

VIRGINIA POLYTECHNIC INSTITUTE
AND STATE UNIVERSITY

ORBITAL LAUNCH VEHICLE TEAM

TRAJECTORY AND AERODYNAMIC ANALYSIS DOCUMENT

Rocket Trajectory Simulation

M. WERNER



Summer – Fall 2020

CONTENTS

1. <i>Time Systems and Transformations</i>	4
1.1 Defining Time	4
1.2 Complications of the Day	4
1.2.1 The Solar Day	4
1.2.2 The Sidereal Day	5
1.3 Calendars	5
1.3.1 The Julian Calendar	5
1.3.2 The Gregorian Calendar	5
1.4 Measuring the Passage of Time	6
1.4.1 Time Systems	6
1.4.2 Julian Dates	7
2. <i>Coordinate Systems and Transformations</i>	9
2.1 Earth-Centered Inertial (ECI) Reference Frame	9
2.2 Earth-Centered Earth-Fixed (ECF) Reference Frame	11
2.3 Spherical Coordinates	13
2.4 Ellipsoidal Coordinates	16
2.4.1 Reduced Latitude	17
2.4.2 Geocentric Latitude	18
2.4.3 Geodetic Latitude	20
3. <i>Earth Model</i>	25
3.1 Reference Ellipsoid	25
3.2 Gravitational Model	26
3.2.1 Newtonian Gravity (Smooth, Nonrotating Sphere)	26
3.2.2 Normal Gravity (Smooth, Rotating Spheroid)	27
3.2.3 Earth Gravity (Rough, Rotating Spheroid)	29
3.3 Magnetic Model	37
3.3.1 World Magnetic Model 2020	38
3.4 The Geoid	40
3.4.1 Determining Height Above Mean Sea Level	40
3.4.2 Accuracy and Resolution	42
3.4.3 Results	42
3.5 Atmospheric Model	46
3.5.1 Gas Properties	46
3.5.2 Layers of the Atmosphere	50
3.5.3 Sourcing a Standard Atmosphere	52
3.5.4 Jet Standard Atmosphere	53
3.5.5 1976 U.S. Standard Atmosphere	54
3.5.6 Reference Atmospheric Models	56
3.5.7 NRLMSISE-00 Atmospheric Model	57

4. Structures	58
4.1 Determination of Characteristics for Fundamental Geometries	59
4.1.1 Nose Cone Configurations	60
4.1.2 Body Configurations	64
4.1.3 Cavity Tuning	66
5. Propulsion	67
5.1 Quasi-Unidimensional Nozzle Flow	68
5.1.1 Subsonic Flow	68
5.1.2 Supersonic Flow	69
5.2 Basic Grain Geometry	72
5.2.1 BATES Grain	73
6. Aerodynamics	75
6.1 Lift	75
6.2 Drag	75
6.2.1 Flow over a Cone	76
6.2.2 Flow over a Circular Cone at Zero Angle of Attack	76
6.2.3 Second-Order Shock-Expansion Theory — Relation to Curved Nose Cones	81
6.2.4 Evaluating the Pressure Drag at Zero Angle of Attack	83
7. Phases of Flight	84
7.1 Prelaunch	84
7.1.1 Initial Conditions	84
7.1.2 Defining the Body Frame	85
7.2 Rail-Guided Flight	86
7.2.1 Fully Rail-Guided Flight	86
7.2.2 Partially Rail-Guided Flight	87
7.3 Free-Flight	88
Appendix	94
A. Mathematical Background of Coordinate Transformations	95
A.1 Vector Spaces	95
A.2 Coordinate Transformations	96
A.3 Notation	98
A.4 Elemental Rotations	99
B. Mathematical Background of Spherical Harmonics	101
B.1 Function Spaces	101
B.2 Spherical Harmonics — switch places with Legendre polynomials	101
B.3 Associated Legendre Polynomials — Switch Places with Sphr Harmonics	103
C. Fluid Dynamics (Possibly Remove Entirely)	105
C.1 Conservation Laws	105
C.2 Similarity Parameters	105
C.3 Flow Regimes	106
C.3.1 Compressibility	106
C.3.2 Rotationality	106
C.3.3 Viscosity	107
C.3.4 Turbulence	107

1. TIME SYSTEMS AND TRANSFORMATIONS

Many various methods and systems are used throughout the world to track the passing of time. Such systems originated and are influenced by local culture, history, convenience, and formal scientific need. Common and scientific time systems and how they relate to one another will be discussed in detail.

1.1 *Defining Time*

Regardless of the time system used to express the time, the *passage* of time, or the difference between two times, can always be measured in the fundamental unit of seconds. One second is defined by the International System of Units (Système International d'unités) to be the amount of time it takes the Cesium-133 (atomic number 55) isotope at absolute zero temperature to undergo exactly 9,192,631,770 periods of radiation produced by transitions between the two hyperfine levels of its unperturbed ground state. This definition provides the foundation for the atomic clock.

Additional, commonly used terms apart from the second are recognized, though not officially designated, by the International System of Units to mark the passage of time. One minute is recognized as the passing of 60 seconds and one hour is recognized as the passing of 60 minutes. Though more complicated than the minute and the hour, one day is accepted to denote the passing of 24 hours (equivalently 86,400 seconds).

1.2 *Complications of the Day*

1.2.1 *The Solar Day*

The common time system that most people use regularly operates on the 24-hour, local time zone clock, in which the sun appears roughly overhead at noon at a given point on Earth and is roughly beneath that same location at midnight. This time system is closely related to (but is not the same as) the *solar day*.

The solar day, or apparent solar day, is defined by the amount of time it takes for the sun to return once to its position in the sky from a given datum, or epoch time. Because Earth is tilted and moving in an elliptical orbit about the sun, each solar day is not the same length of time as the one preceding it. These differences are small, on the order of seconds apart, and indeed periodic throughout successive orbits of Earth about the sun; however, they do accumulate to reach a maximum deviation about 15 minutes from the average length of a solar day. For common time-keeping purposes, the apparent solar day is rather unwieldy as it requires continual adjustment to what defines the length of its day — something that is easily handled by computers but historically is not easily handled by people.

The *mean solar day* is therefore defined to provide a constant amount of time defining the length of a day and of course is the average length of the apparent solar day. This mean solar day, or just day, is the 24-hour, local time zone clock commonly used by people everywhere that is widely accepted as “the time.”

Local time zones are referenced by an hour offset to the historically- and arbitrarily-chosen time in Greenwich, England, whose time is known as Greenwich Mean Time (GMT). The offset spans, usually in increments of one hour between adjacent time zones, from twelve hours behind

(GMT-12) to fourteen hours ahead (GMT+14) of the local time in Greenwich. Conversions between common time zones largely in the northwestern world and slightly into the northeast are summarized below.

Tab. 1.1: Common time zone conversions with daylight savings found in America and Europe.

Time Zone	Offset (Hours)	Name	Example Location
PST/PDT	GMT-8/-7	Pacific Standard/Daylight Time	Los Angeles, California
MST/MDT	GMT-7/-6	Mountain Standard/Daylight Time	Denver, Colorado
CST/CDT	GMT-6/-5	Central Standard/Daylight Time	Houston, Texas
EST/EDT	GMT-5/-4	Eastern Standard/Daylight Time	Washington D.C.
GMT	GMT	Greenwich Mean Time	Greenwich, England
CET/CEST	GMT+1/+2	Central European (Summer) Time	Berlin, Germany
EET/EEST	GMT+2/+3	Eastern European (Summer) Time	Helsinki, Finland

1.2.2 The Sidereal Day

A different measure for the length of a day, which is more for scientific than common time-keeping purposes, is the *sidereal day*. One sidereal day is the amount of time it takes for Earth to complete one rotation about its axis relative to distant stars. While this amount of time is not constant because of the precession and nutation of Earth's spin axis, a widely accepted value is currently 86,164.0905 seconds. Consequently, the rotational rate of Earth is taken as

$$\omega = 0.000\,072\,921\,158\,579 \text{ [rad/s]}. \quad (1.1)$$

Note that the sidereal day is a pure number and has no attachment to any time zones or particular location on Earth. As a notational standard, this symbol ω is sometimes shown with the subscript \oplus or \otimes to emphasize reference to the earth itself if ambiguity arises from the presence of other angular velocity symbols. Any ambiguity should be resolved by context otherwise.

1.3 Calendars

Predicted dates are collected into the familiar modern groups of a week (7 days), a month (several weeks and some days), a year (12 months), a decade (10 years), and so on. These groups form a calendar. Historically, two such calendars have been used extensively throughout time.

