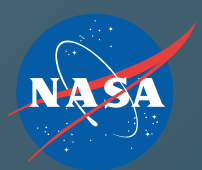
NASA Monographs in Systems and Software Engineering

Harold L. Hallock
Gary Welter
David G. Simpson
Christopher Rouff

ACS Without an Attitude



Springer

NASA Monographs in Systems and Software Engineering

Series editor

Mike Hinchey, Limerick, Ireland

The **NASA Monographs in Systems and Software Engineering** series addresses cutting-edge and groundbreaking research in the fields of systems and software engineering. This includes in-depth descriptions of technologies currently being applied, as well as research areas of likely applicability to future NASA missions. Emphasis is placed on relevance to NASA missions and projects.

More information about this series at http://www.springer.com/series/7055

Harold L. Hallock · Gary Welter
David G. Simpson · Christopher Rouff

ACS Without an Attitude

Springer

Harold L. Hallock
Bowie, MD
USA

Gary Welter
Software Engineering Division
Goddard Space Flight Center
NASA
Greenbelt, MD
USA

David G. Simpson
Software Engineering Division
Goddard Space Flight Center
NASA
Greenbelt, MD
USA

Christopher Rouff
Applied Physics Laboratory
Johns Hopkins University
Laurel, MD
USA

ISSN 1860-0131 ISSN 2197-6597 (electronic)
NASA Monographs in Systems and Software Engineering
ISBN 978-1-4471-7324-3 ISBN 978-1-4471-7325-0 (eBook)
DOI 10.1007/978-1-4471-7325-0

Library of Congress Control Number: 2017934637

© Springer-Verlag London Ltd. 2017
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer-Verlag London Ltd.
The registered company address is: 236 Gray's Inn Road, London WC1X 8HB, United Kingdom

This book is dedicated to the many students/colleagues that attended the lectures, whose active participation greatly improved the lectures and this book.

Preface

If you do a casual search for books that contain the word attitude in their title, you'll find yourself drowning in a sea of over 500 volumes. Of course, most of those books relate more to personal self-improvement than to spacecraft dynamics, but even when the subject is limited to the areas of aerospace and astrodynamics you'll still find a fair number of items from which to choose. So, what distinguishes this text from those many other candidates? This book attempts to de-emphasize the formal mathematical description of spacecraft onboard attitude and orbit applications in favor of a more qualitative, concept-oriented presentation of these topics (whether or not we ultimately achieved that goal is something we'll have to leave to you, dear reader, to decide). As such, it would most likely be described by an attitude control analyst as a (hopefully) amusing light read rather than an essential reference bible. And it certainly would not be the first text an Attitude Control Subsystem (ACS) flight software (FSW) designer would grab if he needed a specification of a Kalman filter algorithm. ACS Without an Attitude is instead intended for programmers and testers new to the field who are seeking a commonsense understanding of the subject matter they're coding and testing in the hope that they'll reduce their risk of introducing or missing the key software bug that causes an abrupt termination in their spacecraft's mission and their careers.

ACS Without an Attitude is organized in four major sections. Section One (Chaps. 1–3) contains the attitude, orbit, and dynamics background material required to understand the downstream spacecraft applications. Section Two (Chaps. 4 and 5) is a survey of the spacecraft sensors and actuators used to measure and control the spacecraft attitude and orbit. Section Three (Chaps. 6–11) examines how sensor data is combined with reference data to measure attitude and orbit, and how desired or commanded attitude parameters are compared with measured attitude parameters to determine what should be done to maintain the current pointing, or modify it to satisfy future needs. Finally, Section Four (Chap. 12) is a survey of mission characteristics and how attitude and orbit geometries are selected to accomplish mission objectives.

The information presented in these sections was originally collected to support an informal set of lectures in 1999 and 2000 instigated by my Branch Chief, Elaine

Shell (Flight Software Branch, NASA Goddard Space Flight Center), who also realized that bullet charts are an ineffective means to document information, hence this book. The following is a list of textbooks and documents I drew on (hopefully not too blatantly) while preparing for my talks, as well as additional references used when writing this book:

1. Spacecraft Attitude Determination and Control, edited by James R. Wertz, Kluwer Academic Publishers (1978).
2. Space Mission Analysis and Design, edited by Wiley J. Larson and James R. Wertz, Microcosm, Inc. and Kluwer Academic Publishers (1992).
3. Reducing Space Mission Cost, edited by James R. Wertz and Wiley J. Larson, Microcosm Press and Kluwer Academic Publishers (1996).
4. Satellite Technology and Its Applications, by P.R.K. Chetty, TAB Professional and Reference Books (1991).
5. An Introduction to the Mathematics and Methods of Astrodynamics, by Richard H. Battin, AIAA Education Series (1987).
6. Modern Inertial Technology Navigation, Guidance, and Control, by Anthony Lawrence, Springer (1998).
7. Modern Control Systems, by Richard C. Dorf and Robert H. Bishop, Addison-Wesley Publishing Company (1995).
8. Feedback Control of Dynamic Systems, by Gene F. Franklin, J. David Powell, and Abbas Emami-Naeini, Addison-Wesley Publishing Company (1991).
9. Modern Control Engineering, by Katsuhiko Ogato, Prentice-Hall (1997).
10. Goddard Trajectory Determination System (GTDS) Mathematical Theory, Revision 1, edited by A.C. Long, J.O. Cappellari, Jr., C.E. Velez, and A.J. Fuchs, NASA/GSFC Flight Dynamics Division Code 550 (1989).
11. Fundamentals of Astrodynamics, by Roger R. Bate, Donald D. Mueller, and Jerry E. White, Dover Publications, Inc. (1971).
12. Classical Mechanics, by Herbert Goldstein, Addison-Wesley Publishing Company (1950).
13. Fundamentals of Spacecraft Attitude Determination and Control, by Landis F. Markley and John L. Crassidis, Springer Science and Business Media (2014). (An excellent book for both subject history and in-depth mathematical analysis.)

In addition, here are some references targeted to specific topics discussed in the later chapters:

1. Hermite Polynomials, Legendre polynomials, and spherical harmonics: Mathematical Methods for Physicists (seventh edition), by G. Arfkin, H. Weber and F. Harris, Academic Press, Inc. (2013), Sect. 18.1.
2. Runge-Kutta integrator: Numerical Methods for Scientists and Engineers by R.W. Hamming, McGraw-Hill (1962).

One of the strengths of my original set of lectures was the participation of several of my NASA/GSFC Guidance, Navigation, and Control (GN&C) colleagues who supplemented (and often, and graciously, corrected) my presentations with material

drawn from their extensive experience here at GSFC. And when my pitches started to drag a bit, they also helped liven things up by recounting some of their many war stories accumulated during their years of applying clean-cut mathematical algorithms to uncooperative real-life spacecraft. My crew of semi-regular experts included

1. Gary Welter
2. Landis Markley
3. Dave Quinn
4. Dave McGlew
5. Bruce Bromberg

As the years have rolled by since the first version of the manuscript, the material in the book has been updated many times, stimulated by new missions and new ACS technologies, as well as new teaching experiences gained repeating the course. Finally, as I reached the point I could no longer bear to edit the material again, Gary Welter, Dave Simpson, and Chris Rouff have ridden to the rescue to co-author with me this final version.

Lou Hallock—2010

Well … life goes on, including delays from other obligations. Lou passed the torch to us, his three amigos, when he retired in 2011, along with encouragement to put our own stamp on the book (sometimes with a "you broke it, you bought it" response to editorial suggestions). We've also corralled a couple of colleagues (Scott Starin and Tim McGee) to provide some review and feedback on the near-final text. (Our thanks to you both.) So, here is the multi-chef result, we think well-flavored—though some of you may find certain sections more "in your face" than "without an attitude". Time to let it go. Enjoy.

Gary Welter, Dave Simpson and Chris Rouff—2016

Contents

Acronyms

ACE	Attitude Control Electronics
ACS	Attitude Control Subsystem
AOS	Acquisition of Signal
AST	Advanced Star Tracker
AU	Astronomical Unit
CCD	Charge-Coupled Device
CDR	Critical Design Review
CGRO	Compton Gamma Ray Observatory
CPU	Central Processing Unit
CSA	Celestial Sensor Assembly
CSS	Coarse Sun Sensor
CXO	Chandra X-Ray Observatory
DCM	Direction Cosin Matrix
DSCOVR	Deep Space Climate Observatory
DoD	Department of Defense
DOY	Day of Year
DSS	Digital Sun Sensor
ECI	Earth-centered Inertial
emf	Electromotive force
EOS	Earth Observing System
ESA	European Space Agency
ESA	Earth Scanner Assembly
FDF	Flight Dynamics Facility
FEEP	Field Emission Electric Propulsion
FES	Fine Error Sensor
FGS	Fine Guidance Sensor
FHST	Fixed-Head Star Tracker
FOG	Fiber Optic Gyro
FOV	Field of View
FSS	Fine Sun Sensor

FSW	Flight Software
GCI	Geocentric Inertial
GEO	Geosynchronous Earth Orbit
GMT	Greenwich Mean Time
GOES	Geostationary Operational Environmental Satellite
GPS	Global Positioning System
GSFC	Goddard Space Flight Center
GTDS	Goddard Trajectory Determination System
HCI	Heliocentric Inertial
HEAO-2	High Energy Astronomy Observatory-2
HGA	High-Gain Antenna
HRG	Hemispheric Resonating Gyro
HST	Hubble Space Telescope
ICOB	Inclined-Center-of-Box
IFOG	Interferometric Fiber Optic Gyro
IFOV	Instantaneous Field of View
IGRF	International Geomagnetic Reference Field
IR	Infrared
IRU	Inertial Reference Unit
ISEE-1	International Sun–Earth Explorer-1
IUE	International Ultraviolet Explorer
JD	Julian Day
JPL	Jet Propulsion Laboratory
JWST	James Webb Space Telescope
KSC	Kennedy Space Center
LISA	Laser Interferometer Space Antenna
LEO	Low Earth Orbit
LMSC	Lockheed Missiles and Space Company
LOS	Loss of Signal
L-R-C	Inductance Resistance Capacitance
LVA	Local Vertical Acquisition
MJD	Modified Julian Day
MT	Magnetic Torquer
MTB	Magnetic Torquer Bar
NASA	National Aeronautics and Space Administration
NCC	Network Control Center
NMR	Nuclear Magnetic Resonance
OBC	Onboard Computer
PD	Proportional–Derivative
PI	Proportional–Integral
PID	Proportional–Integral–Derivative
RFOV	Reduced Field of View
RG	Rate Gyro
RGA	Rate Gyro Assembly
RIG	Rate Integrating Gyro

RLG	Ring Laser Gyro
RPO	Revolution Per Orbit
RWA	Reaction Wheel Assembly
RXTE	Rossi X-ray Timing Explorer
SAMPEX	Solar Anomalous and Magnetospheric Particle Explorer
SDOF	Single-Degree-of-Freedom
SI	Science Instrument
SLP	Solar-Lunar-Planetary
SMM	Solar Maximum Mission
SOHO	Solar and Heliospheric Observatory
ST 5	Space Technology 5
TAI	International Atomic Time
TAM	Three-Axis Magnetometer
TDB	Barycentric Dynamical Time
TDT	Terrestrial Dynamical Time
TDOF	Two-Degree-of-Freedom
TDRS	Tracking Data and Relay Satellite
TDRSS	Tracking Data and Relay Satellite System
TFOV	Total Field of View
TONS	TDRSS Onboard Navigation System
TRIAD	Tri-Axial Attitude Determination
TRMM	Tropical Rainfall Measuring Mission
TT	Terrestrial Time
UT	Universal Time
UTC	Universal Time Coordinate
WMAP	Wilkinson Microwave Anisotropy Probe
YGC	Yaw Gyro Control

Chapter 1
Attitude Conventions and Definitions

Most people who have had even the most casual contact with spacecraft (no, not a Close Encounter of the First Kind) have heard the term "orbit" used and have at least a rough idea of what the word means. By contrast, the first time you hear the phrase "attitude determination" or "attitude control" a whole host of rather exotic images may cross your mind, for example Michael Caine's portrayal of Harry Palmer in the movie *The Ipcress File*. In some respects, these impressions are not totally far-fetched, and there are some similarities between psychiatry and spacecraft attitude control. Although the field of psychoanalysis is possessed of a rich literature documenting its knowledge and a host of psychoanalytical techniques for analyzing a subject have been established, it is difficult even for a talented psychiatrist to determine in the short-term a subject's state of mind. There are many reasons for this problem, but four that strike a resonant chord relative to the aerospace profession are that:

1. many forces can influence a subject's behavior;
2. the relative importance and impact of those forces can be difficult to judge;
3. events from the past must be blended with current events to predict current behavior;
4. there is no direct means to measure the subject's state of mind.

Now replace this image of a patient lying on the couch pouring out his heart to a well-paid psychiatrist with that of a spacecraft in orbit telemetering data to a not-so-well paid Attitude Control Subsystem (ACS) analyst, and let's run through the checklist of difficulties. A spacecraft is subject to a wide variety of perturbations on its pointing (item 1), the effects of which can only be approximated even with the help of complex mathematical models (item 2). One of the standard tools for accurate measurement of attitude errors (i.e., deviations of the measured orientation of the spacecraft in 3-dimensional space from its desired orientation), the Kalman filter, combines a (limited) historical baseline with the most recent measurements, weighting fresher data higher than staler data (item 3). And, unlike the highly controlled conditions of a laboratory or a vacuum chamber, no direct measurements can be

© Springer-Verlag London Ltd. 2017

H.L. Hallock et al., *ACS Without an Attitude*, NASA Monographs
in Systems and Software Engineering, DOI 10.1007/978-1-4471-7325-0_1

made of a spacecraft's orientation in orbit (item 4). One must deduce its orientation from attitude sensor data, i.e., one must see through the "eyes" of the spacecraft rather than seeing the spacecraft state directly. So, as with psychoanalysis, attitude determination and control requires a great deal of analytical rigor in its methodologies along with a somewhat artistic flair in selecting which combination of factors should be evaluated and how those factors should be weighted relative to each other.

Some of you may find Sects. 1.2–1.6 a bit tedious, with its many explanations of different ways to express attitude. One of our reviewers suggested an early plug for Sect. 1.7, which explains why they are all needed. So take heart as you proceed; justification will be forth-coming - we promise. (Thanks Scott.)

1.1 Definition of the Inertial Reference Frame

As is the case for any serious discipline, attitude determination and control possesses its own special vocabulary and conventions that must be absorbed prior to learning the key material that can make its practitioner truly dangerous to his spacecraft. To this end, in this chapter, we'll present the fundamental concepts used to define attitude, along with (to a minimal degree) the various mathematical expressions of attitude and how they relate to each other. To describe the orientation of an object in space, it is first necessary to specify a frame of reference relative to which directions in space and the object's orientation is defined. For spacecraft in orbit about the Earth, it is usually most convenient to define a reference frame whose origin is located at the Earth's center.

The reference frame z-axis is defined to be a unit vector parallel to the Earth's spin axis. By convention, a spin axis direction is defined such that if the fingers of a right hand are curled along the spinning motion, the "thumbs up" thumb direction will be parallel to the positive spin axis direction; this is called a right-handed rotation. By conservation of angular momentum (more on that in Chap. 3), and the fact that the Earth's mass distribution is very nearly spherically symmetric, the direction of the Earth's spin axis is very nearly constant in time …at least over typical human time scales. The reference frame x-axis is associated with another quasi-invariant, the unit vector pointing from the Earth to the Sun at the instant of the vernal equinox. The vernal equinox is the time during the year when the Sun moves along its "orbital" path (as seen from the Earth's point of view) across the Earth's equator from the Southern hemisphere to the Northern hemisphere. This equatorial transit is also called the First Point of Aries, a reference back thousands of years ago to the time when the constellations of the zodiac were originally named and the Sun was in the constellation of Aries at the moment of the vernal equinox. The plane of the Sun's orbit about the Earth (or less parochially, the Earth's orbit about the Sun) is called the *ecliptic*, and is tilted (on average) about 23.44° relative to the Earth's equatorial plane. Finally, taking the cross product of z-axis and x-axis yields the third member of the orthonormal triplet, the y-axis. The reference frame whose axes are defined by this (x, y, z) triplet is called the *Geocentric Inertial (GCI) frame*, where the term

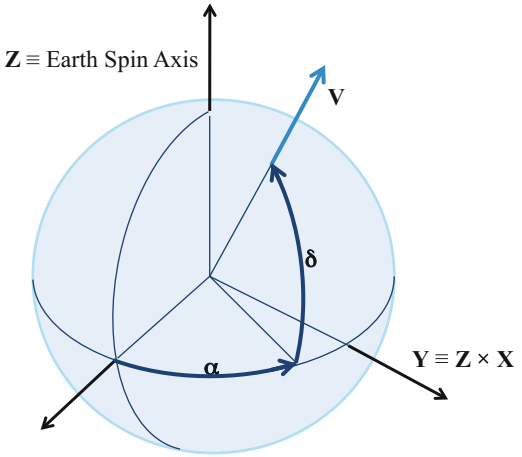


Fig. 1.1 A direction vector **V** relative to the GCI reference frame, rotated from the x-axis based on angles α (along the equator) and δ *above* or *below* the equator

geocentric means "Earth-centered" while the term *inertial* means that the reference frame is fixed relative to the "stationary" stars in the celestial sphere. For folks not into Greek and the Earth goddess Gaia, the reference frame is also called the Earth-centered Inertial (ECI) frame. Figure 1.1 illustrates how a direction vector (**V**) can be defined relative to the GCI frame.

If the Earth's mass distribution were spherically symmetric, the GCI frame described in the preceding paragraph would be truly fixed in space. However, because the Earth is in fact shaped like an oblate spheroid (i.e., it has an equatorial bulge), the gravitational intereaction of the Sun and Moon with the Earth's equatorial bulge causes the Earth's spin axis to precess (i.e., cone) and nutate (i.e., wobble) like a top spinning on the floor. The precession rate (generated by the Sun's torque on the bulge) is 50 arcsec/year, so it takes about 26,000 years for the Earth's spin axis (i.e., the North celestial pole) to complete a single coning period about the North ecliptic pole. (The ecliptic poles define a line perpendicular to the ecliptic plane.[1]) This rotation of the North celestial pole about the North ecliptic pole is also called the *precession of the equinoxes*. The half cone angle is the average angle between the ecliptic pole and the Earth's spin axis, i.e., 23.44°. The amplitude of the wobble (as the spin axis precesses) is about 9.2 arcseconds and arises from the Moon's torque on the Earth's equatorial bulge, with the period of the wobble equal to 18.6 years, which also is the precession period of the Moon's orbital plane due to Solar gravitational influences. Because of these motions, the GCI frame isn't truly "inertial", which is perhaps another reason (in addition to the insistence on viewing the Sun as orbiting

[1] Wertz, Spacecraft Attitude Determination and Control, p. 27.

the Earth rather than the generally accepted Copernican model) for renaming the Geocentric Inertial frame as the Egocentric Inertial frame. In any event, one can't really talk about GCI coordinates without associating those coordinates with a time.

The time identifies the direction of the Earth's spin axis relative to the stars in the Sun's neighborhood, and thereby the orientation of the GCI axes (at that time) relative to the fixed stars. Further, the time identifier comes in different "flavors". If the time is *mean of epoch*, the nutation effects are ignored (i.e., the cone angle is set to 23.44°), but the precession effects are included for the specified epoch time. Nowadays, a standard epoch time is January 1, 2000. So in a Goddard Space Flight Center (GSFC) Flight Dynamics Facility (FDF) ephemeris file, consisting of a series of GCI position and velocity vector pairs spaced at equal time intervals (say, once per minute) over an extended time interval (say, one week), each vector's components are specified relative to the GCI frame defined at the precession angle corresponding to January 1, 2000. Mean of epoch is a popular time choice for star catalogs.

If the time is *true of epoch*, both nutation and precession effects are included, and the same rotation is applied to each vector in the GSFC FDF ephemeris file. If the time is *mean of date*, each vector in the file will include the precession effects associated with the attached time for the vector, so each position and velocity vector pair will be rotated by a slightly different amount. This means, in effect, that the components of each vector pair in the file are specified with respect to a slightly different reference frame. Similarly, with *true of date* times, the nutation and precession effects associated with the attached time for each individual vector pair are included. Unless otherwise requested, GSFC FDF-generated ephemeris files are typically true of date, potentially making them inconsistent with a mean of epoch star catalog. In other words, unless care is taken, the star catalog star vectors may be defined relative to a different GCI frame than the spacecraft position and velocity vectors. This problem, of course, can be avoided by requesting the ephemeris data in mean of epoch format, with the same precession time for both the ephemeris and star catalog data.

To avoid the complexity of having an inertial reference frame (GCI) that's only inertial if you compensate for the fact that it is really moving relative to the celestial objects you're most concerned with (stars in the Sun's neighborhood), one could select a Sun-centered (i.e., Heliocentric) Inertial (HCI) reference frame. However, that choice costs you the convenience gained by choosing as the origin the gravitational body (the Earth) about which most GSFC spacecraft orbit. So for most GSFC missions, it's less work and less confusing to select GCI as the reference relative to which you'll define your spacecraft's orientation. An exception is missions where the spacecraft orbits a Lagrange point (stable or pseudo-stable gravitational equilibrium points relative to two gravitational objects orbiting each other, as will be discussed in Chap. 12). For such cases, for example a Sun-Earth Lagrange point, the "natural" reference frame to pick for many calculations is a Sun-centered rotating (i.e., non-inertial) frame that "freezes" the gravitational objects and the Lagrange point in that frame.

The GCI and HCI reference frames described above are examples of a right-handed reference frames. Mathematically, their orthogonal unit vectors ($\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$) satisfy the relations: $\mathbf{x} \times \mathbf{y} = \mathbf{z}$, $\mathbf{y} \times \mathbf{z} = \mathbf{x}$, and $\mathbf{z} \times \mathbf{x} = \mathbf{y}$. The right-handedness

arises from imagining a right-handed rotation from **x**-to-**y** having a spin axis along **z**, a right-handed rotation from **y**-to-**z** having a spin axis along **x**, and a right-handed rotation from **z**-to-**x** having a spin axis along **y**. One can also have left-handed frames; these satisfy $\mathbf{x} \times \mathbf{y} = -\mathbf{z}$, $\mathbf{y} \times \mathbf{z} = -\mathbf{x}$, and $\mathbf{z} \times \mathbf{x} = -\mathbf{y}$. Right-handed frames are the standard; we will only be using such frames in this book.

1.2 Defining Attitude via Euler Angles (Right Ascension, Declination, and Roll)

Getting back to our favorite (for now) reference frame, the GCI, one of the first things you want to be able to describe, even before talking about the orientation of the spacecraft, is where important celestial objects are located within the frame. Strictly speaking, since most celestial objects (like stars) are so far away, an attitude analyst is primarily interested in the direction of an object in GCI as opposed to its actual position vector (which would include both direction and distance). Just as only two angles, latitude and longitude, are required to specify the location of any point on the Earth's surface, only two angles are needed to specify the location of the center of any celestial object on the celestial sphere (see Fig. 1.1). The GCI counterpart of longitude is called *right ascension*. Right ascension is an azimuthal angle measured within the GCI's *x*-*y* plane and has a range of 0–360°, with a right-handed sense of rotation, and with 0° defined to be the First Point of Aries. Right ascension is computed by projecting onto the *x*-*y* plane the vector directed from the GCI origin to the celestial object center. The angle between the *x*-axis and the vector's projection is the object's right ascension. The GCI counterpart of latitude is called *declination*. Declination is an elevation angle that measures the angle between the *x*-*y* plane and the vector from GCI origin to the celestial object, and has a range of −90° to +90°. For a given right ascension and declination, the equivalent three-dimensional vector is

$$(x, y, z) = [\cos(\alpha)\cos(\delta), \sin(\alpha)\cos(\delta), \sin(\delta)] \tag{1.1}$$

where

(x, y, z) = unitized 3-vector in GCI reference frame
α = right ascension (degrees)
δ = declination (degrees)

The reverse transformation is

$$\begin{aligned} \alpha &= \arctan(y/x) \\ \delta &= \arcsin(z) \end{aligned} \tag{1.2}$$

(The range of the arctan function in Eq. 1.2 is 0–360°, and must be assigned to the correct quadrant according to the signs of x and y. If both x and y are equal to zero (i.e. the declination is $\pm 90°$), then α is undefined.)

Vectors having the z-component zero (i.e., vectors in the x-y plane) have declination zero. The North Celestial Pole is defined to have declination 90°, while the South Celestial Pole has declination $-90°$. Note that if you try to project a vector directed to either pole onto the x-y plane you get a single point at the origin of the GCI frame, not a vector. This condition is referred to mathematically as a singularity, meaning the right ascension is undefined (i.e., cannot be calculated) for objects located at the poles. Alternately, if the declination is $\pm 90°$ the celestial object will be located at the North (South) Pole no matter what value of right ascension is specified.

As illustrated above, the definition of the direction of (ideally) infinitely distant point sources (which, practically speaking, includes the "fixed" stars) in GCI requires only two numbers, right ascension and declination. However, if you want to define the *orientation* of a three-dimensional object in GCI, more information is required.

To help visualize the situation, suppose you wanted to define the orientation of a spinning spacecraft in GCI at a specific instant in time. It is straightforward to specify the spacecraft's spin axis right ascension and declination. The only thing left that we care about is how components mounted on the spacecraft body are oriented with respect to the GCI as they rotate about the spacecraft spin axis at a constant rate (in the absence of perturbations). For example, suppose a camera is mounted on the spacecraft body looking out in a direction perpendicular to the spin axis. You would like to be able to say in what direction in GCI the camera was looking, so if it saw something interesting you could associate what it saw with known celestial objects, or alternately determine the location of a new celestial object so it could be studied in more detail in the future. A convenient way to do this would be to define the angle about the spacecraft spin axis at which the object was observed. (Note, this approach requires that the spin axis be fixed in direction, which is frequently close enough to true.) However, as we saw with right ascension and declination, to use angles to define the orientation of an object in space, we must also define a fiducial (i.e., a reference mark) from which the angle is to be measured.

Continuing our imaginary spacecraft construction exercise, suppose there was a bright object sensor mounted on the spacecraft body looking out in a direction perpendicular to the spin axis. If the spacecraft were in sunlight, as the spacecraft rotated about its spin axis, the bright object sensor would spot the Sun (assuming the spin axis was not pointed directly at, or away from, the Sun) and you could note that time relative to an onboard clock. You could also note relative to an onboard clock when the camera saw its celestial object. Knowing the azimuthal angular separation between the camera and sensor boresights (assume it was measured when the spacecraft was built or was calibrated on-orbit), and knowing the rotation rate of the spacecraft (easily obtained by measuring the time differences between Sun sightings by the bright object sensor, given that the spin axis is fixed in space), you could compute the angle about the spin axis measured from the Sun direction to the direction of the celestial object. This third angle, along with the right ascension and

declination of the spin axis, uniquely "pins" the spacecraft to the celestial sphere at the time of the camera observation.

The spinning spacecraft discussion is a nice practical exercise in visualizing how you can define a specific spacecraft's attitude. What we next want to do is define a more general approach that would be applicable to any spacecraft. Keeping the image of our spinning spacecraft, let's now visualize the spacecraft spin axis (defined to be the spacecraft x-axis) lined up with the GCI x-axis and the camera boresight (defined to be the spacecraft z-axis) lined up with the GCI z-axis. Note that since the spacecraft y-axis is the cross product of the spacecraft z- and x-axes, the spacecraft y-axis will line up with the GCI y-axis as well. Then we can define a unique orientation angle, called the *roll angle* (ϕ), that describes how the spacecraft's camera axis has been rotated from the GCI z-axis in the GCI y-z plane. The roll angle's range of rotation will be 0°–360° (or −180° to +180°), with a right-handed sense of direction.

So now we can imagine putting the spacecraft through a set of three successive "right-handed" rotations about the spacecraft z-, y-, and x-axes with values $(\theta_Z, \theta_Y, \theta_X) = (\alpha, -\delta, \phi)$. The first two get the spacecraft x-axis (the spin axis for our previously imagined spinning spacecraft) pointed at the desired (α, δ) coordinates. (Note that by "right-hand" convention, a rotation that brings spacecraft pointing up into the $+Z$ hemisphere requires a negative rotation about the local Y spacecraft axis.) The final rotation adjusts the roll angle to the desired value (see Fig. 1.2).

Applying these three rotations in that order suffices to reorient the spacecraft from its original attitude (where its axes are aligned with the GCI reference axes) to any general attitude at which the spacecraft could be oriented. Rotating the spacecraft through successive rotations within a fixed reference frame is often called "performing active rotations". The same effect can also be achieved kinematically by conceptually holding the spacecraft fixed and rotating the reference frame (and implicitly the entire universe with it) by the negative of those rotations in the reverse order.

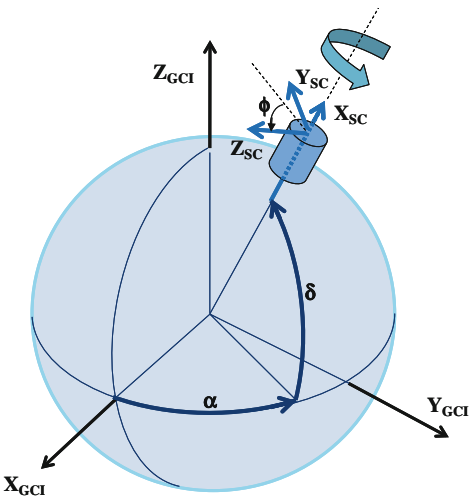


Fig. 1.2 A spacecraft's attitude expressed as a rotation sequence $(\alpha, -\delta, \phi)$

In other words, you get the same result vis-a-vis relative orientation of the spacecraft to the universe by flipping the order and signs of rotation from $(\alpha, -\delta, \phi)$ to $(-\phi, +\delta, -\alpha)$ and imagine them applied to the universe. This approach (rotating the reference frame rather than the object) is often called "performing passive rotations". It is the approach most common in classical dynamics textbooks, although for many the active approach seems easier to visualize.

Whether your rotation convention is active or passive, the set of angles utilized in performing the successive rotations is called an Euler angle sequence. The order of rotation is somewhat arbitrary, as is even the set of axes about which the rotations are performed. Specifically, for a given set of 3 mutually orthogonal rotation axes, there are 12 different rotation combinations that span the space, i.e., that can generate a set of successive rotations that can orient the spacecraft to any desired attitude. The values of the angles, of course, may change depending on the convention selected, so although the definitions of the angle-axis combination is arbitrary, once a convention has been selected all elements of the ground-flight system must follow that convention or pointing interpretation errors will result. The 12 allowed combinations of rotations are 3-1-3, 2-1-2, 1-2-1, 3-2-3, 2-3-2, 1-3-1, 3-1-2, 2-1-3, 1-2-3, 3-2-1, 2-3-1, 1-3-2. In this shorthand notation, "1" stands for the x-axis, "2" stands for the y-axis, and "3" stands for the z-axis.

1.3 Defining Attitude via Euler Angles (Roll, Pitch, and Yaw)

{Roll, Dec, R.A.} is just one of several popular applications of Euler angle formulations to ACS problems. So let's discuss some of the other Euler angle uses. The problem we've examined so far is how do you define the orientation of the spacecraft in the GCI frame, and we've defined a convenient formalism that satisfies that need. Another important problem is how to describe the difference between the way the spacecraft is pointed and the way you'd like to have it pointed. You could do a direct comparison between the measured and desired {Roll, Dec, R.A.}, but the simple differences between the measured and desired angles could give you a rather misleading impression of how large the error is and how much effort is required to correct it. The unreliability of the simple differences grows with the size of the angles (breakdown of first order approximations) and with how close you are to the celestial pole singularities. A better approach to the problem is to define two body reference frames, the desired or commanded body frame and the measured body frame. Of course, there really is only one body frame, and it usually is defined relative to "natural" spacecraft axes, such as the primary moments of inertia, science instrument boresights, etc. But if you substitute for the GCI frame the desired body frame, and substitute for the spacecraft frame the measured body frame, you can similarly define an Euler angle triplet that rotates you from the commanded frame to measured frame. The associated angles are a measure of the angular displacements between the two frames

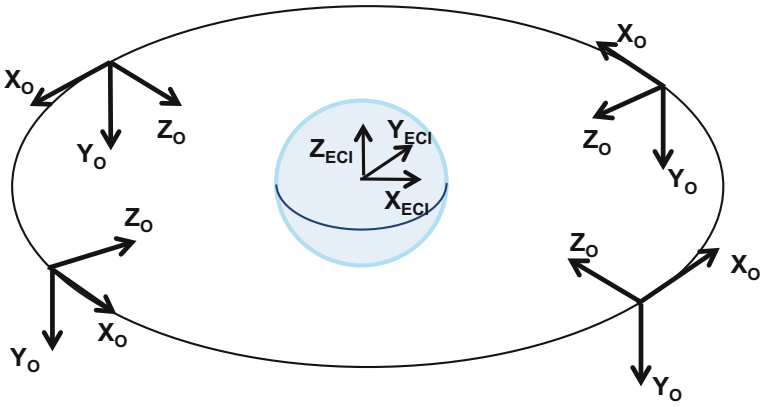


Fig. 1.3 Orbit reference frame; $\mathbf{Z}_0$ towards nadir, $\mathbf{X}_0$ parallel to velocity, $\mathbf{Y}_0 = \mathbf{Z}_0 \times \mathbf{X}_0$

for the selected axes of rotation. The names used for these displacement angles are {Roll, Pitch, and Yaw}, as compared to {Roll, Dec., and R.A.}, where unlike the declination angle's left-handed sense, the pitch angle is measured in a right-handed fashion. The ranges of the angles are the same. The roll and yaw angles have ranges of 0°–360° (or sometimes $-180°$ to $+180°$), and the pitch angle has a range from $-90°$ to $+90°$.

Note that the quantitative description of attitude error is just one use of the {Roll, Pitch, Yaw} formalism. If you substitute for the GCI frame the current commanded body frame and substitute for the spacecraft frame a future commanded body frame, you can obtain an Euler angle triplet that rotates you from where you currently want to be to where you want to go in the future. These Euler angles (and their associated axes) can then be used to control the spacecraft's motion when executing a slew (i.e., a large re-orientation of the spacecraft's pointing).

Another application of {Roll, Pitch, Yaw} is describing an Earth-pointing spacecraft's orientation relative to its orbit about the Earth. In the orbit frame, the z-axis (the yaw axis) is aligned with the direction of the nadir vector, the vector that points from the spacecraft center-of-mass to the Earth center-of-mass (similarly, the zenith vector points from Earth center to spacecraft center). The y-axis (the pitch axis) is aligned with the negative orbit-normal, with the x-axis (the roll axis) forming the last of the orthogonal triplet. Figure 1.3 illustrates how the orbit frame axes are dependent on the spacecraft position in the orbit, rotating once per orbit about the y-axis. For a perfect circular orbit, the x-axis will line up with the direction of the spacecraft orbital velocity vector.

1.4 Defining Attitude via the Direction Cosine Matrix

So an Euler angle description of the spacecraft attitude is an extremely flexible formulation that is easily adaptable to different applications (absolute pointing, relative pointing, and reorientation). It is also ideal for visualizing what's going on with your spacecraft. The problem is how do you convert your visual image to something useful to a flight (or ground) computer that is wonderfully adept at doing exactly what you tell it to do very quickly, but is (at least at this writing) utterly inept at imagining anything at all. That's where the *direction cosine matrix* comes in. The *direction cosine matrix* (DCM) recasts those three Euler angles into a 3×3 matrix that can be used with mathematical rigor to transform any 3-dimensional vector between two reference frames. Specifically, the DCM that transforms vectors from the GCI to the body frame is the attitude matrix. The individual Euler angle rotations themselves can be expressed as 3×3 rotation matrices and, as stated earlier, you can use these elementary single-axis rotations to build up a set of rotations that span the space, although (as discussed at the end of Sect. 1.2) the choice and order of the three rotations is somewhat arbitrary.

OK, so let's construct the attitude direction cosine matrix for a situation in which the spacecraft has been rotated from GCI via an Euler angle sequence (Φ_3, Φ_2, Φ_1), equal to $(\alpha, -\delta, \phi)$, as discussed in Sect. 1.2. First construct the 3×3 matrix that would transform a direction vector from the GCI frame (G) to intermediate frame $F1$ rotated from GCI by angle Φ_3 about axis-3 (i.e., the z-axis):

$$\mathbf{V}_{F1} = \boldsymbol{A}_{F1<G}\mathbf{V}_G = \begin{bmatrix} \cos(\Phi_3) & \sin(\Phi_3) & 0 \\ -\sin(\Phi_3) & \cos(\Phi_3) & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{V}_G$$

where the notation "$F1 < G$" indicates transformation into frame $F1$ from G. Next construct the matrix that would transform a direction vector from frame $F1$ to intermediate frame $F2$ rotated from $F1$ by angle Φ_2 about axis-2 (i.e., the y-axis):

$$\mathbf{V}_{F2} = \boldsymbol{A}_{F2<F1}\mathbf{V}_{F1} = \begin{bmatrix} \cos(\Phi_2) & 0 & -\sin(\Phi_3) \\ 0 & 1 & 0 \\ \sin(\Phi_3) & 0 & \cos(\Phi_2) \end{bmatrix} \mathbf{V}_{F1}$$

Finally, construct the matrix that would transform a direction vector from frame $F2$ to the final spacecraft frame (S) rotated from $F2$ by angle Φ_1 about axis-1 (i.e., the x-axis):

$$\mathbf{V}_S = \boldsymbol{A}_{S<F2}\mathbf{V}_{F2} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\Phi_1) & \sin(\Phi_1) \\ 0 & -\sin(\Phi_1) & \cos(\Phi_1) \end{bmatrix} \mathbf{V}_{F2}$$

For each of the three matrices $\boldsymbol{A}_{F1<G}$, $\boldsymbol{A}_{F2<F1}$, and $\boldsymbol{A}_{S<F2}$, we encourage you to compute a few examples so that you're comfortable that the minus signs are in the

correct positions in the three matrices. Now that you have the individual pieces, put them together as:

$$\mathbf{V}_S = \boldsymbol{A}_{S<F2}\boldsymbol{A}_{F2<F1}\boldsymbol{A}_{F1<G}\mathbf{V}_G = \boldsymbol{A}_{S<G}\mathbf{V}_G \tag{1.3}$$

If you want to go in the opposite direction, i.e., transform a vector from the spacecraft frame to the GCI frame, the appropriate equation is:

$$\mathbf{V}_G = \boldsymbol{A}_{G<S}\mathbf{V}_S = [\boldsymbol{A}_{S<G}]^T\mathbf{V}_S = [\boldsymbol{A}_{F1<G}]^T[\boldsymbol{A}_{F2<F1}]^T[\boldsymbol{A}_{S<F2}]^T\mathbf{V}_S$$

where the superscript T indicates matrix transposition. Direction cosign matrices (a.k.a., rotation matrices) are a special subclass of 3×3 matrices for which the matrix inverse is equal to the matrix transpose.

Equation 1.3 specifies the particular Euler angle convention employed by the ground system for the Hubble Space Telescope (HST) mission, where the order for the roll, declination, and right ascension rotations were arrived at in what passes in aerospace work for a collegial atmosphere. Two analysts (including one of the authors), who were also best friends, argued with and yelled at each other for about an hour until they realized they were saying the same thing (the author was an "activist", the other analyst was a "passivist", but definitely not a "pacifist").[2]

Getting back to our original discussion of attitude formalisms, you might ask how is it that the 9 numbers of the DCM can say the same thing, nothing more and nothing less, as 3 Euler angles. That's 6 more numbers than we started with but no more content (please, no wisecracks about government projects). The reason this can be true is that unlike the independent nature of the 3 Euler angles, which can take on any values within their allowed ranges, the 9 direction cosine numbers are highly correlated with each other. The fact that the DCM is an orthonormal matrix requires that the columns of the matrix, when viewed as three 3-dimensional vectors, be mutually orthogonal (i.e., their dot products with each other yield value zero) and be normalized (i.e., their dot products with themselves yield value unity). That places 6 conditions (3 from orthogonality and 3 from normality) on the 9 numbers, each of which has a range from -1 to 1. Counting independent pieces of information, the 9 numbers of the DCM represent only 3 independent pieces of information, the

[2] While we're on the subject of convention confusions, there's a classic source of misunderstandings between spacecraft builders and spacecraft users that arises entirely from their respective roles. If you're building a spacecraft, you literally see the thing in front of you and have the perspective of a god on the outside looking down on (actually more likely up to, unless it's a very small spacecraft) your creation. So, for example, you see the star trackers attached to the body of the spacecraft and see the sensor's field of view (FOV) from the outside looking in. By contrast, if you're using a spacecraft, it's way above your head in orbit, so you imagine yourself to be a bug sitting inside the spacecraft (or, if you're a Trekkie, you imagine yourself to be Captain Kirk standing on the bridge of the Enterprise) looking out on the heavens through the windows provided by the sensors. This difference in perspective has led to numerous heated arguments between hardware providers and ACS analysts regarding reference frame polarities, and even worse, can produce an incorrect impression of agreement pre-launch, starting a ticking time bomb that will wait until a critical moment in the mission post-launch to rear its ugly head.

same as contained by the set of 3 Euler angles. You can play the same game with the DCM rows, but those 6 conditions can be shown to be redundant with the 6 constraints on the columns. But in return for expressing the same information in a less compact (and therefore less efficient) fashion and abandoning the visual ease of the Euler angle formulation, we've gained a straightforward means to transform vectors quantitatively between reference frames, an essential capability if we are to be able to determine and control the spacecraft pointing accurately in realtime. We also get as a side benefit a very convenient gimmick for computing the inverse pointing. Because a direction cosine matrix is orthonormal, its inverse is the same as its transpose (i.e., you get the inverse by exchanging rows and columns). So if you know the rotation matrix that takes vectors from the GCI frame to the body frame, all you have to do is flip the matrix indices to get the rotation matrix that turns body vectors into GCI vectors.

1.5 Defining Attitude via the Eigenvector and Rotation Angle

Still, utilizing 9 numbers to do the work of 3 seems pretty inefficient, although nowadays flight computer memory and computing power has improved to the point that this inefficiency no longer poses the serious problem it once did to the first generation of onboard computers (OBCs). A useful alternative to these two approaches (direction cosines vs. Euler angles) is obtained by breaking away from the key concept the first two approaches have in common, namely that a general rotation in 3-dimensional space should be built up from a series of 3 rotations about 3 axes. One could equally well view the process of rotating an object in 3-dimensional space as a single, right-handed rotation about a single generalized axis, the generalized axis being any unit vector in the starting reference frame. (That the rotation between any two reference frames can be expressed as a single rotation about a single axis was first demonstrated by Euler.) Because the axis of rotation is an invariant under its rotation, the components of the rotation vector in the starting reference frame will not only have the same numerical values in the ending reference frame, they will also have the same values at any intermediate "snapshot" along the way. For this reason, the rotation axis is also called the eigenvector, where "eigen" means "same" in German. Considering some of the applications we discussed earlier, the rotation angle could be the slew angle if you were maneuvering the spacecraft between attitudes, the angular error between the commanded and measured attitudes, or simply the angular displacement from coincidence with the GCI axes needed to orient the spacecraft at the specified pointing.

So any rotation also can be described by four numbers, one number specifying the angle of rotation and three numbers specifying the axis of rotation. Since the rotation axis is a unit vector, there is a normalization constraint on the rotation axis's three numbers, reducing that information content from three to two. Therefore the

four numbers of the single rotation angle-plus-vector formulation represent just three independent pieces of information, the same as the Euler angle or DCM formulation. As for the Euler angle formulation, one can visualize what the rotation about that one axis will look like, but (as for the Euler angle formulation) you can't easily transform a vector between frames without (for example) first converting from the rotation angle-plus-vector formulation to a DCM. Equation 1.4[3] supplies a messy looking matrix that does that.

$$A = \begin{pmatrix} \cos\phi + e_1^2(1-\cos\phi) & e_1e_2(1-\cos\phi)+e_3\sin\phi & e_1e_3(1-\cos\phi)-e_2\sin\phi \\ e_1e_2(1-\cos\phi)-e_3\sin\phi & \cos\phi+e_2^2(1-\cos\phi) & e_2e_3(1-\cos\phi)+e_1\sin\phi \\ e_1e_3(1-\cos\phi)+e_2\sin\phi & e_2e_3(1-cos\phi)-e_1\sin\phi & \cos\phi+e_3^2(1-\cos\phi) \end{pmatrix} \tag{1.4}$$

where

$\boldsymbol{A}$ = the direction cosine attitude matrix
(e_1, e_2, e_3) = eigenvector
ϕ = slew angle

However, in the case of the single rotation angle-plus-vector formulation, there is a more direct analog to the Euler angle's DCM, the *attitude quaternion*, which is the subject of our next section.

1.6 Defining Attitude via Quarternions

Quaternions seem to generate a wide variety of emotional reactions among people in the aerospace field, such as fear, anger, disbelief, etc. Perhaps the parallels to psychoanalysis aren't so strained after all. Quaternion algebra was first introduced by Sir William Hamilton in the 1840s.[4] For a number of decades, it became the dominant method of expressing much of physics (including kinematics and electromagnetic theory), eventually being replaced by vector analysis starting in the mid 1880s. However, because of its compact form and associated numerical efficiency, the use of quaternion algebra has made a comeback in a number of fields, including attitude control for aerospace systems.

Quaternion algebra may be viewed as an extension of complex algebra, wherein the single imaginary element of the latter ($i = \sqrt{-1}$) is replaced with a triplet of imaginary elements (i, j, k) satisfying Eq. 1.5:

$$i^2 = j^2 = k^2 = -1 \tag{1.5a}$$

$$ij = k;\ jk = i;\ ki = j \tag{1.5b}$$

[3] Wertz, Spacecraft Attitude Determination and Control, p. 413.

[4] W. Hamilton, "On Quaternions, or On a New System of Imaginaries in Algebra", in 18 installments in volumes 25–36 of The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 1844–1850. See http://www.emis.de/classics/Hamilton/OnQuat.pdf.

$$ij = -ji;\ jk = -kj;\ ki = -ik \tag{1.5c}$$

A general quaternion can be expressed as

$$\boldsymbol{q} \equiv (q_1, q_2, q_3, q_4) \equiv iq_1 + jq_2 + kq_3 + q_4 \tag{1.6}$$

where each of q_1, q_2, q_3, and q_4 are real numbers. (The placement of the real term as the fourth component is not a universal convention. Some authors prefer to place it first and call it q_0.) Multiplication of two general quaternions follows from Eq. 1.5, together with normal multiplication of real numbers, as in Eq. 1.7:

$$\begin{aligned} \boldsymbol{pq} &= (p_1, p_2, p_3, p_4)(q_1, q_2, q_3, q_4) && \text{(1.7a)} \\ &= ((p_1q_4 + p_2q_3 - p_3q_2 + p_4q_1), (-p_1q_3 + p_2q_4 + p_3q_1 + p_4q_2), \\ &\quad (p_1q_2 - p_2q_1 + p_3q_4 + p_4q_3), (-p_1q_1 - p_2q_2 - p_3q_3 + p_4q_4)) && \text{(1.7b)} \end{aligned}$$

Ultimately, if you find the reference to hyper-complex numbers off-putting, the basic points are that (a) each quaternion is an ordered quartet of real numbers, and (b) quaternion multiplication can be defined as in Eq. 1.7b.

A quaternion conjugate ($\boldsymbol{q}^*$) is defined analogously to a complex conjugate, i.e., $\boldsymbol{q}^* = (q_1, q_2, q_3, q_4)^* = (-q_1, -q_2, -q_3, q_4)$. Note that $\boldsymbol{q}^*\boldsymbol{q} = \boldsymbol{q}\boldsymbol{q}^* = (\Sigma_i q_i^2)(0, 0, 0, 1)$, so $(\Sigma_i q_i^2)^{-1}\boldsymbol{q}^*$ is the multiplicative inverse of $\boldsymbol{q}$.

Let's turn now to expressing an attitude (or a rotation generally) as a quaternion. Given a rotation expressed as an angle (ϕ) and eigenvector ($\mathbf{e}$) set, the corresponding attitude quaternion is constructed via Eq. 1.8.

$$\boldsymbol{q} = (q_1, q_2, q_3, q_4) = (e_1 \sin(\phi/2), e_2 \sin(\phi/2), e_3 \sin(\phi/2), \cos(\phi/2)) \tag{1.8}$$

Although an attitude quaternion's information presentation is more abstract than the rotation angle-plus-vector format, the information contained is the same and can be extracted from the attitude quaternion fairly easily. The sum of the squares of the four components of the attitude quaternion is equal to the expression $[(e_1^2 + e_2^2 + e_3^2)\sin^2(\phi/2) + \cos^2(\phi/2)]$. Using the fact that the eigenvector is a unit vector and the trigonometric identity $(\sin^2(\phi/2) + \cos^2(\phi/2)) = 1$, it's clear that like the eigenvector the attitude quaternion is unitized as well. Since attitude quaternions are normalized, an attitude quaternion packages only three independent numbers, just as the case for the other attitude formats. Note also that the reverse rotation is represented by the conjugate quaternion, $\boldsymbol{q}^*$, i.e., the same rotation angle about an eigenvector pointed in the opposite direction. Note also that the $-\boldsymbol{q}$ represents a rotation about $-\mathbf{e}$ by an angle $2\pi - \phi$, which in normal 3-D space gets you to the same orientation as $\boldsymbol{q}$. For attitude representation purposes, we therefore have $\boldsymbol{q} = -\boldsymbol{q}$ (though a computer program may take issue with that if you're not careful).

A DCM, $\boldsymbol{A}$, can be formed from the corresponding quaternion components without backtracking through the $(\phi, \mathbf{e})$ representation and Eq. 1.4 as shown in Eq. 1.9.[5]

$$\boldsymbol{A} = \begin{pmatrix} q_1^2 - q_2^2 - q_3^2 + q_4^2 & 2(q_1q_2 + q_3q_4) & 2(q_1q_3 - q_2q_4) \\ 2(q_1q_2 - q_3q_4) & -q_1^2 + q_2^2 - q_3^2 + q_4^2 & 2(q_2q_3 + q_1q_4) \\ 2(q_1q_3 + q_2q_4) & 2(q_2q_3 - q_1q_4) & -q_1^2 - q_2^2 + q_3^2 + q_4^2 \end{pmatrix} \tag{1.9}$$

If you need to rotate a 3-vector $\mathbf{v}$ from one frame to another given the rotation in quaternion format, you can use Eq. 1.9 to construct $\boldsymbol{A}$ and then use $\boldsymbol{A}$ to rotate the vector using $\mathbf{v}' = \boldsymbol{A}\mathbf{v}$. Alternatively, you can perform the entire operation with quaternion algebra as shown in Eq. 1.10.

$$\boldsymbol{V'} = \boldsymbol{q^*Vq} \tag{1.10}$$

where

$\boldsymbol{q}$ = the attitude quaternion equivalent of $\boldsymbol{A}$
$\boldsymbol{V} = (\mathbf{v}, 0) = (v_1, v_2, v_3, 0)$
$\boldsymbol{V'} = (\mathbf{v}', 0)$

Quaternion multiplication is associative, so $\boldsymbol{q^*Vq} = \boldsymbol{q^*}(\boldsymbol{Vq}) = (\boldsymbol{q^*V})\boldsymbol{q}$. Equation 1.10 is sufficiently different from its matrix equation counterpart that we recommend that you experiment with a few examples to verify that $\boldsymbol{V'} = \boldsymbol{q^*Vq}$ does indeed give the same results as $\mathbf{v}' = \boldsymbol{A}\mathbf{v}$. (You may need to refresh your memory on trig half-angle identities along the way.)

Given two "active" rotations: $\phi_a, \boldsymbol{e}_a$ followed by $\phi_b, \boldsymbol{e}_b$, with corresponding quaternions $\mathbf{q}_a$ and $\mathbf{q}_b$ as computed via Eq. 1.8, the combined rotation quaternion is given by the product $\mathbf{q}_a\mathbf{q}_b$. One can see this from Eq. 1.10 by letting $\mathbf{V}_0$ be the quaternion-formatted vector in the initial frame, $\mathbf{V}_a$ be its representation following rotation 'a', and $\mathbf{V}_{ab}$ be its representation following the combined rotation 'a' followed by 'b':

$$\mathbf{V}_{ab} = \mathbf{q}_b^*\mathbf{V}_a\mathbf{q}_b = \mathbf{q}_b^*(\mathbf{q}_a^*\mathbf{V}_0\mathbf{q}_a)\mathbf{q}_b = (\mathbf{q}_b^*\mathbf{q}_a^*)\mathbf{V}_0(\mathbf{q}_a\mathbf{q}_b) = (\mathbf{q}_a\mathbf{q}_b)^*\mathbf{V}_0(\mathbf{q}_a\mathbf{q}_b)$$

In contrast, the corresponding DCM multiplication is in the reverse order

$$\boldsymbol{v}_{ab} = \mathbf{A}_b\boldsymbol{v}_a = \mathbf{A}_b(\mathbf{A}_a\boldsymbol{v}_0) = (\mathbf{A}_b\mathbf{A}_a)\boldsymbol{v}_0$$

There's nothing magic about the quaternion vs. DCM multiplication order reversal here. It is just a matter of the convention(s) one chooses for representing the algebraic operations. One could, for example, have chosen to write vectors as 1×3 row matrices and use the transpose of the preceding equation for vector transformation between frames:

$$[\boldsymbol{v}_{ab}]^T = [(\mathbf{A}_b\mathbf{A}_a)\boldsymbol{v}_0]^T = [\boldsymbol{v}_0]^T[(\mathbf{A}_b\mathbf{A}_a)]^T = [\boldsymbol{v}_0]^T[(\mathbf{A}_a)^T(\mathbf{A}_b)^T]$$

[5] Wertz, Spacecraft Attitude Determination and Control, p. 414.

which has the matrix multiplication, $(\mathbf{A}_a)^T(\mathbf{A}_b)^T$, in the same order as the corresponding quaternion multiplication $\mathbf{q}_a\mathbf{q}_b$. Alternatively, one could follow a different convention with quaternion multiplication. Many authors[6] prefer a convention that, although not expressed as such, essentially redefines Hamilton's hyper-complex commutation relations (Eq. 1.5b above) into

$$ij = -k, kj = -i, ki = -j$$

The net effect is that the product of two quaternions using this alternative convention is equal to the reverse product using Hamilton's convention, with the result that $(\mathbf{A}_b\mathbf{A}_a)$ maps to $(\mathbf{q}_b\mathbf{q}_a)$ when the alternative convention is used. So, when reading about using quaternions, give a care to which convention pertains.

There is yet another caveat we must give regarding quaternion multiplication ordering. Equation 1.7 is linear in the elements of $\boldsymbol{p}$ and $\boldsymbol{q}$ individually. This means that the product can be recast as a matrix equation in various forms, e.g.,

$$\boldsymbol{q} \rightarrow [q_1\ q_2\ q_3\ q_4]^T$$

$$\boldsymbol{pq} \rightarrow \boldsymbol{f}(\boldsymbol{p}, \boldsymbol{q}) = \begin{pmatrix} p_4 & -p_3 & p_2 & p_1 \\ p_3 & p_4 & -p_1 & p_2 \\ -p_2 & p_1 & p_4 & p_3 \\ -p_1 & -p_2 & -p_3 & p_4 \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \end{pmatrix} \tag{1.11}$$

or

$$\boldsymbol{p} \rightarrow [p_1\ p_2\ p_3\ p_4]^T$$

$$\boldsymbol{pq} \rightarrow \boldsymbol{g}(\boldsymbol{q}, \boldsymbol{p}) = \begin{pmatrix} q_4 & q_3 & -q_2 & q_1 \\ -q_3 & q_4 & q_1 & q_2 \\ q_2 & -q_1 & q_4 & q_3 \\ -q_1 & -q_2 & -q_3 & q_4 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix} \tag{1.12}$$

or two similar variations that cast the matrix multiplication in a $[1 \times 4]$ $[4 \times 4]$ form. The underlying meaning remains the same, so all are equally valid, but the apparent ordering of $\boldsymbol{p}$ and $\boldsymbol{q}$ changes in the representation. As it turns out, representing the $\boldsymbol{pq}$ product in the $\boldsymbol{g(q, p)}$ matrix multiplication form is fairly common,[7] which may confusingly simplify to a notation "$\boldsymbol{qp}$". All we can do is (again) caution you to be careful about authors' conventions regarding quaternion multiplication ordering, as they may differ. For that matter, you also need to be careful about the conventions used in quaternion multiplication subroutines when using utility libraries. The meaning of "first" and "second" quaternion in a given coder's documentation will depend on the convention used by the author from which the coder took the equations. This means

[6] See e.g., Markley and Crassidis, Fundamentals of Attitude Determination and Control, Sect. 2.7, and references therein.

[7] Wertz, Spacecraft Attitude Determination and Control, Appendix D.

that before you call a quaternion multiplication utility subroutine, you better carefully read the internal documentation (Stop laughing. Most coder comments nowadays are both readable and consistent with the code they're documenting), otherwise you may end up in software Unsafemode as opposed to spacecraft Safemode. If it turns out that the documentation doesn't make the convention clear, it is always prudent to verify empirically that the output result corresponds with what you expect for the input that you're providing.

So, given that attitude quaternions are more abstract than visual formulations like Euler angle or rotation angle-plus-vector format, given that you can inadvertently multiply your attitude quaternions in the wrong order, and given that you can even mess up the order of the components (the real component being "zeroth" vs. "fourth"), why are attitude quaternions used as the prime attitude formulation aboard most three-axis-stabilized spacecraft? Part of the answer is that flight computers can't visualize anything and find attitude quaternions easy to manipulate mathematically. Specifically, there are no trig functions, no singularities, and only 4 numbers to deal with instead of 9. The other part of the answer is that attitude quaternions have many nice compensating features that even computationally challenged human beings like. First, because attitude quaternions have an association with the eigen-vector-plus-slew-angle format, there are fewer consistency issues than with Euler angle format, where you have to build up your attitude from a non-universal set of 3 rotations. Second, taking an attitude quaternion's inverse is very easy, just negate the vector (hyper-complex) components. The effect mathematically is equivalent to reversing the axis of rotation or negating the angle of rotation (recall Eq. 1.8). Third, it's easy to compute attitude differences between two attitude quaternions. Just take the inverse of one of the attitude quaternions and perform quaternion multiplication with the other one, making sure to get the order right (see Eq. 1.13).

$$\Delta \boldsymbol{q} = \boldsymbol{q}_i^* \boldsymbol{q}_f \tag{1.13a}$$

$$= \boldsymbol{g}(\boldsymbol{q}_f, \boldsymbol{q}_i^*) \tag{1.13b}$$

$$= \begin{pmatrix} q_{f4} & q_{f3} & -q_{f2} & q_{f1} \\ -q_{f3} & q_{f4} & q_{f1} & q_{f2} \\ q_{f2} & -q_{f1} & q_{f4} & q_{f3} \\ -q_{f1} & -q_{f2} & -q_{f3} & q_{f4} \end{pmatrix} \begin{pmatrix} -q_{i1} \\ -q_{i2} \\ -q_{i3} \\ q_{i4} \end{pmatrix} \tag{1.13c}$$

where

$\Delta \boldsymbol{q}$ = the delta attitude quaternion
$\boldsymbol{q}_i$ = starting attitude quaternion = $(q_{i1}, q_{i2}, q_{i3}, q_{i4})$
$\boldsymbol{q}_f$ = final attitude quaternion = $(q_{f1}, q_{f2}, q_{f3}, q_{f4})$

You can then extract the angular difference from the scalar component of the attitude quaternion, and the directional difference (i.e., the eigenvector rotating the first attitude into the second) from the vector components. As discussed previously, the angular difference could be the slew angle if the spacecraft is being slewed to a new target attitude, or an error angle if the two attitude quaternions are the commanded

and measured attitude quaternions. If the angular error is small, first order approximations are valid, and one can replace the sine of an angle with the value of the angle in radians. So for small angular differences, Eqs. 1.1–1.8 says that the vector attitude quaternion components reduce to half the angular difference times the eigenvector. In other words, the first component is half the roll error, the second component is half the pitch error, and the third component is half the yaw error, all specified in radians. Then, using a rule of thumb that an arcsecond is about 5×10^{-6} radians, it's fairly easy to "eyeball" an error quaternion and estimate your pointing error in more human friendly terms. Note that because first order approximations are often utilized in onboard equations, and because the small angle approximation between the sine of the angle and the angle itself only works if the angle's units are radians, it's a good programming practice always to use radians (not degrees, arcseconds, etc.) for all internal computations.

Before allowing you to escape this section, we have one last side comment on quaternions. With the DF-224 flight computer originally flown on HST (it was replaced with a more modern computer on a Servicing Mission, and that computer by now is also obsolete), unitized quaternions having components many orders of magnitude different from each other could introduce scaling-related computational errors that might not be ignorable relative to HST's high pointing accuracy requirements (specifically several milliarcseconds). This typically would occur on HST when dealing with unitized quaternions near identity, i.e. (0, 0, 0, 1), where precision loss could occur when dealing with 4th components that looked like 0.9999 . . . To deal with this problem, the HST Pointing Control Subsystem (PCS) (every other spacecraft calls it ACS) flight software (FSW) utilized what the programmers called "pseudo-quaternions", where they subtracted the identity quaternion (0, 0, 0, 1) from the actual quaternion before performing any downstream application processing. They also introduced a separate pseudo-quaternion multiplication convention defining the rules for operating on pseudo-quaternions. As pseudo-quaternions have not been employed on any other missions, there's no value in burdening the reader with the details for manipulating pseudo-quaternions. However, any reader with an archaeological bent may check out Lockheed Missiles and Space Company's (LMSC's), massive documentation of their original FSW, which also includes a quote from the movie The Vikings placing the "Curse of Odin" on anyone who tampers with a particularly intricate portion of their attitude maneuver control logic.

1.7 Attitude Format Applications

At this point, a legitimate question a reader might ask is "Why do I need all these attitude formats"? Why isn't one enough? Why can't I take a cafeteria approach and learn the ones I like and ignore the rest? The answer is not only do each of the formats described have real aerospace applications, but what's more, each of these applications may even be required to support a single science observation. To illustrate (and convince the skeptical reader of) this point, let's examine (at a

high level) the process of planning, scheduling, and executing a hypothetical science observation for a celestial-pointing spacecraft, like the HST.

The first step, using Euler angle format, is for the astronomer who has authored the observation to specify the direction of the science target to be observed by a given science instrument (SI), and to specify where in the SI's FOV the target should be positioned. The science target's coordinates are defined as right ascension and declination, while the science target's offset from the SI boresight is defined as pitch and yaw angles. In addition, the astronomer indicates whether the observation requires a special orientation about the SI boresight, or whether the spacecraft can be oriented so as to optimize solar power generation. This is done by specifying a deviation from optimal roll, which later in the process is associated with an actual roll angle to go with the already specified target right ascension and declination.

The second step is for the ground system to figure out what spacecraft pointing is needed to support the science observation. Using the SI alignment matrix (a DCM specifying how the SI is oriented relative to the spacecraft body reference frame), the ground software determines what the spacecraft right ascension and declination need to be in order to place the science target at the desired location in the SI FOV. Then, using the deviation from optimal roll and solar ephemeris, the ground software determines the absolute roll angle for the spacecraft that will produce the desired relative Sun geometry. At this point, the three pieces of information needed to specify the spacecraft's orientation in inertial space uniquely are known.

However, the ground system has other responsibilities beyond defining the spacecraft pointing for each science observation. For example, telemetry downlink must be enabled by orienting the spacecraft's antenna properly to establish communications with a Tracking Data and Relay Satellite (TDRS) or with a ground station. To determine the appropriate antenna gimbal angles to command, the ground system must first determine the line-of-sight (LOS) vector from the spacecraft to a TDRS (or a ground station antenna). This requires that ephemeris information in the GCI reference frame be transformed into (first) the spacecraft body frame and then into the antenna frame. The most straightforward way to perform this frame change is to transform the spacecraft attitude from Euler angle format to DCM format, and then use the attitude matrix to rotate the GCI frame LOS vector into the body frame.

Potentially, the ground system could uplink the attitude matrix to the spacecraft in order to direct the spacecraft to its new target attitude, but in practice this is not done for two reasons. First, uplinking the nine numbers of a DCM is a pretty inefficient way to command a spacecraft orientation whose specification requires only three numbers. Second, the format the spacecraft really "wants" for its attitude bookkeeping is a different format, namely attitude quaternion format, which is ideal for figuring out what sort of reorientation maneuver must be executed in order to arrive at its new commanded attitude. So in practice, the ground system either uplinks the attitude quaternion directly (possibly with adjustments for the effects of velocity aberration, which will be discussed in Chap. 6), or uplinks the Euler angles the FSW would need to calculate the commanded attitude quaternion on its own. The first approach was used for the HST mission; the second approach was used for the Rossi X-ray Timing Explorer (RXTE) mission.

Either way, the FSW then will compute the delta attitude quaternion, which defines the reorientation needed to slew the spacecraft from its current attitude quaternion to the new target attitude quaternion. The components of the delta attitude quaternion provide a convenient means for extracting the eigenvector and slew angle associated with the desired slew, from which the FSW can determine the appropriate actuator commands needed to realize the attitude change. Finally, to command appropriate orientations of solar arrays and antennas, the FSW usually transforms the commanded attitude quaternion to DCM format to support calculation of the Sun vector and communication LOS vector in the body frame.

Chapter 2
General Orbit Background

Unlike our first topic, attitude, *orbit* is a word for which most people probably could conjure up a reasonable visual image. They might envision the Earth and the other planets circling the Sun, or the Moon or a spacecraft circling the Earth. But beyond that, their perceptions can get kind of fuzzy, and sometimes pretty inaccurate. For example, our experiences on the surface of the Earth lead us to assume that if you want to catch up with something you should move faster. For example, if you are trailing another car and want to overtake it, your natural reaction is to lean on the accelerator to increase your speed. But when orbiting the Earth, this would be exactly the wrong strategy. As Earth-orbiting objects move from perigee (their point of closest approach to the Earth) to apogee (their point of furthest separation from the Earth), they exchange kinetic energy for potential energy, and as a result fly slower at higher altitudes than at lower altitudes. This counterintuitive aspect of orbital motion is also relevant to astronauts performing rendezvous operations with other spacecraft. In orbit, the effect of the equivalent action (relative to the car chase scenario), firing a thruster in the direction of your motion, will be to raise your orbital altitude, causing you to move slower, and thereby losing ground with your rendezvous partner. What you need to do instead is to fire your thruster opposite to your direction of motion, lowering your orbital altitude, causing you to move faster, and resulting in an eventual overtaking of your quarry. So even though we think of the classical dynamics engendered by Newton's law of gravitation over 300 hundred years ago as being sensible and down to Earth (sorry), the motion of objects in the presence of gravitational fields is not always what you'd expect based on your interactions with the Earth's gravitational field in your daily lives.

2.1 Historical Perspective

For everyday life spent at or near the surface of the Earth, dropped objects appear to fall at a constant acceleration of about $9.8\,\mathrm{m/s^2}$ (or $32\,\mathrm{feet/s^2}$ if you prefer English units). This value is accurate to 1% over a spectrum of locations ranging from the

© Springer-Verlag London Ltd. 2017

H.L. Hallock et al., *ACS Without an Attitude*, NASA Monographs
in Systems and Software Engineering, DOI 10.1007/978-1-4471-7325-0_2

Earth's poles, to its equatorial regions, to points located up to 20 miles above the equator. So Galileo's 400 year old observation (that, in the absence of air resistance, two dropped objects of different mass experience the same acceleration independent of altitude) has been something you could count on pretty reliably (Aristotelian ivory tower philosophizing to the contrary) for most of human existence. However, when people started looking up at the heavens in a serious, quantitative fashion, a handful of very smart people began to realize that if things worked upstairs (in the heavens) the way things appeared to work downstairs (on the Earth), Galileo's law of falling bodies could not be a reliable guide to explaining how celestial bodies interacted with each other, even though they had an instinctive feeling that the force that caused the Sun and planets to follow predictable paths in the sky relative to the Earth was the same force that caused a dropped hammer to fall and hit your big toe. The force that causes your hammer to hit your thumbnail instead of a steel nail is a different one (see poltergeists or gremlins).

Actually, the idea that things should work upstairs the same way they work downstairs would have been considered downright silly, possibly even depraved, 1,600 years ago. Even someone like Aristotle, whose work in the biological sciences was based soundly on observation and measurement, took at times what we would view today as almost a theologically based view of the physical sciences. To Aristotle, and nearly everybody else from his age (Aristarchus of Samos, the birthplace of Pythagoras, being an exception), the real world here on the Earth's surface was (probably with good reason) seen as messy, complicated, and inelegant. By contrast, the heavens were perfect, simple, and elegant. Since the circle was for the Greeks an expression of perfection, it followed that objects that lived in the perfect heavens must follow paths composed of perfect circles.

The Greeks' obstinate commitment to heavenly bodies only being allowed to move in immutable circles (because that's the way things are intended to be as opposed to the way we find them to be, also a not unfamiliar modern day phenomenon), coupled with an unshakable belief that the Earth was the center of the Universe, shackled them to a cumbersome system that required that the known planets of the time follow paths composed of circles within circles. The coming of the Renaissance began a revolution (yes, that pun was intended) from a conceptual standpoint that promoted the Sun as the central body (supported by Copernicus' measurements of planetary motions made about 500 years ago), displacing Aristotle and Company's Earth-centered model. Galileo further raised the discussion (about 400 years ago) from metaphysics to actual physics when he aimed his custom-built telescope at Jupiter and observed its moons in orbit, suggesting that just as that planet's moons orbited the planet, the planets themselves could be orbiting the Sun. However, the key to burying the Earth-centered model was provided by an Odd Couple that together embodied the fundamentals on which modern science is based, namely careful quantitative measurements, elegant mathematics and models, and uncompromising intellectual honesty.

The Odd Couple in question, Tycho Brahe and Johannes Kepler, made their contributions around the same time as Galileo, but worked in a politically safer location, Denmark. Brahe was an expert at making positional measurements (providing major

improvements over those produced by Copernicus) and accurately recording them, but lacked theoretical vision and was awful at mathematics. By contrast, Kepler's optical vision was terrible, but he could see meaning in Brahe's mountain of data and had the extraordinary mathematical skills required to organize the data within a working model. After several years of empirical efforts with Brahe's Mars positions, Kepler concluded that an ellipse was the simplest curve to which Brahe's data could be fit to the level of accuracy demanded by Brahe's highly precise observations. Kepler summarized the impact of his discovery in his Three Laws,[1] namely,

1. Each planet follows an elliptical orbit with the Sun at one focus of the ellipse.
2. A line connecting the Sun and a planet sweeps out equal areas in equal times.
3. The square of a planet's orbital period is proportional to the cube of the major axis of the ellipse.

Kepler's laws and associated model finally ended the science controversy (though not the religious or political ones) regarding who orbited whom, and made quantitative confirmed predictions regarding the relative positions and rates of the objects. On the downside, Kepler's work in no way explained why any of these neat facts should be true. That's where Isaac Newton enters the story. When plague closed down Cambridge University for two years (1665–1667), Newton spent his free time creating the basis for differential calculus and developing his laws of motion and gravitation (not bad for a vacation). But because he had some initial problems explaining details of the Moon's motion, Newton did not publish his *Principia*[2] (describing Newton's Laws of Motion and Gravitation) until 1687. In the *Principia*, Newton's Law of Gravitation (see Eq. 2.1) appears in print for the first time. As applied to the gravitational interaction between two idealized point masses, it states that

1. The magnitude of the force experienced by one of the objects is proportional to the product of the two masses and inversely proportional to the square of their separation distance.
2. The direction of the force on one of the objects is towards the other object and is along the line connecting the centers of the two objects.

$$\mathbf{F}_{12} = \frac{Gm_1m_2}{|\mathbf{R}_{12}|^3}\mathbf{R}_{12} \tag{2.1}$$

where

$\mathbf{F}_{12}$ = the force exerted on object 1 by object 2 (newtons)
G = gravitational constant = $6.67428 \times 10^{-11}\,\mathrm{m^3kg^{-1}s^{-2}}$
m_1 = mass of object 1 (kg)
m_2 = mass of object 2 (kg)

[1] See, e.g., http://www-spof.gsfc.nasa.gov/stargaze/Kep3laws.htm.

[2] I. Newton, Philosophiae Naturalis Principia Mathematica (Mathematical Principles of Natural Philosophy), 1687; cf. the translation by I. Cohen and A. Whitman, University of California Press, Berkeley, Los Angeles, London, 1999.

$\mathbf{R}_{12}$ = position vector from object 1 to object 2 (m)
$|\mathbf{R}_{12}|$ = magnitude of position vector from object 1 to object 2 (m)

Newton further showed that the external gravitational force exerted by an object with a spherically symmetric mass distribution is equivalent to that which would be exerted if all of the mass were concentrated at the object's center. (The demonstration is actually fairly straightforward, but we'll leave it to bored readers who enjoy integrations involving spherical coordinates.) In real world applications, celestial objects are never such perfect spheres, and the problem to be solved usually involves more than two objects (often with non-gravitational perturbing forces thrown into the mix as well). But Newton's Law of Gravitation still is the key to solving the classic problem of orbital motion. To deal with non-spherical masses, one can integrate over the object's mass distribution to obtain the influence of the entire celestial object. Similarly, if more than two objects are interacting gravitationally with each other, just sum the effects of all the other bodies on the one you're looking at. However, because the other bodies will interact with each other, no closed-form solution exists for the N-body problem once N is 3 or larger, hence Newton's problems with the Moon, primarily influenced by the Earth and Sun. For such problems, approximations must be used and the analytical approach used is to integrate the equations of motion–an easy thing to do in our computer age, but a far more difficult problem in Newton's time. Such approximation methods are particularly appropriate once the problem is further complicated by including non-gravitational perturbing influences such as atmospheric drag, solar radiation pressure, and propulsive forces along with gravity when determining the orbital motion of a spacecraft (as opposed to the motion of planets or moons).

2.2 Orbital Shapes

For the time being, let's not worry about the details regarding how you predict an object's orbit; that will be covered in a later chapter (or as we often say in bullet slide presentations, "Great question! Dave will talk about that later today."). Instead, let's consider the best way to describe an object's orbit assuming you have all the information required to define it. First, you have to know where the object is at a given instant in time, say right now. An easy way to define an object's position is by specifying the three-dimensional vector (magnitude and direction) that connects the object to the origin of a three-dimensional coordinate frame. But when discussing an object's orbit, you need to know both where the object is right now as well as where the object is going to be in the future (or where it came from in the past if your interests lean more towards history than current events). In general, if you want to predict an object's future position (relative to its current position) you at least need to know in what direction it's heading and how fast it's going (i.e., its velocity vector relative to its current position). So if you know an object's position and velocity vector relative to the coordinate frame's origin at a given instant in time, you can

guess at what its position will be at some later (or earlier) point in time. If the new time is very close to the original time, you can compute the new position pretty accurately just by multiplying the velocity vector by the time difference and adding it to the starting position (a crude integration over time).

The problem with this solution is that if the object is following a wildly varying path, the error in the position prediction will rise quickly as the time difference increases, or to put it another way, the original information will go stale very rapidly. However, when examining the orbital behavior of two (and only two) spherically symmetric masses acting under the influence of an inverse square law force like gravity, the orbital motion of one of the objects relative to the other will always be a conic section (i.e., circle, ellipse, parabola, or hyperbola). In such a case, the predictive power of those seven pieces of information (time, three position vector coordinates, and three velocity vector coordinates) will be extremely good. For all such trajectories, knowing an object's position and velocity vectors at a given instant in time is sufficient information to predict its motion at all other instants in time, future or past, along that curve.

As a specific example of a conic section, let's imagine an object in an elliptical orbit (recall that Kepler demonstrated that the planetary orbits are ellipses with the Sun at a focus) and ask how many pieces of information you need to define all points on the ellipse as a function of time. First, you need to define the size and shape of the ellipse. Figure 2.1 illustrates how this is done. Given a pair of points, the foci of the ellipse ($F1$ and $F2$ in the figure), and a distance d, an ellipse is defined as the set of points $\{X\}$ for which the sum of lengths $F1$-to-X plus $F2$-to-X is equal to d. The major axis is the longest line segment connecting two points on the ellipse; in Fig. 2.1 it connects points A and B, and passes through both foci. Half the major axis is called the semi-major axis, symbolized by a. The ellipse minor axis bisects the major axis at right angles, connecting points C and D in the figure. Half the minor axis is called the semi-minor axis, symbolized by b. The semi-major and semi-minor axes of an ellipse define its contours in the same manner that the radius of a circle defines its contours. Indeed, a circle is the special case of an ellipse in which the values of the semi-major and semi-minor axes are identical (equal to the circle radius), with the distance d being the circle diameter.

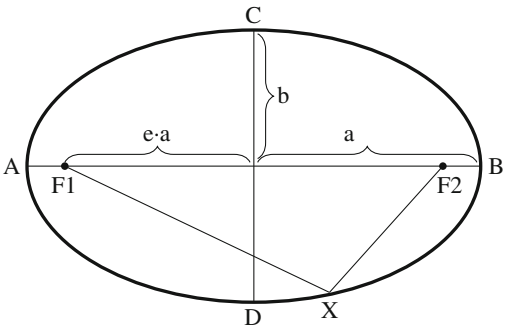


Fig. 2.1 Parameters that shape an elliptical orbit

Another way of describing the shape of the ellipse is via the eccentricity, a unitless parameter that is related to the semi-major and semi-minor axes via Eq. 2.2:

$$e = \frac{\sqrt{a^2 - b^2}}{a} \tag{2.2}$$

where

e = eccentricity (unitless)
a = semimajor axis
b = semiminor axis

Before moving on (in the next section) with parameters that characterize the orientation of an ellipse in space, let's explore some physics associated with various elliptical orbits. Let's consider a sequence of ellipses defined by a small satellite orbiting a massive central body (at $F1$), with the satellite at some initial time located a distance R_0 from the massive body and having velocity $\mathbf{V}_0$ perpendicular to the line segment back to the central body. It is important for this discussion that $\mathbf{V}_0$ is perpendicular to the line back to the central body; it implies that at our initial time the satellite is at either a point of closest approach to or farthest retreat from the central body (perigee or apogee, if the central body is Earth). As it turns out, if $V_0^2 = GM/R_0$, the orbit is a circle. If $V_0^2 < GM/R_0$ (but not zero), then the orbit is an ellipse with $0 < e < 1$, and the initial position is one of furthest retreat. The satellite falls in towards the central body, speeding up as it does so. Because the satellite's initial velocity was perpendicular to the line back to the central body, the satellite at first moves away from that line. Eventually, however, gravity pulls it back towards the line (after it passes the minor axis line), causing the satellite to whip around the central body. After passing the point of closest approach, the satellite rises back up to its initial position. If $V_0^2 = 0$, then we're just dropping the object from rest, and it falls into the central body. Although this orbit is still formally an ellipse, in this limit it is a simple line segment connecting the two foci ($F1$ being the location of the central body, and $F2$ the initial position of the satellite). The minor axis has shrunk to 0, and the eccentricity for the line segment is 1. (Well, strictly speaking, this discussion for the "$V_0^2 = 0$" case should really be taken as the limiting case as V_0^2 approaches infinitesimally close to 0. Since we're dealing with a point mass approximation for the central body, we really can't deal with the infinite forces that would apply as the orbiting body encounters the central point.)

Now let's go in the other direction, one of increasing V_0^2. For $GM/R_0 < V_0^2 < 2GM/R_0$, we again have an ellipse with e increasing from 0 to just under 1. Our initial position now is one of closest approach, the initial velocity being as large as it will ever be. As the body moves away from its point of closest approach, the gravitational force from the central body slows it down until it reaches its point of furthest retreat, after which it falls back to its initial position. For $V_0^2 = 2GM/R_0$, the satellite never reaches its point of further retreat. Indeed, for our initial conditions, the velocity $(2GM/R_0)^{1/2}$ is just enough for the satellite to escape from the central body with zero velocity at infinite distance. Although e was heading towards 1 in

this limit, Eq. 2.2 is only well defined for finite a and b. Both a and b become infinite at $V_0^2 = 2GM/R_0$. Suffice it to say that the orbit is no longer an ellipse, but rather has become a parabola with arms that continue to extend outwards forever. (You can imagine that the location of $F2$, as well as the lengths a and b, have been pulled out to infinity.)

If V_0^2 is increased further, the arms of the orbit pull apart yet further and the orbit becomes that other conic section, the hyperbola. "Hyperbola" comes from the same Greek root as "hyperbole", meaning excess; satellites in hyperbolic orbits may be considered as having excess (greater than zero) energy at infinite distance. Geometrically, with appropriate redefinitions of parameters a, b, and e that we won't go into here, hyperbolas have eccentricity $e > 1$. They accurately model the orbits of uninvited (and usually undesired) guests to our Solar System, as they (hopefully) follow near-miss trajectories relative to the permanent residents. And for departing residents, parabolas and hyperbolas describe the paths followed by objects that have reached escape velocity and are departing from the central body to which they had previously been bound.

2.3 Specifying the Orbit's Orientation in Inertial Space

Getting back to ellipses and how to specify their orientation in space, we now need to specify how the plane defined by the ellipse is oriented relative to a coordinate frame, e.g., GCI for an Earth orbiting satellite (which we'll assume for the remainder of this section). Imagine the ellipse superimposed on the GCI frame's equatorial plane with the minor axis aligned with the vernal equinox. Now tilt the ellipse about a line that runs through the center of the Earth and parallel to the minor axis. The line about which the tilt is done is called the *line of nodes* (only coincidentally parallel to the minor axis in this example). The tilt angle is called the *inclination* and is a right-handed rotation about the line of nodes. The inclination's range of rotation is 0–180°, where 90° will yield a polar orbit and inclinations greater than 90° yield retrograde orbits. But placing the ellipse's minor axis at a special orientation relative to the vernal equinox was, well special. We can generalize further by rotating the ellipse's plane about the GCI z-axis. This angle is called the *right ascension of the orbit's ascending node*, where the *ascending node* is the point on the orbital ellipse where the object rises from the GCI frame's southern hemisphere and enters the northern hemisphere. (The point where the object moves from the northern to southern hemisphere is called the *descending node*, hence *line of nodes* as the term for the line connecting the nodes.) The right ascension of the ascending node is measured from the vernal equinox (the GCI x-axis) and has a range from 0 to 360°. Note that for an equatorial orbit (inclination equal to zero) there is no orbital ascending node or descending node, so the right ascension of the ascending node is undefined. This is an orbital equivalent of the singularity situation we talked about in Chap. 1 when mentioning that the right ascension Euler angle is undefined when the declination is ±90°. Figure 2.2 illustrates how these two angles (right ascension of ascending node

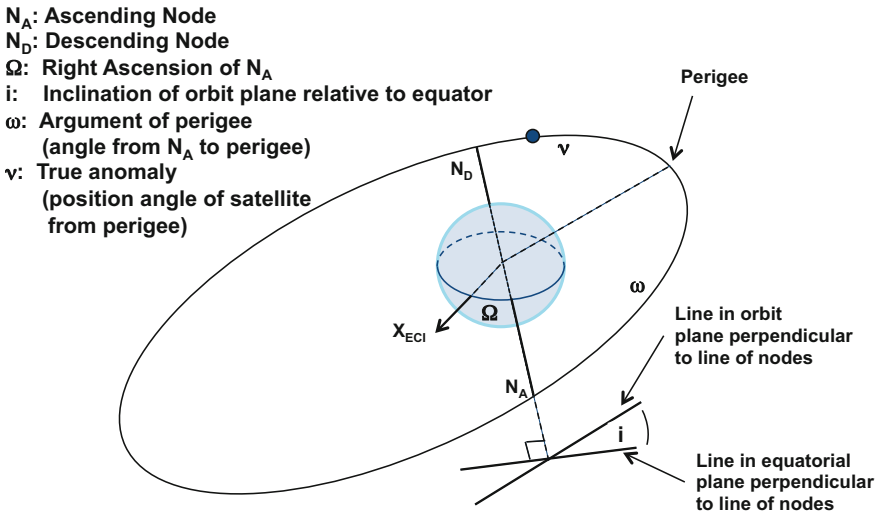


Fig. 2.2 Parameters that orient an elliptical orbit

and inclination) define the orientation of orbital plane in space. It also illustrates two other important angles, to be described in the next paragraph and next section.

