times 40,000 km and the spacecraft's science orbit is 250 km \times 250 km?

APPENDIX A DICTIONARY OF SYMBOLS

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 try to 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	$\hat{\bar{r}}$
Matrix	bold	\mathbf{r}
Matrix (predicted)	bold and over bar	$\bar{\mathbf{r}}$
Matrix (best estimate)	bold and hat	$\hat{\mathbf{r}}$

The following list of symbols is alphabetical—lowercase, then uppercase; Arabic, then Greek letters.

a	Semimajor axis	e_{\oplus}	Eccentricity of Earth's ellipsoidal approximation (see also flattening f)
\hat{a}	Acceleration vector (often with subscripts)	el	Elevation, measured positively from the local horizontal
a_i	$i = 1 \dots 4$ intermediate range-only	f	f function—Kepler problem
\hat{a}_i	Osculating orbital elements	f	Flattening coefficient for the Earth (see also e_{\oplus})
acc_i	Satellite's initial acceleration magnitude	f	Parameter in Encke's formulation
a_i, a_{iu}	Coefficients of Gaussian initial orbit determination ($i = 1, 3$)	f	Moons phase angle
a_p	3-hourly value of the geomagnetic planetary amplitude	f	Frequency of electromagnetic radiation
a_v	Vehicle acceleration (low-thrust)	f	General mathematical function notation, usually with an argument e.g., $f(x)$
b	Semi-minor axis	f_d	Distance function
b	Observation matrix (observations)	f_d	Doppler shifted frequency
\tilde{b}	Residual matrix (observed-nominal)	f_i	Power Series of F_i
c	Chord length between two vectors	f_r	Reference frequency
c	Constant in energy equation, assumed to be zero for astrodynamics	f_r	Retrograde factor for equinoctial elements (+1 for direct orbits, -1 for retrograde orbits)
c	Half the distance between foci in a conic section	g	Gravity
c	Speed of light	g	g function—Kepler problem
c_i	Coefficients of techniques for determining initial orbits ($i = 1, 2, 3$)	$g_{equator}$	Equatorial value of gravity
c_i	Arbitrary orbital element	$g_{l,m}$	Gaussian geomagnetic coefficient
c_2, c_3	Functions of universal variable	g_{normal}	Normal component of gravity
c_D	Coefficient of drag	g_p	Poincaré orbital element
c_R	Coefficient of reflectivity (used with solar-radiation calculations)	g_{pole}	Polar value of gravity
c_v	Control variable (low-thrust)	g_{th}	Theoretical gravity, sometimes with a descriptor as in (th-84) for WGS-84 values
d	General distance	h	Specific angular momentum
d_i	Intermediate values in Gaussian angles-only orbit determination	h	Step size
e	Eccentricity	h	Intermediate argument in Gaussian initial orbit determination
e_e	Eccentricity of Earth's orbit around the Sun	h	Intermediate parameter in third-body accelerations
e_F	Eccentric for minimum eccentricity orbit (fundamental ellipse)	h_e	Equinoctial element

h_{ellp}	Height above the reference ellipsoid	\hat{n}	Node vector, points to ascending node
$h_{l,m}$	Gaussian geomagnetic coefficient	\bar{n}	<i>Mean</i> mean motion
h_o	Base altitude	n_p	Parabolic mean motion
h_{sys}	Specific angular momentum for n -body problem, perpendicular to the invariable plane	p	Semiparameter
i	Inclination	p_λ	General Precession in Longitude
i	Index (general)	p_e	Equinoctial orbital element
j	Index (general)	p_{mf}	Propellant mass fraction
$k_{day2rep}$	Days to repeat a repeat ground track orbit	p_o	Pressure
k_e	Equinoctial element	$p_{parabola}$	Parabolic semiparameter
k_g	Intermediate value used in calculating theoretical gravity	p_{srp}	Solar-radiation pressure
k_i	Intermediate values of numerical integrators	Δp	Change in pressure
k_{int}	Interceptor satellite's additional revolutions	q	Quaternion
k_l	Love numbers	q_e	Equinoctial element
k_p	Geomagnetic planetary index	r	Position vector's magnitude
$k_{rev2rep}$	Revolutions to repeat a repeat ground track orbit (integer value)	r_a	Radius of apogee
k_{tgt}	Target satellite's additional revolutions	r_d	Relative distance function
l	Index (general)	r_{op}	Parabolic mean radius in Battin's initial orbit determination
l	Gravitational index	r_p	Radius of perigee
l	Intermediate value of Gaussian initial orbit determination	r_p	Radius of primary
l_p	Poincaré orbital element	r_s	Radius of secondary
m	General mass; usually with a subscript added	r_K	Position vector's vertical component
m	Intermediate value of Gaussian initial orbit determination	r_δ	Position vector's equatorial component
m_{sat}	Mass of satellite	\dot{r}	Radial component of velocity (<i>not</i> the velocity magnitude)
m_\oplus	Mass of the Earth	\bar{r}_i	Residual in differential correction; observed minus computed value
\dot{m}	Specific mass flow rate (actual mass flow rate \div initial mass)	r_{site}	Matrix of three site position vectors
n	Mean motion	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	x	Independent value in differential correction
t_{acc}	Accumulated time	x	Component of position vector in P -axis direction
t_{drift}	Time to drift between maneuvers for specialized orbit design	x_p	Polar motion component, measured positively from the North pole along $\lambda = 0^\circ$
t_m	Transfer direction of motion between two position vectors ($+1, \Delta v < 180^\circ$) short way, ($-1, \Delta v > 180^\circ$) long way	x_1, x_2	Intermediate values of Gaussian initial orbit determination
t_p	Parabolic flight time between two positions	y	Sector-to-triangle ratio of Gaussian initial orbit determination
t_s	Switching time (continuous thrust)	y	Component of position in direction of velocity vector (relative motion)
u	Argument of latitude (circular inclined orbits); positive in direction of satellite's motion	y	General value of the function in differential correction and numerical integration
u	Argument of Taylor series of f and g series development	y	Component of position vector in Q -axis direction
u	Truncation error in development of process noise	y_c	Vertical component of a circle
u_x, u_y	Instantaneous position of the North Pole (x / y)	y_e	Vertical component of an ellipse
u_M	Argument of latitude (circular inclined orbits) using mean anomaly	y_h	Vertical component of a hyperbola
v	Magnitude of satellite's velocity	y_{he}	Vertical component of a equatorial hyperbola
v	Uncertainty in dynamics model (process noise)	y_p	Polar motion component, measured positively from the North pole along $\lambda = 90^\circ$ W
v_{acc}	Accumulated velocity	z	Rotation angle for precession
v_{esc}	Escape velocity	z	Cross-plane position (relative motion)
v_L	Velocity of the launch site	 	
v_R	Radial velocity in the RSW system	 	
v_S	Tangential velocity in the RSW 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)	 	
		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
		$A_{t,m}$	Intermediate value for third-body potential function
		A_p	Daily value of the geomagnetic planetary amplitude

A_{SR}	Cross-sectional area—solar radiation	D	Constant in the Hill's relative motion problem
A_{\odot}	Cross-sectional area—solar-radiation pressure	D	Intermediate value in Gibbsian orbit determination
B	Parabolic anomaly	D	Diameter of sensor aperture
B	Intermediate vector in Gibbsian orbit determination	D_{\odot}	Mean elongation of the Sun, positive east
B	Intermediate parameter in third-body potential derivation	D_{ζ}	Mean elongation of the Moon, positive east
B	Summation of Legendre polynomials in third-body acceleration	$D_{1,2}$	Middle position determinant, Gaussian initial orbit determination
\vec{B}	Magnetic field	$D_{3,4}$	Middle velocity determinant, Gaussian initial orbit determination
\vec{B}	Vector constant of integration ($= \mu e$) used in deriving the trajectory equation	E	Eccentric anomaly, positive in direction of satellite's motion
B^*	Ballistic coefficient and perigee density (See also BC)	E	Complete elliptic integral of the second kind
$B_{t,m}$	Intermediate value for third-body potential function	$E()$	Expected value
C	Dot product of middle line-of-sight unit vector and middle site vector	E_i	Intermediate vector in range-only orbit determination
C	Jacobian constant used in the three-body problem	F	Forces
C	Constant in the Hills relative motion problem	F	Mean elongation of the Moon
C	$\hat{L}_2 \cdot \hat{r}_{site2}$ in Gaussian angles-only orbit determination	F, F'	Foci of a conic section
C	Intermediate value in double- r iteration	F	Partial derivative matrix (Kalman filter)
C_{\oplus}	Auxiliary variable for site determination	F	Inclination functions of Kaula gravitational potential
$C_{2,0}$	Unnormalized zonal harmonic of degree 2, gravitational coefficient	F_g	Force of gravity
$C_{l,m}$	Unnormalized gravitational coefficient, degree l , order m	F_i	Intermediate parameter of double- r iteration ($i = 1 \dots 4$)
$\bar{C}_{l,m}^*$	Normalized gravitational coefficient used for geoid undulation calculations, degree l , order m . Even zonal harmonics minus geometric value.	F_i	VOP Perturbing function
D	Determinant value	F_{bnp}	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_{bnp}	Eccentricity function	M	Matrix of slant-range vectors in range-only orbit determination
G_p	Poincaré element	M_{fade}	Matrix of fading-memory values for differential correction
H	Hyperbolic anomaly	M_\oplus	Mean anomaly of the Earth, positive in the direction of satellite's motion
H	Scale height	M_\odot	Mean anomaly of the Moon, positive in the direction of satellite's motion
H	Observation matrix at current time	M_\odot	Mean anomaly of the Sun, positive in the direction of satellite's motion
\mathcal{H}	Hamiltonian	M	Illumination
H_d	Delaunay element	N	Number of observations in statistical orbit determination
H_{lpq}	Third-body formulation of Hansen coefficients	N	Ambiguity
H_p	Poincaré element	N	Intermediate vector in Gibbsian orbit determination
H_{MSL}	Height above mean sea level	N	Newton, equal to 1 kg m/s ² .
I	Identity matrix	N_\oplus	Geoid undulation height
I	Polar moment of inertia	N_ϕ	Radius of curvature in meridian—see also C_\oplus
I_{sp}	Specific impulse (usually in seconds)	P	Parabolic anomaly
J_i	Zonal harmonic coefficient ($i = 2$, etc.)	P	General polynomial, usually with a subscript
J_i	Modified Bessel function of the first kind	P_{conj}	Probability of conjunction
J	Cost Function, estimation	P	Integer parameter in resonance calculations
K	Kalman gain matrix	ρ	Orbital period (Keplerian)
K	Complete elliptic integral of the first kind	$P_{l,0}$	Legendre polynomials, degree l
K_p	Planetary index (magnetic field)	$P_{l,m}$	Associated Legendre functions, degree l , order m
L	Line-of-sight vector, usually with a subscript; L is a matrix of three L	ρ_{res}	Resonance period
L_d	Delaunay element	ρ_K	Anomalistic period (also Keplerian period)
L_e	Equinoctial element	ρ_Ω	Nodal period (also nodal period of Greenwich with subscript G)
L_g	Intermediate variable in Gibbs formulation	\bar{P}	Predicted covariance matrix
L_i	Lagrange point for the three-body problem ($i = 1 \dots 5$)	\hat{P}	Best estimate of covariance
L_p	Poincaré element		
M	Mean anomaly, positive in the direction of satellite's motion		
M	Molecular mass, general mass		
M	Matrix of $(\mathbf{L}^{-1}\mathbf{r}_{site})$ in Gaussian angles-only orbit determination		

Q	Accuracy estimate of double-r iteration	T_{TAI}	Julian centuries of atomic time, TAI
Q	Integer parameter in resonance calculations	T_{UT1}	Julian centuries of universal time, UT1
Q	Power density matrix resulting from process noise	$U_{2\text{-body}}$	Gravitational potential (two-body)
Q	Matrix used in Lambert's time formulation	U	Gravitational potential (aspherical), includes two-body portion
R	Disturbing potential ($U - U_{2\text{-body}}$)	V	Magnetic field potential
R	Measurement noise matrix, also the inverse of W matrix	W	Weighting matrix
R	Universal gas constant	\bar{X}	Predicted state vector; sometimes just position and velocity vector, sometimes with solve-for parameters
R	Ratio of orbit's final radius over its initial radius	\hat{X}	Best estimate of the state vector
R_{\oplus}	Radius of Earth	Z_{ij}	Cross-product of vectors i and j
R^*	Ratio of orbit's intermediate radius over its final radius (bi-elliptic transfers)	α	General angle
$[R]$	Rotation matrix	α	Reciprocal of semimajor axis (Kepler's problem in universal variables)
S	Intermediate value in double-r iteration	α	Right ascension, measured positively to the east
S	Alternate matrix for square root formulations	α_{xx}	Coefficients for <i>non-summed</i> numerical integration methods (Adams-Basforth AB , Adams-Moulton AM , Störmer S , Cowell C)
S_{lm}	Unnormalized gravitational coefficient	α_{cop}	Coplanar angle measuring amount of separation, given three vectors
S_{lmpq}	Kaula gravitational function	α_{DMS}	Degree-minute-second angular value in radians or degrees
$\hat{S}_n^I, \hat{S}_n^{II}$	First and second summed back differences	α_L	Lead angle, positive in the direction of the target satellite's motion
S_{\oplus}	Auxiliary function for site determination	α_e, α_h	Intermediate angles (elliptical and hyperbolic) used in Lambert problem
T	Time of perigee passage	α_t	Topocentric right ascension, measured positively to the east
T	Lambert's time function	α_{12}, α_{23}	Separation angle between two vectors
T	Absolute temperature ($^{\circ}\text{K}$)		
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		

β	Azimuth, measured positively east from north	lations for the IAU-76/FK5 and the IAU 2000 resolutions, subscripts are included to indicate either the 1980 or 2000 theory.
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_{SUN}	Shadow (beta) angle	
β_{xx}	Coefficients for <i>summed ordinate</i> numerical integration methods (Adams-Basforth <i>AB</i> , Adams-Moulton <i>AM</i> , Störmer <i>S</i> , Cowell <i>C</i>)	
x	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_{SUN}	Angular separation of the Earth from the Vernal Equinox, measured along the ecliptic	
δ_t	Topocentric declination, measured positively north from the local equator	
$\delta\bar{x}$	Predicted state correction	
$\hat{\delta}\hat{x}$	Best estimate for state correction	
ϵ	Obliquity of the ecliptic (true)	
ϵ	General small parameter	
ϵ	Small parameter in Encke's formulation	
$\bar{\epsilon}$	Mean obliquity of the ecliptic	
ϵ_{tmpq}	Phase lag	
ϵ_Θ	Resonance parameter	
$\Delta\epsilon$	Change in obliquity of the ecliptic, used in nutation calculations. Because there are different formu-	
ϕ	General latitude, specified with subscripts	
ϕ_{as}	Astronomical latitude, measured positively north from the equator	
ϕ_c	control angle, positive in the angular momentum 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_{fpav}	Compliment of flight-path-angle, measured positively from the local vertical 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 solar radiation	
ϕ_{rd}	Reduced latitude, measured positively north from the equator	
ϕ_{ref}	Reflected solar radiation angle	
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	
η_{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	μ_{xy}	Correlation coefficients between x and y
φ	Turning angle for hyperbolic orbits. φ is the supplementary angle.	μ_{\oplus}	Gravitational parameter for the Earth
κ_{\odot}	Gaussian constant of gravitation for the Sun	μ_{\odot}	Gravitational parameter for the Sun
λ	Longitude, measured positively east from the Greenwich meridian	ν	True anomaly
λ	Wavelength of electromagnetic radiation	θ	General angle (angular separation angles, etc.)
λ	Temporary variable for analyzing parabolic time of flight	θ_p	angle of satellite travel in low-thrust transfers
$\Delta\lambda$	Change in terrestrial longitude for a groundtrack, relative to the starting location	θ_{AST}	Apparent sidereal time, measured positively to the east from the Greenwich meridian along true equator (also AST or GAST). Because there are different formulations for the IAU-76/FK5 and the IAU 2000 resolutions, subscripts are included to indicate either the 1982 or 2000 theory.
$\lambda_{ecliptic}$	Ecliptic longitude	θ_{BW}	Beamwidth
λ_{true}	True longitude at epoch (circular equatorial orbits) measured in one plane; positive	θ_{ERA}	Earth Rotation Angle
λ_u	Longitudinal component of satellite position, measured positively from the ascending node to the longitude of a satellite's subpoint	θ_{GMST}	Greenwich mean sidereal time, measured positively to the east from the Greenwich meridian (also GMST). Because there are different formulations for the IAU-76/FK5 and the IAU 2010 resolutions, subscripts are included to indicate either the 1982 or 2000 theory.
$\lambda_{vR}, \lambda_{vS}$	Lagrange multipliers in orbit raising (λ dynamics)	θ_{LST}	Local mean sidereal time, measured positively to the east from the site (also LST)
λ_M	Mean longitude at epoch (circular equatorial orbits) measured in two planes ($\Omega + \omega + M$)	ρ	Range magnitude (site to satellite)
$\lambda_{M\odot}$	Mean longitude for the Sun	ρ	Atmospheric density
$\lambda_{M\odot}$	Mean longitude for the Moon	ρ	Range vector (site to satellite)
$\Delta\lambda_{rev}$	Groundtrack shift parameter per revolution	ρ_n	Density of the atmosphere at night
$\Delta\lambda_{pass}$	Groundtrack shift parameter per pass	σ	Standard deviation, usually with a subscript for estimation theories
$\Delta\lambda_{rev+}, \Delta\lambda_{rev-}$	Maximum and minimum values of groundtrack shift parameter	σ	Rate parameter for mean anomaly
λ_{\odot}	Celestial longitude for the Sun	τ	Scaling parameter in the <i>SIGHT</i> problem
μ	Gravitational parameter, usually with a subscript		
μ^*	Mass ratio for the restricted three-body problem		

τ_i	Delta time between a given time (t) and a middle time ($i = 1, 3$)	ζ	Precession angle
τ_{phase}	Time in a phasing orbit	Φ	Error State transition matrix
τ_{min}	Location of minimum distance in <i>SIGHT</i> algorithm	Φ_S	State transition matrix
τ_{wait}	Wait time for a rendezvous	ϑ	Angle of Hyperbolic asymptotes
τ_{trans}	Transfer time of flight	ϑ	Angle between normal component of velocity vectors in maneuver calculations (positive in direction of the target satellite's motion)
τ_{HMS}	Time of hour-minute-seconds in radians or degrees	ϑ	Phase angle (rendezvous calculations) from target satellite to interceptor satellite. Positive sense is in direction of the target satellite's motion.
τ_{UT}	Universal time in seconds	$\Pi_{t,m}$	Conversion term for normalizing gravitational coefficients
$\Delta\tau_{eclipse-circ}$	Eclipse duration for a circular orbit	$\Pi_{t,mSchmidt}$	Conversion term for normalizing geomagnetic coefficients with Schmidt's method
ω	Argument of periapsis; positive in the direction of a satellite's motion	Θ_{lmpq}	Kaula's gravitational argument
ω_{int}	Angular velocity of a satellite—usually specifies interceptor or target	Θ_{3-body}	Third-body gravitational argument
ω_r	Mean rotation of the Earth with respect to a precessing equinox	Δ	Total range angle
ω_S	Rotation of synodic frame with respect to barycentric frame	Δ_{FOV}	Total sensor ground range visible to a sensor
ω_{\oplus}	Rotational velocity of the Earth	Δ_{FOV}	Sensor ground range visible to a sensor
ω	Longitude of periapsis, measured in two planes, in the direction of a satellite's motion ($\Omega + \omega$)	Θ	Rotation angle for precession
$\tilde{\omega}_{true}$	True longitude of periapsis, measured in one plane, in the direction of a satellite's motion	Θ	Angle between velocity vectors in maneuver calculations
$\tilde{\omega}_M$	Mean longitude of periapsis, using the mean anomaly, measured in one plane in the direction of satellite motion	Σ_i	Summation of previous function values for numerical integrators
ξ	Specific mechanical energy (see also SME)	ς	Angular distance between the Sun and satellite vectors (solar radiation-pressure and <i>PREDICT</i> problem). Specified with a type—civil, nautical, or astronomical—when used as a sunrise sunset parameter
ξ_i	Arbitrary vector component ($i = 1, 2, 3$)	Ω	Right ascension of the ascending node; positive east from the I axis. It is also called the longitude of the ascending node if Earth-fixed coordinates are used.
ξ_{sys}	Specific mechanical energy for n -body problem		
$\xi(x)$	Continued fraction in Battin's initial orbit determination		
ψ	Universal variable argument		

- Ω_{c} Right ascension of the ascending node of the Moon; positive from the I axis. It is also called the longitude of the ascending node of the Moon if Earth-fixed coordinates are used.
- Ψ Precession angle
- $\Delta\Psi$ Change in longitude (nutation). Because there are different formulations for the IAU-76/FK5 and the IAU 2000 resolutions, subscripts are included to indicate either the 1980 or 2000 theory.
- γ Vernal equinox
- φ Parallax angle, sometimes with a subscript for geocentric, horizontal, or heliocentric
- $\nabla_i(\cdot)$ 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

APPENDIX B MODELING THE ATMOSPHERE

B.1 Jacchia-Roberts Atmosphere

B.2 Russian GOST 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 Russian 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} .

$$T_{unc} = T_c \left\{ 1 + 0.3 \left[\sin^2(\theta) + (\cos^2(\eta) - \sin^2(\theta)) \cos^3\left(\frac{\tau}{2}\right) \right] \right\}$$

Now, we 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 ($-180^\circ < \tau < 180^\circ$):

$$\eta = \frac{|\phi_{gd} - \delta_\odot|}{2} \quad \theta = \frac{|\phi_{gd} + \delta_\odot|}{2}$$

$$\tau = LHA_\odot - 37.0^\circ + 6.0^\circ \sin(LHA_\odot + 43.0^\circ)$$

$$\begin{aligned}
LHA_{\odot} &= \frac{180.0^{\circ}}{\pi} \left\{ \frac{r_x r_J - r_y r_I}{|r_x r_J - r_y r_I|} \cos^{-1} \left(\frac{r_x r_I + r_y r_J}{\sqrt{r_x^2 + r_y^2} \sqrt{r_I^2 + r_J^2}} \right) \right\} \\
\phi_{gd} &= \tan^{-1} \left\{ \frac{1}{(1-f)^2} \left[\frac{r_K}{\sqrt{r_I^2 + r_J^2}} \right] \right\}
\end{aligned} \tag{B-1}$$

Notice we don't iterate to determine latitude. We use Eq. (3-11). Both the satellite vector, $\hat{r}(r_I r_J r_K)$, and the Sun's unit vector, $\hat{r}_{\odot}(r_x r_y r_z)$, are in true-of-date (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

$$T_x = 371.6678^{\circ} + 0.051\,880\,6 T_{corr} - 294.3505^{\circ} \exp[-0.002\,162\,22 T_{corr}] \tag{B-2}$$

Using the base value temperature $T_o = 183^{\circ}$ K, Jacchia defines the empirical temperature function for altitudes below 125 km as

$$\begin{aligned}
T(h_{ellp})_{0-125} &= T_x + \frac{T_x - T_o}{35^4} \sum_{n=0}^4 C_n h_{ellp}^n \\
C_0 &= -89,284,375.0 & C_1 &= 3,542,400.0 & C_2 &= -52,687.5 \\
C_3 &= 340.5 & C_4 &= -0.8
\end{aligned} \tag{B-3}$$

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).

$$\begin{aligned} T(h_{ellp})_{125-} = & T_x + \frac{2}{\pi}(T_{corr} - T_x) \\ & \times \tan^{-1} \left\{ 0.95\pi \left(\frac{T_x - T_o}{T_{corr} - T_x} \right) \left(\frac{h_{ellp} - 125}{35} \right) (1 + 4.5 \times 10^{-6}(h_{ellp} - 125)^{2.5}) \right\} \end{aligned} \quad (\text{B-4})$$

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 Evaluating Density

Jacchia used a standard exponential relation to evaluate density. From Eq. (8-33),

$$\rho_{std} = \rho_o \exp\left[-\frac{h_{ellp} - h_o}{H}\right] \quad (B-6)$$

For altitudes below 200 km, we include the *geomagnetic effect* on density:

$$(\Delta \log_{10} \rho)_G = 0.012 K_p + 1.2 \times 10^{-5} \exp[K_p] \quad (B-7)$$

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.

$$(\Delta \log_{10} \rho)_{LT} = 0.014(h_{ellp} - 90) \sin(2\pi T_{1958} + 1.72) \sin(\phi_{gd}) \\ |\sin(\phi_{gd})| \exp[-0.0013(h_{ellp} - 90)^2] \quad (B-8)$$

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

$$(\Delta \log_{10} \rho)_{SA} = (5.876 \times 10^{-7} h_{ellp}^{2.331} + 0.06328) \exp[-0.002868 h_{ellp}] \\ \times \{0.02835 + [0.3817 + 0.17829 \sin(2\pi \tau_{SA} + 4.137)] \sin(4\pi \tau_{SA} + 4.259)\} \quad (B-9)$$

Be careful of the units because the trigonometric terms use radians.

We then apply the correction terms of Eq. (B-7), Eq. (B-8), and Eq. (B-9) to the standard density:

$$(\Delta \log_{10} \rho)_{corr} = (\Delta \log_{10} \rho)_G + (\Delta \log_{10} \rho)_{SA} + (\Delta \log_{10} \rho)_{LT}$$

giving the final corrected density

$$\rho(h_{ellp}) = \rho_{std}(h_{ellp}) 10^{(\Delta \log_{10} \rho)_{corr}} \quad (B-10)$$

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 mainly for calculating reentry and are usually handled by standard atmospheres (See Sec. 8.6.2). The end result is that we use Eq. (B-11), Eq. (B-12), or Eq. (B-13) for the standard density in Eq. (B-10).

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.

$$\rho_{std_{90-100}}(h_{ellp}) = \frac{\rho_{90} T_{90}}{M_{90}} \frac{M(h_{ellp})}{T(h_{ellp})_{125-}} F_1^k \text{EXP}[kF_2] \quad (\text{B-11})$$

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

$$k = -\left(\frac{g_{SL}}{R(T_x - T_o)}\right)$$

using the gravity at sea level ($g_{SL} = 9.80\ 665$ m/s²), and the universal gas constant ($R = 8.314\ 32$ Joules / K·mol).

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 define an auxiliary parameter, f , and solve for F_1 and F_2 in Eq. (B-11) using quadratic expressions and the polar radius ($R_{pole} = 6356.766$ km):

$$\begin{aligned}
f &= \frac{35^4 R_{pole}^2}{C_4} \\
F_1 &= \left(\frac{h_{ellp} + R_{pole}}{90 + R_{pole}} \right)^{P_1} \left(\frac{h_{ellp} - x_{r1}}{90 - x_{r1}} \right)^{P_2} \left(\frac{h_{ellp} - x_{r2}}{90 - x_{r2}} \right)^{P_3} \\
&\quad \times \left(\frac{h_{ellp}^2 - 2x_{r3}h_{ellp} + x_{r3}^2 + x_{i3}^2}{8100 - 180x_{r3} + x_{r3}^2 + x_{i3}^2} \right)^{P_4} \\
F_2 &= (h_{ellp} - 90) \left[fA_6 + \frac{p_5}{(h_{ellp} + R_{pole})(90 + R_{pole})} \right] \\
&\quad + \frac{p_6}{x_{i3}} \text{TAN}^{-1} \left[\frac{x_{i3}(h_{ellp} - 90)}{x_{i3}^2 + (h_{ellp} - x_{r3})(90 - x_{r3})} \right]
\end{aligned}$$

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

$$\begin{aligned}
f(h_{ellp}) &= \sum_{n=0}^4 C_n^* h_{ellp}^n & C_o^* &= \frac{35^4 T_x}{C_4(T_x - T_o)} + \frac{C_0}{C_4} \\
C_n^* &= \frac{C_n}{C_4} \quad (1 \leq n \leq 4)
\end{aligned}$$

for values of C_n given by Eq. (B-3). We need other quantities to complete the next relations:

$$\begin{aligned}
X^* &= -2x_{r1}x_{r2}R_{pole}(R_{pole}^2 + 2x_{r3}R_{pole} + x_{r3}^2 + x_{i3}^2) \\
V &= (R_{pole} + x_{r1})(R_{pole} + x_{r2})(R_{pole}^2 + 2x_{r3}R_{pole} + x_{r3}^2 + x_{i3}^2) \\
U(x_{ri}) &= (x_{ri} + R_{pole})^2(x_{ri}^2 - 2x_{r3}x_{ri} + x_{r3}^2 + x_{i3}^2)(x_{r1} - x_{r2}) \\
W(x_{ri}) &= x_{r1}x_{r2}R_{pole}(R_{pole} + x_{ri}) \left(R_{pole} + \frac{x_{r3}^2 + x_{i3}^2}{x_{ri}} \right)
\end{aligned}$$

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

$$p_2 = \frac{S(x_{r1})}{U(x_{r1})} \quad p_3 = \frac{-S(x_{r2})}{U(x_{r2})} \quad p_5 = \frac{S(-R_{pole})}{V}$$

$$\begin{aligned}
p_4 &= \frac{1}{X^*} \{ B_0 - x_{r1}x_{r2}R_{pole}^2 [B_4 + B_5(2x_{r3} + x_{r1} + x_{r2} - R_{pole})] \\
&\quad + W(x_{r1})p_2 - x_{r1}x_{r2}B_5R_{pole}(x_{r3}^2 + x_{i3}^2) + W(x_{r2})p_3 \\
&\quad + x_{r1}x_{r2}(R_{pole}^2 - x_{r3}^2 - x_{i3}^2)p_5 \} \\
p_6 &= B_4 + B_5(2x_{r3} + x_{r1} + x_{r2} - R_{pole}) - p_5 - 2(x_{r3} + R_{pole})p_4 \\
&\quad - (x_{r2} + R_{pole})p_3 - (x_{r1} + R_{pole})p_2 \\
p_1 &= B_5 - 2p_4 - p_3 - p_2
\end{aligned}$$

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. Intermediate Coefficients. These coefficients help determine the functions that lead to the density between 100 and 125 km.

<i>n</i>	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.

$$\rho_{std}_{100-125}(h_{ellp}) = \sum_{i=1}^5 \rho(100) \frac{M_i}{M_s} \mu_i \left[\frac{T(100)}{T(h_{ellp})} \right]^{1+\alpha_i} F_3^{M_i k f} \text{EXP}[M_i k f F_4] \quad (\text{B-12})$$

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 , a_i , M_i).

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	

We don't consider hydrogen significant in determining density below 125 km, thus the summation from $i = 1$ to 5 in Eq. (B-12). We could evaluate the density at 100 km [$\rho(100)$] by Eq. (B-11), 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 $\rho(100)$.

$$\rho(100) = M_s \sum_{n=0}^6 \xi_n T_{corr}^n$$

$$\xi_0 = 0.198\ 554\ 9 \times 10^{-10} \quad \xi_1 = -0.183\ 349 \times 10^{-14}$$

$$\xi_2 = 0.171\ 173\ 5 \times 10^{-17} \quad \xi_3 = -0.102\ 147\ 4 \times 10^{-20}$$

$$\xi_4 = 0.372\ 789\ 4 \times 10^{-24} \quad \xi_5 = -0.773\ 411\ 0 \times 10^{-28}$$

$$\xi_6 = 0.702\ 694\ 2 \times 10^{-32}$$

Now, find the temperature at 100 km, $T(100)$, using Eq. (B-3):

$$T(100) = T_x + 35^{-4} \sum_{n=0}^4 C_n (100)^n (T_x - T_o) = T_x - 0.945\ 855\ 89(T_x - T_o)$$

The parameters, k and C_n , in Eq. (B-12) are the same as in the 90–100 km case. The functions F_3 , F_4 are slightly different.

$$F_3 = \left(\frac{h_{ellp} + R_{pole}}{R_{pole} + 100} \right)^{q_1} \left(\frac{h_{ellp} - x_{r1}}{100 - x_{r1}} \right)^{q_2} \left(\frac{h_{ellp} - x_{r2}}{100 - x_{r2}} \right)^{q_3}$$

$$\times \left(\frac{h_{ellp}^2 - 2x_{r3}h_{ellp} + x_{r3}^2 + x_{i3}^2}{10000 - 200x_{r3} + x_{r3}^2 + x_{i3}^2} \right)^{q_4}$$

$$F_4 = \left[\frac{q_5(h_{ellp} - 100)}{(h_{ellp} + R_{pole})(R_{pole} + 100)} \right] \\ + \frac{q_6 \tan^{-1} \left[\frac{x_{i3}(h_{ellp} - 100)}{x_{i3}^2 + (h_{ellp} - x_{r3})(100 - x_{r3})} \right]}{x_{i3}}$$

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. We include $\rho_6(h_{ellp})$ for altitudes over 500 km.

$$\rho_{std\ 125-}(h_{ellp}) = \sum_{i=1}^5 \rho_i(125) \left(\frac{T_x}{T(h_{ellp})} \right)^{1+a_i+\gamma_i} \left(\frac{T_{corr} - T(h_{ellp})}{T_{corr} - T_x} \right)^{\gamma_i} \quad (B-13)$$

$$\gamma_i = \frac{M_i g_o R_{pole}^2}{R \ell T_{corr}} \left(\frac{T_{corr} - T_x}{T_x - T_o} \right) \left(\frac{35}{6481.766} \right)$$

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).

The correction for density because of *seasonal variations of helium by latitude* is

$$(\Delta \log_{10} \rho)_{\text{He}} = 0.65 \left| \frac{\delta_{\odot}}{\epsilon} \right| \left[\sin^3 \left(\frac{\pi}{4} - \frac{\phi_{gd} \delta_{\odot}}{2 |\delta_{\odot}|} \right) - 0.35355 \right]$$

with the obliquity of the ecliptic, ϵ , and the declination of the Sun from before. This final correction for helium is

$$[\rho_{\text{He}}(h_{\text{ellp}})]_{\text{corrected}} = \rho_{\text{He}}(h_{\text{ellp}}) 10^{(\Delta \log_{10} \rho)_{\text{He}}}$$

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-13), 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.

Although we can explicitly determine the constituent densities, ρ_i (125) from Eq. (B-12), Draper Laboratory uses a polynomial curve fit approximation in the Goddard Trajectory Determination System (Long et al. 1989:4-49).

$$\rho_i(125) = M_i 10^{\sum_{j=0}^6 \delta_{ij} T_{\text{corr}}^j}$$

Here, δ_{ij} are curve-fit coefficients in Table B-4, and M_i are from Table B-3.

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

B.2 Russian GOST Atmosphere

This model is constructed empirically from observations of the orbital motion of Russian Cosmos satellites. These models have been adopted as official standards in Russia. They include the dependence of the density on solar flux and geomagnetic activity as well as the diurnal and semiannual density variations. Usually called the GOST model, it is valid for satellites in the altitude range of 120–1500 km. Consult Amelina et al. (1996), Volkov (1978, 1984), and Voiskovskii et al. (1973) for more information. Yurasov (1999, 2005) graciously provided this update.

The Russian formula models the actual atmospheric density as a product of several factors, where each factor corresponds to a particular density variation,

$$\rho = \rho_n k_0 k_1 k_2 k_3 k_4 \quad (\text{B-14})$$

where ρ_n is the *night-time density profile* assumed to be exponentially decreasing with respect to altitude, according to

$$\rho_n = 9.80665 \exp[a_1 - a_2 \sqrt{h_{ellp} - a_3}] \text{ kg/m}^3 \quad (\text{B-15})$$

The coefficients in this expression were determined empirically and are listed in Tables B-6, to B-9. The altitude of the satellite (in km) is shown as h_{ellp} .