1.3.1 The Julian Calendar

The Julian calendar was first introduced by Julius Caesar and officially established on January 1, 45 B.C. Unlike the preceding Roman calendar, the Julian calendar boasts 12 months with each year consisting of 365.25 days on average, each day consisting of 86,400 seconds. This fractional representation of an additional quarter day per year indicates that every year evenly divisible by 4 is a *leap year*, which is a year containing an extra day than usual. As such, every year in the Julian calendar is 365 days except every year evenly divisible by 4, which is the leap year and is 366 days.

1.3.2 The Gregorian Calendar

Most countries switched from the Julian calendar to the Gregorian calendar, introduced by Pope Gregory XIII in October 1582 (October 5 by the Julian calendar but October 15 by the Gregorian calendar), before the early 20th century. Unlike the Julian calendar, the Gregorian calendar exhibits a year consisting of 365.2425 days on average (faster than the Julian calendar by

about 11 minutes per year), each day consisting of 86,400 seconds. Every year in the Gregorian calendar is 365 days except every leap year, which is 366 days. Because of the more complicated fractional representation of days in a year compared to the Julian calendar however, the determination of which years are leap years is accordingly more complicated. In the Gregorian calendar, a leap year occurs when the year is evenly divisible by 4 except when also evenly divisible by 100 unless also evenly divisible by 400.

1.4 Measuring the Passage of Time

1.4.1 Time Systems

Many scientific applications require a more precise meaning and standardized form to represent the passing of time. The introduction of precisely-measured time systems (as opposed to time zones) permeating throughout the world simultaneously meet these demands. However represented, though, time systems should connect to the common time zones used by all.

Universal Time (UT or more commonly UT1) is introduced to begin connecting time systems to time zones as

$$\text{UT1} = \text{GMT}. \quad (1.2)$$

The choice of GMT serving as the time for which the entire world can reference is as arbitrary as was choosing GMT to serve as the defining time zone for others to reference an offset, which is what makes it the most convenient choice of all.

The atomic clock naturally stands out as the lead candidate to address precise amounts of passing time. The Navigation and Ancillary Information Facility (NAIF) at JPL provides that International Atomic Time (TAI) is a statistical measure from about 200 atomic clocks of the amount of seconds (as defined by the International System of Units) past January 1, 1958 00:00:00 UT1, the precise moment of midnight in Greenwich, England.

Coordinated Universal Time (UTC) links the common time zone representation of UT1 with the precise time measurement of TAI with the introduction of *leap seconds*. Since UT1 operates on the constant 24-hour mean solar day, which does not account for the nonuniformity of Earth's rotation (explained later), leap seconds must be occasionally introduced into UTC to account for accumulated differences between UT1 and TAI. Consequently, the time difference between UT1 and UTC, denoted DUT1, always satisfies

$$\text{DUT1} < 0.9 \text{ [s]}. \quad (1.3)$$

In other words, UT1 and UTC are kept within one SI second of each other. For applications not requiring extreme precision, these two time systems are essentially indistinguishable.

There was exactly a 10 second difference between UTC and TAI (with UTC behind TAI) before the first leap second was added on December 31, 1971 11:59:60 UTC. As of December 30, 2020, a total of 27 leap seconds have been added to UTC. Accordingly, the relation between TAI and UTC after January 1, 1972 00:00:00 UTC follows

$$\text{TAI} = \text{UTC} + (10 + \text{current total of leap seconds}) \text{ [s]}. \quad (1.4)$$

It is noted, however, for long-term applications requiring extreme precision, the injection of leap seconds, which are always added on either June 30 or December 31 of the current year, is not a predictable process.

The International Astronomical Union defines a standardized time system utilized when taking astronomical measurements from Earth's surface called Terrestrial Dynamical Time (TDT), equivalently called Terrestrial Time (TT). Per NAIF, TT is the "proper" time on Earth at sea level and is related to atomic time via

$$\text{TT} = \text{TAI} + 32.184 \text{ [s]}. \quad (1.5)$$

The constant difference of exactly 32.184 seconds between TT and TAI is a historical artifact ensuring continuity of Barycentric Dynamical Time (TDB), equivalently called Ephemeris Time (ET), when the standard was set to use TAI as the official atomic time in 1976. NAIF indicates that ET is a mathematical idealization for the notion of time such that it is often used as what defines “the time” when mathematically writing equations of motion. ET is very closely related to TT but accounts for relativistic effects and is written

$$ET = TT + 0.001\,657 \sin(E + e \sin E) \text{ [s]}, \quad (1.6)$$

where e is Earth’s orbital eccentricity ($e \sim 0.01671$) and E is Earth’s eccentric anomaly. This expression is describable in terms of calendar dates and of course remains within 1.66 milliseconds of TT.

These time systems can be represented in the familiar format of years, months, days, and time of day that most people would recognize as indicated by the indirect connection of ET to GMT. When represented as so, time systems are often represented like time zones in that the time system is shown as a time zone would be. This notation should cause little confusion since time zones and time systems are connected anyways.

1.4.2 Julian Dates

Time systems can also be represented, however, as pure numbers (in seconds) from a starting time, or epoch, by the definition of TAI. The epoch may be chosen as any convenient time since UTC, TAI, and TT all advance in time at the same rate. Thus, choosing an epoch for one time system chooses the epoch for the others. While ET advances at a slightly different rate from UTC, TAI, and TT, the relativistic correction is merely added onto TT, so choosing an epoch for one time system also chooses it for ET in addition with the others. Both TT and ET are often taken to be zero near the epoch January 1, 2000 12:00:000 GMT, referred to as the “J2000 epoch,” or just “J2000.”

A continuous tracking of time (in days) may be kept by introducing the *Julian date*, denoted JD. The Julian date is a measure of how many days have passed since the epoch January 1, 4713 B.C. 12:00:000 UT1 in the Julian calendar. The *Julian date number*, denoted JDN, of a particular date on the (now deprecated) Julian calendar is given by

$$JDN = 367Y_J - \left\lfloor \frac{7}{4} \left(Y_J + 5001 + \left\lfloor \frac{M_J - 9}{7} \right\rfloor \right) \right\rfloor + \left\lfloor \frac{275}{9} M_J \right\rfloor + D_J + 1\,729\,777 \text{ [days]} \quad (1.7)$$

where Y_J , M_J , and D_J are the corresponding year, month, and day of the month in the Julian calendar. The Julian date in the Gregorian calendar is founded with epoch November 24, 4714 B.C. 12:00:000 UT1 from which the Julian date number is given by

$$\begin{aligned} JDN = & \left\lfloor \frac{1461}{4} \left(Y_G + 4800 + \left\lfloor \frac{M_G - 14}{12} \right\rfloor \right) \right\rfloor + \left\lfloor \frac{367}{12} \left(M_G - 2 - 12 \left\lfloor \frac{M_G - 14}{12} \right\rfloor \right) \right\rfloor \\ & - \left\lfloor \frac{3}{4} \left\lfloor \frac{1}{100} \left(Y_G + 4900 + \left\lfloor \frac{M_G - 14}{12} \right\rfloor \right) \right\rfloor \right\rfloor + D_G - 32\,075 \text{ [days]}, \end{aligned} \quad (1.8)$$

where Y_G , M_G , and D_G are the corresponding year, month, and day of the month in the Gregorian calendar. Regardless of the calendar used, these Julian date numbers are the same and the Julian date is based on the 24-hour mean solar day as

$$JD = JDN + \frac{h - 12}{24} + \frac{m}{1440} + \frac{s}{86\,400} \text{ [days]}, \quad (1.9)$$

where h , m , and s are the hour, minute, and second of the day in UT1. Leap seconds do not play a role in determining the Julian date, but they may be incorporated into the calculation

by changing 86400 to either 86399 or 86401 depending on if there is a negative or positive leap second inserted into the day. The resulting number is called the quasi-Julian date and is not often used in application.

Because the epoch for JD is thousands of years ago, the Smithsonian Astrophysical Society defined the Modified Julian Date (MJD) in 1957 as

$$\text{MJD} = \text{JD} - 2\,400\,000.5 \text{ [days]}. \quad (1.10)$$

The effect of the MJD is to set the beginning of the Julian day to midnight instead of noon and bring the epoch to the 19th century – precisely to November 17, 1858 00:00:00 UT1 (Gregorian calendar). Another variation to the Julian date is the Julian century (JC). It is defined accordingly, most commonly to Julian dates referencing the start of a new century.

$$\text{JC} = \frac{\text{JD}}{36\,525} \text{ [centuries]} \quad (1.11)$$

Note that the Julian century is defined using the year containing 365.25 days in the Julian calendar.