Another parameter is still needed for full generality, because there's no requirement that the ellipse's major axis be aligned so that those points on the ellipse (at the ends of the major axis) take on the largest positive and negative Z-component values. Instead, we can envision a rotation about the orbital plane normal vector that places the major axis in a more general configuration within the (now) fixed orbital plane. For example, a 90° rotation about orbit normal would interchange major axis with minor axis. This angle of rotation within the orbit plane is called the *argument of perigee* for Earth orbits, or in general is called the *argument of periapsis* (again, see Fig. 2.2). By convention, it is the angle measured in the orbital plane from the orbit ascending node to (for Earth orbits) perigee, the object's closest approach to the Earth during an orbital period. The range of the argument of perigee is 0–360°. As with the right ascension of the ascending node, the argument of perigee is undefined for equatorial orbits (i.e., inclination equal to zero) because the ascending node (from which it is measured) is undefined. Although the two angles are undefined when inclination is exactly zero, you can consider any given such orbit as the limiting case for a family of orbits with non-zero but infinitesimal inclination, all with different ascending node right ascension and argument of perigee, but all having the same value for the sum of the two angles, as long as the orbit is not circular.

2.4 The Location of the Spacecraft in the Orbit

So now with five parameters (semimajor axis, eccentricity, inclination, right ascension of the ascending node, and argument of perigee), we're able to fix the orbit ellipse's size and shape, orbit plane orientation, and orbit orientation within the orbit plane. But there's one thing we need to know that has not yet been mentioned, namely the position of the orbiting object within the orbit at the current time, or some other time of interest. That information is supplied via the *true anomaly* at a given epoch time, which is the actual angular position of the object along the orbit relative to perigee. So this geometric model for describing an orbit as a whole requires exactly six parameters and a time to define an object's orbit, just as the previously described "vectorial' perspective (where we looked at what the object was doing while in its orbit) required exactly six parameters (position and velocity vectors) and a time. The *epoch time* is usually the time associated with the start time of an ephemeris file or the beginning of a new orbit, for example as a result of an orbit modification maneuver.

However, the true anomaly is difficult to calculate (it is the root of a transcendental equation), so in practice a more straightforwardly determined parameter, the *mean anomaly*, is provided instead. The mean anomaly is also an angular displacement from perigee, but is determined simply by dividing the time since perigee passage by the orbital period, then multiplying by 2π radians:

$$M = \frac{2\pi}{T}(t - T_P) \tag{2.3}$$

where

M = mean anomaly at time t (radians)
T = orbital period
T_P = time of perigee passage
(t, T, and T_P are all measured in the same time units)

In other words, it is the fraction of the period of the orbit covered since perigee passage. The mean anomaly does not have the true anomaly's concrete geometric meaning (except for circular orbits, where they have the same value), but the two can be related via the following two equations:

$$M = E - e\sin(E) \qquad \text{"Kepler's Equation"} \tag{2.4}$$

$$\tan(\nu/2) = ((1+e)/(1-e))^{1/2}\tan(E/2) \qquad \text{"Gauss's Equation"} \tag{2.5}$$

where

ν = true anomaly
M = mean anomaly
E = eccentric anomaly
e = eccentricity

Since this book is supposed to be the Kinder, Gentler ACS, we'll avoid getting caught up too much in these different anomaly flavors; it is sufficient to consider E an intermediate variable to compute ν from M. (It would have been nice if E could be expressed as an analytic function of M, as opposed to the other way around, but we'll take what we can get.) Let's concentrate now on the mean anomaly at epoch and see under what conditions it becomes undefined, as we did with right ascension of the ascending node and argument of perigee. First, if the orbit is circular, perigee is undefined and so is mean anomaly, but again we have a situation where the orbit may be considered a limiting case for a family of orbits defined by the sum of two angles, in this case it is the sum of the argument of perigee (ω) and the mean anomaly (M) that is well-behaved (for a non-equatorial orbit). This bit of hocus-pocus works because the angle from the ascending node (a well-defined point for non-equatorial orbits) and the spacecraft location (always a well-defined point) is equal to the sum of ω and M. Similarly, if the inclination is zero and the orbit is circular, then the mean anomaly, argument of perigee, and the right ascension of the ascending node will all individually be ill-defined, but the quantity representing their geometric sum (i.e., the angle from the the GCI x-axis to the spacecraft) will be constant for a family of Kepler parameter sets all representing the same orbit.

Although this focus (no pun intended) on singularities seems somewhat esoteric, utilization of these relationships between orbital parameters for "degenerate" orbits can be of great value when transforming an orbital (position, velocity) pair into equivalent Keplerian elements. For example, these singularity relationships were utilized repeatedly when developing the ground system algorithms for computing the uplink parameters required to refresh HST's onboard ephemeris models.

2.5 Keplerian Element Types

Putting all six of these parameters (called Keplerian elements) and an epoch time together, we have all the information required to define the current position of an object in a perfect elliptical orbit and, in the absence of orbit perturbations, have all the information required to determine the orbital position of the object at any other time as well, future or past. It also can be shown (but not here) that specifying the six components of two object position vectors on the ellipse at two different times (as long as the vectors are not parallel) provides the same information as the other two approaches (i.e., Keplerian elements and (position, velocity) vector pair, both with associated times) although, if the angle between the two position vectors is near 0 or 180°, the results of the computation of new position vectors may be numerically unstable. These relationships were also utilized when developing the ground system ephemeris uplink parameter generation algorithms for HST. So we have a similar situation with orbit formulations as we had with the attitude formulations discussed in Chap. 1. There are lots of ways to describe an orbit, each having its own advantages and disadvantages.

The Keplerian elements provide for orbits the sort of visualization advantages that the Euler angles provided for attitude. You can easily grasp the size and shape of the orbit, how the orbital plane is oriented, where perigee is located, and where the spacecraft is located on the ellipse at the epoch time. Also, in the absence of orbit perturbations, only the mean anomaly changes with time. On the downside, we have seen that three out of six Keplerian elements become ill-defined for special geometries, circular and equatorial orbits that tend to be very "popular" for Earth orbiting science missions. And although the (position, velocity) vector formulation is pretty user-hostile from a visualization standpoint, it is ideal for numerical calculations of "real" orbits in the presence of perturbative effects, especially when an ephemeris file with many vector pairs is provided, enabling interpolation between points. So in practice, you pick your formulation for its convenience relative to the problem you're trying to solve or the function you're trying to perform.

There are even two types of Keplerian elements, each having their own uses. Implicitly, the type we've been talking about are Keplerian *osculating elements*, so called because they will generate an ideal orbit that will "kiss" the physical orbit (when orbit perturbation effects are included) at a single point, the point corresponding to the epoch time. However, in the presence of perturbations, the ideal osculating orbit and real physical orbit will diverge over time. A more accurate fit to the physical orbit can be achieved through the use of *Brouwer mean elements*. Brouwer mean elements effectively provide averaged elements over several orbits, thereby averaging out the influence of periodic perturbative effects rather than (as would the osculating elements) incorporating the value of the perturbation value at the epoch time as if it was constant in time. So osculating elements may propagate and amplify the effects of periodic perturbations with time, while mean elements will bound their effects. But without inclusion of explicit time dependence (i.e., adding derivative terms to the basic six-term Keplerian set), neither approach will deal effectively with secular (i.e., ramping) perturbations over long propagation periods. Probably by now the term "perturbative effects" is sounding like some sort of physics *deus ex machina* invoked to intimidate an audience from asking orbit questions. ("Quaternion" is another good term for that purpose when presenting attitude material to a review board.) Just to put that suspicion to rest, we'll spend the rest of this chapter discussing the physical phenomena (and manmade machinery) that perturb orbits and the orbital geometries particularly susceptible to those influences.

2.6 Orbit Perturbations - Oblate Earth

Perturbative effects on Earth orbits are caused by forces other than those associated with a "point Earth" (i.e., spherically symmetric Earth) attracting the object. An analytical approach in which you treat all other influences as perturbations will only be successful numerically if any additional forces are much smaller in magnitude than that generated by the point Earth, but fortunately (for overworked flight dynamics analysts) that is usually the case for spacecraft orbiting the Earth. For the

missions we've worked with here at GSFC so far, there are five forces that exert particularly significant influence on spacecraft orbits, namely non-point Earth gravitational forces, non-two-body gravitational forces, aerodynamic drag, solar radiation pressure, and thrusters.

Back in elementary school you probably learned that the Earth is shaped more like a pear than an apple (disregarding flat bottoms and pointy tops). This is still true today, and this aspect of the Earth's structure, referred to as Earth oblateness, is the key enabler of two major mission classes. If you model the Earth as an oblate spheroid (a good approximation to the slightly pear-shaped distortion mentioned by your teacher), great circles along a meridian on the Earth's surface are replaced by ellipses whose semi-major axes are equal to the Earth's equatorial radius (about 6,378 km) while their semi-minor axes are equal to the Earth's polar radius (about 6,357 km), yielding an ellipse eccentricity of about 0.08. Small circles at constant latitudes will remain small circles, but their radii will be slightly modified (from the spherical Earth model) because of the equatorial bulge relative to the poles.

To get an idea how oblateness influences a spacecraft's orbit, let's start with the spherical Earth. Assume the radius of the sphere is the "real" Earth's polar radius. In the absence of any perturbing influences, the orbit sketched out previously in Fig. 2.2 will be constant in time. Next, replace the spherically symmetric Earth model with one in which some of the mass is moved into an equatorial bulge extending about 20 km out from the immediate vicinity of the equator. When the spacecraft is at its orbital nodes, the force arising from the mass in the bulge exerts an in-plane force on the spacecraft in the direction of the bulge's (and Earth's) center. Visualizing the spacecraft's orbit with orbit normals sticking out of it as if it were a rotating top (where the spacecraft as it orbits is sitting on the external edge of the top), the effect of the bulge at the nodes is just a force applied to the center-of-mass of the top. By contrast, at the points 90° from the nodes, the bulge mass exerts an out-of-plane force that is trying to tip the top down (i.e., reduce the orbit inclination). As we'll see in Chap. 3, that's the same thing as saying the bulge is exerting a torque on the top, whose effect will cause the top's spin axis (i.e., the orbit normal) to precess westwards (when inclination is less than 90°) about the Earth's spin axis. The direction of rotation is determined by the cross product of the spacecraft position vector with the direction of the force. Expressed in terms of Keplerian elements, that's equivalent to saying that the equatorial bulge will cause the orbit's right ascension of the ascending node to rotate.

The rate of the nodal rotation depends mostly on the altitude and inclination of the spacecraft's orbit. As you get farther and farther away from the Earth, the Earth acts more and more like a single point of gravitational attraction. Mathematically, you can show (but we won't for fear of losing what little readership we have left) that the nodal precession rate varies inversely with the 3.5th power of the semi-major axis, so the rate drops off very quickly with altitude. The effect of inclination is also easy to see. If the spacecraft is in a perfect polar orbit (90° inclination), the orbital plane cuts symmetrically through the equatorial plane and therefore the torque generated by the bulge must go to zero (to first order). As you decrease inclination, the influence

of the bulge steadily increases, although you reach a singularity at inclination zero since the right ascension of the ascending node becomes undefined.

To get a more quantitative feel for the size of the nodal rotation induced by the bulge, let's look at a couple of standard orbits. For a 500-km altitude near-circular orbit with an inclination of 28.5° (the inclination you get for "free" by launching from Kennedy Space Center (KSC), whose latitude is 28.5°), the nodal rotation will be about −6.7 deg/day, i.e., a westward nodal rotation. By contrast, if you're in an orbit with inclination 97.4° and altitude 500 km, the nodal rotation rate (eastwards) becomes 1 rotation/year (about 1 deg/day), which keeps the orbit synchronized with the annual movement of the Sun, i.e., Sun synchronous.[3] We'll talk more about these very special, and useful, orbits in the last chapter. Also, keep in mind that this is the simple picture you get from just looking at the first order perturbations, also called the J_2 term in reference to the coefficient of the second term (after the point mass term) in a spherical harmonic expansion of the gravitational potential (that mathematical derivation we mentioned but refused to present earlier). When the inclination is near 90°, the effects of Earth oblateness on the orbit typically become smaller than that produced by less symmetric components of the Earth's mass distribution, and the real evolution of the orbit becomes correspondingly more complicated.[4]

But wait, there's more (don't worry; we're not selling Ginsu knives). In addition to causing the line of nodes to rotate, the Earth's oblateness will also cause the line of apsides (the line joining apogee and perigee) to rotate. Again, imagine we've generated a static orbit as a result of the gravitational force exerted by a spherically symmetric Earth, which we can shrink to a single point at the focus of the spacecraft's elliptical orbit. Again, pull some of that mass into a ring modeling the equatorial bulge. For validity of the approximation, assume the bulge is even smaller than the Earth's 20-km one. Let's also (referring to Fig. 2.2) visualize that the spacecraft orbit is lying in the equatorial plane (i.e., has inclination zero) and that the spacecraft's orbit is highly elliptical. At perigee (in particular), the spacecraft will feel a stronger force than that resulting from the entire bulge mass being concentrated at the Earth's center. The effect is dominated by the portion of the bulge closest to the spacecraft.[5] Since the extra tug at perigee is perpendicular to the velocity vector, it won't add or detract from the spacecraft's kinetic energy, so the height at apogee will stay constant, unlike the case where we can raise (or lower) apogee by adding (or removing) kinetic energy at perigee by thrusting parallel (or anti-parallel) to the direction of the velocity vector. Instead, the extra gravitational pull at perigee will cause a tighter deflection

[3] Wertz, Spacecraft Attitude Determination and Control, pp. 68–69.

[4] In more formal mathematical terms, the secular J_2 term becomes smaller than the periodic terms associated with the zonal, sectorial, and tesseral coefficients that model the more complex (latitude, longitude) dependent mass distributions, as we'll discuss in the section on the Earth's geopotential in Chap. 6.

[5] Using a simple model in which the bulge is represented by a thin, massive wire girdling the equator, one can show that in the limit in which the spacecraft almost grazes the wire at perigee, the gravitational pull from a small fraction of the wire closest to the spacecraft will be inversely proportional to the distance between the wire and the spacecraft (becoming infinite if the separation is zero) and will completely dominate the pull from the rest of the wire.

of the spacecraft's path than would have been the case for an orbit about a point-mass Earth. This will cause the new apogee point to be shifted in a positive orbit sense relative to the previous apogee without changing the distance between apogee and the Earth's center. As a shift in apogee will occur on each orbit, what we end up with is (to first order) a constant rotation of apogee (and its associated perigee) in the direction of the spacecraft's orbital motion.

By contrast, for a polar orbit with perigee initially located at a pole, the spacecraft at perigee (in particular) will feel a little bit weaker force from the bulge than that resulting from all the bulge mass being concentrated at the Earth's center. Note that the Earth's center is closer to the pole than is any point on the equator, so moving mass from the center to an equatorial ring reduces the pull felt at the pole. Once again, the diminished tug at perigee is perpendicular to the velocity vector, so no change in the spacecraft's kinetic energy results. But the reduced gravitational attraction at perigee will cause the spacecraft's path not to be deflected as tightly as would be true if the spacecraft had been orbiting a point-mass Earth, so the new apogee point will be shifted in a negative orbit sense relative to the previous apogee without changing the distance between apogee and the Earth's center. So for polar orbits we get a rotation of perigee in the direction opposite of the spacecraft's orbital motion.

Logically, if perigee rotations at zero inclination are positive and perigee rotations at polar inclination are negative, there should be some "magic" inclination angle between 0 and 90° at which the perigee rotation is nulled, and in fact there is. At an inclination of 63.435° (ignoring terms of higher order than the secular J_2 term), the perigee rotation rate goes to zero, yielding a "frozen" orbit. This type of orbit has very useful communications applications for countries located in high latitudes, like Russia, as we'll see in the last chapter if you survive the intervening 9 chapters of material. As an example of the size of the effect, for a 500-km near-circular orbit with inclination 28.5°, the perigee rotation rate is 10.9° per day in a positive direction.[6]

Nearby the Earth, oblateness effects are the dominant influence on spacecraft orbits. However, as you get further from the Earth, perturbative effects arising from the Sun's and Moon's gravity become more important. As a rule of thumb, if the spacecraft altitude is below 700 km, lunar and solar gravitational effects can be ignored, while at 8000 km, lunar perturbations are comparable in importance to Earth oblateness.[7] For a great many missions, for example those orbiting Lagrange points, two-body approximations break down completely and the spacecraft's orbital behavior can only be defined by analyzing the combined influences of two celestial bodies. Continuing the running gag for this chapter, a discussion of Lagrange points and their associated missions also will be provided in Chap. 12.

[6] J. Wertz, Spacecraft Attitude Determination and Control, p. 69.

[7] Ibid, p. 63.

2.7 Orbit Perturbations - Aerodynamic Drag

Returning closer to home, aerodynamic drag dominates at altitudes below 100 km, and is an important influence on spacecraft orbits up to about 1,000 km.[8] Of the various perturbations discussed here, drag may be the most complicated, as there are so many factors entering into the picture. In particular, the drag force is proportional to atmospheric density (which is affected by altitude, latitude, constituent composition and chemistry, time of day, season of year, point in solar cycle, butterfly migration patterns, etc.), the square of the spacecraft velocity relative to the atmosphere, and the spacecraft cross-sectional area normal to the wind. Like friction, drag is a retarding force, so it acts to oppose the spacecraft direction of motion. Over the long term, drag eventually causes the orbit to decay, leading to a decrease in the semimajor axis. So unlike Earth oblateness, drag actually removes energy from the orbit. Drag also tends to make the orbit more circular for the following reason. Since drag has its greatest effect at the point in the orbit where the atmosphere is densest (drag decreases exponentially with altitude), the effects of drag will be highest at perigee. Therefore at perigee, more energy will be removed than at apogee. But removing energy at perigee (or apogee) means the swingby at apogee (or perigee) will not reach as high an altitude as on the previous orbit. So the perigee and apogee altitudes will be lower on each succeeding orbit, but the amount of lowering will be greater at apogee than at perigee. This also is due to the inverse distance behavior of the gravitational potential energy, which will cause the same amount of energy reduction to produce a larger altitude decrease at apogee than at perigee. So until the two altitudes roughly become identical, the effect of drag will be to make the orbit steadily more circular, i.e., it will steadily lower eccentricity. Note that these changes from energy loss are only in-plane. Drag has no effect on inclination or right ascension of ascending node, nor for that matter on argument of perigee (at least until it circularizes the orbit) .

Before leaving our discussion of atmospheric drag, it's worthwhile to spend a little time talking about how you model the key input to drag computations, atmospheric density. Of course, the higher you go, the thinner the air, following an exponential drop-off. The atmospheric density will also be influenced by temperature (via the Ideal Gas Law), which decreases as we move away from the equator. But the dominant factor determining the molecule's energy content is solar heating, the effects from which vary over several time scales. The highest frequency influence is the daily variation caused by ultraviolet radiation heating the atmosphere by conduction, causing atmospheric density to increase to its maximum level by 2–3 h after local noon. There also is a 27-day solar activity period, the yearly period (i.e., seasonal patterns), and the 11-year solar cycle, which reached a maximum in 2001. Lumped in with these fairly predictable events, we also have short-term fluctuations in atmospheric density arising from the solar wind and completely unpredictable, violent perturbations from solar flares and coronal mass ejections. At GSFC, two models in particular have found popular use, especially as input to GSFC's orbit

[8] Ibid, p. 63.

determination program, the Goddard Trajectory Determination System (GTDS).[9] The simpler of the two is the Harris-Priester model,[10] which accounts for diurnal (a fancy word for *daily*) influences. The Jacchia-Roberts model[11] includes seasonal variations and constituent composition (at least at the level of molecular weight behavior as a function of altitude) as well as atmospheric density changes caused by solar flares and geomagnetic activity.

2.8 Orbit Perturbations - Solar Radiation Pressure

Once you get up high enough that the atmosphere has virtually disappeared and the Earth looks round (i.e., away from the neighborhood that Low Earth Orbit (LEO) spacecraft call home), solar radiation pressure takes over as the major environmental orbital perturbation. Solar radiation pressure is much easier to model, depending largely on solar luminosity, spacecraft reflectivity, distance from the spacecraft to the Sun, and spacecraft cross-sectional area. The cross-sectional area itself is a function of attitude, which for celestial-pointing missions will change whenever the science target changes. So the actual solar radiation pressure orbit perturbations are closely tied to the science observing schedule. In practice, as when atmospheric drag is modeled, GTDS is not set up to deal with this dynamic attitude behavior, so an average cross-sectional area approximately valid for the time duration of interest is what's input. In any event, solar radiation pressure is proportional to the spacecraft cross-sectional area normal to the sunline (as opposed to the cross-sectional area normal to the spacecraft velocity vector as in the case of atmospheric drag). It also is proportional to the momentum flux from the Sun. The flux, in turn, is proportional to the solar luminosity and is inversely proportional to the square of the distance from the Sun. To see that, no pun intended, just imagine light emitted by the Sun evenly distributed over a sphere of radius equal to the separation distance. The constant of proportionality for the pressure calculation is a function of the spacecraft reflectivity. A reflectivity of 2 is perfectly reflective, while a value of 1 denotes perfectly absorbent, and a value of 0 means perfectly transparent. Most spacecraft are pretty shiny, so aluminum's reflectivity of 1.95 is characteristic of most spacecraft.

[9] A. Long, J. Cappellari, C. Velez, and A. Fuchs. Goddard Trajectory Determination System (GTDS) Mathematical Theory, Revision 1. Technical report, NASA/GSFC Flight Dynamics Division Code 550, 1989.

[10] I. Harris and W. Priester, Time-dependent Structure of the Upper Atmosphere, Journal of Atmospheric Science, 19, pp. 286–301, 1962.

[11] C. Roberts, An Analytic Model for Upper Atmosphere Densities Based Upon Jacchia's 1970 Models, Celestial Mechanics, Volume 4, Issue 3, pp. 368–377, Dec 1971.

2.9 Orbit Perturbations - Orbit Maneuvers with Thrusters

The last perturbation we'll talk about, thrust, is completely artificial and, unless something goes badly wrong, planned, predictable, and easily modeled. Depending on the size of the thrusters mounted on the spacecraft, thrust can be the dominant influence on the spacecraft orbit. Thrusters often come in two functional flavors, *orbit thrusters* and *attitude thrusters*. Orbit thrusters, of course, are designed to generate major orbit changes and trims. Although attitude thrusters, as their name implies, are designed to enable attitude control and angular momentum management, misalignments in the thrusters can produce minor, undesirable, but somewhat predictable orbit perturbations. We'll have a lot more to say about thruster hardware and electronics in Chap. 5, but for now we'll just mention briefly some strategies for their use.

Recall in our discussion of drag, frictional removal of energy at perigee will lower apogee while frictional removal of energy at apogee will lower perigee. This same physical effect can be achieved artificially by firing thrusters so as to decelerate the spacecraft at those points (actually on an arc centered on those points). By contrast, firing orbit thrusters so as to accelerate the spacecraft (i.e., artificially adding kinetic energy) at perigee will raise apogee while doing the same at apogee will raise perigee. These kinds of orbit maneuvers are called in-plane maneuvers, and tend to cost much less fuel than out-of-plane maneuvers (nodal rotation or inclination change), which change both the magnitude and direction of the orbital angular momentum vector (in-plane maneuvers can only change the magnitude). If you want to produce pure orbit inclination changes (say, lowering the "automatic" 28.5° you get with a KSC launch to an equatorial orbit), the most cost-effective place to perform your "burns" is at the orbital nodes. Note that because of fuel budget limitations you often accomplish the desired orbital modification simply by initiating a slow, steady drift in the orbital elements in the appropriate direction. At a later time, possibly months later, an inverse orbit maneuver can be performed to cancel the drift, leaving the spacecraft in the desired orbit.

Chapter 3
Angular Momentum and Torque

3.1 Historical Digression

During the 17th century, great physicists and mathematicians such as Descartes, Huygens, Newton, and Leibniz laid the foundations for our current classical understanding of translational motion, orbital motion, momentum, force, and energy, connecting to the physical world via the phenomena of mass and gravitation. But it wasn't until Euler's work in the middle of the 18th century that a formal description of general rotational motion was established. When you look at all of Euler's contributions to the field of mathematics (he played a part in advancing the state of the art in every branch of mathematics known at the time), it's unclear whether his seminal "Theory of the Motions of Rigid Bodies" was a conscious intent to fill an existing void (it was at least in part stimulated by what, at that time, was the unsolved problem of the precession of the equinoxes, which was discussed in Chap. 1) with malice of forethought or a generic effort to tidy things up. Regardless, most of the non-phenomenological stuff we'll talk about in this chapter, in effect the fundamental physics upon which spacecraft attitude control is based, was developed by Euler about 200 years before the first satellite was launched, but 100 years after Newton developed his Law of Gravitation, which even now is sufficiently accurate to describe the motion of spacecraft in their orbits in most situations of practical interest.

3.2 Translational Motion

Most people are very comfortable with the basics of translational motion. Specifying an object's position vector is a mathematical way of saying you know where it is relative to something or somewhere else. An object's velocity vector, which is the first derivative (with respect to time) of its position vector, just tells you how fast the object is going and in what direction, again relative to something or somewhere else. And an object's acceleration vector (the second derivative of its position vector)

© Springer-Verlag London Ltd. 2017

H.L. Hallock et al., *ACS Without an Attitude*, NASA Monographs
in Systems and Software Engineering, DOI 10.1007/978-1-4471-7325-0_3

describes how quickly the object's speed and direction of motion are changing. All these mathematical creations have quite a pedigree, dating back to the time of Galileo (over 400 years ago) and earlier. Translational momentum, force, and energy are somewhat less intuitive concepts, but can be related to everyday experience fairly easily. The translational momentum vector, the product of an object's mass with its velocity vector, is a quantitative way of saying that heavier objects are more difficult to deflect or slow down than light ones (e.g., if you're a cornerback trying to stop a fullback who has a full head of steam, better tackle him low or he'll shrug you off or carry you in for a touchdown along with the football). The force vector, the product of an object's mass with its acceleration vector, and therefore the first derivative of momentum, says that heavier objects deliver bigger wallops than lighter ones (e.g., you don't use a sledgehammer to crack the shell of a hardboiled egg). Note that since force is the first derivative of translational momentum, if the total force acting on a body is zero, its translational momentum is conserved (i.e., is constant with time).

Although each of these five quantities represents different physical things with different units, being vectors they still have many things in common. For example, you can add vectors of the same type (say, two force vectors), a particularly convenient property if you're interested in knowing what is the net force acting on a body (a relevant concern both in physics and in politics). You can also rotate vectors from one coordinate reference frame to another (as we discussed in Chap. 1), which allows you to examine and solve problems in the most convenient reference frame, i.e., the one where the equations look simplest. By contrast, work and energy are rather different from vectors. Like non-relativistic time and mass, they are scalars, not vectors, and at least from a classical perspective are invariants under rotation and translation, as opposed to vectors whose components change when the reference frame changes. Keeping things on the simple side, work is the energy expended when force is applied over a distance – it's a lot harder to carry a non-productive team member for a year than it is to carry him for a month.

Energy itself covers such a wide spectrum of phenomenon that it would take much of this chapter just to list all its manifestations, but there are two generic flavors that should be mentioned at this point. Translational kinetic energy describes an object's energy of translational motion (big surprise) and is equal to half the product of the object's mass and the square of its velocity. So it takes more energy to increase something's speed by 1 mph when it's traveling at 60 mph than when it's moving at 6 mph (even ignoring frictional energy loss and wind resistance). This algebraic definition makes intuitive sense if you imagine you started with a particle at rest and then ask how much energy it would take to increase its momentum up to its current level. A natural way to calculate this is to integrate the particle's momentum over velocity from its starting value (zero) to its current velocity, which gives you the simple formula described previously, and accounts for the factor of one half. This also is really what's going on when you calculate the work performed. By that we mean you're integrating the force vector on the object over the distance vector the object moves, an operation on vector quantities called a line integral, rather than simply multiplying the force times distance. So the mathematically fancier line integral takes into account the actual path traveled rather than just the net change in

Table 3.1 Definitions of translational motion quantities

Quantity	Symbol	Units (MKS system)	Equations
Position	$\mathbf{R}$	meter	$\mathbf{R} = (x, y, z)$
Translational velocity	$\mathbf{v}$	meter/second	$\mathbf{v} = (v_x, v_y, v_z)$
			$\mathbf{v} = d\mathbf{R}/dt$
Translational acceleration	$\mathbf{a}$	meter/second^2	$\mathbf{a} = (a_x, a_y, a_z)$
			$\mathbf{a} = d\mathbf{v}/dt$
			$\mathbf{a} = d^2\mathbf{R}/dt^2$
Mass	m	kilogram	
Translational momentum	$\mathbf{p}$	kilogram-meter/second	$\mathbf{p} = (p_x, p_y, p_z)$
			$\mathbf{p} = m\mathbf{v}$
			$\mathbf{p} = m\, d\mathbf{R}/dt$
Force	$\mathbf{F}$	newton (kilogram-meter/second^2)	$\mathbf{F} = (F_x, F_y, F_z)$
			$\mathbf{F} = m\mathbf{a}$
			$\mathbf{F} = d\mathbf{p}/dt$
			$\mathbf{F} = m\, d\mathbf{v}/dt$
			$\mathbf{F} = md^2\mathbf{R}/dt^2$
Kinetic energy	T	Joule (kilogram-meter^2/second^2)	$T = mv^2/2$
Work	W	Joule (newton-meters = kilogram-meter^2/second^2)	$W = \int \mathbf{F}\, d\mathbf{R}$

position (e.g., running around in circles uses up a lot of energy even if you end up at your original starting point).

Potential energy is like a phenomenological bank account you can draw on for conversion into kinetic energy. For example, gravitational potential energy (as opposed to Newton's force law for gravity) varies inversely with distance, where the sign of the potential term is negative. So an object at a higher altitude will have a higher (less negative) potential energy than an object at a lower altitude. Therefore an object in an elliptical orbit about the Earth, as we've seen, trades off potential energy for kinetic energy (i.e., altitude for speed) as it moves from apogee to perigee. For those of you wondering when we're going to make a reference to some equation numbers, you'll find the ones we've been talking about in the last few paragraphs summarized in Table 3.1.

3.3 Rotational Motion

Getting back to the actual topic of this chapter, there are some straightforward parallels between translational motion and rotational motion. For example, we know from Newton's First Law of Motion that if no forces are acting on an object in motion, the object will continue moving in the same direction at the same rate forever. But

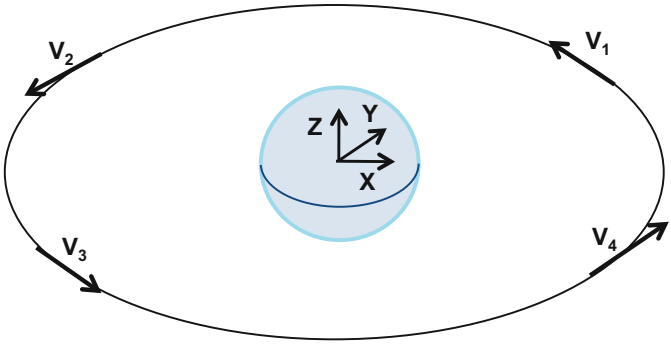


Fig. 3.1 Variation of velocity vector components over a circular orbit

suppose the object is in a circular orbit about a perfectly spherical Earth, with no orbital perturbations. In this case the object will remain in that same orbit forever. Although the magnitude of the object's velocity vector stays the same for all time, the individual components of the velocity vector will display a periodic behavior, with X-, Y-, and Z-components flipping signs every half orbit (for non-equatorial orbits), as illustrated in Fig. 3.1. For example, at Point 1 the velocity vector $\mathbf{V}_1$ has a positive Y-component, while at Point 3 the velocity vector $\mathbf{V}_3$ has a negative Y-component of equal magnitude. The velocity vector's direction must change smoothly over the orbital period because the velocity vector is tangent to the object's orbital path. Still, the magnitude of the velocity vector stays constant. To appreciate what's unchanging about the object's motion and why, we need to leave behind quantities defined with translational motion in mind and work instead with parameters more directly geared towards the physical rotational motion that's actually occurring.

So what is constant? Well, we said a moment ago that the magnitude of the translational velocity vector was constant although its direction (i.e., its individual components) varied periodically. Another constant is the radius of the circle, but that information is really redundant to the magnitude of the velocity vector when viewed from an energy standpoint (remember, the higher the altitude, the slower the velocity), provided we keep the central body's mass constant. We can also package this information as the angular rate (in units of radians/sec), which is equal to the magnitude of the translational velocity divided by the circle's radius. The only other independent constant (unless we count the value of the eccentricity, which is a constant value of zero just because we defined the problem that way) is the orientation of the orbit, which is most naturally defined by the orbit normal. Traditionally, positive orbit normal (what you get by applying right-hand rule with the fingers of your right hand curling in the direction of the orbital motion and the thumb pointing in the normal direction) is the polarity option that is selected (see Fig. 3.2). So two natural quantities describing in essence what the orbiting object is doing are the angular speed (also called the angular frequency or rate) and the positive orbit normal vector. If we multiply the vector and scalar quantities together, we get a vector called the

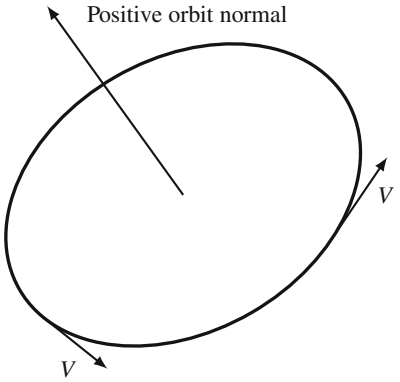


Fig. 3.2 Right-hand rule applied to orbital motion

angular velocity that tells us how fast the object is rotating and what its direction of rotation is – a nice rotational analog to the translational velocity vector.

If we want a quantity that tells us how difficult it is to change the object from doing what it's doing, we'll need a rotational analog to translational momentum. Obviously, it should involve the object's angular velocity and mass. But a simple product of the mass and angular velocity doesn't really capture fully what's going on. That's because you know intuitively, for example from your childhood experience swinging a rock tied to a string in a circle around your head, that the larger the radius (i.e., the longer the string) the more effort you have to exert to keep the object (rock) rotating at the same rate. So it seems natural to define an angular momentum as some kind of product, in fact a vector cross product, between the object's instantaneous position vector (whose magnitude in a circular orbit is the radius of the orbit, a constant) and the object's instantaneous translational momentum vector (whose magnitude in a circular orbit will also be a constant and whose direction is tangent to the circle at the orbiting object's position), as shown in Fig. 3.3.

$\mathbf{L} = \mathbf{R} \times \mathbf{p} = \mathbf{R} \times (m\mathbf{V})$
= angular momentum $(kg * m^2/sec)$
$\mathbf{R}$ = position (m)
$\mathbf{p}$ = translational momentum $(kg * m/sec)$
m = mass of object (kg)
$\mathbf{V}$ = velocity (m/sec)

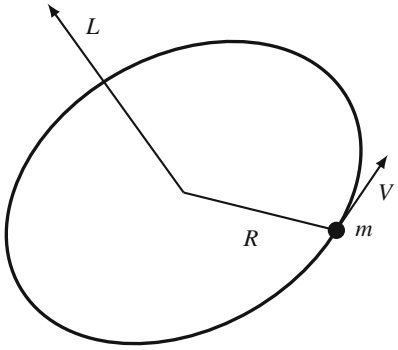


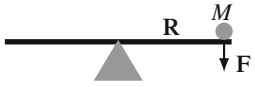
Fig. 3.3 Definition of angular momentum

Note that although the directions of the instantaneous position vector and instantaneous translational momentum vector keep changing, their cross product vector has a constant direction and is the positive normal to the circular orbit. This is exactly the property we were looking for as a rotational analog to the translational momentum vector, a vector quantity whose direction won't change if the rotational motion stays constant and whose algebraic structure suggests that greater efforts would be needed to deflect it from its rotational path if its mass or speed were increased. Written as a vector equation, we can then define the object's angular momentum (symbolized by **L**) to be $\mathbf{L} = \mathbf{R} \times \mathbf{p}$, where **R** and **p** are the instantaneous position and translational momentum vectors, respectively, and where their vector nature has been expressed symbolically through the use of "bold" font. Note that angular momentum is defined relative to a specific (position, velocity) coordinate system origin pair, and so will change if you change either aspect of the reference frame. For our current conversation, we are talking about the angular momentum of a point-mass object about a coordinate origin, e.g., a planet or satellite about the orbit center.

We can similarly find a nice rotational analog for force. Force was defined as the first derivative of translational momentum, so a logical rotational analog to force would be the first derivative of angular momentum, which is called torque. Note that if we take the time derivative of our previous equation for angular momentum ($\mathbf{L} = \mathbf{R} \times \mathbf{p}$), we get a similar expression for torque, $\mathbf{N} = \mathbf{R} \times \mathbf{F}$, where the second term you get when differentiating ($\mathbf{V} \times \mathbf{p}$) is zero since translational momentum has the same direction as translational velocity, and the cross product of two parallel or anti-parallel vectors is zero. This formal definition of torque also makes sense physically if you visualize (or remember painfully) the resulting motion (at least until you hit the ground) if you are sitting at the elevated end of a seesaw and a so-called friend of comparable weight sitting at the other end gets off the seesaw without warning (see Fig. 3.4).

To see the relationships between all these rotational quantities together in one place, see Table 3.2.

Some of the parameters in the rotation table may look a little weird at this point. For example, what's that ***I*** thing that seems to play the same role that mass did for translational motion, and why is it a 3×3 matrix instead of a simple scalar? At the root of that question is one of the big differences between translational motion of an object and rotational motion of the object about its own center-of-mass. For translational motion we can at least imagine that we can squeeze all of an object's



$\mathbf{N} = \mathbf{R} \times \mathbf{F} = \mathbf{R} \times (M\,\mathbf{A})$
$= \text{torque } (kg * m^2/sec^2)$
$\mathbf{R} = \text{position } (m)$
$\mathbf{F} = \text{force } (kg * m/sec^2)$
$M = \text{mass of object } (kg)$
$\mathbf{A} = \text{acceleration } (m/sec^2)$

Fig. 3.4 Definition of torque

Table 3.2 Definitions of rotational motion quantities

Quantity	Symbol	Units (MKS system)	Equations
Angular displacement	$\boldsymbol{\Theta}$	radian	Euler Set= $\boldsymbol{\Theta} = (\theta, \phi, \psi)$
Angular velocity	$\boldsymbol{\omega}$	radian/second	$\boldsymbol{\omega} = (\omega_x, \omega_y, \omega_z)$
			$\omega_x = d\theta/dt$
			$\omega_y = d\phi/dt$
			$\omega_z = d\psi/dt$
Angular acceleration	$\boldsymbol{\alpha}$	radian/second^2	$\boldsymbol{\alpha} = (\alpha_x, \alpha_y, \alpha_z)$
			$\boldsymbol{\alpha} = d\boldsymbol{\omega}/dt$
			$\alpha_x = d^2\theta/dt^2$
			$\alpha_y = d^2\phi/d^2$
			$\alpha_z = d^2\psi/dt^2$
Moment of inertia tensor	$\boldsymbol{I}$	kilogram-meter^2	$\boldsymbol{I} = \{I_{ij}\}$, where $i = 1, 2, 3$ and $j = 1, 2, 3$
Angular momentum	$\mathbf{L}$	kilogram-meter^2/second	$\mathbf{L} = (L_x, L_y, L_z)$
			$\mathbf{L} = \boldsymbol{I}\boldsymbol{\omega}$
			$\mathbf{L} = \mathbf{R} \times \mathbf{p}$
			$\mathbf{L} = m\mathbf{R} \times \mathbf{v}$
Torque	$\mathbf{N}$	Joule (newton-meter = kilogram-meter^2/second^2)	$\mathbf{N} = (N_x, N_y, N_z)$
			$\mathbf{N} = \boldsymbol{I}\boldsymbol{\alpha}$
			$\mathbf{N} = \boldsymbol{I}d\boldsymbol{\omega}/dt$
			$\mathbf{N} = \boldsymbol{\omega} \times \mathbf{L}$
Kinetic energy	T	Joule (kilogram-meter^2/second^2)	$T = \boldsymbol{\omega}\boldsymbol{I}\boldsymbol{\omega}/2$
Work	W	Joule (newton-meter = kilogram-meter^2/second^2)	$W = \int \mathbf{N} \cdot d\boldsymbol{\Theta}$

mass down to a point (called the object's *center of mass*) and just follow where that point goes in order to figure out where the whole object will go. By contrast, if you're not careful about how you toss an oblong object in the air, you'll see it tumble in all directions, even though its center of mass just goes up and comes back down. So in order to describe the object's full range of motions, you need to describe how it's moving relative to its center of mass as well as how the center of mass is moving. And in order to do that, you need to be able to describe what the shape and mass distribution of the object is, not just its total mass. That's where the more complicated quantity $\boldsymbol{I}$ comes in. We'll talk more about $\boldsymbol{I}$ (without getting too egotistical) and motions about the center of mass in the next two sections.

3.4 Motion of the Center of Mass Versus Motion About the Center of Mass

At first glance it would seem that we're in pretty good shape, but what we've come up with must be incomplete by the very nature of the approach we've used. Using our experience with objects in translational motion, we've successfully found rotational analogs for each of the translational vectors. Specifically, angular velocity goes with translational velocity, angular momentum goes with translational momentum, and torque goes with force. And there are others we haven't talked about yet, such as angular acceleration versus translational acceleration that could be added to the list. But there's a subtle question that hasn't been addressed explicitly, namely just what is it that is in motion? When speaking of an object undergoing translational motion or simple orbital rotational motion, what we've really been modeling with equations is the translational or rotational motion of the object's center of mass. So in the preceding discussions, when we've said there's a force on the object, what we've meant is that the force will cause the center of mass of the object to accelerate (assuming the object is not nailed down, or other forces are present that produce a cancelation). And this is the way things work as long as we restrict ourselves to forces acting through the object's center of mass. But what happens if you apply a force to the object in a more general direction? Suppose we define a coordinate system whose origin is located at the object's center of mass. Then a force that is applied directly to the object's center of mass will produce a pure translational motion and, ignoring perturbative effects (such as atmospheric drag), the shape and structure of the object is irrelevant. But a force applied offset to the object's center of mass will generate a non-zero torque on the object, leading to a rotational motion about the object's center of mass as well (see Fig. 3.5).

So what is the object actually doing as a result of the torque applied, i.e., what is its rotational rate and rotational acceleration about its center of mass? At this point we can't answer that question because we haven't said anything about what the object looks like, or how it's put together, or how those issues relate to the problem. Why should those issues matter? Because now we're asking what is the motion of all the component pieces of the object (if we imagine decomposing the object into a great many bite-sized point-masses held together by mathematical constraints), not simply what is the motion of the object's center of mass, which is a kind of average of all the

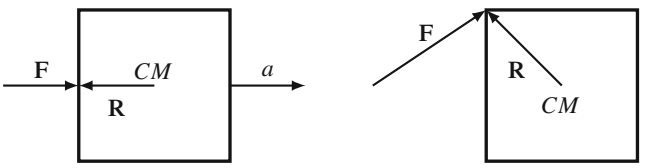


Fig. 3.5 Forces through the Center-of-Mass (C.M.) versus forces offset from the C.M

component pieces. If this was a problem in translational motion, all we'd have to do is take an average of the translational momenta of all the point masses and we could know (at least empirically) what the object's center of mass was doing. And since the center of mass sort of carries the object along with it (ignoring complications like flexible dynamics), we would therefore know what the object itself was doing. We even can describe the rotational motion associated with an object in orbit (provided we can approximate the object as a point mass) by, for example, relating angular velocity in a circular orbit to instantaneous translational velocity as discussed earlier in this chapter. But what if we can't approximate the object by a point mass? Clearly we need to introduce some new quantity that describes the intrinsic nature of the object itself.

3.5 How the Moment of Inertia Tensor Describes the Object's Nature

In the real world, objects are not idealized point masses; they have size, shape, and structure as well. (Maybe that's the three esses of satellites, kind of like the four cees of diamonds, namely carat, color, clarity, and cut. Actually five cees if you count cost, a more important quality for most folks.) So we need some mathematical construct that also expresses those features. As a first step, let's consider again what purpose the mass term in the translational motion equations serves. Comparing translational velocity to translational momentum, multiplication by the object's mass (or more formally, the object's inertial mass) in effect transforms a measure of the object's current motion into a measure of what it takes to change the object's motion. So at an abstracted level, what we now need to define for rotational motion about an object's center of mass is something (put technically, an operator) that will transform the object's angular velocity about its center of mass into its associated angular momentum. Just from a units analysis, it's clear this operator will be much more complicated than a scalar mass multiplication since its units will be the product of mass times the square of a distance. Also, although for the simple circular orbit motions we've been talking about the two vectors are parallel, in general (as we'll see later in this chapter) the direction of an object's angular velocity and angular momentum can be different. So the operator connecting these two physical quantities must at least have the mathematical complexity of a 3×3 linear transformation matrix, a subset of which is the rotation matrices we encountered in Chap. 1. Suppose then we try as a candidate $\mathbf{L} = \boldsymbol{I}\boldsymbol{\omega}$, where $\boldsymbol{I}$ is the desired operator (which we'll call the moment of inertia tensor), and the vector natures of $\mathbf{L}$ and $\boldsymbol{\omega}$ are again indicated via "bold" font.

At this point, some of you might be saying "OK, the word 'moment' kind of invokes 'moment arm', which involves a twisting around some point (appropriate for the topic), and 'inertia' is related to mass and the ease or difficulty of changing an object's motion, but what is a 'tensor'?" Good question. Tensor is a piece of

geometric jargon. It basically refers to a geometrical entity that has a fundamental essence independent of any reference frame, can be expressed in component form relative to any particular reference frame, and can be transformed between reference frames by well defined rules. A position vector, **R**, is an example of a tensor. The moment of inertia tensor is a somewhat more complicated example. Just as **R** is well-defined mathematically independent of any particular reference frame, we will find that $\boldsymbol{I}$ is similarly intrinsically well-defined independent of specific reference frames.

Now we have a name for the operator, and we've made a guess at its structure and units. But what does it really look like? Sometimes when you're hopelessly confused by a problem, and can't think of anyway to dump it in someone else's lap, it helps to simplify the problem as much as possible. So let's consider a very simple object, a hoop of mass M kilograms with radius R meters having negligible thickness, and let's also make the geometry as simple as possible by locating the coordinate system at the hoop's center of mass, with the z-axis perpendicular to the hoop plane. This is a simple extension of Fig. 3.3, used for the definition of angular momentum of a point-mass relative to some coordinate origin. Imagine extending that single mass M over the whole circular orbit there discussed and we've got our hoop. Now, in Sect. 3.3, we explained that **L** was a constant of the satellite's motion about orbit center, independent of specifically where the satellite was at any given time. Smearing M uniformly over the hoop (effectively making many little satellites out of one big one) therefore doesn't change **L** if all of the little satellites are moving at the same angular rate about the origin. For our hoop, with rotation about the z-axis, **L** is therefore given by

$$\mathbf{L} = \mathbf{R} \times \mathbf{P} = \mathbf{R} \times (M\mathbf{V}) = \omega M R^2 \, [0, 0, 1]^T \tag{3.1}$$

where ω is the magnitude of the angular velocity about Z. This implies that for the hoop, in the reference frame with the z-axis perpendicular to the hoop, the zz component of $\boldsymbol{I}$ (i.e., I_{zz}) is mR^2. We don't care at this point what I_{xx} and I_{yy} are, because they can't contribute anything to the angular momentum since the rotation is exclusively about the z-axis. And for now we're ignoring the cross terms (I_{xy}, I_{xz}, I_{yx}, I_{yz}, I_{zx}, and I_{zy}) because this solution for the tensor $\boldsymbol{I}$ gives us the "right" answer. But it's important to recognize that we haven't as yet proven that the cross terms actually are equal to zero for this problem.

So, how do we determine the components I_{xx} and I_{yy}, which we would need if the hoop were spinning about an axis in the plane of the hoop? From symmetry, it should be clear that I_{xx} and I_{yy}, each characterizing rotation about a particular axis in the plane of the hoop, are equal to each other and independent of their actual orientation in the plane. So let's consider rotation about the x-axis. We'll need to explicitly consider the contribution of each mass point individually, since they are at different distances from the axis of rotation. Each small element of the hoop has mass

$$dm = M(d\theta/2\pi)$$

where θ is an angular measure around the hoop, which we can arbitrarily take as measured from the x-axis. Each mass element will be "in orbit" about a fixed point on the x-axis (although different points for different mass elements). A mass element at angular position θ around the hoop will be a distance $R\sin(\theta)$ from the rotation axis. As for the problem of rotation about the z-axis, the angular momentum contribution of each element about the axis of rotation is a constant of the motion. The differential contribution to angular momentum about the x-axis from each mass element is therefore given by:

$$d\mathbf{L} = \omega dm(R\sin(\theta))^2\,[1,0,0]^T$$

where ω is now the magnitude of angular momentum about the x-axis. To get the full value for $\mathbf{L}$, we integrate around the hoop:

$$\begin{aligned}\mathbf{L} &= \int_0^{2\pi} \omega dm(R\sin(\theta))^2\,[1,0,0]^T \\ &= (\omega MR^2/2\pi)(\int_0^{2\pi} \sin^2(\theta)d\theta)\,[1,0,0]^T \\ &= (\omega MR^2/2\pi)(\pi)\,[1,0,0]^T \\ &= (\omega MR^2/2)\,[1,0,0]^T\end{aligned}$$

From this we see that, for the hoop, $I_{xx} = I_{yy} = I_{zz}/2$. This makes qualitative sense; the mean distance of the hoop's mass from the X and Y axes is less than from the z-axis, so for a given angular rate, angular momentum about the x- or y-axis is less than that about the z-axis.

That's fine for this very special problem we've constructed, but obviously most spacecraft are solid, not hollow (let alone a wire hoop), so we're going to need something a bit more general. But we now know that if we're dealing with a set of point masses (or a continuous distribution of differential mass elements), the contribution from a single mass element is easy to compute if we know its mass and its distance from the axis of rotation. We also know that if we can decompose the object into a set of point masses, we can get the moment of inertia (or at least the diagonal elements thereof) for the whole object if we can figure out how to sum together the contributions from each of the individual point masses. In other words, the key to computing the moment of inertia of a complex object is performing an integration over infinitesimal components of the object. The result is:

$$I_{xx} = \int_0^M (y^2 + z^2)dm \tag{3.2}$$

$$I_{yy} = \int_0^M (x^2 + z^2)dm \tag{3.3}$$

$$I_{zz} = \int_0^M (x^2 + y^2)dm \tag{3.4}$$

OK - so now you know how to compute the diagonal elements of $\boldsymbol{I}$ in some reference frame. We still have six more terms that need to be calculated, namely the cross terms (which are called the *products of inertia*) that we intentionally ignored earlier. The key distinction between the diagonal and cross terms lies in the definition of the object's body frame axes. Although we pretty much have the freedom to define the object's coordinate frame however we like, so long as the axes of the frame span a three-dimensional space, it's usually the case that some definitions will be more convenient than others. In other words, some reference frame selections will make solving physics problems easier than others. First, we know from experience that orthogonal frames are easier to work with than non-orthogonal frames. (Our preceding discussion of the diagonal elements of $\boldsymbol{I}$ assumed an orthogonal frame.) Second, instinctively one would think that "natural" body frame coordinate axes "should" line up with the axes of the symmetry of the object, e.g., one body axis for a cylinder "should" be its long central axis, while the natural axes of a rectilinear solid should be parallel to its sides. Third, we should place the origin of the frame at the object's center-of-mass since the problems we're examining involve motions about the object's center-of-mass.

It turns out that all objects have natural axes of symmetry, even those objects for which such symmetry is not immediately apparent. The implication is that when $\boldsymbol{I}$ is computed in said "natural" reference frame, then you will find that its off-diagonal elements are all zero. Following this natural reference frame definition physically aligns the axes of the coordinate system with the symmetry axes of the object, also called the object's *principal axes*. What's physically special about these axes is that if you start the object spinning about a principal axis, and there are no dissipative torques (see Sect. 3.7), the object will continue spinning about that axis (i.e., the object is balanced) at a constant rate. That in turn tells us that for this geometry there is a special relationship between the object's angular velocity and angular momentum. Here's why. Going back to translational motion, we know that in the absence of external forces, translational momentum will be conserved, i.e., the translational momentum is constant, as is the translational velocity since translational momentum is proportional to velocity. Since (in the absence of external torques) angular momentum is also conserved, and since the direction of the angular velocity in this scenario is constant, it follows that the direction of the angular momentum vector is parallel to the angular velocity vector, i.e., the angular momentum is proportional to the angular velocity if the object is rotating about a principal axis. What's different for rotational motion compared to translational motion is that the constant of proportionality for translational motion is the same (the object's mass) regardless what direction the object is moving. But for rotational motion about the center of mass, the constant of proportionality will depend on the structure of the object, and therefore need not (and typically, will not) be the same for rotations about the three principal axes. What this means mathematically is that the moment of inertia tensor (for this scenario) will be diagonal, but its diagonal terms need not, and probably will not, be the same.

As another example, consider an idealized dumbbell (two equal masses of mass *M/2* connected by a massless rod) rotating about an axis perpendicular to the rod

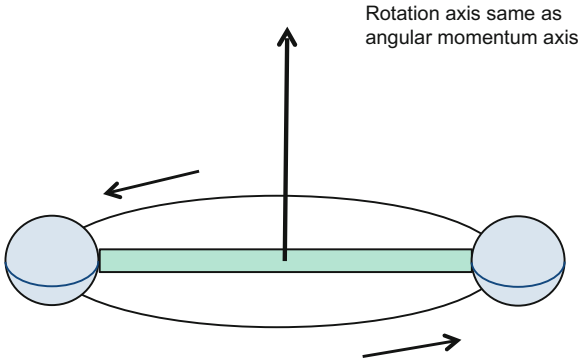


Fig. 3.6 Rotation about a principal axis

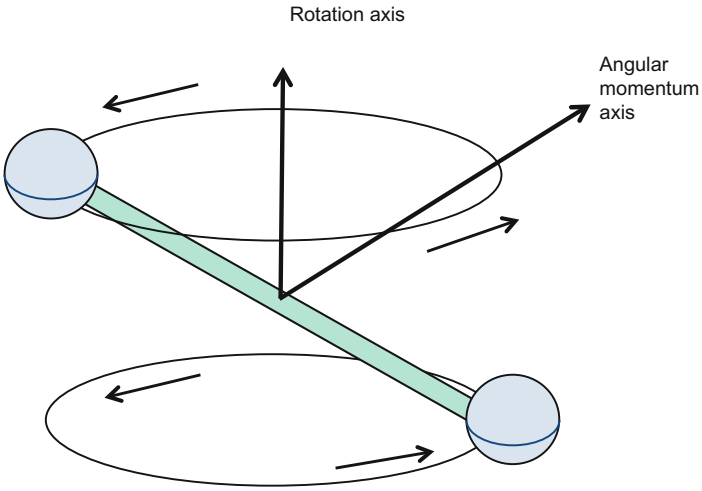


Fig. 3.7 Rotation about a non-principal axis

(see Fig. 3.6). By symmetry, the moment of inertia about any axis perpendicular to the rod is the same as it would be for the hoop previously considered.

So, what if an object (say, our dumbbell of the previous paragraph) is rotating about an axis other than one of its principal axes (see Fig. 3.7)? In this case, we can project the angular velocity vector onto the principal axes and determine the components of the vector relative to the principal axis reference frame. But the motion of the object will no longer be as simple as in the case of motion about a principal axis unless the diagonal terms of the moment of inertia tensor are the same (again, typically not the case). That's because the angular momentum vector will no longer be proportional to the angular velocity vector. Since in the absence of external torques angular momentum is conserved (requiring that its direction be constant in an inertial frame), and since the angular velocity vector will not be in the same direction as the

angular momentum vector if the diagonal elements of $\boldsymbol{I}$ are not all equal, it follows that in an inertial reference frame in the absence of external torques the angular velocity vector must rotate about the angular momentum vector. Alternatively, for the angular velocity vector to maintain a constant direction in an inertial frame, an external torque must be applied to the system, causing the angular momentum vector to rotate about the angular velocity vector (again, see Fig. 3.7).

At this point you're probably saying, wait a minute, what happened to those off-diagonal terms you were supposed to be talking about? Well, having non-zero off-diagonal terms means that the body frame we have selected is not coincident with the object's principal axes, and as a result the components of the angular momentum vector will not be proportional to their corresponding angular velocity vector components. For example, the z-component of the angular momentum vector may depend on the values of all three components of the angular velocity vector. Let's consider what the simplest functional form would be for coefficients multiplying the x and y angular velocity components. Remembering that to compute the moment of inertia tensor elements we have to sum over the volume of the object, it's clear we need a functional form (inside the mass integral) for the off-diagonal elements that will vanish if the z-axis is an axis of symmetry. We can guarantee that to be true if the positional dependence of the off-diagonal elements (inside the mass integral) is odd (so, e.g., mass at negative x positions cancel against mass at positive x positions in the computation of I_{xz}), as opposed to even as was the case for the diagonal terms. From a units analysis, we also know that the off-diagonal terms are of order 2 in position. That implies that the most natural form of the off-diagonal terms (inside the mass integral) would be some combination of things like xy, xz, and yz.

Next, suppose we introduce a small misalignment of the axes, say via a tiny positive rotation of our selected body reference frame about the y-axis of the symmetry axes frame. The body frame mass distribution will still be symmetrically distributed relative to the y-axis, so the y-component of the angular momentum vector will still be proportional to the angular velocity vector. However the z-component of the angular momentum will no longer exclusively depend on the z-component of angular velocity; it should pick up a small term proportional to the x-component of the angular velocity as well. Since the term would go to zero if there was any odd y-dependence in it, it's a reasonable guess that the positional term (inside the mass integral) associated with x-axis angular velocity contribution to the z-axis component of the angular momentum should be xz. This argument also suggests that I_{xz} should be equal to I_{zx}, i.e., the moment of inertia tensor is symmetric.

Both these suppositions indeed are the case, and can be demonstrated mathematically, but the exact form of the off-diagonal elements, unlike the case of the form of the diagonal terms, is not easy to show by simple verbal argument. At this point, though, the math shouldn't be too intimidating. Recall the fundamental equation for angular momentum of an individual mass element, dm:

$$d\mathbf{L} = \mathbf{R} \times (\mathbf{V}dm) = \mathbf{R} \times (d\mathbf{R}/dt)dm \tag{3.5}$$

where $\mathbf{V}dm$ is the momentum of the element.

Let's suppose that we're working in an inertial frame instantaneously aligned with the body coordinate frame (with the origin at the center of mass of the object), and that in said frame the body is rotating at an instantaneous angular rate $\boldsymbol{\omega}$. In that case, $d\mathbf{R}/dt = \boldsymbol{\omega} \times \mathbf{R}$, which allows us to rewrite Eq. 3.5 as

$$d\mathbf{L} = \mathbf{R} \times (\boldsymbol{\omega} \times \mathbf{R})dm \tag{3.6}$$

So, now integrate:

$$\mathbf{L} = \int \mathbf{R} \times (\boldsymbol{\omega} \times \mathbf{R})dm \tag{3.7}$$

You can see that the result is linear in $\boldsymbol{\omega}$, which means that we should be able somehow to pull $\boldsymbol{\omega}$ out to the right side, leaving the rest of the integral as a 3×3 matrix. The required approach is fairly straight-forward algebra. First expand the cross product $(\boldsymbol{\omega} \times \mathbf{R})$ into a vector in terms of the components of $\boldsymbol{\omega}$ and $\mathbf{R}$, and then do the same with $\mathbf{R} \times (\boldsymbol{\omega} \times \mathbf{R})$. We'll leave that bit as an exercise for the reader. With the parts expressed in terms of the vector components, you can reconstruct the total as $\boldsymbol{I}\boldsymbol{\omega}$, with the diagonal terms given via Eqs. 3.2–3.4, and the off diagonal terms via:

$$I_{xy} = I_{yx} = -\int xydm \tag{3.8a}$$

$$I_{xz} = I_{zx} = -\int xzdm \tag{3.8b}$$

$$I_{yz} = I_{zy} = -\int yzdm \tag{3.8c}$$

With the full form of $\boldsymbol{I}$ known, we'd like you to now recall that we mentioned that $\boldsymbol{I}$ is a tensor, i.e., a geometric object with intrinsic meaning independent of reference frames, and a specific rule for its transformation between frames. This implies that if you happen to know $\boldsymbol{I}$ in any arbitrary reference frame, you can compute it in all other frames. If $\boldsymbol{R}_{B<A}$ is the rotation matrix that transforms vectors from frame A to some other frame B (i.e., vectors transform between frames as $\boldsymbol{V}_B = \boldsymbol{R}_{B<A}\boldsymbol{V}_A$), then $\boldsymbol{I}$ is transformed as:

$$\boldsymbol{I}_B = \boldsymbol{R}_{B<A}\boldsymbol{I}_A\boldsymbol{R}_{B<A}^{-1} = \boldsymbol{R}_{B<A}\boldsymbol{I}_A\boldsymbol{R}_{B<A}^{T} \tag{3.9}$$

In particular, this will be true if frame A is defined by the object's principal axes (in which the off-diagonal elements are zero) and frame B is some other frame (say, the official satellite coordinate frame). The off-diagonal elements of $\boldsymbol{I}_B$ capture information regarding the rotation needed to express $\boldsymbol{I}$ as a diagonal (principal axis) matrix.

There's one last thing to mention regarding the moment of inertia tensor and rigid body motion. For clarity of explanation, we've decomposed an object's motion

into two major pieces, namely translations and rotations of the object's center-of-mass and rotations about the object's center-of-mass. Although it's nice to treat each of these phenomena separately when trying to understand what a rigid object can do, when it comes time to describe quantitatively and accurately what the object is actually doing we need to be able to combine all these phenomena together, because ultimately the object couldn't care less about our human desire to compartmentalize. In practice, the object simply responds physically to the aggregate of forces and torques applied to it. Specifically, since angular momentum is a vector, we know that the total angular momentum of a solid rigid body will just be the sum of the angular momentum of the center-of-mass about our reference frame origin (where we imagine concentrating the mass of the object at a single point, the center-of-mass) plus the angular momentum of the body about its center-of-mass.

3.6 Types of Torque-Free Rotational Motion

Now that we've finally defined all the physical quantities describing rotational motion (compare Tables 3.1 and 3.2 for rotational analogs to translational quantities), let's talk about the different ways an object can rotate relative to its center of mass. In this section we consider torque-free motion. The simplest type of such angular motion is called pure rotation (see Fig. 3.8a). For this type of motion, the rotation axis is aligned with a body principal axis, which also happens to be one of the body-frame reference axes. The first condition implies that the angular momentum vector will be aligned with the rotation axis and will remain that way as long as no torques are applied. The second condition is just a convenience for the users, who were clever enough to have aligned their reference axis with a principal axis. (Note, if the mass distribution is such that the moment of inertia matrix is proportional to the 3×3 identity matrix, then all axes are effectively principal axes and any rotation would be pure rotation as defined here.)

The next level of complexity is to allow the body axes to be misaligned relative to the principal axes. This misalignment will not change the physical motion of the object; if the rotation axis is lined up with a principal axis, the angular momentum vector will still be proportional to the angular velocity. However, the body axis corresponding to the rotation axis will now trace out a conical shape about the object's rotation axis (see Fig. 3.8b). This behavior is, not surprisingly, called *coning*. Coning should not be viewed as an entirely artificial mathematical quirk that can easily be removed just by changing body axes. As an example of why it matters, you typically place sensors to "look" in "natural" directions relative to a spacecraft's body axes. If the body axes are then badly misaligned relative to the principal axes of a spinning spacecraft, a key attitude sensor needed for determination of spacecraft rotation rate may fail to generate a sufficiently accurate value until the misalignment is calibrated and compensated for.

The next level of complexity in rotational motion about an object's center of mass is called *nutation* (not to be confused with the same term referring to the wobble

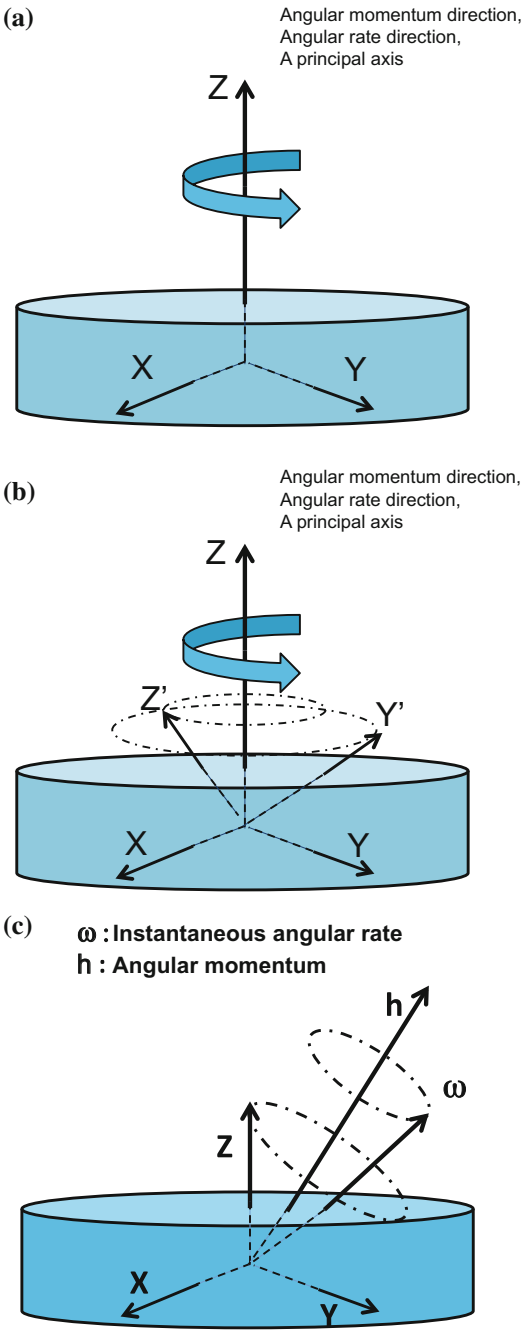


Fig. 3.8 **a** Rotation types: pure rotation. **b** Rotation types: coning of non-principal axis reference frame (X', Y', Z') about rotation axis Z. **c** Rotation types: nutation

of a spinning top). Nutation is present when the instantaneous rotation axis is not aligned with a principal axis. Assuming no external perturbative torques are acting on the object, angular momentum will be conserved, so the direction of the angular momentum vector will remain fixed in inertial space. However, the instantaneous angular momentum vector no longer is proportional to the angular velocity vector, nor are either of them lined up with a principal axis. Viewed from an inertial reference frame, nutation then causes the angular velocity vector and the principal axes to rotate about the angular momentum vector, as illustrated in Fig. 3.8c. The figure is actually drawn for the special case of $I_{xx} = I_{yy} < I_{zz}$, which results in the angular velocity vector, $\boldsymbol{\omega}$, the principal axis $\mathbf{Z}$, and the angular momentum vector, $\mathbf{h}$, being coplanar, with $\mathbf{h}$ directed between $\boldsymbol{\omega}$ and $\mathbf{Z}$. If $I_{xx} = I_{yy} > I_{zz}$, the three vectors would still be coplanar, but $\boldsymbol{\omega}$ would be directed between $\mathbf{h}$ and $\mathbf{Z}$ (see, e.g., Wertz[1]). If we also have misalignments between the body axes and principal axes, the object can experience both coning and nutation simultaneously.

Let's look at how all this applies to a nice symmetrical object like a cylinder. First, let's make the origin of the coordinate system the center of mass of the object and allow no external torques to be present (i.e., angular momentum is conserved). Since we're also going to work with an inertial reference frame, the angular momentum vector will be fixed in our frame. For convenience, we'll make it the inertial reference frame's z-axis. Next, we'll select the body axes so that they line up perfectly with the principal axes in order to make the moment of inertia tensor diagonal. Because of the cylindrical symmetry, selection of the x-axis is arbitrary, but after that selection the y-axis is fixed by the orthogonality condition on the axes. In addition, for a cylinder, two of the principal moments will be identical. Specifically, if the body frame's z-axis lines up with the cylinder's symmetry axis, the moments I_{xx} and I_{yy} will be equal (again, see Fig. 3.8a; we're just elaborating on the same scenario).

Now that we've defined a reference frame in which something can happen, let's put something in the frame that's doing something. Suppose our cylinder is nutating. How can we decompose that rather complex motion in order to understand it better? Well, if the motion was a pure rotation, we'd only be concerned with motions of body frame vectors about a principal axis (say, the z-axis). But when the object is nutating, what we have instead is a superposition of rotation about the principal axis with rotation of the principal axis about the one fixed vector in our reference frame, the angular momentum vector. That means that the instantaneous rotation vector will be a vector sum of these two rotations, which also means that the three vectors (the instantaneous rotation vector, the principal axis vector, and angular momentum vector) will be coplanar (as previously mentioned), and both the principal axis and instantaneous rotation vector will rotate about the fixed angular momentum vector in the inertial reference frame. The angle defining the cone on which the principal axis rotates about the angular momentum vector is called the *nutation angle*, and will be constant. Similarly, the angle defining the cone on which the instantaneous angular velocity vector rotates about the angular momentum vector is called the *space cone*

[1] J. Wertz, Spacecraft Attitude Determination and Control, pp. 489–492.

angle, and also will be constant. If we now imagine ourselves in the non-inertial body frame defined by the principal axes, both the angular momentum vector and instantaneous angular velocity vector will appear to rotate about the principal z-axis. The angle defining the cone on which the instantaneous angular velocity vector appears to rotate about the principal axis is called the *body cone*.

Abstractly, this all is pretty straightforward based on the coplanarity of the three vectors, but envisioning what it looks like physically can be rather difficult. Suffice it to say for now that although the spacecraft may be undergoing an apparently chaotic tumble, there is in fact method in the madness. (See, e.g., Wertz[2] for more detail.)

3.7 How Torques Can Influence an Object's Rotational Motion

All our discussions of physical problems so far have dealt with torque-free motion. But in the real world of spacecraft applications, torques (both intentional and unintentional) are always acting on the spacecraft. The environment in which the spacecraft "lives" exerts on it a wide variety of perturbative torques generically referred to as disturbance torques. To counteract these natural effects, a spacecraft will initiate its own control torques typically via reaction wheels, thrusters, or magnetic torquer bars. Torques come in two other flavors, external and internal. External torques actually change the spacecraft's total angular momentum, for good or ill. So if you want to nullify fully the effects of a disturbance torque, you have to apply a control torque with an attitude actuator like a magnetic torquer bar (MTB) that interacts with the outside, or thrusters that cause part of the spacecraft itself (i.e., thruster fuel) to be cast off along with unwanted angular momentum. By contrast, internal torques may change the angular momentum of a spacecraft component, but do not change the spacecraft's total angular momentum. For example, if you speed up a reaction wheel (thereby increasing the wheel's angular momentum), the spacecraft body's angular rate will change (thereby changing its angular momentum) such that the sum of the wheel's and spacecraft's angular momenta will remain constant. From a practical applications standpoint, typically both types of control torques are utilized onboard. As reaction wheels provide very precise control and are at least moderately powerful, wheels may initially be used to null the effects of disturbance torques. But over time, eventually the wheels may saturate (i.e., reach their maximum allowed speeds), so an external torque attitude actuator is needed to unload excess wheel momenta so the reaction wheels can continue to perform their control function. For this purpose, weak MTBs can be used continuously to keep the wheels far from their saturation limits at all times, or powerful thrusters can be used intermittently to dump wheel momenta whenever their speeds approach saturation levels. We'll discuss wheels, MTBs and thrusters in much more detail in Chap. 5.

[2] ibid.

Torques can affect the spacecraft's angular momentum in either or both of two ways, by changing the angular momentum vector's direction or changing its magnitude. For example, for an object rotating about one of its principal axes, applying a torque perpendicular to the direction of the object's angular momentum vector will cause the momentum vector to precess (like a tilted spinning top precessing, or the Earth's spin axis precessing). A familiar example of this behavior is the precession of a top's spin axis as a response to the torque exerted on it by gravity. If the torque is weak, a slow precession results, which is also called drift. By contrast, a torque applied parallel to the direction of the angular momentum vector will leave its direction unchanged but will instead increase or decrease its magnitude.

As is the case for all vectors, the directions of the angular momentum and torque vectors will change depending on the reference frame selected. For example, although the angular momentum vector will, in the absence of external torques, be constant in both direction and magnitude in an inertial reference frame, the components of the angular momentum vector usually will be time-dependent quantities as seen in the spacecraft body frame, even though the magnitude of the vector will be conserved. Another important conserved quantity (in the absence of external torques and dissipative effects) is the *rotational kinetic energy*. Rotational kinetic energy is the analog to translational kinetic energy. Recall that translational kinetic energy is equal to $m\mathbf{v}^2/2$ (where m is the mass of the object and $\mathbf{v}$ is its translational velocity), which also can be expressed as $\mathbf{p}^2/2m$ (where $\mathbf{p}$ is the object's translational momentum). It should then come as no surprise that the rotational kinetic energy is given by $(L_x^2/I_{xx} + L_y^2/I_{yy} + L_z^2/I_{zz})/2$, where $\mathbf{L} = (L_x, L_y, L_z)$ is the angular momentum in the body frame (where we'll assume the body frame and principal axis frame are coincident) and I_{xx}, I_{yy}, and I_{zz} are the object's principal moments of inertia (i.e., the diagonal elements of its moment of inertia tensor in a body frame where the body axes coincide with the principal axes). Since by conservation of momentum we know that the magnitude of the angular momentum vector (given by $L_x^2 + L_y^2 + L_z^2$) is also a constant, we have two constraints on allowed values of the components of the angular momentum vector in the body frame. Expressed geometrically, conservation of rotational kinetic energy defines the surface of an ellipsoid on which the components of the angular momentum vector in the body frame must lie. Similarly, conservation of angular momentum defines the surface of a sphere on which the components of the angular momentum vector in the body frame must lie. Imposing both constraints, then, is the same as saying that in the absence of external torques the components of the object's angular momentum vector, as seen in the body frame, must lie on the intersection of the energy ellipsoid with the angular momentum sphere.

For simplicity of analysis, let's assume the object once again is cylindrically symmetric, with the axis of symmetry lying along the z-axis (i.e., $I_{xx} = I_{yy}$). Then the intersection of the sphere and ellipsoid will consist of two parallel circles in body frame angular momentum space (see Figs. 3.9a, b). In other words, the angular momentum vector in the body frame rotates about one of the two circles at a constant rate with the z-component of the angular momentum vector held constant. This is equivalent to saying that in the inertial frame the axis of rotation (the z-axis of the

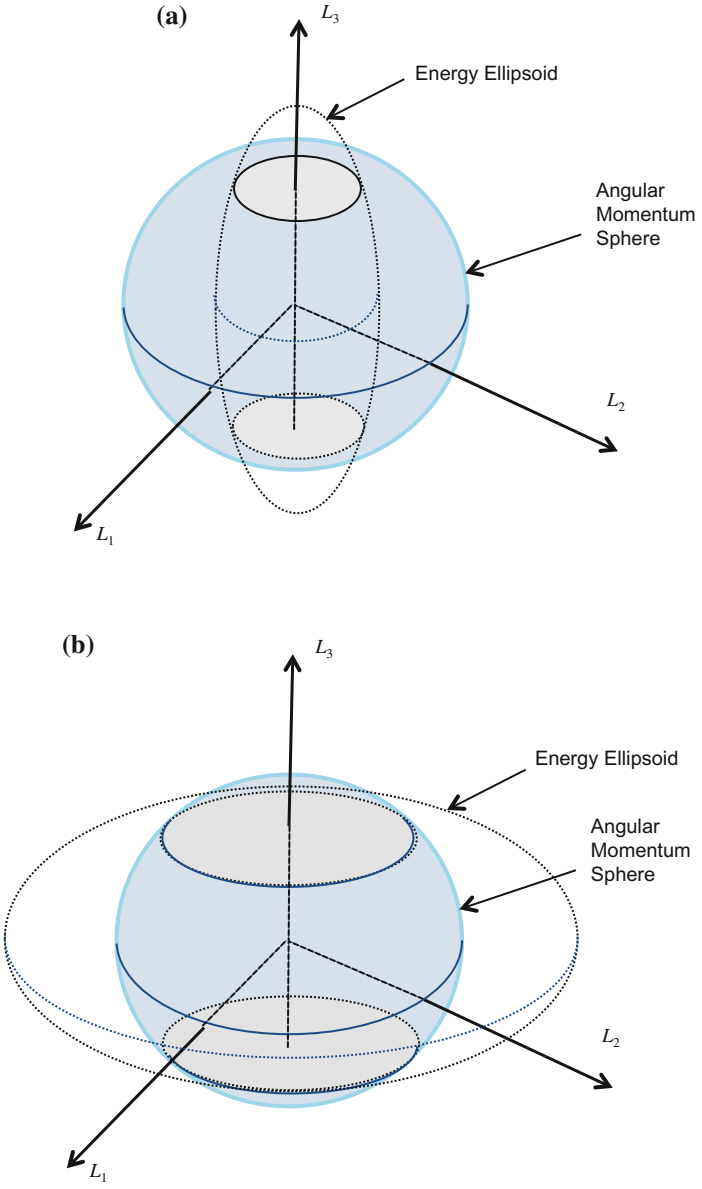


Fig. 3.9 **a** Intersection of the energy ellipsoid and angular momentum sphere when the largest principle moment of inertia is not along the z-axis. **b** Intersection of the energy ellipsoid and angular momentum sphere when the largest principle moment of inertia is along the z-axis

body frame) rotates about the angular momentum vector which is fixed, the nutation effect we discussed earlier.

Now let's introduce dissipative effects that remove rotational energy from the system, e.g., frictional effects from liquid sloshing in a spacecraft nutation damper. How will that affect the object's angular momentum? Since a dissipative effect by its nature has no preferred axis (otherwise we'd be dealing with a systematic effect), its influence on the object will be to shrink down the rotational kinetic energy ellipsoid without changing the ellipsoid's shape. By contrast, since no external torques are present, the angular momentum sphere will retain both its size and shape. This means that while the ellipsoid shrinks under the influence of the dissipative effects, the angular momentum vector, for the case where I_{zz} is the largest of the principal moments, will steadily spiral into smaller and smaller circles, eventually becoming parallel to the z-axis. For the case where I_{zz} is the smallest of the principal moments, the two circles draw together until they coincide along the x-y plane and the angular momentum vector remains fixed on a point on the circle with zero z-component. Once either of these states is reached, the angular momentum vector is now fixed in the body frame as well as the inertial frame, so no nutation is present any longer and the body simply rotates about a fixed principal axis. Energy dissipation also ends at this point.

If there is no symmetry axis (for example, $I_{xx} < I_{yy} < I_{zz}$), when energy dissipation ceases, the angular momentum vector will become aligned with the principal axis having the largest principal moment since for this geometry the rotational kinetic energy will be minimized. So if a spacecraft initially is spinning about the principal axis having the largest principal moment and nutation is introduced as a perturbation (meaning a small effect relative to the primary motion about the principal axis), any frictional effects present will cause the nutation to be damped out with time, hence the utility of a nutation damper filled with sloshing liquid for a spacecraft spinning about its major principal axis. On the other hand, if the spacecraft spin axis is a minor principal axis, energy dissipation will introduce motions about the major axis resulting in an increase in nutation. If continued unabated, the energy dissipation will eventually "suck out" all the rotation about the minor axis while pumping rotation into the major axis, producing a pure rotation about an axis perpendicular to what had been the nominal spin axis, a condition referred to as flat spin. This actually happened on orbit to Explorer 1, the first American satellite, back in February 1958. Explorer 1 was a pencil-shaped spacecraft (see Fig. 3.10) that was spin-stabilized about the principal axis associated with its smallest principal moment. However, vibrational motions built up in the spacecraft antennas that drove the energy dissipation effects we've been talking about, leading to a condition of flat spin and loss of the spacecraft.

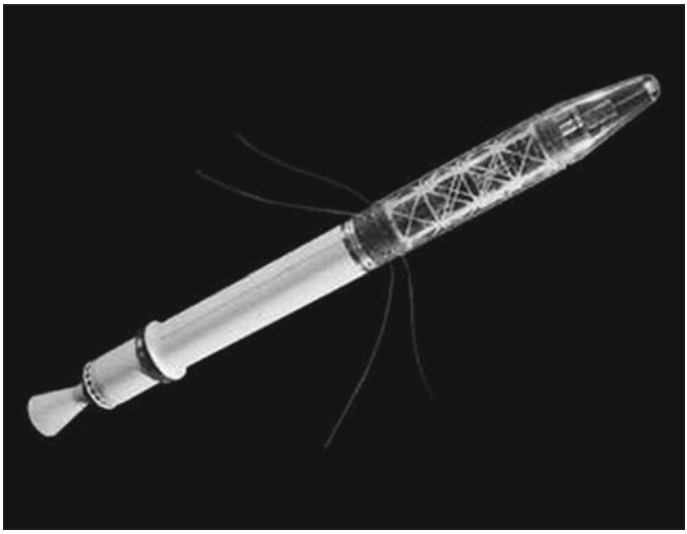


Fig. 3.10 Explorer 1 spacecraft (see https://www.nasa.gov/mission_pages/explorer/explorer-overview.html). Credits: NASA

3.8 Attitude Control Torques

Although many of the first spacecraft relied on spin-stabilization to provide passive attitude control, nowadays most spacecraft utilize active control attitude actuators to maintain their desired orientation, or to reorient the spacecraft to a new pointing. In the next couple of chapters we'll provide a lot of descriptive material regarding attitude actuators, and the attitude sensors that supply the measurements that are processed onboard to determine how the actuators should be commanded. For now though we'll just spend a little time discussing the nature of the control torques generated by the attitude actuators, as well as the environmental disturbance torques for which the control torques must compensate.

An attitude control torque is usually produced by one of three actuator types: thrusters, reaction wheels, or magnetic torque bars. Thrusters are capable of supplying the quickest and most powerful torques, although recently considerable effort has been made to develop very precise low force (micronewton) thrusters to support special high accuracy pointing requirements. Thrusters can supply attitude control torques in any orbit, albeit with some significant negatives (e.g., finite lifetime due to finite fuel supply, and for hydrazine-based thrusters a corrosive influence on sensitive optics). Ideally, to cover both polarities of the 3 degrees of freedom associated with a general rotation, a minimum of 6 attitude thrusters are required for three-axis stabilized spacecraft. For spin-stabilized spacecraft, you only need two thrusters to control the spin rate (spin-up or spin-down) and two more thrusters to precess the spin axis (increasing Sun angle or decreasing Sun angle). However, most missions

require some redundancy in case of hardware failure, so actual spacecraft may fly as many as twice the number of attitude thrusters listed above, although in many cases the redundancy is achieved by using combinations of orbit thrusters that provide the needed torque about the spacecraft center of mass without introducing excessive forces on the center of mass, thereby not significantly perturbing the orbit. By the same token, the attitude thrusters themselves, due to thrust centroid misalignments, may introduce forces on the spacecraft center of mass that can perturb the orbit.