The k factors represent various attributes of the atmospheric density. k_0 relates the daily $F_{10.7}$ solar activity (F_0 in the Tables) to a weighted average, F_{81} (described shortly). The k_1 factor represents the daily effects of the atmospheric density distribution. The term ‘daily’ is used because φ varies 360° in ECEF coordinates. The k_2 factor accounts for the semiannual effects, and k_3 takes into account the changing atmospheric density with a deviation of the daily $F_{10.7}$ from F_{81} . The factor k_4 represents the dependence on geomagnetic activity index (k_p) (Voiskovskii, et al. 1973). The k factors are given below.

$$\begin{aligned} k_0 &= 1 + (l_1 + l_2 h_{ellp} + l_3 h_{ellp}^2)(F_{81} - F_0) \\ k_1 &= 1 + (c_1 + c_2 h_{ellp} + c_3 h_{ellp}^2 + c_4 h_{ellp}^3) \cos^{(n_o + n_1 h_{ellp})} \left(\frac{\varphi}{2} \right) \\ k_2 &= 1 + (d_1 + d_2 h_{ellp} + d_3 h_{ellp}^2) A(d) \\ k_3 &= 1 + \frac{(b_1 + b_2 h_{ellp} + b_3 h_{ellp}^2)(F_{10.7} - F_{81})}{F_{10.7}} \\ k_4 &= 1 + (e_1 + e_2 h_{ellp} + e_3 h_{ellp}^2 + e_4 h_{ellp}^3)(e_5 + e_6 \bar{k}_p + e_7 \bar{k}_p^2) \end{aligned}$$

Several variables are discussed below, and the remaining coefficients are listed in Tables B-6 to Table B-8. Values are chosen by the F_{81} value that is closest to the F_0 values that are listed in the tables. $A(d)$ (Table B-9) describes the semiannual effect on atmospheric density, and d is the number of days from the beginning of the year.

Time lags are used for both solar flux ($t - 1.7$ days) and geomagnetic indices ($t - 0.6$ days when using daily k_p values, and $t - 0.25$ days when using 3-hourly values of k_p). In addition, all indices are interpolated. Solar flux is assumed to be at 20^h UTC (17^h UTC before May 31, 1991). The daily k_p values are assumed to be at 12^h UTC, and the 3-hourly values are at the middle of each interval.

F_{81} is calculated by summing the previous 81 days of $F_{10.7}$ (daily values), and multiplying each by a weighting factor (w_i)

$$F_{81} = \frac{\sum_{i=1}^{81} F_{10.7,i} w_i}{\sum_{i=1}^{81} w_i} \quad w_i = 0.5 + \frac{0.5(i-1)}{80}$$

We divide the result by the summed weighting values to find F_{81} . Note that $i = 1, 2, \dots, 81$ with $i = 81$ (day of interest).

A useful angle is the geocentric angle between the directions of the point where the density is to be found, and the center of the atmospheric bulge (φ). A similar angle (φ_B) is the angular separation of the atmospheric bulge, and the direction to the Sun. Note that the Sun's right ascension and declination are α_\odot and δ_\odot .

$$\cos(\varphi) = \frac{1}{r} \left(r_K \sin(\delta_\odot) + \cos(\delta_\odot) \left\{ r_I \cos(\gamma_1) + r_J \sin(\gamma_1) \right\} \right)$$

$$\gamma_1 = \alpha_\odot + \varphi_B$$

For geomagnetic indices, \bar{k}_p is the average daily k_p -index. It is calculated using either the daily or 3-hourly k_p values. In each case, the time lags are applied to the current time and the k_p values are interpolated at that time.

The GOST 25645.115-84 model is valid for satellites at altitudes of 120-1500 km. Below 120 km, the atmosphere is considered as static. The atmospheric density calculation at altitudes below than 120 km is as follows, using coefficients in Table B-5.

$$\rho_n = A_o \exp[k_1(h_{ellp} - k_3) + k_2(h_{ellp} - k_3)^2] \text{ kg/m}^3 \quad (\text{B-16})$$

TABLE B-5. Russian Density Model Parameters. Coefficients are listed for altitudes below 120 km.

h_{ellp} , km	A_0 , kg/m ³	k_1 , km ⁻¹	k_2 , km ⁻²	k_3 , km
$0 \leq h_{ellp} < 20$	1.2280	-0.90764×10^{-1}	-0.20452×10^{-2}	0
$20 \leq h_{ellp} < 60$	0.9013×10^{-1}	-0.16739	0.62669×10^{-3}	20
$60 \leq h_{ellp} < 100$	0.3104×10^{-3}	-0.13700	-0.78653×10^{-3}	60
$100 \leq h_{ellp} < 120$	0.3660×10^{-6}	-0.18533	0.15397×10^{-2}	100

TABLE B-6. Russian Density Model Parameters. Coefficients are listed for fixed level of solar flux values (F_0) that are valid for altitudes between 120 and 180 km. The first set of values for e_5 , e_6 , and e_7 are for use with daily K_p values. The second set is for use with 3-hourly K_p values.

	$F_0 = 75$	$F_0 = 100$	$F_0 = 125$	$F_0 = 150$	$F_0 = 175$	$F_0 = 200$	$F_0 = 250$
a_1	-1.829910E+01	-1.819080E+01	-1.852090E+01	-1.865220E+01	-1.865860E+01	-1.864950E+01	-1.870740E+01
a_2	7.009000E-01	7.000000E-01	6.419000E-01	6.124000E-01	6.038000E-01	5.974000E-01	5.772000E-01
a_3	1.153430E+02	1.146386E+02	1.159569E+02	1.164154E+02	1.163531E+02	1.162144E+02	1.163395E+02
d_1	-5.101900E+00						
d_2	6.258100E-02	6.258000E-02	6.258000E-02	6.258000E-02	6.258000E-02	6.258000E-02	6.258000E-02
d_3	-1.672100E-04	-1.672200E-04	-1.672200E-04	-1.672200E-04	-1.672200E-04	-1.672200E-04	-1.672200E-04
b_1	-6.828000E-01	-7.804000E-01	-8.220000E-01	-7.376000E-01	-3.150000E-01	-5.161000E-01	-2.531000E-01
b_2	5.576200E-03	7.173000E-03	8.330000E-03	7.597000E-03	2.325000E-03	5.341000E-03	1.929000E-03
b_3	9.523800E-07	-5.578000E-06	-1.233000E-05	-1.209000E-05	2.500000E-06	-8.672000E-06	1.495000E-06
c_1	-4.384000E+00	-4.384000E+00	9.776000E-01	9.776000E-01	-5.632000E-01	-5.632000E-01	4.842000E-01
c_2	8.063000E-02	8.063000E-02	-2.570000E-02	-2.570000E-02	5.743000E-03	5.743000E-03	-1.604000E-02
c_3	-4.925000E-04	-4.925000E-04	2.027000E-04	2.027000E-04	-9.250000E-06	-9.250000E-06	1.405000E-04
c_4	1.042000E-06	1.042000E-06	-4.708000E-07	-4.708000E-07	4.167000E-09	4.167000E-09	-3.375000E-07
n_1	1.500000E+00						
n_2	6.000000E-03						
φ_B	5.411000E-01	5.515000E-01	5.585000E-01	5.585000E-01	5.585000E-01	5.585000E-01	5.585000E-01
e_1	-7.238000E+00	-6.683000E+00	-5.352000E+00	-4.799000E+00	-4.903000E+00	-5.115000E+00	-3.137000E+00
e_2	1.203000E-01	1.101400E-01	8.615000E-02	7.779200E-02	8.270800E-02	8.507500E-02	4.774200E-02
e_3	-6.450000E-04	-5.837500E-04	-4.437500E-04	-4.075000E-04	-4.587500E-04	-4.587500E-04	-2.275000E-04
e_4	1.208000E-06	1.083000E-06	8.125000E-07	7.708300E-07	9.167000E-07	8.750000E-07	3.958300E-07
e_5	-1.200000E-01	-1.200000E-01	-1.000000E-01	-1.000000E-01	-1.200000E-01	-1.100000E-01	-9.000000E-02
e_6	5.000000E-03	2.500000E-02	2.083000E-02	2.750000E-02	4.116000E-02	3.810000E-02	3.118000E-02
e_7	1.500000E-02	7.500000E-03	6.251000E-03	3.800000E-03	1.433000E-03	1.178000E-03	9.662000E-04
e_8	-1.200000E-01	-1.000000E-01	-1.000000E-01	-9.000000E-02	-8.000000E-02	-7.000000E-02	-5.000000E-02
e_9	2.500000E-02	2.625000E-02	2.617000E-02	2.542000E-02	2.333000E-02	2.125000E-02	1.375000E-02
e_{10}	7.500000E-03	5.625000E-03	4.250000E-03	3.125000E-03	2.500000E-03	1.875000E-03	1.875000E-03
l_1	-1.197500E-02	-9.900000E-03	-7.680000E-03	-5.600000E-03	-4.963000E-03	-4.110000E-03	-3.030000E-03
l_2	9.983000E-05	8.212000E-05	6.362000E-05	4.667000E-05	4.136000E-05	3.463000E-05	2.532000E-05
l_3	0.000000E+00	3.125000E-09	3.125000E-09	0.000000E+00	0.000000E+00	-3.125000E-09	-5.556000E-10

The advantage of this analytical model is its simplicity and feasibility for computer implementation. The resulting density computation times are one-two times less than using other time-varying models. At the same time the accuracy of this model is comparable with more sophisticated semi-empirical models (e.g., CIRA). The RMS of relative errors in atmospheric density calculations using the GOST model at altitudes of 200-500 km are about 10% during low solar activity, and about 30% during geomagnetic storms.

Note that the current GOST model is R25645-166-2004, and it is under control of the Federal Agency of the Russian Federation on Technical Regulatory and Metrology.*

TABLE B-7. Russian Density Model Parameters. Coefficients are listed for fixed level of solar flux values (F_0) that are valid for altitudes between 180 and 600 km. The first set of values for e_5 , e_6 , and e_7 are for use with daily k_p values. The second set is for use with 3-hourly k_p values.

	$F_0 = 75$	$F_0 = 100$	$F_0 = 125$	$F_0 = 150$	$F_0 = 175$	$F_0 = 200$	$F_0 = 250$
a_1	-1.556050E+01	-1.564080E+01	-1.522290E+01	-1.697520E+01	-1.730450E+01	-1.826600E+01	-1.927820E+01
a_2	8.248000E-01	7.754000E-01	7.569000E-01	6.736000E-01	6.382000E-01	5.797000E-01	5.118000E-01
a_3	7.691320E+01	6.791620E+01	5.581650E+01	8.544400E+01	8.195960E+01	1.009417E+02	1.165792E+02
d_1	-1.721000E-01						
d_2	5.756000E-03						
d_3	-3.635000E-06						
b_1	-8.607000E-01	-7.540000E-01	-5.700000E-01	-4.760000E-01	-2.920000E-01	-3.113000E-01	-3.307000E-01
b_2	7.861000E-03	6.850000E-03	5.250000E-03	4.400000E-03	2.800000E-03	2.839000E-03	2.878000E-03
b_3	-5.711000E-06	-4.600000E-06	-3.000000E-06	-2.400000E-06	-8.000000E-07	-1.089000E-06	-1.378000E-06
c_1	1.279100E+00	1.279100E+00	1.290300E+00	1.290300E+00	2.057000E-01	2.057000E-01	1.499000E-03
c_2	-1.576000E-02	-1.576000E-02	-1.547000E-02	-1.547000E-02	-2.912000E-03	-2.911000E-03	-2.399000E-04
c_3	6.499000E-05	6.499000E-05	5.964000E-05	5.964000E-05	1.739000E-05	1.739000E-05	7.006000E-06
c_4	-5.145000E-08	-5.145000E-08	-4.503000E-08	-4.503000E-08	-8.565000E-09	-8.565000E-09	-5.999000E-10
n_1	1.500000E+00						
n_2	6.000000E-03						
φ_B	5.411000E-01	5.515000E-01	5.585000E-01	5.585000E-01	5.585000E-01	5.585000E-01	5.585000E-01
e_1	-2.152000E-01	-2.162000E-01	-1.486000E-01	-1.495000E-01	-8.190000E-02	-8.286000E-02	-2.048000E-01
e_2	4.167000E-03	4.086000E-03	3.263000E-03	3.182000E-03	2.358000E-03	2.278000E-03	3.596000E-03
e_3	1.587000E-06	1.270000E-06	3.143000E-06	2.825000E-06	4.698000E-06	4.381000E-06	-1.587000E-06
e_4	-1.651000E-09	-1.587000E-09	-3.429000E-09	-3.365000E-09	-5.206000E-09	-5.143000E-09	3.175000E-10
e_5	-1.200000E-01	-1.200000E-01	-1.000000E-01	-1.000000E-01	-1.300000E-01	-1.100000E-01	-9.000000E-02
e_6	5.000000E-03	2.500000E-02	2.083000E-02	2.750000E-02	4.389000E-02	3.810000E-02	3.117000E-02
e_7	1.500000E-02	7.500000E-03	6.250000E-03	3.750000E-03	1.821000E-03	1.178000E-03	9.662000E-04
e_8	-1.200000E-01	-1.000000E-01	-1.000000E-01	-9.000000E-02	-8.000000E-02	-7.000000E-02	-5.000000E-02
e_9	2.500000E-02	2.625000E-02	2.617000E-02	2.542000E-02	2.333000E-02	2.125000E-02	1.375000E-02
e_{10}	7.500000E-03	5.625000E-03	4.250000E-03	3.125000E-03	2.500000E-03	1.875000E-03	1.875000E-03
l_1	-1.698000E-02	-1.249000E-02	-7.879000E-03	-4.882000E-03	-5.195000E-03	-5.017000E-03	-5.455000E-03
l_2	1.448000E-04	1.111000E-04	7.258000E-05	4.692000E-05	4.664000E-05	4.282000E-05	4.273000E-05
l_3	-9.535000E-08	-7.706000E-08	-3.658000E-08	-1.742000E-08	-2.164000E-08	-2.132000E-08	-2.273000E-08

* See <http://www.gost.ru/sls/gost.nsf/PVP/CA5400452C7991BEC32566DA004601FA?Open-Document&ALT> for additional information.

TABLE B-8. Russian Density Model Parameters. Coefficients are listed for fixed level of solar flux values (F_0) that are valid for altitudes between 600 and 1500 km. The first set of values for e_5 , e_6 , and e_7 are for use with daily K_p values. The second set is for use with 3-hourly K_p values.

	$F_0 = 75$	$F_0 = 100$	$F_0 = 125$	$F_0 = 150$	$F_0 = 175$	$F_0 = 200$	$F_0 = 250$
a_1	-3.322830E+01	-3.277310E+01	-3.167150E+01	-2.975920E+01	-2.884630E+01	-2.629940E+01	-2.366270E+01
a_2	1.784000E-01	1.899000E-01	2.265000E-01	2.948000E-01	3.140000E-01	3.817000E-01	4.231000E-01
a_3	5.550636E+02	5.842450E+02	5.715408E+02	5.283389E+02	5.097230E+02	4.341220E+02	3.364318E+02
d_1	1.020400E+00						
d_2	2.499000E-03						
d_3	-1.519000E-06						
b_1	7.833000E-01	7.250000E-01	6.100000E-01	9.333000E-02	-3.333000E-01	-4.333000E-01	-1.750000E-01
b_2	2.861000E-03	2.675000E-03	2.343000E-03	3.038000E-03	3.522000E-03	3.522000E-03	2.642000E-03
b_3	-1.944000E-06	-1.750000E-06	-1.433000E-06	-1.711000E-06	-1.889000E-06	-1.889000E-06	-1.417000E-06
c_1	-4.400000E+00	-4.400000E+00	-8.980000E+00	-8.980000E+00	-1.578000E+01	-1.578000E+01	-9.750000E+00
c_2	3.024000E-02	3.024000E-02	4.087000E-02	4.087000E-02	5.757000E-02	5.757000E-02	3.383000E-02
c_3	-3.283000E-05	-3.283000E-05	-3.950000E-05	-3.950000E-05	-5.322000E-05	-5.322000E-05	-2.694000E-05
c_4	1.012000E-08	1.012000E-08	1.123000E-08	1.123000E-08	1.512000E-08	1.512000E-08	6.481000E-09
n_1	1.500000E+00						
n_2	6.000000E-03						
φ_B	5.411000E-01	5.515000E-01	5.585000E-01	5.585000E-01	5.585000E-01	5.585000E-01	5.585000E-01
e_1	-3.800000E+00	-3.700000E+00	-3.700000E+00	-4.400000E+00	-3.600000E+00	-3.600000E+00	1.000000E-01
e_2	1.972000E-02	1.783000E-02	1.750000E-02	1.981000E-02	1.694000E-02	1.653000E-02	2.639000E-03
e_3	-1.833000E-05	-1.506000E-05	-1.500000E-05	-1.806000E-05	-1.556000E-05	-1.528000E-05	-2.778000E-07
e_4	4.938000E-09	3.580000E-09	3.704000E-09	4.938000E-09	4.321000E-09	4.321000E-09	-6.173000E-10
e_5	-1.200000E-01	-1.200000E-01	-1.000000E-01	-1.000000E-01	-1.200000E-01	-1.100000E-01	-9.000000E-02
e_6	5.000000E-03	2.500000E-02	2.083000E-02	2.750000E-02	4.116000E-02	3.810000E-02	3.118000E-02
e_7	1.500000E-02	7.500000E-03	6.251000E-03	3.800000E-03	1.433000E-03	1.178000E-03	9.662000E-04
e_8	-1.200000E-01	-1.000000E-01	-1.000000E-01	-9.000000E-02	-8.000000E-02	-7.000000E-02	-5.000000E-02
e_9	2.500000E-02	2.625000E-02	2.617000E-02	2.542000E-02	2.333000E-02	2.125000E-02	1.375000E-02
e_{10}	7.500000E-03	5.625000E-03	4.250000E-03	3.125000E-03	2.500000E-03	1.875000E-03	1.875000E-03
I_1	1.083000E-02	8.317000E-03	4.667000E-03	-1.333000E-02	-3.500000E-03	-1.500000E-03	-8.333000E-04
I_2	6.694000E-05	4.837000E-05	4.606000E-05	7.167000E-05	4.317000E-05	3.250000E-05	2.817000E-05
I_3	-4.277000E-08	-3.039000E-08	-2.722000E-08	-3.518000E-08	-2.056000E-08	-1.389000E-08	-1.129000E-08

TABLE B-9. Russian Density Model Parameters. Coefficients are listed for the semiannual effect on density. The days are measured from the beginning of the year.

<i>d</i> , days	<i>A</i> (<i>d</i>)	<i>d</i> , days	<i>A</i> (<i>d</i>)	<i>d</i> , days	<i>A</i> (<i>d</i>)	<i>d</i> , days	<i>A</i> (<i>d</i>)	<i>d</i> , days	<i>A</i> (<i>d</i>)
0	-0.028	70	0.090	140	-0.037	210	-0.193	280	0.115
10	-0.045	80	0.114	150	-0.086	220	-0.173	290	0.144
20	-0.047	90	0.125	160	-0.128	230	-0.140	300	0.155
30	-0.035	100	0.118	170	-0.162	240	-0.096	310	0.145
40	-0.011	110	0.096	180	-0.185	250	-0.042	320	0.120
50	0.022	120	0.060	190	-0.199	260	0.015	330	0.084
60	0.057	130	0.013	200	-0.202	270	0.070	340	0.044

APPENDIX C MATHEMATICAL FUNDAMENTALS

- C.1 Introduction
- C.2 Vector Fundamentals
- C.3 Matrix Fundamentals
- C.4 Trigonometric Fundamentals
- C.5 Numerical Techniques

C.1 Introduction

The following terms are used throughout the text. They are primarily connected with computer programming, but they specify many operations in equations.

ATAN2 Mathematical procedure which determines the correct quadrant for a trigonometric operation using the sine and cosine arguments.

$$\theta = \text{ATAN2}(-0.9563, -0.2924) \Rightarrow \theta = 253^\circ$$

FRAC Fraction (decimal) of a real number

$$\text{FRAC}(1.57) = 0.57$$

INT Truncation of a real number

$$\text{INT}(1.57) = 1.0$$

MAX Maximum value from a set

$$\text{MAX}(3.2, 1.5, 6.3) = 6.3$$

MIN Minimum value from a set

$$\text{MIN}(3.2, 1.5, 6.3) = 1.5$$

MOD Modulus Operation.

$$8 \bmod 3 = 2 \text{ and } 9 \bmod 3 = 0$$

ROUND Approximation of a real number to nearest integer value

$$\text{ROUND}(1.57) = 2$$

SIGN Returns the sign of an input number, for example

$$\text{SIGN}(-157.93) = -1.0$$

TRUNC Truncation to nearest integer value towards zero.

$$\text{TRUNC}(1.57) = 1$$

The remaining mathematical notations are listed in Appendix A.

Miscellaneous Techniques

The binomial coefficient is often useful in relations.

$$\binom{n}{k} = \frac{n!}{(n-k)!k!} \quad (\text{C-1})$$

The error function is also useful for some applications.

$$\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt \quad (\text{C-2})$$

C.2 Vector Fundamentals

A **vector** is an array of real or complex numbers that has a magnitude and a direction. Although a vector's most general form can have multiple components, I'll present equations in this section as though all vectors have only three components because astrodynamics operates mainly in three dimensions.

First, we can determine the vector's magnitude by taking the square root of the sum of the squares of each component.

$$|\vec{a}| = \sqrt{a_i^2 + a_j^2 + a_k^2} \quad (\text{C-3})$$

Unit vectors describe the direction of a vector. They're often used in angles-only orbit determination and as unit axes for coordinate systems. The magnitude of the unit vector is 1.0, and the vector will always have all three components, although individual components may be zero.

$$\hat{a} = \frac{a_i}{|\vec{a}|}\hat{I} + \frac{a_j}{|\vec{a}|}\hat{J} + \frac{a_k}{|\vec{a}|}\hat{K} \quad (\text{C-4})$$

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 in determining orbital elements.

$$c = \vec{a} \cdot \vec{b} = |\vec{a}||\vec{b}| \cos(\alpha) = a_i b_i + a_j b_j + a_k b_k \quad (\text{C-5})$$

Cross products find 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.

$$\begin{aligned} \vec{c} &= \vec{a} \times \vec{b} = (a_j b_k - a_k b_j)\hat{I} - (a_i b_k - a_k b_i)\hat{J} + (a_i b_j - a_j b_i)\hat{K} \\ |\vec{c}| &= |\vec{a}||\vec{b}| \sin(\alpha) \end{aligned} \quad (\text{C-6})$$

A variant of the cross product involves multiple cross products. Be careful of the parenthesis because the operation is not associative.

$$\begin{aligned} \vec{a} \times (\vec{b} \times \vec{c}) &= (\vec{a} \cdot \vec{c})\vec{b} - (\vec{a} \cdot \vec{b})\vec{c} \\ (\vec{a} \times \vec{b}) \times \vec{c} &= (\vec{a} \cdot \vec{c})\vec{b} - (\vec{b} \cdot \vec{c})\vec{a} \end{aligned} \quad (\text{C-7})$$

When we express a vector in a particular coordinate system, we define the ***direction angles*** (α, β, γ) as the angles between each coordinate axis (IJK) and the individual components of the vector. The ***direction cosines*** are simply the cosines of these angles.

$$\cos(\alpha) = \frac{\hat{a} \cdot \hat{I}}{|\hat{a}| |\hat{I}|} \quad \cos(\beta) = \frac{\hat{a} \cdot \hat{J}}{|\hat{a}| |\hat{J}|} \quad \cos(\gamma) = \frac{\hat{a} \cdot \hat{K}}{|\hat{a}| |\hat{K}|} \quad (\text{C-8})$$

C.3 Matrix Fundamentals

A ***matrix*** is a rectangular array of real or complex numbers. We use a row (r)—col (c) notation for *all* indices of matrices. A ***real matrix*** contains all real numbers. The ***size*** of a matrix is defined to be $\text{row} \times \text{col}$. Matrices are equal if *every* element of the matrices are equal, and of course, if the two matrices are the same size. We refer to those terms with equal row and column indices as the diagonal elements. A ***square matrix*** has equal row and column dimensions, and its size is called the ***order***. The ***rank*** of a matrix is the largest submatrix of a matrix that has a non-zero determinant. The ***trace*** of a matrix is the scalar sum of the diagonal elements of the matrix. A ***scalar matrix*** is sometimes used in astrodynamics. It's a matrix with all the diagonal terms equal to some constant, and all other terms equal to zero. A matrix ***transpose*** simply interchanges the rows and columns.

Matrix inversion usually is mathematically intensive, and time consuming in computer code. For small matrices, simpler options exist. We can invert 2-by-2 matrices by (1) swapping the diagonal elements, (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 also change the sign for each component where $r + c$ is odd. For larger matrices, consult standard texts on mathematics and computer programming. A 3×3 example is shown below.

$$\mathbf{A} = \begin{bmatrix} 21.3 & 2 & 13 \\ -1.1 & 10.5 & 22 \\ -21 & 4 & 37 \end{bmatrix} \quad \mathbf{A}^{-1} = \begin{bmatrix} 0.035\ 913 & -0.002\ 629 & -0.011\ 054 \\ -0.050\ 350 & 0.126\ 814 & -0.057\ 712 \\ 0.025\ 827 & -0.015\ 202 & 0.026\ 992 \end{bmatrix}$$

Several other matrices occur in astrodynamics. A ***symmetric matrix*** results when $\mathbf{A}^T = \mathbf{A}$, and a ***skew-symmetric matrix*** yields $\mathbf{A}^T = -\mathbf{A}$. ***Orthogonal matrices***, $\mathbf{A}^T = \mathbf{A}^{-1}$, occur frequently in coordinate transformations.

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, \mathbf{X} , and arbitrary matrices \mathbf{C}, \mathbf{A} , and \mathbf{B} , and vectors $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$, we have

$$\mathbf{y} = \mathbf{C}^T \mathbf{X} \quad \mathbf{z} = \mathbf{A}^T \mathbf{X} + \mathbf{X}^T \mathbf{B} \mathbf{X} \quad (\text{C-9})$$

The gradients of y and z are

$$\nabla_{\mathbf{x}} \mathbf{y} = \mathbf{C}^T \quad \nabla_{\mathbf{x}} \mathbf{z} = \mathbf{A}^T + 2\mathbf{X}^T \mathbf{B}$$

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. $\det(\mathbf{A}) = \det(\mathbf{A}^T)$

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 .

$$\begin{aligned} a_1x + b_1y &= c_1 \\ a_2x + b_2y &= c_2 \end{aligned}$$

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,

$$D = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} \quad x = \frac{\begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix}}{D} \quad y = \frac{\begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}}{D}$$

C.4 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:

$$\begin{aligned} \cos^2(\alpha) + \sin^2(\alpha) &= 1 \\ 1 + \tan^2(\alpha) &= \sec^2(\alpha) \\ 1 + \cot^2(\alpha) &= \csc^2(\alpha) \end{aligned} \quad (\text{C-10})$$

These relations also exist for hyperbolic functions:

$$\cosh^2(\alpha) - \sinh^2(\alpha) = 1 \quad (\text{C-11})$$

Standard reduction formulas are

$$\begin{aligned} \sin(\alpha) &= \cos(90^\circ - \alpha) = \sin(180^\circ - \alpha) \\ \cos(\alpha) &= \sin(90^\circ - \alpha) = -\cos(180^\circ - \alpha) \\ \tan(\alpha) &= \cot(90^\circ - \alpha) = -\tan(180^\circ - \alpha) \\ \cot(\alpha) &= \tan(90^\circ - \alpha) = -\cot(180^\circ - \alpha) \\ \sin(-\alpha) &= -\sin(\alpha) \\ \cos(-\alpha) &= \cos(\alpha) \end{aligned} \quad (\text{C-12})$$

Selby (1975:228–229) shows a simple way to extend the results of the reduction formulas, establishing a rule so

$$f(\pm\alpha + n90^\circ) = \pm g(\alpha) \quad (\text{C-13})$$

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, and 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 . The second \pm sign is not necessarily the same as the first one, but is determined as follows: For a given function f , a given value of n , and a given choice of the first (\pm) sign, the second side will be the same as the first function for all values of α . Thus, it is only necessary to check the sign for any one value of α , and the formula will be complete.

Half-angle formulas are

$$\begin{aligned} \tan\left(\frac{\alpha}{2}\right) &= \frac{1 - \cos(\alpha)}{\sin(\alpha)} = \frac{\sin(\alpha)}{1 + \cos(\alpha)} \\ \sqrt{2}\cos\left(\frac{\alpha}{2}\right) &= \frac{\sin(\alpha)}{\sqrt{1 - \cos(\alpha)}} \Rightarrow \cos\left(\frac{\alpha}{2}\right) = \sqrt{\frac{1 + \cos(\alpha)}{2}} \\ \sqrt{2}\sin\left(\frac{\alpha}{2}\right) &= \frac{\sin(\alpha)}{\sqrt{1 + \cos(\alpha)}} \Rightarrow \sin\left(\frac{\alpha}{2}\right) = \sqrt{\frac{1 - \cos(\alpha)}{2}} \\ \tan^2(\alpha) &= \frac{1 - \cos(2\alpha)}{1 + \cos(2\alpha)} \\ \sin^2(\alpha + \beta) &= \frac{1}{2} - \frac{1}{2}\cos(2(\alpha + \beta)) \end{aligned} \quad (\text{C-14})$$

It's sometimes useful to expand sum and difference relations:

$$\begin{aligned}
 \sin(\alpha + \beta) &= \sin(\alpha)\cos(\beta) + \cos(\alpha)\sin(\beta) \\
 \sin(\alpha - \beta) &= \sin(\alpha)\cos(\beta) - \cos(\alpha)\sin(\beta) \\
 \cos(\alpha + \beta) &= \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta) \\
 \cos(\alpha - \beta) &= \cos(\alpha)\cos(\beta) + \sin(\alpha)\sin(\beta) \\
 \tan(\alpha + \beta) &= \frac{\tan(\alpha) + \tan(\beta)}{1 - \tan(\alpha)\tan(\beta)} \\
 \tan(\alpha - \beta) &= \frac{\tan(\alpha) - \tan(\beta)}{1 + \tan(\alpha)\tan(\beta)}
 \end{aligned} \tag{C-15}$$

We can express trigonometric terms using other forms:

$$\begin{aligned}
 \sin(\alpha) &= \tan(\alpha)\cos(\alpha), & \cos(\alpha) &= \cot(\alpha)\sin(\alpha) \\
 \tan(\alpha) &= \sin(\alpha)\sec(\alpha), & \cot(\alpha) &= \cos(\alpha)\csc(\alpha) \\
 \sec(\alpha) &= \csc(\alpha)\tan(\alpha), & \csc(\alpha) &= \sec(\alpha)\cot(\alpha)
 \end{aligned} \tag{C-16}$$

Double-angle formulas are useful in reduction of relations and appear often in the perturbation derivations:

$$\begin{aligned}
 \sin(2\alpha) &= 2\sin(\alpha)\cos(\alpha) = \frac{2\tan(\alpha)}{1 + \tan^2(\alpha)} \\
 \tan(2\alpha) &= \frac{2\tan(\alpha)}{1 - \tan^2(\alpha)} \\
 \cos(2\alpha) &= \cos^2(\alpha) - \sin^2(\alpha) = 2\cos^2(\alpha) - 1 \\
 &= 1 - 2\sin^2(\alpha) = \frac{1 - \tan^2(\alpha)}{1 + \tan^2(\alpha)} \\
 \cot(2\alpha) &= \frac{\cot^2(\alpha) - 1}{2\cot(\alpha)}
 \end{aligned} \tag{C-17}$$

Principal values of inverse trigonometric functions are useful for determining the proper quadrant. I often show both sine and cosine expressions to facilitate the use of ATAN2, but the following relations are useful for other applications.

$$\begin{aligned}
 \sin^{-1}(\alpha) &= -90 \text{ to } 90^\circ \\
 \cos^{-1}(\alpha) &= 0 \text{ to } 180^\circ \\
 \tan^{-1}(\alpha) &= -90 \text{ to } 90^\circ
 \end{aligned}$$

Series expansions are especially helpful in universal-variable techniques:

$$\begin{aligned}\sin(\alpha) &= \sum_{k=0}^{\infty} \frac{(-1)^k(\alpha)^{(2k+1)}}{(2k+1)!} = \alpha - \frac{\alpha^3}{3!} + \frac{\alpha^5}{5!} - \frac{\alpha^7}{7!} + \frac{\alpha^9}{9!} - \frac{\alpha^{11}}{11!} + \frac{\alpha^{13}}{13!} - \dots \\ \cos(\alpha) &= \sum_{k=0}^{\infty} \frac{(-1)^k(\alpha)^{2k}}{(2k)!} = 1 - \frac{\alpha^2}{2!} + \frac{\alpha^4}{4!} - \frac{\alpha^6}{6!} + \frac{\alpha^8}{8!} - \frac{\alpha^{10}}{10!} + \frac{\alpha^{12}}{12!} - \dots\end{aligned}\quad (\text{C-18})$$

For hyperbolic functions, the series expressions are

$$\begin{aligned}\sinh(\alpha) &= \sum_{k=0}^{\infty} \frac{\alpha^{(2k+1)}}{(2k+1)!} = \alpha + \frac{\alpha^3}{3!} + \frac{\alpha^5}{5!} + \frac{\alpha^7}{7!} + \frac{\alpha^9}{9!} + \frac{\alpha^{11}}{11!} + \frac{\alpha^{13}}{13!} + \dots \\ \cosh(\alpha) &= \sum_{k=0}^{\infty} \frac{\alpha^{2k}}{(2k)!} = 1 + \frac{\alpha^2}{2!} + \frac{\alpha^4}{4!} + \frac{\alpha^6}{6!} + \frac{\alpha^8}{8!} + \frac{\alpha^{10}}{10!} + \frac{\alpha^{12}}{12!} + \dots\end{aligned}\quad (\text{C-19})$$

A useful miscellaneous conversion is

$$\begin{aligned}\cos(\alpha) &= \cos^2\left(\frac{\alpha}{2}\right) - \sin^2\left(\frac{\alpha}{2}\right) \\ &= 2\cos^2\left(\frac{\alpha}{2}\right) - 1\end{aligned}\quad (\text{C-20})$$

C.4.1 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 geometry for the general equations.

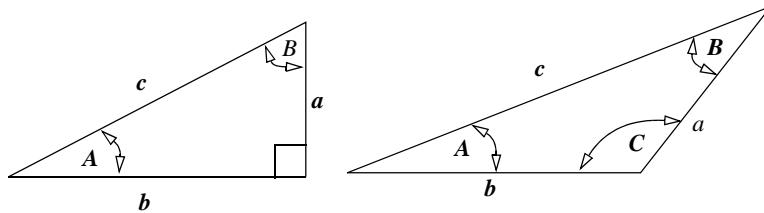


Figure C-1. Planar Triangles. The most commonly used triangle is the right triangle (left), although the oblique triangle (right) appears often.

For the right triangle, the *Pythagorean Theorem* is a convenient relation that permits us to determine the sides:

$$a^2 + b^2 = c^2$$

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.

$$\begin{aligned}c^2 &= a^2 + b^2 - 2ab \cos(C) \\a^2 &= b^2 + c^2 - 2bc \cos(A) \\b^2 &= c^2 + a^2 - 2ac \cos(B)\end{aligned}\tag{C-21}$$

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:

$$\frac{a}{\sin(A)} = \frac{b}{\sin(B)} = \frac{c}{\sin(C)}\tag{C-22}$$

Finally, we can form combinations:

$$\begin{aligned}a &= b \cos(C) + c \cos(B) \\b &= c \cos(A) + a \cos(C) \\c &= a \cos(B) + b \cos(A)\end{aligned}\tag{C-23}$$

C.4.2 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:

$$s = r\theta\tag{C-24}$$

The radius of the circle, r , allows the relation to work, and the units of the radius and the distance will be the same.

Another analysis deals with right spherical triangles, shown in Fig. C-3. Notice that lower-case letters denote all sides, and capital letters signify angles.

Napier's rules apply for all right spherical triangles.

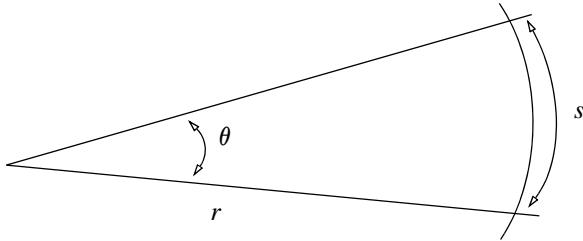


Figure C-2. Range and Angle Geometry. The ground range is related to the central angle by the simple relation $s = r\theta$.

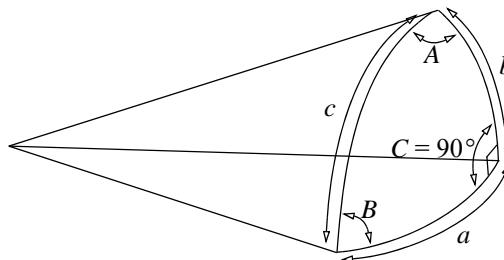


Figure C-3. Right Spherical Triangle. A right spherical triangle has one angle (C in the figure), that equals 90° .

$$\begin{array}{ll}
 \cos(c) = \cos(a)\cos(b) & \tan(a) = \cos(B)\tan(c) \\
 \sin(a) = \sin(A)\sin(c) & \tan(b) = \tan(B)\sin(a) \\
 \tan(b) = \cos(A)\tan(c) & \cos(c) = \cot(A)\cot(B) \\
 \tan(a) = \tan(A)\sin(b) & \cos(A) = \cos(a)\sin(B) \\
 \sin(b) = \sin(B)\sin(c) & \cos(B) = \cos(b)\sin(A)
 \end{array} \quad (C-25)$$

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.