A particularly important Julian date, which serves as a common epoch, is that of January 1, 2000 12:00:00 TT (Gregorian calendar), denoted J2000, and is written

$$\text{J2000} = 2451545.0 \text{ [Julian days]}. \quad (1.12)$$

Other epochs have been produced historically. Although now deprecated, the Besselian year was introduced as an attempt to theoretically deriving an epoch as

$$\text{B} = 1900.0 + \frac{\text{JD} - 2415020.31352}{365.242198781} \text{ [Besselian years]}. \quad (1.13)$$

These kinds of epochs, like B1950, will not be mentioned again, but its use is solidified in records predating the early 1980s and therefore included to establish a relationship between it and the Julian date.

The International Earth Rotation and Reference Systems Service (IERS) provides in Technical Note No. 36 that if the current time t references J2000 as a number of Julian centuries in TT, that is

$$t = \frac{\text{TT} - \text{J2000}}{36525}, \quad (1.14)$$

then the Greenwich Mean Sidereal Time (GMST) is given by

$$\begin{aligned} \text{GMST} = & \text{ERA}(\text{UT1}) + 0.014506'' + 4612.156534'' t + 1.3915817'' t^2 - 0.00000044'' t^3 \\ & - 0.000029956 t^4 - 0.0000000368'' t^5. \end{aligned} \quad (1.15)$$

Here, the earth rotation angle (ERA) is given in terms of the amount of Julian days since J2000 as

$$\text{ERA}(T_u) = 2\pi(\text{UT1 Julian day fraction} + 0.7790572732640 + 0.00273781191135448 T_u), \quad (1.16)$$

where $T_u = \text{UT1 Julian date} - \text{J2000}$ and it is noted that the UT1 Julian day fraction is 0.5 at 0^h (midnight) UT1 [IERS Tech Note 36].

2. COORDINATE SYSTEMS AND TRANSFORMATIONS

In the context of spaceflight, most transformations are simple translations and rotations while reusing the same original basis but now with respect to the newly transformed axes, though there are notably still nonlinear transformations. In particular, transformations to and from spherical coordinates and ellipsoidal coordinates are nonlinear and require some special attention. A thorough look into each relevant coordinate system and frame of reference follows.

2.1 Earth-Centered Inertial (ECI) Reference Frame

The frame of reference defined by the basis

$$\{\hat{I}, \hat{J}, \hat{K}\} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}, \quad (2.1)$$

with respective coordinate variables $\mathbf{X} = (X, Y, Z)$ is known as the Earth-Centered Inertial (ECI) coordinate system and is positioned at Earth's center of mass. Contrary to the given implication, there are actually many ECI coordinate systems from which to choose due to rather subtle dynamical nuances exhibited by the earth. Whichever one is chosen, however, will exhibit this basis and be referred to as the ECI frame.

Complications arise in choosing an appropriate ECI coordinate system because of how, in general, the ECI coordinate system is defined. For a complete, formal definition, referral to the International Earth Rotation and Reference Systems Service (IERS) and its latest technical note is encouraged, particularly to find the International Celestial Reference System and International Celestial Reference Frame.

In these frames, the IERS accounts for and models complicated and subtle dynamics, like the effects of tidal motion, tectonic plate movement, and various corrections in the framework of general relativity, affecting the position of Earth's center of mass and notion of proper time. These effects become increasingly important when simulating interplanetary trajectories, receiving data from quasars or other deep space objects, or otherwise generally performing operations requiring extreme precision.

For rocket flights in atmosphere and very low Earth orbit (VLEO), most of these effects are already overwhelmed by the uncertainty involved in computing drag coefficients and predicting plumes of air rising into the way of suborbital and orbital flight paths. These rather complicated and small corrections will therefore be ignored. With this decision, a general ECI coordinate system is defined by Tab. 2.1.

Tab. 2.1: A general definition of ECI coordinates with complicated and small effects removed from consideration. The origin is located at Earth's center of mass. This definition is consistent with that of the true-of-date ECI reference frame.

Axis	Variable	Basis	Definition
3	$Z (x^3)$	$\hat{K} (\hat{e}_3)$	Aligned with Earth's spin axis now
1	$X (x^1)$	$\hat{I} (\hat{e}_1)$	Aligned along the <i>vernal equinox</i> now
2	$Y (x^2)$	$\hat{J} (\hat{e}_2)$	Complete the right-handed system, $\hat{J} = \hat{K} \times \hat{I}$

The vernal equinox is defined to be the direction of the sun at spring equinox along the intersection of Earth's equatorial and orbital planes. It is denoted by the astrological symbol of Aries Υ , the first in the zodiac, and is typically shown jointly with the symbol indicating the 1-axis. Earth's equatorial plane, which is normal to Earth's spin axis, is inclined to the orbital plane, called *the ecliptic*, by about 23.43° .

Such a nonzero inclination gives rise to a precession, or gyration, of Earth's spin axis, therefore a precession of the equatorial plane, therefore a precession of the vernal equinox, due to a gravitational gradient imparted by the sun (and moon) on the bulging of Earth's equator. This precession follows the motion of Earth's inclination offset to the ecliptic, which is almost but not quite a constant 23.43° , and is coupled with small nutations mainly imparted by the moon having an inclined, elliptical orbit about Earth.

Furthermore, the actual motion of Earth about the sun in the ecliptic is not truly captured in a flat plane as predicted by Kepler's laws since other celestial bodies always provide slight perturbations to the motion of Earth. Also, due to tidal effects, other celestial bodies, and general relativity among other factors, Earth's orbit experiences an apsidal drift, or a changing of the locations of the apses (apoapsis and periapsis), affecting its trajectory about the sun itself.

These effects all amount to shifting the direction of the vernal equinox continuously, something which should be accounted for in a proper treatment of the problem, though the effects of gyration on the coordinate basis become increasingly important as deeper space and is encountered. Additionally, the time scales over which these effects occur are quite large. The period of precession for Earth's spin axis is nearly 26,000 years and, although the period of nutation from the moon is on the order of weeks to months, its overall effects are relatively small. Earth's inclination to the ecliptic varies by about 1.3° with a period of about 41,000 years while its out-of-plane perturbations are also small with a period of about 70,000 years, and the period of Earth's apsidal drift is about 112,000 years. The result is that the vernal equinox precesses with a period of about 22,000 years.

A coordinate system used with intention to track events over even a notable fraction of such a time scale, such as long-term spacecraft tracking or generation of planetary ephemerides, would appropriately implement such nuances into its framework. Such a system is called the true-of-date ECI coordinate frame since its basis correctly points in the directions of the spin axis and vernal equinox as they are at all times. The true-of-date ECI reference frame is what was presented as the general ECI frame in Tab. 2.1. This reference frame is technically correct, but leads to overcomplications in terms of mathematical and computational logic for flights not involving deep space or long time scales.

For this reason, an additional ECI coordinate system called mean-of-date ECI is defined by the intersection at spring equinox of the ecliptic, frozen in time forever at a particular epoch, with the equatorial plane, also frozen in time forever at the same particular epoch. The orientation of this frozen equatorial plane is determined only by the advancement of the spin axis precession at the epoch time and considering no other perturbative effects.

Tab. 2.2: The definition of the mean-of-date ECI coordinate system. The epoch time defining the system is any reference time and its origin is located at Earth's approximate center of mass at this epoch time.

Axis	Variable	Basis	Definition
3	$Z (x^3)$	$\hat{K} (\hat{e}_3)$	Aligned with Earth's spin axis from epoch
1	$X (x^1)$	$\hat{I} (\hat{e}_1)$	Aligned along the corresponding vernal equinox from epoch
2	$Y (x^2)$	$\hat{J} (\hat{e}_2)$	Complete the right-hand system, $\hat{J} = \hat{K} \times \hat{I}$

The epoch can be chosen to be any reference time, but several standard choices of past epochs do exist. The most recent standard epoch is currently taken to be January 1, 2000 12:00:00.000

TT. Because of the precession of Earth's spin axis, the true alignment of the spin axis and the mean-of-date ECI 3-axis slowly drift apart as time increases away from the epoch. For times close to the epoch, the misalignment may essentially be neglected since the precession period is, comparatively, so long. This particular epoch and associated mean-of-date ECI reference frame is referred to as any of IJK, EME2000, or, most commonly, J2000.

The J2000 reference frame and previously mentioned International Celestial Reference Frame (ICRF) are closely related. The ICRF, defined by the International Earth Rotation and Reference Systems Service (IERS), which is the international standard of astronomical inertial coordinate systems, was constructed using 295 extragalactic radio sources to coincide almost exactly with the J2000 coordinate system. Though the ICRF is placed at the solar system's barycenter, the two coordinate systems differ in rotation from one another by less than $0.1''$, or less than 0.000028° , when placed on top of each other [IERS, NAIF].