Unlike thrusters, wheels don't control spacecraft momentum by casting off material that carries some net angular momentum (relative to the spacecraft center of mass) with it; rather, they "merely" redistribute angular momentum between parts of the spacecraft. They tend to be weaker than thrusters, with typical maximum wheel torques ranging from 0.01 to 1 newton-meters.[3] Unlike thrusters, wheels don't push against or interact with the world outside the spacecraft. So the torques that wheels generate are internal ones that cannot change the spacecraft's total angular momentum. Instead, they orchestrate an exchange of angular momentum between the wheels and the spacecraft body. For example, if a pitch wheel spins up positively, a previously "stationary" spacecraft will rotate in the negative pitch direction so as to keep the total angular momentum constant. Of course the spacecraft's spin rate will be much smaller, reflected by the ratio of the wheel moment of inertia and the spacecraft's pitch-axis moment of inertia. The major negative regarding use of wheels for attitude control is that they can saturate. In other words, if you keep commanding torques to the wheel to compensate for environmental torques acting in about the same direction (a secular torque term), eventually the wheel will reach its maximum spin rate and spacecraft attitude control may be lost. So use of wheels as the primary mechanism for attitude control requires employment of a secondary system (thrusters or MTBs) to allow angular momentum to be dumped from the wheels before their spin rates get too high. For a three-axis stabilized spacecraft, covering the three degrees of freedom associated with a general rotation requires employment of a minimum of 3 wheels. In practice, most spacecraft of this type fly 4 wheels, which enables a wider distribution of angular momentum (ameliorating somewhat the saturation problem) and also provides redundancy if a wheel fails. By contrast, if attitude accuracy is not terribly tight, it is possible with spinning spacecraft (e.g., Earth-pointing spacecraft with 1 revolution per orbit (RPO) motion) to use only 2 wheels. One wheel supplies pitch control to enable the 1 RPO motion while the second wheel provides roll and yaw control, taking advantage of quarter orbit roll-yaw coupling (the interchange of roll and yaw attitude errors as viewed in the body frame). For spacecraft designed with this strategy, the roll and yaw errors oscillate over a quarter of an orbit but stay within some bounded magnitude.

MTBs tend to be the weakest and most limited control torque sources. MTBs require a significant geomagnetic field with which to interact. They normally only are usable below synchronous altitude (35,000 km) because of the altitude dependent drop-off in the geomagnetic field. Even at low Earth orbit (LEO), typical maximum

[3] Larson and Wertz, Space Mission Analysis and Design, p. 355, Table 11–11.

torques from magnetic coils range between about 0.0005– 0.2 newton-meters.[4] Their use also can degrade magnetometer accuracy, unless their perturbing influence is compensated for via a coupling matrix. Offsetting these negatives, the advantage of magnetic torquers is that they're operated solely with renewable electrical power, and therefore do not suffer from the lifetime limitations of attitude thrusters. As such, they are a very convenient means to provide continuous dumping of angular momentum from the wheels. They also are a very "patient" mechanism for nulling anomalously high body rates following spacecraft separation from the launch vehicle, or high rates induced by a stuck-on thruster.

3.9 Environmental Torques

Having mentioned the control torques that can be used to fix problems, let's spend the rest of the chapter discussing the environmental torques that can cause problems. Four major sources of external, environmental torques are gravity gradient, Solar radiation, aerodynamics, and flexible dynamics. Gravity gradient torque arises from variations in the gravitational force over a non-spherical object. An equation for calculating torque, approximating the Earth as a spherically symmetric body, is provided in Eq. 3.10.

$$\mathbf{N}_{GG} = (3\mu/|\mathbf{R}|^3)\mathbf{R} \times (\boldsymbol{I}\mathbf{R})/|\mathbf{R}|^2 \tag{3.10}$$

where

$\mathbf{N}_{GG}$ = gravity gradient torque (kg m^2/sec^2)
μ = the Earth's gravitational constant (GM_e) (m^3/sec^2)
$\boldsymbol{R}$ = the geocentric position vector of the object's body reference frame (note that the center of the frame need not be the center of mass) (m)
$|\mathbf{R}|$ = magnitude of the position vector (m)
$\boldsymbol{I}$ = the object's moment of inertia tensor (kg m^2)

Its general features of interest are that the torque is normal to local vertical (so it will cause rotational alignment with local vertical, not rotations around it), it is inversely proportional to the cube of the distance (so it drops off faster than the gravitational force on the center of mass), and it vanishes for a spherically symmetric object. For a fixed-pointing spacecraft where the duration of the pointing is comparable to or longer than the orbital period, gravity gradient torque will consist of two components, a periodic term and a secular term. The secular term is the one that introduces problems for wheel-controlled spacecraft as it will cause the wheel speeds to ramp upward with time, eventually leading to saturation unless angular momentum is dumped. The periodic term will simply cause the wheel speeds to oscillate up and down as the spacecraft follows its orbital path. Recall that Earth oblateness was most important near the Earth because the farther you got away from the Earth

[4] Larson and Wertz, Space Mission Analysis and Design, p. 355, Table 11–11.

the rounder it "looked" to the spacecraft; it also is the case that the farther you get away from the Earth, the more symmetric the spacecraft looks to the Earth and so the less important gravity gradient torque becomes. However, if you get too close to the Earth, the influence of atmospheric drag becomes more important. Trading the relative drop-offs of gravity gradient and atmospheric torques with altitude, gravity gradient's region of dominance is LEO above 400 Km altitude.[5]

Solar radiation torque also arises from asymmetries, but in its case the source is the displacement between the object's center of mass and its center of pressure. The effect of solar radiation pressure can be viewed as a force applied to the object's center of pressure. But because the center of pressure is determined by the object's cross-section displayed to the Sun, while the center of mass is determined by the object's mass structure independent of its orientation relative to the Sun, it can easily be the case that the two points will not coincide. That displacement will result in a torque about the object's center of mass, where the lever arm is given by the displacement between the center of mass and the center of pressure projected into local plane perpendicular to the Sun line.

The magnitude of the solar radiation pressure is driven by the vector difference between the translational momentum associated with the incident and reflected radiation flux. In other words, the amount of momentum per unit time that gets imparted to the object's cross-section defines the force being applied to the object by the solar radiation flux. The flux itself, as experienced by the spacecraft, is influenced by several effects. First, the intrinsic nature of the flux itself is determined by the intensity and spectral distribution of the light emitted by the Sun. Second, the effect of the flux will diminish as the square of the distance between the object and the Sun. This is easy to visualize when considered from an energy standpoint. If we approximate the Sun as a point object emitting light energy in a homogenous, spherical pattern, a given package of energy at a specified post-emission time will be distributed evenly over the surface of a sphere whose radius (r) is the time multiplied by the speed of light. Since the area of the sphere is $4\pi r^2$, the density of the energy distributed on the sphere will diminish with the square of the radius. So the flux felt by the spacecraft will also drop off with the square of the distance separation. A third influence is the geometry of the spacecraft's surface and its associated optical properties. Reflected light imparts twice the momentum to the surface it strikes as does absorbed light, so what the spacecraft is made of and how it's coated plays a large part in determining what influence a given amount of solar flux will have. Fourth, the orientation of the Sun vector relative to the spacecraft determines the size and shape of the cross-section being presented and plays a role in defining the displacement between the center of pressure and the center of mass. Fifth, spacecraft structure shadowing effects can influence solar radiation torque in two ways: by reducing the effective force being applied, and by shifting the location of the center of pressure. Although at LEO altitudes gravity gradient and atmospheric effects vie with each other for dominance of spacecraft attitude perturbations, once you get to high altitudes solar radiation torque is the prime perturbation.

[5] Wertz, Spacecraft Attitude Determination and Control, p. 573.

By contrast, below 400 km, aerodynamic torque dominates all other environmental torques.[6] A fairly accurate way to model the influence of the atmosphere on spacecraft attitude is to treat its effect as if the momenta of any impacting molecules are absorbed by the spacecraft. In other words, instead of visualizing an air molecule as impacting the spacecraft and then bouncing off in an elastic collision, just assume the molecule joins the spacecraft surface molecules, reaches thermal equilibrium with them (the spacecraft), and then just escapes away. Since the surface thermal velocity is much less than the air molecule's incident velocity, it is basically the same as assuming the air molecule gives up all its kinetic energy to the spacecraft at the time of collision. As with solar radiation pressure, aerodynamic torque is influenced by the size of the displacement between the center of mass and the center of pressure (i.e., the longer the lever arm length, the greater the torque). The aerodynamic torque corollary to solar radiation pressure's radiation intensity and spectral distribution is the atmospheric density, which we know from Chap. 2 is pretty complicated to model and is affected by things like altitude, time of day, and day of year. We also have to be concerned with the spacecraft orientation relative to its flight path through the atmosphere. This is called the *angle of attack*, and is defined as the angle between the spacecraft velocity vector and the outward normal of the cross-section presented to the atmosphere. Also, as with solar radiation torque, shadowing effects (relative to atmospheric impingement) can reduce the force being applied and shift the center of pressure.

The last class of attitude torques, flexible dynamics, is more a description of what's happening to the spacecraft than a description of an environmental phenomenon that's acting on the spacecraft. Flexible dynamics refers to motions of spacecraft appendages, such as antenna booms or solar arrays that can impact the accuracy of the spacecraft's pointing. Contributing causes to flexible dynamics perturbations effectively include all the phenomena we've already discussed in this chapter, i.e. gravity gradient forces, heating due to solar radiation, solar pressure, aerodynamic drag, and thrust forces. Gravity gradient forces can affect spacecraft structures in several ways. First, long, flexible components such as antenna booms may be deformed as they "feel" unequal gravitational forces along the length of the structure. Also, as these effects vary throughout the orbit (as a result of non-zero orbit eccentricity), periodic time-dependent variations may be experienced as well. Second, large deformations can even change the spacecraft's moments of inertia, further perturbing the spacecraft's motion. Third, the additional force applied to the structure can introduce an axial tension, which can increase bending stiffness, which in turn will increase the structure's natural resonance frequencies. So when designing the spacecraft's attitude control law, one must consider how the spacecraft may respond in the frequency domain as gravity gradient shifts its resonance locations at different points in the orbit.

Temperature gradients due to unequal solar heating can also introduce deformations. When these effects manifest themselves gradually, temperature variations over an entire orbital period can induce spacecraft motions characterized by the orbital

[6] Wertz, Spacecraft Attitude Determination and Control, p. 573.

frequency. These motions in turn can excite spacecraft natural frequencies that are close to integer multiples of the orbital frequency, producing undesirable resonance behavior. Thermal gradients can also manifest themselves in a more violent fashion, such as the "slip-stick" phenomenon experienced by HST soon after launch. In that case, temperature gradients during day-night transitions led to a build up in mechanical energy within HST's original solar arrays, which when released produced large amplitude transients leading to frequent guide star loss-of-lock. This problem was initially ameliorated via development of a more robust attitude control law, and later was solved by replacing the original solar arrays with those of an improved design during HST's First Servicing Mission.

Structural deformations can also introduce changes in the solar radiation torques experienced by the spacecraft because structural changes can change the angle of incidence of the solar radiation "falling" on the spacecraft. The deformations may also change the location of the center of pressure relative to the center of mass, thereby modifying the lever arm length over which the radiation pressure is being applied. Flexible dynamics effects of this type do not become important (relative to the other three sources) until you get up to geosynchronous altitude. Below 500 km,[7] we're back in the "soup" of the atmosphere, and aerodynamic influences become important. As when we discussed the torques directly inflicted by these sources, aerodynamic drag acts in a similar manner to solar pressure, except the spacecraft is being bombarded by air molecules instead of photons. Finally, operating thrusters can also introduce flexible dynamics effects. As a series of thrust pulses are fired, structural deformation may build up, as for example as a result of thruster plume impingement on a spacecraft boom. Also, the simple expulsion of fuel from the spacecraft will change the spacecraft's moment of inertia tensor and will shift the spacecraft's center of mass relative to its center of pressure, again modifying the lever arm length over which solar radiation pressure and aerodynamic drag are applied.

[7] Wertz, Spacecraft Attitude Determination and Control, p. 17, Table 1–2.

Chapter 4
Attitude Measurement Sensors

One of the original things about the *Columbo* television series was that you got to see the murderer commit a murder, and then got to watch Peter Falk try to learn what you already knew over the course of the rest of the program. Columbo would always go to the scene of the crime and view the carefully preserved murder site with his own eyes. Forensic staff, fully equipped with appropriate state-of-the-art investigative equipment, were there already collecting the on-site information he'd need to solve the crime, and he'd make his own unique, disruptive contributions to the process. Then Columbo would gather more information off-site, by interviewing people in the neighborhood, doing some research, and otherwise making a nuisance of himself. He then analyzed the accumulated information (i.e., he hit his forehead with his hand and smoked a cigar) to develop an initial list of suspects, which then would be narrowed to a prime suspect after collecting more data, irritating more people with his one last question shtick, hitting his forehead again, and smoking some more cigars. Miraculously, the prime suspect was always the guy who we saw commit the murder at the start of the show, and by the end of the show he had always confessed to having committed the crime just so Columbo wouldn't burn anymore holes in his furniture with his cigars.

Without having any direct knowledge of how a real murder investigation managed by real police detectives works, the process described above (with the Columbo comedy removed) actually seems at a high level to be a fairly plausible model for how the problem of determining the guilty party might be solved, and how a whole host of real-world problems in other applications might be solved. Now where does this logical, well-ordered process break down when you're working a spacecraft operational problem? Unfortunately, pretty much everywhere. As opposed to collecting information you need to solve your problem after the fact, you're usually stuck with whatever information happened to be collected before the problem occurred. And unless you're one of a handful of astronauts, it's not likely you'll be personally visiting the scene of the crime to see for yourself what happened and how it occurred.

© Springer-Verlag London Ltd. 2017

H.L. Hallock et al., *ACS Without an Attitude*, NASA Monographs
in Systems and Software Engineering, DOI 10.1007/978-1-4471-7325-0_4

So part of the spacecraft design process must consist of anticipating what problems might happen in-flight, which translates as including information-gathering devices onboard that are capable of generating the data needed to analyze the problems when they happen, installing software that will identify the problems as they happen (placing the spacecraft in at least a safe condition until the problem can be solved and the solution implemented), and recording the information collected onboard in an organized fashion so an orderly analysis of the data can be performed. For spacecraft operations, as opposed to murder investigations (fictional, or otherwise), the most important actions then must be taken before (in fact many years before) the problem occurs if a quick, economical solution is to be achieved. And one must view the crime scene through the "eyes" of the "dumb" electronic information gatherers, not through the eyes of a smart, experienced professional. Fortunately although there are thousands of murderers and victims in this country each year (that probably didn't come out the way we intended), with a fairly wide range of motives and methodologies, there are only a relatively few number of spacecraft mission types that are flown, with a relatively limited number of problems to solve. So there are many fewer potential victims to be monitored and a lot less information and information types that need to be collected.

For spacecraft attitude determination and control purposes, the information gatherers are called attitude measurement sensors. These sensors are used to collect information needed to deal with (hopefully) infrequent inflight anomalies, and more importantly are used to supply the attitude data required to operate the spacecraft when conducting its nominal science activities. There are five major classes of attitude sensors: namely, Sun sensors, Earth sensors, magnetometers, star sensors, and gyros. The first four are absolute attitude sensors, i.e., they provide measurements that can enable determination of the spacecraft's orientation in inertial space. The fifth sensor type, gyros, only measures the change in attitude between the current measurement and the time of the previous measurement. Each sensor type has its own range of accuracy and reliability, its own natural orbital geometry for application, its own responsiveness and processing speed, and its own strengths and weaknesses. When designing the attitude sensor suite to support a specific mission, all these elements must be factored in, along with more mundane issues such as cost and availability.

But before you run out and start designing your own spacecraft, let's spend some time talking about each sensor type, starting first with a comparison of their relative accuracies, without getting into the details of how they work until later. The first sensor type we'll look at is Sun sensors. Sun sensors generally measure the direction of the Sun's center (called the Sun vector) in the sensor frame (transformable to the body frame through multiplication by the sensor alignment matrix), although for spinning spacecraft the measurement is just the angle between the Sun vector and the spin axis. There are two types of Sun sensors typically used: coarse Sun sensors (CSSs) and fine or digital Sun sensors (FSSs or DSSs), with the CSSs not surprisingly being the less accurate of the two types. Sun sensor accuracies range from 3° to 0.005°

(18 arcseconds).[1] In rare cases, typically for solar observing missions, a DSS may (with a lot of effort) be calibrated accurate to a few arcseconds. By contrast, even after calibration, CSS accuracy will always be limited by their less sophisticated construction and by Earth albedo effects, i.e., sunlight reflected from the Earth into the CSS that is interpreted by the CSS electronics as coming directly from the Sun, thereby producing significant measurement errors.

Earth sensors measure quantities from which you can deduce the direction vector to the Earth center (called the nadir vector) relative to the spacecraft. Like Sun sensors, there are also two basic types of Earth sensors: scanning sensors and fixed sensors. Scanner accuracies range from 1.0° to 0.1°, while fixed sensors are good from about 0.25° to 0.1°.[2] Most Earth sensors trigger on the Earth's atmosphere rather than the Earth's surface. They respond to infrared (IR) radiation at what's called the CO_2 level about 40 km above the Earth's surface. The altitude of this level is greatly affected by daily and seasonal factors related to solar radiation, so interpretation of the sensor's data requires use of a horizon radiance model, the accuracy of which limits the effective accuracy of the sensor.

Magnetometers, or more precisely three-axis magnetometers (TAMs) measure the local geomagnetic field vector relative to the spacecraft. Their accuracies range from 3° to 0.5°.[3] When used in single-frame attitude determination (i.e., computing individual attitudes from individual measurement sets as opposed to computing an average attitude over a range of data), the accuracies of the results are limited by the accuracy of the geomagnetic field model. However, better results can be obtained by averaging out field model errors over a swath of data.

Star sensors measure the direction to a star (or set of stars) relative to the spacecraft. Unlike the previously described sensors that have remained basically the same over the past 20 years, star sensors have undergone enormous increases in sophistication, largely due to the inclusion of software functionality and star catalog data directly in the star sensor "box". But more on those details later. For "standard" star sensors, accuracies range from an arcminute (60 arcseconds) to several arcseconds. However, using special customized fine error sensors (FESs), such as the Hubble Space Telescope's (HST's) fine guidance sensors (FGSs), sub-arcsecond accuracies can be achieved, where the absolute (as opposed to relative) accuracy of attitudes derived from the sensor data largely is limited by that of the associated star catalog.

Gyros measure the instantaneous spacecraft angular velocity, and by integration the change in orientation of the spacecraft over a period of time (e.g., since the last star tracker attitude measurement). There have also been broad-based advances in space-qualified gyro technology. Whereas previously the term *gyro* was effectively synonymous with mechanical gyro, the class of gyros has now been expanded to include vibrating gyros and optical gyros. Because gyros only measure relative attitude changes, not absolute ones, small systematic measurement errors when integrated over time will generate very large errors if not regularly corrected. This

[1] Larson and Wertz, Space Mission Analysis and Design, p. 360, Table 11–14.

[2] Ibid.

[3] Larson and Wertz, Space Mission Analysis and Design, p. 360, Table 11–14.

secular, ramping error is called the *gyro drift bias*, and has an associated accuracy for mechanical gyros ranging from 0.003 to 1 deg/hr.[4] Hemispheric resonating gyros (HRGs), a class of vibrating gyro, deliver comparable drift bias errors, ranging down to 0.005 deg/hr, or for a lot of additional cost down as low as 0.001 deg/hr.[5] At the coarse end, there's a particularly rugged class of vibrating gyro that can survive being fired from a gun and still be reliable to 0.1–0.01 deg/hr.[6] The third class of gyro, optical gyros, itself comes in a couple different "flavors", fiber optic gyros (FOGs) and Ring Laser Gyros (RLGs). We'll have a lot more information on these various types of gyros later in this chapter, but for now let's return to the first sensor type we mentioned, Sun sensors, and cover it in more detail.

4.1 Sun Sensors

Since Sun sensors have many positive qualities and very few negatives, just about every GSFC spacecraft flies at least one type of Sun sensor. Unlike the case of Earth sensors, where the apparent size of the Earth varies dramatically with orbit altitude, the large distance from the Earth to the Sun (almost 4 orders of magnitude larger than the spacecraft-to-Earth distance for even high altitude Earth orbits) causes the Sun to appear (from the point of view of a spacecraft in Earth orbit) to be a disk of constant angular diameter, about a half a degree in diameter. For some calculations, the Sun can be approximated simply as a point source. So, unlike many Earth sensors, a Sun sensor does not need to be pre-tuned to the anticipated orbit altitude. Furthermore, unlike the case of star trackers and the star directions they measure, the Sun is so bright that nothing else can be mistaken for it (i.e., there are no potential spoiler sources). Also because of the Sun's extraordinarily high brightness, Sun sensors require little external electrical power. In fact, the energy delivered by solar radiation falling on the sensor (at the Solar separation distances associated with most Earth orbits) is sufficiently high that CSSs are powered by the object they're measuring. In addition to their specialized role in attitude determination, Sun sensors also are used to support the other subsystems' needs by enabling bright object constraint checking (relevant to optics and thermal protection), facilitating solar array positioning (relevant to power maintenance), and providing timing initialization for spinning spacecraft (relevant to thruster firing).

There are three significant issues associated with Sun sensor use. First, when the spacecraft is in orbit night, the Sun is occulted by the Earth, so no Sun sensor data is available. This will particularly impact low-Earth orbiting spacecraft where extended periods of shadow are inevitable. But for most of these missions, the Sun sensors are primarily used during the early operations phase immediately after separation from the launch vehicle, and when the spacecraft enters safemode as the result of an inflight

[4] Ibid.

[5] Lawrence, Modern Inertial Technology, p. 158 and 167.

[6] Lawrence, Modern Inertial Technology, p. 167.

anomaly. Typically, the launch window is selected to avoid shadow periods right at separation because of the need (from a spacecraft health and safety standpoint) to restore the spacecraft battery to a fully charged state, so Sun sensor data is usually available for an extended time period after separation. Once the Sun is acquired via the Sun sensor, there usually is a non-Sun sensor fall-back option (for example, gyro hold or passive gyroscopic stiffness) that will keep the spacecraft properly oriented relative to the Sun during shadow periods. These fall-back options are also used to support Sun-pointing safemodes when the Sun is not in view. The second issue also relates to Sun visibility, but from a more immediate blockage source, the spacecraft itself. Clearly, if the Sun sensor does not have an unobstructed view of the Sun, either a measurement will fail to be made or the measurement will suffer a high level of error. Third, reflections off a spacecraft structure that fall into the Sun sensor could make the Sun appear to be shifted from its true direction. So to avoid being shadowed by masts, booms, etc., the Sun sensors are often mounted at the ends of the vehicle, or directly on its solar array wings (a mounting strategy utilized only for CSSs) where structural bending can limit the accuracy of attitudes computed using the associated Sun sensor data.

Now having talked about why most spacecraft fly Sun sensors, let's describe more fully the three different classes of Sun sensors: Sun presence sensors, analog sensors, and digital sensors. All that a Sun presence sensor measures is whether it sees the Sun or not. So the output signal is above a threshold if the Sun is in the sensor's field of view (FOV) and is below a threshold if the Sun is not in the sensor's FOV. A Sun presence sensor has many important uses independent of attitude determination. It can be used as a bright object detector to protect light sensitive instrumentation. It can be used as a trigger to activate hardware and or software. It also can be used to reorient the spacecraft, possibly as a precursor to higher accuracy activities. For example, a bright object sensor that identifies that the Sun is on the "wrong" side of the spacecraft can initiate and control an attitude reorientation to bring the Sun into the FOV of more sophisticated Sun sensors. And onboard a spinning spacecraft, you can place two Sun presence detectors in a V-shaped configuration to form a 2-slit Sun sensor that can measure the angle between the Sun vector and the spin axis of the spacecraft (see Fig. 4.1).

As opposed to Sun presence detectors, analog Sun sensors (also called coarse Sun sensors (CSSs)) measure the angle between the sun sensor normal and the Sun direction. When the Sun is in a CSS's FOV, its silicon photocell generates an output

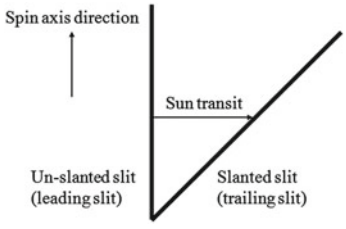


Fig. 4.1 V-slit Sun sensor

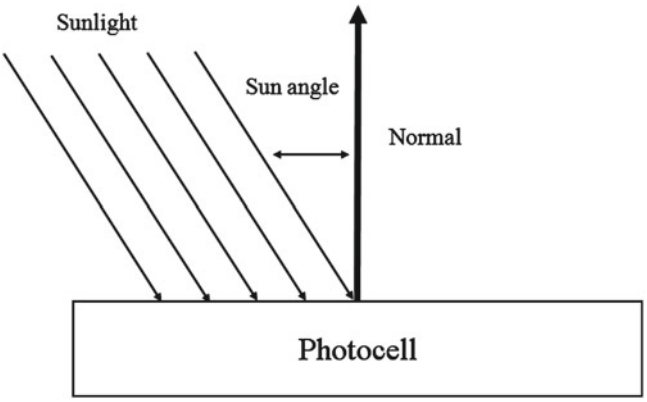


Fig. 4.2 Cosine detector sun sensor

signal that is a continuous, monotonic function of the Sun angle. The output current is proportional to the cosine of the angle between the sensor boresight and the Sunline, which is why the sensor is often called a cosine detector (see Fig. 4.2).

Because the device is so simple, it can't tell the difference between direct light from the Sun and Earth albedo, i.e., sunlight reflected off the Earth. Because the angle output by the detector is a measurement of the center of the light distribution input to the sensor, spurious light generated by Earth albedo will cause the sensor to "believe" the Sun direction is shifted from its true direction. Error due to Earth albedo can be quite large (the Global Precipitation Measurement team saw errors of upwards of 30° when flying near Antarctica). Consequently, you only want to use CSS data for applications having fairly low accuracy requirements, for example last gasp bright object constraint checking, pointing control during safemode, attitude sanity checking, etc. With 2 photocells arranged as shown in Fig. 4.3a, you can difference the output from the two detectors to yield a nice linear signal over a conical FOV (of half angle about 45°) centered on the boresight (see Fig. 4.3b).

With 4 photocells you can measure the angle the sunline makes with respect to two orthogonal planes, which is sufficient information to define the direction of the Sun vector in the spacecraft body frame (the corollary to defining a star direction from two angles, right ascension and declination, as discussed in Chap. 1). The direct output of a CSS is two angles, which the flight software (FSW) processes to convert to vector format for use by other onboard applications and the ground. Note that although the reliable measurement range of a CSS is only a cone of half-angle 45°, since CSSs are so cheap you can economically cover the full sphere about the spacecraft simply by flying several CSSs with overlapping FOVs.

Like CSSs, digital Sun Sensors (DSSs), also called fine Sun sensors (FSSs), measure angles that the FSW can transform into the Sun vector direction. And also like CSSs, DSSs output a Sun presence flag that tells you when the Sun is in the FOV of the sensor. However, to achieve its much higher accuracy compared to that of a

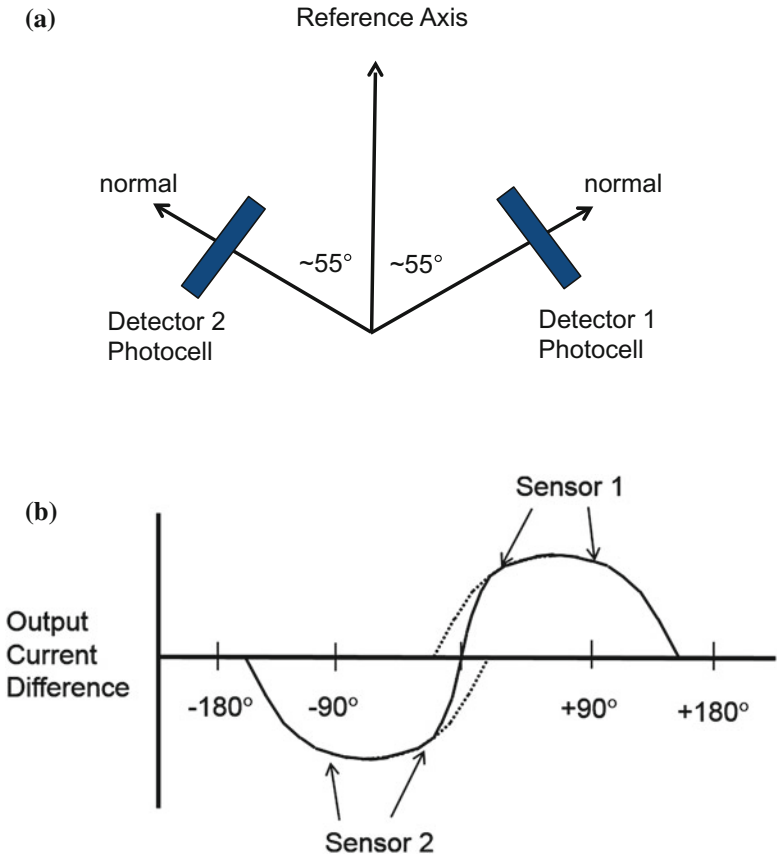


Fig. 4.3 **a** Cosine detectors used together. **b** Output current difference between two cosine detectors

CSS (arcminute accuracy as compared to degrees), the construction of the DSS must be far more complex (and expensive) than that of a CSS. The way a DSS works is shown in Fig. 4.4a, b.

Light from the Sun comes into the DSS entrance slit and illuminates a pattern of slits sitting on top of a series of photocells, with a photocell positioned under each row of measurement slits. As with the CSS, the photocell voltage is proportional to the cosine of the Sun angle. The first row of measurement slits (the slit to the extreme right in Fig. 4.4b) is used as a threshold setting with respect to using data from the other rows. Specifically, measurement slits (and the bits they represent in the Sun angle measurement) are deemed to be "on" if their intensity is greater than the threshold slit output. Since the threshold slit is half the width of the other measurement slits, a slit's bit is therefore considered to be "on" if it is at least half-illuminated by the Sun. The next row of slits to the left of the threshold slit is the sign bit, which (depending on which photocell is illuminated above threshold) defines which side of the sensor

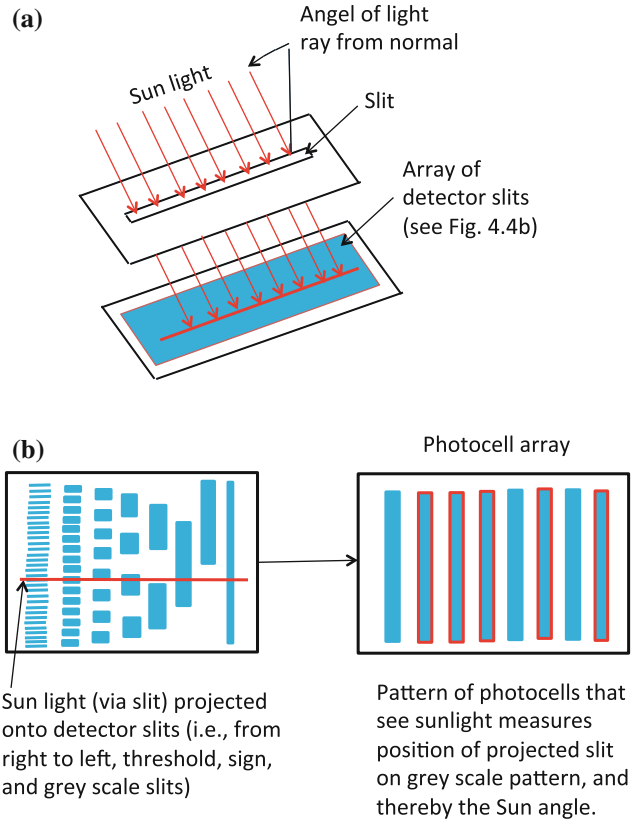


Fig. 4.4 **a** Interpreting digital sun sensor output. **b** Interpreting digital sun sensor output

the Sun is on. The bits defining the magnitude of the Sun angle then are derived from the remaining rows. The central rows are referred to as the "Gray code", and provide Sun angle resolution good to a half a degree, approximately the apparent angular diameter of the Sun as viewed from the Earth (about 0.53° from LEO altitude). The rows of slits that are situated most to the left then generate the fine bits supplying the highest degree of resolution, typically up to 0.125°, although DSSs have been calibrated to even higher levels of accuracy (5 arcseconds in the case of the Solar Maximum Mission (SMM)). Of course, in order to define the direction of the Sun vector we need to measure two angles relative to the sensor's boresight. This can be accomplished via a two-axis DSS.

For more information on Sun sensors, see Wertz, Spacecraft Attitude Determination and Control, Sect. 6.1.

4.2 Earth Sensors

Just as the target of a Sun sensor, the Sun, is easy to find, the target of an Earth sensor, the Earth, is pretty easy to acquire as well. As a matter of fact, at LEO altitudes, the Earth will appear to cover up to 40% of the celestial sphere.[7] And if your mission is an Earth imager, your science instrument will be approximately pointing in the same direction as the Earth sensor (i.e., along the nadir vector), so it is natural to couple the Earth sensor output directly into the fine attitude control law, pointing accuracy constraints permitting. However, use of Earth sensors for pointing control can engender a number of problems not characteristic of Sun sensor use. First, the accuracy of the Earth sensor-inferred attitudes is limited by the accuracy of its reference model. For example, for IR Earth sensors, the associated horizon radiance model typically is only good to about 0.1°, although data averaging can significantly reduce the effects of the modeling errors. Second, as with CSSs, extraneous light can degrade measurement accuracy. In this case, the extraneous light can result from reflections from high altitude clouds. Third, although the large apparent angular diameter of the Earth can make acquisition of the Earth by the sensor easier, the large angular diameter (as opposed to the near-point source size of the Sun) requires identifying the Earth's horizon, rather than measuring the center of a small light distribution, in order to achieve reasonable attitude accuracy.

Identification of the hard edge of the Earth's horizon (as seen in visible light) can be difficult because energy radiated from the Earth decreases gradually due to the presence of atmosphere and variations between illuminated and non-illuminated regions. Simply modeling the visible light terminator is an issue by itself. These problems can be ameliorated somewhat by viewing the Earth in the infrared. Most Earth sensors utilize the CO_2 band, whose wavelength lies between 14 and 16 μm and whose edge is about 40 km above the Earth's surface.[8] Use of the CO_2 band eliminates the terminator modeling and albedo issues pretty successfully as the Earth "looks" pretty uniform across its disk at CO_2 wavelengths (heat stays trapped in the atmosphere overnight so there's no terminator phenomenon between sunlit Earth and shadow), and IR detectors are less susceptible (than visible light detectors, including Sun sensors) to reflected light from spacecraft structures.

An Earth sensor contains three major functions, namely scanning (supplied by the spacecraft itself for spinners), optics, and detection. The optics portion needs both a lens to focus the target image on the detector and a narrow-band filter to limit the observed spectral band of the target (e.g., 14–16 μm). The detector typically is a bolometer. When light hits a bolometer, the incident radiation heats a resistor, thereby changing its resistance. Since the bolometer electronics maintains a constant current through the resistor, the change in resistance can be quantified by measuring

[7] J. Wertz, Spacecraft Attitude Determination and Control, p. 167.
[8] Ibid.

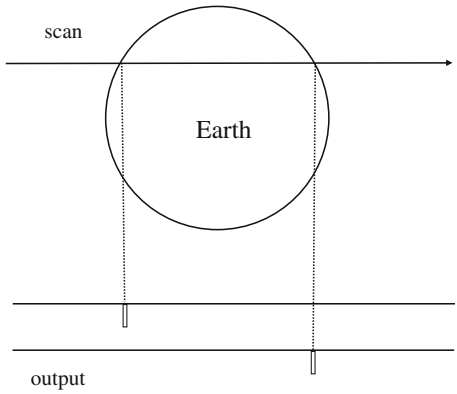


Fig. 4.5 Earth sensor output

the voltage change across the resistor, which is correlated with the change in light intensity incident on the detector. So the detector can "see" the horizon edge by identifying the sharp intensity drop-off as the scanner rotates the bolometer line-of-sight past the Earth and into space. Put more formally, the output of the Earth sensor is a measurement of the time between sensing a reference and the time associated with an electronic pulse created when the detector's measured voltage crosses a threshold (above or below), as illustrated in Fig. 4.5.

For a spinning, body-mounted Earth sensor, the reference usually is the electronic pulse associated with the leading slit of a V-slit Sun sensor "seeing" the Sun. For a wheel-mounted sensor, the reference usually is a magnetic pick-off on the body. When the detector signal increases across the threshold, the detector is said to experience acquisition of signal (AOS), and when the signal decreases across the threshold we have loss of signal (LOS).

There are three generic types of Earth sensors, horizon-crossing indicators, static sensors, and conical scanning sensors. The simple horizon-crossing indicators are mounted on the bodies of spinning spacecraft, like the "old-fashioned" Geostationary Operational Environmental Satellite (GOES)-D through -F series. They directly measured AOS and LOS times. When that information is combined with the known angular size of the Earth based on the spacecraft's distance from Earth, you can calculate the nadir angle, i.e., the angle between the spacecraft spin axis and the direction to Earth's center relative to the location of the spacecraft. See Fig. 4.6a for an example in which the radius of the scan cone (the angle between the spin axis and the Earth sensor direction) is larger than the angular radius of the Earth, resulting in two solutions for the nadir angle. (Note, it is assumed that the spacecraft spin period is sufficiently short that spacecraft translational motion can be neglected during a single spin period.) The nadir angle solutions together with the Sun angle (the angle between the spin axis and the Sun vector) measured by a Sun presence sensor can then be used to determine the direction of the spacecraft spin axis in inertial space

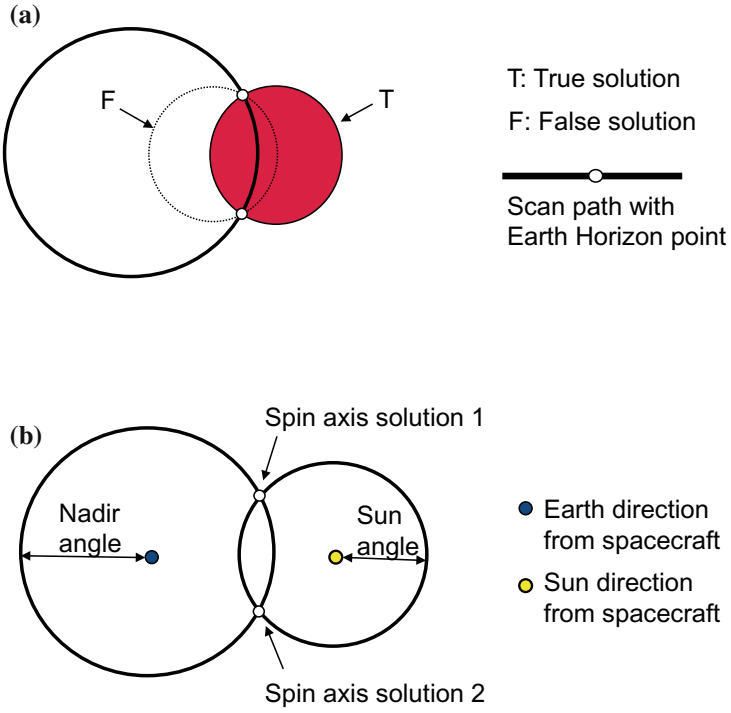


Fig. 4.6 **a** Earth Scanning Horizon Sensor: scan path crossing IR Earth image from AOS to LOS as viewed from satellite location. **b** Determining spacecraft inertial-space spin axis given the Earth nadir angle and the Sun angle relative to the spacecraft spin axis

given the known directions to Earth center and Sun in inertial space; see Fig. 4.6b, for which you're to imagine viewing the Earth (center) and Sun in inertial space and the spacecraft spin axis pointing towards one of the two indicated solutions. This solution also has an 2-fold ambiguity. The total four-fold ambiguity for spin axis direction in inertial space can be resolved by accumulating additional Earth and Sun sensor data over a period long enough for the Earth direction to change significantly, but short enough for the spacecraft spin axis direction to remain essentially constant. The proper solution will remain fixed while the spurious solutions drift as the relative direction to Earth changes.

The second class of Earth sensors, static sensors of the type flown on Landsat-7 and the Tropical Rain Measuring Mission (TRMM) (see Fig. 4.7), employ four oscillating detectors that dither across the horizon edge, thereby generating sets of rectangular pulses. The width and spacing of the pulses depend on the detector position with respect to the horizon. Processing this data through an onboard control law, the ACS FSW for these missions corrects the measured errors in pointing using

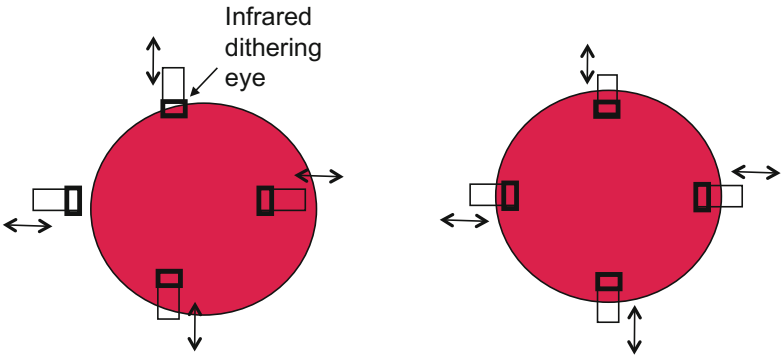


Fig. 4.7 Static Earth sensor: *Left* - Earth not centered relative to eyes. *Right* - Earth centered

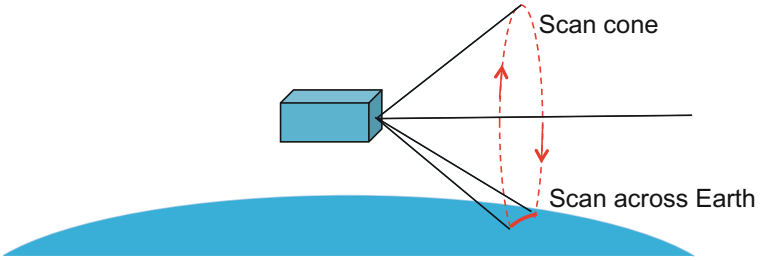


Fig. 4.8 Conical scanning Earth sensor

reaction wheels, thereby centering the Earth in the FOV of the sensor. When the centered state is achieved, the horizon penetration computed for opposite quadrants of the Earth sensor will be the same. Note that for the four quadrants to be able to bracket the Earth horizon, the Earth sensor must be built so as to be tuned to a fairly narrow range of mission orbits and attitudes. So there is not much flexibility for using the sensor if anomalous geometries occur.

Conical scanning Earth sensors, such as the one flown on the Earth Observing System (EOS)-AM mission, constitute the third class of Earth sensor (see Fig. 4.8). By canting the boresight of the sensor 45° from (for example) the direction of flight, the scanner causes the boresight to rotate over a cone of 45° half-angle. When the detector crosses into the Earth, the temperature difference between the Earth and space is sensed by the detector and a pulse is generated by the detector electronics. From those measurements, pitch and roll attitude errors (relative to the orbit frame) can be generated.

For LEO spacecraft, selecting the sensor scanning axis to be the pitch or roll axis (without a cant angle) will guarantee that the sensor FOV will intersect the Earth. At higher altitudes the sensor boresight must be canted down in order to ensure intersecting the Earth. For example, at geosynchronous altitude, a typical cant angle is 40°. As with the static Earth sensor previously described, since the Earth width

is a function of altitude, you have to tune the conical scanning Earth sensor to the expected orbit. Note that we've tacitly assumed up to now that the mission orbit is circular, so the Earth width could be assumed to be constant over the entire orbit. However, for elliptical orbits this is not the case. Therefore for missions employing elliptical orbits, you might need two conical scanners to deal with the variation of Earth widths over the orbit.

4.3 Magnetometers

Of all the types of attitude sensors from which to choose, magnetometers (which measure both the magnitude and direction of the magnetic field in which they're immersed, specifically the geomagnetic field for Earth orbiting spacecraft) are the bargain basement option from three standpoints. First, the instrument itself is cheap. Second, magnetometers are lightweight so they don't affect launch vehicle trades much. Third, they draw very little power, so they don't affect power system trades either. In addition, they are very robust, i.e., they're fairly reliable, have no moving parts, and operate consistently over a wide range of temperatures. On the downside, they're not very accurate. Limited for attitude determination purposes by the accuracy of the geomagnetic field reference, attitudes computed from magnetometer and Sun sensor data typically are only good to about 2°, although using data collected over a significant portion of the orbit can help average out the reference errors and improve results. "Serious" use of magnetometer data above 1,000 km usually is avoided, since at those altitudes uncertainties in the spacecraft's residual dipole moment can dominate the Earth's field.[9]

As you'd expect given its inexpensive cost, a magnetometer is a pretty simple device. It consists of a magnetic sensor and an electronics unit to process the sensor signals. Most magnetometers flying today utilize the principal of magnetic inductance to make their measurements, although there are some designs based on atomic properties such as Zeeman splitting or nuclear magnetic resonance (NMR). Our discussion here will be limited to the magnetic induction type since it is pretty standard on GSFC spacecraft, and that type often is provided at GSFC as an in-house product.

The dual-core fluxgate type is a good example of a magnetic induction magnetometer. The fluxgate converts an external magnetic field (say, from the Earth) into a measurable voltage. For a dual-core fluxgate magnetometer, there are two magnetically saturable cores, each wound by identical coils (see Fig. 4.9), but with opposite winding directions.

By applying a sufficiently high alternating current through the coils (see Fig. 4.10a), you can drive the cores to states of opposite saturation, where saturation means that the magnetic fields generated in the cores (by the current in the coils) reach their maxima at magnetic field values below that generated by the coils.

[9] J. Wertz, Spacecraft Attitude Determination and Control, p. 181.

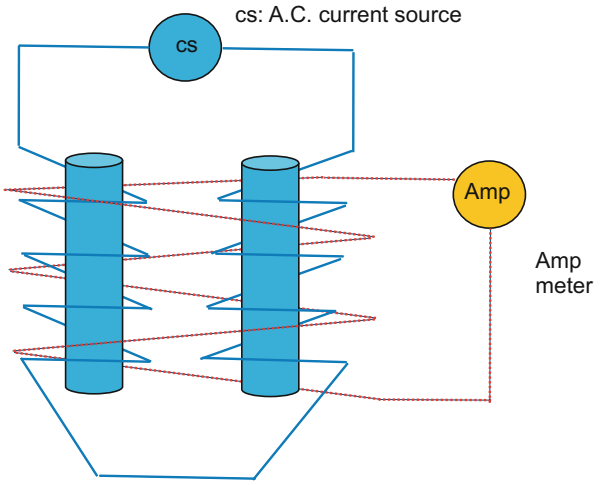


Fig. 4.9 Dual-core flux magnetometer

In the absence of any external magnetic fields, the signs of maximum coil fields will flip-flop at half the period of the alternating current. Now, let's assume the cores are sitting in an external magnetic field, in our case the geomagnetic field. The external field will introduce an asymmetry between the time interval spent at positive saturation versus the time spent at negative saturation (see Fig. 4.10b, c).

For example, assume the direction of the external field is parallel to and smaller than the coil field. If the external field is positive, the net field "felt" at the core (the sum of the external field and the field produced by the current in the coil) for positive current will be larger than if the external field was zero, the core will reach saturation sooner, and maintain saturation longer. By contrast (again assuming the external field is positive), the net field felt at the core for negative current will be smaller than if the external field was zero, the core will reach saturation later, and maintain saturation shorter. So in the presence of an external magnetic field, the times between zero crossings, calculated from zero crossing times identified by the magnetometer's output coil, will no longer be half the period of the alternating current (see Fig. 4.10d).

This in turn means that given a measurement of the time asymmetry there should be some way to calculate the external field that has produced the asymmetry, and in fact Eq. 4.1 supplies a simple formula for doing just that.

$$\Delta H = (1 - 2K_2)H_D \tag{4.1}$$

where

ΔH = magnitude of the external field (gauss).
K_2T = time interval between zero crossings relative to crossings bracketing the time spent at saturation having sign opposite to external magnetic field component

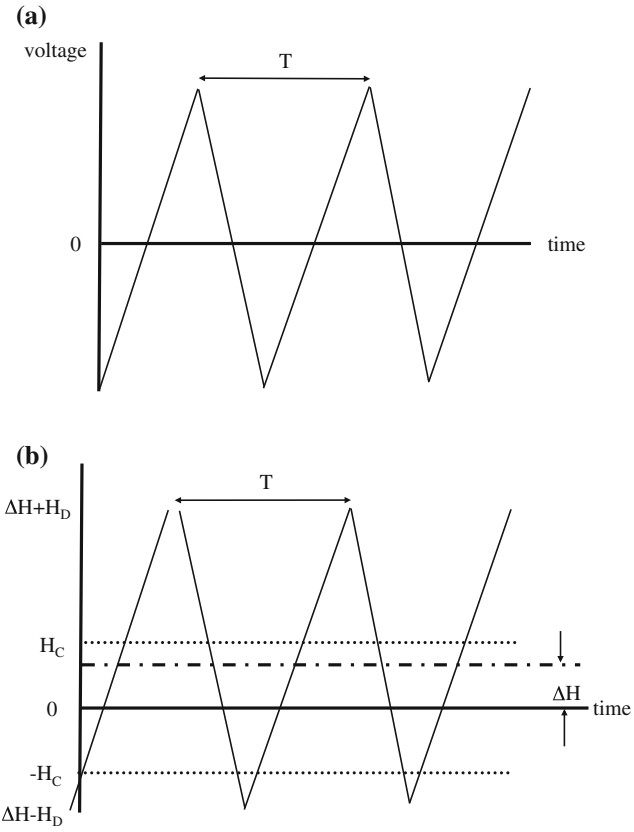


Fig. 4.10 **a** Variation in voltage driven by alternating current. **b** Asymmetric magnetic intensity induced by external field. **c** Asymmetric time durations at positive and negative saturation. **d** Induced voltage in secondary

parallel to the coil (sec), where T is the period of the alternating current applied to the coils.
$(1 - K_2)T$ = time interval between zero crossings relative to crossings bracketing the time spent at saturation having the same sign as the external magnetic field component parallel to the coil (sec).
H_D = magnetic intensity achieved when the coil voltage is at maximum (gauss).

The sign of the external field (whose magnitude is computed via Eq. 4.1) is determined by examining a Fourier expansion of the coil voltage. Note also that by using two cores wound oppositely, the output coil is insensitive to the alternating current frequency.

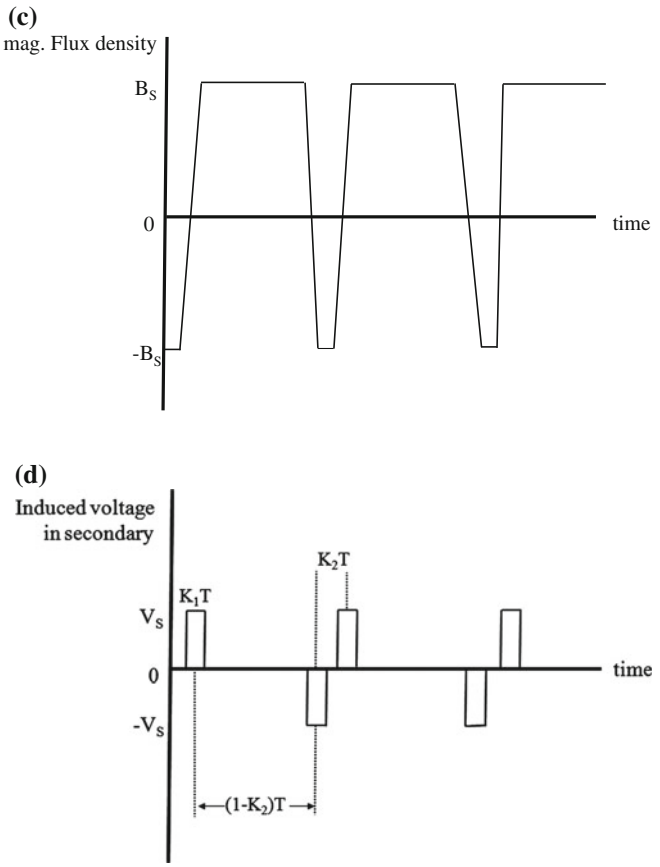


Fig. 4.10 (continued)

4.4 Star Sensors

If what you want is the most accurate way to measure your absolute pointing in inertial space, and you have a reasonable amount of money burning a hole in your pocket, a star tracker would be your sensor of choice. Compared to very bright objects like the Sun or the Earth observed by Sun sensors and Earth sensors, respectively, the stars sensed by star trackers will be much fainter in intensity by many orders of magnitude. Star trackers typically use stars between magnitudes 2 and 6, as compared to magnitude -26.7 for the Sun as seen from the Earth[10] and magnitude -12.7 for a full moon,[11] where magnitude is a linear function of the logarithm of the intensity (see Eq. 4.2).

[10] J. Wertz, Spacecraft Attitude Determination and Control, p. 79.

[11] Larson and Wertz, Space Mission Analysis and Design, p. 808.

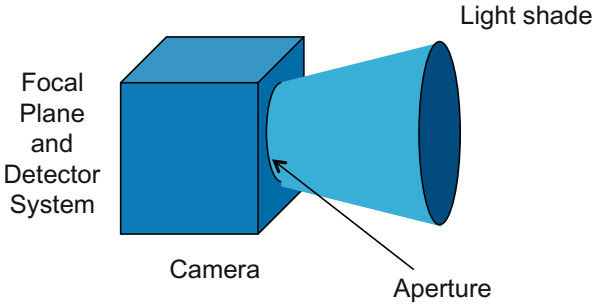


Fig. 4.11 Generic star tracker

$$m = m_0 - 2.5 \log_{10} i \tag{4.2}$$

where

m = object's magnitude
m_0 = scale constant
i = object's intensity (relative to intensity of a source of magnitude m_0)

Put a little more concretely, a handy rule-of-thumb is that a difference of 5 in magnitude is equivalent to a factor of 100 in intensity.[12] Also, the scale factor relating intensity to magnitude is negative, so negative magnitudes are brighter than positive magnitudes, and small positive magnitudes are brighter than large positive magnitudes.

A star sensor is a much more complicated instrument than an Earth sensor or a Sun sensor. For example, a star sensor designer has to be equally concerned with keeping out undesirable light as with bringing in the light from the source it's trying to measure. To satisfy the former need, there's a limiter on the star sensor FOV and a Sun shade to keep stray or reflected light from entering the optical components and degrading the accuracy of the sensor's measurements, or even preventing the sensor from functioning at all. The fixed head star tracker (FHST) that was the NASA standard star tracker back in the 1980s and early 1990s also had a shutter that would snap shut automatically if sensed incoming light exceeded a threshold, thereby protecting sensitive optics electronics from being burnt out. A generic illustration of a star tracker is presented in Fig. 4.11.

The remaining components sound, at a high level, a lot like the Earth sensor we've just examined. There's a lens to focus the light onto a detector and electronics to control and process the detected signal. Specifically, the electronics includes a commendable thresholding capability that rejects signals for further processing if they fall below a specified intensity threshold. More sophisticated star sensors, such as the Hubble Space Telescopes's (HST's) fine guidance sensors (FGSs) also have a

[12] J. Wertz, Spacecraft Attitude Determination and Control, p. 78.

maximum intensity threshold (i.e., signals are rejected if their intensity rises above some maximum value).

Returning briefly to the past, the primary output of the FHSTs, as well as the first generation of charge-coupled device (CCD) star trackers, were star horizontal and vertical coordinates within the FHST FOV, as well as the measured intensity of the star. Nowadays, advanced star trackers (ASTs) come equipped with star catalogs and software that allow the sensor itself to identify detected stars, match measured coordinates in the star tracker frame up with reference coordinates in inertial space (star right ascensions and declinations), and calculate the corresponding transformation from inertial space into the star tracker frame in quaternion format. Flight software (FSW) can then take the output star tracker quaternions, correct for the star tracker alignment relative to the spacecraft body frame, compensate for the effects of velocity aberration (to be discussed further in Sect. 6.6), and generate the spacecraft attitude quaternion (a transformation from inertial space to the body frame). For some ASTs, the FSW does not even have to compute the velocity aberration compensation. Instead, the FSW periodically sends the AST the spacecraft total velocity vector, and the AST performs its own velocity aberration compensations on each star it observes.

There are so many advantages to using star trackers as the prime source for absolute spacecraft orientation information that they are flown on most well funded missions. Star trackers and other star sensors provide the most accurate source of absolute attitude information. FHSTs could measure attitude good to 10–20 arcseconds and modern star trackers are easily capable of providing accuracies to a few arcseconds. And for just a few million more, fine error sensors (FESs), for example HST's FGSs and the science instrument (SI) itself on other missions, can deliver measurements in the milliarcsecond range by utilizing the light gathering capabilities of the science optics itself. In addition, as star trackers have few moving parts, they tend to be very durable and reliable. And, unlike the previous three sensors we've discussed, technological advances just keep improving the quality, capabilities, and performance of the sensors. In the 1990s, we went from the old FHST design that could only track one star at a time, outputting only that star's coordinates and intensity, to ASTs that can track as many as 50 stars at a time and output both an orientation quaternion and the spacecraft rate (in the sensor frame).

On the downside, star trackers used to be both heavy and expensive, though both of these issues have been greatly ameliorated as lighter, cheaper star trackers have been developed. Also, in the past, the need to accelerate electrons in the old FHSTs from the photocathode to the anode required the use of high voltage, causing the FHST to use a lot more electrical power than other attitude sensors. Again, this is less an issue with modern ASTs. However, use of any type of star sensor can introduce the need (on a mission-dependent basis) to deal with geometric issues such as occultation and interference from bright sources.

Occultation can be a big problem for star trackers since the stars they acquire and track can regularly be blocked by the Earth and interfered with by the Moon. For some missions, such as a Lagrange point Earth- or Sun-pointer, the star trackers can be mounted so that occultations will not occur once the spacecraft is on station doing

nominal science. For other missions, such as LEO celestial-pointers, occultations will be a regular occurrence every orbit, unless the sensor happens to be pointed near orbit-normal. Confusion with other similar bright objects (the spoiler problem) also is a star tracker-unique problem. Although it's hard to mistake anything else in the sky for the Earth, and impossible to mistake anything else for the Sun, it's easy to confuse one star with another star of similar magnitude unless you have a good estimate of your current pointing, or you use a star identification algorithm that can compare the pattern of stars (mathematically put, the specification of angular separations between observed stars) you see with actual patterns from a large star catalog.

As successful operation of a star tracker during a mission clearly involves more than just turning it on, there are more commendable features and options than you find on other attitude sensors. Two functions typically are provided to deal with the spoiler problem. First, a thresholding system allows the sensor to reject signals from stars (or Solar System objects) outside the magnitude range of the star you want to guide on, not surprisingly called a guide star. With ASTs, this is less of an issue as powerful star identification algorithms based on distance separations are used to identify stars the same way ground system software used to process FHST full-field map data. So individual guide stars are no longer key to conventional star tracker operations, although guide star strategies often are still employed when utilizing FESs. Second, a reduced field of view (RFOV) capability allows you to isolate the star you're interested in from other stars of similar magnitude. In full FOV mode, the star tracker will search its total FOV (TFOV) until it finds a star (or several stars in the case of modern star trackers) that satisfies the intensity thresholding condition(s). Depending on the commanding, the star tracker then continues tracking that star "forever", or breaks track and resumes searching the TFOV. Using RFOV mode ($1.5^\circ \times 1.5^\circ$ for FHSTs), also called offset mode, the area of the star tracker you scan is much smaller than the total tracker search area ($8^\circ \times 8^\circ$ for FHSTs), so there are fewer stars you can blunder into and mistake for the star you want. Again, this was a more useful strategy when using FHSTs than is now the case for ASTs.

So the advent of the quaternion star trackers has significantly simplified both the associated FSW functionality and mission operations. Before quaternion star trackers were developed, storage limitations on onboard space for star catalogs, central processing unit (CPU) limitations on the sophistication of onboard star identification and attitude determination algorithms, and the inability of FHSTs to track more than one star at a time typically necessitated the use of offset mode with guide stars. Nowadays, quaternion star trackers can track many stars at a time, do their own star identification from their own star catalogs, and compute attitudes directly in quaternion format. As a result, full-field mode has pretty much replaced offset mode operationally.

4.5 Gyros

One thing all the previous four sensor types have in common is that they measure the direction of some entity external to the spacecraft. By contrast, gyros measure the motion, or attempted motion, of an object internal to the spacecraft. Leaving vibrating and optical gyros for later discussions, one can visualize a gyro as if it were a spinning top (supported by a gimbal) whose angular momentum vector is fixed in magnitude and parallel to the top's spin axis. When the spacecraft rotates about the gyro input axis, the gyro (supported by a gimbal) attempts to precess about the output axis (see Fig. 4.12).

This motion, of course, is a consequence of the definition of torque, as we discussed in Chap. 3 (i.e., torque is the cross product of the vector pointing to where the force is applied and the force vector itself). A single-degree-of-freedom (SDOF) gyro has just one gimbal, and so has only one output axis. A two-degree-of-freedom (TDOF) gyro has two gimbals, and consequently has two output axes. With 2 TDOF gyros or 3 SDOF gyros you can therefore provide full three-axis attitude rate information. In actuality, to provide redundancy in the event of hardware failure, more than the minimum gyro complement usually is flown. Also, when more than the minimum required number of gyros is powered-up, data from all the active gyros can be utilized to improve the accuracy of the computed body rate. Rate gyros (RGs) measure spacecraft rates directly, while rate integrating gyros (RIGs) measure angular displacements over a fixed time interval. Either way, unlike the previous four

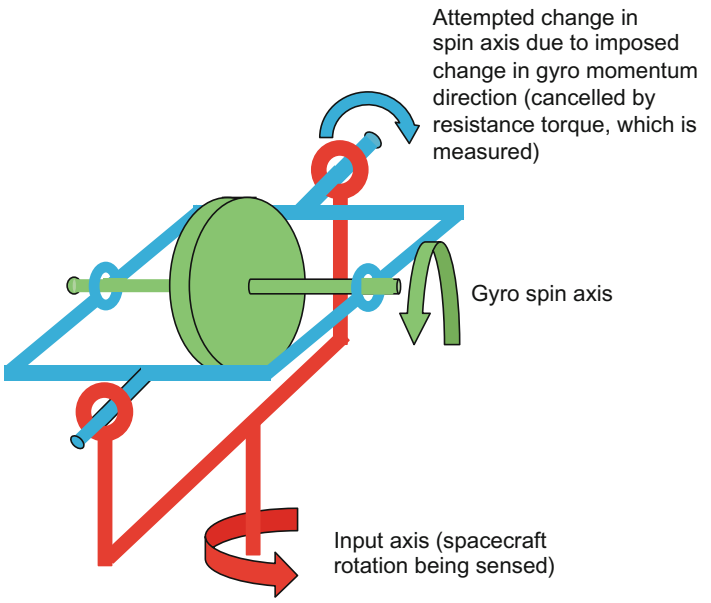


Fig. 4.12 Simple single degree-of-freedom gyro

sensors we've talked about, gyros only measure relative attitude change; they cannot measure the spacecraft's absolute pointing in inertial space.

In practice, the gyro actually is not free to rotate about the output axis as suggested in Fig. 4.12. If it were, the measurement would be the physical precession of the gyro's spin axis, which could be hundreds of degrees. A much better approach is to reduce the amplitude of the motion artificially, and then measure the small change in gyro pointing that is (approximately) proportional to the large, physical vehicle pointing change. This scaling of the gyro motion to the vehicle motion can be achieved through use of viscous damping and spring restraint. A yet better approach is to apply a torque artificially that inhibits gyro spin axis precession entirely. By measuring the torque required to inhibit the gyro motion, you obtain a quantity from which you can "back-out" the vehicle physical motion that must have occurred, thereby necessitating the application of the inhibiting torque. One type of gyro that utilizes this approach is an electromagnetic torque rebalanced gyro, where the current that needs to be applied to null the gyro spin axis precession is proportional to the physical motion of the spacecraft.

The mechanical gyros used on the HST mission are among the highest quality mechanical gyros flown to date. HST's gyros have two modes of operation. Low mode (threshold at 0.00025 arcsec/count, saturation at 38 arcsec/sec) provides greater precision for fine pointing, while high mode (threshold at 0.015 arcsec/count, saturation at 2300 arcsec/sec) provides greater dynamic range during slews. The requirement on the ground system for calibration of the low mode drift bias is 0.006 arcsec/sec (3σ) while the gyro scale factor and alignment has been calibrated as accurate as a third of an arcsecond per degree of slew.

Gyros are ideal partners to star trackers or FESs. They usually have a faster readout than the other sensor and can be extremely accurate, allowing you to "fly" on gyros in-between star tracker or FES updates, and then update both attitude knowledge and gyro drift bias calibrations with absolute attitude sensor information when it is available for processing by, for example, a Kalman filter (we'll discuss Kalman filters in detail in Chap. 7). Gyros also are easy to operate (no occultations or acquisition issues to worry about, and simple onboard processing), are lightweight, and require little power. The down-side is that a mechanical gyro's moving parts make it failure-prone and, since a gyro only can provide relative attitude information and its drift bias ramps with time, gyros must be coupled with an absolute attitude sensor (like a star tracker) to provide a spacecraft with the full onboard attitude knowledge needed to perform a mission, and to maintain the gyro's accuracy throughout the lifetime of the mission. The gyro's scale factor and alignment, by contrast, are more stable and often need to be calibrated only once by the ground system at the start of the mission, then uplinked to the spacecraft for use throughout the rest of the mission.

The vulnerability to failure of the mechanical gyro has stimulated the development of two other gyro classes, vibrating gyros and optical gyros, neither of which utilizes a "spinning top" to define a reference angular momentum vector. A nice way to visualize how a vibrating gyro works is to imagine a vibrating, light-weight string, attached to top and bottom supports, oscillating in a plane (see Fig. 4.13).

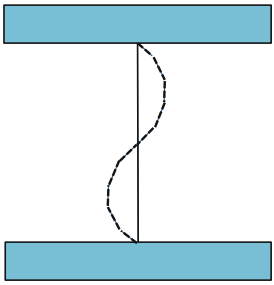


Fig. 4.13 Vibrating string gyro

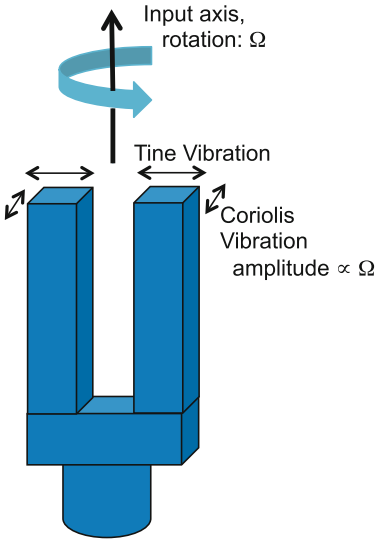


Fig. 4.14 Tuning fork gyro

Now suppose the string supports are attached to a large object that rotates about the axis defined by the line passing through the two support points. Then, by conservation of angular momentum, the plane of the string vibration will remain fixed in inertial space while the object rotates, just as the plane of a swinging Foucault pendulum remains fixed relative to the Earth rotating beneath it. So by observing on the rotating object how the vibrating string plane appears to rotate relative to the object, you can measure how the object itself is rotating relative to inertial space, i.e., you can measure the object's body rate. In practice, a vibrating string would not be very accurate because the momentum stored in the string is small relative to error torques and because of frictional forces at the attach points. The principle illustrated by the vibrating string does, however, apply to two more practical approaches, the tuning fork (see Fig. 4.14 and the hemispheric resonator).

Replacing the vibrating string by a tuning fork deals with the low angular momentum problem by replacing a lightweight string by, for example, metal. If you've ever played with a musical tuning fork by moving it around after striking it (and thereby starting it vibrating), you found that the pure tone you started with changed as the

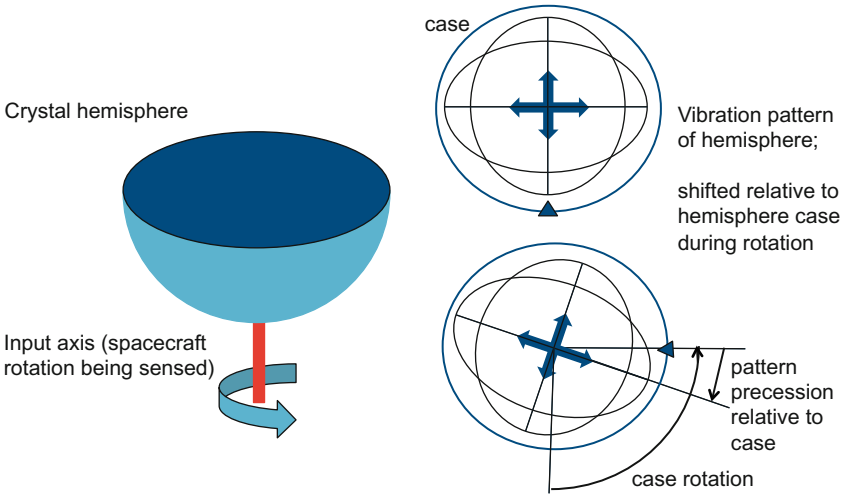


Fig. 4.15 Hemispheric resonating Gyro

tuning fork's plane rotated. Put more technically, rotation of a struck tuning fork about its axis introduces a sinusoidally-varying precession amplitude. Tuning fork gyros naturally have their own problems. First, the elastic moduli of the materials constituting a metal tuning fork can vary with temperature, leading to instability in the gyro's measurements unless very complicated corrections are made. Elaborate calculations of this sort would detract from one of the gyro's prime advantages for onboard use, namely the simplicity of the required data processing facilitating rapid utilization of the gyro's high rate output. Second, as a metal tuning fork vibrates, its tines shift position, thereby changing the tuning fork's center of mass. Again, this theoretically could be compensated for onboard via a sophisticated model of the gyro dynamic behavior, but that would defeat one of gyro's prime advantages. Fortunately, these problems can be dealt with to some degree by using crystalline quartz instead of metal, and in fact Draper Labs has developed a gyro of this type utilizing micro-machined silicon. Its resolution for use onboard a spacecraft is far too coarse at 20 deg/hr, but it is very rugged, having the ability to withstand a shock of 120,000-g and still keep ticking.[13] Presumably Draper sells more of their tuning fork gyros to the Department of Defense (DOD) than to the National Aeronautics and Space Administration (NASA).

At last we arrive at a vibrating gyro concept with practical application potential onboard spacecraft, the hemispheric resonator gyro (HRG). In this case, we replace the vibrating string with something that behaves like a struck wine glass. The high-tech replacement for the wine glass is a vibrating hemispheric shell of fused-quartz supported by a stem along its diameter (see Fig. 4.15).

[13] Lawrence, Modern Inertial Technology, p. 156.

Since nothing is ever cast perfectly, the hemisphere is tuned by removing material from the rim so as to compensate for inhomogeneities in the quartz and errors introduced in the original machining. The shell is excited electrostatically so as to cause it to ring at its fundamental frequency, and then is driven to resonance with a constant oscillation amplitude. To lessen frictional losses, the entire enclosure is maintained at a vacuum. Now let's rotate the stem. This will cause the anti-node axis (i.e., the line connecting the maximum and minimum of the vibration) to precess in the same sense as the rotation of the stem, but by a different amount (see again Fig. 4.15). The difference in rotation is defined by the geometry of the HRG. For the HRG pictured in Fig. 4.15, the gain is 0.30, so a 90° stem rotation yields a 27° anti-node axis precession. Recalling the original comment about compensating for quartz inhomogeneities, one of the problems caused by non-uniformity in the shell is that the nodes will tend to "cling" along an axis of minimum energy. So the larger the non-uniformity in the shell, the more effort it takes to move the nodes out of their potential "well", and the less sensitive the HRG will be to small stem rotations.

As compared to mechanical gyros, HRGs will be less failure-prone as they have no moving parts. They tend to be very rugged as there is only a weak interaction between the shell vibration and gyro mounting. They also require very little power.[14] For example, one Delco HRG design only requires power to be applied every 10–15 min. In the event of an unexpected power hit, an HRG will tend to be very forgiving as the conservation of angular momentum on which it is based and the low-friction environment in which it lives will cause the wine glass vibration to be maintained, just as a mechanical gyro's spinning top will tend to maintain its rotation rate if power is anomalously interrupted for a brief time. On the downside, HRGs are expensive to manufacture because the quality of their performance is limited by the precision with which the shell and housing are machined, and by the quality of vacuum imposed and the success of the seals in maintaining that vacuum. Unlike a mechanical gyro, correlating the gyro output (i.e., movement of the anti-nodes axis) with the physical spacecraft behavior (i.e., the movement of the stem) is a pretty elaborate calculation, requiring its own computer support.

All of the previous gyros we've discussed are fairly directly based on conservation of angular momentum. By contrast, the principal behind an optical gyro is the Sagnac effect, which is more related to the interference phenomenon characteristic of a Michelson interferometer. Because of differences in the paths traveled by the two light beams, you get interference fringes when you rotate a source sending two light pulses out along a ring in opposite directions (see Fig. 4.16).

If the path lengths are constructed to be identical when the table drive wheel isn't rotating, the measured intensity when you recombine the beams will be a maximum, i.e., the light waves remain in phase. Once you start rotating the interferometer (in Fig. 4.16, about the axis perpendicular to page), the light waves paths will differ (the wave in the rotation direction will have a longer path to travel), and the light waves will get out of phase.

[14] Lawrence, Modern Inertial Technology, p. 160.

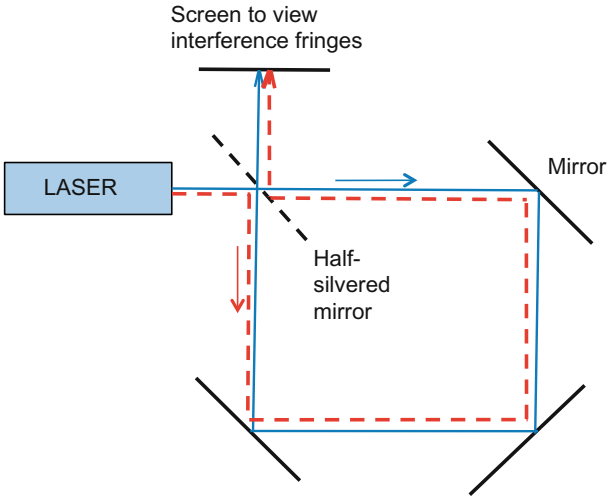


Fig. 4.16 Sagnac optical gyro

You could measure the departure from the maximum to determine how fast the interferometer was rotating, but since the slope of the response curve is zero at zero rate, you'd need to experience large changes from zero rate before you'd obtain a detectable output. Also, you'll encounter a sign ambiguity around maximum since you won't be able to tell the difference between a decrease in intensity due to a negative phase difference versus a decrease in intensity due to a positive phase difference. On the other hand, if you operate the interferometer with the wheel spun-up to the rate at which the phase difference between the two beams is 180°, the intensity will go to zero, an easy to distinguish state.

Unfortunately, you'll still have low sensitivity due to the behavior of the response curve and you'll still suffer a sign ambiguity around null since you won't be able to tell the difference between an increase in intensity due to a phase decrease versus an increase in intensity due to a phase increase. The best plan is to rotate the wheel drive so that zero external rotation of the interferometer as a whole leaves the light beams 90° out of phase. That way, you avoid the sign ambiguity (a phase decrease generates an increase in intensity, while a phase increase generates a decrease in intensity) and get maximum sensitivity by operating in the near-linear region.

Sagnac interferometers tend to be very noisy over short time durations (i.e., seconds), but improve by an order of magnitude over long durations (i.e., hours).[15] The primary noise source is the electrical noise (called *photon shot noise*) caused when light falls on a detector. Another problem is that the interferometer has low sensitivity if the light only travels around the path once. This issue is easily dealt with by replacing the mirrors shown in Fig. 4.16 with an optical fiber coil, by which you can exploit the principle of total internal reflection to send the light beams around

[15]Lawrence, Modern Inertial Technology, p. 174.

their paths many times. By this means, optical gyros have the potential to deliver superior performance to mechanical gyros in high-rate applications.

One class of optical gyro that has flown on GSFC missions is the Interferometric Fiber-optic Gyro (IFOG). Its stability is not as good as high quality mechanical gyros, as its drift biases can get as large as 1 deg/hr. Compared to another class of optical gyro, the RLG (which we'll talk about next), the IFOG is more rugged and has a longer expected lifetime. The limiting noise source is photon shot noise, which is inversely proportional to the square root of the sampling time. The maximum rate measurable by a given IFOG is limited by the angular change (over an interrogation period) needed to shift the IFOG reading from a 90° phase difference to a 180° phase (i.e., near null). As an example, for one IFOG having a coil of area 100 cm^2 with 1000 turns of fiber optic cable, the maximum rate that could be measured accurately and without ambiguity was 150 deg/sec.[16] Clearly this wouldn't be an issue for point-and-shoot spacecraft with slew speeds less than 10 deg/min, but other vehicles operated by non-NASA agencies can experience rates as high as 500 deg/sec.[17] And while we're talking about displacements from the linear region at 90° of phase shift, the use of fiber-optic cable provides a much cleaner way to shift the linear region to the interferometer rest state than spinning up a drive wheel. Instead, you can wrap some of the fiber-optic cable around a piezoelectric cylinder. Then by applying an alternating current to the cylinder you can alternately stretch and relax the cable. By this means, you can selectively generate a path shift of a quarter of a wavelength (of light), thereby shifting the phase difference at rest from 0° to 90° without use of a drive wheel.

The second type of optical gyro, the RLG, uses a laser to generate a standing wave. The pure, coherent light emitted by a laser is produced via stimulated emission. A somewhat oversimplified explanation of the process is that you're trying to generate a light beam consisting of a collection of photons tightly concentrated in frequency, phase, and direction. The photons come from atoms that have been raised to an excited energy state, then induced (by stimulating it with a another photon) to emit a photon of well defined frequency as it (the atom) transitions back to its original energy state. The phase of the emitted photon will be the same as that of the stimulating photon, i.e., the emission will be coherent. So by appropriate choice of target atoms, we can fix the frequency of the emitted photons, and, by appropriate selection of the phase for stimulating radiation, we can fix the phase of the emitted radiation. Then if we place an atomic gas in a resonator whose length is an integral number of half-wavelengths of the photon wavelength we're working with, we can establish a standing wave of that light frequency in the resonator cavity, provided the resonator gain (a measure of photons retained within the cavity versus photons lost) is sufficient to exceed losses from collisions with the walls, collisions with unexcited atoms, etc. Gas lasers designed as described (crudely) above emit very coherent light with small beam divergence, but require high operating voltages (600–1000 V) and are fairly

[16] Lawrence, Modern Inertial Technology, p. 189.

[17] Ibid.

large (2–20 inches long).[18] Further, in order to form the standing wave successfully, a gas laser's mirrors must be aligned to within a few arcseconds and the gas must be extremely free of contamination. By contrast, semiconductor lasers can be smaller, cheaper, and more easily controlled, but suffer a relatively higher level of beam divergence, requiring astigmatic optics to couple their output into fiber-optic cables or waveguides.

An RLG designed by Sperry uses three to five mirrors to bounce light around the resonator, forming standing wave patterns. As "viewed" by the light wave (has sort of an Einsteinian flavor about it), the standing wave is stationary while the spacecraft moves around it. The RLG's optical detector then senses the intensity of the light at a point in the standing wave, so the RLG can interpret changes in the measured output intensity in terms of the rotation of the RLG case relative to the standing wave. The resolution of the measurement can be on the order of an arcsecond. Whereas shot noise was the dominant error source limiting IFOG performance, spontaneous emission is the primary error source for RLGs. Like shot noise, spontaneous emission is inversely proportional to the square root of the sampling, which finds its corollary in gaussian noise being inversely proportional to the square root of the number of samples. Quantization noise also is a significant error source for RLG measurements (not a problem for IFOGs).[19]

[18] Lawrence, Modern Inertial Technology, p. 210.

[19] Lawrence, Modern Inertial Technology, p. 223.

Chapter 5
Attitude Actuators

One of the amusing qualities of human civilization is the hierarchical structure that is common to most problem solving activities in terms of the sophistication of assets applied at different points in the process. At a high level, there are five major steps in solving a problem, namely

1. recognizing there is a problem
2. collecting data to characterize the problem
3. analyzing the data to identify the nature of the problem
4. developing a plan to solve the problem
5. implementing the solution

Here's a good historical example of how this works in real-life. Soon after Balboa discovered the Pacific coast of Panama, commercially-minded people realized (Step 1) that a lot of time could be saved when sailing from the Pacific Ocean to the Atlantic Ocean, or vice versa, if a shortcut in the form of a canal was built across the isthmus of Panama. In 1835, President Andrew Jackson commissioned a Colonel Biddle to report on the feasibility of such a construction, and many other surveys followed (Step 2). After an unsuccessful attempt by a French company to build a canal purely with commercial profit in mind (abortive effort at Steps 3 to 5), the concept of a canal got a new shot in the arm (no Yellow Fever pun intended) as a result of the Spanish American War. In 1901 President Theodore Roosevelt, who was much impressed by the fact it took the Battleship Oregon 67 days to sail from San Francisco to the Caribbean back in 1898, recognized that Americas ability to project force to defend its interests would always be hamstrung until a quicker means to shift ships from Atlantic to Pacific, or vice versa was developed (back to Step 1). From 1901 to 1905, serious work began preparing to do the job (Steps 2 and 3), with important supporting milestones achieved such as instigating (or at least implicitly supporting with a wink and a nod) a revolution to separate Panama from Columbia (1903) and eradicating Yellow Fever in what would be called the Canal Zone (1905). In 1906, the engineering team revived and revised a backup French plan to use locks

© Springer-Verlag London Ltd. 2017

H.L. Hallock et al., *ACS Without an Attitude*, NASA Monographs
in Systems and Software Engineering, DOI 10.1007/978-1-4471-7325-0_5

to cross the Panamanian mountains and deal with the elevation difference between the Pacific and Atlantic sides (Step 4). (The original French plan was to cut a sea-level canal across the entire isthmus.) At that point, American construction efforts began in earnest (Step 5) and were completed in 1914.

Now let's talk about the hierarchical structure mentioned earlier. Ignoring the unsuccessful first effort by the French company, from 1835 to 1901 some of the most senior and prestigious individuals in American society became convinced of the need to attack the problem of slow sea travel between the Atlantic and Pacific Oceans (Step 1). Much of that period (and a few years more) was occupied by data collection and data analysis (Steps 2 and 3). This work was overseen and performed by senior highly educated individuals, i.e., statesmen, diplomats, surveyors, engineers, doctors, etc. At the same time, plans were developed to attack the problem (Step 4), including a diplomatic initiative (the revolution), a medical initiative (Yellow Fever eradication), and an engineering design (including the locks concept). These plans also were developed by senior, highly educated, prestigious individuals. Finally, the engineering plan for building the canal was implemented (Step 5) and carried out. How was Step 5 carried out and by whom? It was dug by a virtual army of uneducated and non-prestigious folks carrying shovels. In the case of the failed French effort, the army totaled 20,000 people (200 died per month from Yellow Fever).