1. Cosine Law:

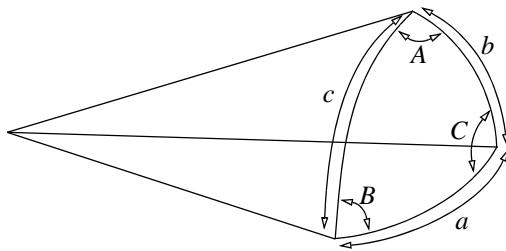


Figure C-4. Oblique Spherical Triangle. Oblique spherical triangles have no 90° angles. Also, angles and distance are interchangeable when dealing with the triangle.

For Sides

$$\begin{aligned}\cos(a) &= \cos(b)\cos(c) + \sin(b)\sin(c)\cos(A) \\ \cos(b) &= \cos(c)\cos(a) + \sin(c)\sin(a)\cos(B) \\ \cos(c) &= \cos(a)\cos(b) + \sin(a)\sin(b)\cos(C)\end{aligned}\tag{C-26}$$

For Angles

$$\begin{aligned}\cos(A) &= -\cos(B)\cos(C) + \sin(B)\sin(C)\cos(a) \\ \cos(B) &= -\cos(C)\cos(A) + \sin(C)\sin(A)\cos(b) \\ \cos(C) &= -\cos(A)\cos(B) + \sin(A)\sin(B)\cos(c)\end{aligned}$$

2. Sine Law

$$\frac{\sin(a)}{\sin(A)} = \frac{\sin(b)}{\sin(B)} = \frac{\sin(c)}{\sin(C)}\tag{C-27}$$

3. Four Parts

For Sides

$$\begin{aligned}\cot(A)\sin(B) &= \sin(c)\cot(a) - \cos(c)\cos(B) \\ \cot(B)\sin(C) &= \sin(a)\cot(b) - \cos(a)\cos(C) \\ \cot(C)\sin(A) &= \sin(b)\cot(c) - \cos(b)\cos(A)\end{aligned}\tag{C-28}$$

For Angles

$$\begin{aligned}\cot(a)\sin(b) &= \sin(C)\cot(A) + \cos(C)\cos(b) \\ \cot(b)\sin(c) &= \sin(A)\cot(B) + \cos(A)\cos(c) \\ \cot(c)\sin(a) &= \sin(B)\cot(C) + \cos(B)\cos(a)\end{aligned}$$

4. Analog to the cosine formula

For Sides

$$\begin{aligned}\sin(a)\cos(B) &= \cos(b)\sin(c) - \sin(b)\cos(c)\cos(A) \\ \sin(b)\cos(C) &= \cos(c)\sin(a) - \sin(c)\cos(a)\cos(B) \\ \sin(c)\cos(A) &= \cos(a)\sin(b) - \sin(a)\cos(b)\cos(C)\end{aligned}\tag{C-29}$$

For Angles

$$\begin{aligned}\sin(A)\cos(b) &= \cos(B)\sin(C) + \sin(B)\cos(C)\cos(a) \\ \sin(B)\cos(c) &= \cos(C)\sin(A) + \sin(C)\cos(A)\cos(b) \\ \sin(C)\cos(a) &= \cos(A)\sin(B) + \sin(A)\cos(B)\cos(c)\end{aligned}$$

C.5 Numerical Techniques

To solve problems in astrodynamics, you often use various numerical techniques because some problems don't have analytical solutions. For this book, we're mainly interested in interpolation techniques to determine precise values of coefficients.

C.5.1 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

$$x_{root_{1,2}} = \frac{-\alpha_1 \pm \sqrt{\alpha_1^2 - 4\alpha_2\alpha_0}}{2\alpha_2}\tag{C-30}$$

Cubic Roots

Astrodynamic 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.

I 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,

$$x_{root_1} = \sqrt[3]{-\frac{b}{2} + \sqrt{\Delta}} + \sqrt[3]{-\frac{b}{2} - \sqrt{\Delta}} \quad (\text{C-31})$$

and two imaginary roots. Be careful to evaluate negative cube roots correctly:

$$x_{root_{2i}} = \sqrt{\frac{-3}{2}} \left(\sqrt[3]{-\frac{b}{2} + \sqrt{\Delta}} - \sqrt[3]{-\frac{b}{2} - \sqrt{\Delta}} \right) \quad (\text{C-32})$$

$$x_{root_{3i}} = -x_{root_{2i}}$$

If the delta value (Δ) is equal to zero, all three roots are real, so

$$\begin{aligned} x_{root_1} &= 2 \sqrt[3]{\frac{b}{2}} \\ x_{root_2} &= x_{root_3} = \sqrt[3]{\frac{b}{2}} \end{aligned} \quad (\text{C-33})$$

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. Note that because $\sin(\phi)$ is always positive, the two quadrant solutions of cosine will also be positive.

$$\cos(\phi) = \frac{-b}{2 \sqrt{\frac{-a^3}{27}}} \quad \sin(\phi) = \sqrt{1 - \cos^2(\phi)}$$

We determine the three roots (for $\Delta < 0.0$) through circular relations:

$$Z_1 = E_0 \cos\left(\frac{\phi}{3}\right) \quad Z_2 = E_0 \cos\left(\frac{\phi}{3} + 120^\circ\right) \quad Z_3 = E_0 \cos\left(\frac{\phi}{3} + 240^\circ\right)$$

Then, for all temporary real roots (Z_i)

$$x_{root_{i=1\dots 3}} = Z_i - \frac{P}{3} \quad (\text{C-34})$$

All answers are real [Eq. (C-31), Eq. (C-33), and Eq. (C-34)], except for two imaginary roots in Eq. (C-32). 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 polynomial form that lacks x^3 :

$$\begin{aligned} y &= x^4 + Px^2 + Qx + R \\ P &= 6h^2 + 3bh + c \\ Q &= 4h^3 + 3bh^2 + 2ch + d \\ R &= h^4 + bh^3 + ch^2 + dh + e \end{aligned}$$

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 ($Z_{root1}, Z_{root1i}, Z_{root2}, Z_{root2i}$). 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 ($k = 1 \dots 4$)

$$\begin{aligned} r_j &= \sqrt{Z_{root_j}^2 + Z_{root_{ji}}^2} & \theta &= \text{ATAN2}(Z_{root_{ji}}, Z_{root_j}) \\ y_k &= \sqrt{r_j} \left(\cos\left(\frac{\theta}{2}\right) \pm i \sin\left(\frac{\theta}{2}\right) \right) \end{aligned} \quad (\text{C-35})$$

If Q is non-zero, we can find parameters similar to the cubic case, along with a new parameter, s :

$$a = \frac{1}{3} \left\{ 3(P^2 - 4R) - 4P^2 \right\}$$

$$s = -\frac{2}{3}P$$

$$b = \frac{1}{27}(16P^3 - 18P(P^2 - 4R) - 27Q^2)$$

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-31), (C-32), (C-33) and (C-34). We find R' as the largest value of $(Z_i + s)$ and then

$$\xi = \frac{1}{2}(P + R' - \sqrt{R'})$$

$$\beta = \frac{1}{2}\left(P + R' + \frac{Q}{\sqrt{R'}}\right)$$

Solving the quadratics gives us

$$\begin{aligned} y_i^2 + \sqrt{R'}y_i + \xi &= 0 & i &= 1, 2 \\ y_j^2 - \sqrt{R'}y_j + \beta &= 0 & j &= 3, 4 \end{aligned} \tag{C-36}$$

The final roots for both cases of Q are found using Eq. (C-35) or Eq. (C-36) and

$$x_{root_i} = y_i + h \quad i = 1 \dots 4$$

C.5.2 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

$$y(t) = y(t_o) + \frac{y(t_1) - y(t_o)}{t_1 - t_o}(t - t_o) \quad (\text{C-37})$$

Refer to Meeus (1991:23–24) or GAMS for more complex interpolation schemes.

Press (1992:389) introduces ***parabolic interpolation*** as a minimization / maximization technique. It has numerous uses in astrodynamics, particularly in applications requiring minimum and maximum values, such as the maximum elevation during a satellite pass over a site, or the minimum distance between two satellites as they approach each other. The technique produces a solution for the minimum / maximum value, x , (for example the time of closest approach) given three points, a , b , and c , (the times of two satellite ephemerides), and the evaluations at each point $f(a)$, $f(b)$, and $f(c)$, (the distances between each satellite).

$$= b - \frac{1}{2} \frac{(b-a)^2 \{f(b)-f(c)\} - (b-c)^2 \{f(b)-f(a)\}}{(b-a)\{f(b)-f(c)\} - (b-c)\{f(b)-f(a)\}} \quad (\text{C-38})$$

C.5.3 Blending and Splining Techniques

These 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.

First define a quadratic polynomial $P_{pb}(T)$ passing through three consecutive points (p_1, p_2, p_3) in Eq. (C-39). 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 .

$$P_{pb}(T) = \alpha_{pb2}T^2 + \alpha_{pb1}T + \alpha_{pb0} \quad (\text{C-39})$$

We'll define the parabolic coefficients as

$$\begin{aligned} \alpha_{pb0} &= p_1 \\ \alpha_{pb1} &= \frac{-3p_1 + 4p_2 - p_3}{2} \\ \alpha_{pb2} &= \frac{p_1 - 2p_2 + p_3}{2} \end{aligned}$$

The position and slope at the central point ($T = 1$) become

$$\begin{aligned} P_{pb}(1) &= p_2 \\ \dot{P}_{pb}(1) &= \frac{p_3 - p_1}{2} \end{aligned}$$

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_{pb2} parabola passes through points p_2, p_3, p_4 . On the central interval, p_2, p_3 , define the cubic $f_c(\tau)$ curve so it linearly weights P_1 and P_2 through the blending function:

$$\begin{aligned} f_c(\tau) &= \alpha_{c3}\tau^3 + \alpha_{c2}\tau^2 + \alpha_{c1}\tau + \alpha_{c0} & 0.0 \leq \tau \leq 1.0 \\ \dot{f}_c(\tau) &= 3\alpha_{c3}\tau^2 + 2\alpha_{c2}\tau + \alpha_{c1} \end{aligned} \quad (\text{C-40})$$

based on end conditions

$$f_c(0) = P_{pb1}(1) \text{ and } \dot{f}_c(0) = \dot{P}_{pb1}(1)$$

$$f_c(1) = P_{pb2}(1) \text{ and } \dot{f}_c(1) = \dot{P}_{pb2}(1)$$

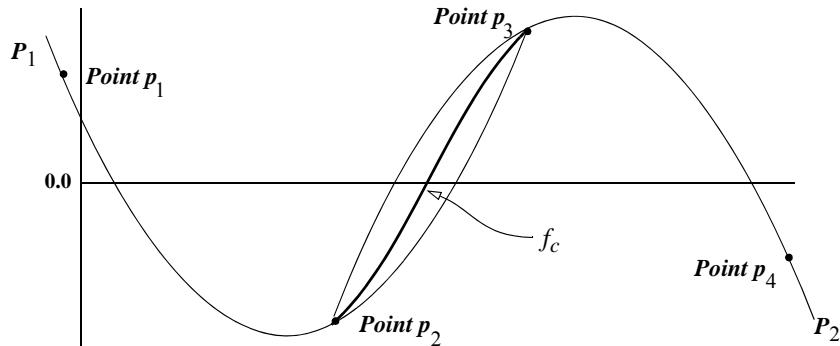


Figure C-5. Parabolic Blending. Parabolic blending involves taking two parabolas (P_i) and matching their curves to approximate the blended cubic curve (f_c). The points don't have to be equally spaced.

Equating the above expressions for the f_c curve gives us

$$\begin{aligned} f_c(0) &= \alpha_{c0} = P_{pb1}(1) = p_2 \\ \dot{f}_c(0) &= \alpha_{c1} = \dot{P}_{pb1}(1) = \frac{p_3 - p_1}{2} \\ f_c(1) &= \alpha_{c3} + \alpha_{c2} + \alpha_{c1} + \alpha_{c0} = P_{pb2}(1) = p_3 \\ \dot{f}_c(1) &= 3\alpha_{c3} + 2\alpha_{c2} + \alpha_{c1} = \dot{P}_{pb2}(1) = \frac{p_4 - p_2}{2} \end{aligned}$$

In terms of the original points, the α_c coefficients reduce to

$$\begin{aligned} \alpha_{c0} &= p_2 \\ \alpha_{c1} &= -\frac{1}{2}p_1 + \frac{1}{2}p_3 \\ \alpha_{c2} &= p_1 - \frac{5}{2}p_2 + 2p_3 - \frac{1}{2}p_4 \\ \alpha_{c3} &= -\frac{1}{2}p_1 + \frac{3}{2}p_2 - \frac{3}{2}p_3 + \frac{1}{2}p_4 \end{aligned} \tag{C-41}$$

For the final answer, proceed as follows (consider a time history of satellite elevation from a ground station).

1. Find the coefficients from Eq. (C-41) using the elevation values.
2. Use these coefficients to find the roots of f_c in the first equation of Eq. (C-40). We can solve this analytically by the methods in Sec. C.5.1.
3. Find new coefficients from Eq. (C-41) using the time values, and then choose real roots from step 2 between 0.0 and 1.0 and substitute into the

- first equation of Eq. (C-40). The solution is the time answer, τ .
4. Use the same process with the second equation of Eq. (C-40). The sign of the rate of f_c in the original evaluation will indicate a minimum or maximum.
 5. Any other parameters at the same time (range, azimuth, etc.) are found using coefficients from Eq. (C-41) formed with the function values, and the root found in step 3.

Cubic Splining

To alleviate the problems in modeling Fourier series and polynomial wiggle we use cubic and quintic spline functions. **Polynomial wiggle** is a phenomena where large variations occur between data points due to the extrema of the polynomial (Matthews, 1987:231). For instance, a fifth-order polynomial may fit each data point, while simultaneously diverging large amounts between data points. In this case, a second-order polynomial is probably a better answer. 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 a 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

$$\begin{aligned}
 P_c(\tau) &= \alpha_{c3}\tau^3 + \alpha_{c2}\tau^2 + \alpha_{c1}\tau + \alpha_{c0} \\
 \alpha_{c0} &= f(t_n) \\
 \alpha_{c1} &= \dot{f}(t_n)\Delta t \\
 \alpha_{c2} &= -3f(t_n) - 2\dot{f}(t_n)\Delta t + 3f(t_{n+1}) - \dot{f}(t_{n+1})\Delta t \\
 \alpha_{c3} &= 2f(t_n) + \dot{f}(t_n)\Delta t - 2f(t_{n+1}) + \dot{f}(t_{n+1})\Delta t \\
 \Delta t &= t_{n+1} - t_n \quad (t_{n+1} > t_n)
 \end{aligned} \tag{C-42}$$

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 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-42) α_c coefficients are shown in Eq. (C-43). 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,

$$\begin{aligned}
 \alpha_{c0} &= p_1 \\
 \alpha_{c1} &= \frac{(\tau_2^3 - \tau_2^2)(p_2 - p_1) + (\tau_1^2 - \tau_1^3)(p_3 - p_1) + (\tau_1^3 \tau_2^2 - \tau_1^2 \tau_2^3)(p_4 - p_1)}{DET} \\
 \alpha_{c2} &= \frac{(\tau_2 - \tau_2^3)(p_2 - p_1) + (\tau_1^3 - \tau_1)(p_3 - p_1) + (\tau_1 \tau_2^3 - \tau_1^3 \tau_2)(p_4 - p_1)}{DET} \\
 \alpha_{c3} &= \frac{(\tau_2^2 - \tau_2)(p_2 - p_1) + (\tau_1 - \tau_1^2)(p_3 - p_1) + (\tau_1^2 \tau_2 - \tau_1 \tau_2^2)(p_4 - p_1)}{DET} \\
 DET &= \tau_1^3 \tau_2^2 + \tau_1^2 \tau_2 + \tau_1 \tau_2^3 - \tau_1^3 \tau_2 - \tau_1^2 \tau_2^3 - \tau_1 \tau_2^2
 \end{aligned} \tag{C-43}$$

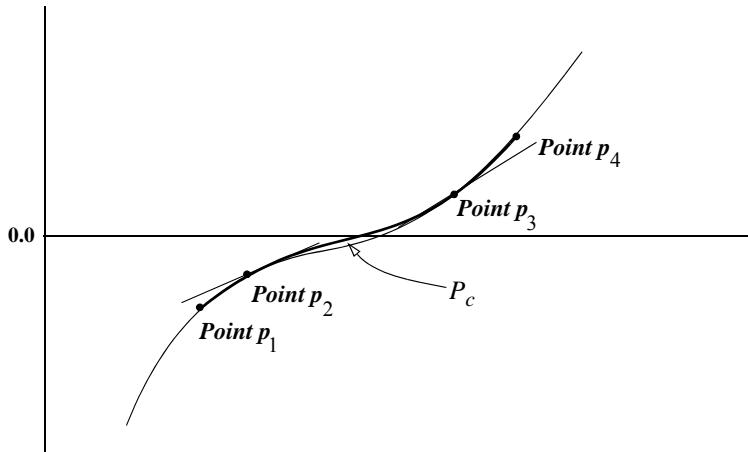


Figure C-6. Cubic Splining. We fit the curve with cubic splining by either examining the function and derivative at the start and end (p_2 and p_3), or by using four points along the curve. The resulting polynomials will perfectly match the point and slope of the function at p_2 and p_3 .

Cubic splining uses Eq. (C-42) or Eq. (C-43) to produce the polynomial which best fits the data. We find the three roots (real and imaginary) using analytical expressions in Sec. C.5.1. 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.5.1. Expressing the cubic equation with nested multiplication yields

$$\{(\alpha_{c3}\tau + \alpha_{c2})\tau + \alpha_{c1}\}\tau + \alpha_{c0} = 0 \tag{C-44}$$

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,

$$\text{IF } \alpha_{c0} > 0.0 \text{ and } \text{MIN}(\alpha_{c1}, \alpha_{c1} + \alpha_{c2}, \alpha_{c1} + \alpha_{c2} + \alpha_{c3}) > -\alpha_{c0}$$

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

$$P_{qr}(T) = \alpha_{qr4}T^4 + \alpha_{qr3}T^3 + \alpha_{qr2}T^2 + \alpha_{qr1}T + \alpha_{qr0} \quad (0.0 \leq T \leq 4.0) \quad (\text{C-45})$$

where

$$\begin{aligned} \alpha_{qr0} &= p_0 \\ \alpha_{qr1} &= (-50p_0 + 96p_1 - 72p_2 + 32p_3 - 6p_4)/24 \\ \alpha_{qr2} &= (35p_0 - 104p_1 + 114p_2 - 56p_3 + 11p_4)/24 \\ \alpha_{qr3} &= (-10p_0 + 36p_1 - 48p_2 + 28p_3 - 6p_4)/24 \\ \alpha_{qr4} &= (p_0 - 4p_1 + 6p_2 - 4p_3 + p_4)/24 \end{aligned}$$

The point, slope, and its derivative with respect to T at mid-point are

$$\begin{aligned} P_{qr}(2) &= 16\alpha_{qr4} + 8\alpha_{qr3} + 4\alpha_{qr2} + 2\alpha_{qr1} + \alpha_{qr0} \\ \dot{P}_{qr}(2) &= 32\alpha_{qr4} + 12\alpha_{qr3} + 4\alpha_{qr2} + \alpha_{qr1} \\ \ddot{P}_{qr}(2) &= 48\alpha_{qr4} + 12\alpha_{qr3} + 2\alpha_{qr2} \end{aligned} \quad (\text{C-46})$$

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_{qi} , ranging between p_3 and p_4 , combines the two quartics to match the P_{qr1} curve to second order at p_3 and to match the P_{qr2} curve to second order at p_4 . This process uses Eq. (C-45) and Eq. (C-46) to determine the function and derivatives at p_3 and p_4 . After some algebraic manipulation, the quintic equation for $f_{qi}(\tau)$ becomes

$$\begin{aligned} f_{qi}(\tau) &= \alpha_{qi5}\tau^5 + \alpha_{qi4}\tau^4 + \alpha_{qi3}\tau^3 + \alpha_{qi2}\tau^2 + \alpha_{qi1}\tau + \alpha_{qi0} & 0.0 \leq \tau \leq 1.0 \\ f_{qi}(\tau) &= 5\alpha_{qi5}\tau^4 + 4\alpha_{qi4}\tau^3 + 3\alpha_{qi3}\tau^2 + 2\alpha_{qi2}\tau + \alpha_{qi1} \end{aligned} \quad (\text{C-47})$$

$$\begin{aligned} \alpha_{qi0} &= p_3 \\ \alpha_{qi1} &= (2p_1 - 16p_2 + 16p_4 - 2p_5)/24 \\ \alpha_{qi2} &= (-p_1 + 16p_2 - 30p_3 + 16p_4 - p_5)/24 \\ \alpha_{qi3} &= (-9p_1 + 39p_2 - 70p_3 + 66p_4 - 33p_5 + 7p_6)/24 \\ \alpha_{qi4} &= (13p_1 - 64p_2 + 126p_3 - 124p_4 + 61p_5 - 12p_6)/24 \\ \alpha_{qi5} &= (-5p_1 + 25p_2 - 50p_3 + 50p_4 - 25p_5 + 5p_6)/24 \end{aligned}$$

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 $f_{qi}(\tau)$ equals zero. The roots of $f_{qi}(\tau)$ 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-44),

$$(((\alpha_{qi5}\tau + \alpha_{qi4})\tau + \alpha_{qi3})\tau + \alpha_{qi2})\tau + \alpha_{qi1})\tau + \alpha_{qi0} = 0$$

no roots exist

IF $\alpha_{qi0} > 0$ and

$$\begin{aligned} \text{MIN}(\alpha_{qi5}, \alpha_{qi5} + \alpha_{qi4}, \alpha_{qi5} + \alpha_{qi4} + \alpha_{qi3}, \alpha_{qi5} + \alpha_{qi4} + \alpha_{qi3} \\ + \alpha_{qi2}, \alpha_{qi5} + \alpha_{qi4} + \alpha_{qi3} + \alpha_{qi2} + \alpha_{qi1}) > -\alpha_{qi6}, \end{aligned}$$

or

IF $\alpha_{qi0} < 0$ and

$$\begin{aligned} \text{MAX}(\alpha_{qi5}, \alpha_{qi5} + \alpha_{qi4}, \alpha_{qi5} + \alpha_{qi4} + \alpha_{qi3}, \alpha_{qi5} + \alpha_{qi4} + \alpha_{qi3} \\ + \alpha_{qi2}, \alpha_{qi5} + \alpha_{qi4} + \alpha_{qi3} + \alpha_{qi2} + \alpha_{qi1}) < -\alpha_{qi6}, \end{aligned}$$

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-45) and Eq. (C-46) for the points $[f_{qi}(0), f_{qi}(0.25), f_{qi}(0.5), f_{qi}(0.75), f_{qi}(1)]$, 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. The remaining process follows the steps outlined in parabolic blending, but using Eq. (C-46) and Eq. (C-47).

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 $f(t)$, $\dot{f}(t)$, and $\ddot{f}(t)$ at t_n and t_{n+1} , define a quintic polynomial $P_{qi}(\tau)$ as

$$\begin{aligned}
 f_{qi}(\tau) &= \alpha_{qi5}\tau^5 + \alpha_{qi4}\tau^4 + \alpha_{qi3}\tau^3 + \alpha_{qi2}\tau^2 + \alpha_{qi1}\tau + \alpha_{qi0} & 0.0 \leq \tau \leq 1.0 \\
 \alpha_{qi0} &= f(t_n) & \tau = \frac{t - t_n}{\Delta t} \\
 \alpha_{qi1} &= \dot{f}(t_n)\Delta t & \Delta t = t_{n+1} - t_n \\
 \alpha_{qi2} &= 0.5\ddot{f}(t_n)\Delta t^2 & t_n \leq t \leq t_{n+1} \\
 \alpha_{qi3} &= -10f(t_n) - 6\dot{f}(t_n)\Delta t - 1.5\ddot{f}(t_n)\Delta t^2 & (C-48) \\
 &\quad + 10f(t_{n+1}) - 4\dot{f}(t_{n+1})\Delta t + 0.5\ddot{f}(t_{n+1})\Delta t^2 \\
 \alpha_{qi4} &= 15f(t_n) + 8\dot{f}(t_n)\Delta t + 1.5\ddot{f}(t_n)\Delta t^2 - 15f(t_{n+1}) + 7\dot{f}(t_{n+1})\Delta t - \ddot{f}(t_{n+1})\Delta t^2 \\
 \alpha_{qi5} &= -6f(t_n) - 3\dot{f}(t_n)\Delta t - 0.5\ddot{f}(t_n)\Delta t^2 \\
 &\quad + 6f(t_{n+1}) - 3\dot{f}(t_{n+1})\Delta t + 0.5\ddot{f}(t_{n+1})\Delta t^2
 \end{aligned}$$

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-48), as with Eq. (C-47).

APPENDIX D CONSTANTS AND EXPANSIONS

- D.1 Gravitational Coefficients
- D.2 Planetary Constants
- D.3 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 EGM-08 model for several reasons. First, it's a highly accurate model derived from numerous satellite measurements. Vetter (1994) compared gravitational models in detail and concluded that JGM-2 was the best for general-purpose and high-accuracy use. EGM-08 is the newest model that has replaced JGM-2 as a standard gravitational model.

TABLE D-1. Selected EGM-08 Zonal Coefficients (Normalized). Notice that the relative magnitude stays about the same for most zonal coefficients. Remember the transformation to unnormalized coefficients from Eq. (8-22), $C_l = \bar{C}_{l,0} / \Pi_{l,0}$. (Source NASA, *Gravitational Models*, Sec. D.5).

l	$\bar{C}_{l,0}$	$C_l = -J_l$	l	$\bar{C}_{l,0}$	$C_l = -J_l$
2	-4.84165143790815E-04	-1.08262617385222E-03	16	-4.71037252266068E-09	-2.70590300449556E-08
3	9.57161207093473E-07	2.53241051856772E-06	17	1.91875988417387E-08	1.13515365593836E-07
4	5.39965866638991E-07	1.6198975991697E-06	18	6.09862871807421E-09	3.70965102525025E-08
5	6.86702913736681E-08	2.27753590730836E-07	19	-3.30313643444747E-09	-2.06280804215613E-08
6	-1.49953927978527E-07	-5.40666576283813E-07	20	2.15591507033563E-08	1.38045920407128E-07
7	9.05120844521618E-08	3.50551795713742E-07	21	6.24977184727142E-09	4.09824946793957E-08
8	4.94756003005199E-08	2.03993125929884E-07	22	-1.07772567310914E-08	-7.22960359850626E-08
9	2.80180753216300E-08	1.22127958919496E-07	23	-2.22592369353670E-08	-1.52601640097366E-07
10	5.33304381729473E-08	2.44390769772693E-07	24	-7.29311730529204E-11	-5.10518211370443E-10
11	-5.07683787085927E-08	-2.43476590998147E-07	66	-3.66362872088076E-10	-4.22510275471911E-09
12	3.64361922614572E-08	1.82180961307286E-07	67	4.76877408077929E-10	5.54081477900513E-09
13	4.17293021685027E-08	2.16831814560722E-07	68	-5.23196166221788E-10	-6.12385412006501E-09
14	-2.26681154094404E-08	-1.22071537346982E-07	69	6.18121399841284E-10	7.28754382675693E-09
15	2.19216154508434E-09	1.22054389282870E-08	70	-5.86830047313794E-10	-6.96822072875664E-09

The values for the sectorial and tesseral coefficients are more numerous because the model is complete to order 360. Unlike the zonal coefficients, notice that the relative magnitude stays about the same for the normalized coefficients, but not for the unnormalized coefficients. See "Data Sources" on page 1050, for methods to obtain electronic files.

TABLE D-2. Selected EGM-08 Sectorial and Tesselal Coefficients. Sectorial and tesselal coefficients occur for all other gravitational coefficients where m is not equal to zero. Both normalized and unnormalized values are shown. Remember the transformation to unnormalized coefficients from Eq. (8-22), $C_{l,m} = \bar{C}_{l,m} / \Pi_{l,m}$. (Source NASA, *Gravitational Models*, Sec. D.5).

l	m	$\bar{C}_{l,m}$	$C_{l,m}$	$\bar{S}_{l,m}$	$S_{l,m}$
2	1	-2.06615509074176E-10	-2.66739475237484E-10	1.38441389137979E-09	1.78727064852404E-09
2	2	2.43938357328313E-06	1.57461532572292E-06	-1.40027370385934E-06	-9.03872789196567E-07
3	1	2.03046201047864E-06	2.19314963131333E-06	2.48200415856872E-07	2.68087089400898E-07
3	2	9.04787894809528E-07	3.09043900391649E-07	-6.19005475177618E-07	-2.11430620933483E-07
3	3	7.21321757121568E-07	1.00583513408823E-07	1.41434926192941E-06	1.97221581835718E-07
4	1	-5.36157389388867E-07	-5.08643560439584E-07	-4.73567346518086E-07	-4.49265432143808E-07
4	2	3.50501623962649E-07	7.83745457404552E-08	6.62480026275829E-07	1.48135037248860E-07
4	3	9.90856766672321E-07	5.92150177639666E-08	-2.00956723567452E-07	-1.20094612622961E-08
4	4	-1.88519633023033E-07	-3.98320424873187E-09	3.08803882149194E-07	6.52467287187687E-09
5	1	-6.29211923042529E-08	-5.38824899516695E-08	-9.43698073395769E-08	-8.08134749120457E-08
5	2	6.52078043176164E-07	1.0552866715599E-07	-3.23353192540522E-07	-5.23297729692969E-08
5	3	-4.51847152328843E-07	-1.49264876546663E-08	-2.14955408306046E-07	-7.10091727223769E-09
5	4	-2.95328761175629E-07	-2.29951140350422E-09	4.98070550102351E-08	3.87811503746114E-10
5	5	1.74811795496002E-07	4.30428041979182E-10	-6.69379935180165E-07	-1.64817193269055E-09
41	1	-7.14389627713801E-09	-2.21805653165695E-09	-3.29362707237955E-09	-1.02261437699097E-09
41	2	3.61396468299455E-09	2.70555988559151E-11	3.35004497864170E-09	2.50797893841895E-11
41	3	3.93976978594286E-09	7.12008946299204E-13	3.14595484298994E-09	5.68547939236019E-13
41	4	-1.64739041632043E-09	-7.19967819600023E-15	3.73031323662575E-09	1.63027868851955E-14
41	5	4.60274032374840E-09	4.87588266270418E-16	-3.02319751991475E-09	-3.20260439139394E-16
70	65	5.29614665832765E-10	5.93965041235761E-123	1.38556278213740E-09	1.55391439875052E-122
70	66	-1.59321539339451E-09	-6.85205624573007E-124	2.24580695324853E-09	9.65870379140892E-124
70	67	1.26460677051399E-09	3.2333400430087E-125	-1.14140348996022E-09	-2.09698508871215E-125
70	68	-3.06107682030793E-09	-2.76395013980299E-126	2.11623889518553E-09	1.91082391379390E-126
70	69	-1.88431012914400E-09	-1.02043644323414E-127	-9.31238136357700E-12	-5.04306324447286E-130
70	70	2.98214665798648E-10	1.36489284609771E-129	-1.40484139457899E-10	-6.42979098371188E-130

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 and Table D-4 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 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. 5.3.1. Knowing the orbital elements for each planet is important when analyzing interplanetary problems. Many algorithms require knowing the physical parameters of the Sun, Moon, and planets.

Table D-5 shows data for the Sun.

D.3 Reduction Coefficients

Trigonometric series are used extensively when forming the nutation coefficients. I list the largest ten to twenty terms of the primary theories in use today. Table D-6 through Table D-9 list the values.

TABLE D-3. 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
Symbol	\mathbb{C}	\mathfrak{M}	\mathfrak{V}	\oplus	\mathfrak{J}
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 ($^{\circ}$)	5.145 396	7.004 986 25	3.394 466 19	0.000 000 00	1.849 726 48
Right ascension of ascending node ($^{\circ}$)	varies	48.330 893 04	76.679 920 19	0.000 000 00	49.558 093 21
Longitude of perihelion ($^{\circ}$)	varies	77.456 119 04	131.563 707 24	102.937 348 08	336.060 233 98
True longitude ($^{\circ}$)	varies	252.250 905 51	181.979 800 84	100.466 448 51	355.433 274 63
Orbital period, (yrs)	0.0748	0.240 844 45	0.615 182 57	0.999 978 62	1.880 711 05
tropical (days)	27.321582	87.9666	224.6906	365.242 189 7	686.9150
Orbital velocity (km/s)	1.0232	47.8725	35.0214	29.7859	24.1309
Equatorial radius (km)	1738.0	2439.0	6052.0	6378.1363	3397.2
Reciprocal flattening				0.003 352 813 1	0.006 476 30
Gravitational parameter (km^3/s^2)	4902.799	2.2032×10^4	3.257×10^5	$3.986 004 415 \times 10^5$	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 ($^{\circ}$)	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-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).

	Jupiter	Saturn	Uranus	Neptune	Pluto
Symbol	\mathcal{J}	\mathcal{S}	\mathcal{U}	\mathcal{N}	\mathcal{P}
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 ($^{\circ}$)	1.303 269 66	2.488 878 10	0.773 196 17	1.769 952 21	17.142 166 67
right ascension of ascending node ($^{\circ}$)	100.464 440 64	113.665 523 70	74.005 947 23	131.784 057 02	110.297 138 89
Longitude of perihelion ($^{\circ}$)	14.331 309 24	93.056 787 28	173.005 159 22	48.123 690 50	224.134 861 11
True longitude ($^{\circ}$)	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)	71,492.0	60,268.0	25,559.0	24,764.0	1151.0
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 Equa- tor to orbit ($^{\circ}$)	3.12	26.73	97.86	29.56	118.0
J_2	0.014 75	0.016 45	0.012	0.004	
J_3					
J_4	-0.000 58	-0.001			
Density (gm/cm ³)	1.33	0.70	1.30	1.76	1.1

* 1 AU = 149,597,870 km.

TABLE D-5. 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 $_{\odot}$	29.784 691 674 9	km /solar s
1.0 AU	149,597,870.7	km	Gravitational Parameter		
a_{\oplus}	149,598,023.0	km	(μ_{\odot}) 1.0 AU ³ / TU $_{\odot}^2$	1.327 124 28	km ³ /(solar s) ² $\times 10^{11}$
Shape			Rotation		
Mass	332,946 Earth masses	kg	25.38		
	1.9891×10^{30}		Time		
			1.0 TU $_{\odot}$	58.132 440 906	solar days

TABLE D-6. Largest 1980 IAU Theory of Nutation Coefficients, Epoch J2000. The units for the longitude terms (A_i , B_i) and the obliquity terms (C_i , D_i) are 0.0001" per Julian century. Data is from Seidelmann (1992:112–113) and McCarthy (1996). There are 106 terms in the complete theory.

i	a_{n1}	a_{n2}	a_{n3}	a_{n4}	a_{n5}	A_i	B_i	C_i	D_i	i	a_{n1}	a_{n2}	a_{n3}	a_{n4}	a_{n5}	A_i	B_i	C_i	D_i
1	0	0	0	0	1	-171996	-174.2	92025	8.9	40	-1	0	2	2	2	-59	0.0	26	0
9	0	0	2	-2	2	-13187	-1.6	5736	-3.1	39	-1	0	0	0	1	-58	-0.1	32	0
31	0	0	2	0	2	-2274	-0.2	977	-0.5	41	1	0	2	0	1	-51	0.0	27	0
2	0	0	0	0	2	2062	0.2	-895	0.5	14	2	0	0	-2	0	48	0.0	1	0
10	0	1	0	0	0	1426	-3.4	54	-0.1	3	-2	0	2	0	1	46	0.0	-24	0
32	1	0	0	0	0	712	0.1	-7	0.0	42	0	0	2	2	2	-38	0.0	16	0
11	0	1	2	-2	2	-517	1.2	224	-0.6	45	2	0	2	0	2	-31	0.0	13	0
33	0	0	2	0	1	-386	-0.4	200	0.0	43	2	0	0	0	0	29	0.0	-1	0
34	1	0	2	0	2	-301	0.0	129	-0.1	44	1	0	2	-2	2	29	0.0	-12	0
12	0	-1	2	-2	2	217	-0.5	-95	0.3	46	0	0	2	0	0	26	0.0	-1	0
35	1	0	0	-2	0	-158	0.0	-1	0.0	15	0	0	2	-2	0	-22	0.0	0	0
13	0	0	2	-2	1	129	0.1	-70	0.0	47	-1	0	2	0	1	21	0.0	-10	0
36	-1	0	2	0	2	123	0.0	-53	0.0	16	0	2	0	0	0	17	-0.1	0	0
38	1	0	0	0	1	63	0.1	-33	0.0	18	0	2	2	-2	2	-16	0.1	7	0
37	0	0	0	2	0	63	0.0	-2	0.0	48	-1	0	0	2	1	16	0.0	-8	0

The calculations for the IAU 2010 conventions equinox-based transformation have numerous coefficients from both nutation and planetary effects.