With this information considered, an inertial coordinate system, from which everything to follow will be defined, is now decided. Atmospheric and VLEO flights, which typically exhibit exceedingly short flight times in comparison of the cosmic scale, essentially leave no time for the vernal equinox to gyrate or nutate, so the true-of-date ECI variety will not be considered. The standardization and recency of the latest epoch from mean-of-date ECI frames however, paired together with exhibiting such a small, practically negligible, rotation from the ICRF, yield strong indications that J2000 should be the inertial coordinate system and reference frame of choice. The defining characteristics of the J2000 basis are defined in Tab. 2.2 using the epoch January 1, 2000 12:00:00.000 TT exhibiting the basis (2.1).

2.2 Earth-Centered Earth-Fixed (ECF) Reference Frame

The Earth-Centered Earth-Fixed (ECEF, or simply ECF) reference frame views the Earth as a nonrotating body with a basis *viewed in its own coordinate system*

$$\{\hat{i}, \hat{j}, \hat{k}\} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\} \quad (2.2)$$

with coordinate variables (x, y, z) and its origin located at Earth's center of mass. With respect to the ECI frame, the ECEF frame rotates with its basis attached to fixed points on Earth. Since the chosen ECI frame, J2000, ignores perturbative effects causing nutations of Earth's spin axis and the assumed flight time is so small in comparison to any periodic perturbations of Earth's spin axis, the angular velocity of Earth, denoted $\vec{\omega}$, is taken to be a true constant along the J2000 3-axis and defined in terms of the rotational rate of Earth given by (1.1) as

$$\vec{\omega} = \omega \hat{K}. \quad (2.3)$$

The quantity is taken to be oriented along the positive 3-axis since Earth rotates west to east, or counterclockwise when viewed from above the North Pole. The ECEF reference coordinate system is then intuitively defined.

Tab. 2.3: The definition of the ECEF coordinate system as based off of the J2000 reference frame. The epoch time defining the system is therefore J2000 by definition and the coordinate system's origin is located at Earth's approximate center of mass at this epoch.

Axis	Variable	Basis	Definition
3	z (x^3)	\hat{k}	Aligned with the J2000 spin axis
1	x (x^1)	\hat{i}	Aligned along the equator towards Greenwich, England
2	y (x^2)	\hat{j}	Complete the right-hand system, $\hat{j} = \hat{k} \times \hat{i}$

The choice of aligning the 1-axis in the direction of Greenwich, England is purely historical to give a convenient way of tracking Earth's rotation using the local time there from which everything is defined. With respect to the ECI frame, the ECEF frame is therefore seen to be rotating, and considering an object \mathfrak{X} , the components relate to one another between the two bases according to

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \underbrace{\begin{pmatrix} \cos \lambda_G & \sin \lambda_G & 0 \\ -\sin \lambda_G & \cos \lambda_G & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\mathbf{T}^{\text{ecf/eci}}} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}. \quad (2.4)$$

The quantity λ_G is called the Greenwich hour angle and is used to track the angular displacement of Greenwich, England from the vernal equinox. It is therefore, by definition, a function of time and is expressed

$$\lambda_G = \lambda_{G_0} + \omega(t - t_0), \quad (2.5)$$

where λ_{G_0} is the initial Greenwich hour angle at an initial time t_0 . This angle is (implicitly) given by relating it to the amount of seconds that have passed in Greenwich since an epoch up until t_0 as

$$\frac{\lambda_{G_0}}{\omega} = 24110.54841 + 8640184.812866 T_0 + 0.093104 T_0^2 - 0.0000062 T_0^3 + \frac{86400}{2\pi/\omega} T_s \text{ [s]}, \quad (2.6)$$

where T_0 is the amount of Julian centuries at t_0 since the epoch, J2000 (January 1, 2000 12:00:00 TT), defined

$$T_0 = \frac{\text{JD}_0 - 2451545}{36525} \quad (2.7)$$

and T_s is the amount of seconds passed in the current day (since midnight) according to UT1. The periodicity of the elementary trigonometric functions, however, may be exploited to provide more information. The number of sidereal days elapsed since the initial time t_0 can be identified by counting the number of discontinuities of

$$\Lambda_G = \lambda_G - \lambda_{G_0} \text{ modulo } 2\pi. \quad (2.8)$$

This alternate representation of an hour angle simply uses the initial time t_0 as the epoch for where Greenwich, England is as the reference and clearly tracks when $\omega(t - t_0) = 2n\pi$, where $n \in \mathbb{Z}$ is the current sidereal day at time t since t_0 . An alternate, but equivalent, definition of the Greenwich hour angle in terms of Greenwich Mean Sidereal Time (GMST) is given by (? eq ?).

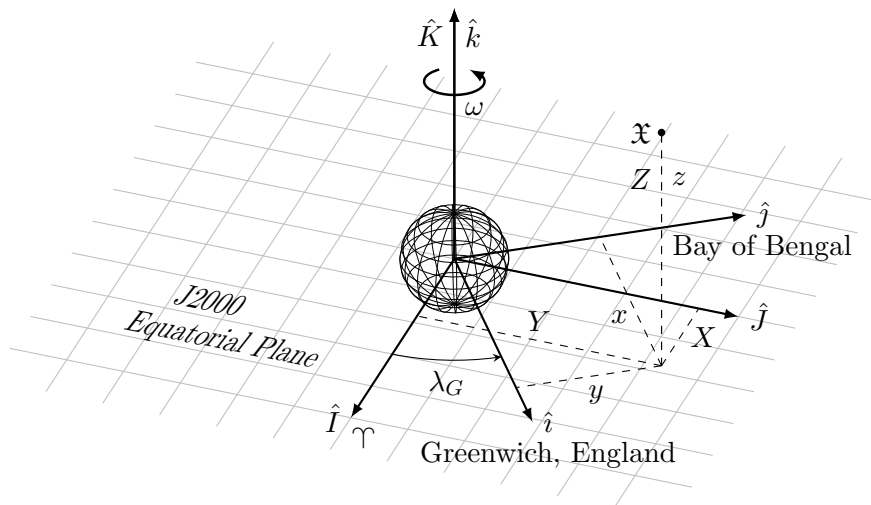


Fig. 2.1: A graphical representation of the relation between J2000 ECI and J2000 ECEF coordinate systems. The earth is depicted as a spherical wire frame of rings corresponding to constant longitudes and latitudes — the 6 unit vectors are exaggerated in length for the sake of illustration.

2.3 Spherical Coordinates

Because of the sphere's perfect symmetry, the most convenient basis is taken to be stationary with respect to the sphere originating from the sphere's center as

$$\{\hat{e}_u, \hat{e}_v, \hat{e}_w\} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}. \quad (2.9)$$

This basis is the usual Euclidean basis but with coordinates (ξ, η, ζ) to describe a point on the sphere of *spherical coordinates* (ρ, θ, λ) via the coordinate transformation

$$u = r \sin \theta \cos \lambda \quad r^2 = u^2 + v^2 + w^2 \quad (2.10)$$

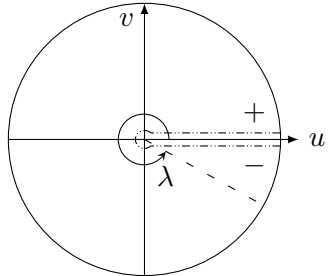
$$v = r \sin \theta \sin \lambda \quad \cos \theta = w/r \quad (2.11)$$

$$w = r \cos \theta \quad \tan \lambda = v/u, \quad (2.12)$$

where the inverse transform is given implicitly since naïvely taking inverse functions may give the wrong answer. The spherical coordinates must satisfy all of

$$0 < r < \infty \quad 0 \leq \theta \leq \pi \quad 0 \leq \lambda < 2\pi. \quad (2.13)$$

Obviously, the sphere's radius r is zero at the origin, but such a value is discluded from spherical coordinates since both θ and λ are then undefined; this kind of sphere is called the null sphere or a singularity. Demanding its positivity indicates that the positive square root is always taken, reflecting with the absoluteness of the notion of distance. The second coordinate θ , called *colatitude*, may be directly obtained by applying the inverse cosine — no checks are needed. In contrast, the final coordinate λ , called *longitude*, is often the source of error. To obtain the correct answer, the inverse tangent needs defining here as the 2-argument inverse tangent with its branch cut on the positive u axis. That is to say that longitude must necessarily be defined

$$\lambda = \begin{cases} \tan^{-1}(v/u) & u > 0, v \geq 0 \\ \pi/2 & u = 0, v > 0 \\ \pi + \tan^{-1}(v/u) & u < 0 \\ 3\pi/2 & u = 0, v < 0 \\ 2\pi + \tan^{-1}(v/u) & u > 0, v < 0 \\ \text{undefined} & \text{otherwise.} \end{cases} \quad (2.14)$$


Geometrically, specifying the radius but not the two angles produces a spherical shell (as opposed to a solid ball) of fixed radius whereas not specifying the radius and specifying the two angles produces a line emanating from the origin in the direction determined by the two angles and so on. In particular application to Earth, specifying a radius r and a colatitude θ produces a circle with radius

$$\rho = \sqrt{r^2 - r^2 \cos^2 \theta} = r \sin \theta. \quad (2.15)$$

In the special case of when $\sin \theta = 1$, the circle reaches its maximum possible size and is referred to as a *great circle*. This great circle is often referred to as *the equator*. The other special case of when $\sin \theta = 0$ corresponds with a null circle, called a *pole*. The $\theta = 0$ pole is often called the *north pole* and the $\theta = \pi$ pole is often called the *south pole*. All circles formed this way are called *parallels*. In contrast, specifying a radius r and longitude λ produces a semi-circle (because of the range of θ) of radius r called a *meridian*. The complement to this meridian is determined by specifying the same radius r but with longitude $\lambda + \pi$ and is called the *anti-meridian*. When considered joined together, every pair of meridian and anti-meridian forms a great circle.