So what does all this have to do with a spacecraft's ACS subsystem? Well, the problem recognition part (Step 1) is pretty much implicit in the fact that we have an ACS (of a specific design) to deal with pointing control in the first place, which derives from extensive trades performed by system engineers early in the mission lifetime. The information gathering component (Step 2) is the set of ACS sensors we discussed in Chap. 4. Some of these sensors are rather sophisticated. For example, a quaternion star tracker can not only provide direction measurements of stars it has acquired and tracked, it also can identify them and determine the transformation from inertial space to sensor space utilizing its own built-in star catalog for reference information. The analysis component may include powerful state estimation algorithms (usually a Kalman filter, details of which will be discussed in Chap. 7) and control laws that estimate the pointing error (Step 3) and figure out how much error should be removed each cycle (Step 4), and by which attitude actuators (Step 4), in order to maintain the current pointing or reorient the spacecraft to a new one. Finally, it comes time to implement the solution (Step 5). That's where the actuators do their bit to give us the spacecraft pointing we want, as opposed to the one we've got, or at least think we've got.

And what do we have to do to make that happen? We can turn a wheel faster (or slower), run some current through a coil of wire, or open up a valve to squirt some propellant out. Doesn't sound very impressive compared to what happened in all the steps leading up to the implementation, does it? In fact, there is a bit more to it than it seems. For example, we do want to be very careful about the way we spin the wheel up (or down) so we don't inadvertently impart jitter to the spacecraft as we carefully model all the frictional effects inherent in a rotating object to make sure we get the motion we intended out of the wheel. Also, we don't want to spin the wheel up too quickly (and overshoot the target) or too slowly (and take too long to

reach the target). And in order to get the thruster to do what we want, we may have to start up a catalyst bed heater, factor in a time delay for things to warm up, model the ramp-up in thruster performance, calibrate the centroid direction, model thruster impingement effects, etc. So nothing really is easy in this business. But probably the guys with the shovels in Panama saw things a bit differently from folks looking back on their accomplishments nowadays. So let's return to the early twenty-first century from the early twentieth century and talk about three types of attitude actuators, namely reaction wheels, magnetic torquer bars, and thrusters.

5.1 Reaction Wheels

A reaction wheel can be viewed as a spinning wheel. The magnitude of its angular momentum is simply expressed as the product of its moment of inertia and its speed (i.e., $L = I_{wheel}\omega_{wheel}$). The direction of the wheel angular momentum vector is the same as that of its angular velocity vector (i.e., right-hand rule applied to the way the wheel is spinning). Conservation of angular momentum is the key to understanding why changes to the wheel's speed produce changes in the spacecraft's motion. Suppose we have a simple box-shaped spacecraft that contains a reaction wheel mounted such that its spin axis is parallel to a spacecraft axis of symmetry and the wheel center of mass is coincident with the spacecraft's center of mass (see Fig. 5.1).

Suppose also that initially the spacecraft is fixed in inertial space and the wheel is not rotating. Now set the wheel rotating at a rate labeled by ω_{wheel}. Since the angular momentum of the entire spacecraft (the spacecraft structure plus the wheel) initially was zero, conservation of angular momentum demands that even when the wheel starts rotating the angular momentum of the entire spacecraft must remain zero. So the spacecraft body will have to start rotating in the opposite direction to the wheel in order to "cancel out" the wheel's angular momentum. How fast will the spacecraft rotate? Well, if the principal moment of inertia of the wheel about its spin axis is I_{wheel} and the principal moment of inertia of the spacecraft about that axis is I_{space}, then to keep the total angular momentum equal to zero we must satisfy the condition

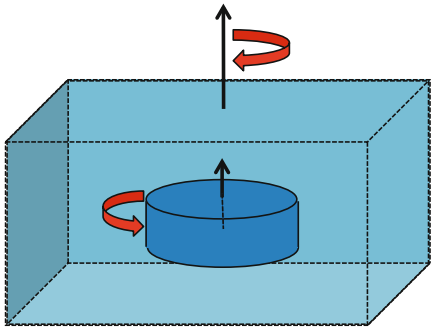


Fig. 5.1 Exchange of angular momentum between wheel and body

$\omega_{space} = (I_{wheel}/I_{space})\omega_{wheel}$ where ω_{space} is the rotation rate of the spacecraft about the wheel's rotation axis.

In fact, that's exactly what we do if we want to slew the spacecraft from one pointing to another one via reaction wheels. If our spacecraft has at least 3 reaction wheels, appropriate torque commands to the reaction wheel set can cause the spacecraft to begin rotating about any eigenvector we choose. To stop the spacecraft rotation when we get to where we want to be, just return the wheel speeds to where they were before the change was commanded. Of course, we really need to do more than this. Errors in modeling wheel frictional effects will cause us not to get the exact response from the wheels we wanted. Errors in wheel positions and alignments will cause the spacecraft not to respond exactly as we expected. And on top of these errors, the rest of the universe will stick an oar in the water via the perturbative torques we discussed earlier (atmospheric, gravity gradient, and solar radiation torques, etc.). That's why for this simple slewing strategy to work properly, we really need a feedback mechanism so that we can measure the error in the spacecraft pointing and order wheel commands to cancel out the errors.

Clearly, once we bring environmental torques into the picture, this feedback system will be needed just to maintain the spacecraft's inertial pointing at a constant attitude, let alone control large angle slewing. For Earth-pointing spacecraft, we're not trying to maintain our current pointing or even slew to a new pointing. Instead, we're trying to maintain a nadir pointing (i.e., stay looking down at the Earth), even though the spacecraft's orbital motion keeps changing the relative orientation between the spacecraft and Earth. For this case, we need to spin-up the wheel and keep it spun-up so that the body of the spacecraft responds by rotating such that it keeps its imaging instrument pointed at the Earth. This is called 1 RPO operation since rotating the spacecraft about the appropriate spacecraft axis (usually defined as the pitch axis) at 1 revolution per orbit (1 RPO) will maintain its pointing geometry relative to the Earth throughout each of its orbits .

The actual structure of a reaction wheel is quite a bit more complicated than Fig. 5.1's simple picture. In addition to a flywheel and its housing, a reaction wheel must have a torque motor and electronics for driving the wheel, as well as a tachometer to measure and output the wheel's speed. Wheel speeds are measured via a magnet mounted on the rotating wheel and a pickoff coil mounted on the fixed housing. Maximum wheel performance is often expressed in terms of maximum wheel angular momentum capacity instead of maximum wheel speed (recall the linear relationship between the two). Typically, wheel capacities range from 0.4 to 40 N-m-sec.[1] Usually these angular momenta are delivered by small, high-speed wheels as opposed to large, slow speed ones because of the advantages gained during spacecraft weight trades. The last piece of the wheel structure is the bearing assembly, which enables the wheel to rotate without seizing up. The bearings can be lubricated via low vapor pressure oils, Teflon, or magnetic suspension.

[1] Larson & Wertz, Space Mission Analysis and Design, p. 355, Table 11-11.

Wheel friction originates from three different sources, namely stiction, Coulomb friction, and viscous friction. *Stiction* is also called static friction. It is the friction that arises when you try to move (either translationally or rotationally) an object at rest. *Coulomb friction* (also called *kinetic friction*) is a constant friction that inhibits the movement of a moving object. Stiction and Coulomb friction can be a particular problem for a spacecraft that needs to be held extremely steady to do its science because even when the control system is very carefully designed (e.g., through use of a momentum bias, as opposed to a zero momentum system, or via a set of selection laws) there still may be cases where a wheel speed passes through zero. Since friction opposes the direction of motion, you get a sudden shift in direction of the frictional torque when the wheel speed changes sign, which can deliver a significant jolt to the spacecraft, upsetting its fine pointing. And a jolt isn't such a great thing for wheel longevity either. *Viscous friction*, by contrast, is linearly proportional to the wheel speed. So viscous friction is not an issue during zero-crossings, but it increases in magnitude as the wheel speed increases.

As hinted at earlier, to be able to maintain and reorient the spacecraft attitude via wheels with full flexibility, the spacecraft needs to employ more than one reaction wheel. Earth-pointers used to employ a two-wheel system, placing one wheel on the pitch axis and the other wheel on either the roll or yaw axis. This system took advantage of the fact that relative to inertial space the spacecraft roll and yaw axes interchange every quarter orbit (see Fig. 5.2). So, for example, a roll-mounted wheel could fully correct spacecraft roll pointing errors (relative to inertial space) at one point in the orbit, while a quarter an orbit later it could fully correct yaw-pointing errors (relative to inertial errors). In actuality, throughout the orbit, corrections are made via the roll wheel to a linear combination of roll and yaw errors that on average over the duration of the orbit would address both errors equally. Meanwhile, the pitch wheel exclusively dealt with pitch errors, i.e., rotational errors about the orbit normal axis.

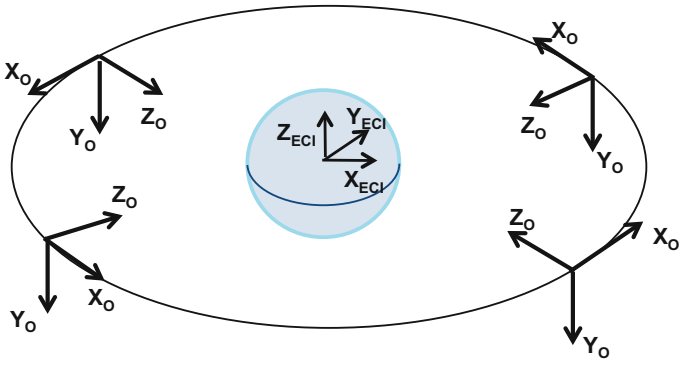


Fig. 5.2 Interchange of roll and yaw axes in 1-RPO motion

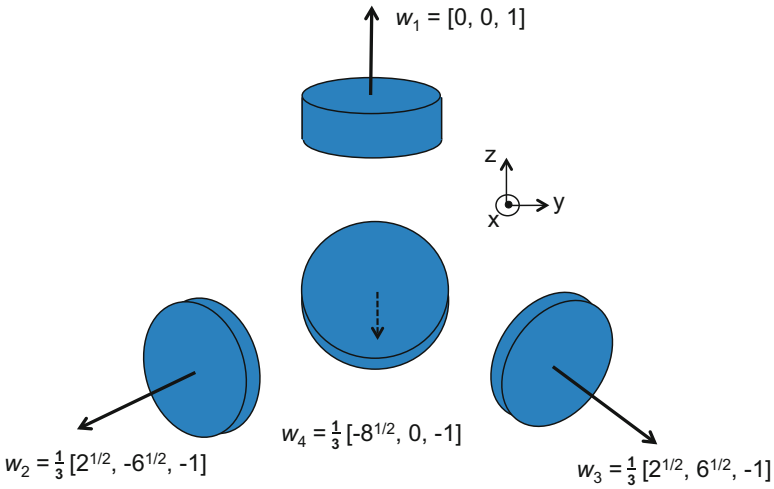


Fig. 5.3 Four reaction wheel system

By contrast, celestial-pointers must carry at least three wheels to be able to slew in any direction as well as correct attitude errors arising from all disturbance torques. Usually celestial-pointers fly at least four wheels (see pyramidal distribution scheme in Fig. 5.3), which both provide redundancy in case of a wheel failure and enables momentum distribution through use of a distribution law so as to inhibit wheel saturation.

Wheel saturation occurs when you've cranked a wheel up to its maximum speed, but still have attitude error that needs to be corrected. Without recourse to some other actuator that can allow dumping of angular momentum from the saturated wheel, the wheel will no longer respond to torque commands and loss of control of the spacecraft can result. So whenever you control the spacecraft attitude with reaction wheels, you always need another attitude actuator, either magnetic torquer bars or thrusters, to protect the wheels from saturation. But before we move on to those other actuators, let's summarize the advantages and disadvantages associated with using reaction wheels, starting with the advantages. Unlike thrusters, whose usable lifetime is limited by their fuel supply, wheels utilize renewable electrical energy, so their lifetime is limited only by mechanical failure. The fact that it is a big moving part might suggest that failures could be frequent, but in fact wheel failures have been infrequent on GSFC missions. Commanding wheels is pretty straightforward. It usually just involves a simple transformation of control law-computed body-frame corrective torques to corresponding torques in the wheel frame, with a distribution law thrown in for grins to help inhibit wheel saturation. For missions that don't change inertial pointing for a long time, it may be desirable to use a more complex set of logic to avoid wheel zero-crossings, or at least pass through them quickly. A system momentum bias that keeps all the wheel speeds high enough that they don't get close to zero may achieve the same results. The biggest disadvantage of

wheels is the saturation problem, which we'll also touch on some more in the subsections on magnetic torquer bars and thrusters. The other major disadvantage is that compared to some types of thrusters, wheels can be very slow, which can be an issue for science missions requiring quick response.

5.2 Magnetic Torquer Bars (MTBs)

An MTB is the simplest of the three attitude actuators. By running an electrical current through a coil you generate a magnetic dipole moment:

$$\mathbf{m} = NIA\mathbf{n} \tag{5.1}$$

where

$\mathbf{m}$ = magnetic dipole moment (ampere-meter^2)
N = number of wire coil loops (integer)
I = current in coil (amperes)
A = coil cross-sectional area (meters^2)
$\mathbf{n}$ = unit vector normal to the plane of the coil

The magnetic dipole can be thought of heuristically as a little bar magnetic (see Fig. 5.4) whose axis is aligned with the coil's symmetry axis (or for a single loop, parallel to the normal to the loop with the polarity of the bar defined by right-hand rule applied to the sense of the current flow).

If you immerse the bar magnetic in an external magnetic field, the bar will "feel" a torque that tries to align the bar with the field direction. By the same token, when you run a current through a coil mounted on the spacecraft (actually the coil is wound around a ferrous bar mounted on the spacecraft), you generate a dipole moment that

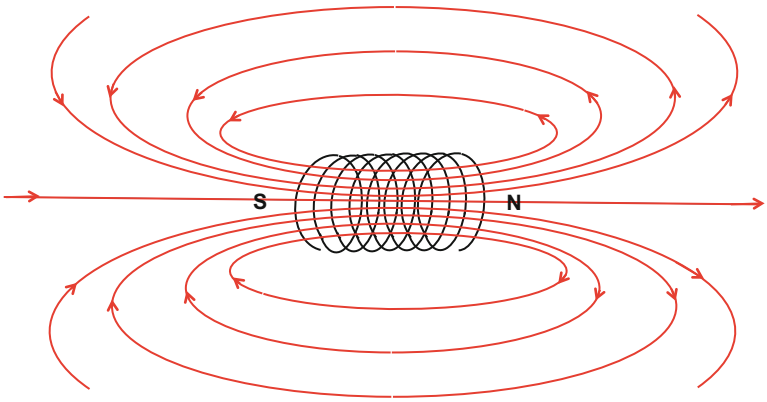


Fig. 5.4 Magnetic field generated by a current coil

"wants" to align itself with the geomagnetic field, the external field in which the spacecraft is immersed. Since the MTB can't move without taking the spacecraft with it, the net effect of commanding the current in the coil is to apply a control torque to the spacecraft (see Eq. 5.2).

$$\mathbf{N} = \mathbf{m} \times \mathbf{B} \tag{5.2}$$

where

$\mathbf{N}$ = torque (Newton-meters)
$\mathbf{B}$ = Earth's magnetic field vector (gauss)

On some small, low cost spacecraft with undemanding attitude control requirements, MTBs have been used as the sole attitude control device, although their performance flexibility is limited by the fact that they interact with the existing geomagnetic field at that time, which constrains the direction in which torques can be applied. Specifically, the direction of the applied torque must always be perpendicular to the direction of the Earth's magnetic field at that time. This is not much of a limitation when using MTBs for angular momentum dumping since we're not too concerned in that application about how much angular momentum we're dumping at a given instant in time. Instead, what's important is how much angular momentum is dumped over an extended time duration, such as the orbital period of a LEO spacecraft.

If you want to dump angular momentum in the most energy efficient fashion, a possible strategy would be just to counter the external torque secular terms with MTBs and allow the wheel speeds to oscillate sinusoidally in response to external torque periodic terms. (For LEO spacecraft above 500-km altitude, the dominant environmental torque is gravity gradient, which has both secular and periodic terms.) This strategy for angular momentum management, called *minimum energy momentum management*, was actually proposed for and implemented in HST's FSW, but was found to be unnecessary even pre-launch when the power subsystem was upgraded with better batteries having superior power margins. In practice for most missions, since electrical energy is usually plentiful, the simplest thing to do from an overall costs standpoint (factoring in the expense of FSW) is to generate an MTB dipole command based on current satellite conditions (attitude and reaction wheel momentum load) and the currently measured geomagnetic field, e.g.,

$$\mathbf{m}_C = k_G(\mathbf{B}_v \times \mathbf{N}_d) \tag{5.3}$$

where

$\mathbf{m}_C$ = magnetic dipole to be commanded
k_G = control gain factor
$\mathbf{B}_v$ = local, TAM-measured geomagnetic field (in body frame)
$\mathbf{N}_d$ = desired torque to reduce reaction wheel momentum load

A third application of MTBs is rate damping using the *B-dot* control law ("dot" being short-had notation for time derivative). B-dot was first used to place the SMM

spacecraft into a controlled rotation after a wheel failure put the spacecraft into a tumble. The B-dot algorithm simply read TAM measurements and generated commands to the MTBs to null the observed change in magnetic field. Of course, the geomagnetic field actually does change (as measured from the spacecraft) as the spacecraft travels through its orbital path, so application of B-dot cannot inertially fix the spacecraft attitude. But it will (very gradually because of the relative weakness of magnetics) convert an extremely complex attitude motion into a controlled, slow, periodic one, roughly having a frequency of twice the orbital one.

Since MTBs have no moving parts, failures are rare. There are no artificial limitations on lifetime either since MTBs are run on electrical power, a renewable resource that usually is readily available. Because they're lightweight, MTBs don't impact weight trades significantly, and at LEO orbits the significant size of the geomagnetic field (about a quarter of a gauss) make MTBs the ideal desaturation partner for reaction wheels. On the downside, MTBs are typically the weakest of the attitude actuators, or at least they are most limited in terms of the maximum torque they are capable of generating. Even at LEO altitudes, the action of MTBs will be extremely slow, and once you get above synchronous altitude, the Earth's field is so weak that MTBs are pretty much useless. Finally, the magnetic field generated by the MTBs can corrupt or interfere with other spacecraft activities, such as the performance of science-oriented magnetometer measurements.

Before closing out this section on MTBs, we find it appropriate to throw in a caveat regarding magnetic conventions. One possible source of confusion in dealing with magnetic quantities is that they follow two different conventions, which give rise to two different definitions of magnetic pole strength and dipole moment, expressed in two different sets of units. In order to keep the units straight, it's important to know which convention you're using. One convention is based on the magnetic induction vector $\mathbf{B}$ (which has units of teslas in SI units). Under this convention, the force $\mathbf{F}$ on a (hypothetical) magnetic pole q^* in a magnetic field $\mathbf{B}$ is given by

$$\mathbf{F} = q^*\mathbf{B} \tag{5.4}$$

Here q^* is called the pole strength, with units of ampere-meters (Am) in this convention. A magnetic dipole can be imagined as two magnetic poles of equal pole strength (but opposite polarity) separated by a fixed distance. The magnetic dipole moment $\mathbf{m}$ is defined as the product of one of the pole strengths multiplied by the distance between the poles; under the $\mathbf{B}$-field convention, this is

$$\mathbf{m} = q^*\mathbf{L} \tag{5.5}$$

and the units of magnetic dipole moment are Am^2. Here $\mathbf{L}$ is the separation vector between the two magnetic poles, and points in the direction from the south pole to the north pole.

Under the other convention, these quantities base their definitions on the magnetic intensity vector $\mathbf{H}$ instead of the magnetic induction field $\mathbf{B}$. In vacuum, the $\mathbf{H}$ field is related to the $\mathbf{B}$ field by $\mathbf{B} = \mu_0\mathbf{H}$, where μ_0 is the permeability of free space,

and is equal to $4\pi \times 10^{-7} N/A^2$. The **H** field has units of *A/m*. Under the **H**-field convention, the force **F** on a magnetic pole Q^* in a magnetic field **H** is given by

$$\mathbf{F} = Q^* \mathbf{H} \tag{5.6}$$

where Q^* is the pole strength, which under this convention has units of webers (Wb). Another common unit for pole strength under the H-field convention is the unit pole. It is used with cgs (centimeter-gram-second) units, and is defined to be 1 dyne per oersted, where the oersted is the cgs unit for **H**. One unit pole is equal to $4\pi \times 10^{-8}$ webers. Under the **H**-field convention, the magnetic dipole moment **d** of two magnetic poles separated by a distance **L** is given by

$$\mathbf{d} = Q^* \mathbf{L} \tag{5.7}$$

where the dipole moment **d** has SI units of weber-meters (Wb m) under this convention. Another common unit for magnetic dipole moment under the **H**-field convention is pole-cm, and is used with cgs units. One pole-cm is equal to $4\pi \times 10^{-10}$ Wb m. To convert magnetic pole strength between the **B**-field and **H**-field conventions, one may use

$$Q^* = \mu_0 q^* \tag{5.8}$$

The conversion between conventions for the magnetic dipole moment is

$$\mathbf{d} = \mu_0 \mathbf{m} \tag{5.9}$$

This is all a somewhat lengthy way of saying that different folks have different strokes (conventions), so be careful that you know what's being used by the teams for any given spacecraft on which you're working.

5.3 Thrusters

At last, we get to talk about something that sounds like rocket science: thrusters. There are three distinct classes of thrusters that we'll talk about here: hot gas, cold gas, and ion. They all have at least one thing in common. They produce thrust through the collective acceleration of propellant molecules. This may occur via a chemical reaction, thermodynamic expansion, or electrostatic fields, hence the three thruster classes.

With hot gas thrusters, the energy for propulsion is derived from a chemical reaction. Relative to the other two classes, hot gas thrusters produce a higher level of thrust and a greater impulse (the time integral of force). For GSFC spacecraft, the

force generated can be greater than 5 newtons,[2] although smaller thrusters are often employed for higher precision in applied force or torque. The fuel for hot gas thrusters is either bipropellant or monopropellant. Monopropellant fuels, such as hydrazine or hydrogen peroxide, use a catalyst or high temperature to cause decomposition of the fuel. The most common fuel for hot gas thrusters flown on GSFC spacecraft is hydrazine with catalytic decomposition.

Let's examine those components in an operational order, starting off with the fuel tanks where the hydrazine and pressurant are stored, with a diaphragm separating the two. The pressurant is nitrogen gas, which helps push the hydrazine out of the tanks. As fuel is depleted from the tank, the gas pressure steadily diminishes, so that the performance of the thruster near end-of-life will be much less than its performance at launch. On leaving the fuel tank, the hydrazine passes through filters and latch valves as it heads towards the thrusters, with a pressure measurement from pressure transducers along the way. Heaters in the fuel lines are used to keep the hydrazine from freezing. Next, the hydrazine is filtered again and its flow through the lines is controlled via solenoid valves. Once the fuel hits the thruster, the propulsion process finally can occur, with thruster heaters used to maintain proper operating temperatures, as measured by thermocouples. And of course, the execution of the entire process is monitored, controlled, and reported on by propulsion system electronics.

There's a ramp-up time before a hydrazine system is operating at nominal performance. Initial performance can be low until the catalyst is fully heated by operation of the thruster. Hydrazine thrusters are fast and powerful, and their use can be orchestrated via relatively simple control laws, although modeling the thrust profile can be quite complex. Of course, as for all thrusters, thruster lifetime is limited by the fuel supply, in this case the supply of hydrazine onboard. Unlike small wheels and MTBs, hydrazine systems are heavy, so the size of the tank and the amount of fuel to be carried are big considerations in weight trade studies. Where you place gas thrusters and how you align them is a big deal also. For example, the plume from the thruster may impinge on spacecraft structure, such as a mast, leading to unexpected and undesired torques on the spacecraft. Even worse for many science spacecraft, the ammonia in hydrazine is very corrosive and can damage sensitive optics. Also, since hydrazine freezes below zero degrees Celsius, you should include a lot of heaters everywhere the hydrazine travels through the propulsion system. Otherwise you can end up with a permanent blockage that could disable the thruster resources.

Although this seems like a lot of stuff to worry about when designing a spacecraft, hydrazine thrusters remain extremely popular and are capable of performing just about every translational or rotational function imaginable. They can control attitude and spin rate. They can damp out nutation and keep reaction wheels from saturating. And they can perform delicate stationkeeping, or insert your spacecraft into an entirely new orbit. For attitude purposes, hydrazine thrusters operate in pulsed-mode, where the thruster is periodically cycled on and off. For orbit maneuvers, the thrusters are operated in continuous mode, with the thrusters starting firing at the beginning of the maneuver and kept in non-stop operation until the end of the maneuver.

[2] Wertz, p. 206.

The alternative to monopropellant fueled thrusters, like hydrazine thrusters, within the hot gas thruster class is bipropellant fueled ones. Since the fuel and oxidizer are stored separately, clearly bipropellant systems will be more complex than monopropellant systems. Bipropellants (such as mono-methyl-hydrazine and nitrogen tetroxide) are hypergolic, a 10-dollar word that means that when the fuel and oxidizer are mixed they combust spontaneously. This enables the bipropellants to deliver much higher thrust levels than monopropellants, up to 500 Newtons of force.[3]

One of the few things hot gas thrusters are not designed for is fine attitude pointing, because the precision and consistency obtainable from hot gas thrusters is not comparable to what you can achieve with reaction wheels. That's where the other thruster classes shine.

In the class of high-precision thrusters, cold gas thrusters have been around about as long as hot gas thrusters. They derive their energy from the latent heat released during a phase change or from simple compression if no phase change occurs. Cold gas thrusters operate more consistently than hot gas thrusters, especially in pulsed-mode, because there are no steady state issues to deal with. And they're simpler to operate since no pressurant is needed. On the downside, the forces delivered will be considerably less, typically below 1 newton.[4] Still, they have been used successfully to support attitude control on missions such as the International Sun-Earth Explorer-1 (ISEE-1), where the pointing requirement was only on the order of a degree.

But to achieve truly fine pointing precision, you have to move on to a somewhat Buck Rogers sounding technology called ion thrusters. In addition to high-precision for attitude control, ion thrusters have relatively high thrust exhaust velocity, which proportionately reduces the mass of onboard propellant required to achieve the same total momentum change (linear or angular). Ion jets produce thrust through the acceleration of individual ionized atoms by electrodynamics, whose energy in turn comes from solar cells or self-contained electric generators. The thruster components include an emitter, accelerator, beam neutralizer, and of course a propellant. The emitter ionizes the propellant, examples of which are mercury and xenon, and separates the now positively charged atoms from the stray electrons knocked off the previously neutral matter. See Fig. 5.5a and b for an illustration of the xenon thruster used in JPL's Dawn (a.k.a. Deep Space 1) mission.

The accelerator then electrostatically accelerates (big surprise) the positive ions to the desired velocities. Naturally, you want to form the propellant beam from the positive ions since that's where almost all the propellant mass is located. Finally, the beam neutralizer recombines ions and electrons prior to the propellant leaving the thruster so that the net charge of the propellant returns to its neutral state. Otherwise, as you ran the thruster you'd steadily build up charge on the spacecraft.

Recently, ion thrusters have taken on particular importance for a new series of missions dedicated to experimental verification of predictions arising from General Relativity Theory, especially those involving the detection and study of gravity waves.

---

[3] J. Wertz, Spacecraft Attitude Determination and Control, p. 206.

[4] J. Wertz, Spacecraft Attitude Determination and Control, p. 206.

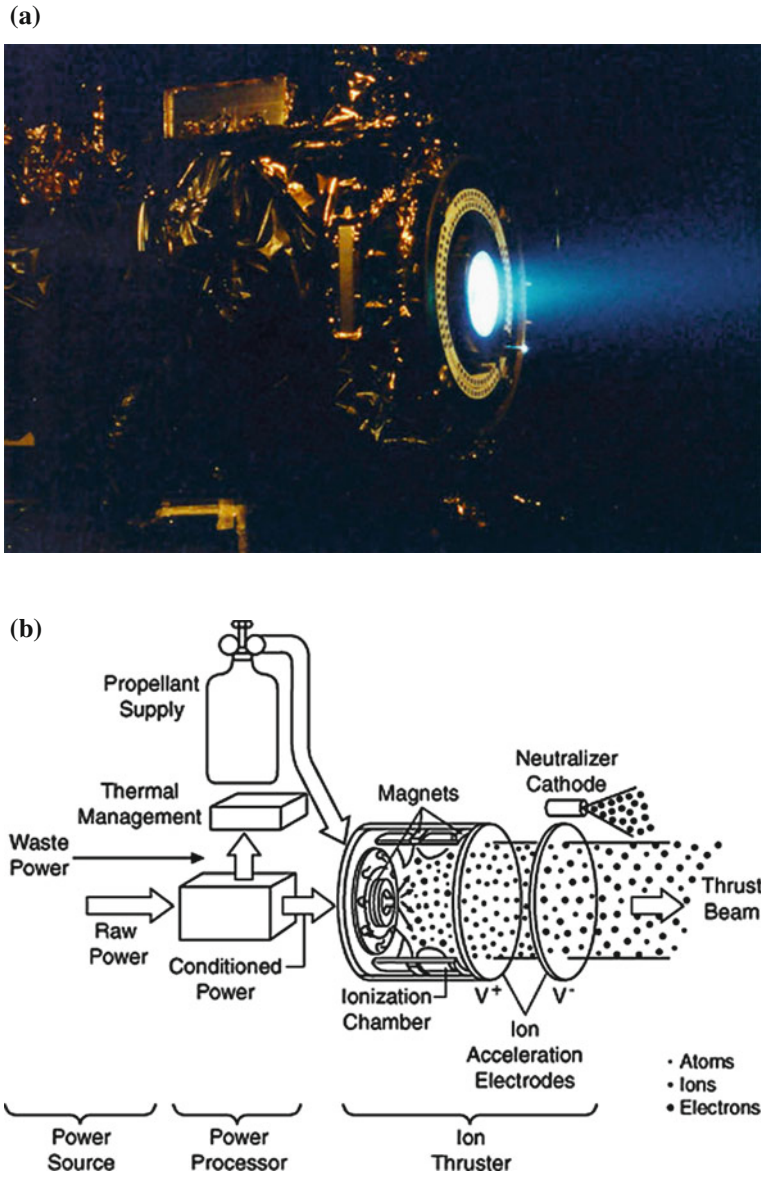


Fig. 5.5 **a** Deep Space 1 (Dawn mission) ion thruster test (Image from http://dawn.jpl.nasa.gov/mission/ion_prop.asp). Credit: NASA/JPL. **b** Deep Space 1 ion thruster schematic (Image from http://dawn.jpl.nasa.gov/mission/ion_prop.asp). Credit: NASA/JPL

For extremely sensitive science of this type, any moving component onboard the spacecraft is a potential noise contributor. So not only aren't any reaction wheels permitted for control purposes, even mechanical gyros are forbidden as well. Future missions such as the Laser Interferometer Space Antenna (LISA) must instead rely on a newly developing micro-newton thruster technology to provide the nanometer level control that will be required for successful detection and study of gravity waves.

One approach for this low-thrust, high-precision ion thruster technology is Field Emission Electric Propulsion (FEEP), under development by the Austrian Research Center Seibersdorf. The FEEPs are part of a larger class of thrusters called "electrospray" thrusters, which operate somewhat like inkjet printers. The propellant in this case is liquid metal indium liquefied by a heater in the thruster (other design concepts utilize liquid cesium). The thruster works as follows. When you apply an electrostatic field between the surface of the liquid indium and an accelerator electrode, the indium responds by forming the cone on the tip of a needle. If the field is made strong enough (a few volts per angstrom), the indium atoms at the tip of the needle are vaporized, ionized, and accelerated to the accelerating electrode. As ions are drawn away, the supply of atoms at the needle tip is replenished continuously by the action of the electrostatic field. The result is a steady stream of ions from the apex of the cone. Unlike the case of the earlier described ion thruster type, the FEEP beam is not made charge neutral prior to exiting the thruster. Instead, a neutralizer may be provided to render the spacecraft charge neutral. Even that may be unnecessary for the LISA mission as it may be the case that other charging effects (photoelectric, solar wind, etc.) may be strong enough to keep the spacecraft neutralized. This hypothesis will likely be tested on LISA Pathfinder.[5] By varying the applied voltage associated with the electrostatic field, you can control the power of the thrust, with minimum levels down as low as fractions of a micronewton. Incidentally, for this class of thruster (as opposed to hot gas or cold gas thrusters), fuel supply is not a limiter on mission lifetime, given that the required FEEP propellant mass is on the order of grams rather than kilograms.

[5]This information on LISA's FEEPs was supplied by Stephen Merkowitz of the LISA Project.

Chapter 6
Reference Models

Tip O'Neill, a former Speaker of the House, once remarked that "All politics is local". One perhaps could generalize his comment to be "Everything is local". In any event, we all certainly have the tendency to view everything that happens around us from our own personal context. And even when we function as physicists, astronomers, or engineers, any measurements we collect are made in the local reference frame associated with the physical space occupied by our measuring device. There are many advantages to viewing and measuring things from a local perspective. If you choose the size of a local region small enough (without getting too tiny and being afflicted by Heisenberg's Uncertainty Principle), things going on in that small region look fairly simple. For example, the surface you're standing on will look pretty flat no matter how complicated the contours of the surface are, so long as you restrict the size of the sampled surface to a small enough area. If you watch how things are changing in the region, the changes will always look linear as long as you sample things fast enough (again, with a nod to Heisenberg, as long as you're not too fast) and don't look at more than a few samples together. Even the laws of physics get simpler (i.e., can accurately be approximated by simpler equations) if the laboratory in which you're conducting your experiment gets small enough. And, of course, the "gedanken experiments" (translation: "thought experiments") that Einstein conceptualized in these tiny laboratories were quite a bit cheaper than the ones most physicists perform in full-sized labs. So given the comfort value of local surroundings, why would you ever leave them? The problem is that eventually, unless you're a hermit living on top the Himalayas, you'll want to communicate your findings to someone or something outside your locality. Or alternatively, you'll want to compare your findings with those obtained by someone or something outside your magic circle. There are some exceptions to this rule of thumb, but if you think about it for a while you'll see there are a lot fewer exceptions than might seem at first glance.

For example, suppose you're sitting in your living room's reclining chair on a cold winter day watching your favorite basketball team get killed. After reaching the resignation state that follows slamming the TV remote control into the floor, your

© Springer-Verlag London Ltd. 2017

H.L. Hallock et al., *ACS Without an Attitude*, NASA Monographs
in Systems and Software Engineering, DOI 10.1007/978-1-4471-7325-0_6

body temperature will begin returning to normal and the room will start feeling cooler, especially a while after you've drowned your sorrows in an ice cold Sam Adams beer. Stated in a more technical sense, you'll have made a temperature measurement in your immediate vicinity and have determined that warmer surroundings would be a pleasant change. But unless you're that stoic hermit on top of the mountain, if you're going to take any practical action to improve your lot in life, you're going to have to communicate the desired change to your house's furnace in terms that the furnace will understand. That means you're going to have to raise the furnace's control dial a couple of degrees higher to get the heat to come on. At a detailed level, that actually means you've sampled data from a personal temperature sensor, processed the data to determine an average current temperature, run a comfort model to determine what the desired temperature is, calculated a delta between the two to determine a desired temperature change in your vicinity, transformed the desired change in your frame to a commanded change in the furnace thermostat frame (utilizing a thermostat model), delivered that command to the thermostat, and moved the thermostat dial or lever to actuate the heating mechanism. And even with this much complexity, we've left out the entire process that takes you from the thermostat to the furnace heating coil, from the heating coil through the fan to the hot air vents, and the closed-loop control loop through which you continuously monitor the results of your action to determine if the thermostat needs to be raised again, if the furnace should be cut off, or if you should kick in the picture tube and put an end to your misery. Maybe this scenario explains why so many people nowadays seem to have living room fireplaces.

In light of this tangential discussion (maybe Tip's adage should have been "No Politics is Simple") you'll probably find that any measurement worth making for any purpose other than self-absorbed enlightenment (and even those exceptions can probably be dealt with by decomposing the person into his/her constituent parts) has little value taken in local isolation. In order to relate that measurement to something else requires a model of the something else to which it is to be related, or a model relating your measurement to someone or something else's measurement. So by a rather roundabout route we arrive at the subject of this chapter, reference models.

In spacecraft attitude work, we're always comparing measurements of where something is to where that something ought to be. For example, a Sun sensor says the Sun direction vector is (x, y, z). The next thing we want to know is what the "correct" direction is. "Correct" in this context can mean two things. First, it can mean the direction we really wanted the Sun vector to have relative to the spacecraft frame, for example normal to a fixed solar array. Ignoring relative alignment issues between the Sun sensor and spacecraft, this comparison does not require a very sophisticated model, and determining the error between measured and desired Sun vector is as simple as a cross-product computation. But if "correct" means what is the actual Sun direction in inertial space, that's a much more complicated question, requiring referral to an analytical model of Sun position, or (for very accurate information) even referral to an almanac. Typically you'll find that the first comparison, measured vs. desired, is the concern of control laws used for error reduction, while the second comparison, measured vs. actual, is the concern of attitude determination algorithms used to define the orientation of the spacecraft relative to an agreed upon reference.

The models used to support calculation of those "actual" vectors are the subject of the rest of this chapter. Specifically, we'll examine the models for describing the Earth's gravitational field, spacecraft ephemerides, solar, lunar, and planetary ephemerides, the Earth's magnetic field, and catalogs specifying star directions, magnitudes, and other exotic information.

6.1 Modeling the Earth's Gravitational Field

Let's start off with how you model the Earth's gravitational field. Equation 2.1 showed Newton's equation of force between two point masses. A conservative force field, which gravity is, can be expressed as the gradient of a scalar function called the potential, i.e., $\mathbf{F} = -\nabla U$. Given the Earth's spheroidal shape, it is natural to describe its gravitational potential (and thereby implicitly its mass distribution) as terms in a spherical harmonic expansion. In this formalism, latitude dependence is characterized by the zonal harmonic coefficients of the expansion and longitude dependence is characterized by the tesseral and sectorial coefficients (see Eq. 6.1).

$$U = -\frac{\mu_e}{r} + B(r, \theta, \phi) \tag{6.1}$$

where

U = Earth's gravitational potential (m^2/s^2)
μ_e = geocentric gravitational constant = $GM_e = 3.986005 \times 10^{14} m^3/s^2$
G = gravitational constant = $6.67428 \times 10^{-11} m^3/(kg\text{-}s^2)$
M_e = Earth mass (kg)
r = distance from the center of the Earth (m)
θ = co-latitude angle (rad)
ϕ = longitude angle (rad)

and

$$B(r, \theta, \phi) = \frac{\mu_e}{r}\left\{\sum_{n=2}^{\infty}\left[\left(\frac{R_e}{r}\right)^n J_n P_{n0}(\cos\theta) + \sum_{m=1}^{n}\left(\frac{R_e}{r}\right)^n (C_{nm}\cos m\phi + S_{nm}\sin m\phi) P_{nm}(\cos\theta)\right]\right\}$$

where

R_e = Earth equatorial radius = 6378.140 km
n, m = indices representing integers
J_n = zonal harmonic coefficients (for example, $J_2 = 0.0010826359$)
P_{nm} = Legendre polynomials
$\{C_{nm}, S_{nm}\}$ = non-zonal spherical harmonic coefficients ("tesseral" for $n \neq m$; "sectoral" for $n = m$)

The low-order zonal terms model the Earth's oblateness and generally are larger than the corresponding tesseral and sectoral terms at comparable orders in the expansion. As a result, the low-order zonal terms are major perturbing sources for near-Earth orbiting spacecraft. Especially at LEO altitudes, the low-order zonals are primary factors influencing orbital period, right ascension of ascending node, and argument of perigee. By contrast, tesseral and sectorial terms can be more important at geosynchronous altitude since those spacecraft "hang" over the same longitude points, so the longitudinal variations do not average to zero over time. Actually, at geosynchronous altitude, solar and lunar gravitational influences are more important than the higher order terms in the (zonal, tesseral, sectorial) expansion, so arguments over relative importance of zonals vs. tesserals and sectorials at geosynchronous take on a bit of a philosophical flavor as far as the spacecraft is concerned. Typically, for attitude-related use (e.g., gravity gradient torque computations), you don't need to expand beyond the J_2 term, i.e., the first zonal term, which is a second-order term (see Eq. 6.1). However, for high accuracy orbit specification, many additional higher order terms need to be kept in the expansion.

6.2 Modeling the Spacecraft's Ephemeris

One good reason for accurately modeling the Earth's gravitational potential, of course, is to enable more accurate modeling of the spacecraft's orbital position, also referred to as the spacecraft's ephemeris. The primary tool for computing spacecraft ephemerides, both definitive and predictive, here at GSFC is the Goddard Trajectory Determination System (GTDS). GTDS utilizes spacecraft range and rate data to determine definitive spacecraft ephemerides. Using the definitive orbit determined from measurements as a baseline, GTDS then uses a collection of models, including the spherical harmonic expansion of the geopotential we just talked about, to generate predicted spacecraft ephemerides. The other models include those of perturbative effects, such as atmospheric drag (with its associated atmospheric density model), solar radiation pressure, and N-body gravitational contributions. For both definitive and predictive orbit runs, the primary output of GTDS is a file specifying spacecraft position and velocity vector pairs in inertial space at a constant, user-specified time interval. An interpolator is used to determine spacecraft position and velocity for times that fall between points in the file. The GTDS output also provides a header that includes information such as the Keplerian osculating elements corresponding to the first vector pair in the file, a set of Brouwer mean elements averaged over the entire file, a parameter characterizing the solar activity at the time the file was generated, and other parameters describing how the file was generated.

Onboard a spacecraft, predictive orbit generators are often used to avoid the space demands presented by a predictive ephemeris file. Instead, a (sometimes) simple mathematical model is used to calculate spacecraft position and velocity vectors from a (sometimes) small number of input parameters uplinked regularly by the ground system. The "freshness" of the orbital predictions is maintained by frequently

uplinking updated input parameters. Many different types of orbit generators have been implemented onboard spacecraft over the decades. Not surprisingly, as the speed and computing power of the onboard computers have increased during this time period, the complexity and sophistication of the orbit propagators have increased correspondingly.

The least complicated orbit propagator you can make is a two-body one treating the Earth as a point mass (i.e., a model having the six Keplerian osculating elements as input), but the inaccuracy of a simple two-body model at LEO altitudes is so high that we aren't aware of any that have been flown on GSFC missions. The next step up is to include Earth oblateness effects (adds in time derivative terms for right ascension of ascending node, argument of perigee, and mean anomaly), but that will still suffer from significant error when atmospheric drag cannot be neglected. The HST onboard orbit propagator dealt with the atmospheric drag problem by adding in a correction term to the mean anomaly that was proportional to the square of the time since epoch, i.e., an acceleration term compensating for frictional effects. It also included an independent input of circular velocity for grins. But to get really accurate onboard orbit prediction, you need something more elaborate mathematically than an elements-based model. For example, the Landsat model utilized Fourier coefficients (good for modeling periodic structures) from which Hermite polynomial coefficients were constructed for interpolation purposes. An extra level of accuracy could be added in the form of a "fudge factor" grid for empirical correction of small, non-periodic-like effects. And in an approach more similar to the GTDS generation of predicted ephemeris vectors, the Solar Anomalous and Magnetospheric Particle Explorer (SAMPEX), and many more recent spacecraft, utilized its own host of physical models (including the Jacchia-Roberts atmospheric density model) to integrate the equations of motion from input of an extended precision position and velocity vector pair. We'll have more on this topic in Chap. 9, as well as a description of the latest approach to onboard orbit determination, measuring your orbit position directly in realtime.

6.3 Modeling Solar, Lunar, and Planetary Ephemerides

Shifting from manmade orbiting objects to natural ones, the orbits of the planets (including the Earth) and our Moon are nearly circular and in the plane of the ecliptic. For example, Mercury has the largest eccentricity at 0.2, while Mars has the largest inclination (from the ecliptic) at 7°. The most detailed specification of planetary and lunar positions is provided in tabular form in *The Astronomical Almanac*.[1] A computer-readable version (at least by large ground system computers and storage facilities) with interpolation capability is provided by GSFC's Solar-Lunar-Planetary (SLP) file, whose information is derived from a Jet Propulsion Laboratory (JPL)

[1] See http://asa.usno.navy.mil/ for a description of The Astronomical Almanac, and contact information for ordering a printed version.

product. Neither of these, of course, is appropriate for onboard use, but the Explanatory Supplement to the Almanac mentioned previously has formatted this information as expansions of orbital elements, which can be utilized in closed-form algebraic expressions in analytic propagators.

6.4 Modeling the Geomagnetic Field

The Earth's magnetic field, **B**, like its gravitational field, can be represented as the gradient of scalar function, with the latter modeled via a spherical harmonic expansion (see Eq. 6.2).

$$\mathbf{B} = -\nabla V \text{ (gauss)} \tag{6.2}$$

where

$$V(r,\theta,\phi) = a\sum_{n=1}^{\infty}\sum_{m=0}^{n}\left(\frac{a}{r}\right)^{n+1}\left[g_n^m \cos m\phi + h_n^m \sin m\phi\right] P_n^m(\theta)$$

where

r = separation distance (m)
θ = co-latitude angle (rad)
ϕ = longitude angle (rad)
a = "conventional" Earth radius = 6371.2 km
n, m = indexes representing integers
g_n^m = Gaussian coefficients
h_n^m = Gaussian coefficients
P_n^m = Legendre polynomials

Note that although a Legendre polynomial expansion is utilized in both Eqs. 6.1 (the geopotential) and 6.2 (the geomagnetic field), the forms of the equations are not the same. This is really just a matter of convention, manifest as a difference in normalization conditions applied to the two polynomial sets. Specifically, the Legendre polynomials P_n^m used in Eq. 6.2 use the Schmidt normalization convention while the Legendre polynomials P_n^m in Eq. 6.1 use the Neumann normalization convention.

Looking at Eq. 6.2 more qualitatively, the first-order term, which models the Earth like a bar magnet, is dominant. The strength of this term, the dipole, drops off as the cube of the distance separation. Except for its title role in a 1950s low-budget science fiction movie (starring Richard Carlson and King Donovan, for the benefit of those who might also have spent their summers watching the Late, Late, Late Show on small black and white TVs), there currently is no experimental evidence of magnetic monopoles, which if they existed would exhibit the inverse square law behavior made famous by Newton's Law of Gravitation. The Earth's dipole term is not constant with time, and appears to be decreasing roughly monotonically at about

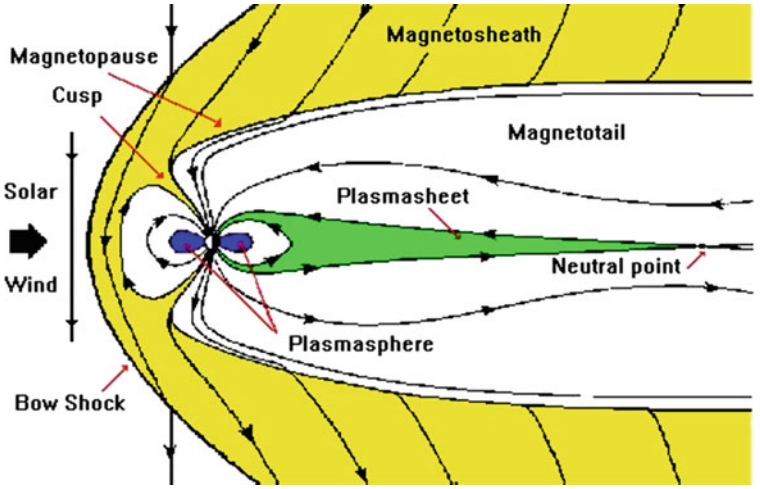


Fig. 6.1 Geomagnetic field as distorted by the Solar wind (image from http://helios.gsfc.nasa.gov). Credit: NASA/Cosmicopia

0.05% per year.[2] This behavior suggests that the polarity of the geomagnetic field may reverse in several thousand years, and in fact there is evidence that several polarity switches may have already occurred, with time spacings of about 70,000–100,000 years between reversals.[3] In terms of the variation in field strength as a function of position relative to the Earth, if you keep your separation from the Earth's center constant, the field strength doubles as you move from the magnetic equator to one of the magnetic poles. This is easy to see qualitatively in terms of the increasing density of field lines as you move towards the poles, as displayed in "lines of force" diagrams first envisioned by Faraday in 1831. Traveling radially from the magnetic equator, the field strength decreases from about a quarter of a gauss at LEO altitudes to just 0.005 gauss by the time you're 4 Earth radii from the Earth's center, roughly 60% of geosynchronous altitude.[4] The actual geomagnetic field is depicted in Fig. 6.1, where the field lines are somewhat distorted from this simple description due to the influence of the solar wind. As a result, Eq. 6.2 is only accurate near the Earth. At large distances from the Earth, the real field will deviate substantially from the spherical harmonic expansion and, because of the solar wind, is a function of time of day.

Figure 6.2 shows the contours of geomagnetic field strength near the Earth's surface, more-or-less the same as for LEO missions. Note that the regions of greatest intensity (towards the extreme northern and southern latitudes) do not coincide exactly with the Earth's rotation poles; the magnetic and rotation poles are offset

[2] J. Wertz, Spacecraft Attitude Determination and Control, p. xxx.
[3] Ibid, pp. 113–114.
[4] Ibid, p. 114, Fig. 5-1.

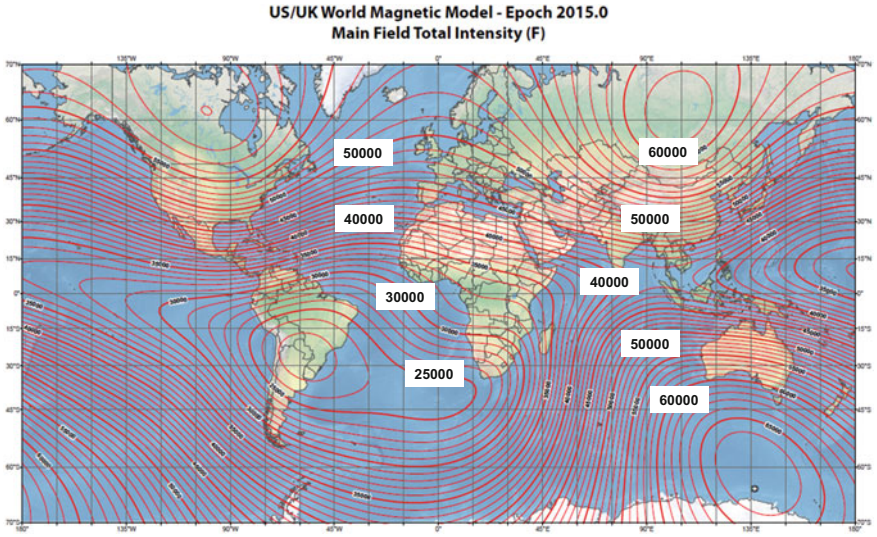


Fig. 6.2 Earth surface geomagnetic field strength, nanoTeslas, epoch 2015. Image from: http://www.ngdc.noaa.gov/geomag/WMM/data/WMM2015/WMM2015_F_MERC.pdf. Credits: NGDC/NOAA (field levels added by authors)

from each other by about 11.5°. Note also the large dip in field strength, down to below 25,000 nano-Teslas, just off the east coast of Brazil. This region is called the South Atlantic Anomaly (SAA), sometimes the Brazilian Anomaly.

Much of this anomalous structure (anomalous relative to a simple dipole picture) can be fit to the second order term in the expansion, the quadrupole term. Since a monopole term varies inversely with the square of the distance separation, and a dipole term varies inversely with the cube of the distance, it won't come as much of a shock to hear that the quadrupole term varies inversely with the fourth power of the distance. The SAA has two features of interest for LEO missions. First, the weak magnetic field in the SAA implies weak coupling of magnetic torquer bars (MTBs) to the Earth's field, and therefore a lowered efficiency for dumping excess angular momentum from spacecraft reaction wheels. This is pretty much just an academic point since MTB-based momentum management is really an orbit-averaged concern; time spent in the SAA is a small fraction of the total orbit. Of more concern is the fact that the weak SAA magnetic field allows the Van Allen radiation belt to penetrate much closer to the Earth's atmosphere there, resulting in high energy (>10 MeV) proton fluxes rising above typical by two to three orders of magnitude. This in turn results in a substantially increased probability for electronics interference, e.g., computer single event upsets.

For most practical applications involving geomagnetic field modeling (for example, coarse attitude determination) accuracy requirements demand that higher order terms in the spherical harmonic expansion be included. The set of expansion coefficients, called Gaussian coefficients (the constants g_n^m and h_n^m appearing in Eq. 6.2),

used here at the GSFC are the standard International Geomagnetic Reference Field (IGRF) ones, which go as high as eleventh order. In addition to modeling the polar, azimuthal, and radial variations via traditional spherical harmonic terms, the IGRF coefficients also take into account secular field variations (recall the comments on pole reversals over time) through the inclusion of first order time derivatives of the coefficients. As new geomagnetic field measurements are made, for example by the Magsat and SAMPEX missions, updated values of the IGRF coefficients incorporating these improvements are published, usually at least as frequently as every 5 years with occasional interim updates between major releases. The range of validity of the IGRF coefficients is roughly from the Earth's surface to altitudes of about four Earth radii (i.e., below 25,000 km),[5] mainly because the strength of the geomagnetic field above that altitude becomes very weak. Furthermore, once you get above geosynchronous altitude, perturbations from the action of the solar wind and interplanetary magnetic field can overwhelm the "naturally" arising geomagnetic field, making IGRF "predictions" quite unreliable.

6.5 Star Catalogs

By contrast to the ephemeris and geomagnetic field modeling problems, the mathematics associated with specifying star-related input to fine attitude determination (or other applications utilizing star data) is pretty simple. In the previous cases, what you're trying to do is package measurements via a mathematical model in such a way that they can be accessed in a more compact form (e.g., elements, spherical harmonics coefficients, etc.) than provided by, for example, an almanac. For star data, what you're trying to do is organize measurements in a manner that makes retrieval straightforward, either when dealing with a set of data for an individual star or when trying to extract a subset of relevant data for a geometric grouping of stars. So what's needed is a catalog to hold the information, a logical grouping of stars in the catalog to facilitate extraction of current data as well as facilitating the addition or updating of new information, a set of tools that give you the power to carry out those functions, and a set of tools that allow you to manipulate the data once it's extracted. That all sounds pretty abstract and nebulous, so let's put some meat on the bones by specifying concretely what this mysterious data is.

First, each star in the catalog must have a unique identification. Next, since what you primarily want to do with star tracker measurements is compare spots on the sky seen by the star tracker with known star candidates in the catalog, the catalog needs to contain the star's right ascension and declination (i.e. the star's direction in inertial space) at the specified catalog epoch time. The origin of the inertial frame is typically our Sun (i.e., the frame and associated data are heliocentric). Another natural measurement to include for each star is its magnitude, and if we're going to

[5] J. Wertz, Spacecraft Attitude Determination and Control, p. 119, Tables 5-2 and 5-3.

relate that to star intensity measurements made by the star tracker (for example, as a secondary check in a star identification algorithm), we're also going to need to know something about the star's color. That way, given knowledge of the star tracker's spectral response, we'll know if a red star actually will look a lot brighter to our star tracker than a comparable yellow star would. There are many more data types in a star catalog than just these items (I haven't mentioned proper motion corrections, nearest neighbor distances, etc.), but let's concentrate just on star directions and magnitudes, and leave the bottom of the iceberg for inquiring minds who want to know.

For those of you who still remember reading Chap. 1 dealing with attitude conventions a long time ago in a swivel chair far, far away, specification of the direction of an object in inertial space carries with it the responsibility of defining which inertial space you're talking about. So we need to specify an epoch time and we need to say whether we're taking into account only the precession of the Earth's spin axis when defining our inertial reference frame (i.e., a mean of epoch specification), or whether we're including both precession and nutation (true of epoch). And if the star is moving significantly relative to our Sun (i.e., the star has a high proper motion), we'll need to adjust the right ascensions and declinations specified at the catalog epoch time for the amount the star has appeared to move at our current time of interest. In other words, we may need to compensate both the amount the "inertial" reference frame has moved relative to the Sun and the amount the individual star has moved relative to the Sun.

6.6 Velocity Aberration

In addition to the corrections mentioned in the previous section, you typically correct the reference data for a physical effect that actually is corrupting the measurement, but it's less computationally onerous to "corrupt" the reference data to make it consistent with all the measurements than to correct the measurements to make them all consistent with the reference. The physical effect we're talking around is called velocity aberration, which can make the observed star direction look shifted from its catalog-specified direction due to relative motion between the observer and the catalog frame. Velocity aberration is one space-related phenomenon that you don't need to be an astronaut to experience. At least you can experience an earthbound effect very much like it, as long as you've been unlucky enough to have been caught in a torrential downpour with just an umbrella for protection, and have been blessed with long enough legs that an umbrella is not a complete solution to your dilemma. Under those circumstances, the logical tendency is to get out of the rain before the lower half of your body is washed out to sea. However, if you start running for cover with the umbrella held straight above your head, you'll find yourself getting wetter and wetter the faster you run, assuming the wind doesn't blow the umbrella inside out. The problem is that, even if the rain is falling straight down (relative to local vertical), your running has created a lateral relative velocity between you and the raindrops. So when you run through the storm, the rain seems to be driving into you at

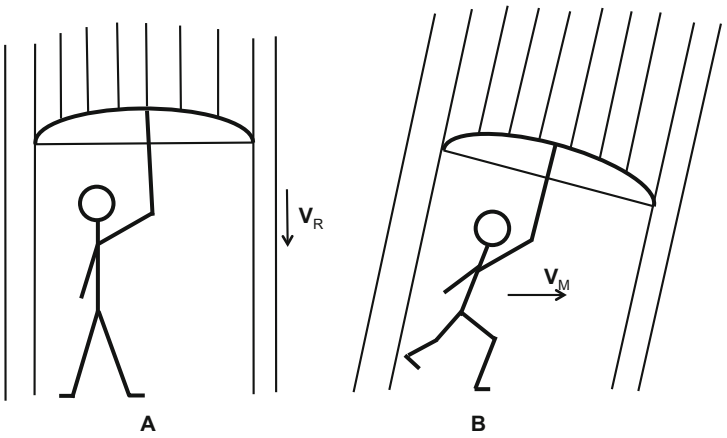


Fig. 6.3 Velocity aberration metaphor – raindrops; **A** Rain falling vertically relative to stationary man, **B** Rain apparently slanted due to forwared velocity ($\mathbf{V}_M$) of man, with angle relative to vertical given by $\tan(\Theta) = \mathbf{V}_M/\mathbf{V}_R$

a slant rather than falling straight down. To deal with the new problem caused by the implementation of your first solution, you have to tilt your umbrella into the apparently slanted falling raindrops (see Fig. 6.3), which also conveniently protects your umbrella from destruction. Of course, a better solution is just to invest in a raincoat as well as an umbrella, but that's not the way engineers typically solve problems.

In any event, out in space you can (in this context) treat light radiated from a star like raindrops falling from a cloud. So if the observer is moving laterally relative to the star, the observer will "see" the incoming light slanted relative to the catalog prediction, and the star direction will consequently also appear shifted (see Fig. 6.4).

Because the velocity of light is so high, the observed shift will always be fairly small (to first order, the ratio of the platform lateral velocity to the speed of light, with units of radians). For a star tracker (or other stellar object measuring device) mounted on a spacecraft in Earth orbit, there are two contributions to velocity aberration. First there's the motion of the spacecraft about the Earth, which for LEO spacecraft can generate a shift as large as 5 arcseconds. The 5 arcsecond shift can occur when the spacecraft's orbital velocity vector is perpendicular to the star direction vector (i.e., all spacecraft orbital motion is lateral relative to the star). By contrast, when the spacecraft's orbital velocity is carrying it directly towards or away from the star, the shift due to velocity aberration is zero. So over the course of the spacecraft's orbital period (about 97 min for LEO spacecraft), to first order the size of the correction will vary sinusoidally.

The second contribution is due to the motion of the Earth around the Sun, which can produce an effect as large as 20 arcseconds when the Earth's orbital velocity vector about the Sun is perpendicular to the star direction vector. And similarly to the spacecraft orbital motion term, when the Earth's orbital velocity vector about the Sun is parallel to the star direction vector, the velocity aberration shift from that

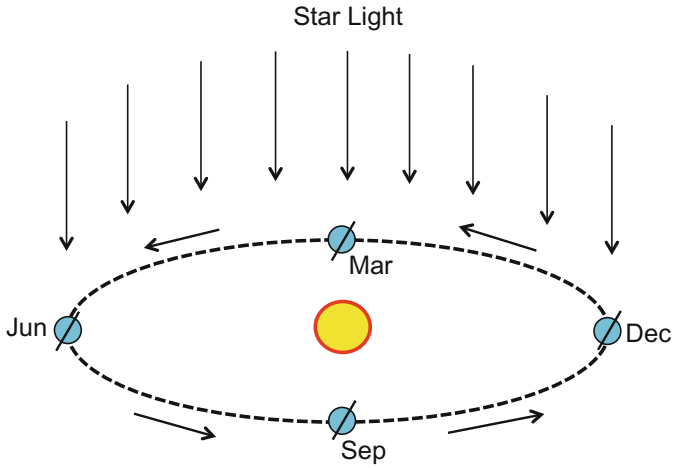


Fig. 6.4 Velocity aberration caused by lateral component of (Earth) velocity relative to incoming star light

source drops to zero. Again, we get a sinusoidal pattern over the course of an orbit, but in this case the period of the orbit is 1 year instead of 97 min. So to keep a science target "fixed" in instrument space, the platform on which the instrument is mounted (i.e., the spacecraft) often must be commanded to follow an elliptical pattern (relative to the "fixed stars" of inertial space) to compensate for the target's apparent motion arising from the platform's relative velocity.

For spacecraft with very high pointing accuracy requirements, like HST, even this level of sophistication in pointing control is not adequate. The extra fly in the ointment for HST is that the star measuring devices used for pointing control do not point exactly in the same direction as the boresight of the science instrument (SI) staring at the science target. So if you adjust the spacecraft attitude to keep guide stars fixed in HST's FGSs, a residual, uncorrected velocity aberration term influencing the apparent direction of the science target will cause the science target to appear to execute a small ellipse relative to the SI. The size of the ellipse is driven by the angular separation between the star sensor and SI boresights. For HST, the maximum size of the effect is 40 milliarcseconds, a negligible effect for nearly every other 20th century mission except HST. The way HST deals with this problem is to "tell" its attitude control function that the guide stars are "supposed" to move in tiny ellipses relative to their FGS FOVs, with the same size but opposite sense as the ellipse that the target star would have appeared to execute. Adding that feed-forward term to the attitude control will cause the target star to appear fixed in the SI FOV, which in turn prevents unnecessary smearing of the target image over long integration periods.

6.7 Parallax

There's still one more correction we may need to make to our star reference data, namely the parallax correction. Parallax is caused by the lateral offset of the observer from the line connecting the object being observed and the origin of the coordinate system in which the object's true direction is specified, typically the Heliocentric Inertial (HCI) reference frame (see Fig. 6.5).

Like velocity aberration, parallax can be experienced without leaving terra firma, and better yet without getting drenched. You can "see" the effects of parallax directly by lining up an object with your thumb using your right eye only, and then closing your right eye and sighting only with your left eye. The object will appear to shift position in your field of vision. With velocity aberration, the shift was due to the observer's relative lateral velocity. With parallax, the shift is due to the observer's lateral position shift. The larger the lateral position shift, the larger the apparent shift in target direction. And if you play the game with one target nearby and one far away, you'll find that the apparent direction shift for the faraway target is much smaller than that for the nearby one.

Over a spacecraft orbit, you usually only can observe this effect if the target is a Solar System object. For example, at LEO altitudes, the apparent direction of the Moon can shift by up to 2° (i.e., from −1° to +1° over an orbital period). The Sun's direction can appear shifted by up to 19 arcseconds (from –9.7 to +9.7 arcseconds). However, if we work with a much larger time scale than 97 min, say 1 year, we'll now be subjected to the change in position of the Earth as it orbits the Sun, trading in a baseline of, say, 7,500 km for one of nearly 300,000,000 km. Now, over the course of a year, you can measure the parallax shifts of nearby stars and, knowing the diameter of the Earth's orbit, deduce the distances to those stars. The largest

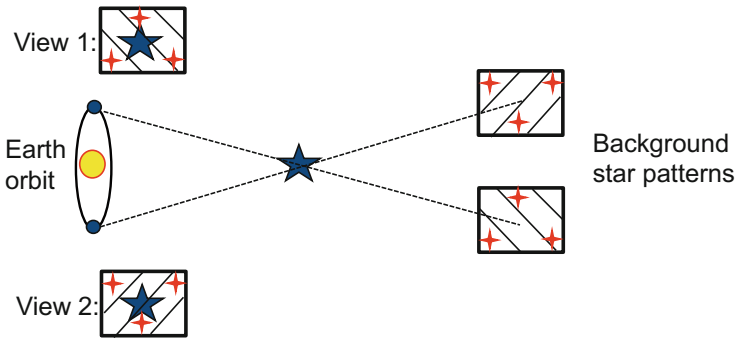


Fig. 6.5 Stellar parallax

stellar parallax observable from an Earth-based observatory is that for our nearest star, Alpha Centauri. Alpha Centauri is about 4 light years away, yielding a parallax shift of about 0.76 arcseconds, half-angle.[6]

6.8 Stellar Magnitude

At this point, we've talked so much about what you have to do to model star directions accurately that you've probably forgotten that we still need to talk about star magnitude. Most star trackers are capable of "seeing" down to magnitude 7. Probably magnitude 9 can be considered beyond the extreme limits of anything visible to a conventional star tracker. That's just as well because at magnitude 9 you're dealing with a pretty crowded sky. Half of the stars brighter than 9th magnitude have another star of equal or greater brightness within 0.2°.[7] Further, about 90% of the 9th magnitude stars have a near-neighbor within 0.35°.[8] And probably you've forgotten that we already discussed back in Chap. 4 the quantitative relationship between star magnitude and the intensity a star tracker measures, so let's go back over that ground again. Star magnitude is a logarithmic function of a star's intensity (see Eq. 6.3).

$$m = m_0 - 2.5 \log i \tag{6.3}$$

where

m = object's magnitude
m_0 = scale constant
i = intensity

Note that a difference of 5 in magnitude is equivalent to a factor of 100 in intensity. First magnitude roughly corresponds to the brightest stars (except for the Sun, which clocks in at magnitude –26.7).[9] Sixth magnitude corresponds to the dimmest stars visible to an average unaided eye. And the star trackers we've spent so much time talking about typically utilize stars between 2nd and 7th magnitude, but are capable of detecting stars as dim as 8th magnitude. So the larger the star magnitude, the dimmer is the star, while really bright objects have negative magnitude. Since magnitude was originally defined based on the response of a particular detector class (i.e., human eyes), the definition depends on the spectral responsivity of that class. Visual (V) magnitude is defined using a frequency band centered at ~550 nm. Various different magnitudes have also been defined, e.g., blue (B) magnitude with a band centered at ~420 nm.

---

[6] J. Wertz, Spacecraft Attitude Determination and Control, p. 31.
[7] Ibid, p. 146.
[8] Ibid, p. 146.
[9] Ibid, p. 79.

Magnitude actually is a lot more complicated a creature than indicated by Eq. 6.3. That's because the intensity measured by a star tracker is dependent on the color of the star being observed. Not all star trackers have the same spectral response. Typically, a star tracker will "see" a red star as brighter than a corresponding blue star, but HST's FHSTs were more neutral in that regard, having spectral characteristics more like that of the human eye. In any event, what all this regarding color means is that it's not enough for your star catalog to list a star's visual magnitude, it must also say something about the star's color. The SKYMAP star catalog, for example (see Sect. 6.9), lists visual magnitude along with the $B - V$ (blue minus visual) difference term. Since star tracker responses do not coincide with either B or V, a combination of the two usually is required to correlate an observation with its associated catalog star. An added complication is that the visual magnitude of some stars fluctuates over time. So the catalog must also specify the variability of the star's magnitude. Fortunately, only about a half a percent of stars brighter than 9th magnitude vary by more than 0.1 of a magnitude,[10] so this is not a practical concern for ground or onboard star identification and attitude determination.

6.9 Star Catalog Examples

Before we finish, let's briefly talk about some of the more popular star catalogs. Back in 1978, two standard astronomical catalogs were the Catalog of Bright Stars and the Smithsonian Astronomical Observatory (SAO) Catalog. The Bright Stars catalog contained 9,100 stars down to magnitude 7.0, with a complete listing down to magnitude 6.0. The SAO is much larger, containing (in 2001) 258,996 stars, and is complete (as of 2001) down to 9th magnitude, but with 4,503 stars fainter than magnitude 10. Back in the 1970s and 1980s, the most used catalog at GSFC for attitude determination purposes was SKYMAP, which (then) contained 255,000 stars down to 10th magnitude, but was being significantly upgraded in the 1990s using data from the Tycho catalog. The Tycho Catalog is populated with data collected by the European Space Agency's (ESA's) Hipparcos satellite, and (as of 2001) contained 2,539,913 objects. For people who can't get enough of a good thing, there's the U.S. Naval Observatory Astrometric Reference Catalog, which contained (as of 2001) 54,787,624 sources providing a grid of reference objects for astrometers over the entire sky. Not so finally (as need for better accuracy and fainter reference stars continues to evolve), ESA's GAIA mission team is scheduled to release in June-2016 its first version of a star catalog with with upwards of 100 million stars, in most sky regions complete down to magnitude 20.[11]

[10] J. Wertz, Spacecraft Attitude Determination and Control, p. 145.

[11] http://www.cosmos.esa.int/web/gaia/science-objectives.

One of the particularly nice qualities of the SKYMAP catalog was that it came equipped with a set of utility subroutines that provided a whole set of functions useful when extracting data from the catalog and manipulating the data once you've extracted what you want. For example, tools allowed you to convert the data stored as mean-of-epoch 2000 to mean-of-epoch Whatever-You-Want after first converting from (right ascension, declination) to unit vector format. It also could transform the star vectors from mean-of-date to true-of-date, and convert from unit vector format back to (right ascension, declination). In addition to getting all the information in the catalog regarding a specific star, its tools also supported extraction of selected data regarding all stars "contained" in a circular cap on the celestial sphere, or even all the stars that would be seen within a star tracker's square FOV. Other tools told you which nearby neighbors a star had within a given angular separation and magnitude range. Some of the other catalogs mentioned may also provide similar software tools, but SKYMAP was a catalog designed specifically to support attitude determination, so its utility functions were tailored to that purpose. This completes a gratis political advertisement for the old SKYMAP catalog, bringing us back full-circle to Tip O'Neill.

Chapter 7
Onboard Attitude Determination

In the movie *Magic Town*, Jimmy Stewart plays a public opinion consultant who's obsessed with finding a cheap, fast way to report how the American public feels about current events, issues, marketable products, election candidates (is that redundant?), etc. Not satisfied with traditional polling methods that canvass large numbers of people scattered across the United States, Stewart instead searches for a small town whose population is a perfect cross-section of the country. That way, all he needs to do is visit the barbershop and other natural gathering places in the town and subtly extract from the residents their views on whatever topic the pollsters are interested in that week. Hardly amazingly (this is Hollywood, after all) the scheme works pretty well until the residents find out what's going on. At that point, being aware that their views have importance, or at least that someone cares what they think, they stop reporting what they honestly believe and start telling the pollsters what they think the pollsters want to hear. As a consequence, the polling results are hopelessly skewed from the norm, the town's views become commercially useless, and everybody's mad at Stewart for making them look like fools. At that point, Stewart (of course) drops all the subterfuges and returns to being the naturally honest, all-American guy he originally was. So there's a happy ending of some sort as life in the town is restored to normal. In many ways, onboard attitude determination is a lot like *Magic Town*, without the star power provided by Jimmy Stewart. In our case, though, we need cheap, fast results less from the standpoint of reducing costs and more from the need to respond to the frequent demands of a hungry realtime control function. If we can't determine quickly and accurately what our pointing and associated error is, we can't maintain the spacecraft stability required to perform science, and may not even have the capacity to keep the spacecraft from tumbling out of control. Fortunately though, the things we poll (the star trackers, Sun sensors, Earth sensors, gyros, and magnetometers) lack self-awareness, so there's no danger of silly responses, unless an inflight anomaly occurs.

Unfortunately, the mathematics needed to make sense of the information provided by the sensors is much more complex than Stewart's data processing methods in *Magic Town*. Also unfortunately, the math in this section is somewhat more demanding than in most others. Although things should be pretty straight-forward at first,

© Springer-Verlag London Ltd. 2017

H.L. Hallock et al., *ACS Without an Attitude*, NASA Monographs
in Systems and Software Engineering, DOI 10.1007/978-1-4471-7325-0_7

the mathematics will get progressively more challenging as we approach the section on Kalman filtering, at which point some of you may find the quote from Dante's Inferno "Abandon all hope ye who enter here" prophetically fitting. For those of you that don't lose hope in the face of equations, we encourage you to view them from two perspectives - both at the detailed level within each equation, and more importantly in the repetition of patterns through the chapter. A number of patterns will recur, both in verbal explanations and associated equations. Try to pay attention to those even if the details are a bit glaze-inducing. For all of you, but particularly those that are a bit math shy, we provide a human-readable synopsis of the basic ideas of the Kalman filter in Sect. 7.13, which is a no-equations version of the Kalman filter's processing steps, along with a figure illustrating that processing. Just as the Kalman filter used for attitude determination is (in principle) an iterative algorithm, best results in understanding this material probably lie in reading through the material multiple times, possibly reading Sect. 7.13 ahead of Sect. 7.9 and then returning to Sect. 7.13 as you come staggering out of Sect. 7.12.

7.1 Attitude Propagation with Gyroscope Data

Many classic works of literature begin *in medea res* ("in the middle of things"). We first encounter Aeneas ship-wrecked on the shores of Carthage, and Victor Frankenstein nearly delirious on a merchant ship in the North Sea. A somewhat more recent example is the release order of the Stars Wars movies, beginning with Episode 4. Imitating art, we'll begin our discussion similarly, with the action already started. Imagine that somehow you've been able to establish an estimate of your spacecraft's attitude at some particular moment in time (we'll return to how that's done in later sections). In this section we discuss how you can use gyroscope data to propagate that estimate to different moments in time.

The quickest way to measure change in attitude (not absolute attitude or attitude error) is with gyroscopes. Typically, one processes the attitude change measured by a set of gyros over the sampling interval to obtain an estimate of the spacecraft body rate. A gyroscope provides a measurement of the projection of the spacecraft's rotation rate onto the axis to which the gyro is sensitive. If there are at least three gyroscopes mounted on the spacecraft, oriented so as to sample the full, three-dimensional rotation vector of the spacecraft, you'll be able to reconstruct that full rotation. Knowing the rotation, i.e., the way attitude is changing over time, you'll be able to propagate a given estimate of attitude either forward or backward in time to construct a history of the spacecraft attitude vs. time over the time span for which you have gyro data.

Conceptually, that's all there is to it. Now let's put those words into equations. Given a set of measurements from N gyroscopes ($N \geq 3$), you can solve for the spacecraft rate using the following equation:

$$\boldsymbol{\omega} = \boldsymbol{T}_{v/g}\boldsymbol{S}(\mathbf{P} - \mathbf{B}) \tag{7.1}$$

where

$\boldsymbol{\omega}$ = spacecraft body rate in the vehicle reference frame (rad/sec), dimension = 3
$\boldsymbol{T}_{v/g}$ = gyro-to-vehicle frame transformation matrix (unitless), dimension = $3 \times N$
$\boldsymbol{S}$ = an $N \times N$ diagonal matrix containing the scale factors (one for each gyro), from raw gyro counts to engineering units ((rad/sec)/count)
$\mathbf{P}$ = raw gyro measurement vector (counts), dimension = $N \times 1$
$\mathbf{B}$ = vector of gyro drift biases (counts), one for each gyroscope

The matrix $\boldsymbol{T}_{v/g}$ is a function of the alignments of the gyros relative to the body. In the simplest case (not uncommon for spacecraft design), engineers will have mounted three gyroscopes, one each along the spacecraft X, Y, and Z axes. For that case, $\boldsymbol{T}_{v/g}$ is just $\boldsymbol{I}_3$, the 3×3 identity matrix. For arbitrary orientations of 3 gyroscopes, $\boldsymbol{T}_{v/g}$ will be the inverse of the matrix constructed from the set of unit vectors that defines their alignments relative to the body (let's call that $\boldsymbol{T}_{g/v}$). If more than three gyros are being used, the math for constructing $\boldsymbol{T}_{v/g}$ from the $N \times 3$ matrix $\boldsymbol{T}_{g/v}$ becomes a bit more complicated. It can be shown that in order to treat all gyroscopes as equally sensitive, the optimal solution for constructing $\boldsymbol{T}_{v/g}$ from $\boldsymbol{T}_{g/v}$ in the general case is given by

$$\boldsymbol{T}_{v/g} = \left[\boldsymbol{T}_{g/v}^T \boldsymbol{T}_{g/v}\right]^{-1} \boldsymbol{T}_{g/v}^T \tag{7.2}$$

$\boldsymbol{T}_{v/g}$ as constructed from Eq. 7.2 is called the *left pseudo-inverse* of $\boldsymbol{T}_{g/v}$. Note that $\boldsymbol{T}_{v/g}\boldsymbol{T}_{g/v} = \boldsymbol{I}_3$, i.e., $\boldsymbol{T}_{v/g}$ acts like an inverse when multiplied on the left side of $\boldsymbol{T}_{g/v}$.

OK - now that you have an estimate for the spacecraft rate, you can determine the *change* in attitude over time. You can do this in quaternion format (i.e., the quaternion that represents the small rotation over short time period Δt) in a very straightforward fashion:

$$\delta\boldsymbol{q} = \left[e_1 \sin\frac{\omega_1 \Delta t}{2},\ e_2 \sin\frac{\omega_2 \Delta t}{2},\ e_3 \sin\frac{\omega_3 \Delta t}{2},\ \cos\frac{|\boldsymbol{\omega}|\Delta t}{2}\right] \tag{7.3}$$

This delta-quaternion can in turn be used to propagate the previous attitude to an updated attitude. If the previous attitude is specified by the quaternion $\boldsymbol{q}_n$, the updated quaternion $\boldsymbol{q}_{n+1}$ is given as:

$$\boldsymbol{q}_{n+1} = \boldsymbol{q}_n \delta\boldsymbol{q} = g(\delta\boldsymbol{q}, \boldsymbol{q}_n) \tag{7.4}$$

where the function g casts $\delta\boldsymbol{q}$ into the 4×4 quaternion matrix multiplier form used in Chap. 1, Sect. 1.6.

Of course, there's nothing magic about using the quaternion representation for attitude and attitude change. You can use the matrix representation of attitudes and rotations, with a small rotation given by

$$\delta \boldsymbol{R} \approx \begin{bmatrix} 1 & \omega_3 \Delta t & -\omega_2 \Delta t \\ -\omega_3 \delta t & 1 & \omega_1 \Delta t \\ \omega_2 \Delta t & -\omega_1 \Delta t & 1 \end{bmatrix} \tag{7.5}$$

and Eq. 7.4 recast as

$$\boldsymbol{A}_{n+1} = \delta \boldsymbol{R} \boldsymbol{A}_n \tag{7.6}$$

In Eq. 7.6, $\boldsymbol{A}_{n+1}$ and $\boldsymbol{A}_n$ are the 3×3 attitude matrices that correspond to the quaternions $\mathbf{q}_{n+1}$ and $\mathbf{q}_n$. There is an important distinction between the differential quaternion representation in Eq. 7.3 and the differential matrix representation in Eq. 7.5. For non-infinitesimal Δt, $\delta \mathbf{R}$ as expressed in Eq. 7.5 isn't normalized and therefore no longer properly represents a rotation. This is in contrast to the form of $\delta \mathbf{q}$ in Eq. 7.3, which by construction is *always* a proper quaternion and therefore always represents some rotation. This may not be magic, but it is a very convenient property, and one of the reasons that quaternions are often the preferred representation for attitude work. If $\boldsymbol{\omega}$ is changing with time, a full rotation (represented in either quaternion or matrix notation) can only properly be found by "integrating" the selected differential form over time. We use quote marks around "integrating" because in this case the integral is formed not as an infinite sum, but rather an infinite product of small rotations.

Note, by the way, that the rotation information represented as $\boldsymbol{\omega} \Delta t$ (or $\delta \boldsymbol{q}$, or $\delta \boldsymbol{R}$) allows you to propagate not just the current attitude estimate forward in time from t_n to t_{n+1} as the the spacecraft rotates, but also any inertial quantities represented in the spacecraft reference frame at time t_n. Thus, if you had made a measurement with a spacecraft-mounted sensor to characterize (in the spacecraft frame at time t_n) some vector that is intrinsically fixed in inertial space, you could use the gyroscope data accumulated between time t_n and t_{n+1} to determine what you would have measured at time t_n if the spacecraft's attitude then were what it was at t_{n+1}. As a particular example, if you make a measurement of the Sun direction using a Sun sensor at time t_n, you can use the gyro data to propagate that Sun direction estimate to the body frame that corresponds to time t_{n+1}. This propagation does *not* take into consideration effects of spacecraft motion other than rotation. Do not misunderstand us to be saying that the rotation yields what would be measured had you made the measurement at time t_{n+1}. For the Sun vector example, the rotation procedure does not include, for example, the effect of parallax due to the spacecraft's motion in space. Similarly, had we chosen to use as our example a measurement of the Earth's magnetic field with an onboard magnetometer, propagation with gyro data would not include the effects of the motion of the spacecraft to a location where the magnetic field was different.

7.2 Reference Attitude

The discussion in Sect. 7.1 explained how you can use gyro data to propagate some known attitude estimate to a future time as the spacecraft rotates. The only complication with this is the source of the previous attitude estimate. Where did that come

from? For stellar pointing missions back in the days before quaternion star trackers, the initial onboard attitude came from the ground system, which processed downlinked star tracker data to get the (then) current spacecraft attitude, which could be uplinked back to the spacecraft to provide the spacecraft with its first attitude reference. That reference quaternion then would be propagated onboard solely with gyro data until the conditions were favorable to let the spacecraft compute its own star tracker attitude updates. Prior to that event, there usually would be additional ground attitude reference updates.

These later ground updates had to deal with an extra complication, namely time synchronization with the propagated onboard quaternion. Here's the way that worked. When the ground computed, say, its second attitude reference update, it processed, for example, 15 min of star tracker and gyro data downlinked from the spacecraft between 14:00 and 14:15 UTC. So the midpoint in time of the ground's computed solution would be 14:07:30 UTC. Suppose the time required to process the data and uplink the solution took another 15 min. Then at 14:30 UTC an attitude reference update arrived onboard. But if what was uplinked was a new estimated quaternion, we'd be replacing a propagated quaternion current to 14:30 UTC with one that was valid back at 14:07:30. Twenty-two and a half minutes of gyro propagation that accurately reflected the actual motion of the vehicle between 14:07:30 and 14:30 UTC would therefore have been lost. A better approach was to uplink a delta quaternion representing the difference between the downlinked onboard estimated quaternion at 14:07:30 UTC and the ground computed attitude valid for that time. That way, the FSW was able to multiply the uplinked delta quaternion with the *current* onboard estimate (14:30 UTC), thereby correcting the onboard estimate with the ground's improved solution while still retaining the value of the additional gyro information known only to the flight system.