Expressions for the Greenwich Sidereal time use another set of coefficients.

The IAU 2010 conventions also use a long list of coefficients for the CIO.

TABLE D-7. Largest IAU-2000 IERS Nutation Coefficients. There are several parameters needed for the IAU-2000 nutation terms. The terms (A_{psl} , A_{pcl} , A_{esi} , A_{eci}) are the sine and cosine multipliers with units of $0.001''$. The argument has 5 nutation and 9 planetary multipliers respectively. Data is from the IERS documents (Kaplan 2005), or online (http://maia.usno.navy.mil/conv2010/chapter5/tabc5_3a.txt, 3b.txt). Note that the 2010 tables include both nutation and planetary coefficients. There are about 1320 total terms in multiple summations.

TABLE D-8. Largest IAU-2010 IERS Greenwich Sidereal Time Coefficients. Mean Greenwich Sidereal Time requires another set of coefficients. The terms (A_{gsi} , A_{gci}) are the sine and cosine multipliers with units of $0.000\ 001''$. The argument has both a_{pi} , nutation, and a_{pli} , planetary multipliers. Data is from online (http://maia.usno.navy.mil/conv2010/chapter5/tabc5_4.txt). There are 34 total terms.

TABLE D-9. Largest IAU-2000 IERS CIO Theory Coefficients. There are several parameters needed for the IAU-2000 x , y , and s terms. The terms (A_{0xsi} , A_{0xci}) are the sine and cosine multipliers with units of 0.000 001". The argument has both nutation (a_{0xi1i} to a_{0x5i}) and planetary (a_{0x6i} to a_{0x14i}) multipliers. Data is from online files (<http://maia.usno.navy.mil/conv2010/chapter5>)/tab5.2a.txt, 2b.txt, 2c.txt). There are 5 summations and different numbers of terms for each parameter, ranging from 1 to 1306.

D.4 Planetary Ephemerides

From Meeus (1991:202–204), the ecliptic orbital elements of the planets, referenced to the mean equator and mean equinox of J2000, are as follows. You can get ω and ν from Eq. (2-88), and from Eq. (2-93) and Eq. (2-52), respectively. These expressions are for DE 245.

Reference J2000 in Degrees

Mercury

$$\begin{aligned} a &= 0.387\,098\,310 \text{ AU} \\ e &= 0.205\,631\,75 + 0.000\,020\,406 T_{TDB} - 0.000\,000\,028\,4 T_{TDB}^2 - 0.000\,000\,000\,17 T_{TDB}^3 \\ i &= 7.004\,986^\circ - 0.005\,951\,6 T_{TDB} + 0.000\,000\,81 T_{TDB}^2 + 0.000\,000\,041 T_{TDB}^3 \\ \Omega &= 48.330\,893^\circ - 0.125\,422\,9 T_{TDB} - 0.000\,088\,33 T_{TDB}^2 - 0.000\,000\,196 T_{TDB}^3 \\ \tilde{\omega} &= 77.456\,119^\circ + 0.158\,864\,3 T_{TDB} - 0.000\,013\,43 T_{TDB}^2 + 0.000\,000\,039 T_{TDB}^3 \\ \lambda_M &= 252.250\,906^\circ + 149,472.674\,635\,8 T_{TDB} - 0.000\,005\,35 T_{TDB}^2 + 0.000\,000\,002 T_{TDB}^3 \end{aligned}$$

Venus

$$\begin{aligned} a &= 0.723\,329\,820 \text{ AU} \\ e &= 0.006\,771\,88 - 0.000\,047\,766 T_{TDB} + 0.000\,000\,097\,5 T_{TDB}^2 + 0.000\,000\,000\,44 T_{TDB}^3 \\ i &= 3.394\,662^\circ - 0.000\,856\,8 T_{TDB} - 0.000\,032\,44 T_{TDB}^2 + 0.000\,000\,010 T_{TDB}^3 \\ \Omega &= 76.679\,920^\circ - 0.278\,008\,0 T_{TDB} - 0.000\,142\,56 T_{TDB}^2 - 0.000\,000\,198 T_{TDB}^3 \\ \tilde{\omega} &= 131.563\,707^\circ + 0.004\,864\,6 T_{TDB} - 0.001\,382\,32 T_{TDB}^2 - 0.000\,005\,332 T_{TDB}^3 \\ \lambda_M &= 181.979\,801^\circ + 58,517.815\,676\,0 T_{TDB} + 0.000\,001\,65 T_{TDB}^2 - 0.000\,000\,002 T_{TDB}^3 \end{aligned}$$

Earth (Ω varies widely in other programs because the inclination is near zero)

$$\begin{aligned} a &= 1.000\,001\,018 \text{ AU} \\ e &= 0.016\,708\,62 - 0.000\,042\,037 T_{TDB} - 0.000\,000\,123\,6 T_{TDB}^2 + 0.000\,000\,000\,04 T_{TDB}^3 \\ i &= 0.000\,000\,0^\circ + 0.013\,054\,6 T_{TDB} - 0.000\,009\,31 T_{TDB}^2 - 0.000\,000\,034 T_{TDB}^3 \\ \Omega &= 174.873\,174 - 0.241\,090\,8 T_{TDB} + 0.000\,040\,67 T_{TDB}^2 - 0.000\,001\,327 T_{TDB}^3 \\ \tilde{\omega} &= 102.937\,348^\circ + 0.322\,555\,7 T_{TDB} + 0.000\,150\,26 T_{TDB}^2 + 0.000\,000\,478 T_{TDB}^3 \\ \lambda_M &= 100.466\,449^\circ + 35,999.372\,851\,9 T_{TDB} - 0.000\,005\,68 T_{TDB}^2 + 0.000\,000\,000 T_{TDB}^3 \end{aligned}$$

Mars

$$\begin{aligned}
a &= 1.523\,679\,342 \text{ AU} \\
e &= 0.093\,400\,62 + 0.000\,090\,483 T_{TDB} - 0.000\,000\,080\,6 T_{TDB}^2 - 0.000\,000\,000\,35 T_{TDB}^3 \\
i &= 1.849\,726^\circ - 0.008\,147\,9 T_{TDB} - 0.000\,022\,55 T_{TDB}^2 - 0.000\,000\,027 T_{TDB}^3 \\
\Omega &= 49.558\,093^\circ - 0.294\,984\,6 T_{TDB} - 0.000\,639\,93 T_{TDB}^2 - 0.000\,002\,143 T_{TDB}^3 \\
\tilde{\omega} &= 336.060\,234^\circ + 0.443\,889\,8 T_{TDB} - 0.000\,173\,21 T_{TDB}^2 + 0.000\,000\,300 T_{TDB}^3 \\
\lambda_M &= 355.433\,275^\circ + 19,140.299\,331\,3 T_{TDB} + 0.000\,002\,61 T_{TDB}^2 - 0.000\,000\,003 T_{TDB}^3
\end{aligned}$$

Jupiter

$$\begin{aligned}
a &= 5.202\,603\,191 + 0.000\,000\,191\,3 T_{TDB} \text{ AU} \\
e &= 0.048\,494\,85 + 0.000\,163\,244 T_{TDB} - 0.000\,000\,471\,9 T_{TDB}^2 - 0.000\,000\,001\,97 T_{TDB}^3 \\
i &= 1.303\,270^\circ - 0.001\,987\,2 T_{TDB} + 0.000\,033\,18 T_{TDB}^2 + 0.000\,000\,092 T_{TDB}^3 \\
\Omega &= 100.464\,441^\circ + 0.176\,682\,8 T_{TDB} + 0.000\,903\,87 T_{TDB}^2 - 0.000\,007\,032 T_{TDB}^3 \\
\tilde{\omega} &= 14.331\,309^\circ + 0.215\,552\,5 T_{TDB} + 0.000\,722\,52 T_{TDB}^2 - 0.000\,004\,590 T_{TDB}^3 \\
\lambda_M &= 34.351\,484^\circ + 3034.905\,674\,6 T_{TDB} - 0.000\,085\,01 T_{TDB}^2 + 0.000\,000\,004 T_{TDB}^3
\end{aligned}$$

Saturn

$$\begin{aligned}
a &= 9.554\,909\,596 - 0.000\,002\,138\,9 T_{TDB} \text{ AU} \\
e &= 0.055\,508\,62 - 0.000\,346\,818 T_{TDB} - 0.000\,000\,645\,6 T_{TDB}^2 + 0.000\,000\,003\,38 T_{TDB}^3 \\
i &= 2.488\,878^\circ + 0.002\,551\,5 T_{TDB} - 0.000\,049\,03 T_{TDB}^2 + 0.000\,000\,018 T_{TDB}^3 \\
\Omega &= 113.665\,524^\circ - 0.256\,664\,9 T_{TDB} - 0.000\,183\,45 T_{TDB}^2 + 0.000\,000\,357 T_{TDB}^3 \\
\tilde{\omega} &= 93.056\,787^\circ + 0.566\,549\,6 T_{TDB} + 0.000\,528\,09 T_{TDB}^2 + 0.000\,004\,882 T_{TDB}^3 \\
\lambda_M &= 50.077\,471^\circ + 1222.113\,794\,3 T_{TDB} + 0.000\,210\,04 T_{TDB}^2 - 0.000\,000\,019 T_{TDB}^3
\end{aligned}$$

Uranus

$$\begin{aligned}
a &= 19.218\,446\,062 - 0.000\,000\,037\,2 T_{TDB} + 0.000\,000\,000\,98 T_{TDB}^2 \text{ AU} \\
e &= 0.0462\,959\,0 - 0.000\,027\,337 T_{TDB} + 0.000\,000\,079\,0 T_{TDB}^2 + 0.000\,000\,000\,25 T_{TDB}^3 \\
i &= 0.773\,196^\circ - 0.001\,686\,9 T_{TDB} + 0.000\,003\,49 T_{TDB}^2 + 0.000\,000\,016 T_{TDB}^3 \\
\Omega &= 74.005\,947^\circ + 0.074\,146\,1 T_{TDB} + 0.000\,405\,40 T_{TDB}^2 + 0.000\,000\,104 T_{TDB}^3
\end{aligned}$$

$$\tilde{\omega} = 173.005\ 159^\circ + 0.089\ 320\ 6T_{TDB} - 0.000\ 094\ 70T_{TDB}^2 + 0.000\ 000\ 413T_{TDB}^3$$

$$\lambda_M = 314.055\ 005^\circ + 428.466\ 998\ 3T_{TDB} - 0.000\ 004\ 86T_{TDB}^2 + 0.000\ 000\ 006T_{TDB}^3$$

Neptune

$$a = 30.110\ 386\ 869 - 0.000\ 000\ 166\ 3T_{TDB} + 0.000\ 000\ 000\ 69T_{TDB}^2 \text{ AU}$$

$$e = 0.008\ 988\ 09 + 0.000\ 006\ 408T_{TDB} - 0.000\ 000\ 000\ 8T_{TDB}^2$$

$$i = 1.769\ 952^\circ + 0.000\ 225\ 7T_{TDB} + 0.000\ 000\ 23T_{TDB}^2 - 0.000\ 000\ 000T_{TDB}^3$$

$$\Omega = 131.784\ 057^\circ - 0.006\ 165\ 1T_{TDB} - 0.000\ 002\ 19T_{TDB}^2 - 0.000\ 000\ 078T_{TDB}^3$$

$$\tilde{\omega} = 48.123\ 691^\circ + 0.029\ 158\ 7T_{TDB} + 0.000\ 070\ 51T_{TDB}^2 - 0.000\ 000\ 000T_{TDB}^3$$

$$\lambda_M = 304.348\ 665^\circ + 218.486\ 200\ 2T_{TDB} + 0.000\ 000\ 59T_{TDB}^2 - 0.000\ 000\ 002T_{TDB}^3$$

The following elements for Pluto are from Seidelmann (1992:316).

Pluto

$$a = 39.481\ 686\ 77 - 0.000\ 769\ 12''T_{TDB} \text{ AU}$$

$$e = 0.248\ 807\ 66 + 0.000\ 064\ 65''T_{TDB}$$

$$i = 17.141\ 75^\circ\ 11.07''T_{TDB}$$

$$\Omega = 110.303\ 47^\circ - 37.33''T_{TDB}$$

$$\tilde{\omega} = 224.066\ 76^\circ - 132.25''T_{TDB}$$

$$\lambda_M = 238.928\ 81^\circ + 522,747.90''T_{TDB}$$

Reference Mean Equinox of Date in Degrees

The semimajor axis and eccentricity will be identical to the degree expressions already shown.

Mercury

$$i = 7.004\ 986 + 0.001\ 821\ 5T_{TDB} - 0.000\ 018\ 09T_{TDB}^2 + 0.000\ 000\ 053T_{TDB}^3$$

$$\Omega = 48.330\ 893 + 1.186\ 189\ 0T_{TDB} + 0.000\ 175\ 87T_{TDB}^2 + 0.000\ 000\ 211T_{TDB}^3$$

$$\tilde{\omega} = 77.456\ 119 + 1.556\ 477\ 5T_{TDB} + 0.000\ 295\ 89T_{TDB}^2 + 0.000\ 000\ 056T_{TDB}^3$$

$$\lambda_M = 252.250\ 906 + 149,474.072\ 2491T_{TDB} + 0.000\ 30397T_{TDB}^2 + 0.000\ 000\ 018T_{TDB}^3$$

Venus

$$i = 3.394\ 662 + 0.001\ 003\ 7T_{TDB} - 0.000\ 000\ 88T_{TDB}^2 - 0.000\ 000\ 007T_{TDB}^3$$

$$\Omega = 76.679\ 920 + 0.901\ 119\ 0T_{TDB} + 0.000\ 406\ 65T_{TDB}^2 - 0.000\ 000\ 080T_{TDB}^3$$

$$\begin{aligned}\tilde{\omega} &= 131.563\ 707 + 1.402\ 218\ 8T_{TDB} - 0.001\ 073\ 37T_{TDB}^2 - 0.000\ 005\ 315T_{TDB}^3 \\ \lambda_M &= 181.979\ 801 + 58519.213\ 030\ 2T_{TDB} + 0.000\ 310\ 60T_{TDB}^2 + 0.000\ 000\ 015T_{TDB}^3\end{aligned}$$

Earth

$$i = 0.0$$

$$\Omega = 0.0$$

$$\tilde{\omega} = 102.937\ 348 + 1.719\ 526\ 9T_{TDB} + 0.000\ 459\ 62T_{TDB}^2 + 0.000\ 000\ 499T_{TDB}^3$$

$$\lambda_M = 100.466\ 449 + 36000/769\ 823\ 1T_{TDB} + 0.000\ 303\ 68T_{TDB}^2 + 0.000\ 000\ 021T_{TDB}^3$$

Mars

$$i = 1.849\ 726 - 0.000\ 601\ 0T_{TDB} + 0.000\ 012\ 76T_{TDB}^2 - 0.000\ 000\ 006T_{TDB}^3$$

$$\Omega = 49.558\ 093 + 0.772\ 092\ 3T_{TDB} + 0.000\ 016\ 05T_{TDB}^2 + 0.000\ 002\ 325T_{TDB}^3$$

$$\tilde{\omega} = 336.060\ 234 + 1.841\ 033\ 1T_{TDB} + 0.000\ 135\ 15T_{TDB}^2 + 0.000\ 000\ 318T_{TDB}^3$$

$$\lambda_M = 355.433\ 275 + 19141.696\ 474\ 6T_{TDB} + 0.000\ 310\ 97T_{TDB}^2 + 0.000\ 000\ 015T_{TDB}^3$$

Jupiter

$$i = 1.303\ 270 - 0.005\ 496\ 6T_{TDB} + 0.000\ 004\ 65T_{TDB}^2 - 0.000\ 000\ 004T_{TDB}^3$$

$$\Omega = 100.464\ 441 + 1.020\ 955\ 0T_{TDB} + 0.000\ 401\ 17T_{TDB}^2 + 0.000\ 000\ 569T_{TDB}^3$$

$$\tilde{\omega} = 14.331\ 309 + 1.612\ 666\ 8T_{TDB} + 0.001\ 031\ 27T_{TDB}^2 - 0.000\ 004\ 569T_{TDB}^3$$

$$\lambda_M = 34.351\ 484 + 3036.302\ 788\ 9T_{TDB} + 0.000\ 223\ 74T_{TDB}^2 + 0.000\ 000\ 025T_{TDB}^3$$

Saturn

$$i = 2.488\ 878 - 0.003\ 7363T_{TDB} - 0.000\ 015\ 16T_{TDB}^2 + 0.000\ 000\ 089T_{TDB}^3$$

$$\Omega = 113.665\ 524 + 0.877\ 097\ 9T_{TDB} - 0.000\ 120\ 67T_{TDB}^2 - 0.000\ 002\ 380T_{TDB}^3$$

$$\tilde{\omega} = 93.056\ 787 + 1.963\ 769\ 4T_{TDB} + 0.000\ 837\ 57T_{TDB}^2 + 0.000\ 004\ 899T_{TDB}^3$$

$$\lambda_M = 50.077\ 471 + 1223.511\ 014\ 1T_{TDB} + 0.000\ 519\ 52T_{TDB}^2 - 0.000\ 000\ 003T_{TDB}^3$$

Uranus

$$i = 0.773\ 196 + 0.000\ 774\ 4T_{TDB} + 0.000\ 037\ 49T_{TDB}^2 - 0.000\ 000\ 092T_{TDB}^3$$

$$\Omega = 74.005\ 947 + 0.521\ 125\ 8T_{TDB} + 0.001\ 339\ 82T_{TDB}^2 + 0.000\ 018\ 516T_{TDB}^3$$

$$\tilde{\omega} = 173.005\ 159 + 1.486\ 378\ 4T_{TDB} + 0.000\ 214\ 50T_{TDB}^2 + 0.000\ 000\ 433T_{TDB}^3$$

$$\lambda_M = 314.055\ 005 + 429.864\ 056\ 1T_{TDB} + 0.000\ 304\ 34T_{TDB}^2 + 0.000\ 000\ 026T_{TDB}^3$$

Neptune

$$\begin{aligned} i &= 1.769\,952 - 0.009\,308\,2T_{TDB} - 0.000\,007\,08T_{TDB}^2 + 0.000\,000\,028T_{TDB}^3 \\ \Omega &= 131.784\,057 + 1.102\,205\,7T_{TDB} + 0.000\,260\,06T_{TDB}^2 - 0.000\,000\,636T_{TDB}^3 \\ \tilde{\omega} &= 48.123\,691 + 1.426\,2677T_{TDB} + 0.000\,379\,18T_{TDB}^2 - 0.000\,000\,003T_{TDB}^3 \\ \lambda_M &= 304.348\,665 + 219.883\,309\,2T_{TDB} + 0.000\,309\,26T_{TDB}^2 + 0.000\,000\,018T_{TDB}^3 \end{aligned}$$

Pluto

$$\begin{aligned} i &= 17.132\,33^\circ \\ \Omega &= 110.406\,5^\circ \\ \tilde{\omega} &= 224.6148^\circ \\ \lambda_M &= 218.887\,35^\circ \end{aligned}$$

D.5 Data Sources

Proper data is essential to accurate flight dynamics calculations. The gravitational and third-body models are well established, but coefficients can differ between models. Sources to assemble a continuous file of Earth Orientation Parameter and solar weather data for use with historical, real-time, and future mission design operations is shown.

Gravitational Models:

It's important to obtain accurate coefficients for any gravity model that is used in a numerical integration program. The following sites may be used to find several popular gravity model coefficients, as well as information concerning their development.

Information on EGM-2008:

<http://earth-info.nga.mil/GandG/wgs84/gravitymod/egm2008/index.html>

Information on EGM-96/WGS-84:

<http://earth-info.nga.mil/GandG/wgs84/gravitymod/index.html>

Information on various JGM, TEG, and EGM models:

<ftp://ftp.csr.utexas.edu/pub/grav/>

Earth Orientation Parameters (EOP):

Final results of smoothed Earth Orientation Parameters (x_p , y_p , ΔUT1 , LOD , $\delta\Delta\Psi_{1980}$ and $\delta\Delta\epsilon_{1980}$, dX and dY) including 180 days of prediction are available from the International Earth Rotation Service (IERS). These products are also available from the National Earth Orientation Service (NEOS) at the U.S. Naval Observatory (USNO). The IERS Bulletins A (published weekly, includes one year predictions for x_p , y_p , ΔUT1) and B (published monthly; includes `finals.data`) are also available. Both centers include separate files for the IAU-76/FK5 and IAU-2000 values. The USNO also provides EOP data in the Bulletins A and B. Bulletin A contains the rapid results while Bulletin B contains the finalized results. Figure D-1. shows the available EOP data^c.

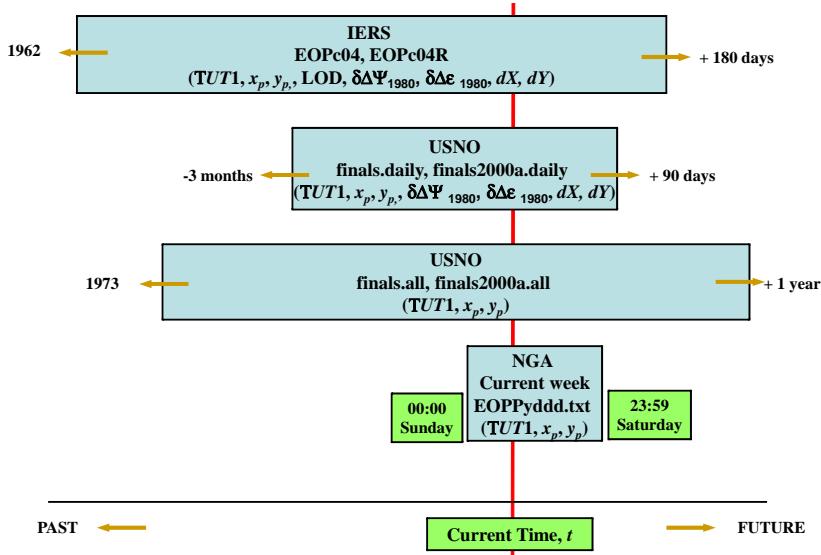


Figure D-1. Earth Orientation Parameter Data Availability. There are several sources of data we can use to populate a database of values for propagation into the future. Real-time operations require seamless integration of this information. File names are included from each source of data.

For the current week, predictions of x_p , y_p , and ΔUT1 are available from National Geospatial Intelligence Agency (NGA) in the form of polynomials. The information is updated every Thursday. The pole positions are accurate to about 0.01" after a month, and ΔUT1 predictions are accurate to a few milliseconds after about 20 days. Although pole positions may be calculated many months into the future, the same operation for ΔUT1 is not recommended.

The NGA Earth Orientation Parameter Prediction (EOPP) files are ASCII files named eoppyww.txt where y is the last digit of the year and ww is the week of the year. eg. EOPP402.TXT is week 02 of 2004. The predictions are calculated on Thursday of each week (Wednesday when a federal holiday falls on a Thursday) to go into effect on the following Sunday. Single daily values are given after the coefficients.

Sample:

```

53154.00   .056293   .000000   .079979   .034045   -.028833   -.113721365.25
435.00    .343292   .000000   .030247   .117765   .083031   .026326365.25435.00
53370.00   -.523010   -.000381   .000514   -.000831   -.022000   .006000
   .000607   .000855   .012000   -.007000  27.5600  13.6600 365.2500 182.6250
 32 5205 53575 53572
53575        -.013034           .417876        -.6043788
53576        -.011681           .418738        -.6046355
53577        -.010318           .419577        -.6049358

```

There is not much information given with the parameters, although the readme file (<http://earth-info.nga.mil/GandG/sathtml/eoppdoc.html>) explains each field. The information is presented below for convenience. There are several constants in the file - the

annual period, AnnPer = 365.25, Chandler period, ChnPer = 435.0, semi-annual period, SAnnPer = 182.625. Using a FORTRAN format statement and the variable names, each line is

```
FORMAT( F10.2, 6(F10.6), F6.2 )    -> ta, A, B, C1, C2, D1, D2, P1
FORMAT( F6.2, 6(F10.6), 2(F6.2) )  -> P2, E, F, G1, G2, H1, H2, Q1, Q2
FORMAT( F10.2, 6(F10.6) )          -> tb, I, J, K1, K2, K3, K4
FORMAT( 4(F10.6), 4(F9.4) )        -> L1, L2, L3, L4, R1, R2, R3, R4
FORMAT( I4, I5, I6 )                -> dat, EOPPWk, teff
```

These coefficients are used with the following equations to obtain estimates of the EOP parameters at any time (t – the current time in MJD).

$$\begin{aligned} x_p(t) &= A + B(t - ta) + \sum_{j=1}^2 C_j \sin\left(\frac{2p(t-ta)}{P_j}\right) + \sum_{j=1}^2 D_j \cos\left(\frac{2p(t-ta)}{P_j}\right) \\ y_p(t) &= E + F(t - ta) + \sum_{j=1}^4 G_j \sin\left(\frac{2p(t-ta)}{Q_j}\right) + \sum_{j=1}^4 H_j \cos\left(\frac{2p(t-ta)}{Q_j}\right) \quad (\text{D-1}) \\ \Delta UT1(t) &= I + J(t - tb) + \sum_{j=1}^4 K_j \sin\left(\frac{2p(t-tb)}{R_j}\right) + \sum_{j=1}^4 L_j \cos\left(\frac{2p(t-tb)}{R_j}\right) \end{aligned}$$

If we examine several runs of the data, we see the results in Fig. D-3. Note in particular the lines for $\Delta UT1$ in which the values increases over time. Extrapolating this for several months or a year is not recommended.

The assembly process uses files as discussed in Celestrak (Vallado and Kelso, 2005):

1. NGA coefficients for the upcoming week, updated weekly (Thursday). Each week a file of the form EOPPyddd.TXT is produced where y is the last digit of the year and ddd is the day of the year.
<ftp://164.214.2.65/pub/gig/pedata/EOPPyyyy/EOPPyddd.TXT>
2. Predicted data ($t_o + 180$ days) from IERS ($x, y, \Delta UT1, LOD, \delta\Delta\Psi, \delta\Delta\epsilon, (dX, dY)$, updated daily:
<http://hpiers.obspm.fr/eoppc/eop/eopc04/eopc04.62-now>
http://hpiers.obspm.fr/eoppc/eop/eopc04/eopc04_IAU2000.62-now
3. Predicted data (to +90 days) from Bulletin A ($x, y, \Delta UT1, LOD, \delta\Delta\Psi, \delta\Delta\epsilon, (dX, dY)$, updated daily:
<http://maia.usno.navy.mil/ser7/finals.daily>
<http://maia.usno.navy.mil/ser7/finals2000A.daily>
4. Predicted Data (to +1 year) from Bulletin A, updated weekly:
<http://maia.usno.navy.mil/ser7/finals.all>
<http://maia.usno.navy.mil/ser7/finals2000A.all>
5. TAI-UTC data, updated periodically with new leap seconds:
<http://maia.usno.navy.mil/ser7/tai-utc.dat>

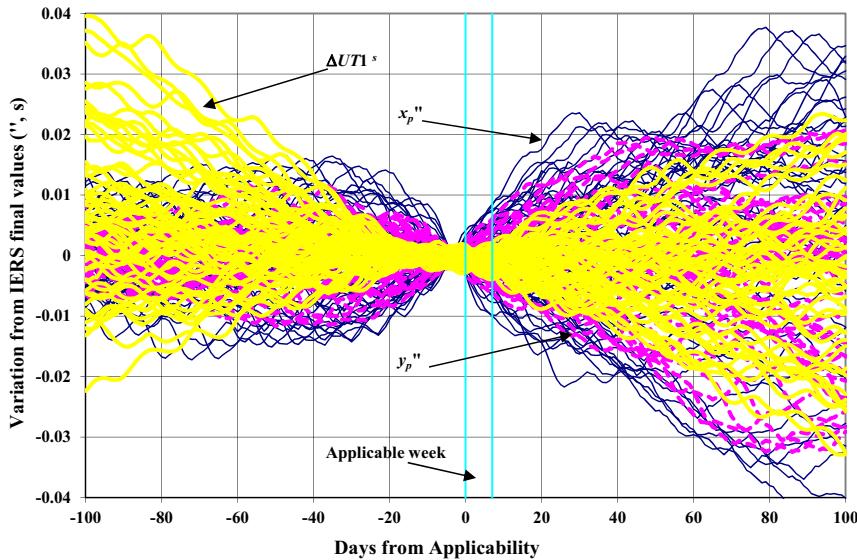


Figure D-2. Long-Term EOP Coefficient Performance. Numerous weekly coefficients were used to calculate values that were then compared to actual EOP measurements. Notice that the week before the week of applicability shows the best match - a consequence of least squares processing.

Atmospheric Parameters:

Space weather data are available through the National Geophysical Data Center (NGDC) in the National Oceanic and Atmospheric Administration (NOAA), but there are several files necessary to piece together a complete data file. Because 3-hourly data are available for the geomagnetic indices, a truly current file requires frequent retrieval and organization of the data. Figure D-3 shows the available data and their time of applicability.

NOAA provides actual and predicted atmospheric parameters. The files for a particular year are designated as yyyy where yyyy is the year. Data within the current year are indicated with a version extension for the month, for instance 2012.v5 includes data through the end of May in 2012. The data are given to within about one month of the current time. Note that the $F_{10.7}$ values are valid at 1700 UT to May 31, 1991. From that time to the present, the values are for 2000 UT, measured at the Penticton site in Canada. All data are adjusted to 1.0 AU.

For the last month or so, solar and geomagnetic data files are required. The quarterly solar data provides only the daily $F_{10.7}$ values. The geomagnetic data provide the 3-hourly K_p and daily a_p values. Notice that the 3-hourly a_p values are not given.

Predicted values of solar activity are very useful. The predictions for the next 45 days are found in the 45df.txt file. There are only single daily values, and no K_p values. Long-range predictions (about 6 years into the future) are found in the predict.txt file. Notice there are no long-range geomagnetic predictions, and the solar flux has only a single

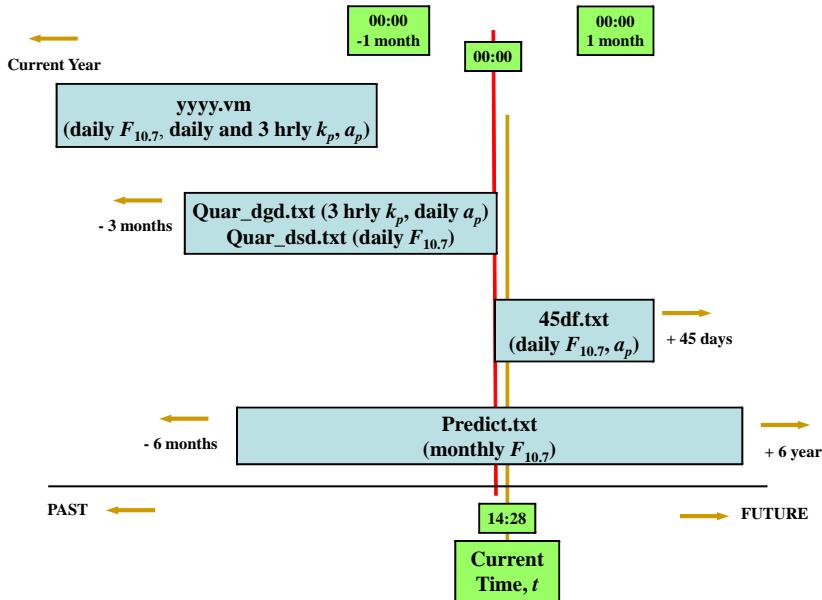


Figure D-3. Space Weather Parameter Data Availability. There are several sources of data from NGDC that we can use to populate a database of values for propagation into the future. Satellite operations require a seamless integration of this information.

value for each month. Given the difficulty in accurately predicting the geomagnetic indices, this is to be expected.

Similar to the Earth Orientation Parameters, we can present a list for assembling the space weather data. The assembly process is from Celestrak (Vallado and Kelso, 2005):

1. BSRN, ND, 3-hourly K_p , 3-hourly a_p , C_p , C9, ISN, $F_{10.7}\text{Adj}$, Q, updated monthly (a little over one month after end of the month). Each month a yyyy.vm file is produced, where yyyy is the year and m is the month. When the year is complete, the file becomes simply yyyy:
http://ftp.ngdc.noaa.gov/STP/GEOMAGNETIC_DATA/INDICES/KP_AP/yyyy.vm
2. $F_{10.7}\text{Obs}$ and $F_{10.7}\text{Adj}$ (in [Penticton_Adjusted](#)) from 1947 Jan 01 until end of previous month, updated monthly (end of month):
http://ftp.ngdc.noaa.gov/STP/SOLAR_DATA/SOLAR_RADIO/FLUX/Penticton_Observed/daily/DAILYPLT.OBS
3. 3-hourly K_p for the current quarter to date, updated every 3 hours.
http://www.spaceweather.noaa.gov/ftpdir/indices/quar_DGD.txt
4. $F_{10.7}\text{Adj}$ for the current quarter to date (previous quarters available) Files:
http://www.spaceweather.noaa.gov/ftpdir/indices/quar_DSD.txt
5. 45-day forecast of daily a_p and $F_{10.7}\text{Adj}$, updated daily:
<http://www.spaceweather.noaa.gov/ftpdir/latest/45DF.txt>
6. Monthly ISN and $F_{10.7}\text{Adj}$, for approximately two years, updated monthly:

<http://www.spaceweather.noaa.gov/ftpdir/weekly/Predict.txt>

The process of assembling these files requires attention to detail. The results of an automated procedure are located at <http://celesttrak.com/SpaceData/>. The space weather data is updated on CelesTrak every three hours at 35 minutes past the hour (e.g., 0035 UTC, 0335 UTC) because the 3-hourly data seem to come out at about 30 minutes past the hour. The Celestrak site provides nearly real-time updates for both sets of files. The formats are given at the top of each file and the files are simple ASCII data.

Solar irradiance values may be useful for some applications. Data exists at the following website:

<http://www.ngdc.noaa.gov/stp/solar/solarirrad.html>

Planetary Ephemerides:

Planetary ephemerides are available through the Jet Propulsion Laboratory (JPL). You can get them in binary or ASCII form through the Internet:

<ftp://ssd.jpl.nasa.gov/pub/eph/planets/ascii/>

ftp://naif.jpl.nasa.gov/pub/naif/generic_kernels/

The files are several megabytes each, and they are grouped by each ephemeris number, 403*.*, 405*.*, etc. A CD containing each ephemeris can be obtained from Willmabell publishing (www.willbell.com). Send requests for additional information to the JPL. The Minor planet center has information on observatories and celestial bodies.

<http://www.minorplanetcenter.net/iau/lists/ObsCodesF.html>

Geomagnetic Values:

Values for the magnetic field of the Earth are available through NASA's Goddard Space Flight Center. The current files are IGRF0x.dat and IGRF0xS.dat. The Internet site is:

ftp://nssdcftp.gsfc.nasa.gov/models/geomagnetic/igrf/fortran_code/

Initial Conditions:

General users may find two-line element set information from

<http://celesttrak.com/NORAD/elements>

Authorized users can also receive some data from AFSPC's website.

<https://www.space-track.org/auth/login>

You can send requests for 2-line element sets to Space Command. The web site contains an application form for special requests.

Precision Orbit Ephemerides:

Precision orbits are sometimes available from satellite owner operators, or for certain SLR-tracked satellites. Several useful sites are listed below.

<ftp://ftp.nga.mil/pub2/gps/>

<ftp://ftp.csr.utexas.edu/pub/>

<http://www.deos.tudelft.nl/ers/precorbs/orbits/>

Observational Data:

Raw observations are difficult to obtain. GPS and SLR data are quite useful.