A new coordinate system with a new basis is established at (r, θ, λ) , denoted $\{\hat{e}_r, \hat{e}_\theta, \hat{e}_\lambda\}$, since this representation of coordinates is actually a coordinate transformation onto the sphere. Inserting all of u , v , and w into (A.7), the required length-preserving (normalized) basis rotation is written

$$\begin{pmatrix} \hat{e}_r \\ \hat{e}_\theta \\ \hat{e}_\lambda \end{pmatrix} = \underbrace{\begin{pmatrix} \sin \theta \cos \lambda & \sin \theta \sin \lambda & \cos \theta \\ \cos \theta \cos \lambda & \cos \theta \sin \lambda & -\sin \theta \\ -\sin \lambda & \cos \lambda & 0 \end{pmatrix}}_{\mathbf{B}^{\text{sphr}/\text{cart}}} \begin{pmatrix} \hat{e}_u \\ \hat{e}_v \\ \hat{e}_w \end{pmatrix}. \quad (2.16)$$

Precisely at (r, θ, λ) , the origin of the *spherical coordinate system*, \hat{e}_r is normal to the sphere and pointing away from the sphere's center, \hat{e}_θ is tangent to the meridian and directed towards south pole, and \hat{e}_λ is tangent to its great circle pointed in the positive (counterclockwise-like) direction. The tuple (r, θ, λ) fixes the origin's location and determines its associated basis $\{\hat{e}_r, \hat{e}_\theta, \hat{e}_\lambda\}$, which is functionally the same as the usual Cartesian system, in which new objects may be expressed with new variables (r', θ', λ') . Despite the similar symbolism to (r, θ, λ) , these new variables (r', θ', λ') are linear coordinates — *they do not represent spherical coordinates*. To emphasize this point, a new basis is defined to simply change symbols and avoid this ambiguity.

$$\{\hat{e}_{r'}, \hat{e}_{\theta'}, \hat{e}_{\lambda'}\} \equiv \{\hat{e}_r, \hat{e}_\theta, \hat{e}_\lambda\} \quad (2.17)$$

Considering an object \mathfrak{X} at (r, θ, λ) in the $\{\hat{e}_u, \hat{e}_v, \hat{e}_w\}$ basis and another object \mathfrak{X}' at (r', θ', λ') in the $\{\hat{e}_{r'}, \hat{e}_{\theta'}, \hat{e}_{\lambda'}\}$ basis, the object \mathfrak{X}' has components (u', v', w') in the $\{\hat{e}_u, \hat{e}_v, \hat{e}_w\}$ basis written

$$\begin{pmatrix} u' \\ v' \\ w' \end{pmatrix} = \begin{pmatrix} u \\ v \\ w \end{pmatrix} + \underbrace{\begin{pmatrix} \sin \theta \cos \lambda & \cos \theta \cos \lambda & -\sin \lambda \\ \sin \theta \sin \lambda & \cos \theta \sin \lambda & \cos \lambda \\ \cos \theta & -\sin \theta & 0 \end{pmatrix}}_{\mathbf{T}^{\text{cart}/\text{sphr}}} \begin{pmatrix} r' \\ \theta' \\ \lambda' \end{pmatrix}. \quad (2.18)$$

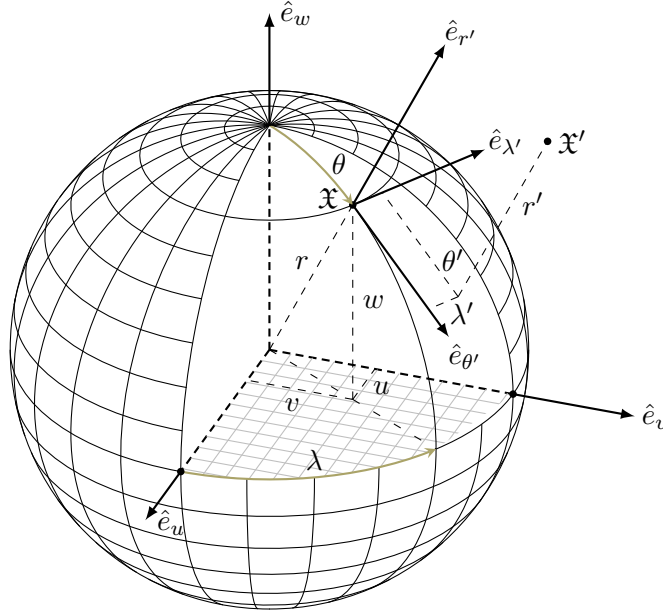


Fig. 2.2: The relationship between the Cartesian system, spherical coordinates, and the spherical coordinate system. The basis $\{\hat{e}_{r'}, \hat{e}_{\theta'}, \hat{e}_{\lambda'}\}$ is shown with individually adjusted lengths of its elements for visualization, but each element is really of unit length to form the orthonormal basis.

Tab. 2.4: Spherical coordinate system basis with its origin to be positioned at (ρ, θ, λ) .

Axis	Variable	Basis	Definition
1	$x^1 (r')$	$\partial_{x^1} (\hat{e}_{r'})$	Normal to the sphere at (r, θ, λ) pointing outwards
2	$x^2 (\theta')$	$\partial_{x^2} (\hat{e}_{\theta'})$	Tangent to the λ meridian pointing southwards
3	$x^3 (\lambda')$	$\partial_{x^3} (\hat{e}_{\lambda'})$	Tangent to the θ parallel pointing eastwards (counterclockwise)

Alternate representations of spherical coordinates indeed exist. The differences can be purely cosmetic or involve very simple coordinate transformations but are still enough to cause confusion when comparing information herein with outside sources. For clarity, the symbols chosen in this representation of spherical coordinates are (r, θ, λ) and will remain so unless otherwise stated; the following is simply a caution that other representations exist and how they relate to ours.

Symbolism is a simple and common source of error when working in spherical coordinates. Spherical coordinates are typically written (r, θ, ϕ) in the studies of both physics and pure mathematics, but this tuple has different meanings in each field. The symbol r in each field denotes the radial distance from the origin; the differences come from the angles. In physics, θ and ϕ respectively denote the colatitude and longitude while in mathematics, θ and ϕ respectively denote the longitude and colatitude. Converting between the two systems is merely a swapping of symbols.

Spherical coordinates are sometimes used with the *latitude* φ to which the colatitude is complementary. That is to say that the latitude and colatitude are related by

$$\varphi + \theta = \frac{\pi}{2}. \quad (2.19)$$

The effect had on the basis from this simple transformation if repeating everything with latitude in place of colatitude is reversing the direction of $\hat{e}_{\theta'}$, thus producing a coordinate system that is not right-handed. Therefore, it is important to use only the scalar relation (2.19) if working with latitude in place of colatitude and requiring the spherical coordinate system.

In some fields, longitude is also called *right ascension* and is abbreviated RA. Latitude is also called *declination*, denoted δ or simply DEC, or sometimes *inclination* i .

When applied to Earth, the Euclidean basis is essentially taken to be the J2000 ECEF basis such that

$$\{\hat{e}_u, \hat{e}_v, \hat{e}_w\} \equiv \{\hat{i}, \hat{j}, \hat{k}\}, \quad (2.20)$$

and the associated radius r is popularly taken as the earth's average equatorial radius because of the equatorial bulge induced by rotation. Doing so ensures that “all” of Earth is contained in the spherical shell. This shell is, however, *not* a good representation of Earth's surface directly as there is about a 22 kilometer difference between the Earth's equatorial and polar radii.