So that's how you used to exploit the ground system's attitude determination capabilities to obtain a starting point, or updated starting point, for onboard attitude propagation. But to function autonomously, the spacecraft really needs to be capable of performing its own attitude determination independently. Nowadays that's as easy as commanding a quaternion star tracker to "Lost-in-Space" mode and letting the star tracker compute an inertial space to star tracker frame quaternion for you. All the flight software needs to do after that is to combine the quaternion with the star tracker-to-body frame alignment matrix, and make appropriate velocity aberration corrections (if the star tracker itself doesn't do that) to get the corresponding estimated attitude quaternion (a transformation from inertial space to the body frame). But prior to the development of quaternion star trackers, the attitude control flight software was responsible for processing absolute attitude measurements from selected attitude sensors (star trackers, Earth sensors, Sun sensors, or magnetometers), mixing in the appropriate reference data, and calculating the estimated attitude or attitude error itself.

A number of approaches to performing this function have been used on GSFC missions. We will consider a few variations on the theme in the following sections, beginning with one of the simplest, known as Triad.

7.3 Minimum Data Attitude Determination

Attitude estimation is typically done by comparing measurements made using onboard sensors of a number of directions in space (e.g., towards a number of stars, the Sun, or the local geomagnetic field direction) with corresponding reference direction information for the same objects (e.g., star direction data from a catalog, Sun direction based on ephemerides of the Earth about the Sun and spacecraft about the Earth, or a geomagnetic field model using the spacecraft position as input).

The smallest number of such measurements that one would need for such an estimate is two: one to pick out a particular direction in space (say, towards the Sun), and the second to determine the orientation of the spacecraft about the selected direction. Indeed, although two measurements are needed, they in fact over-specify the problem. Given that any measurement will (almost) inevitably have errors at some level associated with it, it is equally (almost) inevitable that no attitude rotation will be able to make the two measurements exactly lie on top of the corresponding reference direction vectors. The inability to find an exact fit will be manifest as a difference between the angular separation of measured directions and the corresponding angular separation of the reference directions. This difference is actually a good thing; it provides you with an estimate of how trustworthy the solution is.

The mathematical procedure for converting the qualitative description provided in the preceding paragraph into a quantitative result is called *Triad*, or sometimes the *algebraic method*.[1] It works with pairs of measurements to compute an attitude that is valid for the time the measurements were taken. A popular example uses a measurement of the Sun direction vector and a measurement of the geomagnetic field vector (both rotated from sensor space to the body frame) along with their corresponding reference vectors (both in the GCI frame) to compute the attitude matrix.

Triad's first step is to designate one of the two measurements (say, the Sun direction) as "primary." Triad then will force its attitude solution to rotate this primary measurement vector onto the primary reference direction, thereby using up two of the three degrees of freedom of the attitude solution. The third degree of freedom is used by rotating the secondary vector (in our example, the measured magnetic field direction) about the primary vector until the secondary vector comes as close as possible to the secondary reference vector, subject to the constraint that the measured angular separation from the primary is maintained. The mathematical procedure that realizes this algorithm and provides the attitude matrix $\boldsymbol{A}$ is to use the vector cross product to construct a "triad" of orthogonal unit vectors, as shown in Eq. 7.7:

$$\boldsymbol{A} = \boldsymbol{M}_B \boldsymbol{M}_R^{-1} \tag{7.7}$$

[1] H. Black, A Passive System for Determining the Attitude of a Satellite, AIAA Journal 2(7), pp. 1350–1351, 1964.

where

$\boldsymbol{A}$ = attitude matrix
$\boldsymbol{M}_B = [\mathbf{u}_B\ \mathbf{v}_B\ \mathbf{w}_B]$
$\mathbf{u}_B = \mathbf{S}_B$, where $\mathbf{S}_B$ is the sun unit vector in the body frame
$\mathbf{v}_B = \mathbf{S}_B \times \mathbf{B}_B / |\mathbf{S}_B \times \mathbf{B}_B|$, where $\mathbf{B}_B$ is the geomagnetic field unit vector in the body frame
$\mathbf{w}_B = \mathbf{u}_B \times \mathbf{v}_B$

and where the reference matrix, $\boldsymbol{M}_R$, and its associated vectors are defined similarly.

Note that the vectors $\mathbf{u}_B$, $\mathbf{v}_B$, $\mathbf{w}_B$ are "column" vectors, and $\boldsymbol{M}_B$ is a 3×3 matrix whose columns are these three vectors. Notice also the asymmetry in the procedure between the use of $\mathbf{S}_B$ and $\mathbf{B}_B$. The vector $\mathbf{u}_B$ (and therefore $\boldsymbol{M}_B$) contains the full information content of $\mathbf{S}_B$, whereas $\mathbf{v}_B$ is only sensitive to the component of $\mathbf{B}_B$ perpendicular to $\mathbf{S}_B$. Computationally, it's best to select as the "primary" vector ($\mathbf{u}_B$ in Eq. 7.7) the more accurate measurement, which for this example would be the Sun vector. Examination of Eq. 7.7 shows that the attitude matrix constructed in this fashion will transform the reference vector $\mathbf{u}_R$ into exactly the measured vector $\mathbf{u}_B$. The information from the secondary measurement ($\mathbf{B}_B$ in our example) is therefore only used to supply the phase angle about $\mathbf{u}_B$. Since $\mathbf{S}_B$ and $\mathbf{B}_B$ are each defined by two angles (i.e., they each contain two independent pieces of information), and since the attitude matrix is parameterized by just three independent pieces of information, one piece of measurement information is simply discarded by Triad. The discarded information is the measured component of $\mathbf{B}_B$ parallel to $\mathbf{S}_B$. Of course, if $\mathbf{B}_B$ happens to be parallel, or anti-parallel, to $\mathbf{S}_B$, there is no phase angle information and $\boldsymbol{A}$ cannot be fully determined; the Triad algorithm will not work in this case.

For those of you unhappy with the notion of being forced to give preference to one vector over the other, there is a solution that is almost as simple as plain-vanilla Triad, with just a little preliminary work. Select as the primary vector a normalized vector that points midway between the two, i.e., select as primary a vector $\mathbf{P}_B = (\mathbf{S}_B + \mathbf{B}_B)/|\mathbf{S}_B + \mathbf{B}_B|$, and similarly for its associated reference $\mathbf{P}_R$. For the secondary, pick either $\mathbf{S}_B$ or $\mathbf{B}_B$ (it doesn't matter which), and then proceed with Triad as before. The result will be an attitude matrix that treats the two measurements equally, with each rotated reference vector equally far from its measurement. The equation for $\mathbf{P}$ can also be generalized using weighting factors w_S and w_B for $\mathbf{S}$ and $\mathbf{B}$ that depend on the uncertainty of their respective measurements: $\mathbf{P} = (w_S\mathbf{S} + w_B\mathbf{B})/|w_S\mathbf{S} + w_B\mathbf{B}|$. Appropriate values for w_S and w_B can be determined using a standard assumption of Gaussian distributed noise for the measurements. This leads to $w_S = \sigma_S^{-2}/(\sigma_S^{-2} + \sigma_B^{-2})$ and $w_B = \sigma_B^{-2}/(\sigma_S^{-2} + \sigma_B^{-2})$, where σ_S and σ_B are the uncertainties in the Sun and magnetic field direction measurements.

Because of its simplicity (particularly if one ignores the complication of constructing a weighted primary vector), Triad has been a popular onboard algorithm for coarse attitude determination. Applications for Triad have included attitude initialization following separation from the launch vehicle, attitude re-initialization following safemode entry, antenna steering, bright object checking, and fine attitude sanity checking. In general, Triad was a great tool for getting an attitude quickly and

robustly. Not a whole lot can go wrong in the computations, as long as you don't use vectors that are nearly parallel to each other, and as long as you actually have two measurements. (When a spacecraft is in shadow, there is no Sun vector to use). The only downside is its limited accuracy due to the minimal amount of data used. Any noise gets carried over directly as attitude error since no averaging is performed. Or if there is a "glitch" of some kind in a measurement, there is no way to identify it and edit it out, other than by comparing Triad solutions to each other. For most missions, Triad simply isn't accurate enough to use as the prime attitude determination method, but it can be extremely useful in a subordinate capacity where high accuracy is not the prime concern. And nowadays the presence of quaternion star trackers on most missions means that a high accuracy star tracker attitude is readily available for many of the applications (for example, antenna steering) that might have gotten by with the lower accuracy Triad solution.

7.4 Batch Attitude Determination with Vector Observations

In Sect. 7.3, we discussed attitude estimation using only two measurements. Although both simple and often useful, the Triad algorithm is statistically non-optimal in that (a) it doesn't permit the use of more measurements even if they are available, and (b) it doesn't support the best weighting between the two measurements, at least in its plain vanilla form. In this section we consider a treatment that results in an optimal solution, i.e., one that gives proper weighting to as many vector measurements as you may have available.

No doubt some of you have already jumped from the words "statistical" and "optimal solution" to the expectation that we'll likely be discussing a least-squares method here. Derivations for such methods often begin with the statement of a *loss function* that must be minimized as the formal goal for finding the desired solution. For our attitude problem, i.e., the determination of spacecraft attitude based on the measured direction of a set of N known objects (Sun, stars, local geomagnetic field), an appropriate loss function is given in Eq. 7.8 (first proposed by Grace Wahba[2]):

$$L(\boldsymbol{A}) = \sum_{i=1}^{N} w_i |\mathbf{v}_{B_i} - \boldsymbol{A}\mathbf{u}_{Ei}|^2 \tag{7.8}$$

where

w_i = the ith weight, $i = 1, 2, \ldots N$, derived from estimated measurement and reference uncertainties
$\boldsymbol{A}$ = the attitude matrix
$\mathbf{v}_{B_i}$ = the ith measurement unit vector, expressed in the body ('B') frame
$\mathbf{u}_{E_i}$ = the ith reference unit vector, expressed in the ECI ('E') frame

[2]Wahba, G., A least-squares estimate of satellite attitude, SIAM Rev. 7(3), 409 (1965).

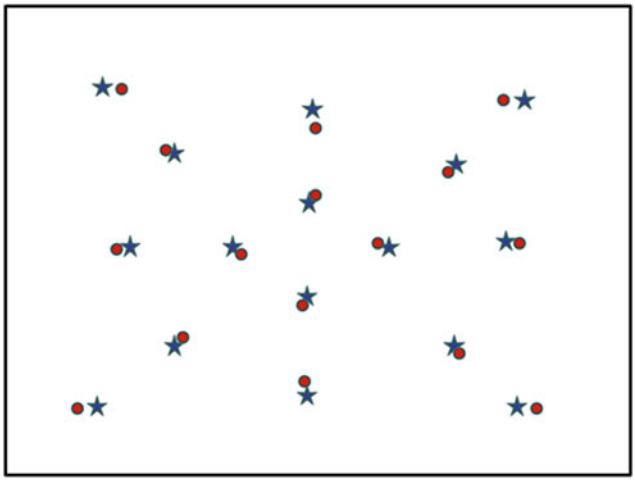


Fig. 7.1 An imagined overlay of a set of observations (*blue stars*) and associated reference vectors from a catalog (*red circles*) after correction for an optimal attitude solution (color figure online)

The idea with Eq. 7.8 is that the attitude matrix that minimizes the loss function $L(\boldsymbol{A})$ will be, in a statistical sense, the optimal solution for the spacecraft attitude.

If all of the measurements are taken with sensors of equal accuracy, the weight factors will be the same for all of the measurements. In that case, you can simply ignore the constant weight term as being just a renormalization of $L(\boldsymbol{A})$, one that won't change the solution. For the moment, let's consider that simplified problem. For ease in visualization, let's further imagine that all of the measurements are clustered together in a limited region. The optimal solution, $\boldsymbol{A}_{opt}$ (the one that results in the smallest value for $L(\boldsymbol{A})$) will rotate the cluster of reference vectors onto the cluster of observations such that (a) the mean position of the cluster of reference vectors will lie on top of the mean of the cluster of measurements, (b) the scatter of the residuals (i.e., the differences between measurements and associated rotated references) will appear to be random, and (c) the size of the residuals will be characteristic of the accuracy of the sensor being used and the reference data against which the measurements are being compared. (Sometimes the measurements can be more accurate than the reference data against which they are being compared, an example being measurements made with HST's fine guidance sensors.) Fig. 7.1 shows an imagined overlay of a set of observations and associated reference vectors from a catalog after correction for an optimal attitude solution - say, in the field of view of a star tracker. The separation between each measurement and reference is indicative of the combined error of star tracker measurements and whatever measuring device was used to generate the catalog.

If $\boldsymbol{A}$ is chosen in Eq. 7.8 to be the true attitude rotation, the quantity $\mathbf{r}_{B_i} \equiv \mathbf{v}_{B_i} - \boldsymbol{A}\mathbf{u}_{E_i}$ (the ith "residual") is a "small" vector representing the difference between the ith measurement and the associated ith reference. (In Fig. 7.1, each residual would

correspond to a vector from a red circle to its associated blue star.) Although still formally a three-vector, $\mathbf{r}_{B_i}$ is essentially perpendicular to $\mathbf{v}_{B_i}$ and as such really only has two degrees of freedom. The typical size of $\mathbf{r}_{B_i}$ is governed by the measuring uncertainty for the sensor (e.g., for the star tracker being used) and the measuring and modeling uncertainty used in the construction of the reference vector. If $\boldsymbol{\sigma}_i$ represents the typical scatter (the *standard deviation*) of $|\mathbf{r}_{B_i}|$, $\boldsymbol{\sigma}_i^2$ (the *variance* of $|\mathbf{r}_{B_i}|$) represents the typical scatter of $|\mathbf{r}_{B_i}|^2$) for an imagined random sampling of measurements from the same sensor. The implication is that the squared error for a relatively poor sensor should be correspondingly de-weighted by that factor relative to a good sensor, i.e., $w_i \propto \boldsymbol{\sigma}_i^{-2}$.

What kind of accuracy should be expected for any algorithm that minimizes $L(\boldsymbol{A})$? If the measurements are fairly tightly clustered, and if all of the measurements have the same uncertainty ($\boldsymbol{\sigma}_i = \boldsymbol{\sigma}$, the same for all measurements), the knowledge captured in $\boldsymbol{A}_{opt}$ representing the direction of the center of the cluster will be good to $\sim \boldsymbol{\sigma}/\sqrt{N}$, where N is the number of measurements. If α is the typical angular distance of a measurement from the center of the cluster, the uncertainty in the knowledge of the rotation about the cluster center will be $\sim (\boldsymbol{\sigma}/\alpha)/\sqrt{N}$, i.e., the tighter the clustering of measurements, the less well known will be the roll angle about the cluster center. In contrast, if you have a great many measurements distributed rather uniformly about the sky, the attitude knowledge will be more or less uniform in all directions, good to $\sim \boldsymbol{\sigma}/\sqrt{N}$. Note that these comments assume that systematic errors have been removed by calibration, so that the remaining errors are truly random. A constant bias present in all sample points cannot be removed by averaging.

OK - so now you know what to expect once you've found the optimal solution for the matrix $\boldsymbol{A}$, but how do you go about finding it? Well, you can always (at least in principle) take a brute force approach of constructing a sufficiently dense grid of values for $L(\boldsymbol{A})$ as a function of any convenient three-variable set that defines $\boldsymbol{A}$, and then pick the location in the grid where $L(\boldsymbol{A})$ is a minimum. Admittedly, this would be a rather laborious task, particularly if you need $\boldsymbol{A}$ to be known quite precisely - but conceptually the approach is straightforward. Knowing that such an approach is in principle possible, you may also take it as adequate for your understanding of the concepts of this section. You can skip ahead to the next section without losing the basic thread of our discussion. For those of you inclined to stick with us a bit, we'll take a brief detour to explore some of the mathematical methods typically used for solving this and similar problems.

Finding the Minima of Loss Functions by Differential Calculus

A minimum of a function of one or more variables corresponds to a location where the function is flat in all directions, or equivalently where all of its partial derivatives are zero. Calculating the derivatives of a loss function like that in Eq. 7.8, and using them to find a solution that minimizes said loss function, turns out to be fairly straightforward if the residuals can be expressed as a linear function of the independent variables for which you are solving (e.g., a set of three Euler angles $\theta_1, \theta_2, \theta_3$ that define $\boldsymbol{A}$). If each $\mathbf{r}_{B_i}$ is a linear function of the variables θ_1, θ_2, and θ_3, L will be a quadratic function of those variables. If L is quadratic, its partial

derivatives $\partial L/\partial\theta_1$, $\partial L/\partial\theta_2$, $\partial L/\partial\theta_3$ are all linear functions of the variables $\theta_1, \theta_2, \theta_3$. (If $F = a_0 + a_1 x + a_2 x^2$, then $dF/dx = a_1 + 2a_2 x$. The basic pattern is similar for functions of multiple variables and their partial derivatives.) Setting the partial derivatives $\partial L/\partial\theta_1$, $\partial L/\partial\theta_2$, $\partial L/\partial\theta_3$ all to 0 produces three linear equations in the unknowns $\theta_1, \theta_2, \theta_3$, which can be solved using standard algebraic equations.

The approach outlined in the preceding paragraph is all well and good, except for the fact that $\boldsymbol{A}$ is not in general a linear function of its independent variables, but rather a fairly complicated trigonometric function. However, if you have a fairly good initial estimate for $\boldsymbol{A}$, let's call that $\boldsymbol{A}_0$, you can recast the problem as solving for a "small" attitude correction matrix, $\boldsymbol{R}$, that minimizes the following modified loss function:

$$L(\boldsymbol{R}) = \sum_{i=1}^{N} w_i |\mathbf{v}_{B_i} - \boldsymbol{R}\boldsymbol{A}_0 \mathbf{u}_{E_i}|^2 \tag{7.9}$$

where

$$\boldsymbol{R} \approx \begin{bmatrix} \mathbf{1} & \theta_3 & -\theta_2 \\ -\theta_3 & \mathbf{1} & \theta_1 \\ \theta_2 & -\theta_1 & \mathbf{1} \end{bmatrix}$$

and $\theta_1, \theta_2, \theta_3$ are small correction Euler angles relative to $\boldsymbol{A}_0$.

Using $\mathbf{u}_{B_{oi}} \equiv \boldsymbol{A}_0 \mathbf{u}_{E_i}$, Eq. 7.9 can be recast as

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{N} w_i |(\mathbf{v}_{Bi} - \mathbf{u}_{Boi}) + \boldsymbol{U}_{Boi}\boldsymbol{\theta}|^2 \tag{7.10}$$

where

$$\boldsymbol{U}_{Boi} = \begin{bmatrix} 0 & u_3 & -u_2 \\ -u_3 & 0 & u_1 \\ u_2 & -u_1 & 0 \end{bmatrix}_{Boi}$$

and

$$\boldsymbol{\theta} \equiv \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}$$

which makes explicit the fact that our approximate loss function L is a quadratic function of $\boldsymbol{\theta}$.

The approximation for the small rotation matrix, $\boldsymbol{R}$, uses the fact that rotations commute for small angles, at least to first order in the angles. $\boldsymbol{R}$ is an approximate solution in two senses: (a) the various elements of $\boldsymbol{R}$ will have errors of order θ^2, and (b) $\boldsymbol{R}$ as a whole will typically be shifted slightly outside of the three dimensional space of rotations and as such is not strictly a well-defined rotation. You can take care of the latter problem by reconstructing $\boldsymbol{R}$ as a set of Euler rotations in any arbitrary order, say $\boldsymbol{R} \approx \boldsymbol{R}_1(\theta_1)\boldsymbol{R}_2(\theta_2)\boldsymbol{R}_3(\theta_3)$. This will guarantee that $\boldsymbol{R}$ is a well

defined rotation, albeit still having errors of $\sim \theta^2$. Thus rather than a final solution, you "merely" have a revised estimate for the attitude: $\boldsymbol{A} \approx \boldsymbol{R}\boldsymbol{A}_0$. With that improved estimate, though, you can repeat the algorithm iteratively until the correction angles $(\theta_1, \theta_2, \theta_3)$ are reduced to sufficiently small values. Thus we see that if you have a way to get a reasonably good initial estimate of the attitude (good to a few degrees), the algorithm outlined above will allow you to refine it into an optimal estimate. Getting that initial estimate can itself be a tricky problem. One approach would be to use coarse sensors (e.g., a sun sensor and magnetometer) and apply the Triad algorithm discussed in Sect. 7.3.

The algorithm just discussed is fairly straightforward, but can be a bit lengthy in the number of computer computations required to converge. This is really not much of an issue these days, as both ground and flight computers have grown considerably in power. However, in the 1970s, when computers were considerably slower, having highly efficient algorithms was of much greater significance. An alternative approach was developed with this issue of computer efficiency in mind. The derivation is considerably more sophisticated and subtle than the approach outlined above. The approach began as "the Q-Method," developed by Paul Davenport[3] of GSFC, and was subsequently refined into Quest (Quaternion Estimator) by Malcolm Shuster,[4] then of Computer Sciences Corporation (CSC). As the name implies, the approach uses the quaternion format for attitude. Along the way, it defines a loss function $L(\boldsymbol{q})$, with the four components of $\boldsymbol{q}$ as the independent variables of the problem. There is a clever trick applied along the way; $L(\boldsymbol{q})$ is minimized relative to changes in the components of $\boldsymbol{q}$ subject to the constraint that $\boldsymbol{q}$ be normalized, i.e., that $\sum_i q_i^2 = 1$.[5] After quite a few twists and turns, the net result is a very fast and robust algorithm - one that any good attitude analyst should be aware of, and perhaps even have in his or her bag of tricks. The Q-Method and Quest are not the only variations for a fast solution to Wahba's equation. See Markley and Crassidis, Chap. 5, for further variations on the theme.

For all of a multi-star algorithm's (e.g., Quest's) advantages over Triad, you can't get more out of any algorithm than the information content of the input data with which you begin. If Quest is applied with input of just two measurement vectors and their associated reference vectors, and if one of the measurements is much more accurate that the other (and used as the primary measurement), Quest effectively reduces

[3]Documented by J. Keat, Analysis of Least-Squares Attitude Determination Routine, DOAOP, Computer Sciences Corporation Technical Memorandum CSC/TM-77/6034, Feb. 1977; see also F. Markley and J. Crassidis, Fundamentals of Spacecraft Attitude Determination and Control, Sect. 5.3.1, 2014.

[4]M. Shuster and S. Oh, Three-Axis Attitude Determination from Vector Observations, Journal of Guidance and Control (Volume 4, Number 1), January/February 1981, see also F. Markley and J. Crassidis, Sects. 5.3.2–5.3.4.

[5]The constraint is applied using "the method of Lagrange multipliers," developed by the French mathematician Joseph Louis Lagrange (1736–1813). Good descriptions of the method are available in many text books. You will likely encounter reference to the method again during the course of your professional career. Tie it in your mind to the notion of minimizing or maximizing a function subject to some constraint. As a simpler example, consider finding the highest point in the Rocky Mountains subject to the constraint that you stay on a particular cross country highway.

to plain-vanilla Triad. The only advantage of Quest over plain-vanilla Triad for a situation with only two measurements is that Quest provides a statistical weighting of the two measurements relative to each other. Even this advantage is eliminated if you use Triad with a properly weighted intermediate primary vector **P**, as discussed in the penultimate paragraph of Sect. 7.3.

As a final note, there is an assumption built into Eq. 7.8 that we kind of glossed over. Specifically, the form of $L(\boldsymbol{A})$ assumes that each device that provides a measurement of the direction to some object does so in a way that is equally accurate in both independent components of the measurement. As an example, suppose you're dealing with a star tracker that provides coordinates (x, y) for a star in its field of view (imagining the z-axis to be the boresight direction of the camera). Equation 7.8 assumes that the uncertainty in x and y are the same ($\sigma_x = \sigma_y$). This is not necessarily the case. Consider a cat's eye during the day, when the pupil contracts to a long, thin, vertical slit (in contrast to a human pupil, which remains circular). In principle, the cat's eye gives better resolution in the vertical direction than in the horizontal direction, angular resolution along any given axis of an optical device being inversely proportional to the diameter of the aperture along that direction. A star camera built with a pupil like a cat's eye would produce star images that are smeared out in the y-direction (vertical) to a lesser extent than they are in the x-direction. That lesser smearing allows a more accurate determination of the center of the image in the y-direction, which implies that $\sigma_y < \sigma_x$. To take this effect into account, you can rewrite Eq. 7.8 as

$$L(\boldsymbol{A}) = \sum_{i=1}^{N} \sum_{k=1}^{2} \sigma_{ik}^{-2} [\boldsymbol{A}_{IB_i} (\mathbf{v}_{B_i} - \boldsymbol{A}\mathbf{u}_{E_i})]_k^2 \tag{7.11}$$

where $\boldsymbol{A}_{IB_i}$ is an alignment rotation matrix that transforms vectors from the body frame to a pseudo-instrument reference frame defined such that the pseudo-boresight is exactly aligned with $\mathbf{v}_{B_i}$ (assumed hereafter to be the z-axis of this pseudo-instrument), and the x- and y-axes are defined based on the instrument accuracy asymmetry. The summation over k is for the two components (x and y) of the vector as measured in the pseudo-instrument frame, the notation $[]_k$ means to extract the kth component for the enclosed vector, and the weighting factors have been replaced on a component-by-component basis with σ_{ik}^{-2}, where $(\sigma_{i1}, \sigma_{i2}) = (\sigma_{ix}, \sigma_{iy})$. We choose to suppress the third component here because direction vectors are 2-D objects, not actually defined beyond the unit sphere. (Suppressing the out-of-sphere component isn't really necessary for the algorithm to work; the fact that both the observation and reference vector are normalized would suffice to render their out-of-sphere difference negligibly small and therefore without effect. However, by explicitly suppressing that component, we formally acknowledge that there is no out-of-sphere measurement and no associated out-of-sphere uncertainty σ_{i3}.) Algorithm adjustments implied by the use of this generalized loss function are straightforward for the method outlined following Eq. 7.10. Equation 7.10 would become

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{N} \sum_{k=1}^{2} \sigma_{ik}^{-2} [-\mathbf{u}_{Ioi} + \boldsymbol{U}_{Ioi}\boldsymbol{\theta}]_k^2 \tag{7.12}$$

where

$$\boldsymbol{U}_{I_{oi}} = \begin{bmatrix} 0 & u_3 & -u_2 \\ -u_3 & 0 & u_1 \\ u_2 & -u_1 & 0 \end{bmatrix}_{Ioi}$$

and

$$\mathbf{u}_{I_{oi}} \equiv \boldsymbol{A}_{IB_i}\boldsymbol{A}_0\mathbf{u}_{E_i}$$

We have dropped any contribution from $(\boldsymbol{A}_{IB_i}\mathbf{v}_{B_i})$ because its x and y components are 0 by definition (the measurement information now having been absorbed into $\mathbf{u}_{I_{oi}}$ and $\boldsymbol{U}_{I_{oi}}$). We can recast the notation of Eq. 7.12 (which will prove useful in subsequent sections) by defining $\boldsymbol{H}_{I_{oi}}$ as the upper 2×3 part of $\boldsymbol{U}_{I_{oi}}$, $\boldsymbol{\varphi}_{I_{oi}}$ as the upper two components of $\mathbf{u}_{I_{oi}}$, and $\boldsymbol{P}_{I_i}$ as the diagonal 2×2 matrix formed from $(\sigma_{ix}^2, \sigma_{iy}^2)$. Doing so allows us to rewrite Eq. 7.12 as

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{N} \left[-\boldsymbol{\varphi}_{I_{oi}} + \boldsymbol{H}_{I_{oi}}\boldsymbol{\theta}\right]^T \boldsymbol{P}_{I_i}^{-1} \left[-\boldsymbol{\varphi}_{I_{oi}} + \boldsymbol{H}_{I_{oi}}\boldsymbol{\theta}\right] \tag{7.13}$$

For those of you wondering why we bothered to include this "final note" - well, that will become clearer when we use Eq. 7.13 as a starting point in some later discussions.

7.5 Attitude Uncertainty: The Covariance Matrix

At the beginning of the previous section, we provided a qualitative, heuristic description of the uncertainty that would result for the three components of an attitude estimate based on minimization of the loss function $L(\boldsymbol{A})$ in Eq. 7.8. The uncertainty can be expressed more formally via the so-called *covariance matrix*, $\boldsymbol{P}$, a 3×3 matrix in the case of our attitude problem. (The letter $\boldsymbol{P}$ is the usual symbol for a covariance matrix; don't confuse it with the "primary vector" used for the modified Triad algorithm in Sect. 7.3) There are two reasons why you'll care about this expression of uncertainty. The first is simply that it is always appropriate when claiming that you know the value for some quantity to also specify how well you know it. The second reason, as will be demonstrated in subsequent sections, is that when combining a number of estimates together to get a more refined estimate, you'll need the uncertainty information to develop appropriate weighting factors - such as was done with the weighting factors for the extended Triad algorithm in Sect. 7.3 and the batch attitude loss function of Eq. 7.8 in Sect. 7.4.

The functional form for a covariance matrix for an attitude estimate, or for any kind of measurement-based estimate for that matter, can be derived given statistical information regarding the precision of the measurements used for the estimate. The procedure for doing so is typically based on some formal assumption about the nature (statistical distribution) of measurement errors, usually that such errors are "normally distributed." Recall the heuristic discussion about attitude accuracy from Sect. 7.4 regarding any algorithm that minimizes the loss function of Eq. 7.8. It bears repeating here: if the measurements are fairly tightly clustered, and if all of the measurements have the same uncertainty ($\sigma_i = \sigma$, the same for all measurements), the knowledge captured in $\boldsymbol{A}_{opt}$ representing the direction of the center of the cluster will be good to $\sim \boldsymbol{\sigma}/\sqrt{N}$, where N is the number of measurements. If α is the typical angular distance of a measurement from the center of the cluster, the uncertainty in the knowledge of the rotation about the cluster center will be $\sim (\boldsymbol{\sigma}/\alpha)/\sqrt{N}$, i.e., the tighter the clustering of measurements, the less well known will be the roll angle about the cluster center. In contrast, if you have a great many measurements distributed rather uniformly about the sky, the attitude knowledge will be more or less uniform in all directions, good to $\sim\boldsymbol{\sigma}/\sqrt{N}$.

The heuristic description given above for the tight clustering case would be expressed mathematically as a covariance matrix for uncertainty relative to a frame defined by the distribution of observations (essentially the instrument reference frame, if you're dealing with a star tracker with a tight field of view, assuming a fairly uniform distribution of observations in that field of view) as

$$\boldsymbol{P}_O \approx \frac{\boldsymbol{\sigma}^2}{N} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \alpha^{-2} \end{pmatrix} \tag{7.14}$$

The off-diagonal elements being zero has specific meaning - namely that there is no correlation between the errors for the various components of the attitude estimate; the uncertainty in roll about the observation cluster center (assumed to be the z-axis in Eq. 7.14) is uncorrelated with uncertainty in "pitch" or "yaw" pointing errors relative to the cluster center, and "pitch" and "yaw" errors are uncorrelated with each other (with "pitch" and "yaw" here being motion relative to the camera reference frame). A rigorous derivation for $\boldsymbol{P}_O$ would provide a formula for the angle α, as opposed to our fuzzy definition of it as "the typical angular distance of a measurement from the center of the cluster." In fact, since α is related to roll uncertainty, its value will be weighted towards observations at greater distance from the center of the cluster of observations because a greater lever arm provides greater roll knowledge. Of course, $\boldsymbol{P}_O$ as expressed in Eq. 7.14 is appropriate in the case of the loss function of Eq. 7.8 with uniform errors. A more complicated (albeit still diagonal) covariance matrix would correspond to the generalized loss function of Eq. 7.11, e.g., for observations made with a "cat's eye" camera.

An attitude covariance matrix, though expressed in numbers relative to a specific reference frame, is a unified geometric object that has meaning even without reference frames being defined. In this way, it is like a vector, which has intrinsic meaning even

without any particular reference frame being defined; i.e., it is a tensor analogous to the moment of inertia tensor discussed in Chap. 3. Like any tensor (including vectors), after you have expressed an attitude covariance matrix relative to one reference frame (say, the 'O' frame defined by the set of observations), you can transform it to another frame (generically the 'F' frame, perhaps in particular the body frame designated by 'B'). The form of the transformation is

$$\boldsymbol{P}_F = \boldsymbol{A}_{F<O}\boldsymbol{P}_O\boldsymbol{A}_{F<O}^T \tag{7.15}$$

where $\boldsymbol{A}_{F<O}$ is the "O-to-F" transformation matrix (i.e., vectors transform between frames as $\mathbf{V}_F = \boldsymbol{A}_{F<O}\mathbf{V}_O$).

After transformation, $\boldsymbol{P}_F$ will not in general be a diagonal matrix. While the diagonal elements of $\boldsymbol{P}_F$ represent the variance (the square of the uncertainty) for each of three small Euler error angles $(\theta_1, \theta_2, \theta_3)$ quantifying the difference between the true attitude and your estimate thereof relative to the 'F' frame, the off-diagonal elements of $\boldsymbol{P}_F$ quantify the expected size of the products $\theta_1\theta_2$, $\theta_2\theta_3$, and $\theta_1\theta_3$. Each off-diagonal element can be positive, negative, or zero, depending on whether the error angles are correlated, anti-correlated, or uncorrelated with each other. For the example of $\boldsymbol{P}_O$ in Eq. 7.14, the transformation of the attitude solution to another reference frame will cause the relatively large uncertainty in roll as perceived in the 'O' frame (essentially the sensor frame) to be distributed across all axes in the new frame. Because the various products $\theta_j\theta_k$ commute, the matrix $\boldsymbol{P}_F$ is necessarily symmetric (i.e., $P_{Fjk} = P_{Fkj}$).

As a final note in this section (actually, an augmentation of the final note of the previous section), the 2×2 matrix $\boldsymbol{P}_{I_i}$ defined for and used in Eq. 7.13 is a covariance matrix defined in (and only meaningful in) the 2-D plane tangent to the measured vector $\mathbf{v}_i$. It characterizes the uncertainty in that measurement, just as the 3×3 attitude covariance matrix characterizes the uncertainty of an attitude estimate. Just as the attitude covariance matrix can be rotated to other 3-D reference frames, $\boldsymbol{P}_{Ii}$ can also be rotated, but only within the 2-D space within which it is defined.

For more information on covariance matrices, see e.g. Sect. 3.6 of Applied Optimal Estimation by Gelb et al.

7.6 Combining Multiple Attitude Solutions

Section 7.4 discussed procedures for estimating spacecraft attitude given a set of direction vector measurements (e.g., towards stars, the Sun, the local geomagnetic field direction, whatever). Suppose we have two or more such attitude solutions, each produced with a distinct set of measurements - say, as could be the case with star trackers pointing in substantially different directions. Can you combine the results to produce a single, unified attitude estimate without beginning from the underlying set of individual direction measurements? The answer is yes and we will discuss the approach in this section. Be warned - most of this section (everything but this

introductory paragraph) will require dealing with some mathematical detail. For those of you who are content with the answer that it is possible to combine multiple attitudes given appropriate information about their relative accuracies, feel free to skip ahead to Sect. 7.7; for the rest . . .

The Mathematics of Combining Attitudes from Different Sensors

Let's suppose we have a set of N star trackers, each of which has provided an attitude estimate based on its internal processing. Actually, it will have produced an estimate for the attitude of the star tracker, which is then transformed into an estimate of the spacecraft body attitude using the instrument-to-body alignment matrix $\boldsymbol{A}_{BI}$. Pick any convenient attitude (say, the one based on the first tracker) as a reference attitude. Relative to that reference attitude, let $(\theta_1, \theta_2, \theta_3)_i$ represent the Euler angles defining an adjustment to get from the reference solution to the solution based on data from the i-th tracker, and let $\boldsymbol{P}_i$ represent the covariance for the set $(\theta_1, \theta_2, \theta_3)_i$. For compactness, we can express $(\theta_1, \theta_2, \theta_3)_i$ in vector notation as $\boldsymbol{\theta}_i$. Indeed, $\boldsymbol{\theta}_i$ is a vector in 3-space. It is in the direction of a rotation eigenvector that could be used for the construction of a quaternion representing the correction quaternion from a convenient initial estimate to the estimated true attitude.

As a vector, $\boldsymbol{\theta}_i$ can be transformed between frames. To keep that fact clear in our minds, let's write $\boldsymbol{\theta}_i$ as $\boldsymbol{\theta}_{Bi}$ when its components are expressed in the body frame, and as $\boldsymbol{\theta}_{Ii}$ when its components are expressed relative to the instrument (tracker) frame. As discussed in the preceding section, $\boldsymbol{\theta}_{Ii}$ is the attitude of the instrument. Its covariance matrix, $\boldsymbol{P}_{Ii}$, is diagonal, or at least very nearly so.

A loss function for estimating the combined attitude solution is

$$L(\boldsymbol{\theta}_B) = \sum_{i=1}^{N} \sum_{k} \sigma_{Iik}^{-2} \left[\theta_{Iik} - (\boldsymbol{A}_{IiB}\boldsymbol{\theta}_B)_k\right]^2 \tag{7.16}$$

where $\boldsymbol{A}_{IiB}$ is the body-to-instrument alignment matrix for the i-th instrument, and $\boldsymbol{\theta}_B$ is the attitude correction for which we are solving. This loss function is specifically cast to look formally similar to the loss function defined in Eq. 7.11. The term $(\theta_{Iik} - (\boldsymbol{A}_{IiB}\boldsymbol{\theta}_B)_k)$ is the residual for the k-th component of the attitude solution vector for the i-th instrument, expressed in the instrument frame (where there is (very nearly) no correlation between the components of the attitude estimate). By multiplying the desired solution vector $\boldsymbol{\theta}_B$ by $\boldsymbol{A}_{AiB}$, we are transforming $\boldsymbol{\theta}_B$ into a frame where $\boldsymbol{\theta}_{Ii}$ has (very nearly) uncorrelated components. The subscript k on $(\boldsymbol{A}_{IiB}\boldsymbol{\theta}_B)_k$ indicates that we are using the k-th element of that vector. The weight coefficient, σ_{Iik}^{-2}, in this frame is just the inverse of the variance of the quantity $(\theta_{Iik} - (\boldsymbol{A}_{IiB}\boldsymbol{\theta}_B)_k)$. The significance of Eq. 7.16 is that it allows us to treat the various measurement residuals as uncorrelated, with each component θ_{Iik} being here treated as an independent measurement. Clearly we're adjusting our focus here as to what we call a measurement. To the tracker's embedded software, the raw direction vectors are measurements. To the flight software outside the tracker that receives $\boldsymbol{\theta}_{Ii}$ (or something equivalent) as output from the tracker, $\boldsymbol{\theta}_{Ii}$ is a measurement.

Equation 7.16 can be recast as

$$L(\boldsymbol{\theta}_B) = \sum_i [\boldsymbol{\theta}_{Ii} - (\boldsymbol{A}_{IiB}\boldsymbol{\theta}_B)]^T \boldsymbol{P}_{Ii}^{-1} [\boldsymbol{\theta}_{Ii} - (\boldsymbol{A}_{IiB}\boldsymbol{\theta}_B)] \tag{7.17}$$

where

$$\boldsymbol{P}_{Ii} = \begin{pmatrix} \sigma_{Ii1}^2 & 0 & 0 \\ 0 & \sigma_{Ii2}^2 & 0 \\ 0 & 0 & \sigma_{Ii3}^2 \end{pmatrix} \tag{7.18}$$

Now, using $\boldsymbol{P}_{Bi} = \boldsymbol{A}_{BIi}\boldsymbol{P}_{Ii}\boldsymbol{A}_{BIi}^T$ (cf. Eq. 7.15), the fact that $\boldsymbol{A}_{BIi}^T = \boldsymbol{A}_{BIi}^{-1} = \boldsymbol{A}_{IiB}$, and the generic matrix relationship that $(\boldsymbol{M}_1\boldsymbol{M}_2)^T = \boldsymbol{M}_2^T\boldsymbol{M}_1^T$, Eq. 7.17 can finally be written as

$$L(\boldsymbol{\theta}_B) = \sum_i (\boldsymbol{\theta}_{Bi} - \boldsymbol{\theta}_B)^T \boldsymbol{P}_{Bi}^{-1} (\boldsymbol{\theta}_{Bi} - \boldsymbol{\theta}_B) \tag{7.19}$$

The least-squares solution that minimizes this loss function is

$$\boldsymbol{\theta}_B = \left[\sum_i \boldsymbol{P}_{Bi}^{-1}\right]^{-1} \left[\sum_i \boldsymbol{P}_{Bi}^{-1}\boldsymbol{\theta}_{Bi}\right] \tag{7.20}$$

$$\boldsymbol{P}_B = \left[\sum_i \boldsymbol{P}_{Bi}^{-1}\right]^{-1} \tag{7.21}$$

Equation 7.20 is a kind of weighted averaging of the set of individual attitude solutions $\{\boldsymbol{\theta}_{Bi}\}$ that takes into account the correlation between the elements of the individual solutions. For simplicity, imagine a case in which we only cared about a single spin angle for a spacecraft, in which case each vector $\boldsymbol{\theta}_{Bi}$ would reduce to a scalar θ_{Bi}, the associated covariance $\boldsymbol{P}_{Bi}$ would reduce to a scalar variance σ_{Bi}^2, and the solution would reduce to a simple weighted average: $\theta_B = \Sigma_i \sigma_{Bi}^{-2}\theta_{Bi}/(\Sigma_i \sigma_{Bi}^{-2})$. If we simplify this example even further, letting all of the individual measurement variances be equal to a single common value σ_{Bc}^2, the solution becomes: $\theta_B = \Sigma_i \theta_{Bi}/N$, $\sigma_B^2 = \sigma_{Bc}^2/N$, where N is the number of measurements.

The reason for this little detour is two-fold. First, these days it is fairly standard to have quaternion star trackers, as opposed to vector star trackers, as the fundamental data source for onboard attitude estimation. Quest, or an analogous algorithm, is likely to exist as an embedded software algorithm within each tracker, but would not be appropriate as the method for combining multiple solutions into a single solution. Second, combining two attitude solutions will prove to be an important part of our discussion of recursive attitude estimation in Sect. 7.9.

7.7 Combining an Attitude Solution with a Vector Measurement

So, we now know how to combine two attitude solutions. But what if what we have is an attitude solution plus some number of direction vector observations? If there are multiple independent vector observations, you can just combine them into a single attitude solution using the methods of Sect. 7.4, at which point we're back to Eq. 7.17, i.e., combining multiple attitude solutions. But what if we have an attitude solution plus a single vector observation? For example, suppose we have a solution from a single quaternion star tracker and a single observation from a high precision sun sensor and we wish to combine the two pieces of information into a single solution. How do we proceed with combining these two disjoint types?

The case of combining a 3-D attitude solution with a 2-D vector measurement is a little bit subtle. If you're willing to accept that it can be done, feel free to jump now to Sect. 7.8 - but with the knowledge that there are well defined procedures that allow you to modify a full attitude solution with information that in itself provides insufficient data to define the attitude fully. This is a specific example of a more general principle, one that we'll visit in more detail in later sections.

The Mathematics of Combining 3D with 2D Attitude Information

For those of you still with us, we'll take you through the exercise of modifying an attitude solution based on a single vector measurement. If you have a bit of trouble following the first time through, the most important point to keep in mind is that this exercise constitutes an existence proof. The fact that it can be done is more important for our purposes than the details of how it is done.

Recall the previous discussion regarding the "cat's eye" camera at the end of Sect. 7.4. We'll be using the generalized loss function for that case here, or at least a component thereof for a single vector measurement. The loss function for combining a prior attitude estimate with a single vector measurement is:

$$\begin{aligned} L(\boldsymbol{\theta}_B) = &(\boldsymbol{\theta}_{Ba} - \boldsymbol{\theta}_B)^T \boldsymbol{P}_{Ba}^{-1} (\boldsymbol{\theta}_{Ba} - \boldsymbol{\theta}_B) \\ &+ (\boldsymbol{\phi}_{Ioi} + \boldsymbol{H}_{Ioi}\boldsymbol{\theta}_B)^T \boldsymbol{P}_{Ii}^{-1} (\boldsymbol{\phi}_{Ioi} + \boldsymbol{H}_{Ioi}\boldsymbol{\theta}_B) \end{aligned} \tag{7.22}$$

The first term on the right hand side corresponds to the full, 3-D attitude estimate. It is in form like any single attitude measurement contribution in Eq. 7.19. The subscript "a" emphasizes that it is a full attitude solution. Since we are assuming that $\boldsymbol{\theta}_B$ is defined relative to our prior attitude estimate (which can be done without loss of generality), $\boldsymbol{\theta}_{Ba} = \mathbf{0}$. We included it in Eq. 7.22 to match the form used in equation (7–19). The second term on the right hand side is of the form of a single vector contribution as expressed in Eq. 7.13, with the subscript i corresponding to the new (i-th) vector measurement. Recall that

$\boldsymbol{\phi}_{Ioi} \equiv$ upper two components of a 3-vector $\mathbf{u}_{Ioi}$
$\mathbf{u}_{Ioi} \equiv \boldsymbol{A}_{IBi}\boldsymbol{A}_0\mathbf{u}_{EI}$

$\boldsymbol{A}_{IBi} \equiv$ body-to-pseudo-instrument alignment matrix constructed from the measured vector $\mathbf{v}_i$
$\boldsymbol{A}_0 =$ initial attitude estimate, in this case the estimate prior to incorporating the effect of $\mathbf{v}_i$
$\mathbf{u}_{Ei} \equiv$ ECI-frame reference vector
$\boldsymbol{H}_{Ioi} \equiv$ upper 2×3 part of a 3×3 matrix $\boldsymbol{U}_{Ioi}$

$$\boldsymbol{U}_{Ioi} = \begin{pmatrix} 0 & u_3 & -u_2 \\ -u_3 & 0 & u_1 \\ u_2 & -u_1 & 0 \end{pmatrix}_{Ioi}$$

$\boldsymbol{P}_{Ii} \equiv 2 \times 2$ matrix covariance for the measurement $\mathbf{v}_i$ defined relative to pseudo-instrument frame

The corresponding least-squares solution $\boldsymbol{\theta}_B$ that minimizes the loss function of Eq. 7.22 (given that $\boldsymbol{\theta}_{BA} = \mathbf{0}$) is

$$\boldsymbol{\theta}_B = [\boldsymbol{P}_{Ba}^{-1} + \boldsymbol{H}_{Ioi}^T \boldsymbol{P}_{Ii}^{-1} \boldsymbol{H}_{Ioi}]^{-1} [\boldsymbol{H}_{Ioi}^T \boldsymbol{P}_{Ii}^{-1} \boldsymbol{\phi}_{Ioi}] \tag{7.23}$$

$$\boldsymbol{P}_B = [\boldsymbol{P}_{Ba}^{-1} + \boldsymbol{H}_{Ioi}^T \boldsymbol{P}_{Ii}^{-1} \boldsymbol{H}_{Ioi}]^{-1} \tag{7.24}$$

Note that the form for these two equations is very similar to that of Eqs. 7.20 and 7.21, with $(\boldsymbol{H}_{Ioi}^T \boldsymbol{P}_{Ik}^{-1} \boldsymbol{H}_{Ioi})$ used in place of what one might have expected to be designated as $\boldsymbol{P}_{Bi}^{-1}$, and a contribution that one might have expected as $(\boldsymbol{P}_{Ba}^{-1} \boldsymbol{\theta}_{Ba})$ in Eq. 7.17 eliminated because $\boldsymbol{\theta}_{Ba} = \mathbf{0}$. Note that if the reference vector, $\mathbf{u}_{Ei}$, when transformed into the pseudo-instrument reference frame exactly aligns with the measurement vector, then $\boldsymbol{\phi}_{Ioi} = \mathbf{0}$, with the implication that the attitude estimate will not change as a consequence of the new vector measurement. This is as it should be.

This may have seemed like a lot of effort to incorporate the use of a single extra vector measurement. We have an ulterior motive, which will become clear in Sects. 7.9 through 7.11. As previously noted, the principal message at this point is simply that there are well defined procedures for incorporating such data. More generally, it is always possible to combine measurements of different types so long as you have a well-defined specification of the uncertainty, i.e., covariance, for each. You can always use a new measurement to improve your estimate of a quantity of interest, even if that new measurement by itself only partially defines the full quantity of interest.

7.8 Measurement Propagation and De-Weighting

The algorithms presented in Sects. 7.4 and 7.6 assume that all of the observations are in, or can be transformed to, a single spacecraft reference frame and that you are only interested in determining a single attitude solution. If the spacecraft was rotating during the period of time in which the observations were collected, you must first rotate the various observations to a common reference frame using gyroscope data and the techniques discussed in Sect. 7.1. This is numerically straightforward, but

there is a complication involved that we did not consider in the earlier discussion. As for any sensor, gyroscope measurements will have some associated noise. Indeed, there are two kinds of noise typically associated with gyroscopes: (a) random noise in the immediate measurement being made, and (b) a slow, random-walk drift in a bias of the device. The latter can typically be kept small by regular calibration, but the former is an intrinsic feature of each measurement.

The implication of there being noise in the gyro measurements is that you really can't fully trust data that have been propagated to some selected common epoch, with the reliability being more and more degraded the greater the time period over which the propagation is done. Let's suppose that we have a set of vector measurements (star directions, Sun direction, magnetic field direction, whatever), all made with the same sensor or sensor type, but over some extended period of time during which the spacecraft is possibly rotating. Even if each measurement has the same intrinsic accuracy at the moment of its being made, the measurements will have different relative accuracies after being rotated with noisy gyroscope data to some common epoch. Not only will the measurements have different relative accuracies, their ratios will change as a function of the selected common reference time to which all data are to be transformed.

If the raw measurements are direction vectors, as in the case of the algorithms discussed in Sect. 7.4, the variance σ_i^2 to be assigned to the ith vector measurement after transformation to the common frame will be something like the following:

$$\sigma_i^2 = \sigma_D^2 + \sigma_R^2 + \sigma_G^2 \tau \tag{7.25}$$

where

σ_D^2 = the intrinsic variance associated with the direction measuring device
σ_R^2 = the variance associated with the reference measurements
σ_G^2 = the gyroscope measurement variance
t_i = the time of the vector measurement
t_c = the common reference time
$\tau = |t_c - t_i|$

The form of Eq. 7.25 is approximate. It models the immediate random noise of a gyro measurement, but ignores the error associated with the slow drift that usually also effects gyroscopes. It is, however, good enough for the purposes of our qualitative discussion. The use of the absolute value in the definition of τ emphasizes the fact that gyro-based propagation can work in either direction in time. Using the weighting scheme implied by Eq. 7.25 (i.e., with weight $w_i \propto \sigma_i^{-2}$), attitude solutions will depend most strongly on data taken close to t_c, the common reference time.

If your "measurements" are attitude solutions as provided by a quaternion tracker (as envisioned in Sect. 7.6), the de-weighting of such a solution due to gyro-measured rotation of the spacecraft is conceptually essentially the same as for the vectors. Just as σ_i^2 represents the variance of a vector measurement, $\boldsymbol{P}_i$ represents the covariance of the attitude components of an attitude solution at time t_i. The corresponding transformation of $\boldsymbol{P}_i$ to the common reference time based on gyro data is given by

$$\boldsymbol{P}_{ci} = \boldsymbol{R}\boldsymbol{P}_i\boldsymbol{R}^T + \sigma_G^2\tau\boldsymbol{I}_3 \tag{7.26}$$

where $\boldsymbol{R}$ is the rotation matrix constructed from the gyro data over the time period between t_i and t_c. Note that the $\boldsymbol{R}\boldsymbol{P}\boldsymbol{R}^T$ part of Eq. 7.26 is just a reference frame transformation, i.e., the equivalent of Eq. 7.15. After the rotation, each component of the currently estimated attitude has its uncertainty augmented by gyro error in just the same manner as would a propagated direction measurement.

To be fair, both Eqs. 7.25 and 7.26 involve a kind of averaging out of gyro noise in all directions. Things could be made to look more complicated if the details of the actual mounting alignments of the gyroscopes were preserved in detail, but doing so would add nothing essential to your understanding or to algorithm accuracy. The fact that the gyro noise factor $(\sigma_G^2\tau)$ is multiplied by $\boldsymbol{I}_3$ in Eq. 7.26 is an indication of the simplifying assumption that gyro noise is isotropic, i.e., uniform and uncorrelated in all directions. The actual mounting direction of the set of gyroscopes has been washed away in this model.

There's another simplification in Eqs. 7.25 and 7.26 that we won't explore in detail. If multiple measurements are propagated to the common reference time from the same side of that time (i.e., its future or its past), those measurements will all have some level of noise correlation due to having been propagated in part with the same noisy gyro data. Incorporation of this correlation can be done, but would make the math messier. Fortunately, that correlation is automatically incorporated in the method of recursive estimation to be discussed in the remaining sections.

So, what is the effect on our knowledge due to this gyro noise? Let's assume that our measurements (whether as direction vectors or Euler angle triplets) were made somewhat uniformly over a given time window and that the window is much longer than the time scale for gyro forgetfulness. We'll then have a situation in the deep interior of the window where there will be a kind of equilibrium for any reference time such that the associated attitude estimate is made with some effective number of fully weighted measurements, N_E. It would be sort of as if you didn't apply the $(\sigma_G^2\tau)$ term, but rather applied a sliding window for inclusion of a set of N_E fully weighted measurements. As you approach either end of the full data window, there are fewer effective measurements available. At the ends, the number of effective measurements drops by a factor of 2 relative to the deep interior, so the accuracy of attitude estimates at the window ends will be degraded by about a factor of about $\sqrt{2}$ relative to that of the deep interior.

This kind of time-symmetric treatment of attitude estimation may be appropriate for ground-based engineers and scientists wanting to process a whole set of previously acquired data that depend for optimal processing on a best possible estimate at any given time. For flight software, which is pretty much always only concerned with the here-and-now for its control purposes, we are really only concerned with that asymmetric moment when we have past and possibly current measurements, but nothing yet in the future. The flight engineers for the SPARTAN program considered placing a wrapper around the Quest algorithm for precisely this purpose. As outlined above, gyro data would be used to propagate past measurements to the current time, with appropriate de-weighting of past data in accordance with Eq. 7.25. Computer

memory is not infinite, so sufficiently old (degraded) measurements would eventually be discarded. This effectively functions as a "poor man's" recursive filter, an approach we'll discuss in some detail in Sects. 7.9 and 7.10.

For more information on propagation of uncertainty, see e.g. Sect. 3.7 of Applied Optimal Estimation by Gelb et al.

7.9 Recursive Attitude Estimation

The basic ideas for recursive attitude estimation were outlined in the last paragraph of the preceding section. Specifically, you must perform the following steps, or something equivalent to them:

1. Transform prior measurements to the current reference frame (e.g., rotate the measured direction vectors using spacecraft rate information based on gyroscope data);
2. De-weight the prior data so past solutions can be gradually forgotten; and
3. Combine the old data with new data at the current time to get a new estimate of the spacecraft attitude

There are two difficulties with the procedure as stated in the preceding paragraph. First, it envisions keeping every past measurement over time, or at least until the acquisition time for a given measurement passes out of a somewhat arbitrarily selected time window. This, at least in principle, implies allocating a potentially very large data store for the purpose of attitude estimation. Second, the procedure requires continuous reprocessing of each measurement, at least until it passes out of the time window. These difficulties are avoided by recognizing that all of the information that you care about from the prior measurements has been captured in the current estimate and its associated covariance matrix. You don't have to keep all of the old measurements if the recursive estimation steps can be applied using the immediately prior solution and covariance, rather than the set of "all" past measurements, hence our use of the phrase "or something equivalent" before listing the three steps of recursion.

With the use of the prior solution in mind, the steps of recursion become:

1. Transform the preceding attitude solution to the current reference frame.
2. Transform and de-weight the covariance matrix of the prior solution in a way that captures the uncertainty in the transformation process.
3. Combine the transformed prior solution together with the new measurement, with each assigned appropriate statistical weighting. The weighting of the prior solution will be some function of the transformed, de-weighted covariance matrix. The weighting of the new measurements will be some function of the variance of the new measurement.

We actually already have covered all of these steps. Step 1, transforming an attitude solution using gyro data, was discussed in Sect. 7.1. Step 2, transforming and

de-weighting the prior covariance matrix, was discussed in Sect. 7.8 and uses Eq. 7.26. The only change is that rather than transforming the prior solution to some common reference time, it is transformed to the next time step (t_{i+1}). Step 3, combining the transformed prior solution with new measurements, is the same as the problem discussed in Sects. 7.6 and 7.7, which covered the cases of new data being in the form of an attitude measurement (e.g., from a quaternion star tracker) or a direction measurement (e.g., from a sun sensor).

For more information on recursive estimation, see e.g., Sect. 4.1 of Applied Optimal Estimation by Gelb et al.

7.10 Recursive Attitude Plus Gyro Bias Estimation

So far we have restricted ourselves to discussing procedures for estimating spacecraft attitude by itself. In practice, it is usually the case these days that onboard attitude estimation is combined with gyroscope drift bias estimation. This is done because gyroscopes typically have a bias that is slightly unstable, i.e., the bias changes slowly but unpredictably with time. As a consequence, your ability to use gyro data to propagate attitude estimates forward in time would slowly deteriorate if you didn't have a method to keep the bias calibrated. By solving for both attitude and gyro bias in a single algorithm, you keep both well known and the (usually) faster rate gyro data can be used with confidence until a new attitude estimate is available. The algorithm typically used is a specific instance of a quite general procedure for state estimation, where "state" can be defined as whatever the user needs it to be (in this instance, attitude plus gyro bias), provided that the measurements are actually sensitive to the state being estimated.

Within the algorithm definition, the unknown state is usually referred to as a *state vector*. This state vector is analogous to the concept of a position vector in 3D position space, although the space within which the state vector resides is an abstract, higher dimensional space with a sufficient number of dimensions to represent all of the state solution for which you are searching. The physical units of the state space elements need not be the same for all elements of the space. In the case of the state vector being the combination of the spacecraft attitude and the gyroscope system bias, the state space will be six-dimensional, with three dimensions having units of radians to characterize the attitude (i.e., Euler angles relative to a reference), and three dimensions having units of radians per second to characterize the bias of the gyroscope system.

There is often a bit of overloading of the concept of state vector. You can use it to refer to

- the full state (e.g., the attitude and gyro bias: $\{\boldsymbol{A}, \boldsymbol{\Omega}_B\}$)
- a correction to an initial estimate of the state (e.g., $\{\boldsymbol{R}(\boldsymbol{\theta}), \boldsymbol{\omega}_B\}$), where $\{\boldsymbol{A}_0, \boldsymbol{\Omega}_{B0}\}$ is your initial estimate, and $\{\boldsymbol{A}, \boldsymbol{\Omega}_B\}$ is equal to $\{\boldsymbol{R}(\boldsymbol{\theta})\boldsymbol{A}_0, \boldsymbol{\omega}_B + \boldsymbol{\Omega}_{B0}\}$); or
- a linearized representation of the correction (e.g., $\{\boldsymbol{\theta}, \boldsymbol{\omega}_B\}$).

The first and second meanings for "state vector" are isomorphic (i.e., informationally equivalent). The third is isomorphic to the other two in a linear approximation - sort of the way a tangent plane is equivalent to a surface to which it is tangent for small regions close to the tangent point. As was illustrated in the preceding sections, state estimation algorithms are often set up as iterative linear approximations, so all three meanings of state vector become isomorphic as the algorithm converges. For our problem, let's use the symbol $\mathbf{X}$ to represent the linearized state vector $\{\boldsymbol{\theta}, \boldsymbol{\omega}_B\}$ in the form of a 6×1 column matrix, and let's use $\delta\mathbf{X}$ to represent small errors in the components of $\mathbf{X}$.

The basic ideas for recursive state estimation were outlined in Sect. 7.9, albeit limited there to the attitude problem. To generalize, you must perform the following steps:

- Transform the preceding state solution to the current time frame.
- Transform and de-weight the covariance matrix of the prior solution in a way that captures the uncertainty in the transformation process.
- Combine the transformed prior solution together with the new measurement, with each assigned appropriate statistical weighting.

Of course, there is an implied step 0, creation of an initial estimate for $\mathbf{X}$. You could do this by producing two independent attitude estimates at time points separated by Δt, and then defining the gyro bias as whatever is needed to make the gyro-propagated first attitude solution agree with the second solution.

At this point, you've pretty much got the gist of how a recursive filter would work for solving for attitude and gyro bias. For those that enjoy math, stick with us; otherwise skip ahead to Sect. 7.11.

Estimating Attitude and Gyro Drift Bias Together

Let's proceed to step 1. We've already discussed transforming an attitude solution using gyro measurements to rotate the attitude using a rotation matrix, or quaternion, constructed from the gyro-provided rate estimate. What about the transformation of the gyro drift bias? The bias is most conveniently defined in the spacecraft body frame, so rotation of the spacecraft does not change the bias; no transformation for it is needed.

Step 2 is the generalized state equivalent of Eq. 7.26, in which the attitude covariance is rotated to a new frame and then de-weighted based on gyro noise. The parallel equation for the generalized state problem can be written as

$$\boldsymbol{P}_X(t_i) = \boldsymbol{\Phi}\boldsymbol{P}_X(t_{i-1})\boldsymbol{\Phi}^T + \boldsymbol{Q}_G \tag{7.27}$$

where $\boldsymbol{P}_X$, a 6×6 matrix, is the covariance of $\mathbf{X}$, $\boldsymbol{\Phi}$ (the "state transition matrix") is a higher dimension counterpart to the rotation matrix $\boldsymbol{R}$ constructed from gyro data, and $\boldsymbol{Q}_G$ is a higher dimensional counterpart to $\sigma_G^2 \tau \boldsymbol{I}_3$. Note that since processing always moves forward in time for this recursive filter, $\tau = t_i - t_{i-1}$, i.e., there is no need for the absolute value symbol.

Deriving the full details of $\boldsymbol{\Phi}$ is beyond the scope of this presentation; see the work by Lefferts, Markley, and Shuster[6] for the original derivation. However, we can provide a heuristic overview. Just as $\boldsymbol{R}$ rotates 3-vectors from one frame to another, $\boldsymbol{\Phi}$ does the equivalent for the error vector $\delta\mathbf{X}$ as a consequence of the passage of time. Over a small time period τ (all changes can be built up as a sequence of small changes over time), the spacecraft rate as measured by the gyroscopes will be constant, so $\boldsymbol{R}$ will be a function of the product $\boldsymbol{\omega}\tau$. $\boldsymbol{\Phi}$ becomes

$$\boldsymbol{\Phi} \approx \begin{pmatrix} \boldsymbol{R}(\boldsymbol{\omega}\tau) & \tau\boldsymbol{R}(\boldsymbol{\omega}\tau/2) \\ \boldsymbol{0}_3 & \boldsymbol{I}_3 \end{pmatrix} \tag{7.28}$$

with each term in the matrix on the right-hand side being a 3×3 matrix. The appearance of $\boldsymbol{R}(\boldsymbol{\omega}\tau)$ in the upper left means that the attitude error component in $\delta\mathbf{X}$ rotates like a vector (not too surprising given that $\delta\boldsymbol{\theta}$ is a vector in 3-space). The appearance of $\boldsymbol{I}_3$ in the lower right indicates that the bias error is not influenced by the rotation. This is because the bias is fundamentally fixed to the frame in which the gyroscopes are mounted, i.e., the body frame. The appearance of $\tau\boldsymbol{R}(\boldsymbol{\omega}\tau/2)$ in the upper right means that a gyro bias error at time t_{i-1} introduces additional attitude error at t_i. The division by 2 in $\boldsymbol{R}(\boldsymbol{\omega}\tau/2)$ indicates that the bias is averaged-in across the rotation, with the effect being to use the bias at approximately the mean point of the rotation. The "$\approx$" in Eq. 7.28 actually pertains only to the $\tau\boldsymbol{R}(\boldsymbol{\omega}\tau/2)$ term. All the other block components are exact. The appearance of $\boldsymbol{0}_3$ (a 3×3 matrix of zeros) in the lower left indicates that the bias error is independent of past attitude error.

Let's now consider the "de-weighting" matrix $\boldsymbol{Q}_G$. We mentioned that it is a higher dimensional counterpart to $\sigma_G^2\tau\boldsymbol{I}_3$. It's actually a bit more complicated. As mentioned in Sect. 7.8, gyro noise is usually characterized for analysis and modeling purposes by two parameters, not just the single σ_G^2. The two parameters are the gyroscope angle measurement variance (σ_{Gm}^2), and the gyroscope drift rate bias random walk variance (σ_{Gb}^2). The former characterizes the immediate gyro measurement noise, while the latter characterizes how the gyro bias randomly drifts away from its last calibrated value. With these two parameters, $\boldsymbol{Q}_G$ is typically given by[7]

$$\boldsymbol{Q}_G = \begin{pmatrix} \left(\sigma_{Gm}^2\tau + \frac{1}{3}\sigma_{Gb}^2\tau^2\right)\boldsymbol{I}_3 & \left(\frac{1}{2}\sigma_{Gb}^2\tau^2\right)\boldsymbol{I}_3 \\ \left(\frac{1}{2}\sigma_{Gb}^2\tau^2\right)\boldsymbol{I}_3 & \left(\sigma_{Gb}^2\tau\right)\boldsymbol{I}_3 \end{pmatrix} \tag{7.29}$$

Note that in the above formula, σ_{Gm}^2 corresponds to what we previously had simply referred to as σ_G^2, defined heuristically as "the gyroscope measurement variance". Since we are here concerned with the gyro bias as part of our state vector now, treating its drift noise carefully is more important, hence the introduction of σ_{Gb}^2. The exact form of Eq. 7.29 is actually not terribly important for our discussion. What

[6] Lefferts, E.J., Markley, F.L., and Shuster, M.D., Kalman filtering for spacecraft attitude estimation, Journal of Guidance and Control Dynamics, 5(5), 417–429 (1982).

[7] ibid.

is important is that, like the term $\sigma_G^2 \tau \boldsymbol{I}_3$ in Eq. 7.26, $\boldsymbol{Q}_G$ captures the de-weighting of prior knowledge as a consequence of gyroscope hardware noise.

There is a caveat that pertains to Eqs. 7.28 and 7.29. For this discussion, we have assumed a convention for gyro bias that it be added as a correction to the gyro-based spacecraft rate estimate. Some folks prefer a convention whereby the bias is subtracted, in which case a minus sign must precede the $\tau \boldsymbol{R}(\boldsymbol{\omega}\tau/2)$ in Eq. 7.28 and the $[(1/2)\sigma_{Gb}^2 \tau^2]\boldsymbol{I}_3$ terms in Eq. 7.29. If sometime in the future you see a variation of Eq. 7.29 using negative off-diagonal elements, don't worry; it's just that author's preferred convention.

We finally come to step 3, the combining of a propagated, de-weighted, prior solution with new measurement data to generate a new, current-time solution. This is basically a higher-dimensional counterpart to our discussion from Sect. 7.6. There we discussed combining an attitude solution with either (a) another attitude solution, or (b) an independent vector measurement. Variation (b) is in a sense most analogous to what we are considering for the attitude-plus-bias problem in that we wish to improve on a higher dimensional estimate by incorporating a new, lower-dimensional data type. That type can be either an attitude solution (say, from a star tracker) or a vector. The corresponding loss functions are:

$$\begin{aligned} L(\mathbf{X}) = &(\mathbf{X}_0 - \mathbf{X})^T \boldsymbol{P}_{X_0}^{-1} (\mathbf{X}_0 - \mathbf{X}) \\ &+ (\boldsymbol{\theta}_{Bi} - \boldsymbol{H'}_\theta \mathbf{X})^T \boldsymbol{P}_\theta^{-1} (\boldsymbol{\theta}_{Bi} - \boldsymbol{H'}_\theta \mathbf{X}) \end{aligned} \tag{7.30}$$

$$\begin{aligned} L(\mathbf{X}) = &(\mathbf{X}_0 - \mathbf{X})^T \boldsymbol{P}_{X_0}^{-1} (\mathbf{X}_0 - \mathbf{X}) \\ &+ (\boldsymbol{\phi}_{Ioi} - \boldsymbol{H'}_{Ioi} \mathbf{X})^T \boldsymbol{P}_{Ii}^{-1} (\boldsymbol{\phi}_{Ioi} - \boldsymbol{H'}_{Ioi} \mathbf{X}) \end{aligned} \tag{7.31}$$

The first term on the right-hand side of each equation reflects the tendency of the state vector $\mathbf{X}$ to remain fixed at the current estimate based on past data, i.e., our estimate for $\mathbf{X}$ will tend to stay equal to $\mathbf{X}_0$ unless the new data forces a change. $\mathbf{X}_0$ is included in the equations to make them appear formally like prior loss function variations that we've presented. However, since $\mathbf{X}$ is defined each cycle to be the change relative to our current estimate, $\mathbf{X}_0 = \mathbf{0}$. The second term in Eq. 7.30 is just the standard loss function contribution of an attitude measurement, with the slightly obfuscating replacement of $(\boldsymbol{H'}_\theta \mathbf{X})$ for $\boldsymbol{\theta}_B$. This substitution is done to make $\mathbf{X}$ appear formally as the variable in the loss function. The matrix $\boldsymbol{H'}_\theta$ is a 3×6 matrix that strips $\boldsymbol{\theta}_B$ out of $\mathbf{X}$, i.e.

$$\boldsymbol{H'}_\theta = [\boldsymbol{I}_3 \; \boldsymbol{0}_3] \tag{7.32}$$

We added the prime merely as a reminder that we're dealing with the 6-D state vector in these equations. Similarly, in Eq. 7.31, $\boldsymbol{H'}_{Ioi}$ is a 6-D augmentation of $\boldsymbol{H}_{Ioi}$ for the vector case:

$$\boldsymbol{H'}_{Ioi} = \left[\boldsymbol{H}_{Ioi} \; \boldsymbol{0}_{2\times 3}\right] \tag{7.33}$$

so $(\boldsymbol{H'}_{Ioi}\mathbf{X}) = (\boldsymbol{H}_{Ioi}\boldsymbol{\theta}_B)$, and the second term in Eq. 7.31 is just the "standard" loss function contribution for a vector measurement.

The solutions corresponding to Eqs. 7.30 and 7.31 are much the same as for Eq. 7.19. Using a new attitude measurement, you have:

$$\mathbf{X} = \left[\boldsymbol{P}_{X_0}^{-1} + \boldsymbol{H'}_{\theta}^{T}\boldsymbol{P}_{\theta}^{-1}\boldsymbol{H'}_{\theta}\right]^{-1}\left[\boldsymbol{H'}_{\theta}^{T}\boldsymbol{P}_{\theta}^{-1}\boldsymbol{\theta}_{Bi}\right] \tag{7.34}$$

$$\boldsymbol{P}_X = \left[\boldsymbol{P}_{X_0}^{-1} + \boldsymbol{H'}_{\theta}^{T}\boldsymbol{P}_{\theta}^{-1}\boldsymbol{H'}_{\theta}\right]^{-1} \tag{7.35}$$

Using a new vector measurement, you have:

$$\mathbf{X} = \left[\boldsymbol{P}_{X_0}^{-1} + \boldsymbol{H'}_{Ioi}^{T}\boldsymbol{P}_{Ii}^{-1}\boldsymbol{H'}_{Ioi}\right]^{-1}\left[\boldsymbol{H'}_{Ioi}^{T}\boldsymbol{P}_{Ii}^{-1}\boldsymbol{\phi}_{Ioi}\right] \tag{7.36}$$

$$\boldsymbol{P}_X = \left[\boldsymbol{P}_{X_0}^{-1} + \boldsymbol{H'}_{Ioi}^{T}\boldsymbol{P}_{Ii}^{-1}\boldsymbol{H'}_{Ioi}\right]^{-1} \tag{7.37}$$

Note how the pattern of loss functions and solution equations is a repeating pattern, reoccurring just in a slightly different key.

To be fair, we should say that the preceding equation pairs provide a solution for minimizing their respective loss functions. Each follows by solving the six simultaneous equations formed by setting the derivatives of L with respect to the components of $\mathbf{X}$ to zero, and each involves taking the inverse of a 6×6 matrix. There is a computationally more efficient approach to use. That will be the topic of the next section.

7.11 The Kalman Filter for Recursive Least Squares

There is nothing thematically new about a Kalman filter. It is just a particular variation on a recursive algorithm for state estimation that happens to be optimal with respect to efficiency. From a black box perspective (i.e., if we didn't let you take a peak under the hood), you would not be able to tell the difference between a Kalman filter and any other recursive algorithm designed to solve the same problem, except for the fact that the Kalman filter would prove to be most efficient with respect to CPU usage - not so important for today's very efficient machines, but extremely important when Rudolf Kalman[8] developed his approach in the early 1960s.

If you don't care about alternative solution methods after you've got one that works, there really is nothing more that you need to hear about the Kalman filter, except perhaps that it has become the flight software algorithm of choice for

[8] R. E. Kalman, A New Approach to Linear Filtering and Prediction Problems, Journal of Basic Engineering, March 1960, pp 35–46; or see also A. Gelb et al., Applied Optimal Estimation, MIT Press, 1974.

attitude-plus-gyro drift rate bias estimation on Goddard missions. Like Xerox and Kleenex, the specific name has become synonymous with the generic type. Folks in the NASA/GSFC flight software branch pretty much use "Kalman filter" to mean a recursive filter for attitude-plus-gyro bias estimation - well, at least until recently. Developments in GPS flight systems leverage Kalman filtering to refine onboard estimates for GPS-provided position and velocity estimates, but that's another story.

The Mathematics of the Kalman Filter

Both Eqs. 7.30 and 7.31 are of the general form

$$L(\mathbf{X}) = (\mathbf{X}_0 - \mathbf{X})^T \boldsymbol{P}_{X_0}^{-1} (\mathbf{X}_0 - \mathbf{X}) + (\mathbf{Z} - \boldsymbol{H}\mathbf{X})^T \boldsymbol{P}_Z^{-1} (\mathbf{Z} - \boldsymbol{H}\mathbf{X}) \tag{7.38}$$

where $\mathbf{X}$ is *any* generic state vector, $\mathbf{X}_0$ a current estimate thereof, $\boldsymbol{P}_{X_0}$ the covariance of $\mathbf{X}_0$, $\mathbf{Z}$ a new measurement, $\boldsymbol{P}_Z$ the covariance of $\mathbf{Z}$, and $\boldsymbol{H}$ transforms $\mathbf{X}$ into measurement space. Kalman derived the following equations that provide a computationally optimal solution for the problem:

$$\mathbf{X} = \left[\boldsymbol{I}_{N_X} - \boldsymbol{K}\boldsymbol{H}\right] \mathbf{X}_0 + \boldsymbol{K}\mathbf{Z} \tag{7.39}$$

$$\boldsymbol{P}_X = \left[\boldsymbol{I}_{N_X} - \boldsymbol{K}\boldsymbol{H}\right] \boldsymbol{P}_{X_0} \tag{7.40}$$

$$\boldsymbol{K} = \boldsymbol{P}_{X_0} \boldsymbol{H}^T \left[\boldsymbol{H}\boldsymbol{P}_{X_0}\boldsymbol{H}^T + \boldsymbol{P}_Z\right]^{-1} \tag{7.41}$$

where $\boldsymbol{I}_{N_X}$ is the identity matrix of size $N_X \times N_X$, N_X being the dimension of $\mathbf{X}$. The matrix $\boldsymbol{K}$ is called the *Kalman gain matrix*. It has dimension $N_X \times N_Z$, where N_Z is the dimension of $\mathbf{Z}$, typically smaller than the dimension of $\mathbf{X}$. The matrix to be inverted in Eq. 7.41 is only of dimension $N_Z \times N_Z$. This can significantly reduce the numerical complexity of the problem being solved (i.e., the number of computer steps required). For the problem discussed in Sect. 7.10, the matrix to be inverted is reduced from 6×6 to only 3×3 or 2×2, depending on whether the new measurement is in the form of an attitude solution or a direction vector. This is significantly more efficient, given that numerical procedures for inverting matrices typically require $\sim N^3$ operations.

It may seem like a mathematical slight-of-hand that the Kalman filter can adjust for changes in all N_X elements of $\mathbf{X}$ while only inverting an $N_Z \times N_Z$ matrix. This seems like solving for N_X (e.g., 6) unknowns with only N_Z (e.g., 2 or 3) equations. The "magic" in Kalman's method is that, within the full N_X-dimensional state space, it effectively finds the N_Z-dimensional subspace into which changes associated with measurement $\mathbf{Z}$ can be projected - i.e., it effectively reduces the scope of the problem (equivalent to reducing the rank of the matrix to be inverted) by restricting the search to that subspace of state space that is observable via the new measurement.

If we left the discussion of the Kalman filter at this point, some of you at some future time would eventually discover that we told a bit of a fib regarding the derivation of the Kalman equations, one that fits nicely into the story that we've been telling in this chapter. In point of fact, Kalman's derivation does not start by minimizing

the loss function defined in Eq. 7.38. Rather, his approach minimizes the trace (sum of diagonal elements) of $\boldsymbol{P}_X$, the resultant covariance matrix. It is not immediately obvious that the two approaches should yield the same solution. Indeed, it is not even obvious that the trace of $\boldsymbol{P}_X$ is a meaningful sum, given that the units of the components of $\mathbf{X}$, and therefore the units of $\boldsymbol{P}_X$, are not required to be the same. (For our attitude-plus-bias problem, the units of components 1–3 of $\mathbf{X}$ are radians, while the units of components 4–6 are radians per second.) However, it can be shown that an alternative derivation is possible (see Appendix B) that starts from Eq. 7.38 and arrives at Kalman's equations.

For those of you wondering about the term "filter," which we haven't previously used in our discussion, it is often applied to any algorithm used to remove (filter out) measurement noise, leaving the user with an optimal estimate for the sought for truth. We could compare any such filters to the lamp of Diogenes in his quest for truth in the form of an honest man - except that, to the best of our knowledge, he never found his sought for truth.

One final note. You will sometimes hear the phrase "extended Kalman filter." The adjective "extended" indicates that the system for which a state solution is being sought is non-linear, i.e., the filter is extended beyond a simple linear filter. If you think about it, you'll realize that this entire chapter has been dealing with non-linear least squares filters (the attitude matrix being a non-linear function of its independent variables), so there's nothing new or magic in the term "extended". It's just another way of saying non-linear, with the implication that iterative processing may be necessary for precise results, although in fact we never bother with iteration at each time step in the FSW context.

7.12 Synopsis

This chapter has covered quite a bit of material on attitude estimation, enough that it seems appropriate to give a brief synopsis. The topics and main points of each section were as follows:

1. Section 7.1 described how gyroscope data can be used to create rotation representations that allow you to transform attitude estimates and vector measurements from one time to another.
2. Section 7.2 served as a motivating introduction for the subsequent sections that deal with various attitude estimation algorithms.
3. Section 7.3 described the Triad algorithm, a simple method for combing two direction measurements to determine attitude.
4. Section 7.4 extended the discussion to include cases with arbitrarily many direction measurements. It presented a standard, simple (although not the most computationally efficient) algorithm for solving for the optimal attitude solution.
5. Section 7.5 introduced the concept of the covariance matrix, the mathematical object that formally describes the uncertainty in your estimate.

6. Section 7.6 described how you can optimally combine two independent attitude solutions together, e.g., solutions from two quaternion star trackers. The significance of the discussion was the use of a different measurement type for the input, i.e., 3-D Euler angle sets, as opposed to 2-D direction vectors.
7. Section 7.7 extended the discussion of Sect. 7.6 to the case of constructing an attitude solution based on an attitude solution and a direction vector measurement, e.g., for combining data from a quaternion star tracker with a high-precision Sun sensor. The significance of the discussion is that the input data types do not have to be identical.
8. Section 7.8 discussed the effects of gyroscope noise as you propagate either direction vector measurements or attitude solutions from one moment in time to another. Specifically, as you move away from the moment when measurements were originally obtained, those data and any solutions obtained from them become less reliable.
9. Section 7.9 introduced recursive attitude estimation, built up from the elements discussed in the preceding sections. This included rotating a prior attitude solution from a prior time, rotating the associated covariance from the prior time, de-weighting the prior covariance, and combining the old solution with new measurement data.
10. Section 7.10 extended the discussion of recursive attitude estimation to include estimation of gyroscope bias. This extension is important because gyro bias is subject to a slow drift. It is customary on most space missions to include continuous onboard adjustment for the changing drift. The mathematical formulation is very analogous to that of Sect. 7.7, where a measurement of dimension lower than that of the state of interest is combined with an old (now degraded) state estimate to keep the estimate optimally tuned.
11. Section 7.11 introduced the Kalman filter equations, which provide a highly efficient algorithm for solving any recursive least-squares problem.

7.13 Mathematics to English Translation of Kalman Filtering

Many battle memoirs were written after the war by World War II veterans, but one such memoir[9] contains a quote that seems particularly appropriate at this point: "If you survive your first day, I'll promote you." The quote comes from a conversation between the author (then a lieutenant) and his battalion commander. The casualties in his unit, especially among officers, had been so high that company commanders were dropping like flies. The author was the latest lucky fellow to receive command of that company, and to give the author something to shoot for (perhaps that could have been worded better), his commanding officer was guaranteeing him a promotion

[9] G. Wilson, "If You Survive: From Normandy to the Battle of the Bulge to the End of World War II, One American Officer's Riveting True Story", Paperback Edition: Ballantine Books, June 1997.

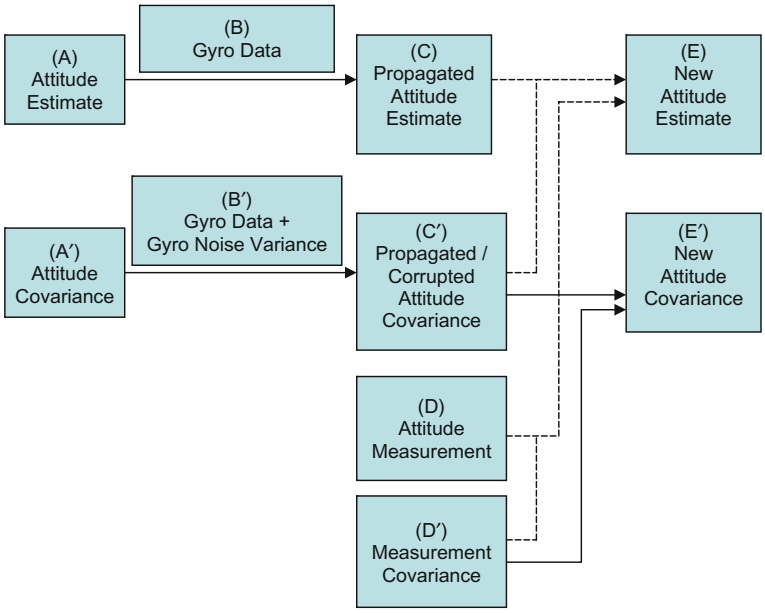


Fig. 7.2 Kalman filter for optimal attitude estimation

from lieutenant to captain (the usual rank for a company commander) if the author's luck proved better than that of his predecessors. If you've reached this section, either you've survived all the hazards of the previous sections, or you've decided to proceed directly to Sect. 7.13 without looking back. Either way, your reward is the math-free description of the Kalman filter algorithm that follows.