<ftp://cddis.gsfc.nasa.gov/>
<http://ilrs.gsfc.nasa.gov/>
<ftp://ftp.csr.utexas.edu/pub/sl/>

Other useful sites:

There are numerous useful Internet sites that distribute computer code, initial conditions, reference material, etc. A few such sites are:

Apollo Lunar Surface Journal	http://www.hq.nasa.gov/alsj
Atmospheric Models	http://nssdcftp.gsfc.nasa.gov/models/atmospheric/ http://www.brodo.de/space/nrlmsise/
Constants	http://hpiers.obspm.fr/eop-pc/models/constants.html
Computer code, papers, data	http://netlib.bell-labs.com/netlib/
Technical papers	http://adsabs.harvard.edu/abstract_service.html http://adsabs.harvard.edu/article_service.html
Misc links (2005)	http://cstl-csm.semo.edu/cobb/ep461/Sat/Sat_main.html
Misc (subscription)	http://tbs-satellite.com/tse/online/

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.

I provide 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 (DO) construct. When you don't know the number of iterations, use a WHILE or a REPEAT-UNTIL (must execute at least one time) construct. Never use a jump to exit a control structure, it is perhaps the worst programming mistake that can be made.

Another common mistake in astrodynamics is the use of REAL variables in a boolean expression. For example, if the eccentricity of an orbit (ECC) is compared as IF ECC = 0, then the resulting boolean evaluation is always ambiguous. Consider the outcome if $e = 0.00001$, or 0.00000000 . Even if the boolean expression is changed to IF ECC = 0.00, the result is still ambiguous. The correct solution in all cases is to bound the test. IF ABS(ECC) < 0.000 000 1.*

A guideline for documentation is shown below. Each main program, procedure (subroutine), and function should have a header block that adequately defines the operation. Assume you [the programmer] will *not* be available for any consultation or questions. In addition, you should assume the future programmer will have *limited* background in astrodynamics!

```
{
-----  

| PROCEDURE <name>  

|  

| <Give a brief description of what the procedure accomplishes. This should  

| include an overview of the process used to solve the problem.>  

|  

| Algorithm :  

| <Describe unique computer science related operations of the routine. For  

| instance, discuss the way pointers operate, or what values are used as  

| exit criteria values. Simpler routines may not need this section.>  

|  

| Author : <who wrote it> <phone, fax, email,...> <Revision Date>  

|  

| Inputs : Description Range / Units Ex Value  

| Var 1 - <Describe what the var is> <Range and units> <example value>  

| LST - Local SIDEREAL Time -2Pi to 2Pi rad 181.690°  

|  

| OutPuts :  

| Var 2 - <Describe what the var is> <Range and units> <example value>  

|  

| Locals :  

| Var 3 - <Describe what the var is> <Range and units> <example value>  

|  

| Coupling :  

| - <List each external routine used by this procedure, and its function>
```

* Incidentally, no satellites have exactly zero eccentricity in the presence of perturbations— $e = 0.0001$ is about as small an eccentricity as you will observe.

```

|      CROSS          CROSS product of two vectors
|
| References   :
|   Author      - <List references for further information. Use pg numbers if possible>
|   Vallado    2007, 200
|
----- }

```

Variables. 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. Common sense is encouraged. Some programmers use all uppercase letters. Although some languages are case sensitive, mixing cases can help manage the overall length of variables (LonPer for instance instead of LONGITUDEOFPERIGEE. It is useful to keep all reserved words as capitals, except when a language (C) dictates otherwise. Finally, you should try to use variable names that are identical to the mathematical documentation. For instance, a ballistic coefficient could be BC.

I have also tried to make the variable names mean more than just the variable. For instance, I use "Dt" as the common variable for time deltas. This corresponds to equations in mathematical formulations where the time step is Δt . However, I supplement the variable name with the units so the programmer can tell what units are being used within a complex expression, or within a loop. Thus, "DtTU" is a time step in time units, while "DtDay" is a time step in days. Units are often the source of debugging errors, and this nomenclature can help reduce the problem.

You should declare all variables within a program. Global variables can be very difficult to trace, and implicitly defined variables can wreak havoc in program operation.

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. Typical routines are between 50 and 100 lines of computer code.

Aesthetics. 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. Today's computers can easily handle the increased space requirements. Also be sure to indent all control structures to enhance readability. Four spaces are usually adequate.

Efficiency. Tremendous effort can be expended trying to make a program run faster. Two axioms apply. First, computers will get faster and it doesn't matter. Second, it's too hard to get all the speed improvements. A compromise is the obvious solution because it can deliver substantial improvements in speed, but expends little additional effort. There are a few techniques.

- Declare one variable for a conversion that is used several times ($\text{TwoPi} = 2.0 * \pi$).
- Minimize the number of transcendental operations. Trigonometric calls are often the culprit, and they are easily replaced with temporary variables. For Multiplica-

tion and Division operations, multiplication is much more efficient, so we transform many operations ($x = x / 10 \Rightarrow x = x * 0.1$). Finally, compilers often use transcendental operations, even when simpler operations exist. For example, when using exponents, be sure to use integer values for integer evaluations. For example, $x = 3.4^2$, not $3.4^{2.0}$.

- You will achieve the best accuracy when evaluating polynomials if you sum from the smallest to the largest terms. This is especially useful in gravitational analysis, and reduction calculations.

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. There are three main considerations when considering testing computer code.

- Test each occurrence for a different dynamic situation. 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 is often not exercised fully. Avoid these shortcomings at all costs.
- Test each computational method. Exercise each structure (WHILE, IF, and so on) in the code. Instead of relying on testing only the dynamic situations above, fully exercise the *code itself*. Also extend the process to test each tolerance employed. Although correct control structures can eliminate many of these types of tests, adequate tolerance checking is necessary to ensure smooth performance over the range of possible values.
- As new operational situations and tests arise, develop new test cases to solve them. Actual operation over time is a valuable tool in refining computer code. These test cases can be incorporated into the standard test cases so subsequent upgrades can quickly match their answers. 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. A wide variety of programs exist for commercial use. One study lists features of several hundred computer programs that provide astrodynamical functions, Boelitz (1995). You should consult this comprehensive study to determine if any software exists that will solve your problems. The report has a few revisions from the Air Force Phillips Laboratory. Probably the most popular commercial programs are Analytical Graphics Inc. Satellite Toolkit (STK, www.agi.com), and the Aerospace Corporations Satellite Orbit Analysis Program (SOAP, soap@zoom.aero.org). I’ve listed some of the better known numerical orbit determination systems in Table D-10. These programs are not generally for commercial sale (except for Free Flyer and ODTK), but their use and long-standing accuracy make them valuable tools for reference. This list is not intended to be exhaustive.

TABLE D-10. Orbit Determination Programs. This table lists some commercial and legacy orbit determination programs. These programs are generally cited in performance comparisons of newer systems as the “truth” simulation. The programs are listed alphabetically by name. The commercial programs have generally comparable accuracy, with specialization in one or more areas. The last 3 programs are military and they concentrate on the problem of generating a complete space object catalog. Organizational acronyms include Goddard Space Flight Center (GSFC), Massachusetts Institute of Technology (MIT), and the Jet Propulsion Laboratory (JPL).

Name	Organization	Comments
DYNAMO	MIT, Lincoln Laboratory	Specializes in Deep Space (GEO) orbits
Free-flyer	ai-Solutions	Batch Least Squares OD
GEODYN	NASA / GSFC	Precision Geodesy applications
GIPSY- OASIS	JPL / NASA	Satellite and Interplanetary
GTDS	Nasa / GSFC	Operational version
GTDS	Draper Laboratory	Includes DSST, Semianalytical technique*
HPOP (STK) and ODTK	Analytical Graphics Inc.	Complete graphical displays and user interface. Kalman filter OD
OCEAN	Naval Research Lab	Precise OD and propagation
Geodynam- ics TRACE	Raytheon	OD, propagation and analysis
TRACE	Aerospace Corporation	Many analysis capabilities
UTOPIA	UT Austin, Center for Space Research	Precise Gravity field development TOPEX POE generation
Special-K	US Navy	Operational System
SPADOC	US Air Force	Operational System
ASW	US Air Force	Workstation System

* Note that basic versions of DSST are being made available from the Open Source Software initiatives at <http://www.orekit.org> and <http://tastrody.unirioja.es>.

References

- The numbers in brackets after each reference indicate the chapter in which they 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. [6]
- Abramowitz, M., and I. Stegun. 1970. *Handbook of Mathematical Functions*. New York: Dover Publications. [6]
- Akella, Maruthi R., and Kyle T. Alfriend. 2000. Probability of Collision Between Space Objects. *Journal of Guidance, Control, and Dynamics*. 23(5): 769-772. [11]
- Alfano, Salvatore. 1982. Low Thrust Orbit Transfer. M.S. Thesis, Department of Aeronautics and Astronautics. Ohio: Air Force Institute of Technology. [6]
- _____. 1991. Personal discussions. U.S. Air Force Academy, CO. [5]
- _____. 1993. Rapid Generation of Site/Satellite Timetables. *Journal of Spacecraft and Rockets*. 30(6): 760-764. [11, C]
- _____. 1994. Determining Satellite Close Approaches. Part II. *Journal of the Astronautical Sciences*. 42(2): 143-152. [11, C]
- _____. 2000. Personal Discussions. [11]
- _____. 2004. Orbital Covariance Interpolation. Paper AAS-04-223 presented at the AAS/AIAA Spaceflight Mechanics Conference. Maui, HI. [10, 11]
- _____. 2005. Relating Position Uncertainty to Maximum Conjunction Probability. *Journal of the Astronautical Sciences*. 53(2): 193-205. [11]
- _____. 2006. Satellite Collision Probability Enhancements. *Journal of Guidance, Control, and Dynamics*. 29(3): 588-592. [11]
- _____. 2007. Beta Conjunction Analysis Tool. Paper AAS 07-393 presented at the AAS/AIAA Astrodynamics Specialist Conference. Mackinac Island, MI. [11]
- Alfano, Salvatore, and Meredith Greer. 2003. Determining If Two Solid Ellipsoids Intersect. *Journal of Guidance, Control, and Dynamics*. 26(1): 106-110. [11]
- 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. [11, C]
- Alfano, Salvatore, David Negron, Jr., and Daniel D. Wright. 1992. The Method of Ratios. *Journal of the Astronautical Sciences*. 40(2): 297-309. [11]
- Alfano, Salvatore, and David Negron, Jr. 1993. The Complete Method of Ratios. *Journal of Spacecraft and Rockets*. 30(6): 374-379. [11, C]
- _____. 1993. Determining Satellite Close Approaches. *Journal of the Astronautical Sciences*. 41(2): 217-225. [11]
- Alfano, Salvatore, and James D. Thorne. 1994. Circle to Circle Constant-thrust Orbit Raising. *Journal of the Astronautical Sciences*. 42(1): 35-45. [6]
- Alfriend, Kyle T., et al. 1998. Single Track Covariance for Space Surveillance Network Sensors. Paper AAS-98-115 presented at the AAS/AIAA Spaceflight Mechanics Conference. Monterey, CA. [10]
- Alfriend, Kyle T. 2006. Personal Discussions. [11]

- 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. [9]
- Almanac for Computers 1990*. Washington, DC: U.S. Government Printing Office. [3]
- Amelina, T., et al. 1996. Comparison of Atmospheric Density Models. Paper presented at the US-Russia Space Surveillance Workshop. Poznan, Poland. [8, B]
- ANSI/AIAA S-131-2010, Astrodynamics – Propagation Specifications, Technical Definitions, and Recommended Practices, Aug 25, 2010.
- Aoki, S. et al. 1982. The New Definition of Universal Time. *Astronomy and Astrophysics*. Vol. 105: 359-361. [3]
- Arora, Nitin. and Ryan P. Russell. 2011. Fast, Efficient, and Adaptive Interpolation of the Geopotential. Paper AAS 11-501 presented at the AAS/AIAA Astrodynamics Specialist Conference. July 31-August 4, Girdwood, AK. [9]
- Ashby, N., and J. J. Spilker Jr. 1996. Global Positioning System. Theory and Applications. AIAA Progress in Astronautics and Aerodynamics. 163:623-697. [11]
- 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, 3, 4]
- Atkinson R d'E, and D. H. Sadler. 1951. On the use of Mean Sidereal Time. Monthly Newsletters of the Royal Astronomical Society. 111:619-623. [3]
- Baker, Robert M., and Maud W. Makemson. 1967. *An Introduction into Astrodynamics*. New York: Academic Press. [7, 8]
- Baker, Robert M. 1967. *Astrodynamics, Applications and Advanced Topics*. New York: Academic Press. [7, 8, 9, 10]
- 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 Specialist Conference. Halifax, Nova Scotia, Canada. [9]
- Barlier, F., et al. 1978. A Thermospheric Model based on Satellite Drag Data. *Annales de Geophysics*. 34(1): 9-24. [8]
- Baron, Sally. 2005. Trust, Risk, and Aerospace Software—Balance and Optimization. Paper AAS 05-202 presented at the AAS/AIAA Spaceflight Mechanics Conference, January 23-27, 2005. Copper Mountain, CO. [11]
- 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. [8]
- Bate, Roger R., D. D. Mueller, and J. E. White. 1971. *Fundamentals of Astrodynamics*. New York: Dover Publications. [2, 6, 7, 8, 9]
- Battin, Richard H. 1987, 1999 2nd ed. *An Introduction to the Mathematics and Methods of Astrodynamics*. New York: AIAA Education Series. [2, 7, 8, 9, 10]
- Beer, Arthur and K. A. Strand. 1975a. *Copernicus*. Oxford: Pergamon Press. [1, 2]
- . 1975b. *Kepler*. Oxford: Pergamon Press. [1, 2, 4, 7]
- Bell, Eric T. 1965. *Men of Mathematics*. New York: Simon & Schuster Inc. [7, 8, 9, 10]

- Bergstrom, S. 2002. An Algorithm for Reducing Atmospheric Density Model Error Using Satellite Observation Data in Real-time. M.S. Thesis. Department of Aeronautics and Astronautics, Massachusetts Institute of Technology. [8]
- Berry, Arthur. 1961. *A Short History of Astronomy*. New York: Dover Publications. [1, 2, 3]
- Berry, Matthew, and Liam Healy. 2001. Implementation of Gauss-Jackson Integration for Orbit Propagation. Paper AAS-01-426 presented at the AAS/AIAA Astrodynamics Specialist Conference. Quebec City, Canada. [8]
- _____. 2002. The Generalized Sundman Transformation For Propagation of High Eccentricity Elliptical Orbits. Paper AAS-02-109 presented at the AAS/AIAA Spaceflight Mechanics Conference. San Antonio, TX. [8]
- _____. 2003a. Comparison Of Accuracy Assessment Techniques For Numerical Integration. Paper AAS-03-171 presented at the AAS/AIAA Spaceflight Mechanics Conference. Ponce, Puerto Rico. [8]
- _____. 2003b. Accuracy and Speed Effects of Variable Step Integration for Orbit Determination and Propagation. Paper AAS-03-664 presented at the AAS/AIAA Astrodynamics Specialist Conference. Big Sky, MT. [8]
- _____. 2004. A Variable Step Double Integration Multi-Step Integrator. Paper AAS-04-238 presented at the AAS/AIAA Spaceflight Mechanics Conference. Maui, HI. [8]
- Berry, Matthew M. 2004. A Variable-step Double-integration Multi-step Integrator. Ph.D. Dissertation. Virginia Polytechnic Institute and State University. [8]
- Bierman, Gerald. 1977. *Factorization Methods for Discrete Sequential Estimation*. New York: Academic Press. [10]
- Blitzer, Leon. 1970. *Handbook of Orbital Perturbations*. Unpublished report. TRW, Inc. [8, 9]
- Boelitz, Carole. 1992. Confidence in Atmospheric Density Forecasts. Paper AAS-92-181 presented at the AAS/AIAA Spaceflight Mechanics Meeting, Colorado Springs, CO. [8]
- _____. 1995. Orbit Analysis Software Survey. Phillips Laboratory Technical Report. PL-TR-95-1139. [D]
- Boers, Jonathan, et al. 2000. Accuracy Assessment of the Naval Space Command Special Perturbations Cataloging System. Paper AAS-00-183 presented at the AAS/AIAA Spaceflight Mechanics Conference. Clearwater, FL. [10]
- 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. [8]
- Bowman, B. 1971. A First Order Semi-Analytic Perturbation Theory for Highly Eccentric 12 Hour Resonating Satellite Orbits. NORAD document, Colorado Springs. [9]
- Bowman, B. et al. 2006. A New Empirical Thermospheric Density Model JB2006 Using New Solar Indices. Paper AIAA-2006-6166 presented at the AIAA/AAS Astrodynamics Specialist Conference. Keystone, CO. [8]
- Bretagnon, P. 1982. Théorie du mouvement de l'ensemble des planètes Solution VSOP82. *Astronomy and Astrophysics*. Vol. 114: 278-288. [5]
- Bretagnon, P., and G. Francou. 1988. Planetary Theories in Rectangular and Spherical Variables, VSOP87 Solutions. *Astronomy and Astrophysics*. Vol. 114: 309-315. [5]

- 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. [9]
- Brouwer, Dirk, and G. M. Clemence. 1961. *Methods of Celestial Mechanics*. New York: Academic Press, Inc. [1, 2, 8, 9]
- 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. [9]
- Brumberg, Victor A., and Eugene V. Brumberg. 1999. *Celestial Dynamics at High Eccentricities*. New York: Gordon and Breach Science Publishers. [11]
- Bryson, Arthur E., and Y. Ho. 1987. *Applied Optimal Control*. Washington, DC: Hemisphere Publishing Corporation. [6, 10]
- Buhler, Walter K. 1981. *Gauss, A Biographical Study*. New York: Springer-Verlag. [7, 11]
- Burns, Rich, et al. 2000. Solar Radiation Pressure Effects on Formation Flying of Satellites with Different Area to Mass Ratios. Paper AIAA-2000-4132 presented at the AIAA/AAS Astrodynamics Specialist Conference. Denver, CO. [8]
- Burnham, W. F., and R. Sridharan. 1995 (March 30). Geosynchronous Surveillance with Space Based Sensor. Paper STK-235, Vol. 1 presented at the Space Surveillance Workshop at MIT Lincoln Laboratory, MA [4]
- Burton, James R., and Robert G. Melton. 1992. Variation of Parameters using Battin's Universal Functions. Paper AAS-92-125 presented at the AAS/AIAA Spaceflight Mechanics Meeting. Colorado Springs, CO. [9]
- 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 Specialist Conference. Halifax, Nova Scotia, Canada. [7]
- Cappellucci, David A. 2005. Special Perturbations to General Perturbations Extrapolation Differential Correction in Satellite Catalog Maintenance. Paper AAS 05-402 presented at the AAS/AIAA Spaceflight Mechanics Conference, January 23-27, 2005. Copper Mountain, CO. [10]
- Capitaine, N., and A. M. Gontier. 1993. Accurate Procedure for deriving UT1 at a sub-milli-arc-second Accuracy from Greenwich Sidereal Time or from the Stellar Angle. *Astronomy and Astrophysics*. Vol. 275: 645-650. [3]
- Capitaine, N., P. T. Wallace and J. Chapront. 2003. Expressions for IAU 2000 Precession Quantities. *Astronomy and Astrophysics*. Vol. 412: 567-586. [3]
- Caprara, Giovanni. 1986. *The Complete Encyclopedia of Space Satellites*. New York: Portland House. [11]
- Caputo, Michele. 1967. *The Gravity Field of the Earth*. New York: Academic Press. [8]
- Carlton-Wipperf, Kitt C. 1993. An Introductory Analysis of Satellite Collision Probabilities. Paper AAS-93-748 presented at the AAS/AIAA Astrodynamics Specialist Conference. Victoria, BC, Canada. [11]
- 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. [8, B]
- Carter, David. 1991. When is the Groundtrack drift rate zero? CSDL Memorandum ESD-91-020. Cambridge, MA: Charles Stark Draper Laboratory. [11]

- _____. 1996. Personal correspondence. [7]
- Casotto, Stefano. 1989. Nominal Ocean Tide Models for TOPEX Precise Orbit Determination. Ph.D. Dissertation. University of Texas at Austin. [8, 9]
- Cayley, A. 1861. Tables of the Developments of Functions in the Theory of Elliptic Motion. *Memoires of the Royal Astronomical Society*. Vol. 29: 191-306. [9]
- 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. [9]
- _____. 1997. Personal Discussions. [9]
- 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. [9]
- 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. [9]
- 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. [9]
- Cefola, Paul J., and D. J. Fonte. 1996. Extension of the Naval Space Command Satellite Theory to include a General Tesselar m-daily Model. Paper AIAA-96-3606 presented at the AIAA/AAS Astrodynamics Specialists Conference. San Diego, CA. [9]
- 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/AIAA Spaceflight Mechanics Meeting. Austin, Texas. [9]
- Cefola, P. J., and A. I. Nazarenko, et al. 1999. Neutral Atmosphere Density Monitoring Based on Space Surveillance System Orbital Data. Paper AAS-99-383 presented at the AAS/AIAA Astrodynamics Specialist Conference. Girdwood AK. [8]
- Chao, C. C., et al. 2000. *Independent Verification and Validation for Analytical Graphics Inc. of three Astrodynamical Functions of the Satellite Toolkit*, Ver 4.1.0. Aerospace TR 2000 (7605)-1. El Segundo, CA. [8]
- Chao, (George) Chia-Chun. 2005. *Applied Orbital Perturbations and Maintenance*. AIAA Education Series. New York: AIAA Press. [9, 11]
- 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. [8]
- Chan, Ken. 2001. A Simple Mathematical Approach for Determining Intersection of Quadratic Surfaces. Paper 01-358 presented at the AAS/AIAA Astrodynamics Specialist Conference. Quebec City, Canada. [11]
- _____. 2003. Improved Analytical Expressions for Computing Spacecraft Collision Probabilities. Paper AAS-03-184 presented at the AAS/AIAA Spaceflight Mechanics Conference. Ponce, Puerto Rico. [11]
- _____. 2009. *Spacecraft Collision Probability*. The Aerospace Press. El Segundo, CA. [11]
- Chetty, P. R. K. 1988. *Satellite Technology and its Applications*. Blue Ridge Summit, PA. Tab Books. [7]

- Chilulli, Roy M. 1994. *International Launch Site Guide*. The Aerospace Press. El Segundo, CA. [6]
- Cioma, Vince. 1997. Orbit Estimation using Track Compression and Least Squares Differential Correction. M.S. Thesis. GA/ENY/97-D-1. Dayton, OH: Air Force Institute of Technology. [7, 10]
- . 2001. Personal correspondence. [7]
- Chobotov, Val A. 2006. *Orbital Mechanics*. 2nd edition, 3rd printing. AIAA Education Series. New York: AIAA Press. [2, 9]
- Clohessy, W. H., and R. S. Wiltshire. 1960. Terminal Guidance System for Satellite Rendezvous. *Journal of the Aerospace Sciences*. 27(9): 653-674. [6]
- Coffey, Shannon, and A. Deprit. 1980a. Short Period Elimination in Satellite Theory. Paper AIAA 80-1657. *Journal of Guidance and Control*. [9]
- . 1980b. Fast Evaluation of Fourier Series. *Astronomy and Astrophysics*. Vol. 81: 310-315. [3]
- Colliers Encyclopedia*. 1992. Vol. 19. New York: MacMillan Educational Co. [3]
- Cook, G. E. 1962. Luni-Solar Perturbations of the Orbit of an Earth Satellite. *Geophysical Journal of the Royal Astronomical Society*. Vol. 6: 271. [9]
- Cook, G. E., and D. W. Scott. 1967. Lifetimes of Satellites in Large Eccentricity Orbits. *Planetary Space Science*. Vol. 15: 1549-1556. [9]
- Coppola, Vincent, John H. Seago and David A. Vallado. 2009. The IAU 2000a and IAU 2006 Precession-Nutation Theories and their Implementation. Paper AAS 09-159 presented at the AAS/AIAA Spaceflight Mechanics Conference, February 9-12. Savannah, GA. [3]
- Coppola, Vincent, Sylvain Dupont, Kevin Ring, and Frank Stoner. 2009. Assessing Satellite Conjunctions for the Entire Space Catalog using COTS Multi-core Processor Hardware. Paper AAS-09-374 presented at the AAS/AIAA Astrodynamics Specialist Conference, August 9-13. Pittsburgh, PA. [11]
- Cortright, Edgar M. 1975. Apollo Expeditions to the Moon. NASA SP-350. Washington DC. [12]
- COSPAR Working Group. 1965. *COSPAR International Reference Atmosphere*. Amsterdam: North Holland Pub. Co. [8]
- Coster, A. J., R. Abbot, L.E. Thornton and D. Durand. 2001 (April). Exploiting AFSCN Ranging Data for Catalog Maintenance. Paper STK-256, presented at the Space Surveillance Workshop at MIT Lincoln Laboratory, MA [4]
- Curtis, Anthony R. 1994. *Space Satellite Handbook*. Houston, Texas: Gulf Publishing Co. [11]
- Cutting, E., G. H. Born, J. C. Frautnick. 1978. Orbit Analysis for SEASAT-A. *Journal of the Astronautical Sciences*. 26(4): 315-342. [11]
- Danby, John M. A. 1992. *Fundamentals of Celestial Mechanics*. Richmond VA: Willmann Bell Inc. [1, 2, 8]
- Danielson, Donald A. 2003 2nd ed. *Vectors and Tensors in Engineering and Physics*. Reading, MA: Addison-Wesley Publishing. [3]
- 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. [9]

- Danielson, Donald A. et al. 1995. *Semianalytical Satellite Theory (SST): Mathematical Algorithms*. Technical Report NPS-MA-95-002. Monterey, CA: Naval Postgraduate School. [9]
- Davis, G. W., and 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 Astrodynamics Specialist Conference. Victoria, BC, Canada. [7]
- Defense Mapping Agency. 1983. *Geodesy for the Layman*. DMA-TR 80-003. Washington, DC: Headquarters, Defense Mapping Agency. [3]
- . 1987a. *Department of Defense World Geodetic System 1984 DMA*. DMA-TR 8350.2. Washington, DC: Headquarters, Defense Mapping Agency. [3]
- . 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. [3]
- de La Cotardière, Philippe. 1987. *Astronomy*. Facts on File Inc. New York, New York. [5]
- de Lafontaine, Jean. 1986. Orbital Dynamics in a Stochastic Atmosphere and a Nonspherical Gravity Field. Ph.D. Dissertation. University of Toronto Canada: Institute for Aerospace Studies. [8]
- Deprit, Andre. 1969. Canonical Transformations Depending on a Small Parameter. *Celestial Mechanics*. 1(1): 12-30. [9]
- . 1981. The Main Problem in the Theory of Artificial Satellites to Order Four. *Journal of Guidance and Control*. 4(2): 201-206. [9]
- 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. [9]
- 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. [8]
- . 2012. Private communication. [7]
- Der, Gim, and Roy Danicheck. 1996. Conversion of Osculating Orbital Elements to Mean Orbital Elements. Internal TRW report. Publication Pending. [10]
- 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. [9]
- 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. [9]
- Edelbaum, Theodore N. 1961. Propulsion Requirements for Controllable Satellites. *American Rocket Society Journal*. Vol. 31: 1079-1089. [6]
- 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. [6]
- 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 Astrodynamics Specialist Conference. Scottsdale, AZ. [8]

- Escobal, Pedro R. [1965] 1985. *Methods of Orbit Determination*. New York: John Wiley & Sons. (Reprint edition. Malabar FL: Krieger Publishing Co.) [1, 2, 3, 8, 9, 10, 11, C]
- . [1968] 1979. *Methods of Astrodynamics*. New York: John Wiley & Sons. Reprint: Malabar FL: Krieger Publishing Co. [6, 7, 8]
- Explanatory Supplement to the Astronomical Ephemeris and the American Ephemeris and Nautical Almanac*. 1961. London: Her Majesty's Stationery Office. [3]
- Fairhead, L., and P. Bretagnon. 1990. An Analytic Formula for the Time Transformation TB - TT, *Astronomy and Astrophysics*. v229, pp. 240-247. [3]
- Fehlberg, Erwin. 1968. Classical Fifth-, Sixth-, Seventh-, and Eighth-Order Runge-Kutta Formulas with Stepsize Control. NASA Technical Report, TR-R-287. [8]
- . 1969. Low-Order Classical Runge-Kutta Formulas with Stepsize Control and their Application to some Heat Transfer Problems. NASA Technical Report, TR-R-315. [8]
- Ferguson, Jack. 1978. *Statistical Orbit Determination*. Astro 422 Course reading. Colorado Springs, CO: US Air Force Academy. [10]
- Fieger, Martin. 1987. An Evaluation of Semianalytical Satellite Theory Against Long Arcs of Real Data for Highly Eccentric Orbits. M. S. Thesis, CSDL-T-938. Department of Aeronautics and Astronautics, Massachusetts Institute of Technology. [9]
- Finkelman, David, and Daniel Oltrogge. 2006. Progress in International Space and Astrodynamics Standards. Paper AAS 06-234 presented at the AAS/AIAA Spaceflight Mechanics Conference, January 22-26, 2006. Tampa, FL. [11]
- Fischer, Jack. 1998. The evolution of Highly Eccentric Orbits. M. S. Thesis, CSDL-T-1310. Department of Aeronautics and Astronautics, Massachusetts Institute of Technology. [10]
- Fisher, R. A. 1912. On an Absolute Criterion for Fitting Frequency Curves. *Messenger of Mathematics*. Vol. 41: 155. [10]
- Fitzpatrick, Philip M. 1970. *Principles of Celestial Mechanics*. New York: Academic Press. [8, 9]
- Fliegel, H. F., et al. 1992. Global Positioning System Radiation Force Model for Geodetic Applications. *Journal of Geophysical Research*. 97(B1): 559-568. [8, 9]
- Flohrer, Tim, et al. 2009. Improving ESA'S Collision Risk Estimates by an Assessment of the TLE Orbit Errors of the US SSN Catalogue. Paper S4.4 presented at the 5th European Space Debris Conference. Darmstadt, Germany. 30 March - 2 April 2009. [10]
- Fonte, Daniel J. Jr. 1993. Implementing a 50×50 Gravity Field Model in an Orbit Determination System. M.S. Thesis, CSDL-T-1169. Department of Aeronautics and Astronautics, Massachusetts Institute of Technology. [9]
- . 1994a. PC Based Orbit Determination. Paper AIAA 94-3776 presented at the AIAA/AAS Astrodynamics Specialist Conference. Scottsdale AZ. [10]
- . 1994b. Evaluation of Orbit Propagators for the HI-CLASS Program. Technical Report PL-94-1017. Kirtland Air Force Base, NM: Phillips Laboratory. [9]
- . 1995 (March 30). HI-CLASS. Paper STK-235, Vol. 1 presented at the Space Surveillance Workshop at MIT Lincoln Laboratory, MA [10]
- 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. [10]
- Frank, Mark. 1984. *The New FK5 Mean of J2000 Time and Reference Frame Transformations*. DVO Technical Report 84-2. [3]

- Fraser-Smith, A. C. 1972. Spectrum of the Geomagnetic Activity Index A_p . *Journal of Geophysical Research*. 77(22): 4209-4220. [8]
- Fricke, W. 1982. Determination of the Equinox and Equator of the FK5. *Astronomy and Astrophysics*, Vol. 107: L13-L16. [3]
- Foster, J. L. Jr., and Frisbee. 1998. Position Error Covariance Matrix Scaling Factors for Early Operational ISS Debris Avoidance. NASA, Johnson Space Center/DM33. [11]
- Fujimoto, Kohei, and Daniel J. Scheeres. 2010. Correlation of Covariance Based track Association Approaches with Simulated Radar Data. Paper AAS 10-319 presented at the AAS/AIAA Spaceflight Mechanics Conference. San Diego, CA. [4]
- Gabor, Michael J. 1993. Orbit Determination for the GEOSAT Exact Repeat Mission Utilizing TRANET Doppler Data. M.S. Thesis. University of Texas at Austin. [10]
- . 1999. GPS Carrier Phase Ambiguity Resolution Using Satellite-Satellite Single Differences. Ph.D. Dissertation. University of Texas at Austin. [11]
- Gammon, Richard H. *Chemistry Between the Stars*. U.S. Government Printing Office: September 1976. [4]
- Gaposchkin, E. M. 1994. Calculation of Satellite Drag Coefficients. Technical Report 998. MIT Lincoln Laboratory, MA. [10]
- 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. [9]
- . 1987. Evaluation of Recent Atmospheric Density Models. Paper AAS-87-557. *Proceedings of 1987 AAS/AIAA Astrodynamics Conference*. San Diego, CA: AAS Publications Office. [9]
- 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, 8]
- Gedeon, G. S. 1968. *Luni-Solar Perturbations*. TRW Report No. 3414.5-59. Redondo Beach, CA: TRW, Inc. [9]
- Gelb, Arthur. 1989. *Applied Optimal Estimation*. Cambridge, MA: MIT Press. [10]
- Geyling, Franz T., and R. H. Westerman. 1971. *Introduction to Orbital Mechanics*. Reading, MA: Addison-Wesley Publishing. [8, 9]
- Gibbs, Josiah W. 1889. On the Determination of Elliptic Orbits from Three Complete Observations. *Memoires National Academy of Science*: 4 (2): 79-104. [7]
- Gim, Dong-Woo, and Kyle T. Alfriend. 2003. State Transition Matrix of Relative Motion for the Perturbed Noncircular Reference Orbit. *Journal of Guidance, Control, and Dynamics*. 26(6): 956-971. [6]
- Giorgini, Jon, et al. 1995. Magellan Aerobrake Navigation. *Journal of the British Interplanetary Society*. (48) 111-122. [12]
- Gooding, R. H. 1988. On the Solution of Lambert's Orbital Boundary-Value Problem. Royal Aerospace Establishment. Technical Report TR 88027. Farnborough, Hampshire. [7]
- . 1990. A Procedure for the Solution of Lambert's Orbital Boundary-Value Problem. *Celestial Mechanics and Dynamical Astronomy*. 48(2): 145-165. [7]
- . 1993. A New Procedure for Orbit Determination Based on Three Lines of Sight (Angles-only). Defense Research Agency DRA TR 93004. Farnborough, Hampshire. [7]

- . 1997. A New Procedure for the Solution for the Classical Problem of Minimal Orbit Determination from Three Lines of Sight. *Celestial Mechanics and Dynamical Astronomy*. 66(1): 387-423. [7]
- Gorochov, Y. P. and Nazarenko, A. I. 1982. Methodical Points in Building Models of the Fluctuation of the Atmosphere Parameters. Astronomicheskii Sovet Akademii Nauk SSSR, Vol. 80. A copy was obtained by the MIT Lincoln Laboratory Library in January 2004 from the CISTI Document Delivery Service and translated by MIT Aeronautics and Astronautics Department graduate student Kalina Galabova in February 2004. [8]
- Granholm, George. 2000. Near-real time Atmospheric Density Correction Using Space Catalog Data. M.S. Thesis, CSDL-T-1380. Department of Aeronautics and Astronautics, Massachusetts Institute of Technology. [8]
- Granholm, George, and Paul Cefola, Andrey Nazarenko, Vasilii Yusarov. 2000. Near-real time Atmospheric Density Correction Using NAVSPASUR Fence Observations. Paper AAS-00-179 presented at the AAS/AIAA Spaceflight Mechanics Conference. Clearwater, FL. [8]
- Green, Andrew Joseph. 1979. Orbit Determination and Prediction Processes for Low Altitude Satellites. MIT Ph.D. Dissertation. #CSDL-T-703. Cambridge, MA: Charles Stark Draper Laboratory, Inc. [10]
- Green, Robin M. 1988. *Spherical Astronomy*. New York: Cambridge University Press. [3, 5]
- Griffin, Michael D., and James R. French. 1991. *Space Vehicle Design*. Washington DC: AIAA Press. [1]
- Guinot, B. 1979. Basic Problems in the Kinematics of the Rotation of the Earth. A79-53001 24-89. Time and the Earth's rotation-Proceedings of the Eighty-Second Symposium. San Fernando, Spain. 7-18. [3]
- 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. [9]
- Halsell, C. Allen, et al. 2003. Trajectory Design for the Mars Reconnaissance Orbiter Mission. *Advances in the Astronautical Sciences*. (114) AAS-03-211. [12]
- 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 Sachsischen Gesellschaft für Wissenschaft*. 2(3): 183-281. [9]
- 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. [8]
- Hart, Andrew. 1991. Application of the Homotopy Continuation Method to Low-Eccentricity Preliminary Orbit Determination. M.S. Thesis. Massachusetts Institute of Technology. [7]
- Hartman, Paul G. 1993. "Long-term SGP4 Performance." Space Control Operations Technical Note J3SOM-TN-93-01. US Space Command, USSPACECOM/J3SO. Colorado Springs, CO. [9]
- Hathaway, David H. 2010. The Solar Cycle. *Living Rev. Solar Physics*. v7:1. [8]
- Haupt, Randy L., and Sue Ellen Haupt. 1998. *Practical Genetic Algorithms*. NY: John Wiley & Sons, Inc. [7]
- Healy, Liam M. 1995. Close Conjunction Detection on Parallel Computer. *Journal of Guidance, Control, and Dynamics*. 18(4): 824-829. [11]
- 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. [8]