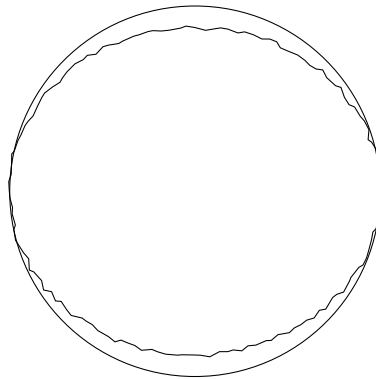


Fig. 2.3: An exaggerated two-dimensional visualization of the containment of Earth in a sphere of average equatorial radius in the cross section along a meridian/anti-meridian great circle.

2.4 Ellipsoidal Coordinates

An *ellipsoid* is a three-dimensional surface satisfying the equation

$$\frac{u^2}{r_u^2} + \frac{v^2}{r_v^2} + \frac{w^2}{r_w^2} = 1, \quad (2.21)$$

where r_u , r_v , and r_w are positive constants. Thus, the ellipsoid is a generalization of the sphere and it is therefore no coincidence that the tuple (ξ, η, ζ) reappears from Sec. 2.3 with the basis

$$\{\hat{e}_u, \hat{e}_v, \hat{e}_w\} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\} \quad (2.22)$$

located at the ellipsoid's center.

If any two axes are of equal length, then the ellipsoid is called a *spheroid*. Supposing the spheroid takes the form

$$\frac{u^2 + v^2}{a^2} + \frac{w^2}{b^2} = 1 \quad (0 < b < a), \quad (2.23)$$

then the spheroid is *oblate* such that a is the *semi-major axis* and b is the *semi-minor axis*. The oblate spheroid can be thought of as a sphere squashed from the poles. The amount of squashing is described by the *flattening* of the spheroid from the sphere.

$$f = \frac{a - b}{a} \quad (2.24)$$

An accompanying quantity to the flattening is the *eccentricity*, also called *ellipticity*, motivated by relating the semi-major axis to distances between the center and a focus in the cross section of a meridian as shown in Fig. 2.4. The eccentricity therefore is like a scale factor of how far away the ellipse foci are from the geometric center and is defined

$$e = \sqrt{\frac{a^2 - b^2}{a^2}}, \quad (2.25)$$

which has resemblance to the flattening. They are indeed related simply via

$$e = \sqrt{1 - (1 - f)^2} = \sqrt{2f - f^2}. \quad (2.26)$$

For completeness, the *semi-latus rectum* is defined as the distance from a focus to the ellipse in the semi-minor axis direction. Its expression is given graphically in Fig 2.4 and is written

$$p = a(1 - e^2). \quad (2.27)$$

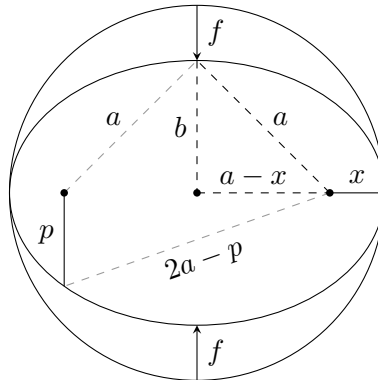


Fig. 2.4: A two-dimensional visualization of the containment of the oblate spheroid inside of a sphere showing the definitions of the flattening (above/below ellipse), eccentricity (top-half of ellipse), and semi-latus rectum (bottom-half of ellipse). Here, $x = a(1 - e)$.

Fitting an ellipsoid to the surface of Earth produces a spheroid (2.23) due to rotational symmetry, which will be represented by the slightly modified notation

$$\frac{x_s^2 + y_s^2}{a^2} + \frac{z_s^2}{b^2} = 1 \quad (2.28)$$

under the ECEF basis $\{\hat{i}, \hat{j}, \hat{k}\}$. The semi-major axis a is fixed to be the average equatorial radius and the semi-minor axis b is fixed as the polar radius. Each coordinate variable carries the subscript s to indicate specifically that this tuple (x_s, y_s, z_s) , which uses the basis $\{\hat{i}, \hat{j}, \hat{k}\}$, resides on the surface of the spheroid representing Earth — not on the surface of Earth itself.

Because Earth's spheroid is a sphere flattened from its poles, the concepts of parallels and poles being circles remain the same, but meridian and anti-meridians are no longer semicircles and their pairs no longer form great circles. This geometric asymmetry can be more easily visualized upon inspection of the mathematical symmetry of (2.28). The spheroid is reduced in dimension to an ellipse, which is the same at every longitude λ , by introducing a new variable χ such that

$$\frac{\chi_s^2}{a^2} + \frac{z_s^2}{b^2} = 1. \quad (2.29)$$

The corresponding basis element is therefore

$$\hat{e}_\chi = \cos(\lambda) \hat{i} + \sin(\lambda) \hat{j} + 0 \hat{k}. \quad (2.30)$$

Though meridians are no longer semicircles, the flattening does not alter their positioning relative to those of the originating sphere — *the longitude λ is determined exactly as it was in spherical coordinates using (2.14)*. Parallels, however, remaining circles as with spherical coordinates, have their positionings scaled vertically inwards because of the flattening. This transformation sources complications in determining a point's latitude on and around the spheroid.

2.4.1 Reduced Latitude

The most intuitive attempt at parameterizing the coordinate variables is probably to copy the form of the parameterization used in spherical coordinates. Doing so with constant radii on each axis, of course, is the equivalent of measuring angles with respect to circles that encircle each axis. In the λ cross-section, these circles are of radius a and b and centered with the ellipse. An object \mathfrak{X}_s on the surface of the spheroid at (x_s, y_s, z_s) is then described by its position on these circles using the *reduced latitude*, or *parametric latitude*, denoted β , and longitude λ . The position on the spheroid and inverse relation are therefore

$$x_s = a \cos \beta \cos \lambda \quad (2.31)$$

$$y_s = a \cos \beta \sin \lambda \quad \sin \beta = z_s/b \quad (2.32)$$

$$z_s = b \sin \beta \quad \tan \lambda = y_s/x_s. \quad (2.33)$$

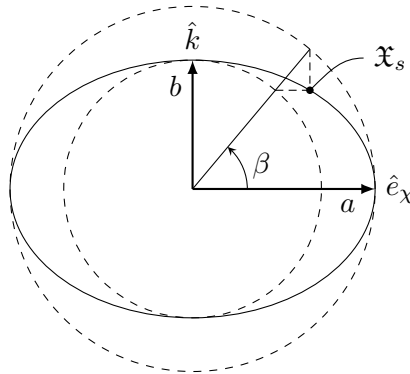


Fig. 2.5: A two-dimensional visualization of the reduced latitude with inscribing circles of each axis. This cross-section is cut along a λ meridian of the three-dimensional spheroid

2.4.2 Geocentric Latitude

Describing a point on the spheroid using a line from the origin directly to the spheroid's surface requires the use of a radius r_s that is variable in *geocentric latitude* denoted φ . This radius is easily obtainable by substituting polar coordinates (r_s, φ_s) into (2.29). Thus, the radius satisfies

$$\frac{r_s^2 \cos^2 \varphi_s}{a^2} + \frac{r_s^2 \sin^2 \varphi_s}{b^2} = 1 \quad (2.34)$$

and is written

$$r_s = \frac{ab}{\sqrt{a^2 \sin^2 \varphi_s + b^2 \cos^2 \varphi_s}} = \frac{b}{\sqrt{1 - e^2 \cos^2 \varphi_s}}. \quad (2.35)$$

The position of an object \mathfrak{X}_s on the spheroid then follows the form of spherical coordinates

$$x_s = r_s \cos \varphi_s \cos \lambda \quad r_s^2 = x_s^2 + y_s^2 + z_s^2 \quad (2.36)$$

$$y_s = r_s \cos \varphi_s \sin \lambda \quad \sin \varphi = z_s / r_s \quad (2.37)$$

$$z_s = r_s \sin \varphi_s \quad \tan \lambda = y_s / x_s. \quad (2.38)$$

These coordinates $(r_s, \varphi_s, \lambda)$ are closely related to those of spherical coordinates (r, θ, λ) . The only difference is that the radius r_s is constrained to the spheroid's surface and the geocentric latitude φ_s is the complementary angle to the geocentric colatitude θ_s . As such, each transformed variable satisfies

$$b \leq r_s \leq a \quad -\pi \leq 2\varphi_s \leq \pi \quad 0 \leq \lambda < 2\pi. \quad (2.39)$$

The geocentric colatitude θ_s may be used instead of the geocentric latitude φ_s if preferred to reflect more closely with the representation found in spherical coordinates. If done, then this use of colatitude *is* spherical coordinates such that the radius is merely constrained to the spheroid's surface. Indeed, removing this restraint to describe an object \mathfrak{X} at an arbitrary (x, y, z) so that all subscripts s are deleted simply results in the definition of spherical coordinates under the basis $\{\hat{i}, \hat{j}, \hat{k}\}$.