The Kalman filter is a recursive filter, which means that it is convenient to start our description in the middle of things (rather like *Star Wars* but without the excitement). Somehow, some way, you have an estimate (perhaps only of moderate accuracy) for the attitude of the spacecraft (box A in Fig. 7.2) at the current time. (For details regarding constructing an attitude estimate based on a set of direction vector data, see Sects. 7.3 and 7.4)

In addition to the attitude estimate itself, you need an estimate for the *covariance* of that estimate (box A'). The covariance is a mathematically precise way to express the uncertainty of your knowledge. When an estimate consists of just a single number (say, your height), the term *variance* is used, which is equal to the square of the typical error or *standard deviation*. When the estimate pertains to a conglomerate concept (such as an attitude, with its three degrees of information), the term *covariance* is used, which captures both the variance of the individual components and the various correlations between the errors of the components. For further details regarding the meaning of an attitude covariance, see Sect. 7.5.

Time passes and the attitude changes as a consequence of torques imposed by the outside world or due to control torques imposed by the attitude control system. The

attitude determination system is a passive part of the overall attitude control system. It doesn't really care why the attitude is changing; it merely observes that it does. The change in attitude is typically monitored using gyroscopes. Using gyroscope data over some short period of time (box B), the Kalman filter determines how the previous attitude estimate should be changed as a consequence of the rates measured by the gyroscopes, the result being a propagated attitude estimate (box C). For details regarding attitude propagation using gyroscope data, see Sect. 7.1.

The Kalman filter also needs to determine the covariance of the propagated attitude estimate. There are two aspects to this. First, even if the gyroscope data were perfect, the attitude covariance would change as a consequence of the attitude itself changing. The Kalman filter basically rotates (propagates) the previous covariance into a new reference frame to follow the changing attitude. Second, because the gyro data are in fact not perfect, the Kalman filter adds a noise corruption term to the rotated attitude covariance matrix to account for the gyro measurement noise. The gyro data used for rotation and gyro noise used for corruption are indicated in box B'. The final propagated and corrupted covariance is indicated in box C'. For details regarding rotating and corrupting a covariance matrix using gyroscope data and associated gyroscope noise parameters, see Sect. 7.8.

If no other data were available, the Kalman filter would just continue this process over time, with the gyro noise continuing to corrupt the attitude estimate. As time passed, the attitude estimate would continue to degrade (the covariance would continue to increase). To avoid this, you must periodically introduce new measurements of the spacecraft attitude (say, from a quaternion star tracker), as indicated in box D. This measurement actually doesn't have to be sufficient by itself to determine a complete attitude solution. A measurement that by itself only partially determines attitude (say, of the Sun direction from a Sun sensor) can be used to improve the degraded estimate partially. Along with the measurement, you will need an estimate of the covariance of that measurement, as indicated in box D'. The new measurement is combined with the propagated prior attitude estimate to produce a new attitude estimate (box E), with each input given a weighting factor based on how well it is known. Greater knowledge is associated with smaller covariance, so the weight assigned to each input component is basically inversely proportional to that component's covariance. The two covariances are combined together to produce a covariance for the new attitude estimate (box E'). For detail regarding optimally combining an attitude estimate with a new measurement, either a complete attitude measurement or a partial attitude measurement in the form of a direction vector, see Sects. 7.6 and 7.7.

Having constructed a new, improved attitude estimate (along with its covariance), you're ready to begin the cycle again. More gyro data are used to propagate and corrupt the new solution across another time step, and another new attitude measurement is used to rein in the corruption. For typical mission scenarios, this process eventually converges to an equilibrium state, with the amount of gyro-related corruption introduced in each cycle cancelled out by the incorporation of the new attitude data. For a discussion of combining the pieces of a recursive attitude filter into a unified whole, see Sect. 7.9.

At this point, those of you who didn't slog through the previous sections' math are perhaps wondering what happened to the determination of the gyro drift rate bias mentioned in the first paragraph of this section. Well, the algorithm just presented remains the same from a high-level perspective. All you have to do is mentally modify Fig. 7.2 so that everywhere that it refers to an attitude estimate or an attitude covariance (i.e., boxes A, A', C, C', E, and E'), you read it as a state estimate or a state covariance, and then define "state" to mean attitude and gyro drift rate bias together. The mathematical details for solving for this extended state are more complex, but the high-level concepts remain unchanged. Note that boxes B, B', D, and D' remain unchanged; the input measurements remain as envisioned in the previous discussion. For details regarding both the attitude and gyro bias in the "state" to be solved for in a recursive filter, see Sect. 7.10.

Regarding the distinction between a recursive filter and a Kalman filter, the former is the more general term with the latter being a specific realization thereof. From a black box perspective, there is no distinction. However, in situations where the state for which one is solving is more complex than the individual measurements being used as input, the Kalman filter is the most efficient realization of a recursive algorithm. For this reason, use of a Kalman filter for combined attitude and gyro bias estimation has become standard for GSFC flight software. For the mathematical details that show how a Kalman filter is more efficient than a naively realized recursive filter, see Sect. 7.11.

Chapter 8
Spacecraft State Estimation More Broadly

There's an old physics joke that compares the underlying philosophies of three major physics theories to attitudes (no pun intended) of three distinctly different baseball umpires. The first umpire is Albert Einstein who, when asked his basis for calling balls and strikes, says "I call them the way I see them". This approach is consistent with the equivalence of inertial reference frames in special relativity, which holds all inertial reference frames are equivalent and the measurements of time and distance in one inertial frame are just as valid as those made in another inertial frame, even if the numbers associated with the measurements are not the same in both frames. The second umpire is Isaac Newton who says "I call them the way they are", which harkens back to the pre-relativity days when physicists believed there was such a thing as absolute time and distance measurements, and a preferred frame in which these measurements could be specified. The third umpire is Neils Bohr who says "They ain't nothing until I call them", which is a somewhat eccentric way of describing the role played by superposition of states in quantum mechanics and the collapse of the wave function to a specific state when a measurement is made. With respect to the ACS calibration function, Isaac Newton is the spacecraft manufacturer who can make direct measurements of sensor calibration parameters in the preferred reference frame of the manufacturer's lab(s). Albert Einstein is the ground system ACS analyst who relies on downlinked sensor data for his relative alignment calibrations. Unlike the manufacturer, the ACS analyst cannot make a direct measurement (and therefore cannot make an absolute alignment calibration relative to a manufacturer specified absolute reference frame), but instead can only compare data from different sensors (say, two star trackers relative to each other or gyros relative to star trackers) and calibrate the alignment of one sensor relative to another. For this purpose, any of the sensors can serve as the fiducial sensor relative to which the alignments of the other sensors can be specified. Neils Bohr in this context is the spacecraft itself, for whom none of these calibration parameters have any relevance until its sensors measure something.

© Springer-Verlag London Ltd. 2017

H.L. Hallock et al., *ACS Without an Attitude*, NASA Monographs
in Systems and Software Engineering, DOI 10.1007/978-1-4471-7325-0_8

In this chapter, we all get to be Einstein as we struggle to interpret Bohr's measurements and peer through a glass darkly (or at least through a distorted lens) to determine how influences like launch stresses have changed Newton's absolute calibrations. While Chap. 7 focused on attitude estimation, here we'll broaden our perspective to include other examples of interest to attitude control engineers, though for the most part not currently realized in flight software. The examples presented below are not intended to be an exhaustive list, but rather just enough to give a larger perspective on the use of the least-squares techniques discussed in Chap. 7 as related to other aspects of spacecraft control.

8.1 Attitude-Related Least Squares Problems

Chapter 7 focused on various techniques for determining spacecraft attitude, with gyro bias brought in as an auxiliary part of the state vector in Sect. 7.10. Along the way we touched upon pretty much all of the important techniques associated with the least-squares methodology. If you are employed as a flight software attitude system developer (engineer, programmer, whatever), it is fairly likely that you'll encounter other related problems dealing with the calibration of attitude hardware. Currently, most of such calibrations are done with software utilities hosted in the spacecraft ground system, although this can change depending on mission needs, the time scale for variation of the various calibration parameters, and the power of the onboard computer(s). Since we have already covered the basic least-squares techniques, our goal is just to provide a synopsis of other related calibration problems that we have encountered in our careers as attitude systems engineers.

8.1.1 Star Tracker Relative Alignments

Determining the relative alignment of star trackers is pretty much the same problem as determining the attitude of a spacecraft based on star tracker data. If you can determine the attitude of two or more star trackers relative to the ECI frame at a given moment of time, then you can construct their alignment relative to each other at that moment. If you make the assumption that the alignments are constant over some extended period of time, then a series of temporally coincident attitude measurements can be used together - either via batch or recursive processing - to determine the relative alignments with improved precision. Something that's not so obvious, but should be clear with a bit of thought, is that if you are using vector star trackers (as opposed to quaternion trackers), simultaneous measurements of a single star in each tracker provides useful information for determining the relative alignments, although in isolation this is insufficient to provide a complete solution for determining the relative alignments or the attitudes of the star trackers individually. If, however, you obtain another similar data set, i.e., one star measurement per tracker but

with each measurement in a different field of view location, you would have sufficient information to determine the relative alignments of the trackers. This is true even if the spacecraft attitude has changed significantly between measurements, thereby precluding you from determining the attitude of the spacecraft from individual tracker measurements at either attitude. Of course, after you know the relative alignments of the tracker set, you can treat the tracker set as a unified sensor and reprocess the data set to determine the attitude of the sensor set relative to ECI at either attitude. By defining the reference frame of the spacecraft relative to that of one of the star trackers, you thereby also know the spacecraft attitude.

8.1.2 Star Tracker Internal Calibrations

The "internal" calibrations of a star tracker are those for the tracker considered in isolation from any other sensors. Specifically, these pertain to field of view scale, field of view distortion, sensitivity to star brightness, and sensitivity to star spectral distribution. For scale and distortion calibration, you need a series of measurement sets with each set consisting of at least two essentially simultaneous star direction measurements for stars in well-separated positions in the field of view. Of course, you also need a reference catalog against which to compare the measurements. (Strictly speaking, the statement that you need a reference star catalog isn't true for distortion calibration, although significantly more data under special observing conditions is required. For the HST program, such a catalog-free approach was designed for the distortion calibration of the observatory's fine guidance sensors.) If all you need is scale calibration, a single pair of star measurements is in principle adequate, with best results obtained using widely separated stars. For a good estimate of distortion, you need good coverage across the whole field of view. If the measured star positions are sensitive to local magnetic flux (they can be), and if the spacecraft is moving in a region of significant magnetic flux (as those in low Earth orbit typically are), you will need to include local magnetic field measurements in the calibration model.

Although a star tracker's sensitivity to stellar brightness and spectrum can in principle be a function of star position in the field of view, in practice one does not find such field dependence. To perform these calibrations, you need measurements and associated star catalog reference data spanning the ranges of interest for star brightness and spectrum.

8.1.3 Gyroscope Calibration

Gyroscopes are calibrated for the following parameters: alignment, scale, distortion, and bias. Typically, a "measurement" for gyro calibration is the attitude change perceived by the spacecraft gyroscope system over a maneuver, where the term "maneuver" in this context can include an extended period of constant attitude. The reference

information is the actual change in attitude as determined by a set of calibrated star trackers. If distortion is negligible for the gyroscopes of interest, full observability for gyroscope alignment, scale factor, and bias can be determined using four independent measurements–say one maneuver about each of the spacecraft principal axes, plus a period of constant attitude to fix the gyroscope bias. Gyroscope non-linearity (i.e., distortion) has been found to be an issue for some missions. Although significant, the effect is typically not large, so the inclusion of a small quadratic correction (or something effectively equivalent) for the conversion of gyroscope counts to rate about a gyro axis is sufficient. With quadratic corrections added, you need to use a set of maneuvers that sample both positive and negative rotations about each gyro axis, so a minimum measurement set will consist of seven rather than four independent measurements, say one positive and one negative maneuver about each of the spacecraft principal axes plus the period of constant attitude. This approach was used for HST and achieved calibration accuracies in the scale factor and alignment of $\sim$0.3 arcsecond per degree of slew when ignoring scale factor non-linearity, and $\sim$0.04 arcsecond per degree of slew when taking scale non-linearity into account. The data also showed a significant scale factor drift over periods of order half a year,[1] which emphasizes the need to periodically recalibrate the sensors over the course of the mission.

8.1.4 Sun Sensor Calibration

Sun sensors can be calibrated for alignment and distortion, where the latter pertains to the apparent Sun direction in the sensor field of regard.[2] The reference information for such calibration is typically the actual spacecraft attitude as measured from, say, a set of star trackers, and the true direction to the Sun in ECI coordinates as obtained from models of the Earth's orbit about the Sun and the spacecraft's orbit about the Earth. For a good calibration, you must obtain measurements of the sun ranging across the whole field of regard of the sensor.

8.1.5 Magnetometer Calibration

Magnetometers can be calibrated for alignment, distortion, bias, and sensitivity to magnetic torquer bars (if present). The bias is often not so much a bias in the sensor, as a magnetic field of the spacecraft as a whole as perceived by the sensor - but it

[1] G. Welter, J. Boia, M. Gakenheimer, E. Kimmer, D. Channel, & L. Hallock, *Variations on the Davenport Gyroscope Calibration Algorithm* in the proceedings of the 1996 GSFC Flight Mechanics/Estimation Theory Symposium.

[2] field of regard - from www.theFreeDictionary.com: "all of the points of the physical environment that can be perceived by a stable eye at a given moment" - so, just replace "eye" with "sensor" in our current context.

amounts to essentially the same thing as far as use of the sensor data are concerned. The reference information for such a calibration is typically the actual spacecraft attitude as measured from a set of star trackers, and the true direction of the Earth's magnetic field at the location of the spacecraft (if we're dealing with a low Earth orbiter) at the time of the measurement. For a good calibration, you must obtain measurements of the Earth's magnetic field across its full range as detectable by the magnetometer, as well as measurements of the MTB field as perceived by the magnetometer over the operating range of the bars.

8.1.6 Wavefront Calibration

For some missions (e.g., the Hubble Space Telescope and the James Webb Space Telescope), spacecraft fine pointing is determined using a sensor or group of sensors placed in the telescope's main optical path. Calibrating the sensor(s) then becomes intrinsically involved with calibration of the telescope optics. This calibration may then not be just a matter of the determining the sensor's operating characteristics, but rather may be used to provide feedback to correct the alignment of the optical elements of the observatory. (Imagine a visit to an optometrist who is flipping those lenses and asking which image looks better.) For calibration of an observatory's optical wavefront, you can compare the observed image of a star (including the pattern of diffraction rings and axial spikes) with the image predicted based on the design of the observatory, where the state vector is given by the displacements of the various optical elements from their design positions. As the optics engineer determines these various displacements, he'll use them to send commands to the optical element actuators to adjust their positions. Afterwards, new star image measurements will be made to verify that the image has in fact improved (e.g., tighter focus, or smaller astigmatism or coma[3]).

8.2 General Issues

The rest of this chapter explores aspects of state estimation broadly. For readers anticipating encountering other state estimation problems in their careers (e.g., calibration of various sensors), you may want to stick with us for this discussion. You won't need it, though, for later chapters. If you feel you've had enough of state vectors, covariances, transition matrices, and the like, feel free to jump ahead to Chap. 9; we'll meet you on the far side. For those of you game to continue, the rest of this chapter will be devoted to some further considerations on estimation theory.

[3] Astigmatism and coma are aberrations in an optical system, the former resulting when the system has different foci for rays that propagate in two perpendicular planes, and the latter manifest as a variation in magnification across the entrance pupil.

8.2.1 Observability

We first want to address the issue of state vector observability, i.e., whether the data that you have accumulated has sufficient information content to actually allow you to solve for the complete state vector. None of the algorithms discussed in Chap. 7 have built-in constraints enforcing the logical requirement that you provide adequate information to support the solution goals. You must make sure that the measurements you obtain contain sufficient information to determine the full set of state vector elements. As a simple example, if you continually use measurements of the same star direction and nothing else, you will have no knowledge of spacecraft rotation about that vector. The algorithm won't be able to cobble it out of thin air.

The example just used is fairly obvious. Sometimes lack of information is more subtle. Suppose, for example, that you have a pair of star cameras and want to use observations of stars, together with a catalog of the star coordinates, to determine both the relative alignment of the cameras and the optical distortion to which they are subject. It is fairly typical to characterize distortion as a two-dimensional polynomial across the field of view of the camera. However, if the polynomial contains linear terms, changes in those terms would be indistinguishable from alignment shifts. Trying to solve simultaneously for both types of parameters would result in an unsolvable set of redundant equations, manifest as a singular matrix with no inverse. You have to be careful to avoid defining a state vector in such a way that some of its elements are linear combinations of other elements.

Perhaps more problematic is when some of the elements are almost, but not quite, linear combinations of other elements, at least relative to the measurement set being used. In such a case, rather than having a well-designed matrix inversion routine catch the singularity (an attempt to divide by zero), your software will provide a solution, albeit possibly one that looks radically different from the initial estimate (possibly an estimate that you know to be fairly good). In this case, the algorithm is, at least in part, just fitting noise - i.e., it is finding the formally best solution vis-à-vis loss function minimization from within a region of state space for which all of the solutions have essentially correspondingly small loss function values. If you're lucky, you'll catch such behavior during algorithm development. This may merely lead to excessive consumption of alcoholic beverages or over-the-counter headache nostrums as you try to code up a fix to stabilize your solution. An alternative, stumbling into the problem during operations, may leave you erroneously thinking that your hardware has significantly physically changed somehow, a cause of possible serious confusion, angst, and yet more consumption of alcohol.

Of course, you may actually be interested in studying a problem for which interesting and distinct state parameters cannot be readily resolved with your available data set. Ultimately, the only real solution is to study the problem carefully and design a complimentary experiment that, together with the original data set, allows you to isolate variations in the state vector terms.

8.2.2 State Vector Selection

As mentioned earlier, the state vector is the set of observables for which you're solving. For folks in our field, it typically consists of things like sensor-related parameters (biases, misalignments, etc.), orbital parameters (e.g., the in-track error), attitude propagation parameters, and the like. The nature of the state vector elements will also vary depending on how you plan to use them. For example, if you are using a kinematic model to represent attitude change, the state vector elements do not include internal or external torque parameters. With nothing in the state vector explicitly describing how the angular momentum vector changes with time, you can compute the modeled attitude directly from the state vector elements. By contrast, if you are using a dynamic model, some elements of the state vector will contain torque-related parameters (e.g., the spacecraft magnetic dipole moment). Therefore, for a dynamic model, you do have information in the state vector pertinent to how the spacecraft angular momentum vector will change with time, and in order to utilize this information you need to integrate the spacecraft equations of motion for your optimal attitude prediction.

From all this discussion, it's clear that there's a large candidate population to choose from when constructing a state vector, so the obvious question is how do you select the elements of your state vector? Well the first guideline is don't get overly ambitious and try to solve for too much. As already noted, elements that are not fully independent can be difficult to distinguish from each other. That in turn can lead to convergence problems, or even convergence to rather peculiar looking solutions when one set of state vector elements "takes off" in one direction and another group starts moving oppositely such that the overall residuals are minimized but the solution is very data dependent (the "noise fitting" problem previously discussed).

Of course, if the input data is of insufficient quantity to fix all the parameters in your state vector, or you lack the right kind of data to make one of your state vector elements observable, the result is solution divergence (i.e., the singular matrix problem). Divergence, and other bizarre problems, can also result if the computer you're using lacks sufficient precision to perform the required calculations accurately. This situation can arise, even for large ground-based computers, if there's an extremely large dynamic range between state vector elements. So you want to be judicious in selecting state vector elements, and not just load it up with everything but the kitchen sink.

With these "don'ts" in mind, here are a few rules of thumb for including something in the state vector. First, it (obviously) should represent a real, physical quantity. Second, since the state is supposed to be observable, changes in state should cause a change in at least some of the set of measurements. Lastly, the value of a state component should be either constant over a given time interval, or capable of being propagated based on data or a model. If this is not possible, data taken at different times cannot be mixed together successfully when solving for the state vector element.

From a practical standpoint, the importance of a candidate state vector element can vary according to the accuracy requirements under which you're operating. For example, a parameter may have too small an effect to notice its behavior vis-à-vis your accuracy domain. Or, the variations in the parameter values may have little effect on your solution at a given accuracy domain, or over a short enough time interval. The influence of the parameter, of course, may become more important if higher accuracy is demanded or longer time periods are examined. Even in this case, the parameter will not necessarily be observable, i.e., capable of having its value estimated over that time interval. To be observable, you must also be able to distinguish the influence of the parameter from that of other elements of the state vector. If, however, several elements are highly correlated, you will only be able to observe a subset of them, or a constrained set of allowed combinations of those parameters.

8.2.3 Observation Model

In the discussion in Chap. 7, we used to the term "measurement" to refer to a measured quantity (e.g., a star direction output from a star tracker), as opposed to the corresponding reference quantity (e.g., the direction vector for the same star based on catalog coordinates). Some folks prefer to define some more precise terminology, using "measurement" to refer to a raw reading from a device, and "observation" to refer to said reading transformed into a quantity suitable for insertion into the least squares algorithm. For example, a measurement for a CCD star camera could consist of the photon counts as a function of focal array position in the tracker, while the corresponding observation could be either the x- and y-position of the flux weighted mean, or perhaps the unit direction vector constructed from it in the camera reference frame. The transformation of the measurement into an observation will typically require some model for the sensor. In the case of the star camera example this could involve the bias of each CCD pixel, distortions in the camera optics, local environmental conditions such as temperature or magnetic field, etc. This transformation can be called a sensor model: $\mathbf{S}(\mathbf{z}; \{p_s\})$, where $\mathbf{z}$ represents the raw measurement, and $\{p_s\}$ represents whatever "fixed" parameters are included for defining the model. There will be uncertainties or simplifications in the model, all of which contribute to the observation covariance. Of course, one person's measurement is another person's observation. For the star camera example, a person interested in calibrating a distortion function for the camera will typically be interested in defining observations as fairly close to the raw measurements, whereas a person interested in attitude estimation could well consider application of a distortion correction as part of the observation model.

An observation is derived or computed from a measurement, and what constitutes an observation and how it is defined is at the discretion of an analyst. Of course, if the processing is embedded directly in the sensor, the sensor could also output the observation directly, as would be the case for a quaternion star tracker. So both measurements and observations are closely coupled to the actual output from a sensor.

A sensor model transforms measurements into observations using your understanding of the hardware, the environment, and possibly your current estimate for the state.

8.2.4 Least Squares Filters

With the exception of plain-vanilla Triad, all of the algorithms discussed in Chap. 7 (as well as all that are used for the calibrations mentioned in Sect. 8.1) are variations of least squares filters. The "filter" aspect of the algorithms is that they separate and remove high-frequency noise from the low-frequency physical information.

Most of the least-squares problems that you encounter are non-linear, and therefore solved iteratively. For iterative problems, knowing when to turn off the process when performing an individual update is a key consideration. The technical term for this is *filter convergence*, and the usual criterion involves monitoring when the differences between state vector elements from one iteration to the next drop below preset thresholds. A simpler criterion is to watch changes in the loss function and stop the process when its step-to-step changes become sufficiently small (this being a single number, rather than a number for each state vector element). To prevent endless loops, you also usually specify a maximum number of iterations, after which the calculation ceases. Another reason to stop the algorithm is if any of the diagonal elements of the covariance matrix become negative. Since the diagonal elements are related to the squares of the uncertainties associated with the state vector elements, a negative diagonal element is physically impossible and therefore indicative of a problem in the processing significant enough that you should stop the algorithm and not take its output seriously.

In Chap. 7 we discussed both batch and recursive methods for state estimation. Let's now consider some practical distinctions between them. Since batch filters usually are processing a large mass of input data, they tend not to be sensitive to a few bad points. Furthermore, removing bad points entirely is usually fairly straightforward because you have all of the residuals together as a single statistical batch, making it easy to flag and delete any obvious outliers. A standard rule of thumb for this purpose is that an outlier is any data point having a residual that exceeds 3 times the weighted root mean square residual (i.e., the standard deviation). This form of data editing has acquired the name "sigma editing" since the traditional symbol for the standard deviation is the Greek letter sigma (σ).

Clearly, since we use recursive filters rather than batch ones onboard, you might expect batch filters to have some disadvantages as well, and they do. First, depending on the size of the state vector, the number of observations, the complexity of the sensor or observation models, and the number of iterations needed to achieve convergence, batch filters can demand a lot of computing power and processing time. These resource requirements arise from the fact that batch filter algorithms have to invert an $N_X \times N_X$ matrix, where N_X is the size of the state vector, or perhaps even an $(N_X + K) \times (N_X + K)$ matrix if K constraints are being applied. The complexity of an $N_X \times N_X$ matrix inversion is comparable to performing N_X-cubed (N_X^3)

mathematical operations. If you have a large state vector to solve for, the onboard computer may not be fast enough or powerful enough to support the application in realtime. In contrast, as discussed in Sect. 7.11, use of the Kalman approach for recursive processing reduces the size of the matrix to be inverted to $N_Z \times N_Z$, where N_Z is the size of the observation vector.

The second contrast between batch and recursive filters pertains to memory requirements. A batch filter takes in all of the data up front, processes it all simultaneously (albeit possibly iteratively), and then is done. This requires sufficient memory to hold the whole batch of input measured and modeled observations. In contrast, a recursive filter works continuously and indefinitely, settling into an equilibrium wherein degradation of past knowledge is balanced by new knowledge from new observations, and retaining only the information content of the state vector and its covariance. Because the most important reason for running a filter onboard is to determine error signals for input to a control law, it's clear that a Kalman filter is ideal for onboard use, which is hardly an accident since that's the type of application for which Kalman designed it.

As is probably clear from Sect. 7.11 and the material leading up to it, Kalman filters can be complicated computationally. The logical reaction (other than "how can I push this work off on someone else?") is to ask the question "What can go wrong?". First, you can have a divergence problem. In nominal operations, a converged, steady-state situation will occur when the state vector corrections become consistent from update to update, and the error covariance matrix is stable. Divergence occurs when the state vector moves away from the true state. Possible causes of divergence include linearization errors (i.e., there are significant non-linear phenomena that are not amenable to being approximated as linear), modeling errors (not understanding fully what you've launched your spacecraft into), unknown noise statistics (ditto), and cumulative round-off and truncation errors (the "butterfly effect"). All but the last item are routinely experienced on-orbit, so you often spend a fair amount of time after launch trying to figure out what's going on and how to adjust things so that you get the desired level of performance.

On the other hand, a Kalman filter possesses many excellent qualities needed for onboard ACS applications. First, you don't need to re-cycle through previously observed data. Next, you can estimate the current state vector in realtime. Because of its weighting scheme, you can easily utilize data from any or all sensors generating useful attitude data. And finally, its signature quality, a Kalman filter is sensitive to state changes reflected in the most recent data. These characteristics have led to the use of a Kalman filter (cf. Sect. 7.11) as the standard approach for onboard attitude and gyro bias estimation for most modern spacecraft.

Chapter 9
Onboard Orbit Computations

In addition to reviving the classic "buddy picture" genre, as well as making Paul Newman and Robert Redford two of the top box office draws in Hollywood, *Butch Cassidy and the Sundance Kid* is particularly memorable for a long, drawn-out chase scene where ultimately nothing substantive happens. Butch and Sundance have just robbed another (their last) train, and as they're trying to pick up all the loose cash flying away as a result of the dynamite overkill applied to the train's safe (a less than subtle commentary on the dangers of exceeding one's requirements), a second train ominously pulls up behind their target and disgorges a whole posse of heavily armed bounty hunters. After a couple of B-class actors are shot, the Hole-in-the-Wall Gang takes off with the bounty hunters riding after them. Trying to evade their pursuers, Newman tries one ploy after another without much success, including leading the bounty hunters through some godforsaken wilderness in the dark. When they stop for a rest, Sundance asks Butch whether he thinks they've lost the posse, to which Butch replies, "I don't know where we've been, and I've just been there."

That line of dialogue summed up very nicely the state of onboard orbit determination until the recent advent of the Global Positioning System (GPS) and the Tracking Data and Relay Satellite System (TDRSS) Onboard Navigation System (TONS), which enable a spacecraft to measure its own position and velocity in real-time. In the past, only the ground could measure where the spacecraft was, and that information was not really current because its orbit knowledge was the end product of the processing of tracking data (range and rate data) distributed over at least several hours worth of past contact opportunities. In other words, by the time the ground orbit determination software delivered its product, the results were at least many hours old. For Low Earth Orbit (LEO) spacecraft, whose orbital periods are around 95–100 min in duration, that's like saying "here's where you were a few orbits ago." So all the ground knew (directly from tracking data) was where the spacecraft had been, not where it actually was at the current time, and the spacecraft didn't even know that much. However, by utilizing this past ephemeris history as input, as well as employing physical models for the Earth's gravitational field, the Earth's atmospheric

© Springer-Verlag London Ltd. 2017

H.L. Hallock et al., *ACS Without an Attitude*, NASA Monographs
in Systems and Software Engineering, DOI 10.1007/978-1-4471-7325-0_9

density, solar radiation pressure, and the gravitational contributions from other celestial objects, ground software could integrate the spacecraft's equations of motion to predict where the spacecraft would be in the future. The accuracy of these predictions clearly could not be as reliable as direct measurements, and they certainly would be expected to degrade with time, but the orbit knowledge requirements for most GSFC missions typically are not as severe as the attitude knowledge ones, which almost universally demand real-time attitude solutions.

So ground-based ephemeris processing and propagation enables the ground system to predict the spacecraft's orbital position, an essential capability for look-ahead scheduling systems, but doesn't address the spacecraft's need to know where it currently is in order to fulfill its real-time pointing responsibilities. One solution to that problem would be to migrate the ground's physical models and integration algorithms to the spacecraft and let the flight processor do the same calculations onboard. This approach has in fact been utilized on many recent missions, enabled by the improved performance of modern flight computers. But in the past, real-time calculations of this complexity could not be performed onboard due to processor limitations, from the standpoint both of computing power and of onboard storage. For these earlier missions, the spacecraft flew simple, efficient orbit models driven by a small number of ground-computed input parameters that were best fits to the ground's predicted ephemeris. The uplink parameters were valid, i.e., generated spacecraft position and velocity vectors that met accuracy requirements, for a mission-specific time duration that was dependent on the model and orbit geometry. For example, for a LEO orbit, perturbations from atmospheric drag would cause the errors in the propagated vectors to grow quickly with time. However, since the uplink parameters can (in principle) be updated as frequently as you want, an onboard model that is much simpler than the ground-based models can (at least for a limited time) deliver pretty accurate orbit predictions. In fact, with frequent uplink parameter refreshing, the dominant error source in the onboard predictions will be the original error resident in the ground-based predictions from which the uplink parameters were generated.

Onboard orbit modeling is not restricted just to predicting the spacecraft's position and velocity, which is needed to support such diverse applications as velocity aberration compensation, geomagnetic field modeling, communications, and nadir-pointing. In addition, for many spacecraft utilizing TDRSS, the predicted positions of TDRS spacecraft must be computed onboard in order to calculate the line-of-sight vector for spacecraft antenna pointing. Onboard ephemerides for celestial objects such as the Sun and Moon usually also are required. Analytic models for solar and lunar positions are used to support bright object constraint checking, while velocities derived from solar position vectors are frequently used (less frequently in the case of the Moon) to support velocity aberration compensations. And for spacecraft in orbits beyond Geosynchronous Earth Orbit (GEO), solar and lunar propagators also supply the input needed to factor in N-body gravitational perturbations when the equations of motion are integrated onboard to yield spacecraft ephemerides.

With respect to onboard modeling of spacecraft ephemerides, there has been a fairly rich population of onboard propagators utilized over the decades. As a representative sampling, we've selected five models to examine in some detail,

namely those used on the Compton Gamma Ray Observatory (CGRO), the Hubble Space Telescope (HST), Landsat, the Rossi X-ray Timing Experiment (RXTE), and the Wilkinson Microwave Anisotropy Probe (WMAP).

9.1 CGRO Onboard Orbit Models

CGRO utilized TDRSS, so it had to model the positions of CGRO (accurate to 6 km, 3σ) as well as two TDRS spacecraft (accurate to 30 km, 3σ). For onboard velocity aberration compensations, CGRO also needed to compute the Sun's velocity (relative to the Earth), for which it used a standard analytic model. CGRO's orbit geometry was a fairly typical LEO (near-circular orbit with 28.5° inclination), although its 400-km altitude was at the low end of standard LEO range. TDRS orbits are geostationary, so their orbital altitudes are synchronous (orbital radius of 42,164 km), and their orbit eccentricities and inclinations are near zero (well, if you consider an inclination of up to ∼14° to be "near zero").

Because the orbit types were so different (LEO vs. GEO), the uplink parameters for the two applications were quite different. The primary input to the CGRO orbit model was one day of ground-computed predicted position and velocity vectors spaced 20 min apart. To compute a predicted vector pair for some general time within the one-day file duration, the onboard algorithm interpolated between the file vectors using Hermite polynomials. By contrast, reflecting the less volatile nature of geosynchronous orbits (as compared to LEO orbits), the primary input to the TDRS orbit model was sets of Fourier power series coefficients for each component of position and velocity for each TDRS. For a given TDRS, there were five sets of Fourier coefficient triplets from which five position vectors were computed onboard, the analogs of the CGRO position vectors spaced 20 min apart. Similarly, for a given TDRS, there were three sets of Fourier coefficient triplets from which three velocity vectors were computed, the analogs of the uplinked CGRO velocity vectors. As with the uplinked CGRO vector pairs, the TDRS vectors computed from the uplinked Fourier coefficients were transformed onboard to Hermite polynomial[1] coefficients, which could then be used to interpolate predicted TDRS vectors corresponding to an input time. Note that there was no direct modeling of the atmospheric drag or Earth oblateness perturbations influencing CGRO's orbit, or of the solar radiation pressure perturbations influencing the TDRS orbits. Instead, these physical effects were implicitly packaged into the uplink parameters when the ground software performed its least-squares fit relative to its predicted ephemerides, which explicitly modeled the phenomena.

[1] For more on Hermite polynomials, see e.g., G. Arfkin, H. Weber and F. Harris, Mathematical Methods for Physicists, Academic Press, Inc., seventh edition, Sect. 18.1.

9.2 HST Onboard Orbit Models

Just as CGRO tried to reduce the size of its flight code by utilizing the same interpolator for both the CGRO and TDRS orbit models, HST tried to minimize the size of its flight code by utilizing the same onboard model for HST, TDRS, solar, and lunar orbit models. Later in HST's mission, following the replacement of its original DF-224 onboard computer (OBC) with a 486 computer, space issues became less important and better, more accurate models replaced the original TDRS, solar, and lunar models. The original algorithm was retained to model the HST orbit as it was tailored to LEO orbits like HST (590 km nominal altitude, near-circular orbit with a 28.5° inclination).

The input to HST's spacecraft orbit model is an extended set of Keplerian elements, 14 in number, along with the epoch time associated with the parameters. In addition to the six parameters informationally equivalent to the standard Keplerian elements, three of the inputs are first derivatives of the elements, namely the rate of change of the right ascension of the ascending node, the first derivative of the argument of perigee, and the orbital rate (i.e., the rate of change of the mean anomaly). These terms, when computed as best-fit parameters by the ground system, effectively model perturbations due to Earth oblateness. A fourth input, the second derivative of the mean anomaly, provided a means to model (roughly) the influence of atmospheric drag. To avoid the FSW having to perform unnecessary repetitive calculations, the ground also uplinked the eccentricity in several functional forms. By refreshing the onboard parameters every other day, the onboard ephemeris accuracy could be maintained to within a couple of kilometers of the "truth" vectors used in the ground-based fit, easily satisfying the mission requirement of 10 km.

Of course, the performance of the model when applied to the Sun and Moon, applications for which the model was not designed, was much poorer. So when a new computer was installed in HST during one of the Servicing Missions, the solar and lunar propagators were upgraded to standard analytical models. The TDRS orbit model also was replaced since by the time of the Servicing Mission, Flight Dynamics was no longer providing to users actual TDRS ephemeris files based on tracking data, but instead files characterizing the center of the orbital "box" in which the TDRS spacecraft was maintained. The Flight Dynamics change permitted a major simplification to the onboard model because although the location of the center of the orbital box varies with time when viewed in the GCI reference frame (i.e., it follows a geostationary orbital path), as viewed from the Earth the center of the box is fixed azimuthally relative to the Earth. As a result, you can define the orbit with fewer parameters relative to the Earth-rotating reference frame than are required relative to GCI. Furthermore, because the center of the box is an ideal point towards which you're commanding (as opposed to the actual physical location of the TDRS spacecraft), the values of the parameter set to first order are constants not requiring regular updating.

In fact, only five parameters are needed to specify the motion of the center of the box in GCI (which we'll call TDRS from now on for simplicity), namely the TDRS

inclination (ideally zero), the TDRS longitude (assigned by the Network Control Center (NCC)), the distance from the Earth center to TDRS (nominally 42,164 km), the TDRS rotation rate in the GCI frame (nominally the Earth's rotation rate), and the Greenwich Hour Angle at the epoch time (i.e., the right ascension of the Greenwich Meridian at the epoch time). The TDRS inclination actually does require calculation because the TDRSS inclinations are allowed to drift up to 14° to conserve fuel and because of the time dependent nature of the GCI reference frame arising from the precession and wobble behavior of the Earth's spin axis relative to the "fixed stars". Operationally, exploiting the Flight Dynamics Facility's (FDF's) inclined-center-of-box (ICOB) modeling, the advantage of using this so-called "TDRS-on-a-stick" model is that the number of uplink parameters (including the epoch time) is reduced from 15 to 6 (the 5 specified above plus the epoch time), and the frequency for parameter updates is reduced from once a week to once a year, or less.

9.3 Landsat Orbit Model

Unlike the previous two spacecraft (CGRO and HST), which occupied 28.5° inclination LEO orbits, the Landsat series flew in near-polar Sun synchronous orbits. For a typical Landsat mission, the orbit was near-circular, 900–950 km in altitude, and 99.0° inclination. The Landsat orbit model was similar to CGRO's TDRS orbit model, with enhancements. Actually, since the Landsat series pre-dated CGRO, it probably would be more accurate historically to say that CGRO's TDRS model was a simplified version of the Landsat orbit model. In any event, the uplink parameters for Landsat were six sets of Fourier coefficients along with position and velocity residuals at a set of grid points spaced 8.7 min apart. The Fourier coefficients "drew" the basic shape of the orbit, with corrections for orbital perturbations supplied by the position and velocity residual "fudge" terms. When the spacecraft wanted to know Landsat's position and velocity vectors at a given time point, a Hermite polynomial interpolator (sound familiar by now?) was used. Thanks to the modeling capabilities of the ground system performing the fit to the Flight Dynamics ephemerides, the uplink parameters implicitly modeled three-body effects, Earth oblateness, and atmospheric drag, enabling the onboard model to determine Landsat's orbital position to an accuracy of better than 1 km.

9.4 RXTE Orbit Models

The previous onboard orbit models we've discussed have been "dumb" ones, all the "smarts" being contained in the ground system models used to calculate fit parameters for uplink to the onboard propagators. By contrast, RXTE's spacecraft models (for the RXTE and TDRSS spacecraft), which were adapted from those used on the

SAMPEX mission, utilized physical models resident onboard to integrate the spacecrafts' equations of motion to determine the position and velocity. RXTE's orbit is similar to HST's with a lower inclination. It's a near-circular orbit with a nominal altitude of 580 km and an inclination of 23°. Since it's flying in a LEO orbit, RXTE explicitly models Earth oblateness and atmospheric drag. Specifically, to incorporate Earth oblateness effects, RXTE carries in its flight database the values of the 2nd, 3rd, and 4th zonal harmonic coefficients, as well as the first sectoral harmonic coefficient (see Eq. 6.1 for full spherical harmonic expansion of the geopotential).

To compensate for atmospheric drag, a Jacchia–Roberts atmospheric density model is used, which requires database input of an atmospheric density constant and a drag coefficient. Drag corrections are only performed when calculating the RXTE ephemeris; drag was turned off for TDRS orbit modeling as those effects are insignificant at geosynchronous altitude. Other database input parameters are the Earth's gravitational constant, the Earth's radius, and the Earth's rotation rate. The parameters that "seed" the algorithm, and need to be uplinked frequently (once a day for the RXTE orbit model and once a week for the TDRS orbit model), are the initial position and velocity pair, the Greenwich Hour Angle at the start time (to connect the Earth rotating frame to the GCI frame when calculating the sectoral corrections), and the associated start time. With these models, the ACS flight software calculates the position, velocity, and acceleration vectors for both RXTE and TDRS, with a position accuracy better than 30 km. Solar and lunar positions are calculated via standard analytic propagators.

9.5 WMAP Orbit Models

Although each of the models described so far have been significantly different from each other, the problem they were designed to solve was identical, namely how to predict LEO spacecraft positions and velocities within (mission-dependent) required accuracies. The differences in the algorithms employed reflected more the limitations of the associated onboard flight data systems than the nature of the output product being calculated. In WMAP's case, the geometric characteristics of the mission itself led to the algorithmic approach utilized on-orbit. Unlike the previous four LEO missions, WMAP was maneuvered (over a three-month time duration after launch) into a halo orbit about the Earth-Sun L2 Lagrange point.[2]

The halo orbit, as viewed within a reference frame with Earth fixed relative to the Sun, has a period of 6 months. Since L2 is located about 1.5 million km from the

[2]The five "Lagrange points" (L1–L5) are named in honor of Joseph-Louis Lagrange, based on his 1772 publication "An Essay on the Three-Body Problem". The essay considered the so-called circular restricted three-body problem in which two massive bodies (e.g., the Sun and a planet) orbit their mutual center of mass, and a small third body moves in the mutual gravitational field of the two large bodies. The Lagrange points are stable or semi-stable positions in the field. Some further information is provided in Sect. 12.1.

Earth, the earlier dominant perturbative effects (Earth oblateness and atmospheric drag) are irrelevant. At L2 the orbit's dominant complexity arises from the combination of gravitational influences contributed by more than one celestial body, namely the Sun, Earth, and Moon. So the WMAP ACS FSW uses solar and lunar analytic models to feed input to its WMAP spacecraft orbit model. Additionally, the spacecraft orbit model is regularly seeded with an extended precision WMAP position and velocity vector pair, and an associated epoch time. The model then sums the gravitational influence of the Earth, Sun, and Moon on the spacecraft to compute the net acceleration acting on WMAP and propagates the ephemeris once per minute.

9.6 Onboard Orbit Measurement with GPS

The GPS constellation is located at about half-geosynchronous altitude (12 h orbital period) and consists of 21 spacecraft (of which 3 are spares) distributed in 6 orbital planes (see Fig. 9.1). GPS enables a spacecraft equipped with a GPS receiver to measure its own orbital position in realtime far more accurately than the ground can calculate it from tracking data. In addition to its current position, a GPS-equipped

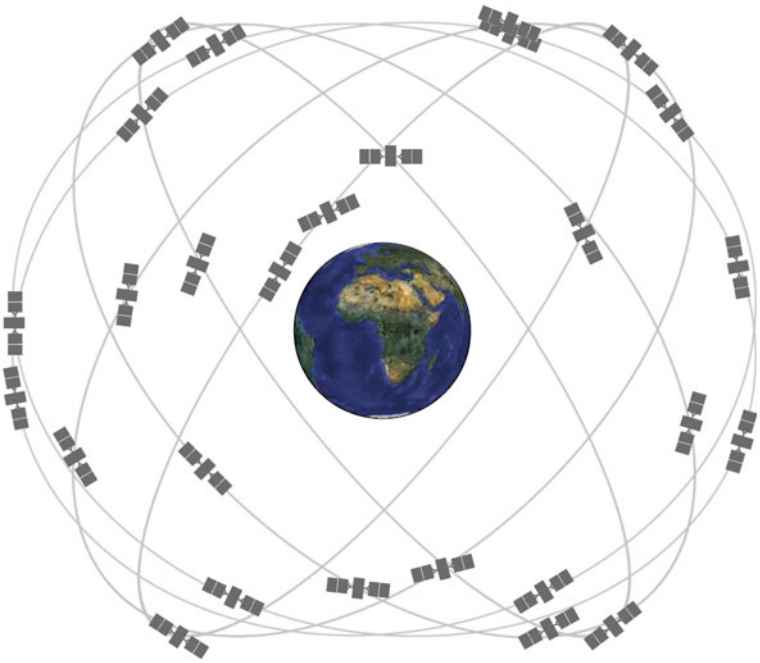


Fig. 9.1 GPS Constellation. Image from: http://www.gps.gov/systems/gps/space/. Credit: NOAA/U.S. Government

spacecraft also gets a time reference for use in maintaining onboard clock correlations. The input to the GPS calculations is pure range data (as opposed to range and rate data), determined from the time difference between the current GPS receiver time and the time attached to the GPS transmission, multiplied by the speed of light. Knowing the time delays relative to three GPS spacecraft, as well as the orbital positions of the GPS spacecraft, you can set up a set of simultaneous equations that triangulate the position of the GPS receiver mounted on the spacecraft. Add in a fourth GPS spacecraft, and the additional time delay information can be used to correct errors in the spacecraft clock. The GSFC-built GPS Navigator is successfully meeting mission requirements for both low-flying and high-flying spacecraft: the Global Precipitation Measurement Core Observatory (GPMCO) and the Magnetospheric Multiscale (MMS) constellation, respectively. The GPMCO ACS FSW, with its 10-Hz cycle, includes a module to propagate the 0.2-Hz Navigator-provided orbit solution between updates, maintaining position accuracy to a few tens of meters. To support the MMS constellation through extended periods without contact with the GPS fleet, the MMS Navigator utilizes an in-built Kalmen filter with detailed dynamics modeling to propagate and combine GPS data.

9.7 Onboard Orbit Measurement with TONS

TONS enables onboard orbit estimation accurate to 150 m/axis, 3σ, by processing one-way forward-link S-Band Doppler measurements. During scheduled TDRSS contacts with the spacecraft, a spacecraft transponder extracts these measurements, which are then run through a nine-state extended Kalman filter (implemented in the spacecraft's ACS FSW) to estimate the spacecraft's position and velocity vectors, an atmospheric drag scale factor, a master oscillator frequency bias, and a TDRS measurement bias. On a given computation cycle, TONS outputs either the spacecraft's orbital elements or the frequency control word. So far, TONS has been used experimentally on GSFC's Extreme Ultraviolet Explorer (EUVE) spacecraft and has been the prime orbit determination scheme for EOS Program spacecraft.

Chapter 10
Control Laws: General Qualities

Your college years were probably your first exposure to the academic equivalent of ethnic humor, i.e., the jokes the various departments tell about their rival branches. For example, if you were a physics major, you heard the jokes physicists tell depicting mathematicians as slaves to logical formalism at the expense of common sense. One old chestnut went: how does a mathematician boil water? He takes a pot from the cupboard, fills it with water from the sink, places the pot on one of the stove's burners, turns the heat on the burner to high, and waits for bubbles to appear. Now suppose you put another water-filled pot on a burner and ask the mathematician what procedure he'll use to boil water given the new set of initial conditions. The stereotypical, purist mathematician would immediately remove the pot from the burner, pour the water into the sink, replace the pot in the cupboard, and inform you that the solution is trivial as it now reduces to the previous case. Not surprisingly, mathematicians (or theoretical physicists) had their own jokes depicting experimental physicists as unenlightened, tool-using technicians who didn't fully appreciate the formal discipline of mathematical proofs. As for their treatment of engineers …well, see Sheldon's treatment of Howard on *The Big Bang Theory*.

Once you get past the cheap shots, there is a grain of truth embedded in the humor, namely that in addition to developing the appropriate toolkit to deal with the kinds of problems you're likely to encounter, your thinking also takes on a mind-set consistent with the kinds of goals you're trying to achieve. A mathematician is concerned with establishing quantifiable relationships between abstract objects of his own devising, so he needs a highly disciplined and formal, logical process for proving or disproving those relationships. A physicist is concerned with establishing quantifiable relationships between physical objects found in nature, so he must first understand conceptually the nature of those objects (which typically evolve over time, unlike the mathematician's eternal abstractions), and then must use mathematics both to characterize the objects quantifiably and to define their interrelationships. Although many view engineers as applications-oriented guys who utilize mathematics, physics, chemistry, etc. as mere tools within the context of building a physical

© Springer-Verlag London Ltd. 2017

H.L. Hallock et al., *ACS Without an Attitude*, NASA Monographs
in Systems and Software Engineering, DOI 10.1007/978-1-4471-7325-0_10

object of his choosing (i.e., to a theoretical mathematician, he's even more of an ungrateful and unappreciative knuckle-dragger than the physicist), in some respects, the engineer can be seen as having a world-view that straddles the other two. Like the physicist, he's driven to gain a complete understanding of a physical object, but like the mathematician, the object he's trying to understand is one (at least to some degree) of his own devising, not something (as in the physicist's case) he's discovered in nature and must both understand and relate to other objects serendipitously encountered that he already thinks he understands to some degree.

So, what does all that mean in terms of our favorite flavor of engineer, namely you, our up-and-coming ACS engineer? First, you have a clear and concrete concept for the physical object (an ACS, including both hardware and software) you're trying to build, one component of which is the control laws that define how pointing and rate errors should be reduced. Second, once you've developed a plausible way to build the thing, you need to develop a quantifiably complete understanding of how your physical object is likely to behave in the Real World, or more accurately in our aerospace context, off the Real World. So until your creation is satisfactorily created, you're constantly toggling between two roles. On the one hand, you're the omnipotent mathematician, creating an object of your choosing, designed to perform in a manner that suits your whim. On the other hand, you're the curious physicist, suddenly presented with an object you don't fully understand, but must master if you're to achieve your goal, and not get downsized in the process. Maybe this repetitive mental toggling explains why it's so natural for an engineer to move directly to iterative algorithms to solve problems rather than search for the elegant, closed-form solutions cherished by both physicists and mathematicians. Or maybe it's just that an engineer's version of the Real World is a lot messier than the equivalent for the other two fields.

In the previous chapters, we've conducted a survey of the more tangible pieces of the control engineer's ACS, namely the sensors and actuators. We've also examined the various models for viewing the spacecraft environment, the algorithms used to interpret the sensors' impressions of that environment and the spacecraft's place in that environment, and how you compare the model output to the sensor output in order to quantify errors in spacecraft's orientation and motion. What still needs to be done is to discuss how the controls engineer must deal with those errors in order to maintain a desired spacecraft attitude profile, and how to change the current profile to a new one. The general engineering discipline that is concerned with that task is control theory.

There are two basic types of control law processes, *open-loop* control and *closed-loop (feedback) control*. Open-loop control uses an actuator for direct control of a process, without sensor feedback. Closed-loop control feeds back processed sensor output into the calculation so as to characterize quantitatively the results from the previous actuator commanding, and uses that information to modify current actuator commanding in order both to accomplish the objective of the current iteration and to clean up errors left over from the previous iteration. From breakfast experience, it appears that a toaster is a good example of open-loop control, although a co-worker claimed that his toast was intelligently manufactured by a device with an optical

sensor that evaluated the darkness of the bread to determine when it was done. If so, how it successfully handles pumpernickel rye is a mystery. A good example of closed-loop control is driving a car, where you use visual feedback (and sometimes audio feedback, delivered at a lower frequency, but higher amplitude by other motorists) to fine-tune actuator (i.e., steering wheel and accelerator) control. More accurately, this was a good example of closed-loop control if you were driving a car back in the 1950s. Sometime after 1990, it seems that a large number of motorists abandoned closed-loop control in favor of open-loop control, at least when traveling in the Greater Washington Metropolitan area.

One thing about control theory that confuses you the first time you dig into it is that there's a new set of technical terms you have to learn that define quantitatively the characteristics of the systems being (or trying to be) controlled. At a high level, they describe how stable the system is, how the system responds if you "kick" it, how accurately and reliably the system achieves the desired final state, etc. So before getting into how a control law works, let alone how you construct one, it's probably a good idea to spend a little time defining some terms.

10.1 Definition of Control Law Terms

Let's say that by some miraculous process, we already have the optimal control law that perfectly meets our needs. Now suppose that some significant element of the system needing to be controlled, called the *plant*, changes. If that happened, the first thing a lazy person (that is, anybody who works for a living who has to meet deadlines) would try to do is tweak some of the database parameters that drive the algorithm to see if the control law still performed acceptably. If that worked, then a natural question to ask is how much you had to change the parameters in order to compensate for the change in the plant. Now, let's invert the question. Suppose, instead, we ask: if I tweak the database parameters a little, does the way the system behaves change by a little, or change by a lot? Control laws whose behavior changes by a lot if you change their database parameters by small amounts are said to have *high sensitivity*. On the other hand, control laws whose behavior doesn't change much if you change their database parameters by large amounts are said to have *low sensitivity*. However, unlike what appears to be true in many human relationships, it is not the case that control laws display a high level of insensitivity if they fail to adjust themselves autonomously to your current needs.

Next, let's go back to our original situation where we started off with an optimal control law, and then made a significant change in the plant. Suppose we make no change in the database parameters and simply ask how well does the original control law perform when used to control the modified plant? Control laws that continue to perform well under those circumstances are said to be *robust*, while those that don't are said (big surprise) not to be robust. Typically, a control law will be robust if it has low sensitivity, is stable over a range of database parameter variations, and continues to satisfy requirements over a range of change in the system parameters.

So sensitivity and robustness are measures of the "twitchyness" and flexibility of a control law as it functions. Another thing that would be nice to characterize is how well the control law gets you where you want to go, or keeps you where you want to stay. Suppose you give your nice stable, functioning control law a little kick. The technical term for the kick is a *disturbance signal*, which simply is any unwanted input signal (for example, a perturbative torque) that affects the system's output signal (what the sensors measure). If your control law is a good one, you'd like to see the influence of that kick on the spacecraft (as measured by the sensors) damp out quickly. Or, put in the arcane language of control theory, the control law's ability to reduce the disturbance signal's effect on the output signal is a measure of the control law's disturbance rejection capability. As the control law deals with the impact the kick has had on the system, it gradually loses interest in and forgets about the original kick entirely. This control law behavior is measured by its *transient response*, which decays with time. For non-ideal control laws, after the transient response has completely decayed, there may remain some lingering trace of the original kick in the output signal that the control law has been unable to remove. This residual error is called the control law's *steady state error*.

10.2 Closed-Loop Control Laws

Having established a new mini-lexicon of control law terms, let's use them to expound on a pretty obvious fact, that if you have enough information and money closed-loop control laws are nicer than open-loop ones, but say it in a way that will leave your friends and associates nodding their heads sagely. As compared to open-loop control, feedback control:

1. decreases the sensitivity of a system to variations in the parameters defining the behavior of system components;
2. makes control of system transient response easier;
3. improves system noise and disturbance rejection; and
4. reduces system steady state error.

That's enough on buzzwords for now. Let's start talking about how a closed-loop control law works by listing the input we should provide it in order to make it do what we want. Obviously, to get it to do what we want we must first tell it what we want done, i.e., it must be provided with a set of commanded values, e.g., a target attitude, a desired body rate, etc. If we not only care about what commanded values we're targeting, but also care what path we take to get there, or are particular about satisfying some other constraints in the process, we may need, for example, to supply a slew profile, which might be best understood by the control law by translating the profile into a set of *feedforward* torques. Feedforward information says what we want the control law to do this control cycle to deal with things we anticipate happening (or desire to happen) during the current control cycle. By contrast, feedback uses information generated during this control cycle (say, from attitude sensors) that describes

what happened in the previous control cycle and that we now want to compensate for in the current control cycle. Feedforward also is a way to deal with disturbance signals in a premeditated fashion. Without feedforward, when a disturbance kicks in, the control law will try to deal with it in a generic fashion, i.e., the disturbance handling will be dictated by issues such as the amplitude and frequency of the disturbance, without any special insight into the phenomenon that's engendering it. By contrast, by utilizing feedforward, you can characterize anticipated disturbances (for example, gravity gradient, magnetic, and atmospheric torques) with accurate models, anticipate and compensate for the disturbance's effects via feedforward, and thereby improve the responsiveness of the control law.

Now that you've given the control law all this input, what does it give you in return? Generically, the output of a control law is an error signal that both tells you how successfully you've achieved your commanded goals and what you should do next in order to make things better. For specific applications, there usually also will be processing that takes the basic error signal and translates it into specific actions you should take with the actuators in order to compensate for the residual error. For our ACS application, that means the control law may output control torques that should be applied with reaction wheels, MTBs, or attitude thrusters. So the control torques ordered by the control law should take into account things like maximum wheel torque capacity (don't ask for more help than the actuator can supply in a given control cycle) and maximum MTB effectiveness (limited by the geomagnetic field direction and magnitude at the time the MTB commands are executed). The control law's "orders" also need to take into account important constraints on spacecraft operation, such as avoiding inducing excessive jitter (handled via torque limiting), avoiding inducing vibrational modes (handled internal to the control law via filters), and avoiding reaction wheel saturation (handled via a distribution law).

Having discussed at a high level how a control law operates when things are going right, let's talk about the things that can cause things to go wrong, or at least cause the control law to work harder (i.e., have lower responsiveness) than it would otherwise. First, you may have omitted modeling effects that could easily have been addressed, such as certain external disturbances, spacecraft dynamics, or electronic time delays. Second, you may have tried to provide feedforward compensation with models, but your models themselves introduced significant errors. A laundry list of such failings would include:

1. unmodeled (high frequency) dynamics;
2. neglected non-linearities (typically, linear approximations are utilized throughout control law computations);
3. intentional (and inappropriate) reduced-order modeling;
4. plant parameter value changes due to environmental or aging factors;
5. improperly treated noise;
6. quantization errors.

10.3 Laplace Transforms and Transfer Functions

Recalling from the laundry list that significant error can be induced by linearizing the control law equations, one might ask the obvious question, why do that? One answer is exactly the same as that which motivated our introduction of the (iterative) linearized solution of the attitude determination problem in Chap. 7, namely, linearizing reduces the complexity of the calculations that must be performed, which is a big deal when you need to do real-time onboard processing in a flight computer. In particular, when you're dealing with a linear system, or a linearized approximation to a non-linear system, you are able to exploit the principle of superposition. In other words, if your input signal can be expressed as the sum of a set of more elementary signals, the response can be decomposed into the sum of the (presumably simpler) individual responses associated with the elements of the input set.

In addition to this standard reason for linearizing any onboard processing when possible, linearizing the control law problem enables you to use a special technique called the Laplace transform when performing control law analysis of a continuous system. What a Laplace transform does is to take you from the time domain to a frequency-like domain, although in a somewhat different manner than application of a Fourier transform. The Laplace transform $F(s)$ of a function $f(t)$ is given by Eq. 10.1:

$$F(s) = \int_0^\infty f(t) \exp(-st) dt \tag{10.1}$$

Note that s has units of inverse time. It is complex, with an imaginary part representing the frequency of system oscillations, and a real part representing the time scale for growth or decay of disturbances. We'll sometimes refer to s as simply a frequency variable, but remember the double encoding it can contain. Readers familiar with the Fourier transform may note that it is very similar in form to the Laplace transform. Indeed, if s is selected to be pure imaginary ($s = i\omega$), and if $f(t)$ is non-zero only for $t \geq 0$, then the Laplace transform of $f(t)$ is in fact the same as the Fourier transform, and therefore captures the same frequency decomposition information.

A particularly important property of the Laplace transform is that it causes derivatives in the time domain to be replaced with multiplication by s in the frequency domain. That last sentence may sound a bit intimidating, so we'll make it more so by writing it out as Eqs. 10.2 and 10.3; given

$$g(t) = \frac{df(t)}{dt} \quad \text{and} \quad F(s) = \mathcal{L}\{f(t)\} \tag{10.2}$$

then the Laplace transform of $g(t)$ is

$$G(s) = \mathcal{L}\{g(t)\} = sF(s) - f(0^+) \tag{10.3}$$

where

$\mathcal{L}\{\}$ indicates the Laplace transform operation, and
$f(0^+)$ is the limit of the function $f(t)$ as t approaches 0 from the positive side

For those of you about to bolt in fear and loathing, Eq. 10.3 simply represents how a time derivative (df/dt) gets converted into multiplication in frequency space $(sF(s))$. That trailing bit, i.e., "$-f(0^+)$", turns out to be useful when considering disturbances that don't exist before a certain time, e.g., the sudden activation of satellite thrusters, or perhaps giving the system a thwack with a hammer. Similar to the Laplace transform of a derivative, we have the following formula for the Laplace transform of an integral; given

$$h(t) = \int_a^t f(\tau)d\tau, \quad \text{and} \quad F(s) = \mathcal{L}\{f(t)\} \tag{10.4}$$

then the Laplace transform of $h(t)$ is given by

$$H(s) = \mathcal{L}\{h(t)\} = 1/s \left(\mathcal{L}\{f(t)\} - \int_0^a f(\tau)d\tau\right) \tag{10.5}$$

Note the inverse symmetry here; d/dt maps to multiplication by s in frequency space, while $\int dt$ maps to dividing by s in frequency space. Given that integration is an inverse operation to taking a derivative, it is not terribly surprising that these operations map into inverse relationships in frequency space, although that the transformed relationship would be simple multiplication and division may not have been obvious.

The Laplace transform has been much studied by mathematicians and engineers, with the result that tables have been generated containing $(f(t), F(s))$ pairs for a great many useful functions. A short table of Laplace transforms is given in Table 10.1.[1]

So, how does this help? Well, you can apply the Laplace transform not just to a function, but to both sides of an equation. The implication is that you can use the Laplace transform to convert a differential equation describing the behavior of a spacecraft into a polynomial equation in frequency, often a much easier object to work with mathematically than the original differential equation. Other functions and operations have simple Laplace transform analogs. For example, trig and exponential functions get converted to ratios of polynomials in the variable s (see Table 10.1). So working in the frequency domain (as opposed to the time domain) via the Laplace transform allows you to substitute relatively simple algebraic expressions for more complicated differential equations.

So now we've got the Laplace transform available as a tool that can help us figure out what's going on. The not-as-stupid-as-you-might-think question is, what do we want to know? If you're a theoretical physicist, your answer is that you

[1] For more extensive tables, see, e.g., C. Wylie, Advanced Engineering Mathematics, Chap. 7, © 1960 by McGraw-Hill, or just do an Internet search for "Laplace transform".

Table 10.1 Short table of laplace transforms

$f(t)$	$F(s) = \mathcal{L}\{f(t)\}$
$u(t) = \begin{cases} 0 \ (t < 0) \\ 1 \ (t \geq 0) \end{cases}$	$\frac{1}{s}$
t	$\frac{1}{s^2}$
t^n	$\frac{n!}{s^{n+1}}$
$\delta(t)$	1
$\sin \omega t$	$\frac{\omega}{s^2+\omega^2}$
$\cos \omega t$	$\frac{s}{s^2+\omega^2}$
e^{at}	$\frac{1}{s-a}$
te^{at}	$\frac{1}{(s-a)^2}$
$t^n e^{at}$	$\frac{n!}{(s-a)^{n+1}}$
$f'(t)$	$sF(s) - f(0^+)$
$f''(t)$	$s^2F(s) - sf(0^+) - f'(0^+)$
$af(t)$	$aF(s)$
$f(t) + g(t)$	$F(s) + G(s)$
$\int f(t)dt$	$(1/s)(F(s) - \int f(t)dt)$

want to define a simple equation (or set of equations), derived from a few basic principles, that quantitatively describes the behavior of a physical system. Einstein's General Theory of Relativity (describing the motion of objects in gravitational fields) is a perfect example of that Grail-like quest. Only after having spent many years deriving a self-consistent, and aesthetically pleasing, set of tensor equations did Einstein ask the question, what physical predictions can I make? By contrast, the practical engineer often takes an almost inverted point of view. Treating the physical phenomena influencing his "world" like a black box, he asks the question, if I send in an input signal that looks like this, what does my output signal look like? In extreme terms, as long as he can predict (or even better, control) what's going to happen, who cares why it does what it does? The selection of a fair number of the tools and methodologies utilized by a controls engineer has that (somewhat caricatured) worldview at its roots.

If that's your outlook on life, a reasonable object to analyze to death is the ratio of the output signal and the input signal. Considering a few trivial cases, if the ratio is always unity, your black box is (in generic terms) a perfect conductor; the signal passes through the box totally unaffected. If the ratio is always zero, your black box is a perfect absorber; the signal is totally blocked. If the ratio is infinite, stop reading immediately and call the patent office; you've either found a cure for conservation of energy or invented perpetual motion. Now let's get a bit more mathematical and add the Laplace transform to the mix, if for no other reason than to justify our earlier discussion. If the ratio of the output signal to the input signal seems of interest, surely

the ratio of the Laplace transform of the output signal and the Laplace transform of the input signal (with all initial conditions assumed to be zero) would be utterly fascinating, so much so that we're hardly surprised that controls engineers give it a special name: the *transfer function* of the control law's plant.

Since it utilizes the Laplace transform, the transfer function analysis approach is only applicable to linear, constant parameter systems. And since it only examines the system's input and output, it can provide no insight to the system's internal structure. Still, it provides a very useful mechanism for getting a feel for the system by varying the input and studying the resulting output. Because of the principle of superposition, the input and output signals usually (for our applications) can be decomposed into combinations of fairly simple mathematical functions. As a result, the transfer function often will have the form of a polynomial in s (frequency) divided by another polynomial in s. (If each of the individual input terms transforms into a ratio of polynomials in s, similar to the right side of Table 10.1, then one can construct from the sum of such ratios of polynomials a single ratio whose denominator is the product of the denominators of the individual ratios.) Values of s for which the numerator goes to zero are called (not surprisingly) the transfer function *zeroes*. Values of s for which the denominator goes to zero are called the transfer function *poles*, a term originating from complex variables analysis (recall that the roots of a polynomial equation generally will have real and imaginary parts, as shown in Fig. 10.1's illustration of the complex plane), which is a major tool of electrical engineering.

The equation resulting from setting the denominator equal to zero is also called the *characteristic equation* because it defines the character of the system's time response, as we'll discuss in the next two paragraphs. A plot of the zeroes and poles of a simple transfer function is provided in Fig. 10.2.

One of the nice things about working with a function that's a ratio of two polynomials is that you often can perform a partial fraction expansion (by factoring the numerator and denominator), which will yield a sum of relatively simple individual terms that can be inverse Laplace transformed separately to yield the transfer function in the time domain. (This presumes that you have one of those tables of $(f(t), F(s))$

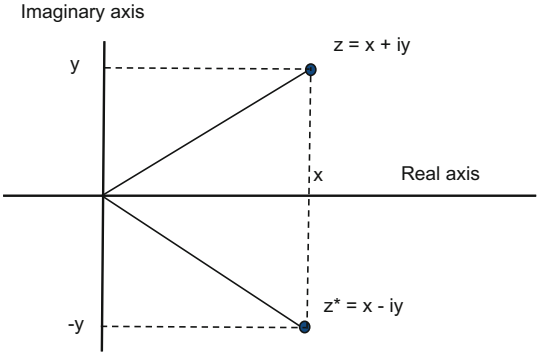


Fig. 10.1 The complex plane

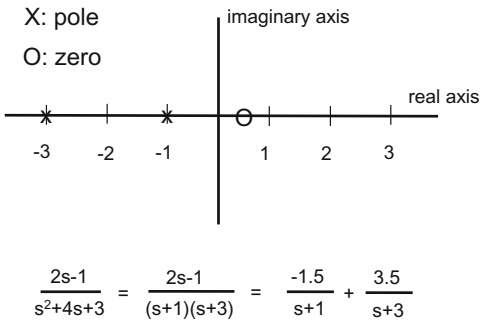


Fig. 10.2 Zeroes and poles of a transfer function

pairs available, such as Table 10.1.) Suppose after doing a partial fraction expansion, all the individual terms are of the form $c_1/(s + c_2)$, where both c_1 and c_2 are complex numbers having both real and imaginary parts. What does having the transfer function given in this form mean for the dynamic system under analysis? We turn now to the book of shadows - er, the book of $(f(t), F(s))$ pairs - to gain some insight. The inverse Laplace transform of $c_1/(s + c_2)$ is $c_1 \exp(-c_2 t)$. This goes to zero as time goes to infinity if the real part of c_2 is positive, which corresponds to suppression of disturbances in the dynamic system. If the real part of c_2 is negative, you'd have uncontrolled growth of disturbances. The imaginary part of c_2 corresponds to oscillatory behavior of the dynamic system (recall that the exponential of a pure imaginary number times time can be re-expressed as a combination of sine and cosine terms).

Glancing back at Fig. 10.2, we see that if the real part of c_2 is positive, the pole (equal to $-c_2$) will be plotted on the left side of the complex plane. Using the principle of superposition to view the entire transfer function as the sum of individual components, it's clear that if all the transfer function poles are located on the left side of the s-space complex plane, all oscillatory terms will damp-out with time (as with noise-suppressing headphones), and the associated control law will be stable. However, if any of the poles are on the right side of the complex plane, the amplitude of at least one oscillatory term will grow quickly with time (like audio feedback occurring when a microphone is too close to the output speakers), and the associated control law will go nuts, as will the controls analyst who has designed it.

There's a bit more general information regarding transfer functions worth noting. Recall that d/dt (the derivative operator) maps to multiplication by s in frequency space, while $\int dt$ (the integration operator) maps to dividing by s. If we take $s = i\omega$, where ω is the frequency of the input signal, the magnitude of a transfer function component associated with d/dt will be an increasing function of frequency, i.e., a derivative component in a control law will have a relatively greater effect at higher frequencies (it is a "high-pass filter"). In contrast, because of the division by s associated with $\int dt$, an integral component in a control law will have relatively greater effect at lower frequencies (it is a "low-pass filter"). The two operations also have opposing phase effects on the output. Let's suppose that the input is a pure sine wave: $\sin(\omega t)$, for which the derivative and anti-derivative are $\cos(\omega t) = \sin(\omega t + 90°)$,

and $-\cos(\omega t) = sin(\omega t - 90°)$, respectively - i.e., the two operations produce an output with a leading and lagging phase, respectively. In frequency space this maps as

$$(d/dt, \int dt) \rightarrow (s, 1/s) = (i\omega, 1/(i\omega)) = (i\omega, -i/\omega) \tag{10.6}$$

Leading and lagging phase corresponds to positive and negative imaginary component for the transfer function.

10.4 Control System Response and Behavior

One thing that both physicists and engineers agree on is that an ideal application on which to try out your ideas is a perfect harmonic oscillator, often pictured as a mass on a spring, as illustrated in Fig. 10.3, wherein M represents the value of the mass (kg), K the spring constant (kg/s^2), f a coefficient of friction (kg/s), and $r(t)$ is an external driving force imposed on the oscillator.

With damping and a driving term included, you have all the basic elements needed to study a wide array of problems in classical dynamics, electromagnetism, etc. As a way to get a little hands-on familiarity with transfer functions and how to interpret them, we've included a worked-out example in Sect. 10.5. For folks who are interested in some of the nitty-gritty math, walking through the problem on your own will be a useful exercise. For those suffering nervous afflictions when exposed to excessive equations, you won't lose much by skipping the example as most of what you might care about, like good poles versus bad poles, has already been discussed qualitatively in the previous paragraphs.

There are some features that are worth understanding qualitatively, even for those of you choosing to avoid the math. The oscillator has a "natural frequency" for its motion, given by:

$$\omega_0 = (K/M)^{1/2} \tag{10.7}$$

If pushed away from equilibrium, the system will tend to vibrate at this frequency. This tendency will be modified by the effects of friction, which damps out the natural

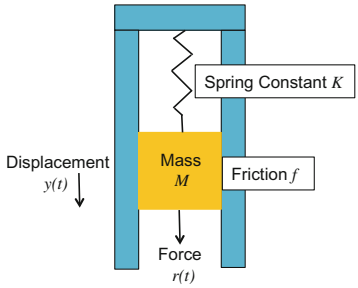


Fig. 10.3 A harmonic oscillator

oscillations. The dimensionless "damping ratio" is defined as:

$$\xi = (f/2)/(KM)^{1/2} \tag{10.8}$$

If $\xi = 0$ (i.e., no friction), the behavior is purely oscillatory. If $0 < \xi < 1$, the signal is *underdamped* and the transient response will oscillate around the equilibrium point with decreasing amplitude over time, reaching equilibrium as time goes to infinity. If $\xi = 1$, the signal is *critically damped*, doesn't oscillate, and asymptotically approaches the equilibrium point as quickly as possible without overshooting. And finally, if $\xi > 1$, the signal is overdamped and decays without ever overshooting its equilibrium (i.e., no oscillations at all), but rather asymptotically approaches its equilibrium position.

When evaluating a system's response, there's a lot more to consider than just how quickly it damps out and what its final value is. That means we need to define some more buzzwords to describe the nature of this other stuff we want to talk about. If we're dealing with oscillatory behavior, a useful thing to know would be how long it takes for it to reach (and overshoot) its eventual equilibrium position, also called the steady-state solution. *Rise time* is the technical term for this time duration (see Fig. 10.4 for an illustrative chart given a step function input); it is inversely proportional to the natural frequency. This is to be expected since the natural frequency tells you how fast the mass moves as it oscillates. The faster the oscillation speed, the less time it should take to overshoot equilibrium, so the rise time decreases. Also, as the damping ratio increases, the rise time will increase until you reach critical damping, at which point the rise time goes infinite. Again for oscillatory situations, you'd want to know how long it takes for the system to hit its maximum amplitude on overshoot. That's called the *peak time*, and as was the case for the rise time, it also is inversely proportional to the natural frequency, and for the same reason. And similar to the rise time, as the damping ratio increases, the peak

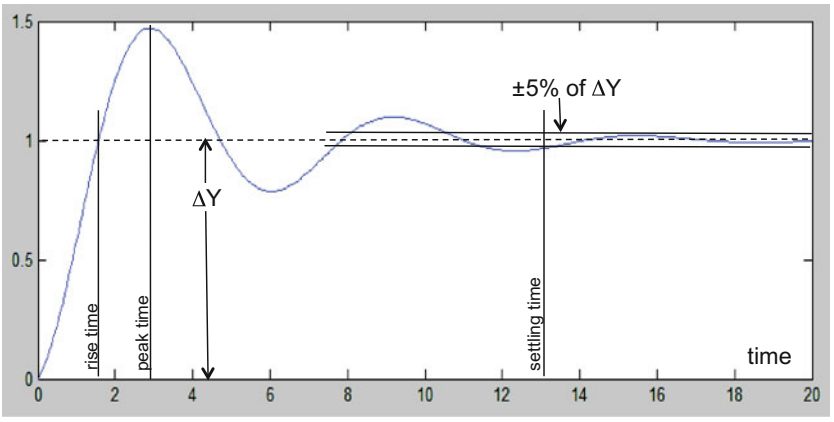


Fig. 10.4 Control system response to step input starting at time = 0

time will increase until you reach critical damping, at which point the time again goes infinite because you asymptotically approach steady state. While we're talking about overshoot, you'd also like to have a feel for how big the first overshoot will be. A good normalized way to express this is *percentage overshoot*, which is simply the percentage of the ratio (peak output - steady state)/(steady state). Naturally, lower damping (when underdamped) means greater overshoot, and when critically damped the overshoot goes to zero.

The previous three terms (rise time, peak time, and percentage overshoot) are useful in characterizing the behavior of the system during its first oscillatory cycle. Clearly, we also need to describe the system behavior after the oscillations all die out. First, we want to know how long it takes until all significant motion stops. That's specified by the *settling time*, which is the time required for the system oscillations to reduce in amplitude to a specified band about the steady-state solution. As you might expect from our previous discussions, the size of the settling time is inversely proportional to both the damping ratio and natural frequency. The steady-state solution need not, and in real life won't be, exactly the desired output. The steady-state error is a measure of the discrepancy between the desired signal and what the system actually is doing after all the transients have decayed. We haven't talked about control law gains yet (see next chapter), but increasing gain is one way to reduce the steady-state error, to the potential performance detriment of some of these other terms (for example, overshoot and settling time). In general, one way to view these five system response characteristics we've talked about is as a way to define quantitatively the system's required performance. And the damping ratio, natural frequency, and gains (on which the response characteristics depend) therefore provide a mechanism to adjust the control law's behavior to meet these requirements.

So far, our insights into system response characteristics have been gained by looking at how the system behaves when a step function is supplied as input. Remembering that engineers often think with a "black box" mentality, it seems reasonable to see if we can glean additional information by "tickling" the system with another simple form of input, like a sine or cosine function. One obvious thing to investigate is how the system's response changes as we change the frequency of the input sinusoid. Put in fancier terms, observing the output produced from sinusoidal input allows us to determine a system's frequency response. For example, the magnitude of the transfer function is a measure of the ratio of the output sinusoid amplitude to the input sinusoid amplitude, while the phase angle of the transfer function (remember, it's complex) measures phase differences between output and input. And for more complicated functions than simple sine or cosine waves, the principle of superposition allows us to Fourier decompose the input signal into a series of sinusoids, whose individual behaviors can be examined separately.

In addition to asking very straightforward quantitative questions like how much reduced from the input amplitude is the output amplitude, and how large a phase difference has been introduced by the operation of the control law, you also can ask a somewhat more qualitative question like, how closely does the output resemble the sinusoidal shape of the input? Again, there's a special vocabulary to provide this initially fuzzy concept with quantitative specificity. For example, a system's

bandwidth defines the maximum input frequency for which the output will continue to track the input sinusoid satisfactorily. The quantitative definition of "satisfactorily" usually sets the bandwidth to the frequency at which the output is attenuated to 0.707 times the input (also called being "down" 3 dB). So bandwidth is a measure of the system's speed of response in the frequency domain, just as rise time and peak time are measures of the system's speed of response in the time domain.

In most of the previous discussions of control law behavior there's actually been an implicit assumption that we were dealing with "good" control laws, i.e., ones that were stable. Obviously, if this was always the case, or easily achieved, there'd be a lot less need for controls engineers, and they'd be a lot poorer paid. So let's conclude this section with a paragraph regarding what formally constitutes stability. A system is considered *stable* if the output is bounded when the input is bounded. In mathematical terms, that means a system will be stable if the poles of the transfer function all have negative real components, yielding exponentially decaying behavior. If, by contrast, the output increases indefinitely with time, the system is considered *unstable*. This will be the case if any of poles have positive real components, yielding exponentially increasing behavior. (Of course, any such exponential growth only pertains to the linearized model. Such growth in any real system eventually saturates, possibly tearing the system apart, as with the collapse of the Tacoma Narrows Bridge in 1940, for example.) And there are a few in-between states of stability as well. For example, if the output oscillates without ever reaching steady-state, but the amplitude of the oscillations is bounded, the system is called *marginally stable*. This behavior is characteristic of poles that are purely imaginary, yielding undamped oscillatory behavior. And if the output does converge, but converges to some value other than the desired one, the system is said to possess *limited stability*. Lastly, if the system is only stable for finite ranges of control parameter values, it is called *conditionally stable*. Stability of the engineers who design control laws is a separate discussion topic (see Freud).

10.5 The Harmonic Oscillator in Detail

In this section we present the harmonic oscillator example worked out in detail. Consider the spring-mass-damper system of Fig. 10.3. The mass starts at rest with the spring extended due to gravity's action on the mass. We define the zero point for position to be where the spring force is zero, so the offset equilibrium position due to the gravitation force is given by:

$$\Delta y K = Mg \tag{10.9}$$

where

Δy = extension in spring at start time (units of meters)
K = spring constant (kg/s^2)

M = mass of object at end of spring (kg)
g = acceleration due to gravity (m/s^2)

The time domain equation of motion is:

$$My''(t) + fy'(t) + Ky(t) = (K\Delta y)r(t) \tag{10.10}$$

where

$y(t)$ = output, i.e., object position as a function of time (m)
$y'(t)$ = first time derivative of y (m/s)
$y''(t)$ = second time derivative of y (m/s^2)
f = friction constant (kg/s)
$r(t)$ = normalized input driving function (dimensionless)

The corresponding Laplace s-space domain equation (assuming zero initial conditions) is:

$$Ms^2Y(s) + fsY(s) + KY(s) = (K\Delta y)R(s) \tag{10.11}$$

where

s = s-space free parameter (1/s)
$Y(s)$ = Laplace transform of $y(t)$ (m-s)
$R(s)$ = Laplace transform of $r(t)$ (s)

The transfer function is given by:

$$\begin{aligned} G(s) &= \frac{Output}{Input} \\ &= \frac{Y(s)}{R(s)} \\ &= \frac{K\Delta y}{Ms^2 + fs + K} \\ &= \frac{\Delta y\omega_0^2}{s^2 + 2\xi\omega_0 s + \omega_0^2} \end{aligned} \tag{10.12}$$

where

$\omega_0 \equiv (K/M)^{1/2}$, i.e., "natural frequency" (1/s)
$\xi \equiv (f/2)/(KM)^{1/2}$, i.e., "damping ratio" (dimensionless)

The response to an impulse function input (i.e., a quick thwack with a hammer) is found by making the input $r(t)$ proportional to a "delta" function, so $r(t) = \alpha\delta(t)$, where α is the proportionality constant. The Laplace transform of $r(t)$ will be $R(s) = \alpha$, so we have:

$$Y(s) = G(s)R(s) = \frac{\alpha\Delta y\omega_0^2}{s^2 + 2\xi\omega_0 s + \omega_0^2} \tag{10.13}$$

The poles of the output $Y(s)$ are:

$$s_1 = -\xi\omega_0 + \omega_0(\xi^2 - 1)^{1/2} \tag{10.14a}$$

$$s_2 = -\xi\omega_0 - \omega_0(\xi^2 - 1)^{1/2} \tag{10.14b}$$

Doing a partial fraction expansion of Eq. 10.13, the output in the s-domain is:

$$Y(s) = \frac{k_1}{s - s_1} + \frac{k_2}{s - s_2} \tag{10.15}$$

where k_1 and k_2 are complex constants. The corresponding output in the time domain is:

$$y(t) = k_1 \exp(s_1 t) + k_2 \exp(s_2 t) \tag{10.16}$$

Note that at $t = 0$, $y(0) = k_1 + k_2$. If $\xi < 1$, s_1 and s_2 are complex conjugates of each other, as are k_1 and k_2. For this range of ξ, it is customary to write the quantity $\omega_0(\xi^2 - 1)^{1/2}$ from Eqs. 10.14a and 10.14b as $i\omega_d$, where $i = \sqrt{-1}$, and ω_d is called the "damped natural frequency". For the limiting case of $\xi = 0$, you find

$$y(t) = 1/2\left[\exp(i\omega_0 t) + \exp(-i\omega_0 t)\right] \tag{10.17}$$

$$= \cos(\omega_0 t) \tag{10.18}$$

i.e., the system oscillates at its natural frequency, ω_0.

The behavior of the output $y(t)$ varies as the damping ratio is changed. When $\xi > 1, = 1$, or < 1, you'll have an over damped, critically damped, or under-damped system (respectively). For the latter situation, you'll find a transient response that exhibits oscillations with decreasing amplitude. As ξ approaches zero, the transient oscillations never die out.

Part of our discussion in Sect. 10.4 considered an impulse force applied to the oscillator at $t = 0$, i.e., $r(t) = \alpha\delta(t)$, for which the Laplace transform $R(s) = \alpha$. We can also consider the effect of applying a step function force:

$$r(t) = \alpha u(t) = \begin{cases} 0 \ (\text{for } t < 0) \\ \alpha \ (\text{for } t \geq 0) \end{cases} \tag{10.19}$$

$$R(s) = \frac{\alpha}{s} \tag{10.20}$$

that would change the equilibrium position. Equation 10.13 changes to:

$$Y(s) = G(s)R(s) = \frac{\alpha\Delta y\omega_0^2}{s(s^2 + 2\xi\omega_0 s + \omega_0^2)} \tag{10.21}$$

which changes only in having an extra factor of s in the denominator, i.e., one extra pole at $s = 0$. Since the original poles of the transfer function don't change, the fundamental behavior in terms of natural oscillation frequency and damping remain the same. The only change is that the phenomena discussed in Sect. 10.4 of rise time, peak time, percent overshoot, and setting time come into play (see Fig. 10.4). The pole at $s = 0$ represents a shift in the equilibrium position for the oscillator. This is sort of a compromise between a disturbance fading away (as happens if $\text{Real}(s) < 0$) and growing indefinitely (as happens if $\text{Real}(s) > 0$).