- Hedin, A. E. 1987. MSIS-86 Thermospheric Model. *Journal of Geophysical Research*. Vol. 92: 4649-4662. [8]
- Herrick, Samuel. 1971. *Astrodynamicics: Orbit Determination, Space Navigation, Celestial Mechanics* (Vol. I). London: Van Nostrand Reinhold Co. [2, 7, 8]
- . 1972. *Astrodynamicics: Orbit Correction, Perturbation Theory, Integration* (Vol. II). London: Van Nostrand Reinhold Co. [8, 9]
- 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. [9]
- Hiebert, Darren. 2011. Personal communication. [5]
- 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. [6]
- Hill, Louis A., Jr. 1981. *Structured Programming in FORTRAN*. Englewood Cliffs, NJ: Prentice Hall. [C]
- Hill, Keric, Kyle T. Alfriend, and Chris Sabol. 2008. Covariance-based Uncorrelated Track Association. Paper AIAA 2008-7211 presented at the AAS/AIAA Astrodynamics Specialist Conference, August 18-21. Honolulu, HI. [6]
- Hill, Keric., et al. 2010. Covariance Based Network Tasking of Optical Sensors. Paper AAS 10-150 presented at the AAS/AIAA Space Flight Mechanics Conference, February 14-17. San Diego, CA. (Also Paper USR 10-S5.2 presented at the US/Russian Space Surveillance Workshop, April 18-23, 2010. Maui, HI.) [4, 11]
- Hill, K. 1995. *Matrix-Based Ellipse Geometry, GraphicsGems V*. Academic Press, New York. [11]
- Hilton, C. G. 1963. "The SPADATS Mathematical Model." Report ESD-TDR- 63-427, Aeronutronic Publ. U-2202. [9]
- Hilton, C. G., and J. R. Kuhlman. 1966. "Mathematical Models for the Space Defense Center." Philco-Ford Publication No. U-3871, 17-28. [9]
- Hobson, E. W. [1931] 1965. *The Theory of Spherical and Ellipsoidal Harmonics*. New York: Chelsea Publishing. [8]
- Hofmann-Wellenhof, B., H. Lichtenegger, and J. Collins. 1994. *GPS Theory and Practice*. New York: Springer-Verlag. [11]
- 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. [6]
- Hoots, Felix. R. 1982. An Analytical Satellite Theory using Gravity and a Dynamic Atmosphere. Paper AIAA-82-1409 presented at the AIAA/AAS Astrodynamics Specialist Conference. San Diego, CA. [9]
- . 1998. A history of Analytical Orbit Modeling in the United States Space Surveillance System. Third US/Russian Space Surveillance Workshop. Washington, D.C. [9]
- . 2013. Semi-Analytical Orbit Prediction Model - Radiation Pressure Effects. Aerospace Report ATM-2013(8265)-1. Chantilly, VA. [9]
- 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, 9]

- 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. [11]
- Hoots, Felix R., and R. G. France. 1987. An Analytic Satellite Theory using Gravity and a Dynamic Atmosphere. *Celestial Mechanics*. 40(1): 1-18. [9]
- Hoots, F. R., P. W. Schumacher, and R. A. Glover. 2004. History of Analytical Orbit Modeling in the U. S. Space Surveillance System. *Journal of Guidance, Control, and Dynamics*. 27(2): 174-185. [9]
- Hovanessian, S. A. 1988. *Introduction to Sensor Systems*. Norwood, MA: Artech House. [7]
- Hujasak, Richard. 1979. A Restricted Four-body Solution for Resonating Satellites with an Oblate Earth. AIAA Paper 79-136. [9]
- . 1994. Gravity Process Noise for Constellations of Satellites. Paper AAS-94-175 presented at the AAS/AIAA Spaceflight Mechanics Meeting. Cocoa Beach, FL. [10]
- . 1995. Using a Single SLR Site to Validate GPS Receiver Navigation. Paper AAS-95-369 presented at the AAS/AIAA Astrodynamics Specialist Conference. Halifax, Nova Scotia, Canada. [10, 11]
- Hujasak, R. S., and G. C. Gilbreath. 1998. Sequential Orbit Determination for GPS-35 and GPS-36 Using SLR Data from NRL at SOR. SPIE Proceedings, 3380 (32). [10]
- Hujasak, R. S., and J. F. Dicello. 1999. Sequential Orbit Determination for the Catastrophic Decay of Object 13390. Paper presented at the Space Control Workshop. MIT/LL, MA. [10]
- Jablonski, [Boelitz] Carole A. 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. [8, 11]
- . 1992. Confidence in Atmospheric Density Forecasts. Paper AAS-92-181 presented at the AAS/AIAA Spaceflight Mechanics Meeting. Colorado Springs, CO. [8]
- 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. [8, 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. [8, B]
- Jane's Space Directory*. 1997. Alexandria, VA: Jane's Information Group. [6, 11]
- Junkins, John L. 1975. Optimal Estimation Theory. NSWC/DL Technical Note. TN-K-45/74. Dahlgren, VA: US Naval Weapons Laboratory. [10, C]
- . 1977. *An Introduction to Optimal Estimation of Dynamical Systems*. Sijhoff-Noordhoff. [10]
- Junkins, John L., Maruthi R. Akella, and Kyle T. Alfriend. 1996. Non-Gaussian Error Propagation in Orbital Mechanics. *Journal of the Astronautical Sciences*. [9, 10]
- Jursa, Adolph S., ed. 1985. *Handbook of Geophysics and the Space Environment*. USAF Document ADA 167000. U.S. Air Force Geophysics Laboratory. [8, 9]
- Kalman, Rudolf E. 1960. A New Approach to Linear Filtering and Prediction Problems. *ASME Transactions - D, Journal of Basic Engineering*. 82D(1): 35-45. [10]
- Kane, Thomas R., Peter Likens, and David Levinson. 1983. *Spacecraft Dynamics*. New York: McGraw Hill. [9]

- Kaplan, G. H. (ed.). 1981. The IAU Resolutions on Astronomical Constants, Time Scales, and the Fundamental Reference Frame. USNO Circular No. 163, U.S. Naval Observatory, Washington DC. [3]
- Kaplan, G. H. 2005. The IAU Resolutions on Astronomical Reference Systems, Time Scales, and Earth Rotation Models, Explanation and Implementation. US Naval Observatory Circular No. 179. <http://aa.usno.navy.mil/kaplan/Circular.html> [3]
- Kaplan, G. H., et al. 1989. Mean and Apparent Place Computations in the New IAU System. III. Apparent, Topocentric, and Astrometric Places of Planets and Stars. *Astronomical Journal*. 97(4): 1197-1214. [4]
- Kaplan, Marshall H. 1976. *Modern Spacecraft Dynamics and Control*. New York: John Wiley & Sons. [6, 7, 8, 9]
- Kaula, William M. 1959. Statistical and Harmonic Analysis of Gravity. *Journal of Geophysical Research*. 64(12): 2401-2421. [8, 10]
- . 1966. *Theory of Satellite Geodesy*. Waltham MA: Blaisdell Publishing Co. [8, 9]
- . 1969. Tidal Friction and Longitude-Dependent Amplitude and Phase Angle. *Astronomical Journal*. 74(9): 1108-1114. [8, 9]
- Kaya, Denise, and Dan Snow. 1999. Element Set Prediction Accuracy Assessment. Paper AAS 99-425 presented at the AAS/AIAA Astrodynamics Specialist Conference. August 16-19, Girdwood, AK. [11]
- Kelso, T.S. 2004. "Frequently Asked Questions: Two-Line Element Set Format." (See <http://CelesTrak.com/columns/v04n03/>) [9]
- . 2007. Personal Discussions. [6]
- Kelso, T.S., and S. Alfano. 2005. Satellite Orbital Conjunction Reports Assessing Threatening Encounters in Space (SOCRATES). Paper AAS-05-124 presented at the AAS/AIAA Spaceflight Mechanics Conference, January 23-27, 2005. Copper Mountain, CO. [11]
- Kennedy, J. 1993. Debris gets in Your Space. *Guardian, The Magazine of Air Force Space Command*. 1(1): 4-5. [11]
- Khutorovsky, Z. N. 1996. Low-perigee Satellite Catalog Maintenance: Issues of Methodology. Paper presented at U.S.-Russian Space Surveillance Workshop. Poland. [10]
- . 1998. Monitoring of GEO satellites in Russian Space Surveillance Center. Third US/Russian Space Surveillance Workshop. Washington, D.C. [9, 10]
- King-Hele, D. 1964. *Theory of Satellite Orbits in an Atmosphere*. London: Butterworths. [3]
- . 1987. *Satellite Orbits in an Atmosphere*. London: Blackie. [9, 10]
- 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. [9]
- Klinkrad, Heiner. 2009. Monitoring Space – Efforts Made by European Countries. ESA/ESOC, Robert-Bosch-Strasse 5, D-64293 Darmstadt, Germany. Internet: <http://www.fas.org/spp/military/program/track/klinkrad.pdf>. [4]
- Knapp, Barry. 2005. Personal communication.
- Knocke, P. C., J. C. Ries and B. D. Tapley. 1988. Earth Radiation Pressure Effects on Satellites. Proceedings of the AIAA/AAS Astrodynamics Conference. Washington DC. 577-586. [8]

- Knowles, Stephen H. 1995. A Comparison of Geocentric Propagators for Operational Use. Paper AAS-95-429 presented at the AAS/AIAA Astrodynamics Specialist Conference. Halifax, Nova Scotia, Canada. [8, 9]
- Koestler, Arthur. 1959. *The Sleepwalkers*. New York: MacMillan Co. [2]
- Koskela, P. E. 1967. Astrodynamical Analysis for the Advanced Orbit/Ephemeris Subsystem (AOES). Aeronutronic Publication No. U-4180, PHILCO-Ford Corporation. [10]
- Kozai, Yoshihide. 1959. The Motion of a Close Earth Satellite. *Astronomical Journal*. 64(1274): 367-377. [9]
- . 1962. Second-Order Solution of Artificial Satellite Theory Without Drag. *Astronomical Journal*. Vol. 67: 446. [9]
- Kreyzig, Erwin. 1983. *Advanced Engineering Mathematics*. 5th ed. New York: John Wiley Publishing. [3, 8, 9]
- Krogh, F. T. 1974. Changing stepsize in the integration of differential equations using modified divided differences. Proceedings of the Conference on the Numerical Solution of Ordinary Differential Equations. vol 362, Lecture Notes in Mathematics. New York: Springer-Verlag. [7]
- Kwok, Johnny. 1987. The Artificial Satellite Analysis Program (ASAP) Version 2.0. JPL EM 312/87-153. Jet Propulsion Laboratory, Pasadena, CA. [9]
- Lagrange, J. L. 1873. *Collected Works*. Vol. 6. Paris: Gauthier-Villars. [9]
- Lambeck, Kurt. 1988. *Geophysical Geodesy*. Oxford: Clarendon Press. [3, 8, 9]
- 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. [7]
- Lane, M. H. 1965. "The Development of an Artificial Satellite Theory Using Power-Law Atmospheric Density Representation." AIAA Paper 65-35. [9]
- Lane, M. H., and K. H. Cranford. 1969. "An Improved Analytical Drag Theory for the Artificial Satellite Problem." AIAA Paper No. 69-925. [9]
- Langley, Richard B. 1999. Dilution Precision. *GPS World*. May 1999: 52-59. [11]
- Larson, Wiley J., and J. R. Wertz. 1999. *Space Mission Analysis and Design*. 3rd ed. Norwell, MA and Torrance, CA: Kluwer Academic Publications and Microcosm, Inc. [6, 11]
- Lawden, D. F. 1952. Orbital Transfer via Tangential Ellipses. *British Interplanetary Society Journal*. 11(6): 278-289. [6]
- 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. [11]
- Lee, Deok-Jin, and Kyle T. Alfriend. 2000. Effect of Atmospheric Density Uncertainty on Collision Probability. Paper AAS-00-181 presented at the AAS/AIAA Spaceflight Mechanics Conference. Clearwater, FL. [10]
- Levit, Creon, and William Marshall. 2011. Improved orbit predictions using two-line elements *Advances in Space Research Journal*. Vol 47: 1107. 1 Apr. [10]
- Lieske, J. H., et al. 1977. Expressions for the Precession Quantities Based upon the IAU (1976) System of Astronomical Constants. *Astronomy and Astrophysics*. Vol. 58: 1-16. [3]
- Lisowski, Ronald. 1995. Personal discussions. U.S. Air Force Academy. [6]

- 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. [9]
- . 1974. Satellite Motion about an Oblate Earth. *AIAA Journal*. Vol. 12: 1511-1516. [9]
- Logsdon, Tom. 1992. *The Navstar Global Positioning System*. New York, NY: Van Nostrand-Rinehold. [11]
- 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. [9, 10]
- 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. [7, 8, 9, 10, 11, B]
- Lovell, Alan. 2006. Personal Discussions. [6]
- 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 Specialist Conference. Halifax, Nova Scotia, Canada. [8]
- Lundberg, J. et al. 1985. Surfaces of Zero Velocity in the Restricted Problem of Three Bodies. *Celestial Mechanics*. 36(2): 191-205. [1]
- Luthcke, S. B., et al. 1997. Enhanced Radiative Force Modeling of the Tracking and Data Relay Satellites. *American Astronomical Society Journal*. Vol 45:349-370. [8]
- Lyddane, Robert H. 1963. Small Eccentricities or Inclinations in the Brouwer Theory of the Artificial Satellite. *Astronomical Journal*. 68 (8): 555-558. [9]
- 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. [8]
- Maia. April 2011. <ftp://tai.bipm.org/iers/conv2010/chapter5/tabc5.2a.txt, 2b.txt, 2d.txt, 2e.txt, 3a.txt, 3b.txt>. [3]
- 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. [9]
- Marcos, Frank A., C. R. Baker, J. N. Bass, T. L. Killeen, and R. G. Roble. 1993. Satellite Drag Models: Current Status and Prospects. Paper AAS-93-621 presented at the AAS/AIAA Astrodynamics Specialist Conference. Victoria BC, Canada. [8, 9]
- Marcos, Frank A. et al. 2006. Accuracy of the Earth's Thermospheric Neutral Density Models. Paper AIAA-2006-6167 presented at the AIAA/AAS Astrodynamics Specialist Conference. Keystone, CO. [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 Specialist Conference. Halifax, Nova Scotia, Canada. [7]
- Marshall, P. M. 1999. Least Squares Solutions in Statistical Orbit Determination Using Singular Value Decomposition. MS Thesis. Naval Postgraduate School. Monterey, CA. [10]
- Matthews, John H. 1987. *Numerical Methods*. New Jersey: Prentice Hall Inc. [8]
- 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. [8]
- Maybeck, Peter S. 1979. *Stochastic Models, Estimation, and Control*. New York: Academic Press. [10]

- McCarthy, Dennis. 1992. *IERS Technical Note #13*. U.S. Naval Observatory. [3]
- . 1996. *IERS Technical Note #21*. U.S. Naval Observatory. [3]
- McCarthy, Dennis D., and Gerard Petit. 2003. *IERS Technical Note #32—IERS Conventions (2003)*. U.S. Naval Observatory. [3]
- 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. [9]
- . 1987. Eccentricity Control and the Frozen Orbit Concept. AAS/AIAA Paper No. AAS 87-516 presented at Kalispell, Montana, Aug. 10-13. [11]
- . 1990. *Single Station Orbit Determination for Landsat 6*. Paper AIAA-90-2923 presented at the AIAA/AAS Astrodynamics Specialist Conference. Portland OR. [8, 11]
- . 1990. The Nodal Period. Unpublished interoffice memo. [11]
- . 1992. *Semianalytic Artificial Satellite Theory*. Vol. 1. Charles Stark Draper Laboratory. [9]
- McReynolds, Steve. 1995. Filter Smoother Consistency Test. Interoffice Memo. Martin Marietta. [10]
- Meeus, Jean. 1991. *Astronomical Algorithms*. Richmond, VA: Willmann Bell Inc. [2, 3, C, D]
- Melton, Robert G. 2000. Time-Explicit Representation of Relative Motion Between Elliptical Orbits. *Journal of Guidance, Control, and Dynamics*. 23(4): 604-610. [6]
- 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. [9]
- . 1975. Numerical Integration of the Differential Equations of Celestial Mechanics. TR 74184. Royal Aircraft Establishment. Farnborough, Hants, UK. also DTIC AD B004645. [8]
- Metzinger Richard. 1993. *Validation of the Workstation Version of R&D GTDS*. Cambridge, MA: Charles Stark Draper Laboratory. [8]
- Milani, Andrea, Anna Maria Nobili and Paolo Farinella. 1987. *Non-Gravitational Perturbations and Satellite Geodesy*. IOP Publishing Ltd, Adam Hilger. Bristol, England. [9]
- Moe, M. M. 1960. Solar-Lunar Perturbations of the Orbit of an Earth Satellite. *ARS Journal*. Vol. 30: 485. [9]
- Molotov I., Agapov V. et al. 2008. International Scientific Optical Network for space debris research, Advances in Space Research, Volume 41, Issue 7, 1022-1028. [4]
- Montenbruck, Oliver, and Thomas Pfleger. 1994. *Astronomy on the Personal Computer*. New York: Springer Verlag. [7, 8]
- Montenbruck, Oliver, Tom van Helleputte, Remco Kroes, and Eberhard Gill. 2005. Reduced dynamic orbit determination using GPS code and carrier measurements. *Journal of Aerospace Science and Technology*. 9:261–271, Elsevier. [10]
- Montenbruck, Oliver, and Eberhard Gil. 2005. *Satellite Orbits: Models, Methods, and Applications*. New York: Springer-Verlag.
- Morgen, Walter L., and G. D. Gordon. 1989. *Communications Satellite Handbook*. New York: John Wiley & Sons. [7]
- Moritz, Helmut, and I. Mueller. 1987. *Earth Rotation—Theory and Observation*. New York: Ungar Publishing Company. [3, 8]

- Morring, Frank Jr. 2012. Sat Operators Team for Situational Awareness. *Aviation Week and Space Technology*. March 27, 2012. [11]
- Moulton, Forest R. 1914. *An Introduction to Celestial Mechanics*. New York: Dover Publications. [1, 7]
- Muldoon, Alana R., et al. 2009. Improved orbital debris trajectory estimation based on sequential TLE Processing. Paper IAC-09.A6.2.9 presented at the 60th International Astronautical Congress. Daejeon, South Korea. [10]
- Muolo, Michael J. 1993. *Space Handbook, A Warfighter's Guide to Space*. Vol. I. AU-18. Maxwell AFB, AL: Air University Press. [10]
- Musen, P. 1960. On the Motion of a Satellite in an Asymmetrical Gravitational Field. *Journal of Geophysical Research*. 65(9): 2783-2792. [9]
- National Geophysical Data Center. *Geomagnetic Indices Bulletin*. Boulder, CO: National Geophysical Data Center. [8]
- National Geophysical Data Center. *Solar Indices Bulletin*. Boulder, CO: National Geophysical Data Center. [8]
- Nayfeh, A. 1973. *Perturbation Methods*. New York: Wiley-Interservice Publications. [9]
- Nazarenko, Andrey. 1999. Atmospheric Density Tracking Studies. Report prepared by the Scientific Industrial Firm NUCOL for the Charles Stark Draper Laboratory, CSDL-C-6505. [8]
- . 1999. The Description of a Theoretical Fundamentals of Universal Semianalytical Method. Development Formulas for Perturbation Account. Report prepared by the Scientific Industrial Firm NUCOL for the Charles Stark Draper Laboratory, CSDL-C-6502. [9]
- 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. [3, 8, 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 Specialist Conference. Halifax, Nova Scotia, Canada. [3, 9]
- Neta, Beny. 1996. Personal discussions. [8, 9, 10]
- Neta, Beny, and David A. Vallado. 1998. On Satellite Umbra/Penumbra Entry and Exit Positions. *Journal of the Astronautical Sciences*. 46(1): 91-103. [5, 8]
- Neter, John, William Wasserman, and G. A. Whitmore. 1982. *Applied Statistics*. Boston, MA: Allyn and Bacon Inc. [10]
- Newcomb, Simon. [1906] 1960. *A Compendium of Spherical Astronomy*. Reprint edition. New York: Dover Publications. [3]
- Newton, Sir Isaac. [1687] 1962. *Mathematical Principles of Natural Philosophy*. Reprint edition. Berkeley, CA: University of California Press. [1, 2, 6]
- NGDC. January 2013.
ftp://ftp.ngdc.noaa.gov/STP/SOLAR_DATA/SOLAR_RADIO/FLUX/Penticton_Adjusted/daily/DAILYPLT.ADJ
ftp://ftp.ngdc.noaa.gov/STP/SOLAR_DATA/SOLAR_RADIO/FLUX/Penticton_Observed/daily/DAILYPLT.OBS[8]
- NIMA. 2000. *Department of Defense World Geodetic System 1984*. NIMA-TR 8350.2, 3rd ed, Amendment 1. Washington, DC: Headquarters, National Imagery and Mapping Agency. [3]

- 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. [8]
- Nguyen, Bao U. and Murray E. Dixson. 2003. Computation of Effective Ground Range Using an Oblate Earth Model. *Journal of the Astronautical Sciences*. 51(3): 291-305. [11]
- Ochoa, Sergio I., and John E. Prussing. 1992. Multiple Revolution Solutions to Lambert's Problem. Paper AAS-92-194 presented at the AAS/AIAA Spaceflight Mechanics Conference. Colorado Springs, CO. [7]
- O'Conner, John J. 1983. Methods of Trajectory Mechanics (Second Edition). RCA International Services Corporation. Report # ESMC-TRR-84-01. [8]
- Oltrogge, Daniel L. 2008. AstroHD: Astrodynamics Modeling with a Distinctly Digital Flavor. Paper AIAA 2008-7065 presented at the AAS/AIAA Astrodynamics Specialist Conference, August 18-21. Honolulu, HI. [9]
- Oltrogge, Daniel L. and Chia-Chun Chao. 2007. Standardized Approaches for Estimating Orbit Lifetime after End-of-Life. Paper AAS 07-261 presented at the AAS/AIAA Astrodynamics Specialist Conference. Mackinac Island, MI. [10]
- 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. [8]
- Pannekoek, Anton. 1989. *A History of Astronomy*. New York: Dover Publications. [1, 2, 3, 4, 9]
- Parker, Jeffrey S. 2007. Low Energy Ballistic Lunar Transfers. Ph.D. Thesis. Boulder, CO: University of Colorado. [12]
- Patera, Russell A. 2001. General Method for Calculating Satellite Collision Probability. *Journal of Guidance, Control, and Dynamics*. 24(4): 716-722. [11]
- . 2005. Calculating Collision Probability for Arbitrary Space-Vehicle Shapes via Numerical Quadrature. *Journal of Guidance, Control, and Dynamics*. 28(6): 1326-1328. [11]
- Patt, Frederick S., Hoisington, Charles M., Gregg, Watson W., and Coronado, Patrick L. 1993. NASA Technical Memorandum 104566, Vol. 11 "Volume 11, Analysis of Selected Orbit Propagation Models for the SeaWiFS Mission" available at <http://library.gsfc.nasa.gov/Data-bases/Gtrs/Data/TM-1993-104566v11.pdf>, Sep 06.
- Peterson, Glenn E. 2005. Determining Object Sizes from Radar Cross Section for Collision Avoidance. Paper AAS 05-307 presented at the AAS/AIAA Astrodynamics Specialist Conference. Lake Tahoe, NV. [11]
- Petit, Gerard, and Brian Luzum. 2010. *IERS Technical Note #36—IERS Conventions (2010)*. Verlag des Bundesamts fur Kartographie und Geodasie. Frankfurt am Main 2010. [3]
- Phillips, Charles L., and H. T. Nagle Jr. 1984. *Digital Control System Analysis and Design*. Englewood Cliffs NJ: Prentice Hall Inc. [8]
- Phillips Laboratory. 1995. 1995 Success Stories. Kirtland Air Force Base, NM: Phillips Laboratory History Office. [10]
- Plummer, H. C. [1918] 1960. *An Introductory Treatise on Dynamical Astronomy*. Reprint edition. New York: Dover Publications. [2, 8]

- 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. [8]
- Pon, Wing Y. 1973. *Theory for the Advanced Orbit Determination Ephemeris Generation System*. Air Force Contract F-04701-73-C-0029. Data Dynamics Inc, CA. [2, 3]
- Press, William H. et al. 1992. *Numerical Recipes in Fortran*. Cambridge, England: Cambridge University Press. [6, 8]
- Proulx, Ron J. et al. 1981. A Theory for the Short-Periodic Motion Due to the Tesserel Harmonic Gravity Field. Paper AAS-81-180 presented at the AAS/AIAA Astrodynamics Specialist Conference. Lake Tahoe, NV. [9]
- Prussing, John E. and B. A. Conway. 1993. *Orbital Mechanics*. New York: Oxford Press. [3, 7]
- Register, Justin E. 2003. Contribution of Individual Forces to Orbit Determination Accuracy. Paper AAS-03-227 presented at the AAS/AIAA Spaceflight Mechanics Conference. Ponce, Puerto Rico. [9]
- Reid, Gary J. 1983. *Linear System Fundamentals*. New York: McGraw-Hill, Inc. [C]
- Ries, John C. and Byron D. Tapley. 1999. Centimeter Level Orbit Determination for the Topex/Poseidon Altimeter Satellite. Paper AAS-99-142 presented at the AAS/AIAA Spaceflight Mechanics Conference. Breckenridge, Colorado. [10]
- Roberts, Charles E., Jr. 1971. An Analytic Model for Upper Atmosphere Densities Based upon Jacchia's 1970 Models. *Celestial Mechanics*. 4(314): 368-377. [8, B]
- Rogers, D. F., and Adams, J. A., *Mathematical Elements for Computer Graphics*, 2nd ed., McGraw-Hill, New York, 1990. [11]
- Roy, Archie E. 1988. *Orbital Motion*. New York: John Wiley & Sons. [2, 8, 9]
- Ross, I. Michael. 1999. Private communication. [5]
- Ross, S. D. 2004. Cylindrical Manifolds and Tube Dynamics in the Restricted Three-Body Problem. Ph.D. Thesis. Pasadena, CA: California Institute of Technology. [12]
- Sabol, Chris, and Scott Carter and David Vallado. 1999. A Fresh Look at Angles-only Orbit Determination. Paper AAS-99-363 presented at the AAS/AIAA Astrodynamics Specialist Conference. Girdwood, AK. [10]
- 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. [9, 11]
- Sadler, D. H. 1974. *Interpolation and Allied Tables*. London: Her Majesty's Nautical Almanac Office. [8, 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. [8]
- Schatten, Kenneth H., and S. Sofia. 1987. Forecast of an Exponentially Large Even-numbered Solar Cycle. *Geophysics Research Letters*. Vol. 14: 632-635. [8]
- _____. 1988. A Model for Solar Constant Secular Changes. *Geophysics Research Letters*. Vol. 15: 121-124. [8]
- 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 Astrodynamics Specialist Conference. Halifax, Nova Scotia, Canada. [9, 10]

- Schumacher, Paul W., Jr., et al. 1998. Design for Operational Calibration of the Naval Space Surveillance System. Paper AAS-98-116 presented at the AAS/AIAA Spaceflight Mechanics Conference. Monterey, CA. [4]
- Schweighart, Samuel A., and Raymond J. Sedwick. 2002. High Fidelity Linearized J2 Model for Satellite Formation Flight. *Journal of Guidance, Control, and Dynamics*. 25(6): 1073-1080. [6]
- Seago, John et al. 2011. Sequential Orbit-Estimation with Sparse Tracking. Paper AAS 11-123 presented at the AAS/AIAA Spaceflight Mechanics Conference. New Orleans, LA. [10]
- Seago, John, and Mark Davis and Anne Clement. 2000. Precision of a Multi-Satellite Trajectory Database Estimated from Satellite Laser Ranging. Paper AAS-00-180 presented at the AAS/AIAA Spaceflight Mechanics Conference. Clearwater, FL. [10]
- Seago, John and David Vallado. 2000. Coordinate Frames of the U.S. Space Object Catalogs. Paper AIAA 2000-4025 presented at the AIAA/AAS Astrodynamics Specialist Conference. Denver CO. [3, 9]
- Seidelmann, Kenneth. 1982. 1980 IAU Theory of Nutation: The Final Report of the IAU Working Group on Nutation. *Celestial Mechanics*. 27: 79-106. [3]
- . 1992. *Explanatory Supplement to the Astronomical Almanac*. California: University Science Books. [2, 3, D]
- Seidelmann, P. K., and J. Kovalevsky. 2002. Application of the new concepts and definitions (ICRS, CIP and CEO) in fundamental astronomy. *Astronomy and Astrophysics*. Vol. 392: 341-351. [3]
- Selby, Samuel M., ed. 1975. *Standard Mathematical Tables*. 23rd ed. Cleveland, OH: CRC Press, Inc. [2, 7, 8, 10, C]
- Sellers, Jerry J. 1994. *Understanding Space: An Introduction to Astronautics*. New York: McGraw-Hill, Inc. [1]
- Shah, Naresh. 1997. Automated Station-Keeping for Satellite Constellations. M.S. Thesis, CSDL-T-1288. Department of Aeronautics and Astronautics, Massachusetts Institute of Technology. [11]
- Shampine, L. F. and M. K. Gordon. 1975. *Computer Solution of Ordinary Differential Equations*. San Francisco, CA: W. H. Freeman and Company. [8]
- Sharma, Jayant, et al. 2002. Toward Operational Space-Based Space Surveillance. *Lincoln Laboratory Journal*. 13(2): 309-334. [4, 10]
- Shaver, Jeffrey S. 1980. Formulation and Evaluation of Parallel Algorithms for the Orbit Determination Problem. MIT P.h.D. Dissertation. #CSDL-T-709. Cambridge, MA: Charles Stark Draper Laboratory, Inc. [10]
- Shen, Haijun, and Panagiotis Tsotras. 2003. Using Battin's Method to Obtain Multiple Revolution Lambert's Solutions. Paper AAS-03-568 presented at the AAS/AIAA Astrodynamics Specialist Conference. Big Sky, MT. [7]
- Sherman, S. 1958. *Non-Mean-Square Error Criteria*, IRE Transactions on Information Theory, Vol. IT-4. [10]
- Shuster, Malcolm. 2008. The Nature of the Quaternion. *Journal of the Astronautical Sciences*. 56(3): 359-373. [3]
- Siewert, C. E. and E. E. Burniston. 1972. An Exact Analytical Solution of Kepler's Equation. *Celestial Mechanics*. 6(3): 294-304. [2]

- Simon, Dan. 2006. *Optimal State Simulation*. New York: John Wiley & Sons. [10]
- 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. [9]
- Small, Robert. [1804] 1963. *An Account of the Astronomical Discoveries of Kepler*. Reprint edition. Madison, WI: University of Wisconsin Press. [1, 5]
- Smith, D. E. 1962. The Perturbation of Satellite Orbits by Extra-Terrestrial Gravitation. *Planetary and Space Science*. Vol. 9: 659-674. [9]
- Smith, James E. et al. 1999. An Operational Approach for Generating Near-Optimal Station Keeping Strategies Via Parallel Genetic Algorithms. Paper AAS-99-387 presented at the AAS/AIAA Astrodynamics Specialist Conference. Girdwood, AK. [11]
- Smith, James E. 1999. Application of Optimization Techniques to the Design and Maintenance of Satellite Constellations. M.S. Thesis, CSDL-T-1336. Department of Aeronautics and Astronautics, Massachusetts Institute of Technology. [7, 11]
- Smith, John C. Jr., and Julia L. Bell. 2005. 2001 Mars Odyssey Aerobraking. *Journal of Spacecraft and Rockets*. 42(3): May-June. [12]
- 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. [7]
- . 1986. *Study of a Homotopy Continuation Method for Early Orbit Determination with the Tracking and Data Relay Satellite System (TDRSS)*. NASA Technical Memorandum 86230. [7]
- Sobel, Dava, and William J. H. Andrews. 1998. *Longitude*. NY: Walker & Co. [11]
- Sorenson, Harold W. 1970. Least-squares Estimation: from Gauss to Kalman. *IEEE Spectrum*. 7(7): 63-68. [10]
- Space-travel. December 2010.
[http://www.space-travel.com/reports/
Chaotic space traffic needs rules less secrecy US general 999.html](http://www.space-travel.com/reports/Chaotic_space_traffic_needs_rules_less_secrecy_US_general_999.html) [11]
- Spence, L. B. 2000. Proceedings of the 2000 Space Control Workshop. Report STK-253. Mass Institute of Technology, Lincoln Laboratory. [4]
- Standish, Myles. 1990. The Observational Basis for JPL's DELOD, the Planetary Ephemerides of the Astronomical Almanac. *Astronomy and Astrophysics*. Vol. 233: 252-271. [8]
- Stephenson, Bruce. 1987. *Kepler's Physical Astronomy*. New York: Springer-Verlag. [1, 2, 7]
- Sterne, 1960. *An Introduction to Celestial Mechanics*. New York: Interscience Publications Inc. [9]
- Stiefel, Eduard L., and G. Scheifele. 1971. *Linear and Regular Celestial Mechanics*. New York: Springer Verlag. [8]
- Storz, Mark. 1999. Satellite Drag Accuracy Improvements Estimated from Orbital Energy Dissipation Rates. Paper AAS 99-385 presented at the AAS/AIAA Astrodynamics Specialist Conference. Girdwood, AK. [8]
- Sundman, Karl F. 1913. Mémoire sur le problème des trois corps. *Acta Mathematica*. Vol. 36: 105-179. [2]

- Sutton, Eric K., Jeffrey M. Forbes and Steven Nerem. 2006. Atmospheric Wind Measurements Deduced from Accelerometer Data. Paper AIAA 2006-6170 presented at the AAS/AIAA Astrodynamics Specialist Conference. Keystone, CO. [8]
- Svalgaard, Leif. 2009 (Jan 2013). The Radio Flux Prediction Graph. http://www.leif.org/research/The_SWPC_Solar_Flux.pdf. [8]
- Swerling, Peter. 1959. First Order Propagation in a Stage-wise Smoothing Procedure for Satellite Observations. *Journal of the Astronautical Sciences*. 6(3): 46-62. [10]
- Szebehely, Victor. 1967. *Theory of Orbits*. New York: Academic Press. [1, 8]
- . 1989. *Adventures in Celestial Mechanics*. Austin, TX: University of Texas Press. [8, 10]
- Taff, Laurence G. 1985. *Celestial Mechanics, A Computational Guide for the Practitioner*. NY: John Wiley & Sons. [1, 2, 3, 8, 9]
- Tapley, Byron D., and Bob E. Schutz and George H. Born. 2004. *Statistical Orbit Determination*. Burlington MA: Elsevier. [10]
- Thompson, William T. 1986. *Introduction to Spacecraft Dynamics*. New York: Dover Publications. [6]
- 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. [7]
- . 2004. Lambert's Theorem - A Complete Series Solution. *Journal of the Astronautical Sciences*. 52(4): 441-454. [7]
- . 2007. Personal discussions. [7]
- 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. [2, 7]
- 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. [6]
- 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. [8]
- Tisserand, F. 1889. *Traité de méchanique céleste*, 4 vols. Paris: Gauthier-Villars. [9]
- Ullman, Richard E. 1992. SOLVE Program. NASA Contract NAS5-29393, Task 503, HSTX/G&G-9201. Hughes STX Corporation Lanham, MD. [10]
- Uphoff, C. 1985. Software for Conversion of Osculating to Mean Elements. JPL IOM 312/85.2-927. Jet Propulsion Laboratory, Pasadena, CA. [9]
- U.S. Standard Atmosphere*. 1976. Washington, DC: U.S. Government Printing Office. [8]
- Vallado, David A. 1991. Orbital Phasing. USAFA Astro-321 course reading. Colorado Springs, CO: U.S. Air Force Academy. [5]
- . 1991a. Reconnaissance Mission Planning. Astro 422 course reading. Colorado Springs, CO: U.S. Air Force Academy. [11]
- . 1992a. Statistical Orbit Determination. Astro 422 Course Reading. Colorado Springs, CO: U.S. Air Force Academy. [4, 8, 9, 10]
- . 1992b. Methods of Astrodynamics—A Computer Approach Ver 4.0. Colorado Springs, CO: U.S. Air Force Academy. [all]