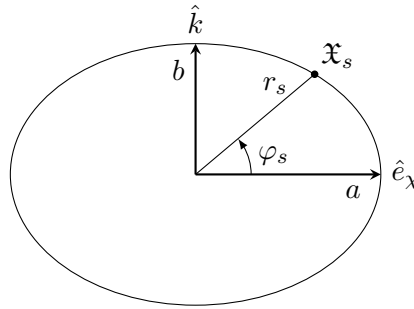


Fig. 2.6: The geocentric latitude and radius constrained to the spheroid's surface at a longitude λ .

On the spheroid's surface, however, the approaches used with the reduced and geocentric latitudes both describe the same object \mathfrak{X}_s at (x_s, y_s, z_s) , so a relationship to transform between the two forms of latitude is establishable. Equating all three components yields two independent equations.

$$\cos \beta = \frac{b}{\sqrt{a^2 \sin^2 \varphi_s + b^2 \cos^2 \varphi_s}} \cos \varphi_s \quad (2.40)$$

$$\sin \beta = \frac{a}{\sqrt{a^2 \sin^2 \varphi_s + b^2 \cos^2 \varphi_s}} \sin \varphi_s \quad (2.41)$$

Upon division is obtained the simple relation

$$\tan \varphi_s = (1 - f) \tan \beta. \quad (2.42)$$

Though (2.42) is undefined when either angle is on the boundary of its domain, intuition and visual inspection of Fig. 2.7 correctly indicate that *at the poles*,

$$\varphi_s = \beta = \text{sgn}(z_s) \frac{\pi}{2}. \quad (2.43)$$

These two points are not the only ones that have this property. Its straightforward from (2.42) that these angles are also equal whenever either is zero. In fact, the rays themselves are identically the same in this case.

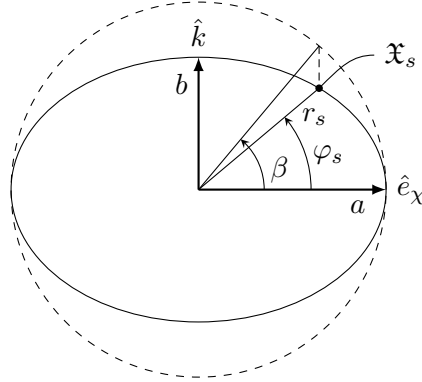


Fig. 2.7: The reduced latitude and geocentric latitude methods compared together at a longitude λ in the spheroid. The effect of the flattening in the relationship between the two angles is seen by the vertical squashing of the circle (sphere) into the ellipse (spheroid).

These three cases, at each pole and along the equator of the spheroid, are the only points at which the reduced latitude β and geocentric latitude φ are the same. At all other points, reduced latitude always outpaces geocentric latitude for their given domains. Because of their domains, taking the regular inverse tangent in favor of either angle is acceptable. The graphic description of the two angles is shown in Fig 2.8 with the flattening as a parameter of four different spheroids.

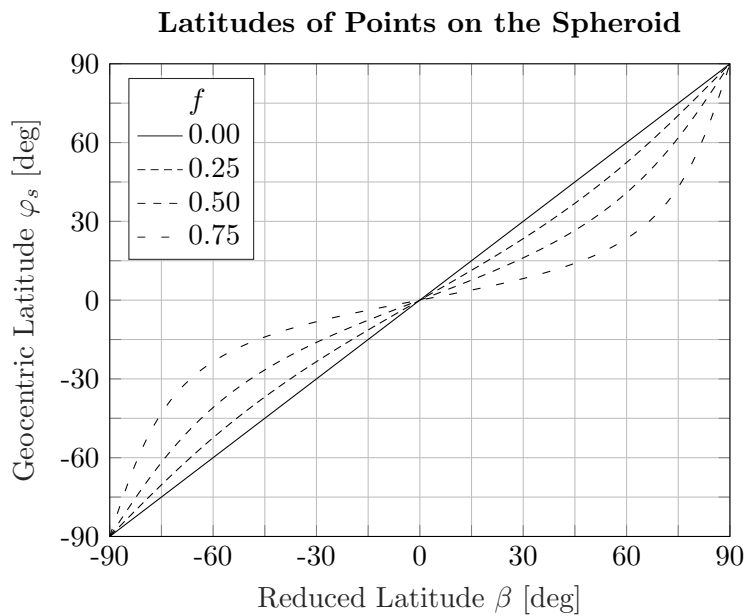


Fig. 2.8: Reduced and geocentric latitude relationship as parameterized by the flattening f .

2.4.3 Geodetic Latitude

Geodetic coordinates are commonly used in geodesy and are what GPS satellites use in mapping Earth's surface with reference to a spheroid. As such, the term *latitude* in real-world applications typically refers to the *geodetic latitude*, also called *geographic latitude*, which, in the λ -plane, is the angle made between the major axis and the ray extending from the minor axis normally to a point on the spheroid's surface. This ray is called the *prime vertical radius (of curvature)* which has a length denoted by R_n .

The geodetic latitude, denoted ϕ , at a point \mathfrak{X}_s described by the tuple (x_s, y_s, z_s) on the surface of the spheroid at a longitude λ is found from (2.29) as

$$\tan \phi = \frac{z_s/\chi_s}{1 - e^2}. \quad (2.44)$$

Using the inverse tangent function here to find the latitude except at the poles is permissible since the relation, in this case, is surjective. At the poles, the latitude follows by correct intuition

$$\phi = \text{sgn}(z_s) \frac{\pi}{2}. \quad (2.45)$$

The equation of the ray in the λ -plane which pierces normally through the spheroid's surface at the point (x_s, y_s, z_s) is therefore

$$z(\chi) = z_s + \frac{z_s/\chi_s}{1 - e^2}(\chi - \chi_s). \quad (2.46)$$

Consequently, the prime vertical radius at this point is of length

$$R_n = \sqrt{\chi_s^2 + z_s^2/(1 - e^2)^2}. \quad (2.47)$$

Despite the breakdown of (2.46) at the poles, the expression for R_n is continuous and will therefore be extended to include points at the poles where $\chi_s = 0$.

The statements (2.29) and (2.46) remain quite powerful generators of additional relations. A short study of (2.46) yields that the portion of prime vertical radius connecting its two intercepts is precisely of length $e^2\ell$. Inserting then a pseudo-polar coordinate pair involving the prime vertical radius and geodetic latitude into (2.29) as

$$\frac{R_n^2 \cos^2 \phi}{a^2} + \frac{(1 - e^2)^2 R_n^2 \sin^2 \phi}{b^2} = 1 \quad (2.48)$$

consequently results in obtaining the prime vertical radius at a geodetic latitude.

$$R_n = \frac{a}{\sqrt{1 - e^2 \sin^2 \phi}} \quad (2.49)$$

A height h normal to the spheroid's surface may be included as the third coordinate to form the coordinate variables (λ, ϕ, h) . Such an introduction allows for the description of Earth's topography everywhere with respect to a given reference spheroid. That is, the value h for a location on Earth's surface at a longitude λ and latitude ϕ with respect to a chosen reference spheroid is given by GPS satellites as that location's "elevation." Thus, this tuple fully determines the components (x, y, z) with respect to the $\{\hat{i}, \hat{j}, \hat{k}\}$ basis of any object \mathfrak{X} from the surface of the spheroid, either outwards or inwards, having semi-major axis a and semi-minor b . The coordinate transformation is then given by visual inspection of Fig. 2.9.

$$x = (R_n + h) \cos \phi \cos \lambda \quad \tan \lambda = y/x \quad (2.50)$$

$$y = (R_n + h) \cos \phi \sin \lambda \quad \left(1 - e^2 \frac{R_n}{R_n + h}\right) \tan \phi = \frac{z}{\sqrt{x^2 + y^2}} \quad (2.51)$$

$$z = ((1 - e^2)R_n + h) \sin \phi \quad h = \sqrt{x^2 + y^2} \sec \phi - R_n \quad (2.52)$$

The inverse transformation cannot be determined exactly and, in general, requires an iterative procedure instead (except for the longitude λ) as given by Alg. 1 for numerical stability.

Algorithm 1 Inverse transformation from (x, y, z) to (h, ϕ, λ)

Require: a , f , and (x, y, z) . The position is not at the origin.