Besides the insights learned from an impulse force $\delta(t)$ and unit step force $u(t)$, one can also gain useful insights by sending a sinusoid through the system. This allows one to determine the system's frequency response. In this case you would use:

$$r(t) = \alpha u(t) \sin(\omega t) \tag{10.22}$$

$$R(s) = \frac{\alpha\omega}{(s^2 + \omega^2)} \tag{10.23}$$

$$Y(s) = G(s)R(s) = \frac{\alpha\omega\Delta y\omega_0^2}{(s^2 + \omega^2)(s^2 + 2\xi\omega_0 s + \omega_0^2)} \tag{10.24}$$

where ω is the frequency of the input force. Now we have two new poles at $s = \pm i\omega$. Since these are purely imaginary, the associated disturbance neither grows nor diminishes. A damped harmonic oscillator driven forever at frequency ω will respond at the same frequency forever, with any disturbances at ω_0 fading away over time. This is not to say that the oscillator doesn't retain knowledge of its fundamental parameters ω_0 and ξ. These govern the strength and phase of the response. If $\omega \approx \omega_0$, the oscillator is being driven in resonance, meaning the response is very strong.

10.6 Adjusting Gains: The Root Locus Diagram

When considering a dynamic system from the perspective of a controls engineer, it is natural to ask how the behavior changes as you change various tunable parameters, where "tunable" might mean turning a control knob, changing a software parameter, or replacing a piece of hardware. A tool that control engineers regularly use to examine such behavior is the so-called *root locus diagram*. In Fig. 10.5 we present a sample root locus diagram for the rate gain parameter (i.e., the damping ratio ξ) for the harmonic oscillator example discussed in the previous section. (If you didn't read Sect. 10.5, don't worry; the example is just for illustration.)

The diagram shows how the poles change in the complex plane as the gain, ξ, changes from 0 to $\pm\infty$, with the poles for this specific example being as given by Eqs. 10.14a and 10.14b. The plot is normalized so that the system's natural frequency is set to unity (i.e., $\omega_0 = 1$). For generality we have extended the plot to include negative damping ratio (i.e., $\xi < 0$) which would correspond to positive rate feedback

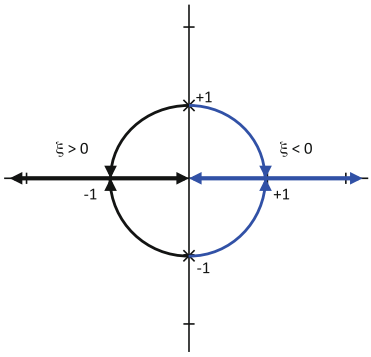


Fig. 10.5 Root locus diagram for variable rate gain

- not something that would be realized via friction, so just think of it as a generic illustration.

For the example, the poles are at $\pm\omega_0$ (the two $\times$'s in the figure) when $\xi = 0$. As discussed in Sect. 10.5, this corresponds to pure oscillations with no damping. As ξ increases from 0 to 1, the two poles move down to the real axis, merging at $s = -1$ (i.e., the point of being critically damped). As ξ continues to increase, one pole moves towards $s = 0$, asymptotically reaching 0 as ξ reaches $+\infty$, while the other pole moves towards $s = -\infty$.

The positive feedback curves in the figure (i.e., those for $\xi < 0$) are mirror images of the negative feedback curves. Having Real$(s) > 0$ everywhere, these curves represent unstable systems that would amplify without bounds any disturbance whatsoever. For $-1 < \xi < 0$, the amplification would involve oscillations with ever increasing amplitude. For $\xi < -1$, the disturbance would simply grow exponentially with no oscillation.

Although the light blue portions of the root loci curves are nonphysical for the damped harmonic oscillator example, having root loci in the positive-real part of the complex s-plane may well make physical sense in some control system designs - imagine the runaway screech in an audio system with the speaker too close to the microphone. Quite generally, it's a bad thing when modification to a gain parameter causes a root locus branch to cross the frequency space imaginary axis, though perhaps not as bad as crossing the streams in the movie *Ghostbusters*. In our case it just means we have to restrict the gain value, not deal with the consequences of all of a spacecraft's protons repelling from each other at the speed of light. Logically (leaving sci-fi cinema behind), the further away the poles are from the imaginary axis (while remaining on the negative side of the real axis), the more stable the system will be. However, stability isn't the only consideration that one has when designing a control system. Having a system that is ultra-ultra stable may translate into uselessly sluggish and non-responsive. One looks for a reasonable compromise between stability and responsivity.

Control engineers have technical characterizations for stability safety margins in their systems. These characterizations go by the names *gain margin* and *phase margin*. Their technical meaning is not as simple as "so far to the left of zero in

frequency space," but they roughly translate into that. If you hear either term in a control systems presentation regarding a new or modified control law, the speaker is referring to how much he had to back off from a marginally stable configuration to achieve an acceptably robust design. Gain margin is a multiplicative factor typically expressed in decibels. A rule of thumb is that the lowest acceptable gain margin is 3 dB, i.e., a factor of $\sqrt{2}$. Phase margin is an angular measure; the rule of thumb is that the lowest acceptable phase margin is 30°. Use of lower margins may indicate that the designer enjoys bridge diving as a hobby.

10.7 Discrete Systems and the *Z* Transform

All of our discussion in this chapter, and all the tools and methods presented, pertain to continuous systems. Unfortunately, all of the onboard applications we work with are discrete systems. For example, attitude sensors output data at fixed time intervals, not continuously, and even if they did output their data continuously, they would only be interrogated at fixed times in the control cycle (the sampling period). Our onboard calculations utilizing the sensor data, as well as onboard error computations, onboard torque command generation, etc., are also performed within well-defined time intervals, and the actuator commands are issued (and executed) at set times in the control cycle too. One thing that should be immediately clear is that any control system based on discrete measurements will be unable to meaningfully detect and respond to disturbances with time scales shorter than its own control cycle time, Δt. Indeed, the system will be fairly myopic even at time scales equal to and slightly greater than Δt. Generally speaking, disturbances will only be well controlled at frequencies at least an order of magnitude or so below the sampling frequency. Even for those frequencies there will be a slight phase lag in control because the system will be able to introduce corrections for disturbances no sooner than a cycle after an error is detected.

Another implication when doing careful analysis on discrete systems is that the continuous integral Laplace transform must be replaced with a summation expression called the *Z* transform expressly designed for discrete systems. The *Z* transform takes you from "sample space" (where each sample is associated with a discrete sample time) into "z space". Formally, a *Z* transform has the form:

$$F(z) = \mathcal{Z}\{\{f_n\}\} \equiv \sum_{n=0}^{\infty} f_n z^{-n} \tag{10.25}$$

where $\{f_n\}$ is an infinite series, just as $f(t)$ was a function defined over all time $t > 0$. If you compare this equation with Eq. 10.1, you'll see that the *Z* transform is similar to the Laplace transform with $z = e^{st}$ and $t = n\Delta t$. For our limited purposes, we don't actually care a great deal about the *Z* transform except for one detail. When drawing control system diagrams (examples to be shown in the next chapter),

control law engineers use a symbol from Z transform theory when indicating that a term proportional to rate is to be used in a feedback loop. Specifically, when the current cycle's measurement of some quantity, f_n, is to be combined with a prior cycle's measurement, f_{n-1}, to form a rate-like term, $(f_n - f_{n-1})/\Delta t$, the operation is indicated by a symbol of the form "$-[z^{-1}]\mapsto$". This usage stems from the fact that the Z transform of the shifted sequence $\{f_{n-1}\}$ (with an assumed extra leading term of $f_{-1} = 0$) is given by:

$$\mathcal{Z}\{\{f_{n-1}\}\} = \sum_{n=0}^{\infty} f_{n-1} z^{-n} = \sum_{n=0}^{\infty} f_n z^{-n-1} = z^{-1}\mathcal{Z}\{\{f_n\}\} \tag{10.26}$$

Just as differential equations for continuous systems can be handled via Laplace transforms, difference equations for discrete systems can be dealt with using Z transforms. We are, however, not going to present any examples of such analysis. At this point we really only want to alert you to the fact that the Z transform exists for discrete systems and that the associated symbol "$-[z^{-1}]\mapsto$" indicates a first derivative, "$-[z^{-2}]\mapsto$" indicates a second derivative, and so on; and that (by symmetry) the symbol "$-[z]\mapsto$" indicates an integral term. Some of you may note a possible oddity in this usage. The control system doesn't have access to the whole infinite sequence $\{f_n\}$, so it really can't construct a Z transform. If the system's logic is just reusing the prior value of f, why not represent that as "$-[f_{n-1}]\mapsto$"? All we can really say is "when in Rome, learn to speak Latin" - or in this case, when talking with control law engineers, learn to understand their jargon.

Chapter 11
Control Laws: Attitude Applications

There's nothing that excites a scientist more than discovery of something new in nature. Sometimes, when the object or phenomenon is first predicted by theory and then observed later experimentally, it's a cause for great celebration, at least among the proponents of the theory. Other times, the new object or phenomenon just comes out of left field. In a rather famous example from theoretical physics, no one predicted (or had a theory that required) the existence of the muon, an elementary particle that appears to be identical to an electron except it's over 200 times heavier. When it was discovered anyway in cosmic rays in 1947, the general head-shaking in the physics community over the apparent lack of purpose to the particle led the Nobel Prize winning physicist I.I. Rabi to complain, "Who ordered this?"

It is somewhat ironic that when one of these "surprise" discoveries happens, the first thing the science community tries to do is shoehorn it into some previously understood pattern. For example, in 1911 Rutherford demonstrated in a scattering experiment that an atom of gold behaved like an object having a very small "hard" positively charged nucleus surrounded by a much larger electron cloud containing an equal amount of negative charge, rendering the atom electrically neutral. This experimental result was greatly at variance with the then-popular "plum-pudding" model, which viewed the atom as a mixture of heavy positively charged objects and light negatively charged objects that somewhat uniformly filled-out the atomic volume. Two years later, Bohr took Rutherford's results and the empirical formula for hydrogen spectral lines developed by Balmer in 1885, and came up with a model theorizing that an atom having a radius of about an angstrom (10^{-10} m) and central mass of 1.7×10^{-27} kg works pretty much the same as our Sun–Earth system having a radius of 93,000,000 miles (1.5×10^{11} m) and central mass of 2×10^{30} kg, with Newton's inverse square law for gravity replaced by Coulomb's inverse square law for electrostatics. Considering that things typically don't scale up or down very successfully in nature over an order of magnitude or two (for example, the design of a greyhound's leg doesn't look much like an elephant's), Bohr's conceptual parallel between atomic systems and solar systems might be considered by a rational person to be breathtakingly reckless.

© Springer-Verlag London Ltd. 2017

H.L. Hallock et al., *ACS Without an Attitude*, NASA Monographs
in Systems and Software Engineering, DOI 10.1007/978-1-4471-7325-0_11

In fact, the models don't match up at all without adding in two additional constraints, namely Bohr's First Postulate that requires that only a discrete set of electron orbital radii are legal and Bohr's Second Postulate that requires that (in spite of what classical electrodynamics says happens for all accelerating charged objects) no electron radiates energy unless it leaves its current orbit. With just those two "kluges" in place (it's pretty remarkable that only two tinkerings were needed), Bohr was able to replicate Balmer's formula for the wavelengths of spectral lines and other known results, and successfully predict new spectral relationships not previously measured. Eventually, Bohr's simple model started falling apart, to be replaced by a quantum mechanically consistent model (pioneered by, of course, Bohr), but in its day Bohr's planetary model for the atom seemed the perfect tool for explaining the then new phenomena of atomic physics.

11.1 Equivalence of *L-R-C* Circuits and Harmonic Oscillators

In this respect (i.e., the shoehorning of old models into new phenomena), engineers' mental processes have much the same tendencies as those of the physicist. What the solar system is to physicists, an *L-R-C* (inductance - resistance - capacitance) circuit is to engineers. Whenever they have to dream up something new, they always seem to fall back on their old friend from electrical engineering. In classical circuit design, L, R, and C are the only elements that change the electromotive force (emf). They show up as coefficients to terms in a differential equation for electrical charge as follows. L multiplies the second derivative of charge, R multiplies the first derivative of charge, and $1/C$ multiplies the charge, which now makes the equation for emf look a lot like our friend from Chap. 10, the harmonic oscillator, the physicist's favorite idealized model. In the electrical circuit case, L takes on the role of the mass on the spring (M), R provides the damping effect (like the friction constant, f), $1/C$ is the spring constant (K), and the natural frequency of the circuit is given by $1/\sqrt{LC}$, as compared to $\sqrt{K/M}$ for the harmonic oscillator. Of course, since the form of the differential equation is the same for both a harmonic oscillator and an *L-R-C* circuit, a purist mathematician would insist that the two problems do not resemble each other; they are identically equal to each other.

Actually electrical engineers normally work with the first derivative of charge, the electrical current, in which case L multiplies the first derivative of the current, R multiplies the current, and $1/C$ multiplies the integral of the current. In this form, the electrical engineer's favorite idealized model maps directly onto the controls engineer's favorite idealized model, the PID (proportional-integral-derivative) control law. But we're getting a bit ahead of the game. Before getting into the different types of control laws, let's first talk about how an attitude control law functions and about some of the standard elements that go into constructing one.

11.2 PID Control Laws

For spacecraft flying gyros, which provide only attitude change information, and an attitude sensor (or sensors) providing absolute attitude information, the attitude control law combines the two types of information needed to estimate required pointing corrections relative to the desired pointing (or pointing profile). The control law determines the command torques designed to

1. maintain observation of the science target;
2. maintain spacecraft stability;
3. slew to new science targets;
4. perform small attitude maneuvers under fine pointing control;
5. perform small attitude maneuvers to locate science targets;
6. compensate for time-varying changes in apparent target direction, such as those arising from velocity aberration;
7. follow moving targets, such as asteroids, planetary moons, comets, etc.

This is a pretty long laundry list of activities the control law must be able to support, many of which seem like opposites of each other, such as both being able to change the pointing by a large amount and maintain it steady to a high level of precision. That suggests either that you're going to need a lot of different, specialized control laws to support all these functions, or you're going to need an awfully complicated single control law that can do anything. In fact, both of these approaches have been used, and which approach gets implemented on a given mission seems to be as much dependent on the personal preferences of the mission's controls analyst as on the intrinsic characteristics of the mission itself. What that suggests is that control theory is an extremely powerful tool, and there are a lot of satisfactory approaches to skinning the same cat. Although characterized by high levels of sophistication and flexibility, the basic structure of an attitude control law can be built from some fairly simple functionality. For example, you can

1. use gain multipliers to scale different error contributions;
2. limit the contributions of terms to prevent overflows, satisfy hardware constraints, or avoid inducing excessive jitter;
3. smooth, average, or accumulate error terms to avoid "whipsaw" response and extract secular trends from noise;
4. leak-in corrections over time when delicate handling is required

The next step up in complexity from these basic mathematical operations is to construct a feedback loop to pick up errors previously not compensated for. The three flavors of feedback we'll concentrate on here are proportional, integral, and derivative, which of course map to the pieces that constitute a PID control law (see Fig. 11.1a). For grins, weve included a diagram showing the same control law in Z-jargon notation (see Fig. 11.1b).

The general form of proportional feedback is $u_p = K_p e(t)$, where u_p is a control torque, K_p is a gain (i.e., an adjustable parameter used to modify the response), and e

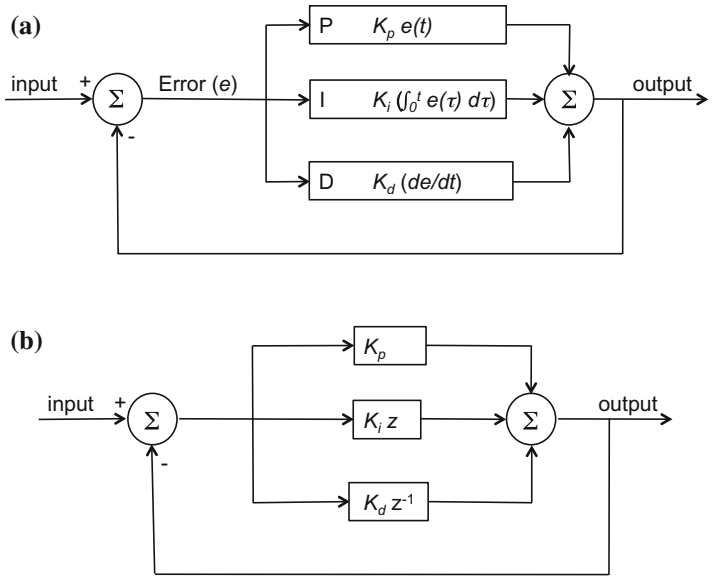


Fig. 11.1 **a** Simplified PID control law. **b** Simplified PID control law in Z-transform notation

is the error signal determined by differencing processed sensor data relative to desired values. So if the error signal falls to zero, the control torque goes to zero as well. Integral feedback looks like $u_i = K_i \int e dt$, where u_i is a second control torque, and K_i is the integral term gain (a second adjustable parameter). Whereas the proportional term used an instantaneous snapshot of the error signal to determine a corresponding instantaneous control torque, the integral term sums the influence of a swath of error terms over time to determine its current, instantaneous control torque. That means that unlike the proportional term, which falls to zero as soon as the error term does, the integral term will maintain a constant value if all successive error signals are zero. Therefore, if you have a constant disturbance affecting your spacecraft (say, a steady solar radiation torque), an integral feedback term could build up to the appropriate value to null it out. Finally, the general form of the derivative feedback is given by $u_d = K_d(de(t)/dt)$, where u_d is a third control torque, and K_d is the derivative term gain (a third adjustable parameter), and $(de(t)/dt)$ is the rate error. Depending on the sensor suite, $(de(t)/dt)$ can either be the derivative of e as defined above (e.g., a time difference of attitude error as determined from two consecutive star tracker measurements), or it may be rate error estimate based on a distinct sensor (e.g., $(de(t)/dt)$ being the difference between spacecraft rate measured with gyros and a currently desired rate). So if the change in error signal drops to zero, the derivative term will drop to zero, which means that the derivative term will be insensitive to steady-state error. Figure 11.2a and b show how the three pieces of the puzzle (proportional, integral, and derivative control) are put together to build a PID control law for a typical LEO spacecraft mission, with attitude control torque commands

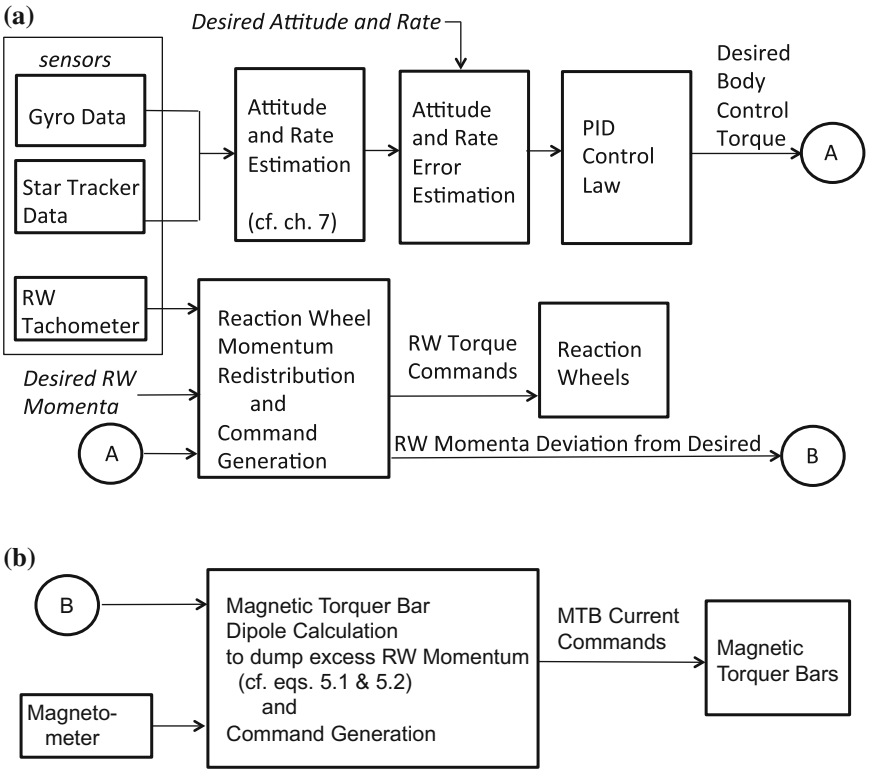


Fig. 11.2 **a** PID control law for attitude control in context. **b** Control of reaction wheel momentum with magnetic torquer bars

going out to the reaction wheel assembly (RWA). The figure includes the extra level of detail associated with reaction wheel momentum management (i.e., saturation avoidance) realized via magnetic torquer bar (MTB) coupling to the geomagnetic field.

There is a caveat that you should keep in mind with respect to the preceding discussion. All measurements are somewhat stale, even the ones most currently available for use. ST measurements require a finite exposure; gyroscope measurements require a finite rotation to occur; data signals require a finite time to be delivered to the CPU for processing. Thus the error(s) used in any closed loop control law (e.g., PID control) are not current error(s) relative to what should be the attitude state now, but rather slightly stale error(s) relative to what should have been the case a little while ago (at least half of a measurement cycle). Similarly, there is a slight lag in getting the resultant commands out to the actuators, which compounds the lag in the system response. This control system lag will itself be a contributor to the errors that the control system is trying to correct. The noise content of such lag-induced errors will

be at frequencies of $1/\tau$ and higher, where τ is a few times longer than the ACS control cycle time.

For clarity, let's take a step back and look at control system responsivity as progressively more detail is added, starting from a very primitive "bang-bang" system, working our way through proportional and proportional-derivative systems, and finally returning to proportional-integral-derivative systems.

11.2.1 Bang-Bang Control

Bang-bang just uses part of the information utilized by proportional control. Instead of scaling the size of the response relative to the error signal using the gain, bang-bang only cares about the sign of the error signal. Bang-bang control torques only take on three values, a fixed positive number, a fixed negative number, and zero. This kind of control law has been used (infrequently) in the past for some missions employing only magnetic torquer bars as actuators. A somewhat better approach is *bang-bang plus deadband control*. As with bang-bang, there are only three command torque values, but the non-zero values are only commanded when the magnitude of the error exceeds a threshold. The deadband protects against whipsawing effects by simply driving the error into a "close enough for government work" region rather than aiming at attaining "perfect" error reduction. Sometimes even more protection against unnecessary, wasteful torquing is provided through the specification of two sets of limits. A hysteresis construction of this sort will support an algorithm that only corrects the error until the error gets below a lower limit and, once settled, doesn't start correcting again until the error rises above a higher limit. This bang-bang with deadband and hysteresis control approach is often used on GSFC missions where thrusters supply the control torques.

11.2.2 Proportional Control

As mentioned previously, the system response under proportional control is a function of the gain. As the gain is increased, the system responsiveness improves, i.e., the response time decreases. However, if the gain is too high, high frequency oscillations with large overshoots can occur and the system can become unstable. That's because proportional control can be too responsive to the instantaneous error term, which may be greatly corrupted by noise. So instead of convergence to a desired value, you may end up with wild oscillations that fail to converge at all. On the other hand, if the gain is too small, proportional control will not be responsive enough to the error signals that must be nulled and the system again can go unstable. The bottom line is that there's a limit to the amount of error you can take out using proportional control alone, hence the need to introduce other, more sophisticated, feedback devices.

11.2.3 Proportional-Derivative Control

If the biggest problem with proportional control is that when the gain is too high, the control law becomes "twitchy" (i.e., responds too quickly and too massively), a logical first step would be to add in a term that slows the response down, without having to reduce the gain. Recall from our harmonic oscillator and *L-R-C* circuit examples that the term supplying the damping in the differential equation is the first derivative term (first derivative of position for the harmonic oscillator, first derivative of charge for the *L-R-C circuit*). So if we add a derivative term, also called a *rate term*, to the mix, we should introduce a damping effect that will reduce oscillations originating from the proportional term, thereby improving stability. This kind of control law is called *proportional-derivative* (PD). When undamped, the system oscillates without restraint, i.e., you have pure proportional control. As damping is added, overshoot is reduced and oscillations are decreased. If the system is overdamped, overshoots and oscillations are eliminated, but system response will be slow, or even unable to reduce the steady-state error to an acceptable value. If underdamped, the system will oscillate with decreasing amplitude, gradually reaching steady state. Figure 10.4 shows typical performance for an underdamped PD control law. If the system is critically damped, you'll hit the target perfectly, going straight to steady-state without overshoot, undershoot, or oscillations. So PD control, if the damping is set properly, sounds at first like a pretty good deal. However, because the derivative term is differencing error signals, a PD can undesirably amplify high frequency noise. Also, as mentioned earlier, the derivative term is insensitive to steady-state error, so a PD can leave you with a large uncorrected error at the end. Nevertheless, PDs have been used in flight, as for example with missions utilizing gyros for attitude error and rate information, and reaction wheels for control torque application.

11.2.4 Proportional-Integral-Derivative Control

To deal with the problem of large uncorrected steady-state error, we need to add more sophistication and move up in class to a full-fledged *proportional-integral-derivative* (PID) controller described earlier. Unfortunately, the improvement can come at the cost of reduced stability. In our attitude control applications, this problem can display itself in actuator saturation. If too large an error is detected, the control torque will be limited since control actuators have maximum output capabilities (i.e., the actuator will saturate). Because the commanded torque is limited, the error signal cannot be zeroed and the integral term will continue to grow. If saturation continues for an extended duration, the integral term will therefore grow very large. After saturation has ceased, you'll still have a very large integral term, which can produce large overshoots even though the current instantaneous error signal now is not very big. One solution to this problem is to stop integrating as long as the actuator is saturated. In addition to its ability to reduce the steady-state error, a PID provides a generally

higher level of flexibility simply from its ability to tune three control parameters (the gain, integral time, and derivative time). For example, increasing K_p or K_i allows you to reduce system errors, although potentially at the risk of damaging stability. And increasing damping by increasing K_d can help increase stability.

11.3 Hubble Space Telescope Pointing Control Law

Section 11.2 provides a reasonably accurate (albeit simplified) description of basic control law processing for many NASA missions, with Fig. 11.2a, b capturing the main highlights for most LEO missions (well, ignoring those that also use thrusters). We'd like now to look in a bit more detail at a mission with more-than-generic pointing requirements, i.e., the Hubble Space Telescope (HST). We're not about to subject you to the HST control law in all of its detail (which can be pretty intimidating), but there are some extra details worth discussing at a high level.

The sensors of interest for HST pointing control are its three fixed-head star trackers (FHSTs), three fine guidance sensors (FGSs), and its gyroscopes. The actuators are its reactions wheels (RWs) and magnetic torquer bars. Of these five components, only the FGSs are unique to HST. The FGSs sample light from the HST primary optical system, thereby taking full advantage of the telescope's light gathering ability and resolution. Each FGS full field of view (FOV) consists of an arc around the main optical axis with an azimuthal range of 82° and a radial range extending from ∼10 to 14 arcminutes. The three FGS FOVs (would you prefer FsOV?) together form a three-quarter annulus (more-or-less a horse-shoe shape), interior to which all of the HST science instrument FOVs reside. The angular area of the FGS FOVs combined is only about 0.026 square degrees, in contrast to each 8° × 8° FHST FOV. The magnitude range for guide stars usable by the FGSs is ∼9–16m_V. When fully calibrated, the FGSs have an accuracy for determining relative star position of ∼0.003 arcsecond, though they must be regularly recalibrated to preserve this measurement accuracy. This high-precision is possible based on the FGS design as an amplitude interferometer using Koesters prisms (don't worry about what they are, though you can find discussions on the web) combined with photomultiplier tubes. It may be compared with HST's optical resolution (∼0.050 arcsecond) and the FHST measurement accuracy (∼20 arcseconds).

An FGS can measure the position of only one star at a time. Each FGS has a 5 × 5-arcsecond instantaneous FOV (IFOV) that can be commanded to a selected position within the FGS full FOV. The FGS sees the star if it falls within the IFOV. A star image falling within the inner 0.020 arcsecond of the IFOV will produce a significant interferometric signal. The FGS is said to be in "fine lock" when so measuring a star's direction. A second mode of FGS operation, "coarse track", is also available. In this mode, the center of the IFOV is commanded to rotate about the true star position in such a way that the edges of the IFOV cut across the image of the star in a symmetric pattern (no interferometric processing). The accuracy of determining star positions using coarse track mode is ∼0.020 arcsecond. Coarse track mode is less sensitive

to spacecraft jitter than is fine lock mode, particularly for faint stars; it is available to support science observations in situations in which high pointing precision is not required.

With that overview of FGS functionality under your belts, let's now turn to the functioning of the HST control system. Let's start with HST having just completed an observation and about to move on to another at a new desired target, typically in some quite different direction in the sky. The first task is to reorient the observatory. There is a problem, however; if the observatory were commanded to swing as quickly as possible to the new pointing, with nearly instantaneous acceleration and deceleration, the sudden start and stop would induce an overall ringing of the spacecraft structure that would take considerable time to die out, thereby making it impossible for the FGSs to immediately lock onto guide stars (thereby delaying the observation), and also blurring science images even if FGS lock could be obtained. To avoid (or at least minimize) this structural ringing, HST follows a smooth slew profile when changing attitude, with no discontinuous changes in angular acceleration anywhere along the way. The slew profile, characterized by a small number of parameters (slew angle, maximum rate, maximum acceleration, maximum acceleration derivative [a.k.a., jerk]), is computed in the form of desired slew angle and rate as smooth, analytic, functions of time. HST's control logic then uses gyro-only measurements to determine true rate and propagated attitude through the slew. These measurements are then compared at each moment through the slew with the profile-specified desired rate and attitude to generate rate and attitude errors for use in PD control law processing. (In Fig. 11.2a, imagine the star tracker(s) temporarily inactive, the desired attitude computed onboard as a function of time since slew start, and the PID control law replaced with a PD law.)

There is a further bit of sophistication that some satellite control laws, including HST's, apply to further reduce jitter introduced by slewing. Recall our earlier discussion of how the intrinsic staleness of observations results in a lag in control law response, and how this lag itself results in control errors at frequencies comparable to the control cycle frequency and above. To somewhat suppress such lag-induced noise, the HST control law adds onto its PD-computed torque an extra "feed-forward" torque based upon the slew profile acceleration. Without the feed-forward, the control law could only respond to errors after the fact and would tend to overshoot a bit whenever the desired (profiled) rate changed. By applying feed-forward, the control law is able to anticipate changes in rate as they are needed, thereby reducing the PD response to mere tweaks to an otherwise open-loop commanding of the slew. (So the control law box in Fig. 11.2a actually becomes for slew processing "PD Control Law with Feed Forward.")

Upon arrival at the new (i.e., post-slew) attitude, the HST star trackers are engaged to estimate any actual attitude error that may have developed over the course of the slew as a consequence of any miscalibration of the gyroscopes. The HST FHSTs are 1970 s technology, so they don't have the smarts to determine attitude on their own. What really happens is that the ground-based mission scheduling system will have determined for each of two FHSTs where an isolatable bright star should appear when the slew is complete. The onboard logic then compares these predicted positions with

the observed positions for the two stars and generates a small attitude adjustment quaternion to apply to the post-slew attitude estimate that integration of the gyro data had produced. With the attitude estimate corrected, the desired attitude known, the control law uses gyro data as input to remove the residual post-slew error. (So we have another variation on Fig. 11.2a, now with the star trackers active for a single cycle, and the "desired attitude" transformed to a specification of what the trackers should observe.)

So, with the telescope having completed its slew and post-slew adjustment, it is then pointing in nearly the proper direction for science, to within the accuracy of the FHSTs and their alignment calibration with the FGSs. Scientists with precise FGS FOV maps will have selected an isolatable guide star for each of two FGSs (sometimes just one) that can be placed in appropriate places in the FGS FOVs such that the desired science target will fall in a desired location in the science instrument FOV. However, given the FHST accuracy of ∼20 arcseconds, and the FGS IFOV size only 5×5 arcseconds, the selected guide stars will typically not immediately be within the FGS IFOVs. The FGSs are therefore commanded to execute spiral search patterns with the IFOVs until the stars are found (a few 10s of seconds), after which they are commanded to transition to coarse track, followed usually by fine lock. For our purposes, it doesn't really matter whether coarse track or fine lock is used. (The logic of this paragraph would result in a number of variations on Fig. 11.2a: (a) holding spacecraft attitude fixed based on gyro data while simultaneously controlling the FGS IFOV search for the guide stars, (b) a post-search attitude adjustment analogous to the post-slew attitude adjustment, and finally maybe (c) again holding the attitude fixed using gyro data while simultaneously executing the FGS fine lock procedure.)

With guide star lock established, the next step for the FSW is to compare the observed positions of the guide stars with the desired FGS coordinates that the scientist had selected to determine the attitude error in both pointing and roll if two guide stars are being used, or just pointing if only one guide star was specified. With the attitude error determined, the FSW commands a small slew to move the telescope to put the guide stars where they need to be to get the science target where it belongs in the science instrument FOV. Sound familiar? This is basically (at least on a logical level) just a repeat of the problem of correcting the post-slew attitude using FHST data. There are a few complications for how this is realized – for example, the optimal attitude for finding the guide stars in the FGS fields is generally not exactly where the telescope needs to be pointing for science, so there may be some extra pointing tweaks along the way – but you get the general idea.

There is a complication on "the general idea." Recall our discussion of velocity aberration from Sect. 6.6; the apparent position of a star shifts by an amount in radians of $\sim(v/c)\sin(\theta)$, where v is the velocity of the spacecraft relative to the Sun, and θ is the angle between the direction to the star and the direction of spacecraft velocity. Motion of HST about the Earth, of Earth about the Earth/Moon center-of-mass, and of the Earth/Moon center-of-mass about the Sun all contribute to velocity at a level significant relative to HST measurement accuracy; indeed, the accuracy requirements are such that the computations must be done using Einstein's relativity theory. The largest contribution to the amplitude of the velocity is the motion about the Sun

($v/c \approx 20$ arcseconds, with a period of 1 year), but the largest contribution to the rate of change of velocity is the motion about the Earth ($v/c \approx 5$ arcseconds, with period ~90 min). Over the course of a single orbit, HST needs to execute a small oscillation in pointing, roughly through an ellipse of size 5 arcseconds, to keep proper pointing at the science target. Furthermore, because of the small differences in θ for the guide stars versus the science target, although the guide stars and target object appear to move in lock step at the arcsecond-level, there are slight differences at the level of a few tens of milliarcseconds. To keep the target properly positioned, the desired guide star tracks follow small ellipses of semi-major axis ~40 milliarcseconds relative to the FGS FOVs, this effect superimposed on top of the larger 5-arcsecond elliptical motion. So, cast in control law parlance, the PID components of the control law work on FGS errors relative to those desired 40-milliarcsecond elliptical trajectories, while a feed-forward torque component is applied to the reaction wheels to open-loop follow the bulk of the 5 arcsecond motion. (Yes, there are further variations on Fig. 11.2a, but at this point consider it an exercise for the reader to think through these changes. By the way, all these variations would be [in fact, were] manifest using Z-transform symbology in the HST control diagrams provided by the HST system engineers at Lockheed-Martin.)

Note the "I" in PID in the previous paragraph. With the FGSs continuously engaged, use of an integral term on the FGS error signal during this process allows the control system to correct for uncompensated gyro drift bias. The astute among you, those that were paying attention to our discussion of the attitude+gyro-rate recursive (Kalman) filter (c.f. Chap. 7), might ask whether keeping the gyro bias well calibrated was suppose to have been the job of a Kalman filter. Well, it turns out that HST doesn't use an optimal, recursive least-squares (Kalman) filter for attitude estimation. With exception of the control law integral term, the HST control law uses only instantaneous, unfiltered FGS measurements. This method does not yield the statistically optimized results of a Kalman filter, but enables the generation of more compact flight code than would have been true of a Kalman approach, a vital concern given HST's narrow CPU and storage margins with its original flight computer. Besides compensating for gyro drift errors, the integral term also attenuates low-frequency disturbance torques, such as gravity gradient torque. To avoid potentially introducing a large error at a new science target, the HST FSW resets the integral term to zero whenever the spacecraft slews to a new attitude.

11.4 Control Law Tuning

Now that we've examined an attitude control law that worked on-orbit (although required post-launch modification due to unexpected solar array dynamic behavior), let's look at how you go about tuning a control law pre-launch to work on-orbit. First, you need to specify the system performance requirements that must be satisfied. Expressed in the time domain, that means defining values for the desired peak time or rise time, maximum overshoot, settling time (for example, for a step input), and

maximum steady-state error (look back at Fig. 10.4 for a pictorial representation of these parameters).

The most natural first approach to take in satisfying these requirements is the PID, with which you have available three free parameters (proportional (K_p), integral (K_i), and derivative (K_d)) for adjustment in order to achieve your goals. Specifically, if the rise time is too slow, you can increase K_p, which will also raise the natural frequency. To improve stability and decrease overshoot, you can increase K_d, thereby increasing the damping effect. If you've got a problem with a transient that persists too long, you can move the pole to the left (i.e., make the real part more negative) by appropriate adjustments to K_p and K_d. And to reduce steady-state errors, you can increase K_p or decrease K_i. If none of these kinds of changes to the three free parameters gets the job done, the next logical thing to try is adding a compensating network, also called a filter. This was the approach used to deal with HST's unexpected solar array "twanging", which resulted from sudden differential heating or cooling of the arrays as HST made night-to-day or day-to-night transitions. To finish off this chapter, we'll examine a few such filters qualitatively, staring with a generic first order filter.

So, what is this notion of filter order? The order of a filter is the highest polynomial order of either the numerator or denominator of the associated transfer function. Thus a first order filter transfer function can be written in the form:

$$G(s) = K(s+z)/(s+p) \tag{11.1}$$

where K is a gain, and z and p are the negatives of the filter transfer function zero and pole, respectively. We assume for this discussion that both z and p are both real and positive. (Recall that if the real part of p were negative, than the associated pole would be positive, and the filter would be unstable.) We'll also assume $s = i\omega$, i.e., we'll be concentrating on the effect of a pure sine wave input. G/K can be written as

$$\begin{aligned} G(\omega)/K &= (z+i\omega)(p-i\omega)/(p^2+\omega^2) \\ &= ((zp+\omega^2)+i\omega(p-z))/(p^2+\omega^2) \end{aligned} \tag{11.2}$$

for which the magnitude and phase are

$$M(\omega) = (((zp+\omega^2)^2+\omega^2(p-z)^2)/(p^2+\omega^2)^2)^{1/2} \tag{11.3}$$

$$\Phi(\omega) = \tan^{-1}(\omega(p-z)/(zp+\omega^2)) \tag{11.4}$$

There are two distinctly different behaviors, depending on whether $z < p$ or $p < z$. In the former case where $z < p$, (see Fig. 11.3), $M(\omega)$ starts out small (at z/p) for ω near 0, increases slowly until ω becomes comparable to z, increases relatively quickly for ω between z and p, and then asymptotically approaches 1 for large ω. Meanwhile, $\Phi(\omega)$ starts out at 0, increases slowly until ω becomes comparable to z, changes relatively quickly between z and p (going through a maximum at $\omega = (zp)^{1/2}$), and then asymptotically returns to 0 for large ω. This behavior is both

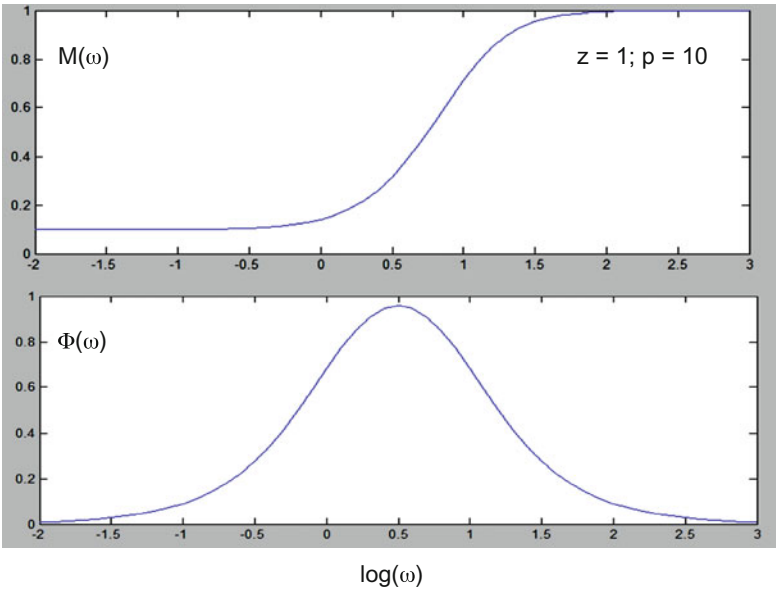


Fig. 11.3 Example of a first-order, high-pass, phase-lead filter

high-pass and phase-lead, with the phase-lead behavior pertaining particularly for the region between z and p. In this way the filter behavior is qualitatively similar to the effects of a derivative component (simple multiplication by s).

Relative to the four performance requirements we're trying to satisfy, a phase-lead filter will lower rise time and decrease transient overshoot. In many ways, a phase-lead filter acts like a PD control law, but alleviates the high frequency amplification characteristic of a PD. Since you can keep the gain lower at high frequency, a phase-lead filter will make the control law less sensitive to sensor noise. You can attenuate the noise still further by adding more than one pole. Reduction of the high frequency gain by adding poles is also a way to avoid possible resonances in the higher frequency regions, a technique called *amplitude stabilization*. There's also a technique called *phase stabilization* that, for cases where a resonance frequency location is well known, reduces the gain at that frequency with a notch filter, which we'll discuss later. Note that some of these potential high frequency resonance problems take care of themselves in real life because high frequency modes require a lot of energy to get excited (i.e., at high enough frequencies, neither the ACS nor external perturbations can provide enough energy to get the mode excited). In general, the phase-lead filter gain is tuned to accomplish two objectives. First, you must have a minimum low-frequency gain to reduce the steady-state error. Second, the high-frequency gain must be constrained to a maximum value in order to enable acceptable noise performance and to ensure a low probability for instabilities arising from plant modeling errors.

Let's now consider the case where $p < z$ (see Fig. 11.4). The effect is qualitatively a mirror image of that seen for the phase-lead filter previously discussed. $M(\omega)$

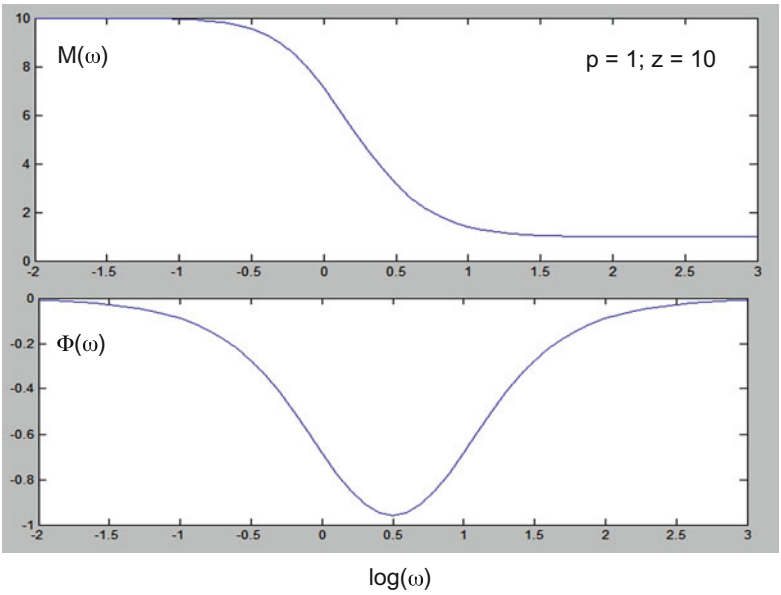


Fig. 11.4 Example of a first-order, low-pass, phase-lag filter

starts out large (at z/p) for ω near 0, decreases slowly until ω becomes comparable to p, decreases relatively quickly for ω between p and z, and then asymptotically approaches 1 for large ω. Meanwhile, $\Phi(\omega)$ starts out at 0, decreases slowly until ω becomes comparable to p, changes relatively quickly between p and z (going through a minimum at $\omega = (zp)^{1/2}$), and then asymptotically returns to 0 for large ω. This behavior is both low-pass and phase-lag, with the phase-lag behavior pertaining particularly for the region between p and z. In this way the filter behavior is qualitatively similar to the effects of a integral component (simple division by s).

A phase-lag filter acts like a PI control law, but when used in conjunction with a PI can improve the system stability. Since a phase-lag filter acts like a PI and a phase-lead filter acts like a PD, putting the two filters together should (and does) give you something with the characteristics of a PID, which is why a PID sometimes is called a lead-lag filter. Lastly, by reducing the phase-lag filter's high frequency gain you can improve the phase margin, and by increasing the low frequency gain you can obtain better error reduction characteristics.

Assuming that you like the frequency space behavior of a first order filter that you have in mind, you might reasonably ask how to realize it in time space. Let $v_i(t)$ and $v_o(t)$ be your input and output signals respectively, with corresponding Laplace transforms $V_i(s)$ and $V_o(s)$. Equation 11.1 implies

$$V_o(s)(s + p) = K V_i(s)(s + z) \tag{11.5}$$

for which the inverse Laplace transform yields

$$dv_o(t)/dt + pv_o(t) = K(dv_i(t)/dt + zv_i(t)) \tag{11.6a}$$

$$v_o(t) = [K(dv_i(t)/dt + zv_i(t) - dv_o(t)/dt]/p \tag{11.6b}$$

Note the inclusion of a $(dv_o(t)/dt)$ term on the right-hand side; to construct a filter of this form requires feedback of the derivative of the output back into the filter. This may require some iteration to achieve. Another variation would be to write Eq. 11.5 as

$$V_o(s)(1 + p/s) = K V_i(s)(1 + z/s) \tag{11.7}$$

for which the inverse Laplace transform yields

$$V_o(t) + p \int v_o dt = K(v_i(t) + \int v_i dt) \tag{11.8a}$$

$$V_o(t) = K(V_i(t) + z \int v_i dt) - p \int v_o dt \tag{11.8b}$$

One still needs feedback of the output signal into the filter, but it may prove more convenient to realize in this integral form.

Although the easiest to analyze, first order filters don't have a lock on the phase-lead (high-pass), or phase-lag (low-pass) behavior. One can adjust the filter behavior by increasing the order of the filter, adding extra zero and pole terms as needed. The HST solar array motion compensator, for example, was a sixth order filter. For either low-pass or high-pass filters, one can sharpen the cut-off/cut-on edge by use of progressively higher order terms. (Butterworth[1] is the name associated with low-pass filters that are maximally flat (for a given order) up to a selected cut-off frequency, and thereafter drop to zero very quickly.) One can combine low-pass and high-pass filters into a "notch" filter (one that suppresses signals within some selected frequency band), a "band-pass" filter (one that enhances a signal within a selected frequency band), or more complex variations on the theme. For example, if the satellite structure has certain characteristic frequencies at which it would naturally tend to ring if those frequencies were excited, one may want to modify one's preliminary PID control law design with a set of notch filters to suppress generation of reaction wheel torque commands near those frequencies to suppress excitation of such ringing. Alternatively, if an "outside" system occasionally imposes a disturbance on the satellite at specific frequencies (e.g., the HST solar arrays occasionally shaking the observatory body), one may want to add enhanced responsivity at those frequencies (a set of band-enhancing filters) to move such disturbances quickly from the body to the reaction wheels.

[1] S. Butterworth, On the Theory of Filter Amplifiers, Wireless Engineer, vol. 7, 1930, pp. 536-541.

Our discussion of control law and filter transfer functions has all been simplified via the use of Laplace transforms and the implied sensitivity to arbitrarily high frequencies. In practice, satellite sensors are read discretely and control law generated commands are issued discretely. The satellite is blind to, and unable to compensate for, disturbances with frequencies at and above $\sim 1/(2\pi\,\Delta t)$, where Δt is the cycle time for the attitude control system. Even at slightly lower frequencies, there will be lag in control law responsivity of order Δt (not to be confused with phase-lag discussed above). As much as filter order and the locations of various poles and zeroes, an appropriate selection of Δt must be considered part of control law design.

Chapter 12
Mission Characteristics

Our final chapter is largely a tabular listing (see Tables 12.1, 12.2, 12.3, 12.4, 12.5, 12.6, 12.7 and 12.8 at the end of the chapter) of the characteristics of four mission types, namely celestial-pointers, Earth-pointers, "fixed"-pointers, and survey missions. Probably the most relevant information (with respect to relating the way the spacecraft has been built and launched in order to satisfy its mission requirements) is provided in

1. Table 12.2: Pointing Behavior (really just another way of describing the mission categories)
2. Table 12.3: Pointing Accuracy (strongly correlates with mission expense)
3. Table 12.4: Orbit Characteristics (dictated by science needs or launch vehicle cost budget)
4. Table 12.5: Orbit Modeling Accuracy (typically most demanding for nadir-pointers such as LEO Earth-pointers that image the Earth)

But for folks who like a little prose along with the dry data, we've also included discussion in the following subsections on

1. how mission orbits are selected
2. how celestial-pointers compare with nadir-pointers (specifically, Earth-pointers)
3. how safemodes and rate-nulling modes work
4. how mission attitude is acquired.

12.1 Mission Orbit Selection

One of the earliest key decisions in mission design is selection of the spacecraft's orbit. Most of the GSFC missions are launched into one of three basic orbit types, Low Earth Orbit (LEO), Geosynchronous Earth Orbit (GEO), and Earth-Sun Lagrange

© Springer-Verlag London Ltd. 2017

H.L. Hallock et al., *ACS Without an Attitude*, NASA Monographs
in Systems and Software Engineering, DOI 10.1007/978-1-4471-7325-0_12

Point Orbit. With the exceptions of missions whose orbits are selected to satisfy special science needs, many GSFC celestial-pointers end up in near-circular LEO orbits having inclinations near 28.5°. That orbit is the "cheapest" you can get, from a launch vehicle cost standpoint, if you launch due-East from Kennedy Space Flight Center (KSC), whose latitude is 28.5°N. The altitude of the orbit is usually between about 500 and 600 km, high enough to get you away from the worst of the atmospheric drag, but low enough to avoid the Van Allen radiation belts and to be (well, to have been) serviceable by the Space Shuttle. The Hubble Space Telescope (HST), Compton Gamma-ray Observatory (CGRO), and Rossi X-ray Timing Explorer (RXTE) are good examples of spacecraft that fall in this category. The downside to the cheap LEO orbit is that the scientific instruments will often be blocked (or occulted) by the Earth, reducing operational efficiency. Also, the spacecraft will regularly cycle between orbit day and orbit night, which induces large temperature fluctuations; communications will suffer frequent outages even if the Tracking and Data Relay Satellite System (TDRSS) is utilized; and after a few years the spacecraft will need an orbit re-boost to continue its mission (or a controlled de-orbit to avoid hitting anything important, like your house).

To deal with the communications problem, the International Ultraviolet Explorer (IUE) was placed in a GEO elliptical orbit, which provided round-the-clock communications opportunities, thereby enabling realtime ground scheduling and operation, colloquially referred to as "joystick" operation. "Joystick" in this context means that IUE observations were scheduled within "block time", as is the case for ground-based observatories. That system allowed the astronomer to issue commands in realtime to perform science within the time block assigned to him, the only condition being that he return the spacecraft to a standard configuration at the end of his allotted time. GEO also can increase a spacecraft's operational lifetime (drag is very low) and places the spacecraft above the Van Allen belts. Sometimes GEO isn't far enough away. To avoid Earth shine and infrared radiation from Earth, the James Webb Space Telescope (JWST) will orbit the second Earth-Sun Lagrange point (L2) (more on that later), eliminating the occultation problem and ensuring a benign thermal environment. Typically, the orbit knowledge and control accuracies for these missions are not very tight; attitude accuracy requirements are the more important concerns.

By contrast, Earth-pointers (because their imaging devices follow their orbital "footprint") tend to be much concerned about the details of the orbit, and the accuracy with which the orbit is known and controlled. In order to maintain consistent lighting, temperature, and visual (vis-a-vis cloud cover) conditions, LEO Earth-pointers tend towards Sun-synchronous orbits, i.e., near-polar orbits, whose inclination and altitude are set to cause the orbital plane to follow the Sun's apparent motion. The Earth Observing System (EOS) series and Landsat series are good examples of imaging spacecraft whose Sun synchronous orbits ensure consistency between multiple observations taken over several repeat cycles. LEO altitudes also have the advantage of enabling smaller, cheaper optics (i.e., the closer you are to what you're photographing, the smaller the lens needs to be to obtain the same resolution). On the downside, as a Sun-synchronous orbit rotates relative to the (also) rotating Earth below, the time

between observations of the same target will typically be about a couple of weeks, not a good number if you want to watch the development of a storm front.

For those kinds of applications, a GEO orbit is a much better idea. At GEO, an Earth-pointer can image the same region of the Earth 24 h per day. And since it's hanging over the same point on the Earth's surface, communications can also be continuous. A good example of this kind of mission is the Geostationary Operational Environmental Satellite (GOES) series. Of course, a GOES orbits much further from the Earth's surface (over 30,000 km further) than a LEO Earth-pointer, so its optics need to be much more powerful and expensive, and it needs more delta-V (i.e., change in velocity) to get on-station. However, although a GEO Earth-pointer has an excellent view of the low latitudes (being in an equatorial orbit, i.e., having inclination zero), its view of the high latitudes will be pretty poor. So GEO is a great orbit for U.S. communications satellites (e.g., the TDRS series), but is a bad orbit for Russian communications satellites. Their communications spacecraft are placed in Molniya orbits, whose inclinations are selected to "freeze" perigee at a desirable location. Specifically, to service Russian users best, perigee is frozen in the southern hemisphere. The orbit then is made extremely elliptical so that most of the potential contact time is in the high northern latitudes. Other advantages are that the delta-V is lower than would be required for a circular orbit having the same apogee altitude, and the high eccentricity (for other missions) allows sampling of a wider range of altitudes. Geometry is also the name of the game for an Earth-pointer orbiting the first Earth-Sun Lagrange point (L1). A satellite at L1 can view a sunlit Earth 24 h per day. It also provides a unique geometry for viewing the action of sunlight on the Earth.

Perhaps Earth-pointers at L1 might more properly be listed under the third category of mission, the "fixed"-pointers. That's because unlike spacecraft in LEO or GEO orbits, Earth-pointers at L1 do not execute the characteristic 1 revolution per orbit (1 RPO) motion required to keep an imaging instrument pointed along the spacecraft's nadir vector. A 1 RPO motion for a LEO spacecraft means the attitude rotates with a period of between 96 and 100 min, while for a GEO the period is 24 h. But as viewed from inertial space, the line defined by the Sun, L1, and Earth completes a full rotation about the Sun over the course of a year. So for an Earth-pointer at L1 (ignoring the complication of the halo orbit), the attitude of the spacecraft only needs to rotate 360° over a one year period to stay aligned with the nadir vector, a sufficiently slow rotation that for control purposes we can consider the attitude to be fixed. The inverse of this situation is that for the Solar and Heliospheric Observatory (SOHO) spacecraft, which looked "up" at the Sun from L1 instead of "down" to Earth as the Triana mission was intended to. (Triana was cancelled near the end of the implementation phase, but came back rechristened as the Deep Space Climate Observatory (DSCOVR).) Another "Sun worshipper" was the Solar Maximum Mission (SMM), which occupied a LEO, mid-inclination, near-circular orbit. Also in the fixed-pointing category is the ST-5 mission, for which three spin-stabilized spacecraft were launched into a highly elliptical, high altitude orbit to enable sampling of the entire geomagnetic field over the mission lifetime.

The last of the four categories, the survey mission, is kind of an oddball one for which we'll only mention one example, the Wilkinson Microwave Anisotropy Probe (WMAP). WMAP occupied an L2 halo orbit in order to take advantage of L2's benign thermal environment. Its attitude motion was a combination of spin-axis rotation and spin-axis precession designed to sample the microwave background of the entire celestial sphere at least twice over the course of the mission's minimum required lifetime.

Having touched on a fair number of different orbit types in the course of discussing why the different types of missions end up in these orbits, let's do a little clean-up on a special, but increasingly popular, orbit type given short shrift in the previous discussions, halo orbits about Lagrange points. Lagrange points are points of stable or unstable equilibrium created by the gravity of two celestial objects acting on a third object whose orbit is defined relative to the other two. The way the physics works is that the gravitational attraction of the two bodies combines to cancel the centrifugal acceleration "felt" by the third body in the rotating reference frame in which the third body appears stationary. Alternately, you can say that the object's gravitational potential energy exactly is canceled by its kinetic energy. For all the Lagrange point missions discussed in this chapter, the two celestial objects are the Sun and the Earth, while the third object is the spacecraft.

For any given pair of celestial objects, there will be five Lagrange points (see Fig. 12.1). Two of the points are located off-axis (L4 and L5). These points exhibit stable equilibrium, i.e., a test mass placed at L4 or L5 subjected to a small perturbative force will tend to return to its point of initial placement. For the Sun-Jupiter system, the off-axis Lagrange points are natural "clumping" spots for the Trojan asteroids. The other three points (L1, L2, L3) are located on the line connecting the two celestial bodies. These points exhibit unstable equilibrium, i.e., a test mass placed at L1, L2, or L3 subjected to a small perturbative force will tend to follow the force away from

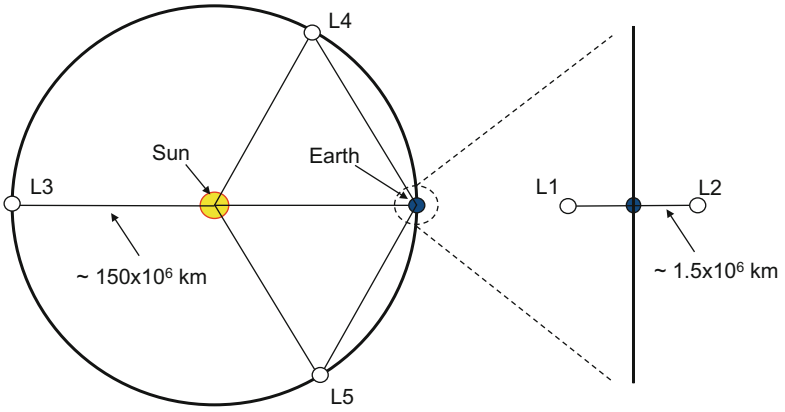


Fig. 12.1 The five Earth/Sun Lagrange Points. Note that the distances from Earth to L1 and L2 are only about 1% of the distance from the Earth to the Sun

its point of initial placement. For the Earth-Sun system, L1 and L2 are located about 1.5 million km from the Earth, so the Moon is close enough to both points to exert a significant perturbative force on any spacecraft located there.

Since the axial symmetric Lagrange points are unstable, you can't simply place a spacecraft on the point and expect it to stay there. Instead, you place the spacecraft in an orbit about the Lagrange point (which in that respect plays the role of a small, imaginary central body), performing occasional, regular stationkeeping maneuvers to maintain the orbit. This is the *halo orbit* referred to previously, which is an elliptical orbit that usually is tilted out of the ecliptic plane in order to facilitate communications with a ground station on the Earth. Typically, the halo orbital period is 6 months, with stationkeeping performed every 3 months. Although 1.5 million km seems pretty far away, it's still close enough that communications with spacecraft at L1 or L2 does not constitute a major challenge. And, as mentioned previously, L1 and L2 offer many nice geometric advantages, including benign thermal environments, lack of occultation issues, and (at L1) an uninterrupted view of a sunlit Earth.

12.2 Celestial-Pointers Versus Earth-Pointers

Since most of GSFC's missions nowadays fall into the celestial-pointer or Earth-pointer categories, we'll spend some time comparing the two mission types for a host of non-orbit related topics, including attitude behavior, attitude sensor and actuator suites, and acquisition and safemode control modes. The trademark attitude behavior of a celestial pointer is to slew to a target, take observations, and then slew to the next target, which is why they are sometimes called "point-and-shoot" missions. Depending on how much time you need to spend photon-counting on target (a function of target intensity and scientific instrument design) and how large an area you're viewing at that target attitude, the slewing frequency can be as high as 20 times per day or as low as once every two weeks. But in practice, the attitude behavior of a celestial-pointer is usually much more complicated than a simple-minded point-and-shoot. For example, you don't really point at a celestial target; you actually point at the direction on the celestial sphere from which light from the target appears to be arriving. That direction may be a function of velocity aberration, parallax, and science instrument characteristics such as distortion. Also, because the target direction may not be known with high enough accuracy to permit immediate post-slew observation, the spacecraft may instead slew to the target's near-neighborhood and then execute attitude scans or "dithers" in order to locate the target precisely enough to permit target acquisition to proceed. Another complication (primarily a scheduling issue) is that the target's direction is defined relative to the GCI reference frame, whose z-axis is the Earth's spin axis. Unfortunately, the Earth's spin axis is not inertially fixed, so the target's coordinates are time dependent. Since the movement of the spin axis is very slow (precesses about 50 arcsec/year and nutates about 0.5 arcsec/year), this doesn't mean you have to adjust the pointing while viewing the target. It just means that you have to define all quantities that "live" in the GCI frame, such as space-

craft pointing, target coordinates, guide star information, and ephemeris data, to be consistent with each other. Usually, the approach taken is to use the target catalog and guide star catalog information as the standard, which nowadays means defining everything with respect to how the GCI reference frame was oriented relative to the "fixed" stars at the start of year 2000, taking into account the precession but not the nutation.

Whereas celestial-pointers are usually trying to keep their attitude fixed, Earth-pointers by contrast usually keep changing their attitude in order to keep imaging the Earth beneath them. This produces their trademark 1 RPO rotation of the bore-sight of spacecraft's imaging instrument. For LEO orbits, the rotation rate is around 4 deg/min, while for GEO orbits the rate is 15 deg/hr. However, the attitude behavior looks a lot more like a celestial-pointer's if you simply switch to a more appropriate frame of reference. For a perfectly circular orbit, a natural set of orthogonal axes to pick places the spacecraft x-axis (the roll axis) along its velocity, the y-axis (the pitch axis) along negative orbit normal, and the z-axis (the yaw axis) along the nadir vector. If the orbit is not perfectly circular, everything still works if you just set the x-axis equal to the cross-product of y- and z-unit vectors. In this non-inertial reference frame, an imaging instrument mounted on the body of the spacecraft and aligned with the spacecraft z-axis will remain nadir-pointing as the spacecraft orbits the Earth. For GEO Earth-pointers that are spin-stabilized, the imaging instrument often is mounted on a despun section that tracks the Earth, just as the spacecraft itself did in the previous example. Although spin-stabilized spacecraft were very popular in the past (e.g., GOES-1 through GOES-6), the extremely high electrical power demands of large modern spacecraft mitigate against spinners because body-mounted solar cells on spinners are so much less efficient than an articulated solar array. Some of the smaller, cheaper recent spacecraft, though, have found spin-stabilization to constitute an effective trade (e.g., ST-5).

Whether your spacecraft is an Earth-pointer or celestial-pointer, in order to achieve the attitude profiles described above, you need to be able to measure the spacecraft's attitude (or attitude change) to see if it's what you want, and then must be able to change the spacecraft's attitude if it isn't. Often you'll see almost the same kinds of attitude sensors and actuators on celestial-pointers and Earth-pointers. The choice of hardware usually relates more to the pointing accuracy requirements and the orbit occupied by the spacecraft than to which way the spacecraft is looking (up or down). But there are a few general differences that usually hold true. Since a celestial-pointer looks up, not down, it doesn't make much sense to fly an Earth sensor on a celestial-pointer to support science operations directly, although for some missions (the Chandra X-Ray Observatory (CXO), for example) they have been included to facilitate initial orbital verification. On the other hand, since what an Earth-pointer does for its science is look at the Earth, flying an Earth sensor used to be mandatory. Nowadays, star tracker hardware costs have dropped far enough that they are used along with spacecraft ephemerides to define for the FSW where the Earth is.

Similarities between the missions are much more common than the differences – a commonality that was exploited for significant cost savings when the RXTE and Tropical Rainfall Measuring Mission (TRMM) spacecraft were designed. Since

both types of missions are dead if they don't have electrical power, both fly some sort of Sun sensor, if for no other reason than to be able to find the Sun during safemode, and thereby be able to reorient to a power-positive attitude. Medium to large missions of both types tend to utilize gyros, tend to fly magnetometers and torquer bars when performing their science at LEO altitudes, tend to use reaction wheels for primary attitude control, and include propulsion systems if maintaining a specific orbit to a reasonable degree of accuracy is needed, or if their orbit is so high that thrusters must be used for momentum dumping due to the ineffectiveness of torquer bars in weak geomagnetic fields. Clearly orbit stationkeeping is much more important for an Earth-pointer, where a specific footprint and target distance must be maintained, than for a celestial-pointer. Also, the hydrazine that fuels the most common propulsion systems can be extremely corrosive to some celestial-pointers' optics. HST got around this problem by using the Space Shuttle for orbit re-boosts, which were occasionally needed because of HST's LEO orbit. For those reasons, most Earth-pointers fly propulsion systems, while most LEO celestial-pointers do not. On the other hand, celestial pointers tend to have the most demanding pointing accuracy requirements because of the nature of the science targets they're observing (e.g., X-ray experiments such as the High Energy Astronomy Observatory-2 (HEAO-2), RXTE, and CXO). So you almost always will see star trackers or fine error sensors (FESs) included in a celestial-pointer's attitude sensor suite.

12.3 Safemodes

Now that we've seen what sort of orbits are typical of the various missions, and what sort of attitude sensor and actuator combinations are employed to support their science operations, let's look at the flavors of safemode (i.e., a robust control mode used in response to onboard anomalies) flown on celestial-pointers and Earth-pointers to help ensure they're on-orbit long enough to accomplish their respective tasks. The number and complexity of safemodes implemented is largely a function of the mission's expense and the uniqueness of its science data. Missions that are cheap and whose science data is fairly repeatable (e.g., a spacecraft that measures the geomagnetic field over a period of years) typically just fly one or two safemodes. The safemodes themselves tend to be simple in construction, and modest in goal. The usual philosophy is just keep the bird alive. Don't worry about losing some science. Don't try to "fly through" the anomaly. Just go into safemode, let the ground analyze and solve the problem, then restart science again when you're sure everything's okay. These sorts of bargain basement missions rarely are designed to handle more than one show-stopping failure, i.e. they don't provide protection against more than single-point failures.

Missions that are expensive and whose science data are unique often have the assets to implement a layered series of safemodes. The mildest level of response to an anomaly may not even be a safemode entry. There may instead be an autonomous hardware reconfiguration designed to allow the spacecraft to continue making its

science observations. If the spacecraft can't fly through the anomaly, there might be a transition to a fairly benign level of safemode that expedites resumption of nominal operations after the ground fixes the problem. At the lowest level of safemode, there will still be a safety net mode whose sole intent is just to keep the spacecraft alive. Depending on the nature of the hardware anomaly that has placed the spacecraft in these dire straits, there may be several safety net modes, each tailored to a specific problem (e.g., zero-gyro for multiple gyro failures, spin-stabilized for multiple wheel failures, and hardware safemode for processor failure). Some of these safemodes may be designed to address multiple problems, i.e., they may selectively protect against more than single-point failures.

For celestial-pointers, two popular safemodes are *inertial hold* and *sunpoint*; on much rarer occasions, you run into a celestial-pointer that utilizes *spin-stabilized* safing. Inertial hold typically uses only reaction wheels and gyros to maintain the spacecraft attitude at the pointing associated with the onset of the anomaly. Since celestial-pointers typically follow a carefully orchestrated schedule, holding the current attitude (which, since it was a "legal" observing attitude, is probably a safe one to be at) improves the chances of being able to restart the interrupted science if it doesn't take too long to fix the problem, or at least facilitates intercepting the original mission schedule. On one mission, RXTE, a more sophisticated version of inertial hold was implemented called *stellar hold*. During early operations before the spacecraft was calibrated, RXTE used its stellar hold mode to process star tracker position data without trying to identify the stars. RXTE dates back to a post-FHST era but pre-quaternion star tracker era where ground-computed attitudes were still needed to initialize the onboard attitude quaternion.

Nowadays, advanced star trackers with "Lost-in-Space" capability ensure that the spacecraft can always identify stars and compute their own attitude autonomously. Although RXTE could not determine a fine attitude at this point, the Kalman filter could be run with the processed star tracker data and gyro data in order to calibrate the gyro drift bias inflight. That way, more reliable gyro data became available at an earlier stage of the mission, enabling more accurate pointing. And since RXTE's high-gain antennas (HGAs) could be pointed using the FSW's coarse attitude determination algorithm (Triad) run with Sun sensor and magnetometer data, the initial unavailability of a fine attitude did not deter use of the higher speed data links. Nowadays, quaternion star trackers make this stellar hold mode superfluous. And the implementation of inertial hold itself has become less frequent as missions strive to reduce costs by reducing the number of spacecraft control modes.

Just about every mission type utilizes one or more forms of sunpoint safe mode. In its most complicated form, sunpoint run in the main processor can utilize sun sensor data (from either a fine or coarse sun sensor), with or without gyro data. In its simplest form, sunpoint-based only on coarse sun sensor data can be run in analog mode in a backup processor, although nowadays (to reduce costs) a separate Attitude Control Electronics (ACE) box may be omitted. In any form, sunpoint is designed to keep the spacecraft in a power-positive condition while satisfying system thermal constraints. Since sunpoint also keeps a spacecraft axis pointed at the Sun (at least while the spacecraft is in orbit day), communications should at least be predictable,

if not facilitated. There is no concern regarding how quickly transition from sunpoint back to a science operational mode can be executed. The objective is just to keep the spacecraft alive, maintaining autonomous control indefinitely.

While sunpoint without gyros was designed as a response to a multiple gyro failure problem, HST's spin-stabilized safemode was intended to deal with multiple reaction wheel failures. Spin-stabilized mode uses torque bars and gyros to keep the spacecraft in a power-positive configuration that can also support low data rate communication with the ground. Since HST is (was) Shuttle-serviceable, spinning the spacecraft about one of its primary moment axes supplies gyroscopic stability to help keep the spacecraft alive until a Shuttle rescue could be organized. Usually, even for asset-rich missions, there is not enough funding to provide a spin-stabilized control mode in the flight software prior to launch. But it may be developed after launch in response to hardware failures experienced on-orbit, or if the mission is long-lived enough (and well-staffed enough) to allow its development by a long-term flight software maintenance team.

There are two other control mode types, rate-null mode and B-dot, that are neither operational modes nor safemodes. They are instead designed to reduce the unusually high body rates that may be experienced on separation from the launch vehicle (also referred to as high tip-off rates), or suffered as a result of a major propulsion system failure, such as a stuck-on thruster. The rate-null mode employed on both celestial-pointers and Earth-pointers uses reaction wheels and gyros to null moderately high body rates, typically up to the sizes of those arising from the gyro drift bias errors present at launch. Since nearly all GSFC Earth-pointers fly a propulsion system, while many GSFC LEO celestial-pointers do not, one difference between the two mission types is that Earth-pointers may also allow for a contingency thruster-based rate-null mode as well. B-dot is almost exclusively for LEO missions, either celestial-pointer or Earth-pointer. It uses torque bars and magnetometers to reduce anomalously high body rates (incurred either on separation, or following a thruster malfunction or venting) steadily, but slowly. The B-dot control law works by assuming that any change in the magnetic field measured by the magnetometers is due to body rates. Since the geomagnetic field is actually changing as the spacecraft flies through it, B-dot does not reduce the body rates to zero, but instead creates a steady-state situation where the spacecraft follows the geomagnetic field, resulting in a twice-orbital frequency vehicle rotation rate. B-dot was created and first utilized to place the SMM spacecraft in a controlled rotation state that kept it alive until a Shuttle rescue mission could be scheduled, and facilitated its capture once a Shuttle rendezvous was achieved. Following the SMM experience, B-dot was included for many years as a standard control mode on most GSFC LEO missions, often packaged with rate-null as a single rate-reducing mode.

12.4 Mission Attitude Acquisition

That covers the safemodes and the special modes designed to deal with the high body rate conditions that can be experienced right after tip-off. These control modes, along with the science operations modes, are usually all the tools a celestial-pointer needs to be successful on-orbit. That's because once a celestial-pointer has been calibrated, it can point extremely accurately, and the locations of the point-sized targets it needs to point to are known very accurately as well. The situation with an Earth-pointer used to be a bit more difficult. After separation (or after inflight anomalies producing large attitude errors), an Earth-pointer had to "find" the Earth, which (given the way a LEO-tuned Earth sensor works) was not necessarily a simple task, and had to get the Earth sensor properly locked on the Earth to commence science observations. To get all this done, LEO Earth-pointers (such as EOS and Landsat) use a three-step process. First the Earth sensor needs to find the Earth, which is accomplished by using coarse data from the Earth sensor quadrants to point the spacecraft yaw axis at the Earth. Data for control of body rates (at this step, called Earth search mode) comes entirely from the gyros. The moderate-level attitude errors remaining after Earth search mode has successfully acquired the Earth are reduced in the next step, local vertical acquisition (LVA) mode. The LVA control law (a PD type on Landsat-7) uses data from both the Earth sensor and gyros to point the yaw axis at the Earth, and uses just Earth sensor data to control attitude errors about the other two axes (roll and pitch). Rates about all three axes are controlled via data collected just from the gyros. The final step is yaw gyro control (YGC), which deals with the small pointing errors left over after a successful LVA. The YGC control law (a PID on Landsat-7) maintains yaw-axis pointing towards the Earth, and uses both Earth sensor and gyro data to measure attitude errors and rate errors. Of course nowadays, if the Earth-pointer is flying an advanced star tracker, the tracker provides the spacecraft attitude, the spacecraft ephemeris tells you where the Earth is, and the FSW can easily compute the Earth-pointing attitude.

12.5 Spacecraft Comparisons

Table 12.1 Spacecraft comparisons: launch dates

- Celestial Pointers
  - JWST: 2018 (planned)
  - RXTE: 1995
  - HST: 1990
  - CGRO: 1990
- Earth Pointers
  - DSCOVR (Triana): 2015
  - EOS-AM: 2000
  - Landsat-7: 1999
  - GOES D-F: Early 1980s
- Fixed Pointers
  - ST-5: 2004
  - SMM: 1980
- Sky Survey
  - WMAP:2001

Table 12.2 Spacecraft comparisons: pointing behavior

- Celestial Pointers
  - JWST: small number of slews per day
  - RXTE: slews up to 20 times per day
  - HST: slews several times per day
  - CGRO: slewed after about 2 weeks
- Earth Pointers
  - DSCOVR (Triana): Earth pointing with small re-orientations
  - EOS-AM: 1 RPO pitch option for Earth pointing
  - Landsat-7: 1 RPO pitch option for Earth pointing
  - GOES D-F: spin axis fixed at negative orbit normal
- Fixed Pointers
  - ST-5: spin axis fixed normal to sunline
  - SMM: Sun pointing with small re-orientations
- Sky Survey
  - WMAP: spins at 0.5 rpm; precesses at 1 rev/hr;
    half cone angle of 22.5°;
    maps celestial sphere twice per year

Table 12.3 Spacecraft comparisons: pointing accuracy

- Celestial Pointers
  - JWST: $\sim$ 0.005 arcsec (relative to guide stars)
  - RXTE: control: 42 arcsec pitch and yaw; 60 arcsec roll (3σ);
    knowledge: 8 arcsec pitch and yaw; 44 arcsec roll
  - HST: $\sim$ 0.005 arcsec (relative to guide stars)
  - CGRO: control: 0.28 deg/axis; determination: 0.024 deg/axis
- Earth Pointers
  - DSCOVR (Triana): control: 0.05° pointing; 1° roll
  - EOS-AM: required control: 150 arcsec/axis;
    required knowledge: 90 arcsec/axis;
    expected knowledge: $\sim$10 arcsec/axis
  - Landsat-7: control: 180 arcsec/axis;
    knowledge: 135 arcsec/axis
  - GOES D-F: required accuracy for apogee motor firing: 2°;
    required accuracy in drift orbit: 0.5°;
    accuracy using image data: 0.1°
- Fixed Pointers
  - ST-5: controls spin axis to within 5° of ecliptic pole;
    knowledge of spin axis: 1°;
    knowledge of spin phase: 1°
  - SMM: knowledge: 0.5 arcsec pitch and yaw; 0.1° roll
- Sky Survey
  - WMAP: knowledge: 1.3 arcmin (RSS, 3σ)

Table 12.4 Spacecraft comparisons: orbit characteristics

- Celestial Pointers
  - JWST: L2 elliptical halo orbit
  - RXTE: LEO, 23° inclination, near circular
  - HST: LEO, 28.5° inclination, near circular
  - CGRO: LEO, 28.5° inclination, near circular
- Earth Pointers
  - DSCOVR (Triana): L1 elliptical halo orbit
  - EOS-AM: Sun synchronous LEO, near polar, near circular
  - Landsat-7: Sun synchronous LEO, near polar, near circular
  - GOES D-F: Geostationary, equatorial, near-circular
- Fixed Pointers
  - ST-5: 200 km $\times$ 35,790 km elliptical orbits;
    3 spacecraft in constellation
  - SMM: LEO; 28.5° inclination, near circular
- Sky Survey
  - WMAP: L2 elliptical halo orbit

Table 12.5 Spacecraft comparisons: orbit modeling accuracy requirements (3σ)

- Celestial Pointers
  - JWST: requirements likely to be loose
  - RXTE: 30 km over 1 day
  - HST: 10 km over 1 day (actually good to about 2 km)
  - CGRO: 3 km
- Earth Pointers
  - DSCOVR (Triana): 100 km for Earth targeting
  - EOS-AM: 150 m/axis Ground track repeatability (orbit control) 20 km
  - Landsat-7: 20 m / axis
  - GOES D-F: no OBC model
- Fixed Pointers
  - ST-5: as needed to maintain spacing between spacecraft
- Sky Survey
  - WMAP: 7500 km

Table 12.6 Spacecraft comparisons: attitude sensors

- Celestial Pointers
  - JWST: FGS, CCD ST, IRU, CSS
  - RXTE: CCD ST, IRU, DSS, TAM, CSS
  - HST: FGS, FHST, RGA, TAM, CSS
  - CGRO: FHST, IRU, FSS, TAM, CSS
- Earth Pointers
  - DSCOVR (Triana): CCD ST, MIMU, DSS, CSS
  - EOS-AM: SSST, IRU, Earth Scanner Assembly (ESA), FSS, TAM, CSS
  - Landsat-7: CSA, IMU, ESA, TAM, CSS
  - GOES D-F: 5° FOV Earth sensor, 2-slit Sun sensor
- Fixed Pointers
  - ST-5: 2-slit Sun sensor, TAM
  - SMM: FHST, IRU, FSS, TAM, CSS
- Sky Survey
  - WMAP: CCD ST, IRU, DSS, CSS

Table 12.7 Spacecraft Comparisons: attitude actuators

- Celestial Pointers
  - JWST: RWA, RCA
  - RXTE: RWA, MT
  - HST: RWA, MT
  - CGRO: RWA, MT, RCA
- Earth Pointers
  - DSCOVR (Triana): RWA, RCA
  - EOS-AM: RWA, MT, RCA
  - Landsat-7: RWA, MT, RCA
  - GOES D-F: RCA, passive nutation damper
- Fixed Pointers
  - ST-5:RCA, passive nutation damper
  - SMM: RWA, MT
- Sky Survey
  - WMAP: RWA, RCA

Table 12.8 Spacecraft comparisons: telemetry capture

- Celestial Pointers
  - JWST: ground station
  - RXTE: TDRSS
  - HST: TDRSS
  - CGRO: TDRSS
- Earth Pointers
  - DSCOVR (Triana): ground stations
  - EOS-AM: TDRSS and ground stations
  - Landsat-7: ground stations
  - GOES D-F: ground stations
- Fixed Pointers
  - ST-5: ground stations
  - SMM: ground stations
- Sky Survey
  - WMAP: ground stations

Appendix A
Time Measurement Systems

There are many different systems for measuring time currently in use in spacecraft attitude work - the least convenient of which is the Gregorian calendar system common in civil use (years, months, days, hours, minutes, and seconds). This civil system, while familiar, has the drawback of making it difficult to find the time difference between two different dates and times. It also makes it inconvenient to create plots that include a time axis, since the months have different lengths.

One commonly used alternative method for measuring dates is the *day of year* (DOY). In this system, one simply counts the number of days elapsed during the current calendar year, with January 1 being DOY 1, and December 31 being DOY 365 (or 366 during leap year).

The day of year is a convenient scale when working with time intervals on the order of a few weeks to a few months, but it has the disadvantage that it rolls over to 1 at the beginning of each calendar year. For longer time periods covering several years, a more convenient method for measuring dates is the *Julian Day* (JD). The Julian Day is the total number of days elapsed since *noon* on January 1, 4713 BC (this date being chosen for historical reasons). As an example, January 1, 2000 at time 00:00 is Julian Day 2451544.5. (The fraction 0.5 occurs because each new Julian Day starts at noon.) Julian Days are very convenient, since they provide a continuous, uniform time scale, and make it simple to compute the time elapsed between two dates; one simply computes the difference between the Julian Days. Algorithms and tables are available for converting dates on the standard Gregorian calendar to the corresponding Julian Day and back. One may add a day fraction to the Julian Day in order to include the time of day.

A variation of the Julian Day is the *Modified Julian Day* (MJD), which is the Julian Day minus 2400000.5. The idea with the Modified Julian Day is to discard the first couple of significant digits, which rarely change: JD 2400000 was in 1858, and JD 2500000 won't be until the year 2132. Also notice that the Modified Julian Day subtracts an extra fraction of 0.5 day from the Julian Day, so that the Modified Julian Day begins at *midnight*, just as dates in our civil calendar do. The Modified Julian Day was 0 at midnight at the beginning of November 17, 1858.

© Springer-Verlag London Ltd. 2017

H.L. Hallock et al., *ACS Without an Attitude*, NASA Monographs
in Systems and Software Engineering, DOI 10.1007/978-1-4471-7325-0

Universal Time (UT) is a term that is sometimes used informally to mean the same thing as *Greenwich Mean Time* (GMT), i.e. civil time in Greenwich, England. More precisely, though, it refers to one of several different time scales based on the rotation of the Earth. The Universal Time scale called UT0 is a simple time scale based on the rotation of the Earth relative to the distant stars, and depends on the location of the observer on the Earth. Time scale UT1 is UT0 corrected for shifts in the longitude of the observer due to polar motion ("Chandler wobble") and is independent of the observer. Another Universal Time scale called UT2 includes additional corrections, but is seldom used.

International Atomic Time (TAI) is time based on cesium (atomic) clocks, rather than on the rotation of the Earth. Since the Earth rotates at an irregular rate, cesium clocks are a more reliable time standard than Earth rotation, and are the basis for all precise time measurements. The TAI time scale is defined so that TAI exactly equaled UT1 at the beginning of 1958; since then the two time scales have drifted apart because of the Earth's irregular rotation.

Terrestrial Time (TT), formerly known as *Terrestrial Dynamical Time* (TDT), is also based on cesium clocks, and differs from TAI only by a constant:

$$TT = TAI + 32.184\,\text{s} \tag{A.1}$$

The 32.184 s offset is included to make TT continuous with an earlier time scale called *Ephemeris Time* (ET). At the beginning of 1958, ET was 32.184 s ahead of UT1 (and therefore also 32.184 s ahead of TAI, since TAI and UT1 were equal at that instant), so this definition makes TT continuous with ET.