- _____. 1992c. Software Standard, *PASCAL and FORTRAN*, Colorado Springs, CO: U.S. Air Force Academy. [D]
- _____. 1993a. A Study in Relative Motion. Phillips Laboratory, Kirtland AFB, NM. [6]
- _____. 1993b. A Study in Intercept and Rendezvous Techniques. Phillips Laboratory, Kirtland AFB, NM. [6, 7]
- _____. 1998. An Analysis of the Behavior of the J2000 Reduction Matrices. Paper AAS-98-155 presented at the AAS/AIAA Spaceflight Mechanics Conference, February 9-11. Monterey, CA. [10]
- _____. 2003. Covariance Transformations for Satellite Flight Dynamics Operations. Paper AAS-03-526 presented at the AAS/AIAA Astrodynamics Specialist Conference, August 4-7. Big Sky, MT. [10]
- _____. 2005. An Analysis of State Vector Propagation for Differing Flight Dynamics Programs. Paper AAS 05-199 presented at the AAS/AIAA Spaceflight Mechanics Conference, January 23-27. Copper Mountain, CO. [8, 9]
- Vallado, David A. and Salvatore Alfano. 2011. Curvilinear Coordinates for Covariance and Relative Motion Operations. Paper AAS 11-464 presented at the AAS/AIAA Astrodynamics Specialist Conference. July 31-August 4, Girdwood, AK. [6]
- Vallado, David A., and Salvatore Alfano. 1999. A Future Look at Space Surveillance and Operations. Paper AAS 99-113 presented at the AAS/AIAA Space Flight Mechanics Conference, February 7-10. Breckenridge, CO. [4]
- Vallado, David A. and Jacob D. Griesbach. 2011. Simulating Space Surveillance Networks. Paper AAS 11-580 presented at the AAS/AIAA Astrodynamics Specialist Conference. July 31-August 4, Girdwood, AK. [4]
- Vallado, David A. and Vladimir Agapov. 2010. Orbit Determination Results from Optical Measurements. Paper AIAA 2010-7525 presented at the AAS/AIAA Astrodynamics Specialist Conference, August 2-5. Toronto, Canada. [10]
- Vallado, David A., and David Finkleman. 2008. A Critical Assessment of Satellite Drag and Atmospheric Density Modeling. Paper AGI-UTH-Vallado presented at the AGI Users Conference, October 7-9. Chicago, IL. (updated from AIAA 2008-6442) [8, 9]
- Vallado, David A., and Paul Crawford. 2008. SGP4 Orbit Determination. Paper AIAA 2008-6770 presented at the AIAA/AAS Astrodynamics Specialist Conference, August 18-21, 2008. Honolulu, HI. [10]
- Vallado, David A., et al. 2006. Revisiting Spacetrack Report #3. Paper AIAA 2006-6753 presented at the AIAA/AAS Astrodynamics Specialist Conference, August 21-24, 2006. Keystone, CO. [9]
- Vallado, David A., John H. Seago, and P. Kenneth Seidelmann. 2006. Implementation Issues Surrounding the New IAU Reference Systems for Astrodynamics. Paper AAS 06-134 presented at the AAS/AIAA Spaceflight Mechanics Conference, January 22-26. Tampa, FL. [3]
- Vallado, David A., and T. S. Kelso. 2005. Using EOP and Solar Weather Data for Real-time Operations. Paper USR 05-S7.3 presented at the US/Russian Space Surveillance Workshop, August 22-26. St Petersburg, Russia. (Also AAS 05-406) [3, 8, D]
- Vallado, David A. and Scott S. Carter. 1997. Accurate Orbit Determination from Short-arc Dense Observational Data. Paper AAS 97-704 presented at the AAS/AIAA Astrodynamics Specialist Conference. Sun Valley, ID. [10]

- 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. [3, 10]
- Van Allen, J. A., G. H. Ludwig, E. C. Ray, and C. E. McIlwain. 1958. Observations of high intensity radiation by satellites 1958 Alpha and Gamma, *Jet Propulsion*. Vol 28:588-592.
- Vetter, Jerome R. 1994. The Evolution of Earth Gravity Models used in Astrodynamics. *APL Technical Digest*, John Hopkins. 15(4): 319-335. [3, 8]
- _____. 2001. Private communication. [8]
- Vincenty, T. 1966. Transformation of Geodetic Data between Reference Ellipsoids. *Journal of Geophysical Research*. 71:2619. [11]
- _____. 1975. Direct and Inverse Solutions of Geodesics on the Ellipsoid with Application of Nested Equations. *Survey Review*. XXII(176): 88-93. [11]
- Voiskovskii, M. I., et al. 1973. An Aspherical Model for the Upper-Atmosphere Density. *Cosmic Research*, 11(1): 70-79. [8, B]
- Volkov, I. I., Ye. I. Knyazeva, and B. V. Kugayenko. 1982. Refining a Model of Atmospheric Density for Ballistic Calculations. Nablyudenija Iskusstvennykh Nebesnykh Tel (Moskva), No. 80, Soviet Union. pp. 126-135. [8, B]
- _____. 1978. *Model of the upper atmosphere for ballistic accounts*. GOST 22721-77. Moscow, Publishing House of the Standards. [B]
- Volkov, I. I. 1984. The Upper Atmosphere of the Earth. Model of Density for Ballistic Maintenance of the Earth Artificial Satellite Flights. GOST 25645-115-84. Publishing House of the Standards, Moscow. [8]
- Von Zeipel, H. 1916. *Recherches sur le mouvement des petites planètes*. Arkiv. Mat. Astron. Fys Col 11(1), 11(7), 12(9). [9]
- Wallace, P. T., and N. Capitaine. 2006. Precession-nutation Procedures Consistent with IAU 2006 Resolutions. *Astronomy and Astrophysics*. Vol. 459: 981-985. [3]
- 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. [9, 11]
- Walpole, Ronald E. 1974. *Introduction to Statistics*. New York: Macmillan Publishing. [10]
- Weeden, Brian C., and Paul J. Cefola. 2010. Computer Systems and Algorithms for Space Situational Awareness: History and Future Development. Paper ISCOPS 10-420 presented at the 12th International Space Conference of Pacific Basin Societies, July 27-30, 2010. Montreal, Canada. [11]
- Weisstein, Eric W. June 2011. "Cylindrical Equidistant Projection." From MathWorld--A Wolfram Web Resource. <http://mathworld.wolfram.com/CylindricalEquidistantProjection.html>
- Wertz, James R., et al. 2011. *Space Mission Engineering: The New SMAD*. Hawthorne, CA: Microcosm, Inc. [11]
- Wertz, James R. 1978. *Spacecraft Attitude Determination and Control*. Dordrecht, Holland: D. Reidel Publishing Company. [3, 8, 9]
- Wheeler, J. A. 1990. *A Journey into Gravity and Spacetime*. San Francisco, CA: W. H. Freeman & Company. [11]
- Wiesel, William E. 1983. Modern Methods of Orbit Determination. MC 7.31 Course Notes. AFIT Press. [10]

- Wiesel, William E., and Salvatore Alfano. 1985. Optimal Many-revolution Orbit Transfer. *Journal of Guidance*. 8(1): 155-157. [6]
- Wilkins, Matthew P., et al. 2000. Transitioning from a General Perturbations to a Special Perturbations Space Object Catalog. Paper AIAA-2000-4238 presented at the AIAA/AAS Astrodynamics Specialist Conference. Denver, CO. [9]
- Wilkins, Matthew P. et al. 2006. Practical Challenges in Implementing Atmospheric Density Corrections to the NRL-MSISE-00 Model. Paper AAS-06-170 presented at the AAS/AIAA Spaceflight Mechanics Conference. Tampa, FL. [8]
- 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. [6]
- Wilson, Beth L. 2004. Space Surveillance Network Automated Tasker. Paper AAS-04-126 presented at the AAS/AIAA Spaceflight Mechanics Conference. Maui, HI. [10]
- Woodburn, James, Vincent Coppola and Frank Stoner. 2009. A Description of Filters For Minimizing The Time Required For Orbital Conjunction Computations. Paper AAS-09-372 presented at the AAS/AIAA Astrodynamics Specialist Conference, August 9-13. Pittsburg, PA. [11]
- Wright, James R. 1981. Sequential Orbit Determination with Auto-Correlated Gravity Modeling Errors. *AIAA Journal of Guidance and Control*. 4(2): 304. [10]
- _____. 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. [10]
- _____. 1994b. Analytical Expressions for Orbit Covariance due to Gravity Errors. Paper AAS-94-3722 presented at the AIAA/AAS Astrodynamics Specialist Conference. Scottsdale, AZ. [10]
- _____. 2002. Optimal Orbit Determination. Paper AAS-02-192 presented at the AAS/AIAA Spaceflight Mechanics Conference. San Antonio, TX. [10]
- Wright, James R. and Sergei Tanygin. 2004. Removal of Arbitrary Discontinuities in Atmospheric Density Modeling. Paper AAS-04-176 presented at the AAS/AIAA Spaceflight Mechanics Conference. Maui, HI. [8]
- Wright, James R. and James Woodburn. 2004. Simultaneous Real-time Estimation of Atmospheric Density and Ballistic Coefficient. Paper AAS-04-175 presented at the AAS/AIAA Spaceflight Mechanics Conference. Maui, HI. [8,10]
- Wright, James R., et al. 2008. Orbit Gravity Error Covariance. Paper AAS 08-157 presented at the AAS/AIAA Space Flight Mechanics Conference, January 28-31. Galveston, TX. [10]
- Yeomans, Donald K. 1991. *Comets*. New York: John Wiley & Sons. [2, 7]
- Yurasov, V. 1996. Universal Semianalytical Satellite Motion Propagation Method. Paper presented at U.S.-Russian Space Surveillance Workshop, Poland. [9, 10]
- _____. 1999. Personal correspondence. [8]
- Yurasov, V., and A. Moscovsky. 1996. Geostationary Orbit Determination and Prediction. Paper presented at U.S.-Russian Space Surveillance Workshop, Poland. [10]
- Yurasov, V., et al. 2006. Density Corrections for the NRLMSISE-00 Model. Paper AAS-05-168 presented at the AAS/AIAA Spaceflight Mechanics Conference. Copper Mountain, CO. [8]

- Zarchan, Paul. 1994. *Tactical and Strategic Missile Guidance*. 2nd ed. Washington, DC: AIAA Press. [10]

- f and g functions**, 82
- 2-line element sets** 707
 - converting time, 201
- 3-hourly index**, 556
- 4-simplex geometry**, 897
- A**
- Aberration**
 - annual, 311
 - correction for, 311
 - diurnal, 311
 - light time, 311
 - planetary, 312
 - secular and stellar, 311
- Acceleration**
 - centripetal, 167
 - Coriolis, 167
 - inertial
 - in three-body system, 34, 35
 - of interceptor satellite, 391
 - inertial, 22
 - tangential, 167
- Accuracy**, 600, 704
- Adams, John Couch**, 609, 610
- Adams-Basforth-Moulton method** *See Integrators, numerical*
- ADBARV elements**, 109
- AFSCN, Air Force Satellite Control Network**, 251
- Air Force Satellite Control Network (AFSCN)**, 251
- Air Force Space Command (AFSPC)**, 105
- Airy, Sir George Biddell**, 611
- Albedo**, 583
- ALCOR, Advanced Research Projects Agency (ARPA) - Lincoln C-Band Observables Radar**, 253
- d'Alembert, Jean le Rond** 609
- Alfano transfer**, 381, 385
 - definition of, 372
- Alfano-Negrone Close Approach Software (ANCAS)**, 924–936
- Almagest**, 5
- Along-track variations**, 156
- Alphonsine tables**, 7
- Altitude**
 - in coordinate system, 154
 - of an orbit, 18
- Ambiguity**, in signal coordination, 903
- Anderle, Richard**, 732
- Angular momentum, specific**, 23
- Annular eclipse**, definition of, 301
- Anti-Satellite Test**, 919
- Anti-spoofing (A-S)**, 901
- a_p , **planetary amplitude**, 556
- Apoapsis**, definition of, 17
- Apollonius**, 3
- Apollo-Soyuz program**, 318
- Apollo-Soyuz Test Project**, 572
- Apsidal rotation**, 661
- Argument of latitude**, definition of, 102
- Argument of perigee**
 - definition of, 99
 - frozen, 886
- Aries, first point of**, 147
- Aristarchus**, 3
- Aristotle**, concepts of motion in, 20
- ARPA Long-Range Tracking and Instrumentation Radar (ALTAIR)**, 253
- Ascending node** *See Node, ascending*
- Aspect ratio**, 931
- Aspherical-potential function**
 - common notations for, 544–545
 - deriving
 - MacCullagh's method, 540–542
 - partial derivatives for, 549
 - spherical harmonics method, 547
 - deriving, 538–550
 - orbital elements method, 641–643
- Associated Legendre functions**, 543
- AST**, *See Sidereal time, apparent*
- Astrodynamics**, definition of, 1
- Astrolabe**, description of, 6
- Astrology**, 1
- Astronomical unit, AU**, 131
- Asymptotic expansions**
 - definition of, 614
 - Theory of, 618
- Atmospheric density**, determining, 1001–1015
- Atmospheric models** 565–572, 1001–1004
 - static, 553
 - time-varying, 554
 - types of
 - CIRA, 568
 - DTM, 571
 - exponential, 565
 - Harris-Priester, 568
 - Jacchia-Roberts, 569, 1001–1010
 - NRLMSIS, 571
 - Russian Dynamic Calibration, 572

Russian GOST, 571
 Soviet Cosmos, 1011–1015
 standard atmosphere, 568
 variations in, 553–555

Atomic time, international, 190

Attitude dynamics, definition of, 1

Attitude, of a satellite, 103, 598

AU, *See also Astronomical unit* 131

Australian Geodetic Datum, 143

Auxiliary circle, for eccentric anomaly, 44

Avogadro’s number, 1010

Azimuth-elevation

- Azimuth, definition of, 154
- conversions for, 261–268

B

B1950 system, 234

Babylonians, 2

Ballistic coefficient (BC), 106, 552

Barker, Thomas, 51, 68

Barycenter, definition of, 145, 149

Barycentric Celestial Reference System, BCRS, 151

Basic kinematic equation, 167

Batch processing, definition of, 734

Bayes filter, 772

Bayes Theorem, 772

BC, *See Ballistic coefficient*

Beamwidth, definition of, 245

Benjamin Franklin, 181

Bessel functions, 72, 674–675

Bessel, Frederick, 312

Besselian year, definition of, 234

Beta angle, 305

Bias, definition of, 740

BIH, *See Bureau International de l’Heure*

Blending techniques 1031–1038

- parabolic, 1031
- quartic, 1036

Blue Moon, 294

Board of Longitude, 895

Bode’s law, 425

Boresight angle

- definition of, 856
- solving for, 857

Borgi, Jorst, 517

B-Plane, 961

Brahe, Tycho, 9, 41, 42, 241

Bretagnon, method of, 296

Brouwer, Dirk, 612

Brouwer’s method, 694–696

Brown, Ernest, 286, 610

Brun’s theorem, 143

Bstar, 106

Buran, 420–421

Bureau International de l’Heure (BIH), 209

C

c(ψ) functions, 62–63

Calibration, 252

Canonical elements, 110–111

Canonical transformations, in equations of motion, 638–640

Canonical units, development of, 237–238

Cavendish, Henry, 130

Cayley-Hamilton theorem, 812

Cefola, Paul, 612

Celestial

- equator, 145
- latitude and longitude, 147
- mechanics, 1st use of, 425
- mechanics, definition of, 1
- poles, 145
- reference frames, 145
- sphere, geometry of, 145

Celestial Intermediate Pole, 207, 603

Center for Space Standards and Innovation (CSSI), 921

Central differencing, 764

Central limit theorem, definition of, 744

Central-body effects 647–671

- categories of, 657
- long- and short-periodic, 662
- m-daily, 663
- resonance, 665–670
- secular, 658–662

Centripetal acceleration, 167

Ceres, 425, 731

Chaldeans, 1

Charge Coupled Device, CCD, 255

Chauvenet’s criterion, 750

Cholesky decomposition, 820

CIO, *See Conventional International Origin*

Circular Restricted Three-Body Problem (CRTBP), 968

CIS, *See also Coordinate systems*

- Conventional Inertial 151

Clairaut, Alexis Claude, 609

Classical orbital elements, 95–105

- illustration of, 96

- Clohessy-Wiltshire equations** *See Hill's equations*
- Close Approach processing**, 919–936
- Close-approach determination**
definition of, 839
- Close-approach determination**, 919–936
type I and Type II errors, 923
- CNES, Centre National d'Études Spatiales**, 250
- Coarse-Acquisition code (C/A), from GPS**, 901
- Collinear Lagrange points**, 971
- Collision Avoidance**, 919
- Colosseum**, 888
- Commission, errors of**, 795
- Condition of osculation**, 620
- Conic sections**
basic parameters, 12–20
definition of, 12
directrix, 14
foci of, 12
geometry of, 12–20
- Conjugate coordinates** *See Canonical elements*,
- Conjunction**
inferior, 314
superior, 315
- Conjunction assessment**, 919
- Conjunction processing**, 919–936
- Conjunction Sieve, CSIEVE**, 922
- Consider parameters**, 748
- Constants, physical**, 237–238, 1043
- Continued fractions**, 494, 668
- Contour grids**, 144
- Control law**, for short-time orbital transfer, 382
- Conventional International Origin (CIO)**, 209
- Coordinate rotations** 160–166
using quaternions, 165
using vectors, 163
- Coordinate systems** 145–159
barycentric, 149
Body-fixed (ITRF), 152
common transformations, 167–169
Conventional Inertial (CIS), 151
Earth-based, 151–154
Earth-Centered Inertial (ECI), 151
Earth-centered, Earth-fixed (ECEF), 152
Earth-Fixed Greenwich (EFG), 152
equinoctial (EQW), 158
- Fundamental Katalog (FK5), 151
Geocentric Celestial (GCRF), 152
geocentric equatorial (IJK), 151
heliocentric (XYZ), 150
IAU-76/FK5, 151
ICRF, 151
inertial, 22
interplanetary, 150
mean of date, 227
perifocal (PQW), 155
pseudo-inertial, 204
quasi-inertial, 205
roll-pitch-yaw (RPY), 156
rotating transformations, 166–167
satellite (NTW), 157
satellite (RSW), 155
satellite-based, 154–159
summary of, 159
synodic, 149
topocentric equatorial, 154
topocentric horizon (SEZ), 153
transforming celestial and terrestrial, 203–236
true of date, 223
- Coordinate time** 193–194
barycentric, 194
geocentric, 194
- Coordinate transformations**, 159–169
- Coordinated universal time (UTC)**, 180
- Copernicus, Nicholas**, 7–8
- Coriolis acceleration**, 167
- Corner-cube retro-reflectors**, use of, 243
- Correlation coefficients**, 747
- COSEA, Committee on Extension to the Standard Atmosphere**, 568
- Cost functions**, definition of, 372, 735
- Council of Trent, and Gregorian calendar**, 4
- Covariance matrix**, 746
- Cowell, Phillip Herbert**, 518
- Cowell's formulation**, 518, 525–526
- Cowell's method**, 526
- Critical inclination**, 660
- Cross-sectional area**, of a satellite, 552
- Cross-track variations**, 156
- Cubic roots** *See Roots*
- Culmination**, definition of, 830
- D**
- Data batches**, definition of, 734
- Data**, for orbit determination, 828

- Day of the year**
 deriving, 200
 values for, 184
- Daylight Savings Time, DST**, 181
- DCA**, *See Russian Dynamic Calibration of the Atmosphere*
- DChandler period**, 210
- Debris, orbital**, 848, 850
- Declination** 147
 geocentric, 256–259
 topocentric, 259–261
 transformations with
 azimuth and elevation, 266–268
 ecliptic latitude and longitude, 268–270
 topocentric and geocentric, 261
- Deep Space Network (DSN)**, 251
- Deep-space satellites**
 definition of, 31
- Deferent**, definition of, 4
- Deflection of Light**, 314
- Deflection of the vertical**, 135, 427
- Del operator (gradients)**, use of, 522
- Delaunay arguments**
 reduction theories, 210
- Delaunay elements**, 111, 694
- Delaunay, Charles E.**, 610
- Density, atmospheric**
 definition of, 552
 magnetic field models for, 556
- Density, atmospheric**, 678
- Deprit, Andre**, 610, 661
- Desaturation of momentum wheels**, 965
- Differential correction** *See also Estimation of orbits* 453, 734, 756
- Digital gravity models**, 717
- Dilution of precision (DOP)**, 907
- Dilution region**
 conjunction probability, 933
- Direct effect**, in three-body motion, 35
- Direction cosines**, 1019
- Directrix**, of a conic section, 14
- Displacements** *See also Relative positions, rendezvous* 156, 398–411
 along-track 156
 cross-track 156
 in-track 157
 radial 156
- Disturbing functions**, 522–523
- Divergence**, in data filters, 800
- Doppler shift** 244, 245
 second-order, relativistic correction for, 898
- DORIS**, 249, 250
- Drag**, aerodynamic, 551, 671–679
 coefficient of, 551, 679
 drag paradox, 676, 677
 effects on orbits, 873, 875
 equations for, 672
 spacecraft attitude effects, 598
- DRAO, Dominion Radio Astrophysical Observatory**, 560
- Draper Semianalytical Satellite Theory (DSST)**, 699–704
- Drift (data)**, definition of, 740
- DSCS, Defense Support Communication Satellite**, 849
- DSN** *See Deep Space Network*
- DSP, Defense Support Satellite**, 849
- DSST** *See Draper Semianalytical Satellite Theory*
- DST**, *See also Daylight Savings Time* 181
- Dynamical time**, 190–193
 barycentric, 191
 calculating, 195
 terrestrial, 190
- E**
- Earth**
 eccentricity of, 132
 fundamental parameters, 131–133
 location parameters, 133
 mean equatorial radius, 132
 orbital eccentricity, 278
 period of rotation, 4
 radius (ER)
 definition of, 237
 estimation of, 3
 rotational velocity, 132
 semimajor axis, 132
 shape of
 models for, 134
 shape of, 134–141
 zonal gravitational coefficient, 547
- Earth Gravity Model (EGM-96)**, 1039–1040
- Earth Orientation Parameters (EOP)**, 191, 603, 1050
- Earth Rotation Angle**, 212
- Earth-Moon barycenter**, 151
- Earth-Moon system**, regions of motion in,

- 973, 975
- Easter**, date of, 6
- East-west stationkeeping**, 851
- Eccentric anomaly** 44
 - alternate formulation, 49–50
- Eccentricity**
 - in computer applications, 98
 - of a conic section, 14
 - of an orbit
 - definition of, 97
 - frozen, 886
 - of Earth, 132
 - of Earth's orbit, 278
- ECEF**, *See also Coordinate systems* 152
- ECHO 1**, 846
- ECI**, *See also Coordinate Systems* 151
- Eckert, Wallace**, 610
- Eclipses** 2, 299–301
 - lunar, 300
 - penumbra region, 299
 - solar, 301
 - umbra region, 299
- Ecliptic latitude and longitude**
 - transformations for, 268–275
- Ecliptic of date**, definition of, 207
- Ecliptic system**
 - transformation to equatorial, 268
- Ecliptic**, obliquity of, 146
- EFG**, *See also Coordinate systems* 152
- EGM-96**, 601, 603
- Einstein, Albert**, 898
- EKF, Extended Kalman Filter**, 793, 797
- Electromagnetic spectrum**, 243
- Element set**, definition of, 95
- Elements**
 - 2-line element sets, 691, 707
 - ADBDRV, 109
 - classical orbital, 95–105
 - converting orbital 708–711
 - equinoctial, 108
 - flight, 109
 - Geographic, 110
 - mean, 616, 694, 703, 707
 - orbital
 - converting with position and velocity vectors, 112–116
 - osculating, 95, 616, 702, 707
 - two-body, 95, 707
 - two-line element sets, 105–108
- Elevation**, definition of, 154
- Ellipse**, fundamental, definition of, 473
- ELLIPSO**, 698
- Ellipsoidal function, for satellite orbits**, 928
- Ellipsoidal models**, 134
- Emissivity**, 583
- Encke, Johann Franz**, 518
- Encke's formulation**, 523–525
- Encounter region**, 926
- Energy integral**, 26
- EOP** *See Earth Orientation Parameters*
- Ephemerides**
 - definition of, 600
 - planetary, 295–298
- Ephemeris time (ET)**, 190
- Epicycle system**, 4, 5
- Epoch**, definition of, 174
- Equant**, 5
- Equation of the center**
 - for satellite motion, 74
- Equation of the Equinoxes**, 186, 224
- Equation of Time**, 178
- Equations of motion**
 - n*-body, 33–972
 - polar form, 375
 - three-body
 - barycentric form, 36
 - center of mass in, 38
 - constant, 970
 - equilibrium solutions, 971
 - general series representation, 35
 - mass ratio, 970
 - phase space, 971
 - relative formula, 35
 - restricted, 40–977
 - ten known integrals for, 37
 - two-body 20–33
 - assumptions for, 23
 - basic, 23
 - derivation of, 21
 - relative form, 23
 - specific angular momentum, 23
 - specific mechanical energy, 25
- Equator crossings**
 - nodal drift effect on, 871
- Equator**, definition of, 134
- Equatorial plane**, definition of, 134
- Equatorial system**
 - transformation to ecliptic, 269
- Equidistant Cylindrical system**, 412
- Equilateral hyperbola**, use of, 54
- Equilibrium solutions**, for three-body problem, 971

- Equinoctial elements**, 108
Equinox
 definition of, 146
 vernal, 147
Equinoxes, equation of, 186, 224
EQW, *See also Coordinate systems*
 equinoctial 158
ER, *See Earth, radius*
Eratosthenes, 3
Error
 analysis, in data estimation, 739–748
 ellipsoid, 920, 928
 instrument/measurement, 740
 time-varying, 740
Error budget, for perturbation analysis, 705
Escape velocity, definition of, 33
Estimation of orbits
 availability, quantity of data, 829
 convergence issues, 816
 deterministic, 733
 EKF, nonlinear systems, 793–795
 error analysis in, 739–748
 geometric constraints, 830
 Kalman filtering
 summary, 798–800
 Kalman filtering, 777–800
 KF, linear systems, 785–789
 linear least squares
 unweighted, 734
 weighted, 750–752
 linear least squares, 734–752
 LKF, nonlinear systems, 790–792
 mathematical modeling for, 741
 Maximum likelihood, 732
 nonlinear least squares, 752–759
 post-processing vs. prediction, 821
 predicting, filtering, smoothing, 733
 probability techniques for, 742–750
 propagation models, 813
 propagator selection, 831
 sequential-batch least-squares techniques, 772–776
 stochastic, 733
ET, *See also Time, ephemeris* 190
Euclid, 3
Euler angle rotation, 160
Euler Angles, 160
Euler integrator, 528
Euler, Leonhard, 12, 424, 609, 610
European Datum, 143
European Space Agency (ESA), 250
Evection, 287
Excentric system, 4, 5
 moveable, 5
Extreme ultra-violet (EUV) radiation, 560
- F**
- Fading-memory filters**, definition of, 819
Fast variables, of orbital elements, 615
Fermat, Pierre de, last theorem, 17
Fictitious Mean Sun (FMS) 178
 right ascension of, 178
 use of, 176, 178
Fiducial direction, 146
Field of view
 calculating, 859
 definition of, 855
 geometry for, 855, 856, 858
Filters
 fading-memory, 819
 Kalman, 777–800
 square-root, 820
 tuning, 785
 types of, 785
Finite differencing, 763, 808
Fisher, Ronald A., 732
Fit spans, definition of, 815–816
Fixed-Dv *See Orbital Transfers*
FK4, 234
FK5 system *See Coordinate systems, IAU-76/ FK5*
Flamsteed, John, 423
Flattening
 of a conic section, 15
 of the Earth, 132
Flight elements, 109
Flight-path angle 103–105
 geometry of, 19, 25
FMS, *See Fictitious Mean Sun*,
Forces, disturbing *See Perturbations*
Formation Flying, 388–419
Free Core Nutation, 217
Free-return trajectories, 977
Frenet system *See Coordinate systems, satellite, NTW*,
Frozen orbits, 847, 849, 880, 885–888
Functions, disturbing-potential, 522
Fundamental arguments
 reduction theories, 209
Fundamental ellipse, 473, 494

- Fundamental Katalog system**, 151
Fundamental period, 658
- G**
- Gabbard classes**, 846
Galilei, Galileo, 8, 10
Galle, Johann Gottfried, 611
Gauss, Karl Friedrich, 130, 425, 731, 896
Gaussian coefficients, 587
Gaussian constant of gravitation, 130
Gaussian coordinate system *See Coordinate systems, satellite, RSW*
Gaussian variation of parameters, 319, 628–637
Gauss-Jackson method, 534
GCRF, *See also Geocentric Celestial Reference Frame* 152
Gemini program, U.S., 318
General perturbation techniques *See Propagation of orbits, analytical*
definition of 613
Generalized momenta *See Canonical elements*,
Genetic algorithms, 513
GEO, Geosynchronous Earth Orbit, 31
Geocentric Celestial Reference Frame, GCRF, 151, 152
Geocentric Celestial Reference System, GCRS, 151
Geocentric values, from topocentric values, 261
Geodesy, 131
GEODSS, *See Ground-based Electro-Optical Deep Space Surveillance*
Geographic elements, 110
Geoid
definition of, 134, 142
undulation, 142
Geopotential surfaces, 142–143
Geops, definition of, 142
GEOS-III, 769
Geostationary satellites
definition of, 30
Geosynchronous orbits
semimajor axis for, 31
Geosynchronous satellites
definition of, 30
stationkeeping, 851–852
Geosynchronous satellites, 669–671
GFZ, Geo Forschungs Zentrum, 248
- GHA**, *See also Hour angle, Greenwich* 148
Gibbs method, 454–461
Gibbs, Josiah, 426
Global Positioning System (GPS) 248, 740, 798, 895–908
determining position from, 906
differential techniques for, 907
dilution of precision, 907
double-difference processing, 908
signals, 901–908
system description, 899–901
GLONASS, 849
GMST, *See Sidereal time, Greenwich Mean*
GMT, *See also Greenwich Mean Time* 178
Goddard Earth Model (GEM), 601
Goddard Trajectory Determination System (GTDS), 600, 612
Goddard, Robert, 317
GPS *See Global Positioning System*
Gradient
definition of, 522
in synodic acceleration, 969
Grand tour of the planets, 987
Gravitation, law of 129
Gravitational acceleration, 142
Gravitational coefficients
normalizing
example of, 546–547
normalizing, 545
sectoral table, 1040
using, 601
zonal table, 1039
Gravitational constant, 130
Gravitational model, 141–144
Gravitational parameter, measurement of, 130
Gravitational potential, deriving, 539–547
Gravity assist trajectory, 958
Gravity, mean value of theoretical, 141
Gravity, theoretical, 141
Great circle, definition of, 145
Greenwich Mean Sidereal Time, GMST, formula for, 187
Greenwich Mean Time (GMT), 178
Greenwich meridian, definition of, 134
Gregorian calendar, 5
Grissom, Virgil, 318
Ground-based Electro-Optical Deep Space Surveillance (GEODSS), 252
Groundtracks
definition of, 120

- determining period, 122
eccentricity and perigee location effects, 124
examples of, 120–123
inclination effects, 125
repeat, 869–879, 888–894
shift parameter, 871
- H**
- Halley, Edmond**, 11, 423
Halley's comet, 423–424, 518
Hamilton, William Rowen, 637
Hansen coefficients, 642, 644
Hansen, Peter A., 609
Harmonics, spherical, 547–549
Harmony of the spheres, definition of, 3
Harrison, John, 895
Height
 ellipsoidal, 143
 orthometric, 142
Heliocentric system, parallax in, 312
Herg, 238
Herrick, Samuel, 426
Herrick-Gibbs method, 461–467
High Precision Orbit Propagator (HPOP), 712
Hill, George W., 610
Hill's equations
 solutions for, 393, 396, 410
Hill's equations, 388–419
Hinge symmetry for groundtracks, 122
Hipparchus, 3
Hohmann transfer *See Orbital transfers*
Hohmann, Walter, 322
Homotopy continuation method, 453
Hooke, Robert, 11
Hour angle
 Greenwich, 148
 local, 148
Hour circles, definition of, 145
Hubble Space Telescope, 239, 397, 410
Hydrostatic delay, 247
Hydrostatic equation, 553
Hyperbolic anomaly
 definition of, 54
 formulation for, 52–57
Hyperbolic excess velocity, 27
- I**
- IAU Theory of Nutation, 1980**, 224
IAU, International Astronomical Union, 130
ICRF, *See also Coordinate systems* 151
Ideal rocket equation, 373
Ideal-gas law, 553
IERS Reference Meridian, 209
IERS, *See International Earth Rotation Service*
Ignorable coordinate, definition of, 639
Illumination, ground, 309
Inclination
 changing, 342–345, 348, 351–353
 critical, 660
 definition of, 98
Indirect effect, in three-body motion, 35
Inertial acceleration, for three-body problem, 969
Inertial formula, for three-body motion, 33
Inferior planets, definition of, 6
Information matrix, 738
Initial Orbit Determination *See Orbit determination, initial*
Integrals of the motion 37
Integration, numerical
 determining step size, 536
 fixed step size, 527
 implementing, 535–537
 methods of, 526–537
 order of, 529
 ordinate, summed-ordinate forms, 533
 regularization, 537
 s-integration, 537
 single-step, 526
 t-integration, 537
 time-regularized Cowell, 537
 variable step size, 527
Integrators, numerical
 Adams-Basforth-Moulton, 531
 Burlisch-Stoer, 528
 Double integration, 527
 Euler, 528
 multi-step, 527
 non-summed forms, 527
 predictor-corrector, 527
 Runge-Kutta, 526, 528
 Sampine-Gordon, 531
 Shampine-Gordon, 528

- Single integration, 527
 Sum-squared, Cowell-Stormer, Gauss-Jackson, 527
 summed forms, 527
- Intercept orbits**, 499
- International Celestial Reference System (ICRS)**, 151
- International Earth Rotation Service (IERS)**, 208
- International Geomagnetic Reference Field (IGRF)**, 587
- International GPS Service (IGS)**, 900
- International Laser Ranging System (ILRS)**, 248
- International Polar Motion Service (IPMS)**, 208
- International Reference Pole**, 209
- International Scientific Optical Network (ISON)**, 251
- International Space Station**, 921
- International Terrestrial Reference Frame (ITRF)**, 145, 152
- Interpolation**, linear, 1030
- In-track variations**, 157
- Invariable plane of Laplace**, 39
- Ionosphere**
 correcting for, 246
 model, 246
- IRM**, *See also IERS Reference Meridian* 209
- ISS, International Space Station**, 726
- J**
- Jacchia-Roberts atmospheric model**, 569, 1001–1010
- Jacobi integral equation**, 970
- Jacobi, Carl**, 12
- Jacobi's constant**, 970
- Jacobian equation**, 970
- Jacobian matrix, 779
- JD**, *See Julian Date*
- Jet Propulsion Laboratory**, 277, 286, 1055
- JGM**, *See Joint Gravity Model*
- Joint Gravity Model (JGM)**, 601
- Joseph algorithm**, 784
- Julian calendar**, 4
- Julian century**, 4
- Julian date** 182–184
 formula for, 182
 modified, 183
 to Gregorian date, 202
- Julian period**, 182
- Julian year**, 4, 182, 202
- K**
- Kalman filtering** *See also Estimation of orbits*
 777–800
 extended, EKF 793–795
 linear, KF 785–789
 linearized, LKF 790–792
 non-linear system use 790–795
 summary 798–800
- Kalman gain matrix**, 783
- Kalman, Rudolf E.**, 732
- Kalman-Bucy method** *See Kalman filtering, extended*
- Kaula gravitational coefficient**, 656
- Kepler, Johann** 9
 calculating Mars orbit, 41–42
 Rudolphine tables, 423
- Kepler's equation** 45, 46
 solutions for
 elliptical, 64–67
 hybrid, 74–76
 hyperbolic, 70–72
 Newton-Raphson iteration, 64–65, 74, 453
 Newton's method, 64
 parabolic, 67–70
 series, 72–74
 solutions for, 43–79
 summary of formulas, 77–79
 universal formulation, 58–63
- Kepler's Laws** 10, 43, 129
 first, 27–29
 second and third, 29–32
- Kepler's problem** 43, 79–95
 solutions for
 first-order approximation, 690
 orbital-element technique, 80–81
 series forms of f and g , 86
 universal variables technique, 87–95
 using f and g functions, 82–87
- Keplerian elements** *See Classical orbital elements*,
- Keplerian period** *See Period of a satellite, anomalistic*
- Kolmogorov, Andry N.**, 732
- Kozai, Yoshio**, 611
- Kozai's method**, 692–694
- K_p** , planetary index, 556
- Kutta, Wilhelm**, 528