Barycentric Dynamical Time (TDB) is a time scale that includes relativistic corrections due to the motion of the Earth in the gravitational potential of the solar system. TDB oscillates about TT with an amplitude of about 1.6 ms and a period of one year.

An important aspect of TDB is that it appears in the definition of the astronomical epoch called *J2000*, which is an instant in time used as a convenient reference point for many astronomical calculations. Epoch J2000 is defined to be the instant of January 1, 2000, at 12:00:00 TDB. Since finding the difference between a given time and epoch J2000 requires that both times be measured on the same scale, it is useful to know the time of the instant of epoch J2000 in several time scales, shown here:

- 12:00:00.000 TDB
- 11:59:27.816 TAI
- 11:58:55.816 UTC Julian Day 2451545.0

Coordinated Universal Time (UTC) is the time scale used as the basis for our everyday civil time (corrected to the local time zone). UTC is a compromise between the Universal Time scales based on Earth rotation and time scales based on the more precise cesium clocks. Due to the tidal drag on the oceans due to the Moon, the Earth's rotation is gradually slowing; this causes the Earth's rotation to drift with respect to time kept by cesium clocks. In order to have a time scale that has the precision of an atomic clock, *and* is also in step with the Earth's rotation, UTC

introduces extra *leap seconds* at irregular intervals, intended to keep UTC to within 0.9 s of UT1. When the decision is made to introduce a leap second, an extra second is added at the end of a day (generally a June 30 or December 31), which has the same effect as setting the clock *back* one second to allow the Earth's rotation to catch up. During a leap second, a clock will read 23:59:60 before rolling over to 00:00:00 of the following day. Leap seconds are introduced at irregular intervals as needed, but on average occur about every 18 months. Announcements of upcoming leap seconds are made by the International Earth Rotation Service a few months in advance of their occurrence.

UTC is offset from TAI by the number of leap seconds LS that have occurred up to that point:

$$UTC = TAI - LS \tag{A.2}$$

UTC has the advantage of having both the precision of cesium clocks and of being in step with the Earth's rotation, but it has the disadvantage of requiring the use of a table of leap seconds to convert to other time scales. Also, since variations in the Earth's rotation rate are unpredictable, it is not possible to accurately express times in the UTC time scale more than a few months in the future: there is no way of knowing how many leap seconds may occur in the future, or when exactly they will occur. For that reason, it is generally better to use some other time scale such as TAI to represent times of future events (ephemeris predictions, for example).

Appendix B
Variation on Deriving the Kalman Gain

This appendix provides an alternative derivation of the Kalman equations, alternative relative to the standard approach of minimizing the trace of the covariance matrix, $\boldsymbol{P}_X$ (see, e.g., Gelb[1]). For whatever reason, one of us has always been most comfortable with least-squares derivations that begin with a loss function defined as a sum of normalized residuals. The following shows an approach to deriving the Kalman equations based on minimizing such a loss function. The target audience for this appendix is really folks comfortable with least-squares fitting, the use of the method of Lagrange multipliers to impose constraints, and matrix inversion, i.e., a bit beyond the target audience for the bulk of this book. We're including it for completeness, and because to the best of our knowledge it is not available elsewhere.

Let's suppose we've got a state vector solution $\mathbf{X}_0$, and its associated covariance $\boldsymbol{P}_{X_0}$, obtained though whatever mechanism (e.g., a standard batch least-squares based on minimization of a sum-of-weighted-residuals loss function). We can pose the question of how $\mathbf{X}$ would change from $\mathbf{X}_0$ given some hypothetical ideal (noise-free) measurement $\mathbf{Z}_I$. This is the standard least squares problem subject to constraints, i.e., minimize

$$L(\mathbf{X}) = [\mathbf{X} - \mathbf{X}_0]^T \, \boldsymbol{P}_0^{-1} \, [\mathbf{X} - \mathbf{X}_0] \tag{B.1}$$

subject to the constraint set

$$\mathbf{X} - \boldsymbol{H}\mathbf{Z}_I = \mathbf{0} \tag{B.2}$$

You can do this using the method of *Lagrange multipliers*,[2] which produces the matrix equation

$$\begin{pmatrix} \boldsymbol{P}\mathbf{X}_0^{-1} & \boldsymbol{H}^T \\ \boldsymbol{H} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} \boldsymbol{P}_{X_0}^{-1}\mathbf{X}_0 \\ \mathbf{Z}_I \end{pmatrix} \tag{B.3}$$

[1] Applied Optimal Estimation, MIT Press, 1974, Arthur Gelb et al.

[2] If you're not familiar with the method of Lagrange multipliers, see e.g., the Wikipedia article at http://en.wikipedia.org/wiki/Lagrange_multipliers.

© Springer-Verlag London Ltd. 2017

H.L. Hallock et al., *ACS Without an Attitude*, NASA Monographs
in Systems and Software Engineering, DOI 10.1007/978-1-4471-7325-0

where $\boldsymbol{\lambda}$ is the set of Lagrange multipliers. The solution is

$$\mathbf{A} = \boldsymbol{B}\mathbf{C} \tag{B.4}$$

where

$$\mathbf{A} = \begin{pmatrix} \mathbf{X} \\ \boldsymbol{\lambda} \end{pmatrix}$$

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{P}_{X_0} - \boldsymbol{P}_{X_0}\boldsymbol{H}^T\left[\boldsymbol{H}\boldsymbol{P}_{X_0}\boldsymbol{H}^T\right]^{-1}\boldsymbol{H}\boldsymbol{P}_{X_0} & \boldsymbol{P}_{X_0}\boldsymbol{H}^T\left[\boldsymbol{H}\boldsymbol{P}_{X_0}\boldsymbol{H}^T\right]^{-1} \\ \left[\boldsymbol{H}\boldsymbol{P}_{X_0}\boldsymbol{H}^T\right]^{-1}\boldsymbol{H}\boldsymbol{P}_{X_0} & -\left[\boldsymbol{H}\boldsymbol{P}_{X_0}\boldsymbol{H}^T\right]^{-1} \end{pmatrix}$$

$$\mathbf{C} = \begin{pmatrix} \boldsymbol{P}_{X_0}^{-1}\mathbf{X}_0 \\ \mathbf{Z}_I \end{pmatrix}$$

This yields for $\mathbf{X}$:

$$\mathbf{X} = [\boldsymbol{I} - \boldsymbol{K}\boldsymbol{H}]\,\mathbf{X}_0 + \boldsymbol{K}\mathbf{Z}_I = \mathbf{X}_0 + \boldsymbol{K}\,[\mathbf{Z}_I - \mathbf{Z}_0] \tag{B.5}$$

$$\boldsymbol{K} \equiv \boldsymbol{P}_{X_0}\boldsymbol{H}^T\left[\boldsymbol{H}\boldsymbol{P}_{X_0}\boldsymbol{H}^T\right]^{-1} = \boldsymbol{P}_{X_0}\boldsymbol{H}^T\boldsymbol{P}_{Z_0}^{-1} \tag{B.6}$$

$$(\mathbf{Z}_0,\, \boldsymbol{P}_{Z_0}) \equiv (\boldsymbol{H}\mathbf{X}_0,\, \boldsymbol{H}\boldsymbol{P}_{X_0}\boldsymbol{H}^T) \tag{B.7}$$

This shows how $\mathbf{X}$ varies as a function of some hypothetical ideal measurement. Now we imagine introducing an actual measurement $\mathbf{Z}_m$ with covariance $\boldsymbol{P}_{Z_m}$. We can ask what the optimal estimate is for $\mathbf{Z}$ given measurements $(\mathbf{Z}_m, \boldsymbol{P}_{Z_m})$ and $(\mathbf{X}_0, \boldsymbol{P}_{X_0})$, of which the latter projects to $(\mathbf{Z}_0, \boldsymbol{P}_{Z_0})$ in $\mathbf{Z}$ space. The optimal estimate minimizes the loss function

$$L(\mathbf{Z}) = [\mathbf{Z} - \mathbf{Z}_0]^T\,\boldsymbol{P}_{Z_0}^{-1}\,[\mathbf{Z} - \mathbf{Z}_0] + [\mathbf{Z} - \mathbf{Z}_m]^T\,\boldsymbol{P}_{Z_m}^{-1}\,[\mathbf{Z} - \mathbf{Z}_m] \tag{B.8}$$

This yields

$$\mathbf{Z} = \left[\boldsymbol{P}_{Z_0}^{-1} + \boldsymbol{P}_{Z_m}^{-1}\right]^{-1}(\boldsymbol{P}_{Z_0}^{-1}\mathbf{Z}_0 + \boldsymbol{P}_{Z_m}^{-1}\mathbf{Z}_m) \tag{B.9a}$$

$$= \boldsymbol{P}_{Z_0}\boldsymbol{P}_{Z_0}^{-1}\left[\boldsymbol{P}_{Z_0}^{-1} + \boldsymbol{P}_{Z_m}^{-1}\right]^{-1}(\boldsymbol{P}_{Z_m}^{-1}\boldsymbol{P}_{Z_m}\boldsymbol{P}_{Z_0}^{-1}\mathbf{Z}_0 + \boldsymbol{P}_{Z_m}^{-1}\mathbf{Z}_m) \tag{B.9b}$$

$$= \boldsymbol{P}_{Z_0}\left[\boldsymbol{P}_{Z_m} + \boldsymbol{P}_{Z_0}\right]^{-1}(\boldsymbol{P}_{Z_m}\boldsymbol{P}_{Z_0}^{-1}\mathbf{Z}_0 + \mathbf{Z}_m) \tag{B.9c}$$

$\mathbf{Z}$ in Eq. B.9a is our best estimate for the projection of $\mathbf{X}$ into measurement space. Using that as your "ideal" solution, i.e., inserting $\mathbf{Z}$ as $\mathbf{Z}_I$ in Eq. B.5, you have

$$\begin{aligned}
\Delta\mathbf{X} &= \mathbf{X} - \mathbf{X}_0 \\
&= \boldsymbol{P}_{X_0}\boldsymbol{H}^T\boldsymbol{P}_{Z_0}^{-1}(\boldsymbol{P}_{Z_0}\left[\boldsymbol{P}_{Z_m} + \boldsymbol{P}_{Z_0}\right]^{-1}(\boldsymbol{P}_{Z_m}\boldsymbol{P}_{Z_0}^{-1}\mathbf{Z}_0 + \mathbf{Z}_m) - \mathbf{Z}_0) \\
&= \boldsymbol{P}_{X_0}\boldsymbol{H}^T(\left[\boldsymbol{P}_{Z_m} + \boldsymbol{P}_{Z_0}\right]^{-1}\boldsymbol{P}_{Z_m}\boldsymbol{P}_{Z_0}^{-1}\mathbf{Z}_0 - \boldsymbol{P}_{Z_0}^{-1}\mathbf{Z}_0) + \boldsymbol{P}_{X_0}\boldsymbol{H}^T\left[\boldsymbol{P}_{Z_m} + \boldsymbol{P}_{Z_0}\right]^{-1}\mathbf{Z}_m \\
&= \boldsymbol{K}(\mathbf{Z}_m - \left[\boldsymbol{P}_{Z_m} + \boldsymbol{P}_{Z_0}\right])\boldsymbol{P}_{Z_0}^{-1}\mathbf{Z}_0 + \boldsymbol{K}\mathbf{Z}_m \\
&= -\boldsymbol{K}\mathbf{Z}_0 + \boldsymbol{K}\mathbf{Z}_m
\end{aligned}$$

$$= -\boldsymbol{K}\boldsymbol{H}\mathbf{X}_0 + \boldsymbol{K}\mathbf{Z}_m \tag{B.10}$$

$$\mathbf{X} = \left[\boldsymbol{I}_{N_x} - \boldsymbol{K}\boldsymbol{H}\right]\mathbf{X}_0 + \boldsymbol{K}\mathbf{Z}_m \tag{B.11}$$

$$\boldsymbol{K} = \boldsymbol{P}_{X_0}\boldsymbol{H}^T\left[\boldsymbol{P}_{Z_m} + \boldsymbol{P}_{Z_0}\right]^{-1} = \boldsymbol{P}_{X_0}\boldsymbol{H}^T\left[\boldsymbol{H}\boldsymbol{P}_{X_0}\boldsymbol{H}^T + \boldsymbol{P}_{Z_m}\right]^{-1} \tag{B.12}$$

This corresponds to Kalman's equations for $\mathbf{X}$ and $\boldsymbol{K}$ (i.e., Eqs. 7.39 and 7.41), which shows the desired correspondence.

Glossary

Acceleration Second derivative (with respect to time) of the position vector. Describes the rate of change of the spacecraft's velocity vector.

Advanced Star Tracker A more advanced version of the Fixed-Head Star Tracker that incorporates its own internal star catalog and star recognition software.

Aerodynamic Drag Retarding force arising from collisions between the spacecraft and gas molecules comprising the Earth's atmosphere. A frictional force acting to oppose the spacecraft's direction of motion.

Aerodynamic Torque Torque induced by a difference between the spacecraft center of mass and center of pressure (with respect to a cross section perpendicular to the direction of motion).

Albedo Defines how much incident light an object reflects. Is the ratio of reflected energy divided by incident energy. The Earth's albedo has a value of about 0.3.

Alignment Matrix The matrix that defines the orientation of an attitude sensor or actuator relative to the spacecraft reference frame.

Altitude Height of a spacecraft above the Earth's surface.

Angstrom A unit of length equal to 10^{-8} cm.

Angular Acceleration The second derivative (with respect to time) of the spacecraft's orientation. Describes the rate of change of the spacecraft's angular velocity vector.

Angular Frequency See *Angular Velocity*.

Angular Momentum For a point particle, the angular momentum of the particle relative to the reference frame origin is equal to the cross product of the particle's position and momentum vectors. For a system of particles, the angular momentum of the system is equal to the sum of the angular momenta of each of the individual particles.

Angular Rate See *Angular Velocity*.

Angular Velocity The first derivative (with respect to time) of the spacecraft's orientation.

Antisolar Point Point on the celestial sphere defined by the negative of the vector pointing from the spacecraft to the Sun.

Aphelion Point of farthest distance from the Sun for a Solar orbit.

© Springer-Verlag London Ltd. 2017

H.L. Hallock et al., *ACS Without an Attitude*, NASA Monographs
in Systems and Software Engineering, DOI 10.1007/978-1-4471-7325-0

Apocenter General term for the farthest distance of an orbiting body from its parent body.

Apogee Point of farthest distance from the Earth for an Earth orbit.

Arc Length The angle between two vectors.

Arc minute One 60th of a degree.

Arc second One 60th of an arc minute.

Argument of Perigee For an Earth orbit, the angle between the right ascension of the ascending node and perigee, as measured in the orbital plane.

Argument of Perihelion For a solar orbit, the angle between the right ascension of the ascending node and perihelion, as measured in the orbital plane.

Ascending Node The point in the orbit where a spacecraft crosses the equatorial plane as it moves from the southern hemisphere to the northern hemisphere.

AST See *Advanced Star Tracker*.

Astigmatism An aberration in an optical system in which the system has different foci for rays that propagate in two perpendicular planes. If an optical system with astigmatism is used to form the image of a cross, the vertical and horizontal lines will be in sharp focus at two different distances.

Astronomical Unit (AU) The length of the Earth's semimajor axis with respect to its orbit about the Sun. It is equal to 1.49597870×10^8 km.

Attitude The orientation of the spacecraft in inertial space. Usually, attitude defines the orientation of all three spacecraft axes with respect to an inertial reference frame, although for spin-stabilized spacecraft it often is the case that only the orientation of the spin axis is specified.

Attitude Acquisition The initial establishment of a desired spacecraft orientation.

Attitude Actuator A control hardware component that generates control torques required to maintain spacecraft stability, null attitude errors, and reorient the spacecraft to a new orientation, etc.

Attitude Control The mechanism for establishing and maintaining a desired spacecraft orientation.

Attitude Control Accuracy A quantitative measurement of the success with which the spacecraft attitude is maintained at its desired orientation.

Attitude Control Torque Torques intentionally applied to the spacecraft to maintain or establish a desired spacecraft orientation.

Attitude Determination The computation of the spacecraft orientation relative to a specified reference frame. For celestial-pointing spacecraft, the reference frame is usually either the geocentric inertial (GCI) frame or the heliocentric inertial frame.

Attitude Determination Accuracy An estimate of the error magnitude of the computed spacecraft attitude.

Attitude Dynamics Study of motion about a spacecraft's center of mass.

Attitude Maneuver A commanded change in the spacecraft's desired attitude.

Attitude Matrix A specification of the spacecraft orientation in direction cosine matrix (DCM) format.

Attitude Perturbations Influences on a spacecraft's attitude behavior, typically arising from environmental effects.

Attitude Propagation Adjustment of the last absolute attitude solution to reflect later relative changes in the spacecraft orientation measured by gyro data.

Attitude Quaternion A specification of the spacecraft orientation in quaternion format.

Attitude Sensors Spacecraft hardware providing measurements that can be used to determine the spacecraft orientation or changes in the spacecraft orientation.

Attitude Stabilization Maintaining a desired spacecraft orientation.

Autumnal Equinox The moment in time at which the Sun, in its apparent orbit about the Earth, crosses the Earth's equatorial plane as it moves from the Northern hemisphere to the Southern hemisphere. The first day of Northern hemisphere Fall.

Azimuth Angle Angle (as measured on the Earth's equatorial plane, or on the equatorial plane of the reference frame in question) between a reference point on the equator and a given point on the equator.

Ballistic Coefficient A measure of the effect of atmospheric drag on the spacecraft.

Bandpass Filter A data processing algorithm that amplifies frequency response at selected frequencies, thereby providing disturbance rejection.

Bandwidth The frequency delta over which the frequency response has dropped off by 3 dB. Colloquially in spacecraft communications, bandwidth is a measure of how much data per unit time you can reliably get down from the spacecraft.

Bang-Bang Control Law For bang-bang control, the magnitude of the control torque is fixed. Only the sign of the control torque varies, depending on the sign of the error term.

Bang-Bang-Plus-Dead-Zone Control Law Bang-bang control, except if the magnitude of the error falls below a threshold, the control torque is zero.

Barycenter The position of the center of mass of two or more objects (e.g., the Earth-Moon Barycenter).

Batch Estimator A state estimation process where all the observations are processed simultaneously to yield a single state solution (opposite of a sequential or recursive estimator).

Batch Least-Squares Estimator A batch estimator which operates by minimizing the squares of the differences between observed and modeled values.

B-Dot Control Law A control law that treats measured changes in the geomagnetic field as if they arose entirely from attitude changes. Since the geomagnetic field varies over an orbital period, B-dot will cause the attitude to oscillate with a frequency equal to twice the orbital frequency (the B-dot natural frequency). (The name arises from the shorthand notation $\dot{B} \equiv dB/dt$ for the time derivative of the magnetic field.)

Bias Systematic measurement error.

Bipropellant Gas Jet A gas jet utilizing a fuel system where the fuel and oxidizer are stored separately.

Bode Diagram A graphical means of observing control system response. It graphs the logarithm of the transfer function magnitude vs. the logarithm of the frequency.

Body-Mounted Fixed to the spacecraft body.

Body Rate The angular velocity of the spacecraft.

Bolometer A thermistor that detects infrared (IR) radiation. Often used in Earth sensors.

Boresight The line of sight of sensor or science instrument.

Brahe, Tycho (1546-1601) Danish astronomer.

Brazilian Anomaly See *South Atlantic Anomaly (SAA)*.

Bremsstrahlung Radiation emitted when a charged particle (e.g., an electron or a proton) is decelerated suddenly. Comes from the German words *Bremsung*, meaning deceleration, and *Strahlung*, meaning radiation.

Cartesian Coordinates An ordered set of three real numbers defining the position of a point relative to three equal scaled, orthogonal coordinate axes; typically referred to as (x, y, z) coordinates.

Celestial Coordinates Defining directions on the celestial sphere, typically using right ascension and declination.

Celestial Equator The great circle on the celestial sphere perpendicular to the Earth's spin axis.

Celestial-Pointer A spacecraft whose fine-pointing science instruments are oriented towards "fixed" points on the celestial sphere. For example, a spacecraft that slews from one attitude to another to observe a series of X-ray point-sources would be a member of the class of celestial pointers.

Celestial Poles The North and South directions on the celestial sphere.

Celestial Sphere The surface of a sphere of unit radius, whose points constitute directions in space. Not surprisingly, the spacecraft-centered celestial sphere places the spacecraft at the origin of the coordinate system.

Center of Mass Average position of a system of objects, weighted in proportion to their masses.

Center of Pressure For an object exposed to solar radiation (for example), the point on the object at which the cumulative effect of all the solar flux can be viewed as being applied as a single force. If there is a difference between the center of pressure and the center of mass, that force will generate a torque on the object.

CGRO See *Compton Gamma Ray Observatory*.

Characteristic Frequency A frequency at which a lightly damped system will resonate when driven by an external force (assuming that the force contains a component at that frequency).

Charge Coupled Device (CCD) Star Tracker A star tracker that detects stars by digitally scanning an array of photosensitive components (pixels). The star tracker integrates the electrical charge in the pixels "struck" by (for example) starlight. The measurement is performed by reading-out the pixel output line-by-line.

Circle A conic section whose eccentricity is zero. An ellipse whose semimajor axis and semiminor axis are equal.

Circular Velocity Velocity of a spacecraft within a perfect circular orbit of a given radius.

Closed-Loop Control A control law utilizing error feedback (i.e., the difference between the measured output and the desired output).

CO_2 Absorption Band Spectral region of Earth's atmosphere associated with the carbon dioxide absorption band. Solar radiation is absorbed by carbon dioxide molecules, and then re-emitted uniformly in all directions. Because of its relative stability, most Earth sensors are designed to trigger on the CO2 band.

Coarse Attitude Determination (CAD) Spacecraft attitude determination (typically) at an accuracy level of 1 degree or worse, typically based on Sun sensor and magnetometer data.

Coarse Pointing Control Spacecraft attitude control (typically) at accuracy levels of a degree or worse.

Coarse Sun Sensor (CSS) Measures photocell output as a function of Sun angle, with photon flux proportional to the cosine of the Sun angle relative to photocell surface normal.

Cold Gas Thrusters Thrusters that generate their energy by releasing a phase change's latent heat (or, if there's no phase change, from the work of compression).

Colored Noise Noise with a non-uniform frequency distribution, in contrast to "white" noise.

Column Matrix An $N \times 1$ matrix, where N is the number of entries in the matrix. Column matrices are matrix representations of vectors.

Coma An aberration in an optical system manifest as a variation in magnification across the entrance pupil. The typical result of comatic aberration is that the image of point source of light that is offset from the optical axis will appear like a tear drop with the narrow end pointing away from the optical axis.

Commanded Attitude Desired spacecraft orientation.

Command Quaternion Desired spacecraft orientation expressed in quaternion format.

Component An element of a vector or matrix.

Compton, Arthur (1892-1962) American physicist and Nobel laureate.

Compton Gamma Ray Obervatory (CGRO) A NASA satellite that detects cosmic sources of gamma radiation

Cone Angle Equivalent to arc length. Frequently used when referring to the angle between a spacecraft's spin axis and an attitude sensor boresight.

Cone-Cone Intersection Method used to solve for the attitude of a spinning spacecraft, where a snapshot of data from two attitude sensors is used to determine the spin axis. Similar to the triangulation method used to determine the location from which an earthquake has originated.

Conic Sections The 4 types of curves resulting from intersecting a plane with a cone: specifically the circle, ellipse, parabola, and hyperbola.

Coning The attitude motion of a rotating spacecraft resulting when the geometric axes are not co-aligned to the principal axes.

Consumables Non-reusable spacecraft resources.

Control Law A process for controlling the spacecraft attitude, utilizing sensors to determine the attitude error and actuators to generate the control torques required to null the attitude error.

Control Torque Torques generated by spacecraft actuators for the purpose of controlling the attitude.

Convergence The state when an estimator has reached a stable solution.

Coordinated Universal Time (UTC) A time scale based on atomic clocks, but which includes leap seconds to keep the time synchronized with the Earth's rotation.

Cosine Detector See *Coarse Sun Sensor*.

Covariance A statistical measure of how the errors (noise) for two measured quantities are interdependent.

Covariance Matrix A square matrix whose diagonal elements are the variances of the state vector elements, and whose off-diagonal elements are the covariance of pairs of state vector elements.

Critically Damped Control law behavior where steady state is reached without overshoots or oscillations. Condition at the boundary between under-damped and over-damped.

Cross Product A vector product yielding a vector whose direction is perpendicular to the input vectors and whose magnitude is proportional to the sine of the angular separation between the vectors.

Cross-track Error The error in spacecraft position with respect to the plane of the orbit.

Damping Retarding motion, or when the amplitude of an oscillation is reduced over time.

Data Dropout The intermittent absence of data.

Data Filter An algorithm that separates desired low-frequency information from the accompanying high-frequency noise.

Data Flagging Identifying and removing invalid data.

Data Smoothing Removing noise from data.

Data Validation Examining data to determine if it is acceptable for downstream processing.

Data Weighting Quantitatively rating some data higher than other data to improve the accuracy of algorithm using all the data as input.

dB see *Decibel*.

Deadband An error region (less than a given threshold) in which active attitude control is inhibited.

Decibel A logarithmic scale for expressing the ratio of a quantity x to a reference level x_{ref}: $dB = 10 \log_{10}(x/x_{ref})$. Since $10 \log_{10} 2 \approx 3$, doubling x is equivalent to adding approximately 3 dB to the decibel measure of x, while halving x is equivalent to subtracting approximately 3 dB from the decibel measure of x.

Declination Angle between an object and the plane of the Earth's equator as seen from the center of the Earth.

Definitive Orbit The best estimate of a spacecraft's orbit during a time period in which tracking data, or other direct orbit measurements, have been made.

Degeneracy A pathological attitude geometry condition, as for example (in coarse attitude determination) when the Sun vector and Geomagnetic field vector are parallel. Mathematically, is usually characterized by the presence of a singularity (i.e., a divide by zero) in the algorithm.

Delta Function The delta function $\delta(x)$ is a "Narrow peak" function; it is zero everywhere except at $x = 0$. At $x = 0$ it is infinitely large, but defined so that the total area under the curve is equal to 1. $\int_{-\infty}^{\infty} \delta(x)dx = 1$.

Delta V Change in velocity. Also, the name of a control mode associated with using thrusters to adjust the orbit by changing the spacecraft velocity.

Descartes, Rene (1596-1650) French philosopher, mathematician, and scientist.

Descending Node The point in the orbit where a spacecraft crosses the equatorial plane as it moves from the Northern hemisphere to the Southern hemisphere.

Diagonal Matrix A square matrix whose off-diagonal elements are all zero.

Differential Velocity Aberration The small difference in light direction shift due to velocity aberration due to slight differences in star or target direction.

Digital Sun Sensor (DSS) Measures output from a series of photocells to determine Sun angle. If a photocell's output is greater than a threshold, it is considered to be "on". The pattern of "on" photocells is directly associated with the Sun angle. Note that this scheme provides a digital representation of the Sun angle.

Dipole The first order term in a spherical harmonic expansion.

Dipole Model Approximating a physical phenomenon by keeping just the first order term in a spherical harmonic expansion, as for example a dipole model of the geomagnetic field.

Dipole Moment For a wire coil wrapped around a bar, the dipole moment is a vector quantity proportional to the magnetic moment vector, with the constant of proportionality reflecting the nature of the material of which the bar is composed.

Direction Cosine Matrix A 3×3 square matrix used (in our context) to represent rotations, spacecraft attitude, or relative alignments between devices.

Distribution Law For reaction wheel systems comprising more than 3 reaction wheels, an algorithm to assign torque commands to individual wheels so as both to reduce the attitude error and prevent wheel run-up.

Disturbance Torque An environmental torque perturbing the spacecraft.

Diurnal Variations Variations (for example, in atmospheric density) driven by local time.

Divergence The state when an estimator's solution disagrees with truth by a continuously increasing amount.

Doppler Effect The apparent change in the frequency of waves (light, sound, etc.) being emitted by an object due to the relative radial velocity between the emitting object and the receiving platform.

Dot Product A vector product yielding a scalar that is proportional to the cosine of the angle between the two vectors.

Downlink Telemetry transmission from the spacecraft to the ground.

Drag Retarding of the spacecraft orbital motion due to atmospheric friction.

Drag Coefficient A parameter describing the spacecraft's response to atmospheric drag.

Dual-Slit Sun Sensor A digital Sun sensor designed for a spinning spacecraft. One slit is canted relative to the other slit such that the time delay in Sun pulses in the two slits can be directly associated with the Sun angle relative to the spacecraft spin axis.

Dynamic Equations of Motion Equations relating the first derivative (with respect to time) of the spacecraft angular momentum vector to the torques being applied to the spacecraft.

Earth Oblateness Bulging of Earth (from a perfect sphere) at the equator. Earth oblateness is a major perturbation on low Earth orbits.

Earth-Pointer A spacecraft whose fine-pointing science instruments are oriented along the nadir vector towards the Earth.

Earth Sensor An attitude sensor that detects reflected light from the Earth, typically in the IR range. Because the disk of the Earth is quite large (as opposed to point objects, like stars), Earth sensors reporting a nadir vector must do fairly elaborate centroid calculations to report the direction of the Earth center. In the case of narrow field of view (FOV) sensors on spinning spacecraft, the Earth sensor simply may report Earth-in and Earth-out times, leaving it to ground or onboard processing to determine the Earth angle with respect to the spin axis.

Eccentric Anomaly Angle between perigee and the projection of spacecraft position onto a circle in the plane of the orbit circumscribed about the orbit.

Eccentricity A scalar parameter quantitatively describing an ellipse's departure from a perfect circle.

Eclipse An occultation of one celestial object by another. For example, a Solar eclipse observed on the Earth occurs when the Moon occults the Sun.

Ecliptic The plane of the Earth's orbit.

Eigenaxis The axis about which a slewing spacecraft rotates.

Eigenvector For rotations, the vector about which the rotation occurs.

Elevation Angle above the Earth's equator, or above the equator of the reference frame in question.

Ellipse A conic section whose eccentricity is less than 1. The general form of an ideal bound orbit. A circle is a special case of an ellipse for which the eccentricity is zero.

Energy Dissipation Loss of energy, typically due to frictional effects.

Environmental Torques External influences applying torques to the spacecraft. Examples are aerodynamic, magnetic, gravity gradient, and Solar radiation.

Ephemeris A listing of positions (either of celestial objects or spacecraft) at regular time intervals. Also, colloquially referred to as an empirical (as opposed to algorithmic) specification of a spacecraft's orbital position (either historical or predicted).

Epoch Time The start time of an ephemeris file, or the reference time of a set of (Keplerian) orbital elements.

Equator In Earth-centered coordinates, the great circle perpendicular to the Earth's polar axis.

Equinoxes The two days in the year when day and night are of equal duration. See *Autumnal Equinox and Vernal Equinox*.

Escape Velocity The velocity required to place a spacecraft in a parabolic orbit.

Euler, Leonhard (1707-1783) Swiss mathematician and physicist.

Euler Angles A specification of the attitude, or a change in attitude, by a series of rotations. The angles of rotation are called the Euler angles.

Euler Axes The axes, relative to which, the Euler angles are defined (see *Euler Angles*).

EUVE See *Extreme Ultraviolet Explorer*.

External Torque A torque applied to a spacecraft that does not originate from the spacecraft (for example, a gravity gradient torque), or that couples the spacecraft to the outside "world" (for example, magnetic torques arising from magnetic torquer bars coupling with the geomagnetic field). External torques will change the total angular momentum of the spacecraft.

Extreme Ultraviolet Explorer A NASA satellite that detects cosmic sources of extreme ultraviolet radiation.

Fading Memory The process in (for example) a Kalman Filter by which more recent data is weighted higher than older data.

Failure Correction A flight software correction of an inflight problem, either due to a hardware anomaly or a malfunctioning of a flight software element,

Failure Detection Identification of an inflight problem by flight software, either due to a hardware anomaly or a malfunctioning of a flight software element,

Failure Detection and Correction (FDAC) See the individual entries on failure detection and failure correction. It is also referred to as fault detection and correction with the same acronym (FDAC), or the acronym FDC.

Feedback Control Use of realtime measured errors to maintain attitude stability or conduct accurate slews.

Feed-Forward Control Use of onboard models (e.g., gravity gradient torque) to anticipate environmentally-induced errors or planned adjustments and compensate for them in advance.

FHST See *Fixed-Head Star Tracker*.

Fine Error Sensor (FES) A star sensor having greater accuracy than a conventional star tracker, often using the optics of the primary imaging science instrument (for example, HST's fine guidance sensors (FGSs)). Often, an FES will be built into the science instrument itself, or constitutes a special operational mode of the science instrument.

Fine Pointing Control Flight software control of the spacecraft attitude to an accuracy (typically) of a few tens of arcseconds or better, although nadir-pointing missions may have looser requirements.

Fine Sun Sensor (FSS) A high precision digital Sun sensor (DSS).

First Point of Aries The point the Sun passes through on the vernal equinox.

Fixed-Head Star Tracker (FHST) Purely electronic star acquisition and tracking devices. Initially, the NASA Standard FHST was an image dissector tube (IDT) device, but starting in the early 1990s charge coupled devices (CCDs) became the new standard.

Flag A software variable specifying the results of a validity check.

Flat Spin For spin-stabilized spacecraft initially rotating about a body axis other than the major principal axis, energy dissipation will cause angular momentum to be pumped out of the spin axis and into rotation about the major principal axis. When the rotational motion of the spacecraft has been converted entirely into

rotation about the major principal axis, the spacecraft is said to be in a condition of flat spin.

Flexible Modes Natural spacecraft vibrational frequencies. Care must be taken in control law design to avoid exciting these frequencies, thereby producing a destabilizing resonant behavior.

Flexible Spacecraft Dynamics The rotational behavior of spacecraft in response to the bending characteristics of spacecraft flexible components due to the influence of external torques.

Fluxgate Magnetometer A device that detects external magnetic fields by measuring currents induced (by external fields) in two saturable coils.

Focus Each ellipse has two foci. The foci are located relative to points on the ellipse such that the sum of the distances from the point to the foci is the same for all points on the ellipse. For elliptical orbits, the center of mass of the gravitational system is located at one of the foci.

Force The product of mass times the acceleration.

Fourier, Joseph (1768-1830) French mathematician.

Fourier Expansion An approximate representation of an algebraic function by a Fourier Series.

Fourier Series The representation of an algebraic function in terms of sine and cosine functions.

Fourier Transform A mathematical procedure converting a function in the time domain to a function in the frequency domain. Specifically, it decomposes a function into sinusoidal functions covering a spectrum of frequencies. The inverse procedure transforms from the frequency domain to the time domain.

Frequency The inverse of period. A measure of angular rate.

Frequency Response In control systems, the steady-state response of a control system to a sinusoidal input.

Friction A dissipative force converting organized (usually mechanical) energy into heat.

Gain A multiplier on a term in a control law enabling you to amplify or diminish the term's contribution.

Gamma Rays Light having wavelength less than 0.1 angstroms. Gamma rays are produced by nucleus and particle decays.

Gas Jet A thruster utilizing gas propellant, either hot or cold.

Gauss, Carl Friedrich (1777-1855) German mathematician and scientist.

Gaussian Coefficients Scalar quantities in a spherical harmonic expansion used for empirical modeling of the geomagnetic field.

Gaussian Measurement Errors Random errors that follow a bell-shaped (i.e., Gaussian) distribution.

Gaussian Noise Noise that follows a bell-shaped (i.e., Gaussian) distribution.

Gauss's Equation Expresses the true anomaly in terms of the eccentric anomaly.

Geocentric Inertial (GCI) An Earth-centered, inertial (i.e., non-rotating with respect to the "fixed" stars) reference frame. The Euler angles specifying spacecraft orientation within the GCI are right ascension, declination, and roll.

Geomagnetic Field The magnetic field of the Earth.

Geomagnetic Field Model A mathematical representation of the Earth's magnetic field specifying the magnitude and direction of the field at a given point in an Earth-rotating reference frame.

Geometric Spacecraft Axes "Natural" axes defined relative to the spacecraft's structure. For example, the x-axis (or roll axis) of a stellar pointer is usually defined to be parallel to the boresight of its primary imaging instrument.

Geostationary Orbit An orbit that maintains a spacecraft directly above the same point on the Earth's surface at all times. This is made possible by the orbit altitude being precisely chosen so that its associated period is 24 hours, matching the Earth's rotational rate. All geostationary orbits are equatorial (i.e., inclination equals zero).

Geosynchronous Orbit An orbit having the same altitude as a geostationary orbit, but not necessarily maintaining the spacecraft above the same point on the Earth at all times. Geosynchronous orbits may have non-zero inclinations.

Gimbal A rotatable element of a piece of spacecraft hardware, for example a gimbaled high gain antenna (HGA) or solar array.

Global Positioning System (GPS) A constellation of low Earth orbiting satellites supporting realtime, onboard measurement of spacecraft position.

Goddard Trajectory Determination System (GTDS) The primary ground software system at Goddard for computing definitive and predictive spacecraft ephemerides.

Goodness-of-Fit Function A statistical algorithm for specifying the validity with which a model portrays actual data.

Gravitation The phenomenon of mutual attraction of massive objects. In our context, the force of gravity is well-modeled by Newton's Law of Gravitation.

Gravitational Constant The constant of proportionality G in Newton's Law of Gravitation. It is equal to $6.67428 \times 10^{11} m^3 kg^{-1} s^{-2}$

Gravity-Gradient Stabilization A passive attitude control strategy applicable to spacecraft with elongated (e.g., rod-shaped) structures. Exploits the behavior of an elongated body to be pulled by gravity so as to be oriented radially relative to the gravitational object.

Gravity-Gradient Torque The torque arising from the difference in gravitational force "felt" at opposite ends of an elongated body. The resulting torque acts to align the elongated body radially relative to the gravitational object.

Gray Code Bit pattern characterization of output from a digital Sun sensor.

Great Circle A circle drawn on the surface of a sphere, where the plane of the circle intersects the center of the sphere (i.e., the center of a great circle and the center of the sphere coincide).

Gyro Originally, a mechanical sensor containing a spinning mass that exploits conservation of angular momentum to measure changes in attitude. Recently gyros have been employed on-orbit that use non-mechanical structures (for example, hemispheric resonating gyros (HRGs) and laser ring gyros).

Gyro Drift A ramping error (with time) in the gyro's output.

Gyro Drift Bias Scalar calibration offset correcting for the effects of gyro drift in gyro measurements.

Gyro Scale Factor & Alignment Calibration A series of large slews are executed to generate gyro data that the ground system uses to determine the gyro scale factors and gyro alignments relative to absolute attitude sensors (such as star trackers).

Gyroscope See *Gyro*.

Heliocentric Coordinates Similar to GCI coordinates, except the Sun is at the center of the coordinate system instead of the Earth.

Hermite, Charles (1822-1901) French mathematician.

Hermite Polynomials An infinite set of polynomials in the independent variable x that satisfy the differential equation $y'' - 2xy + 2ny = 0$, with n an integer. The Hermite polynomials are orthogonal with respect to the weight function $\exp(-x^2)$ over the interval $(-\infty, +\infty)$. They have proven useful for representation of orbits for onboard ephemeris algorithms.

Horizon Radiance Model A mathematical model describing the behavior of the Earth's CO_2 band as a function of time. Provides reference data in defining the horizon when using horizon sensors for attitude determination.

Horizon Sensor An attitude sensor that measures spacecraft orientation relative to the Earth.

Hot Gas Thrusters Thrusters that generate their energy by a chemical reaction.

HST See *Hubble Space Telescope*.

Hubble, Edwin (1889-1953) American astronomer, discovered the expansion of the universe.

Hubble Space Telescope A NASA satellite used to study celestial bodies emitting radiation from the near infrared to the near ultraviolet.

Huygens, Christiaan (1625-1695) Dutch mathematician, astronomer, and physicist.

Hydrazine A monopropellant fuel (i.e., fuel and oxidizer are mixed together) with chemical formula N_2N_4.

Hydrazine Thruster A gas jet utilizing hydrazine for fuel.

Hyperbola A conic section whose eccentricity is greater than 1.

Hysteresis In the context of electricity and magnetism, the property of ferromagnetic substances in which the behavior of the substance in the presence of a magnetic field depends on the substance's previous exposure to magnetic fields.

Identity Matrix A square matrix which when multiplied by (or multiplies) another matrix returns that matrix. It is a diagonal matrix with 1s along the main diagonal.

Impulse The time integral of the force or torque. So the impulse is the change in momentum (translational or angular) over a specified (typically short) period of time.

Inclination Keplerian element describing the tilt angle of the orbit relative to the equatorial plane.

Induction Magnetometers Attitude sensors measuring external magnetic fields by utilizing Faraday's Law of Magnetic Induction, which relates the line integral of the electric field induced in a coil to the time variation in the magnetic field to which it's exposed.

Inertial Frame A reference frame that's hat's neither rotating nor accelerating with respect to the "fixed" stars.

Inertial Reference Unit (IRU) See *Gyro*.

Infrared Light Light roughly at wavelengths between 0.76 μm and 2 mm.

Injection Placing a spacecraft in its orbit.

In-Plane Orbit Maneuver Orbit maneuvers that don't change the inclination.

Input Axis Axis of rotation with respect to which the gyro can sense motion.

Instantaneous Rotation Axis The axis about which a spacecraft is rotating at a given time.

Instrumental Star Magnitude The brightness of a star as observed by a star tracker, as opposed to the brightness of the star as it would be observed by an ideal sensor. Note that the star brightness in either case is reduced by the separation distance between the star and the observer, so neither is a measurement of the star's intrinsic brightness.

Internal Torque A torque generated by one component of a spacecraft that (as a result of conservation of momentum) causes an equal, but opposite torque to be applied to the body of the spacecraft (for example, a reaction wheel torque applying an inverse torque to the body). Internal torques cannot change the total angular momentum of the spacecraft.

International Geomagnetic Reference Field (IGRF) A set of Gaussian coefficients that empirically model the Earth's magnetic field as a function of both position and time.

In-Track Error The error in spacecraft position along the arc of the orbit (an azimuthal error).

Invertible Matrix A square matrix possessing an inverse.

Ion Thrusters Thrusters that generate thrust by accelerating ionized molecules.

J_2 Perturbations Gravitational perturbations arising from Earth oblateness.

J_2 Term The first-order term in a harmonics expansion of the Earth's gravitational potential.

Jacchia-Roberts Atmospheric Density Model A model of the Earth's atmospheric density, taking into account altitude, molecular mixing, temperature, time of day, seasonal variations, Solar cycles, etc.

Julian Day The number of days since noon on January 1, 4713 BCE.

Kalman, Rudolf (1930-) Hungarian-American mathematician and electrical engineer.

Kalman Filter The most computationally efficient recursive filter, commonly used for attitude and gyro bias estimation on GSFC spacecraft.

Kepler, Johannes (1571-1630) German mathematician and astronomer.

Kepler's Equation Expresses the mean anomaly of a spacecraft's orbit in terms of the eccentric anomaly.

Kepler's Laws Pre-Newtonian descriptions of planetary motion that stated that (for two-body problems) an orbiting object follows an elliptical path (an unbound object follows a hyperbolic path), that the line joining the two celestial objects sweeps out equal areas in equal times, and that the square of the orbital period is

proportional to the cube of the semi-major axis. Newton later showed that Kepler's laws could be derived from Newton's laws of mechanics and gravitation.

Keplerian Orbit An ideal two-body orbit parameterizable by Kepler's orbital elements.

Keplerian Orbital Elements Parameters describing the size, shape, and orientation of the orbit, as well as the location of the orbiting body at a specific time (the epoch time of the orbit). Kepler's elements are the semimajor axis, eccentricity, inclination, right ascension of the ascending node, argument on perigee, and mean anomaly.

Kinetic Energy The energy of motion, defined as the amount of work required to bring a body from rest to a final velocity v. It is equal to $mv^2/2$, where m is the mass of the body.

Lag Filter A filter designed to improve steady-state accuracy. Has similarities to a proportional-integral (PI) control law, but when used in conjunction with a PI can improve system stability.

Lagrange, Joseph (1736-1813) Italian mathematician and astronomer.

Lagrange Point Orbit An orbit having a Lagrange point as a focus. The orbit is referred to as a halo orbit.

Lagrange Points Positions of stable or pseudo-stable gravitational equilibrium for a small body under the gravitational influence of two large bodies orbiting each other in nearly circular orbit.

de Laplace, Pierre (1749-1827) French mathematician.

Laplace Transformation A mathematical procedure for converting a function in the time domain to a function in the frequency domain. Laplace transforms are used to convert differential equations (in the time domain) into equivalent algebraic equations (in the frequency domain) whose solution (mechanically) is easier than the original differential equation. The inverse procedure transforms from the frequency domain to the time domain. The inverse transformation can be used to take the solution to the algebraic equation (in the frequency domain) and convert it back to the desired (time domain) solution to the differential equation.

Latitude Angle above the Earth's equator, or above the equator of the reference frame in question.

Latus Rectum A chord through a focus of an ellipse, and perpendicular to the major axis.

Launch Window A time duration during which a spacecraft can be launched and successfully achieve mission orbit. Launch windows may also be constrained by attitude issues, such as shadow avoidance or Sun angle at orbit insertion.

Lead Filter A filter designed to lower rise time and decrease transient overshoot. Has similarities to a position-derivative (PD) control law, but with reduced sensitivity to sensor noise.

Lead-Lag Filter A combination of lead and lag filters. Has similarities to a proportional-integral-derivative (PID) control law.

Leak-In Only responding to a fraction of the computed error on a given control cycle, hence leaking the error in over time.

Leap Second An extra second added to the UTC time scale to keep UTC's atomic-based time in synchronization with the Earth's rotation.

Least-Squares Filter A filter that minimizes the squares of the residuals between measured and modeled quantities.

Leibniz, Gottfried Wilhelm (1646-1716) German philosopher and mathematician.

Legendre, Andrien-Marie (1752-1833) French mathematician.

Legendre Polynomials An infinite set of orthogonal functions over the range $[-1, 1]$, used for constructing latitude dependence in spherical harmonic expansion.

Libration For gravity-gradient stabilized spacecraft, an oscillatory motion of the spacecraft attitude about its equilibrium pointing. With respect to satellite orbits, oscillatory behavior associated with equilibrium points in a gravitational system (see *Lagrange Points*).

Libration Damping Suppression of the libration motion, either actively or passively.

Libration Points See *Lagrange Points*.

Lifetime The time duration (starting at initial orbit insertion) before atmospheric drag causes the spacecraft to reenter.

Light Year The distance light travels in one Julian year (365.25 days).

Limit Checking A validation procedure in which a data point value is checked against a threshold (upper, lower, or both).

Limited Stable System Output converges to a value other than the initial (pre-perturbation) steady-state value.

Line of Apsides The line connecting the apogee and perigee of an orbit. It also contains the orbit's two foci.

Line of Nodes The line connecting the ascending node and the descending node of an orbit.

Linear Combination A sum of a set of objects, where each object can be multiplied by a scalar.

Linear Independence A property of a set of N vectors whereby no linear combination of $(N - 1)$ of the vectors will yield the remaining vector.

Longitude Angle (as measured on the Earth's equatorial plane, or on the equatorial plane of the reference frame in question) between a reference point on the equator and a given point on the equator.

Low Earth Orbit (LEO) Orbits for which atmospheric drag is the dominant perturbation. A popular LEO altitude range is 500 to 600 km.

Magnetic Coil Electrical current run through a wire loop that generates a magnetic field. A magnetic coil is a series of wire loops wrapped about a ferromagnetic core.

Magnetic Disturbance Torque A spacecraft torque generated by the interaction of the spacecraft's residual dipole with the geomagnetic field.

Magnetic Moment For magnetic coils, a vector normal to the wire loops comprising the coil, equal to the number of turns of wire multiplied by the current

in the coil and the cross-sectional area of the coil. The direction of the vector is given by right-hand rule, with the fingers curling in the current direction.

Magnetic Pole Strength The magnetic analog of electric charge.

Magnetic Stabilization See *B-dot Control Law*.

Magnetic Torque Torque on the spacecraft arising from interaction of fields generated by the spacecraft's magnetic coils and the geomagnetic field. If the torque arises from an interaction between a spacecraft residual dipole moment and the geomagnetic field, the torque is called the magnetic disturbance torque.

Magnetic Torquer Bar (MTB) See *Magnetic Coil*.

Magnetometer An attitude sensor measuring the strength and direction of the magnetic field external to the spacecraft. Note that this measurement will include not only the geomagnetic field (for spacecraft in Earth orbit), but also contributions from MTBs and the spacecraft residual dipole moment.

Magnitude (1) The square-root of the sum of the squares of a vector. (2) A logarithmic specification of an astronomical object's brightness. Magnitude is actually a measure of *dimness*, since larger numbers correspond to dimmer objects. The brightest stars in the night sky have an apparent magnitude around 0, while the dimmest visible to the human eye have an apparent magnitude of about 5. A magnitude 0 star is 100 times brighter than a magnitude 5 star, so a difference of one magnitude represents a factor of $100^{1/5}$ change in brightness. The absolute magnitude is a measure of intrinsic brightness; it is the apparent magnitude a body would have if it were at a distance of 10 parsecs from Earth.

Major Principal Axis The spacecraft axis associated with the largest principal moment of inertia.

Marginally Stable System When output displays continuous, bounded oscillatory behavior.

Matrix Inverse The square matrix which when multiplying a given square matrix, or multiplied by the given square matrix, yields the identity matrix.

Mean The average of N numbers, computed as the sum of the numbers divided by N.

Mean Anomaly Angle corresponding to the fraction of an orbit (360°) defined by the ratio of the time since perigee passage and the orbital period.

Mean of Date When defining coordinates of directions in the GCI reference frame, taking into account the precession of the Earth's spin axis relative to inertial space, but not nutation of the spin axis.

Mean Orbital Elements Average values of the orbital elements over a given time duration.

Measured Attitude The current best estimate of the spacecraft attitude.

Measurement Output from an attitude (or orbit) sensor.

Measurement Covariance Matrix In estimation theory, a matrix whose elements describe the reliability of the measurements.

Meridian A great circle on the celestial sphere that passes through the poles. Meridians are perpendicular to the equator.

Minor Principal Axis A spacecraft axis associated with either of the two non-largest principal moments of inertial.

Misalignment An error in the alignment of a sensor, actuator, or science instrument.

Misalignment Matrix The matrix characterizing the error in alignment of a sensor, actuator, or science instrument.

Mode See *Control Mode*.

Modified Julian Day The Julian Day minus 2400000.5; it is therefore the number of days since midnight on November 17, 1858.

Molniya Orbit An orbit designed to support Russian communications satellites. It freezes perigee in the southern hemisphere, tilting the orbit to place the orbit curve at apogee at the northern latitudes, ensuring maximum communication opportunities between spacecraft and ground stations.

Moment of Inertia Tensor A 3×3 matrix describing an object's mass distribution. For angular motion, the moment of inertia tensor is the parallel concept to the mass in linear motion. The increase in complexity (9 numbers in a matrix as compared with a single scalar) is needed to describe the motion of a 3 dimensional object rotating about its center-of-mass, as opposed to the much simpler translational motion of the center of mass itself.

Momentum For a point particle, the product of the particle's mass and velocity vector. For a system of particles, the sum of linear momenta of each of the individual particles. Often in the aerospace business the term "momentum" is used (technically incorrectly) interchangeably with the term "angular momentum", as in the phrase "momentum dumping".

Momentum Bias Operation of a wheel or system of wheels at non-zero angular momentum (e.g., a Sun-pointing safemode may build up a non-zero total angular momentum to provide gyroscopic stiffness for stability enhancement).

Momentum Dumping For spacecraft controlled by reaction or momentum wheels, angular momentum must be removed from the wheel system or the wheels can saturate (i.e., wheel speeds can be run up to maximum values). Angular momentum dumping utilizes other attitude actuators to shift or remove angular momentum from the wheels (e.g., applying current to magnetic coils thereby applying a desirable spacecraft torque by coupling to the geomagnetic field).

Momentum Wheel A large wheel typically flown on spinning spacecraft or (in the past) on nadir-pointing spacecraft.

Monopropellant Gas Jet A gas jet utilizing a fuel system where the fuel and oxidizer are stored together.

Monotonic A curve that either is continuously increasing or continuously decreasing.

Nadir Directly downwards towards the closest point on the Earth's surface. (Due to Earth oblateness, nadir can be slightly offset from the direction to Earth center.)

Nadir Angle For spinning, Earth-orbiting spacecraft, the angle between the spacecraft spin axis and the nadir vector.

Nadir-Pointer A spacecraft whose primary science instrument is directed along the nadir vector.

Nadir Vector For Earth-orbiting spacecraft, the vector directed towards the Earth's surface.

Natural Frequency The characteristic frequency at which a control system will oscillate if the damping is zero.

Nichols Chart A graphical presentation of the relationship between open-loop and closed-loop frequency response.

Newton, Isaac (1643-1727) English physicist, mathematician, astronomer, natural philosopher, alchemist, and theologian.

Nodal Regression See *Nodal Rotation*.

Nodal Rotation The rotation of the line of nodes due to orbital perturbations caused by Earth oblateness.

Nodes For Earth-orbiting spacecraft with non-zero inclination, the two points where the spacecraft orbit intersects the Earth's equator.

Notch Filter A filter that produces relative low gains at selected frequencies, thereby providing rejection capability.

Nutation The attitude motion resulting when the spacecraft's rotation axis is not parallel to a principal axis. For spin-stabilized spacecraft, a rotational motion of the spacecraft attitude about its equilibrium pointing.

Nutation Damping Suppression of the nutation motion, either actively or passively.

Nyquist, Harry (1889-1976) Dutch-American electrical engineer.

Nyquist Stability Criterion A condition imposed on the roots and poles of a feedback control system which, if satisfied, guarantees that the control system will be stable. The examination naturally lends itself to graphical representation for verification.

Oblate The shape of a spheroid having an equatorial radius greater than its polar radius (i.e. "bulging" at the equator).

Oblateness See *Earth Oblateness*.

Obliquity of the Ecliptic The dihedrall angle between the Earth's equatorial plane and the Earth's orbital plane.

Observability The capacity of a state vector element to be solved for reliably, given the amount and type of data available.

Observation An attitude (or orbit) relevant quantity constructed from sensor measurements.

Observation Model Algorithm used to generate predicted values of observations.

Observation Model Vector The set of predicted values of observations at a time snapshot during an estimation calculation.

Observation Residual Vector The set of differences between the modeled observations and the actual observations.

Occultation Geometric condition when the view of the spacecraft sensor or science instrument is obstructed by a celestial body.

Open-Loop Control A control law not utilizing error feedback.

Orbit Acquisition Achieving mission orbit following launch vehicle separation or completion of a major inflight orbit maneuver.

Orbit Decay For Earth-orbiting spacecraft, the decrease in apogee due to atmospheric drag.

Orbit Defined Coordinate System A reference frame in which spacecraft orientation is defined by three Euler angles, roll, pitch, and yaw. The yaw axis is the nadir vector. The pitch axis is negative orbit normal. The roll axis is cross-product of the pitch and yaw axis, and for a circular orbit is approximately parallel to the spacecraft velocity vector.

Orbit Determination Computation of the spacecraft (or celestial body) orbit in inertial space.

Orbit Determination Accuracy A quantitative estimate of the error magnitude of the computed spacecraft orbit.

Orbit Dynamics Study of the motion of the spacecraft's center of mass.

Orbit Generator An algorithm that computes predicted spacecraft position and velocity vectors.

Orbit Maneuver A commanded change in the spacecraft's orbit. If applied to null errors in the actual orbit relative to the desired orbit the maneuver is called a stationkeeping maneuver.

Orbit Normal A vector perpendicular to the orbital plane. The normal vector obtained by right hand rule, with the fingers curled in the direction of spacecraft motion, is positive orbit normal. The opposite vector is negative orbit normal.

Orbit Perturbations Lower order influences on a spacecraft's orbit behavior, typically arising from environmental or celestial effects. The largest of the set of perturbations (for a given orbit geometry) is referred to as the dominant perturbation, or the perturbation that dominates at that geometry.

Orbit Propagation Extrapolation of spacecraft position and velocity vectors from actual measurements at an earlier time.

Orbital Elements A set of parameters specifying the size, shape, and orientation of the orbit in inertial space, as well as the location of the spacecraft at a given time (the epoch time). An example is the osculating Keplerian elements.

Orthogonal Two vectors are orthogonal if they are perpendicular to each other.

Orthonormal Matrix A square matrix whose inverse is equal to its transpose. (A transpose matrix swaps column and row entries relative to a given matrix.)

Osculating Orbital Elements A set of Keplerian orbital elements that give the correct spacecraft position and velocity at a given instant in time.

Out-of-Plane Orbit Maneuvers Orbit maneuvers that change the orientation of the orbit plane by changing the inclination or ascending node.

Overdamped Control System Damping in the control law that causes the system to approach steady state asymptotically as time goes to infinity.

Overshoot Control law behavior where the system response exceeds steady state, leading to oscillatory behavior.

Parabola A conic section whose eccentricity is equal to 1.

Parallax A shift in the direction of an observed object when the platform (i.e., spacecraft) observing it moves.

Parallels Small circles on the celestial sphere that lie in planes parallel to the equatorial plane.

Parameterization Characterization of a system, phenomenon, or individual piece of hardware by a set of variables that plug into a model associated with the system, phenomenon, or piece of hardware.

Parking Orbit The initial orbit into which a spacecraft is launched. Later orbit maneuvers by the second stage of the launch vehicle, an attached motor, and or the spacecraft's own propulsion system then place the spacecraft in its mission orbit.

Parsec The distance for which stellar parallax is 1 arcsec, given a shift by the observer of 1 Astronomical Unit (AU). It is equal to 3.26 light-years, or 3.086×10^{13} km.

Passive Nutation Damper Hardware that removes spacecraft motions about other than the major principal axis by converting the undesired rotational energy to heat.

Pattern Match Identifying star tracker stars by associating the stars observed in the star tracker with a similar pattern of stars in the star catalog by comparing relative angular separations.

Peak Time In response to step input, the time required for a control system to reach its maximum output value.

Penumbra Part of the eclipse cone where only part of the Sun is blocked.

Pericenter General term for the closest approach of an orbiting body to its parent body.

Perifocal Distance Separation between the barycenter and the point of closest approach.

Perifocus Point of closest approach to the barycenter.

Perigee Point of closest approach to the Earth for an Earth orbit.

Perigee Rotation For an Earth orbit, rotation of the line of apsides due to orbital perturbations caused by Earth oblateness.

Perihelion Point of closest approach to the Sun for a Solar orbit.

Period An orbital period is the time it takes a spacecraft (or celestial object) to complete an orbit.

Perturbation Displacements of attitude or orbit from their "true" values due to external influences of lesser magnitude to the dominant factor. For example, at low Earth altitudes, atmospheric drag is a major perturbation to an ideal two-body orbit.

Phase Angle See *Azimuth Angle*.

Pitch Relative to the body frame, a right-handed rotation about the spacecraft y-axis. Relative to orbit coordinates, a right-handed rotation about negative orbit normal.

Plane Change An orbit maneuver that changes the orbital plane (e.g., inclination change or nodal change maneuvers).

Plant The physical object being controlled.

Plume The thruster exhaust envelope.

Plume Impingement Intersection of the thruster plume with a spacecraft component. Plume impingement can result in undesired perturbative torques being applied to the spacecraft.

Pole A zero in the denominator of a transfer function.

Potential Energy Energy stored in a system. For example, an object at rest above the surface of the Earth has (near the Earth's surface) a potential energy equal to its mass times the gravitational acceleration times it's height. If the object is released, its potential energy gets converted to kinetic energy (ignoring frictional effects).

Power Energy per unit time.

Precession The change in direction (but not magnitude) of the spacecraft angular momentum vector due to application of a torque.

Precession of the Equinoxes Due to gravitational perturbations by the Sun and Moon on the Earth's equatorial bulge, the Earth's spin axis rotates about the ecliptic pole, which causes the directions of the Autumnal and Vernal Equinox to rotate in inertial space (i.e., relative to the "fixed" stars).

Predicted Orbit Extrapolated spacecraft position and velocity (as a function of time) determined by applying mathematical models of physical processes (e.g., the Earth's gravitational potential) to actual measurements of the spacecraft orbit at an earlier time.

Pressure Force divided by the area over which the force is applied.

Prime Meridian The 0° longitude meridian, used as a reference for defining longitude or azimuth angles.

Principal Moment of Inertia The moment of inertia associated with a principal axis.

Principal Spacecraft Axes A body axis is a principal axis if rotations about that axis yield an angular momentum vector that is parallel to the axis. This is true of a spacecraft's natural axes of symmetry.

Prograde Orbit For an Earth orbit, an orbit whose sense of rotation is in the same sense as the Earth's rotation.

Prolate The shape of a spheroid having a polar radius greater than its equatorial radius (i.e. "football-shaped").

Propagation For orbits, extrapolation of a spacecraft position and velocity from a known starting value. For attitudes, use of relative attitude information (from gyros) to extrapolate spacecraft absolute pointing from a starting, measured absolute attitude (for example, from star tracker data).

Propellant Thruster fuel.

Propellant Slosh For liquid-fueled thrusters, propellant motion inside the fuel tank. If the fuel tank(s) is (are) not properly balanced, propellant slosh can produce undesired disturbance torques. When centered properly about the spin-axis of a spin-stabilized spacecraft, propellant slosh serves as a nutation damper as it will convert motions that are not about the spin axis into heat, thereby removing those undesired motions.

Proper Motion In star catalogs, the individual motion of an individual star.

Proportional Control Law A control law with one term, namely one proportional to the coordinate being controlled. Without additional terms to moderate the system's performance, the control law will directly track all measured noise and errors, producing large undesirable system oscillations.

Proportional-Derivative (PD) Control Law A control law with two terms: a proportional (or position) term and a derivative (or rate) term. Each term has its own adjustable gain. The rate term damps out oscillations generated by the proportional term.

Proportional-Integral (PI) Control Law A control law with two terms: a proportional (or position) term and an integral term. Each term has its own adjustable gain.

Proportional-Integral-Derivative (PID) Control Law A control law with three terms: a proportional (or position) term, an integral term, and a derivative (or rate) term. Each term has its own adjustable gain.

Pseudo-inverse See *matrix pseudo-inverse*.

Q Method/Quest A batch, least-squares algorithm created by Paul Davenport (GSFC) and refined by Malcolm Shuster (CSC) that computes statistically optimal three-axis attitudes from vector observations (for example, star tracker input).

Quantization Error Measurement error due to the finite size of the least significant bit (LSB).

Quarter Orbit Coupling Geometric aspect exploited when performing attitude determination with Earth sensors. A roll error at one point in the orbit becomes a yaw error of 90° (a quarter of an orbit) later.

Quaternion A parameterization of the three pieces of information supplied by Euler angles in terms of an ordered set of four real numbers. The quaternion lacks the singularity issues present in Euler angle formulations and is more compact than the nine elements of the DCM. Quaternions also are easily manipulated to determine (or incorporate) changes in attitude, a useful feature when supporting attitude slews.

Quaternion Inversion The quaternion inverse is obtained by negating the first three elements (in one popular formulation) of the original quaternion. The product of the quaternion and its inverse is the identity quaternion, (0, 0, 0, 1) (in that formulation).

Quaternion Multiplication A mathematical algorithm for combining the rotational instructions of two quaternions.

Radial Error Error in spacecraft position with respect to distance from the focus.

Radiation Pressure Force per unit area applied to the spacecraft body by incident Sunlight. When a spacecraft's center-of-mass and center-of-pressure are not coincident, solar radiation pressure will generate a torque on the spacecraft.

Random Error The precision error or error due to measurement fluctuation.

Range and Range Rate Orbit Measurements Measurements of a spacecraft's distance and radial velocity used to determine the spacecraft's orbit.

Rate Gyroscope Provides a direct measurement of the spacecraft angular velocity.

Rate-Integrating Gyroscope Measures spacecraft angular displacements over a brief, fixed time interval. Division by the time interval yields the spacecraft angular velocity.

Reaction Wheel A flywheel that stores angular momentum. Reaction wheels are attitude actuators that can be commanded to null spacecraft pointing

errors or slew the spacecraft to a new commanded attitude. Because the angular momentum of a closed system is conserved, changing the rotation rate of a spacecraft's reaction wheel (and thereby changing the angular momentum of the reaction wheel) will initiate an opposite movement in the spacecraft body so as to keep the total angular momentum of the entire system constant.

Recursive Estimator A state estimation process where the state vector is updated after each new observation is received.

Redundancy Built-in back-up capabilities provided to ensure that spacecraft functionality (in that area) will be retained in the event of a spacecraft hardware failure or anomaly. Spacecraft may be fully redundant, partially redundant, or single string, where single string spacecraft (in order to minimize mission costs) provide no redundancy in the event of an inflight hardware problem.

Reference Data Inertial frame information from a model, catalog, etc. that can be combined with attitude observations in the body in order to determine the spacecraft orientation.

Reference Vector The predicted direction of the observation vector in the inertial frame.

Regression of Nodes A rotation of the line of nodes due to orbital perturbations caused by Earth oblateness. When the spacecraft inclination is less than 90°, the line of nodes rotates opposite to the spacecraft orbital rotation, and the nodal rotation is called regression of the nodes.

Residual The difference between the measured quantity and the modeled quantity.

Resolution A quantitative measure of the capacity of a sensor or science instrument to distinguish between two nearby observable objects.

Resonance Frequency Input frequency (associated with an external force driving a control system) yielding maximum output amplitude.

Retrograde Orbit For an Earth orbit, an orbit whose sense of rotation is in the opposite sense as the Earth's rotation.

Revolution A complete orbit.

Right Ascension Angle (as measured on the Earth's equatorial plane) between the vernal equinox and a given point on the equator.

Right Ascension of the Ascending Node The angle between the vernal equinox and the orbit's ascending node, measured in the Earth's equatorial plane.

Rigid Spacecraft Motion The rotational behavior of spacecraft (in the absence of spacecraft flexible components) due to the influence of external torques.

Rise Time For step input, the time required to attain a response equal to the input magnitude.

Robust Control A control system capable of delivering desired levels of performance even in the presence of significant, unexpected plant behavior.

Roll In celestial coordinates, the Euler angle specifying the rotation angle of the spacecraft body about a specific body axis (for celestial pointers, usually the boresight of the prime imaging instrument). Relative to the body frame, a right-handed rotation about the spacecraft x-axis. Relative to orbit coordinates for circular orbits, the rotation angle about the spacecraft velocity axis.

Root A pole (or zero in the denominator of a transfer function).

Root Locus Diagram Plot of a control system's poles (or roots) in the complex plane.

Rossi, Bruno (1905-1993) Italian-American experimental physicist.

Rossi X-Ray Timing Explorer (RXTE) A NASA satellite used to study X-Ray sources.

Rotation Angular motion of a spacecraft about a vector in the body frame.

Round-Off Error Error due to finite word length of software variables.

Row Matrix A $1 \times N$ matrix, where N is the number of entries in the matrix. Row matrices can be used to represent vectors. A matrix with only one row.

Runga Kutta Method A numerical integration algorithm. Developed around 1900 by the German mathematicians C. Runge and M. W. Kutta.

RXTE See *Rossi X-Ray Timing Explorer*.

Safemode In ACS FSW applications, also called Safehold Mode. Safemode is a spacecraft control mode designed to guarantee electrical power, and to meet thermal and bright object constraints, in the extended absence of contact with the ground. Most spacecraft achieve these capabilities via a Sun-pointing Safehold Mode.

Saturation Limit The highest speed at which a reaction or momentum wheel can be operated.

Scalar A number, as opposed to a vector. A scalar has a magnitude; a vector has both a magnitude and a direction.

Scalar Product See *Dot Product*.

Sectoral Harmonic Coefficients Coefficients of terms in a spherical harmonic expansion for which $m = n$ (cf. Eq. 6.1). Graphically, sectoral patterns are bands of longitude.

Secular Terms Terms in an expansion that have a net change over long time periods, as opposed to being periodic or constant with time.

Semimajor Axis Half the length of an ellipse's long axis.

Semiminor Axis Half the length of an ellipse's short axis.

Sensitivity A measure of how output response is related to changes in input. If small changes in an input produce large changes in output, the system has a high sensitivity to that input.

Sensor In an ACS context, a measuring device whose output can be used to determine the spacecraft's attitude, either in absolute or relative terms.

Settling Time In control theory, the time required for the system oscillations to reduce in amplitude to a specified band about the steady-state solution.

Short Period Variations In an orbital context, fluctuations in a parameter's value during an orbital period.

Sidereal Time Time measurement derived from the Earth's rotation relative to the "fixed" stars.

Single-Axis Attitude For spin-stabilized spacecraft, the orientation of the spacecraft's spin axis. Can be parameterized by two Euler angles, right ascension and declination.

Single-Axis Stabilized Typically, a spin-stabilized spacecraft. Gyroscopic stiffness helps the spacecraft to resist disturbance torques that would otherwise cause that axis's orientation in inertial space to fluctuate or drift.

Single-degree-of-freedom Gyroscope A gyro sensitive to motion in only one direction.

Single-String Spacecraft A spacecraft with no built-in hardware back-up capability designed to deal with an inflight failure or anomaly in spacecraft hardware functionality.

Singularity Mathematically, a divide-by-zero condition. Often caused by trying to solve for more information (parameters) than available measurements can support.

Singular Matrix A square matrix for which an inverse cannot be defined.

Singularity Condition A geometric state for which one or more coordinates become undefined. For example, when the declination is 90°, the right ascension becomes undefined, yielding a singularity condition.

Skew-Symmetric Matrix A square matrix whose transpose is the negative of the original matrix.

SKYMAP A star catalog designed to support attitude determination.

Slew A large attitude maneuver.

Solar-Lunar-Planetary (SLP) Ephemeris File Originally a JPL product adapted by GSFC for attitude and orbit applications, the SLP file supplies position information for the Sun, Moon, and planets.

Small Circle A circle on the surface of a sphere where the plane of the circle does not pass through the center of the sphere.

Smoothing Reducing random noise from measured data or computed parameters.

Solar Radiation Pressure Force per unit area applied to the spacecraft body by incident sunlight. When a spacecraft's center-of-mass and center-of-pressure are not coincident, solar radiation pressure will generate a torque on the spacecraft.

Solar Radiation Torque A torque induced on a spacecraft by incident Sunlight when the spacecraft's center-of-mass and center-of-pressure are not coincident.

Solar Time Time measurement derived from Earth rotation relative to the Sun.

South Atlantic Anomaly (SAA) A region of diminished magnitude in the geomagnetic field. The SAA is located off the coast of Brazil. Because of the reduced field strength in this region, proton and electron flux energies (as a function of altitude) are higher in the SAA.

Spherical Harmonic Expansion A polar coordinates expansion. The surface spherical harmonics are expressed as polynomials in terms of powers of sines and cosines of azimuth and elevation angles.

Spin Rotation about a spacecraft axis.

Spin Axis The axis of rotation.

Spin Axis Precession Changing the direction of the axis of rotation.

Spin Rate Speed of rotation about the spin axis.

Spin Stabilization Exploitation of gyroscopic stiffness to maintain the commanded attitude in the presence of disturbance torques.

Spoiler For missions utilizing guide stars to determine and maintain attitude, a star (nearby the actual guide star) that could be detected and tracked instead of

the desired star. Depending on the number of guide stars, acquisition of a spoiler star can introduce large attitude errors.

Square Matrix A matrix having the same number of rows as columns.

Stable System When bounded input yields bounded output.

Star Catalog A file containing a set of data associated with each star in the catalog. The associated data includes unique ID, right ascension, declination, and magnitude. Star catalogs are used to support star identification and provide the reference data required for use of star tracker (or fine error sensor) measurements in attitude determination.

Star Identification Association of stars detected in a star tracker (or fine error sensor) with catalog stars.

Star Tracker A star detecting device used for attitude determination and control.

State Estimation Formal, statistical method for determining values of a set of observable quantities called a state vector.

State Transition Matrix A mathematical object which when operating on the known values of a state at a given time predicts what the values of the state should be at a specific (typically) future time.

State Vector The set of parameter values that describe the system.

Station Keeping Maneuver An orbit maneuver performed to null errors in the actual orbit relative to the desired orbit.

Steady State In (for example) Kalman filter processing, when changes in the computed state vector no longer vary much from time step to time step and the error covariance matrix is stable.

Steady-State Error The difference between the actual control law output and the desired control law output after the transient response has died off.

Stellar-Pointer A spacecraft that conducts its science by pointing a science instrument (or instruments) at targets on the celestial sphere, i.e., stars, galaxies, etc.

Stiction The larger frictional force (between two surfaces) "felt" up until one of the objects starts to move relative to the other one.

Strapdown Torque Rebalanced Gyroscope A mechanical gyro for which the gyro gimbal is torqued to maintain near-zero deflections. The torques are applied by magnetic fields created by commanded currents. Therefore, measuring the current required to keep the gimbal position near null measures the attitude change of the spacecraft in which the gyro is mounted.

Subsatellite Point For Earth-orbiting spacecraft, the point on the Earth towards which the nadir vector points.

Sun Angle Angle between the Sun vector and a specific vector in the body frame (for example, the spacecraft spin axis).

Sun-Pointer A spacecraft whose attitude is maintained so as to keep its science instrument(s) oriented towards the Sun.

Sun Presence Detector An attitude sensor that outputs a value above some threshold when the Sun is in its field of view (FOV).

Sun Sensor An attitude sensor that detects the presence of the Sun and, in the case of analog and digital Sun sensors, also outputs measurements from which the Sun vector or Sun angle can be computed.

Sun Shade A device attached to a star tracker designed to prevent the entrance of Sunlight into the star trackers light detecting electronics.

Sun-Synchronous Orbit A spacecraft orbit whose plane rotates at the same rate as the Earth's orbital rate about the Sun. That way, the spacecraft orbital plane motion matches the apparent motion of the Sun (relative to the Earth). Sun-synchronous spacecraft imaging the Earth will therefore (for example) always image points on the Earth's equator at the same local time.

Symmetric Matrix A square matrix whose transpose is equal to the original matrix.

Synchronous Satellite For Earth orbiting spacecraft, a spacecraft whose orbital motion matches the Earth's rotational motion. If the spacecraft inclination is zero, the spacecraft will appear to hover over the same point on the Earth's surface.

Systematic Error Non-fluctuating error arising from a hardware inadequacy or flaw, a faulty calibration, etc. Systematic errors, unlike random errors, can be measured and compensated for via improvements to calibrations, models, etc.

Tachometer Devices that measure the speeds of reaction wheels and momentum wheels.

Terminator The day/night boundary on a planet or moon.

Tesseral Harmonic Coefficients Coefficients of terms in a spherical harmonic expansion for which $m \neq n$ (cf. Eq. 6.1). Graphically, tesseral patterns are mosaic-like "tiles" along bands of longitude.

Thermistor A resistor whose resistance is temperature dependent, and thereby serves as a thermometer.

Three-Axis Attitude A specification of the orientation of the spacecraft body axes with respect to a reference frame, typically expressed as a transformation from the GCI reference frame to the body frame.

Three-Axis Magnetometer (TAM) An attitude sensor that measures the magnitude and direction of the magnetic field in which the spacecraft is immersed.

Three-Axis Stabilized Spacecraft all of whose body axes are controlled, as compared with spin-stabilized spacecraft which only control the orientation of the spin axis.

Three-Body Interactions Orbit "problems" involving three objects (e.g., solving for the moon's orbit about the Earth, as influenced by the Sun).

Thruster Spacecraft hardware that generates thrust by expelling propellant.

Torque The rate of change of the angular momentum. Torques are generated when forces are applied to a body other than through the body's center-of-mass. Such forces introduce rotations about the object's center-of-mass.

Torque-Free Motion Rotational motion for which the angular momentum is conserved (i.e., unchanged).

Trace The sum of a matrix's diagonal elements.

Tracking and Data Relay Satellite (TDRS) A communications satellite in the TDRSS. TDRSs are approximately geostationary spacecraft. TDRS spacecraft are used as relays to uplink commands and table loads from the ground station to its spacecraft, and are used as relays to downlink telemetry from the spacecraft

to the ground station. In the future, spacecraft constellations may utilize TDRS spacecraft as relays to support intra-constellation communications.

Tracking and Data Relay Satellite System (TDRSS) The collection of TDRSs, their ground stations, etc.

Trajectory A spacecraft's orbital path.

Transfer Function The Laplace transform of a system's output divided by the Laplace transform of the system's input.

Transfer Orbit Intermediate orbit used as a stepping stone to get to mission orbit.

Transient Response Control system response as a function of time.

Transpose Matrix A transpose matrix swaps column and row entries relative to a given matrix.

TRMM Tropical Rain Measuring Mission, a NASA satellite.

True Anomaly Angle measured at the primary focus between perigee and spacecraft position.

True of Date When defining coordinates of directions in the GCI reference frame, taking into account both the precession and nutation of the Earth's spin axis relative to inertial space.

Tumbling Uncontrolled rotational motion.

Two-Body Interactions Orbit problems where the mutual gravitational effects of only two objects are considered. For example, approximating the behavior of a spacecraft in low Earth orbit by only considering the gravitational influence of the Earth on the spacecraft.

Two-Degree-of-Freedom Gyroscope A gyro sensitive to motion along two axes.

Ultraviolet (UV) Light Light roughly at wavelengths between 100 and 3900 angstroms. Ultraviolet light is produced by excited atoms.

Umbra Part of the eclipse cone where the whole Sun is blocked.

Unstable System When output increases without bound.

Uplink Command transmission from the ground to the spacecraft.

UTC See Coordinated Universal Time.

Vector (1) A quantity with direction and magnitude. (2) See "state vector" for alternative use of term.

Velocity The rate of change of position.

Velocity Aberration Apparent shift in the direction from which an object is radiating light due to the relative lateral velocity between the radiating object and the observing platform.

Vernal Equinox The time when (in the Sun's apparent orbit about the Earth) the Sun crosses the Earth's equatorial plane as it moves from the Southern hemisphere to the Northern hemisphere. The time of this crossing is defined to be the first day of Northern hemisphere Spring.

Visible Light Light observable by the naked eye, roughly at wavelengths between 3900 and 7600 angstroms.

Visual Magnitude See *Magnitude*.

Weight A quantitative specification of the relative importance of a given observation.

Weight Matrix A square matrix used to assign quantitatively different levels of importance to different observations.

White Noise Uncorrelated noise.

Wilkinson, David (1935-2002) American cosmologist.

Wilkinson Microwave Anisotropy Probe (WMAP) NASA satellite used to study the microwave background radiation.

WMAP See *Wilkinson Microwave Anisotropy Probe*.

X-rays Light roughly at wavelengths between 0.1 angstroms and 100 angstroms.

Yaw Relative to the body frame, a right-handed rotation about the spacecraft z-axis. Relative to orbit coordinates, a right-handed rotation about nadir vector.

Zenith The vector anti-parallel to the nadir vector (i.e., directly above the spacecraft as opposed to directly below).

Zenith-Pointer A spacecraft whose primary science instrument(s) is (are) directed above local vertical (i.e., along the zenith vector).

Zonal Harmonic Coefficients Coefficients of terms in a spherical harmonic expansion for which the second index of the Legendre polynomials is zero (cf. Eq. 6.1). Graphically, zonal patterns are bands of latitude.

References

1. Arfkin G, Weber H, Harris F (2013) Mathematical methods for physicists, 7th edn. Academic Press Inc., New York
2. Bate RR, Mueller DD, White JE (1971) Fundamentals of astrodynamics. Dover Publications, New York
3. Battin RH (1987) An introduction to the mathematics and methods of astrodynamics. AIAA education series, AIAA, New York
4. Black H (1964) A passive system for determining the attitude of a satellite. AIAA J 2(7):1350–1351
5. Butterworth S (1930) On the theory of filter amplifiers. Wirel Eng 7:536–541
6. Chetty P (1991) Satellite technology and its applications. TAB Professional and Reference Books, Blue Ridge Summit
7. Dorf RC, Bishop RH (1995) Modern control systems. Addison-Wesley Publishing Company, Reading
8. Franklin GF, Powell JD, Emami-Naeini A (1991) Feedback control of dynamic systems. Addison-Wesley Publishing Company, Reading
9. Gelb A (ed) (1974) Applied optimal estimation. MIT Press, Cambridge
10. Goldstein H (1950) Classical mechanics. Addison-Wesley Publishing Company, Reading
11. Hamilton W (1844–1850) On quaternions, or on a new system of imaginaries in algebra. Lond Edinb Dublin Philos Mag J Sci 25–36
12. Hamming RW (1962) Numerical methods for scientists and engineers. McGraw-Hill Book Company Inc., New York
13. Harris I, Priester W (1962) Time-dependent structure of the upper atmosphere. J Atmos Sci 19:286–301
14. Kalman RE (1960) A new approach to linear filtering and prediction problems. J Basic Eng 82:35–46
15. Keat J (1977) Analysis of least-squares attitude determination routine DOAOP. Technical report CSC/TM-77/6034, Computer Science Corporation
16. Lagrange J-L (1772) An essay on the three-body problem
17. Larson WJ, Wertz JR (eds) (1992) Space mission analysis and design. Microcosm, Inc. and Kluwer Academic Publishers, Torrance and New York
18. Lawrence A (1998) Modern inertial technology navigation, guidance, and control. Springer, Berlin
19. Lefferts EJ, Markley FL, Shuster MD (1982) Kalman filtering for spacecraft attitude estimation. J Guid Control Dyn 5(5):417–429

© Springer-Verlag London Ltd. 2017

H.L. Hallock et al., *ACS Without an Attitude*, NASA Monographs
in Systems and Software Engineering, DOI 10.1007/978-1-4471-7325-0

20. Long AC, Cappellari JJO, Velez CE, Fuchs AJ (1989) Goddard trajectory determination system (GTDS) mathematical theory, revision 1. Technical report, NASA/GSFC Flight Dynamics Division Code, p 550
21. Markley LF, Crassidis JL (2014) Fundamentals of spacecraft attitude determination and control. Springer Science and Business Media, New York
22. Newton I (1999) Philosophiae Naturalis Principia Mathematica (Mathematical principles of natural philosophy), 1687. cf. the translation by I. Cohen and A. Whitman, University of California Press, Berkeley, Los Angeles, London
23. Ogato K (1997) Modern control engineering. Prentice-Hall, Englewood Cliffs
24. Roberts C (1971) An analytic model for upper atmosphere densities based upon Jacchia's 1970 models. Celest Mech 4(3):368–377
25. Shuster M, Oh S (1981) Three-axis attitude determination from vector observations. J Guid Control 4(1):70–77
26. Wahba GA (1965) Least-squares estimate of satellite attitude. SIAM Rev 7(3):409
27. Welter G, Boia J, Gakenheimer M, Kimmer E, Channel D, Hallock L (1996) Variations on the Davenport gyroscope calibration algorithm. In: Proceedings of the GSFC flight mechanics/estimation theory symposium
28. Wertz JR (ed) (1978) Spacecraft attitude determination and control. Kluwer Academic Publishers, Berlin
29. Wertz JR, Larson WJ (1996) Reducing space mission cost. Microcosm, Inc. and Kluwer Academic Publishers, Torrance and New York
30. Wilson G (1997) If you survive: from Normandy to the battle of the bulge to the end of World War II, one American officer's riveting true story. Paperback Edition: Ballantine Books

Index

© Springer-Verlag London Ltd. 2017

H.L. Hallock et al., *ACS Without an Attitude*, NASA Monographs in Systems and Software Engineering, DOI 10.1007/978-1-4471-7325-0