L

- LAGEOS** *See Laser Geodynamics Satellites*
- Lagrange brackets**, 622
- Lagrange multipliers**, 375
- Lagrange planetary equations** 621, 626
 - equinoctial form, 628
 - Gaussian form, 636
- Lagrange points**, 971
- Lagrange, Joseph**, 12, 424, 610
- Lagrange's interpolation formula**, 436
- Lambert, Johann H.**, 424, 467
- Lambert's problem** 318, 424, 467–498
 - Battin method, 493–498
 - Gaussian solution
 - first and second equations, 477
 - Gaussian solution, 475–479
 - intercept and rendezvous, 498–513
 - Legrange time equation, 471
 - minimum eccentricity, 470, 473
 - minimum energy, original formulation
 - fundamental ellipse, 473
 - minimum energy, original formulation, 470–475
 - Thorne's solution, 480–489
 - universal-variable solution, 489–493
- Lambertian diffusion**, 582
- Landgrave William IV**, 241
- Landsat**, 665, 726, 838
- de Laplace, Pierre-Simon** 425, 517, 609
- Laplace transform**, 394
- Laser clearinghouse problem**, 938
- Laser Geodynamics Satellite, (LAGEOS)**, 252, 726
- Laser, Light Amplification by Stimulated Emission of Radiation**, 243
- Latitude**
 - astronomical, 135
 - definition of, 134
 - ecliptic, 147
 - geocentric
 - definition of, 134
 - determining, 173
 - geodetic
 - definition of, 134
 - determining, 137, 170–173
 - geometry of, 136
 - reduced, 135
 - transformations between, 140
- Launches**
- direct**, 337
- locations for**
 - velocity of, 339
- locations for**, 337, 339–341
- windows for**, 339–341
- Law of Gravitation**, 21, 129
- Le Verrier, Urbain-Jean-Joseph**, 610
- Lead angle** 361
- Leap seconds**, use of, 180
- Leap year**, 5
- Least squares techniques**, implementing, 765–772
- Legendre functions**, 543
- Legendre polynomials**, 540
- Legendre, Adrian Marie**, 518, 731
- Length of day, LOD**, 191, 222
- Lenhart adjusted data**, 561
- LEO, Low Earth Orbit**, 31
- Leonardo of Pisa**, 241
- Leonov, Alexei**, 318
- LHA**, *See also Hour Angle, Local* 148
- Light**, correcting for deflection of, 314
- Lighting conditions**, 305–309
- Lilius, Aloysius**, 5
- Line of nodes**, definition of, 99, 146
- Line of sight (LOS)**
 - determining, 305–309
 - geometry for, 306
 - with Sun, 308
- Line symmetry for groundtracks**, 121
- Linear interpolation**, 1030
- Linear least squares** 734–752
 - unweighted, 734
 - weighted, 750–752
- LKF, Linearized Kalman Filter**, 790
- Local horizontal**, definition of, 18
- Local Time, LT**, 181
- Longitude**
 - definition of, 133
 - ecliptic, 147
- Longitude of periapsis**, 101
- Look angles, predicting satellite**, 908–913
- LOS**, *See also line of Sight* 305
- Love numbers**, 646
- LST**, *See Sidereal time, local*
- Luna-1 and Luna-2**, 317
- Lux**, 311
- LVLH (local vertical, local horizontal)**, 156

M

- MacCullagh's formula**, 542
Macro models, 584, 599
Main Problem, theory of, 618, 661–662
Maneuver
 detection of, 832
 recovery of, 832
Maneuvers
 combined, 350–357
 coplanar, 320–337
 fixed-Dv, 354–357
 intercept and rendezvous
 relative motion in, 388–419
 intercept and rendezvous, 357–371, 499–513
 maximum efficiency, 335–336
 minimum-inclination, 351–353
 noncoplanar, 337–349
 orbital-maintenance, 875–879
 orbital, definition of, 317
Maneuvers *See also Orbital transfers*
Manifold techniques, for analyzing regions of motion, 978
Manifold, 977
Mathematical modeling
 in data processing, 741
Matrices, fundamentals of, 1019
Matrix inversion lemma, 782
Matrix inversion, 1019
Matrix Riccati equation, 781
Matrix, 1019
Maximum likelihood, 732, 777
Maximum Probability, 931
m-daily variations, 663–664
Mean anomaly, definition of, 45, 46
Mean elements, 616, 694, 703, 707
Mean equator of date, 207
Mean latitude, longitude at epoch, 103
Mean longitude, 103, 108
Mean motion, of a satellite 97
 converting units, 45
 definition of, 45
 distribution of, 847
Mean Sea Level, determining, 142
Mean sidereal day, definition of, 180
Mean, statistical, definition of, 742
Mechanized Algebraic Operations, MAO, 610
MEO, Mid Earth Orbit, 31
Meridian, definition of, 134
Meteoroid and Space Debris Terrestrial Environment Reference (MASTER), 849
Midcourse Space Experiment (MSX), 253
Millstone Hill Radar, 732
Minkowski, Herman, 898
MIR, 318
Mistagged satellites, definition of, 841
MJD, *See also Modified Julian Date* 183
MOD, *See mean equinox of date* 226
Modified Julian Date, MJD, 183
Molniya orbit 714
Molniya orbit
 definition of, 848
Molodensky formulas, 143
Monte-Carlo methods, 742
Moon
 blue moon, 294
 gibbous, 294
 illumination, 310
 phases of, 294
 waxing and waning, 294
Moonrise, Moonset, determining, 290–294
MOSS, Moron Optical Space Surveillance System, 253
MOTIF, Maui Optical Tracking and Identification Facility, 253
Motion, relative *See Relative motion of satellites*
MSL, *See Mean Sea Level*
 MSSS, Maui Space Surveillance Station, 253
MSX, Midcourse Space Experiment, 726
Multipath signals, definition of, 905

N

- Nadir point**, definition of, 856
Napier, John, 517
Napier's rules, 1024
NASA, National Aeronautics and Space Administration, 251
NASA's Navigation and Ancillary Information Facility, 605
Navigation, space-based, history of, 896
NAVSPASUR, Navy Space Surveillance, 252, 253
Navstar *See Global Positioning System*,
n-body equations *See Equations of motion*
Near-identity transform, 700
Neutral atmosphere, 246
Newcomb, Simon, 610

- Newton, Isaac** 41, 129, 317, 609
 developing his laws, 10–11
 method for orbit determination, 423–424
- Newton's Laws** 11
 law of gravitation, 21, 129
 second, 20
- Newton-Raphson iteration**, 64–65, 74, 453
- NOAA, National Oceanic and Atmospheric Administration**, 560
- Nodal drift, effect of**, 865–869
- Nodal period**, 869
- Nodal regression**, 206, 659
- Node, ascending**
 definition of, 99
 longitude of, 99
 right ascension of
 changing, 345–347
- Node, descending**, 99
- Noise (data)**, definition of, 740
- Nonlinear least squares**, 752–759
- Normal equations**, 738
- North American Datum**, 143
- North-south stationkeeping**, 851
- NTW**, *See also Coordinate systems* 157
- Numerical Integration**, *See Integration, numerical*
- Nutation**
 coefficients, 1043
 definition of, 206
 transformations for, 224
- Nyquist critical frequency**, 536
- O**
- Oberon**, orbit of, 103
- Oberth effect**, 950
- Oberth, Herman**, 317
- Oblate models**, definition of 134
- Oblate spheroid**, as Earth model, 134
- Oblers, Heinrich Wilhelm**, 425
- Obliquity of the ecliptic**, 146, 224
- Observability**
 criteria, 752
 definition of, 831
 requirement, 738
- Observables**
 carrier-phase, in GPS, 903
 ionospheric correction for, 905
 linearly combined, 905
- Observation Association**, 842
- Observation Correlation**, 273
- Observations**
 angles-only
 double-r iteration, 443–449
 Gauss's method, 439–443
 Gooding's method, 449
 Laplace's method, 435–439
 angles-only, 433–449
 mixed
 range and range-rate processing, 450
 range-only processing, 452
 mixed, 450–454
 range, azimuth, and elevation, 426–432
 short- and long-arc, 247
- Occultations**, 315
- ODTK, Orbit Determination Toolkit**, 795
- of date**, 211
- of epoch**, 211
- Ohio State University model**, 601
- Omission, errors of**, 795
- Orbit determination**
 close-approach, 839
 differential correction for, 759–772
 initial
 data for, 242–249
 definition of, 423
 Gibbs method, 454–461
 Herrick-Gibbs method, 461–467
 Lambert's problem *See also Lambert's problem* 467–498
 range only processing, 452–454
 range, range rate processing, 450–452
 three position vectors and time, 454–467
 initial, 423–513
 partial derivative matrix for, 762–764
 segmented BC, 814
 sensitivity matrix for, 801
 state-transition matrix, 778–780
 statistical, 731–825
 update intervals, 815
 variational equations for, 803–811
- Orbit raising** 374–381
 accumulated velocity change for, 378
 control law for, 376
 λ dynamics in, 375
- Orbital dynamics**, definition of, 1
- Orbital elements**
 definition of, 95
 fast and slow variables, 615

- from position and velocity vectors, 112–116
 - mean, 616, 694, 703, 707
 - osculating, 616
 - periodic and secular changes, 614
 - single- and double-averaged, 616
 - Orbital transfers**
 - Alfano, 381
 - bi-elliptic, 323, 325, 327–330, 336
 - changing inclination and longitude of ascending node, 348–349
 - continuous-thrust, 371–386
 - coplanar
 - comparison of 336
 - coplanar, 320–337
 - critical limits, 330
 - fixed-D_v, 351
 - general, 337
 - Hohmann
 - comparison to bi-elliptic, 328–330
 - critical limits, 330
 - Hohmann, 322–330, 336
 - inclination-only, 342–345
 - low-thrust, noncoplanar, 381–386
 - node-only, 345–347
 - noncoplanar, 337–349
 - nontangential, 321
 - one-tangent burn, 331–337
 - ratio for, 329
 - tangential, 320, 321
 - Orbits**
 - circular, 16
 - closed and open, 13
 - direct or prograde, 99
 - elliptical
 - apoapsis and periapsis, 17
 - elliptical, 15
 - equatorial, 98
 - frozen 849
 - frozen-eccentricity, 880
 - frozen, 847, 885–888
 - geostationary, 30
 - geosynchronous, 31, 847
 - hyperbolic, 18
 - inclined, 98
 - low-Earth, 31, 847
 - maneuvering for, 317–419
 - minimum-altitude variation, 849, 880–885, 894
 - mission design for, 862–895
 - mission types, 846–851
 - Molniya, 848
 - parabolic, 17
 - phasing
 - use of, 367
 - polar, 99
 - rectilinear, 12
 - repeat-groundtrack 847, 849, 869–879
 - example of, 888–895
 - initial elements for, 888–893
 - retrograde, 99
 - semi-synchronous, 30, 847
 - Sun-synchronous, 847, 849, 862–865
 - transfers between *See Orbital transfers*
- Orthogonal matrix**, 163, 1019
- Osculating elements**, 95, 616, 702
- Osiander, Andreas**, 8
- P**
- Parabolic anomaly**
 - definition of, 51
 - formulation for, 50–52
- Parabolic interpolation**, 1031
- Parabolic mean point radius**, 494
- Parallax**
 - annual, 312
 - correction for, 312
 - definition of, 312
 - diurnal, 312
 - example values, 312
 - geocentric and heliocentric, 312
 - horizontal, 312
 - Moon's, 287
- Parameter estimation**, 739
- Pass**, satellite, definition of, 121, 911
- Patched-conic approximation**, 944–987
- Payload angle 355**
- Payload Assist Module**, 420
- Penalty functions**, definition of, 372
- Penumbral eclipse**, definition of, 300
- Performance indices**, definition of, 372
- Periapsis**
 - definition of, 17, 99
 - longitude of, 101
 - true longitude of, 101
- Period**, of a satellite
 - anomalistic, 869
 - equation for, 30
 - nodal, 869
- Period**, of a satellite, 29
- Periodic effects**, in orbital elements, 615

Perturbation effects

central-body
 categories of, 657
 sectorial and tesseral, LP/ m -daily, 664
 sectorial and tesseral, LP/ m -daily/SP
 linear combination, 665
 sectorial and tesseral, m -daily, 663–664
 sectorial and tesseral, resonance, 665–670
 zonal, long-periodic, 662
 zonal, LP/SP/beat, 663
 zonal, secular, 658–663
 zonal, short-periodic, 663
 central-body, 647–671
 drag, 675–679
 solar-radiation, 688–689
 summary, 727, 728
 third-body, 682–685

Perturbation techniques

accuracy in, 705–706
 analytical *See also Perturbation techniques, general* 613
 Cowell's method, 526
 Encke's method, 523–525
 general
 2-body, 79–95
 Brouwer's method, 694–696
 Kozai's method, 692–694
 method of, 617–618
 PKEPLER, 690–692
 general, 689–698
 initial conditions for, 706–708
 numerical
 complex acceleration model, 596–599
 order of, 617
 simplified acceleration model, 593–596
 numerical, 526–537
 physical data sources
 atmospheric and solar, 604, 1053–1056
 earth orientation parameters, 603
 physical data sources, 601, 1050–1056
 practical considerations, 599–604, 704–711
 semianalytical
 Draper Semianalytical Satellite The-

ory, 699–704

semianalytical, 613, 698–704
special *See also Perturbation techniques, numerical* 517–604

Perturbation techniques *See Variation of Parameters***Perturbations** 538–591

atmospheric drag
 acceleration, 551–552
 atmospheric drag, 551–571
 categories of, 657
 central-body
 deriving aspherical-potential, 538–546
 partial derivatives, 549–550
 conservative and nonconservative, 521
 definition of, 519
 magnetic field of the Earth, 587–589
 solar-radiation pressure
 shadowing analysis, 301–584
 solar-radiation pressure, 578–584
 summary of, 724–725
 third-body effects
 acceleration, 574–576
 partial derivatives, 577–578
 third-body effects, 574–578
 thrust, 589
 tidal, 585, 645

Perturbations, linearized equations

central-body, 647–654
 drag, 672–675
 solar-radiation pressure, 685–688
 third-body, 679–682

Phase angle 358

definition of, 856

Phase lag, 646**Phase space**, in regions of motion, 971, 978**Phasing, orbital**

circular coplanar, 358–365
 circular noncoplanar, 365–371

Piazzi, Giuseppe, 425**Pinch Point, Geosynchronous**, 272**Pisa, leaning tower of**, 10**Planck's constant**, 578**Plane change** *See Inclination, changing***Planetary amplitude**, 556**Planetary constants**, 1041, 1042**Planetary ephemerides**, 604, 1046–1050

calculation of, 295–298

Planetary index, geomagnetic, 556

- Planetary symbols**, 1041
Plato, 3
Pochhammer symbols, 481
POE, *See also Precision Orbit Ephemerides* 1055
Poincaré elements, 111
Poincaré, Henri, 12
Poisson brackets, 626
Poisson series, example of, 616
Polar motion
 transformation for, 207, 223
Polar Motion, IAU 2000
 transformation for, 212
Polar orbits, 99
Pole tide, 586
Pole, 134
Polynomial solutions, 1027–1030
Polynomial wiggle, 1034
Population mean, 742
Porkchop plot, 954
Position vector
 converting to latitude and longitude, 169–173
 Moon, 286–289
 Sun, 277–281
Positive definite, 738
Potential function
 central-body, COE's, 641–643
 general, 522
 third-body, COE's, 643–645
 tidal, COE's, 645
Potential function,
Power-density matrix, 742
PPT3, Position Partials and Time, 697
PQW, *See also Coordinate systems* 155
Precession
 general, 206
 luni-solar, 206
 planetary, 206
 transformation for, 226–228
Precession of the ecliptic, 206
Precession of the equator, 206
Precession, 1976 IAU Theory, 226
Precession, P03 model, 204
Precession-Nutation, IAU 2000
 transformation for, 213–215
Precise conversion of elements, 710
Precise Positioning Service, 901
Precision code, from GPS, 901
Precision Orbit Ephemerides (POE), 600
Precision, 600, 704
Prediction problem
 definition of, 909
Primary hour circle, 148
Prime meridian, definition of, 134
Probability density function, 743
Process noise
 definition of, 741, 780
 second moment of, 742, 781
 techniques for modeling, 799
Prolate models, definition of, 134
Propagation of orbits 79–95
Propagation of orbits *See also Perturbation techniques*
 analytical 689–696
 example 614
 numerical 591–599
 example 520
 weak time dependance 595
 semianalytical 698–704
 stability 597
 two-body 79–95
 validation 599
Propellant mass fraction, 374
Proper motion, correcting for, 314
Pseudoinertial systems, 204
Pseudorandom noise, 901
Pseudorange
 definition of, 902
 effects on, 904
Ptolemaic tables, 6
Ptolemy, 4, 5, 41
Pythagorean theorem, 15, 1024
Pythagorous, 3
- Q**
- Quadratic, quartic roots** *See Roots*
Quasi-inertial systems, 205
Quaternions, 169
- R**
- Radar, Radio Detecting and Ranging**, 243
RADARSAT, 838, 864
Radial component, of position vector 26
Radial variations, 156
Radial velocity, correcting for, 314
Radioisotope thermal generators, RTG, 943
Radius of apoapsis, periapsis, 18
Radius of convergence, 756
Radius of curvature in the meridian, 138
Reconnaissance from space, definition of,

- 853
- Rectification point**, 524
- Rectilinear orbits**, 12
- Recursions** *See also Legendre polynomials* 597
- Redshift correction, for relativistic effects**, 898
- Reduction of coordinates**, 203–236
- Reduction, 2000 IAU Theory**, 211
- Reference frames**
 - celestial, 145
 - ICRS, 151
 - inertial, 204
 - pseudo-inertial, 204
 - quasi-inertial, 205
 - terrestrial, 145
- Reflectivity**, 580
- Refraction**, correcting for, 246
- Regularization**, 537
- Relative formula**, for three-body motion, 33
- Relative motion of satellites** *See also Hill's equations*
 - analyzing 388–419
 - coordinate system geometry for 389
 - transformation to inertial coordinates 411–419
 - trend analysis for 398–411
- Relative positions**
 - coordinate system for, 156
 - rendezvous, 412
- Relativistic corrections**, 314
- Relativity, general and special**, 898
- Rendezvous, circular** 357–371
 - circular coplanar phasing, 358–365
 - circular noncoplanar phasing, 365–371
 - relative motion in, 388–419
- Repeat Groundtracks**, 869
- Residual**, definition of, 735
- Resonance**
 - anomalistic, 666
 - deep, 667
 - effects of, 665–670
 - parameter, 666
 - repeat ground-track, 666
 - shallow, 667
- Restricted three-body problem**, 967–977
- Retrograde factor**, in equinoctial elements, 108
- Retrograde motion**, 6, 99
- Retro-reflectors**, corner cubes, 243
- Revolution**, satellite, definition of, 121
- Rheticus, Georg**, 8
- Right ascension**
 - geocentric, 256–259
 - topocentric, 259–261
 - transformations with
 - azimuth and elevation, 266–268
 - ecliptic latitude and longitude, 268–270
 - topocentric and geocentric, 260
- Right ascension of the node**, 99
- Right ascension**, 147
- Right-handed coordinate system**, 145
- Rise/set problem** 913–919
 - definition of, 839, 909
- RMS** *See Root Mean Square*
- Root mean square, RMS**, definition of, 743
- Root sum square, RMS**, definition of, 743
- Roots of polynomials**
 - cubic, 1027
 - quadratic, 1027
 - quartic, 1029
- Rotating acceleration**, for three-body problem, 969
- Rotational velocity of the Earth**, 132
- RSS** *See Root Sum Square*
- RSSS, Russian Space Surveillance System**, 250
- RSW**, *See also Coordinate systems* 155
- RTN (radial, transverse, and normal)**, 156
- RTOD, Real Time Orbit Determination**, 795
- Runge, Carl**, 528
- Runge-Kutta methods**, 528
- Russian Dynamic Calibration of the Atmosphere**, 572
- Russian Space Surveillance System (RSSS)**, 250
- S**
- Safety zone**, 920
- sample mean**, 742
- Saros cycle**, 2
- Satellite Laser Ranging (SLR)**, 248, 251, 314, 450
- Satellite populations, distribution of**, 845
- Satellites, types of missions**, 846
- Scalar matrix**, 1019
- Scale height**, 566
- Scaliger, Joseph**, 182
- Scheiner, Christopher**, 8

- Schmidt functions and coefficients**, 587
Schur's identity, 775
Schwarzschild geometry, 898
Schwarzschild radius, 192
Schwarzschild, Karl, 898
Second, SI, *See Système International (SI) second* 190
Sectoral harmonics, 548
Selective Availability (S/A), for GPS signals, 901
Semianalytical techniques *See also Propagation of orbits, semianalytical*
definition of 613
Semi-diameter, 315
Semilatus rectum *See Semiparameter*,
Semimajor axis, 13, 95
equation for, 31
Seminor axis 13
of Earth, 132
Semiparameter, definition of, 17, 96
Semiperimeter, definition of, 470
Semi-synchronous satellites
definition of, 30
Sensitivity matrix, 801
Sensor sites, geometry of, 429
Sensor systems, 249–734
Sensors
calibration, 255, 828
data processing techniques, 733
dedicated, collateral, contributing, 249, 253
limitations of, 254
optical, 433
representative ranges, 256
staring, scanning, 855
tasking, 843
Sequential-batch least-squares techniques, 772–776
Sequential-batch routines, definition of, 773
Series solutions, for eccentric, true, and mean anomalies, 72
Sexagesimal number system, 242
SEZ, *See also Coordinate systems* 153
SGP4, Simplified General Perturbations, 696
Shadow geometry, 301
Shannon-Sampling theorem, 536
Sidereal time 184–189
apparent, 186
converting to solar, 180
definition of, 176, 185
Greenwich Mean, 185
local, 185
mean, 186
transformation for, 223
Signals
active, passive, retransmitted, 243
bi-static, mono-static, 244
transponder, 243
Similarity transformation, definition of, 746
Simple harmonic oscillator equation, 396
Single point conversion, 710
Singular elements, definition of, 643
Singularities, definition of, 760
Site position vector, 135–140, 428
Skew-symmetric matrix, 1019
SLR *See Satellite Laser Ranging*
SME, Specific Mechanical Energy, 25
Smugness, in data filters, 777, 798, 800
SOI, Sphere of Influence, 26
Solar Cycle, 555
Solar flux, 553, 560
Solar incidence angle, 581
Solar maxima and solar minima, 560, 563
Solar pressure, definition of, 580
Solar sail, 581
Solar time 177–184
apparent, 177
converting to sidereal, 180
daylight savings time, 181
definition of, 175
mean, 178
Solar-radiation effects 685–689
equations for, 685–688
shadow analysis, 301–584
Solar-rotation cycle, 554
Solve-for parameters, 760, 829
South American Datum, 143
Soviet Cosmos atmospheric model, 1011–1015
Space Situational Awareness (SSA), 839
Space support, definition of, 251
Space surveillance
definition of, 853
space-based, 853–860
Space Surveillance Network (SSN), 250, 270
Space surveillance, ground-based, 249
Special perturbation techniques *See also Propagation of orbits, numerical*
definition of 519
Specific angular momentum
in two-body equation of motion, 23

- Specific force**, 522
Specific mechanical energy
 derivation of, 25
Speed of light, 242
Sphere of Influence, 945
 for a planet, 26
Spherical harmonics 547–549
 addition theorem of, 542
 sectoral, 548
 tesseral, 548–549
 zonal, 547
SPICE files, 605
Splining techniques 1031–1038
 cubic, 1034
 quintic, 1038
SPOT, 838
Sputnik I, II, 317, 611, 732, 837
Square-root filters, definition of, 820
SSN *See Space Surveillance Network*
Stability, in recursion algorithms, 597
Standard deviation
 definition of, 743
 probability values for, 745
Standard Positioning Service, 901
Standards of Fundamental Astronomy, SO-FA, 236
Star magnitudes, cataloging, 4
State equations, definition of, 371
State representations
 of satellites, 95–112, 735
State space or vector, definition of, 95, 735, 760
State-transition matrix, 82, 778–780
State-transition matrix, error 809
 definition of, 779
Stationary systems, definition of, 811
Stationkeeping maneuvers, 850–852
Statistical Orbit Determination *See Differential correction*
Stokes, Sir George Gabriel, 518
Störmer-Cowell method, 534
Sturm-Liouville theorem, 547
Sub-latitude point, of a satellite
 determining, 169–172
Sum-squared method, 534
Sun synchronous, *See Orbits, Sun synchronous*
Sun, illumination, 310
Sundman transformation, 58, 537
Sundman, Karl F., 58, 967
Sunrise, sunset, determining, 281–285
Sun-spot cycle, 555
Superior planets, 6
Super-synchronous satellites
 definition of, 31
Surveillance from space, definition of, 853
Surveillance, 270
Swerling, Peter, 732
Symmetric matrix, 1019
Synodic coordinate frame, in three-body problem, 968
Synodic frame, definition of, 36
Synodic period 362
Synodic system, definition of, 145
Système International (SI) second, 175, 190
- T**
- Tangential acceleration**, 167
Targeting problem, 498–513
Targets, uncorrelated, 841
Tasking, sensors, 843
Taylor series, 64, 528
Taylor, Brook, 528
TCG, *See also Time, geocentric coordinate*
 194
TDB, *See also Time, barycentric dynamical*
 190
TDRS, Tracking and Data Relay Satellite,
 249, 726
TDT, *See also Time, terrestrial dynamical* 190
TEME system, 231
Temperature, exospheric, 1001, 1002
Ten known integrals, for equations of motion,
 37
Terminator, definition of, 910
Terrestrial latitude, 133
Terrestrial parameters, definition of, 133
Terrestrial time, (TT), 190
Tesseral harmonics, 548
Thales, 2
Third-body effects 679–685
 equations for, 679–682
 on geosynchronous satellites, 683
 partial derivatives for, 577–578
Three-body equations *See Equations of motion*
Tide-gauge measurements, use of, 142
Tides
 ocean, 586
 solid-Earth, 585

- TIMATION Program**, 896
- Time**
- coefficients of, 191
 - conversions for
 - DMS to RAD / Rad to DMS, 197
 - HMS to Rad / Rad to HMS, 198
 - HMS to UT / UT to HMS, 199
 - YMD to Day of Year, 200
 - YMDHMS to Days, 201
 - conversions for, 195, 197–203
 - equation of, 178
 - GPS, 181, 191
 - systems of
 - apparent solar time, 177
 - atomic (TAI), 177, 190
 - barycentric coordinate (TCB), 194
 - barycentric dynamical (TDB), 191
 - coordinate, 193–194
 - coordinated universal (UTC), 180
 - dynamical, 177, 190–193
 - ephemeris (ET), 190
 - geocentric coordinate (TCG), 194
 - mean solar, 178
 - sidereal *See also Sidereal time* 184
 - solar *See also Solar time* 177
 - terrestrial dynamical (TDT), 190
 - universal (UT), 176, 178
 - systems of, 174–196
- Time dependence**, weak, definition of, 595
- Time determination, from GPS**, 902
- Time of flight**, calculating, 123–127
- Time unit (TU)**, definition of, 238
- Time zones**, UTC offsets for, 182
- TIROS 1**, 847
- TOD**, *See true equinox of date* 224
- Tokyo Datum**, 143
- TOPEX**, 248, 250, 665, 726, 741, 799
 - orbit determination for, 822
- Topocentric declination**, 154
- Topocentric right ascension**, 154
- Total Electron Count (TEC)**, 246
- Tracking and Data Relay Satellite** *See TDRS*
- Tracking network**, definition of, 250
- Tracking**, 270
- Trajectory equation**, 29, 50
- Transfer method**, 468
- Transit Satnav System**, 896
- Transits**, 315
- Transponder**, 243
- Triangular points**, finding, 972
- Triangulation**, 252
- Triaxial ellipsoid** 134
- Trigonometric relations**, basics of, 1020–1023
- Trigonometry**
- planar, 1023
 - spherical, 1024–1027
- Trilateration**, 450
- Triton**, orbit of, 102
- Trojan planets**, 12, 972
- Tropical year**, 184
- Troposphere Model**, 247
- Troposphere**, correcting for, 247
- True anomaly**
- definition of, 18, 100
 - hyperbolic limitations on, 54
- True equator of date**, 207, 224
- True longitude at epoch**, 103
- True-equator, Mean-equinox**, 106
- Tsiolkovsky, Konstantin**, 317
- TT**, *See also Time, terrestrial* 190
- Tuning the filter**, 785
- Turning angle**, in hyperbolic orbit, 53
- Twilight**, definitions of, 281
- Two-body equation** *See Equations of motion*
- Two-line element set (TLE)**, 105–108
 - translating, 107
- U**
- Umbra and penumbra**, geometry for, 282, 299
- Uncorrelated targets**, definition of, 841
- Uncorrelated tracks**, definition of, 841
- Universal time**, 178
- Universal variable**
- derivation for Kepler's equation, 58
- Update intervals**, 815
- UT0, UT1, UT2**,
- definitions of, 178–180
- UTC**, *See also Time, coordinated universal* 180
- V**
- Validation**, 599
- Variance, statistical**, definition of, 743
- Variation of parameters (VOP)**, *See also Perturbation techniques* 619–621
 - equinoctial form 628
 - Gaussian 628–637
 - Hamiltonian (canonical) form 637–640
 - Lagrangian 621–628

- Variational equations** 803–811
 - equation for, 804
- Vectors** 1018
 - fundamentals of, 1018
 - position and velocity
 - from orbital elements, 116–120
 - transforming, 166
- Velocity**
 - change, accumulated, 373
 - equation for, 32
 - radial, 314
- Verification**, 599
- Vernal equinox**, 146
- Vertical datums**, 142
- Very Long Baseline Interferometry (VLBI)**, 151, 226, 900
- v -infinity**, definition of, 27
- Visibility**
 - satellite to Sun, satellite to satellite, 305–309
 - to a sensor site, 908–913
- Vis-viva equation**, 26, 27
- VLBI** *See Very Long Baseline Interferometry*
- von Braun, Wernher**, 317
- vonZach, Franz Xavor**, 425
- VOP**, *See also Variation of Parameters* 619

W

- Wait time in orbital rendezvous**, 362
- Weak time dependance**, 595
- Weekday**, equations for, 181
- Weiner, Norbert**, 732
- Wet delay**, 247
- World Geodetic Survey (WGS)**, 601
- Wren, Sir Christopher**, 11
- WWV**, 180, 604
- WWVH**, 181

Y

- Year of confusion**, 4
- Young, John W.**, 318

Z

- Zenith distance**, definition of, 154
- Zonal harmonics** 547
 - effects on orbital elements, 647–654
- Zulu time** *See Coordinated universal time*

Julian Date Values

This table lists some common Julian dates for selected years. Remember that the JD is at 12^h UT. The table also gives the Greenwich Mean 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	GMST (°) 0 ^h UT	JD 12 ^h UT	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
2000*	99.9677947	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	100.5376276	245	5198	5229	5257	5288	5318	5349	5379	5410	5441	5471	5502	5532
2011	100.2989171	245	5563	5594	5622	5653	5683	5714	5744	5775	5806	5836	5867	5897
2012*	100.0602067	245	5928	5959	5988	6019	6049	6080	6110	6141	6172	6202	6233	6263
2013	100.8071437	245	6294	6325	6353	6384	6414	6445	6475	6506	6537	6567	6598	6628
2014	100.5684335	245	6659	6690	6718	6749	6779	6810	6840	6871	6902	6932	6963	6993
2015	100.3297233	245	7024	7055	7083	7114	7144	7175	7205	7236	7267	7297	7328	7358
2016*	100.0910132	245	7389	7420	7449	7480	7510	7541	7571	7602	7633	7663	7694	7724
2017	100.8379505	245	7755	7786	7814	7845	7875	7906	7936	7967	7998	8028	8059	8089
2018	100.5992406	245	8120	8151	8179	8210	8240	8271	8301	8332	8363	8393	8424	8454
2019	100.3605307	245	8485	8516	8544	8575	8605	8636	8666	8697	8728	8758	8789	8819
2020*	100.1218209	245	8850	8881	8910	8941	8971	9002	9032	9063	9094	9124	9155	9185
2030	100.6916694	246	2503	2534	2562	2593	2623	2654	2684	2715	2746	2776	2807	2837
2040*	100.2728782	246	6155	6186	6215	6246	6276	6307	6337	6368	6399	6429	6460	6490
2050	100.8457422	246	9808	9839	9867	9898	9928	9959	9989	0020	0051	0081	0112	0142

*Leap years occur every four years, and in centuries that are evenly divisible by 400.

Constants & Conversions

We can use the four fundamental defining parameters (shaded cells here are from WGS-84/EGM-96) 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. Be aware that different gravitational models will have different conversions. Selected constants and conversions are also provided for general use.

Mean Equatorial Radius			Shape		
(R_{\oplus}) 1.0 ER	6,378.137	km	Mass	$5.973\,332\,8 \times 10^{24}$ kg	
	20,925,646,327,545.93	ft	(b_{\oplus}) Semi-minor axis	6,356,751,600,562.9 km	
	3963.190 591 943	miles	(e_{\oplus}) Earth eccentricity	0.081 819 221 456	
	3443.918 466 523	nm	e_{\oplus}^2	0.006 694 385 000	
Time			(f) Earth flattening		
1.0 TU	13,446,852 063 74	solar min	J_2	0.001 082 626 7	
	806,811 123 824 29	solar s	J_3	-0.000 002 532 7	
	0.009 338 091 710 93	solar day	J_4	-0.000 001 619 6	
	0.009 363 658 561 01	sidereal days	$\bar{C}_{2,0}$	-484,165,371 7 $\times 10^{-6}$	
1.0 Solar Day	1.002 737 909 350 795	sidereal days	Conversions & Constants		
Speed			1.0 ft	0.304 8 m exact	
1.0 ER/TU	25,936,239 235 611 378	ft / solar s	1.0 mile	1.609 344 km exact	
	7,905 365 719 014	km / solar s	1.0 mile	5280 ft exact	
Gravitational Parameter			1.0 nm	1.852 km exact	
(μ) 1.0 ER ³ / TU ²	398,600,441 8	km ³ / (solar s) ²	$\pi/2$	1.570 796 326 794 90	
	14,076,441,757,205,109.60	ft ³ / (solar s) ²	π	3.141 592 653 589 79	
mean value of theoretical gravity at equator	9.797 643 222 2	m / s ²	2π	6.283 185 307 179 59	
	32.144 498 76	ft / s ²	1.0 radian	57,295 779 513 082 30°	
Rotation			1.0 kg	0.453 592 37 lb	
ω_{\oplus}	0.000 072 921 158 553 0	rad / solar sec	1.0 ltr	3,785 411 784 gal	
	0.058 833 599 801 549 19	rad / TU	1.0 ft ² / lb	0.204 816 143 6 m ² / kg	
	0.250 684 477 337 462 15	deg / solar min	Speed of light, c	$2.997\,924\,58 \times 10^8$ m/s	
1.0° / solar s	1.0/0.071 015 111 031 520 1	rad / TU	Planck's constant, h	$6.626\,075\,5 \times 10^{-34}$ Js	
	6,300 388 098 665 74	rad / solar day	Boltzmann constant, k	$1.380\,658 \times 10^{-23}$ J/K	
Solar day	86400 sec				
Sidereal day	86164.090 517 sec		$23^h\,56^m\,4.090\,517^s$		
Tropical year (equinox)	365.242 189 7		$365^d\,5^h\,48^m\,45.2^s$		
Aonomalistic year (perigee)	365.259 636		$365^d\,6^h\,13^m\,52.6^s$		
Sidereal year (fixed star)	365.256 363		$365^d\,6^h\,9^m\,9.8^s$		