Ensure: $R_n \sim R_{n_g}$, $h \sim h_g$, $\phi \sim \phi_g$, and $0 \leq \lambda < 2\pi$

```

1:  $b \leftarrow a(1 - f)$ 
2:  $e \leftarrow (2f - f^2)^{1/2}$ 
3:  $\chi \leftarrow (x^2 + y^2)^{1/2}$                                 # Projected distance onto the  $xy$  plane
4: if  $\chi$  is zero, then                                     # Positioned over the north or south pole
5:    $h \leftarrow |z| - b$                                      # Height (exact)
6:    $\phi \leftarrow \text{sgn}(z)(\pi/2)$                            # Latitude (exact)
7:    $R_n \leftarrow a/f$                                        # PVR (exact)
8: else
9:    $\phi_{g_o} \leftarrow \infty$                                 # Latitude (old guess)
10:   $\phi_g \leftarrow \tan^{-1}(z/\chi)$                           # Latitude (guess)
11:  while  $|\phi_g - \phi_{g_o}|$  is large, do                  # Do until convergence to a small tolerance is met
12:     $R_{n_g} \leftarrow a(\cos^2 \phi_g + (1 - e^2) \sin^2 \phi_g)^{-1/2}$  # PVR (guess)
13:     $h_g \leftarrow \chi \sec \phi_g - R_{n_g}$                    # Height (guess)
14:     $\phi_{g_o} \leftarrow \phi_g$                                # Cycle
15:     $\phi_g \leftarrow \tan^{-1}(z/(\chi(1 - e^2 R_{n_g}/(R_{n_g} + h_g))))$  # Latitude (new guess)
16:  end while
17:   $\phi \leftarrow \phi_g$                                      # Latitude (converged)
18:   $R_n \leftarrow a(\cos^2 \phi + (1 - e^2) \sin^2 \phi)^{-1/2}$  # PVR (converged)
19:   $h \leftarrow \chi \sec \phi - R_n$                              # Height (converged)
20: end if
21:  $\lambda \leftarrow \text{ATAN2}(\text{PI}(y, x))$                         # Longitude (according to (2.14), exact)

```

Regardless that the inverse transformation is numerical by nature, the ellipsoidal coordinates (λ, ϕ, h) still must satisfy the constraints

$$0 \leq \lambda < 2\pi \quad -\pi \leq 2\phi \leq \pi \quad -(1 - e^2)R_n < h < \infty. \quad (2.53)$$

As an aside, the longitude on Earth has been historically represented to range from -180° to 180° , where a negative value indicates westward travel from the *prime meridian* at 0° longitude, which is arbitrarily chosen to pass directly through Greenwich, England. Conversion to this domain is permissible if desired, though potentially unforeseen problems with the discontinuity on the prime antimeridian may arise which may be more easily handled with the discontinuity set along the prime meridian.

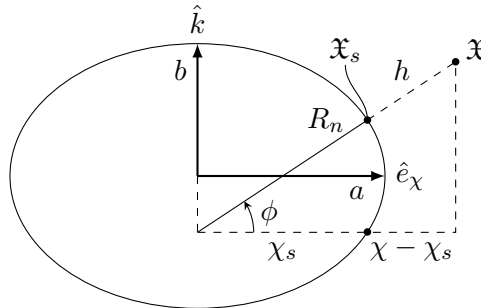


Fig. 2.9: The geodetic coordinates of an object \mathfrak{X}_s on the surface of the spheroid at a longitude λ and latitude ϕ and another object \mathfrak{X} at the same longitude λ and latitude ϕ , but at a nonzero height h normal to the spheroid's surface.

The basis rotation then follows (A.7) in the limit as $R_n \rightarrow 0$ with normalization of each element.

$$\begin{pmatrix} \hat{e}_\lambda \\ \hat{e}_\phi \\ \hat{e}_h \end{pmatrix} = \underbrace{\begin{pmatrix} -\sin \lambda & \cos \lambda & 0 \\ -\sin \phi \cos \lambda & -\sin \phi \sin \lambda & \cos \phi \\ \cos \phi \cos \lambda & \cos \phi \sin \lambda & \sin \phi \end{pmatrix}}_{\mathbf{B}^{\text{env/ecf}}} \begin{pmatrix} \hat{i} \\ \hat{j} \\ \hat{k} \end{pmatrix} \quad (2.54)$$

This basis is often called the Local Horizontal, Local Vertical (LHLV), or also the East, North, Vertical (ENV), coordinate system, which is to have its origin located at (x, y, z) . From this new reference frame, new *linear* variables (with no correlation to (λ, ϕ, h)) may be defined as (λ', ϕ', h') , though the notation of its basis elements may easily lead to confusion as to which variables are actually being referenced. It is therefore motivating to define new symbols for the ENV basis for consistency between itself and its coordinate variables as

$$\{\hat{e}_{\lambda'}, \hat{e}_{\phi'}, \hat{e}_{h'}\} \equiv \{\hat{e}_\lambda, \hat{e}_\phi, \hat{e}_h\}. \quad (2.55)$$

Since this reference frame is rather important for observations from Earth's surface, the ENV basis elements may also be referenced by another, albeit intuitive, choice of elements that do a better job of not shrouding information behind abstract symbolism.

$$\{\hat{E}, \hat{N}, \hat{V}\} \equiv \{\hat{e}_{\lambda'}, \hat{e}_{\phi'}, \hat{e}_{h'}\} \quad (2.56)$$

If the ENV coordinate system were placed at the point (x_s, y_s, z_s) , then $\hat{e}_{\lambda'}$ is tangent to the ϕ -parallel in the positive counterclockwise-like (eastward) direction, $\hat{e}_{\phi'}$ is tangent to the λ -meridian in the direction of the north pole, and $\hat{e}_{h'}$ is outward normal to the spheroid. Raising the basis up from the surface by a height h to the point (x, y, z) (where it belongs) changes nothing about the directions of the basis elements and how they relate to the parallels and meridians of the spheroid.

Considering an object \mathfrak{X} at (λ, ϕ, h) in the $\{\hat{i}, \hat{j}, \hat{k}\}$ basis and another object \mathfrak{X}' at (λ', ϕ', h') in the $\{\hat{e}_{\lambda'}, \hat{e}_{\phi'}, \hat{e}_{h'}\}$ basis, the object \mathfrak{X}' has components (x', y', z') in the $\{\hat{i}, \hat{j}, \hat{k}\}$ written

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \underbrace{\begin{pmatrix} -\sin \lambda & -\sin \phi \cos \lambda & \cos \phi \cos \lambda \\ \cos \lambda & -\sin \phi \sin \lambda & \cos \phi \sin \lambda \\ 0 & \cos \phi & \sin \phi \end{pmatrix}}_{\mathbf{T}^{\text{ecf/env}}} \begin{pmatrix} x_\lambda \\ x_\phi \\ x_h \end{pmatrix} \quad (2.57)$$

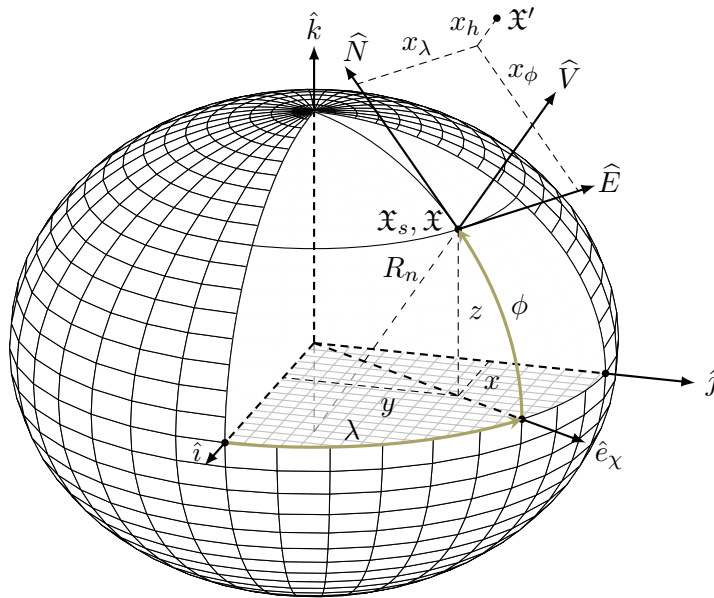


Fig. 2.10: The ENV coordinate system at a position on the surface of the spheroid ($h = 0$).

Tab. 2.5: ENV coordinate system defined to be located at (λ, ϕ, h) .

Axis	Variable	Basis	Definition
1	$x^1 (\lambda')$	$\partial_{x^1} (\hat{e}_{\lambda'})$	Tangent to the ϕ parallel pointing eastward
2	$x^2 (\phi')$	$\partial_{x^2} (\hat{e}_{\phi'})$	Tangent to the λ meridian pointing northwards
3	$x^3 (h')$	$\partial_{x^3} (\hat{e}_{h'})$	Normal to the spheroid at (u_s, v_s, w_s) pointing outwards

Since this coordinate system is derived from the spheroid, a smooth mathematical surface, the vertical direction $\hat{e}_{h'}$ is taken normal to the *spheroid's* surface at a longitude λ and latitude ϕ , *not* necessarily normal to the local surface of Earth at this location. The opposite direction of the local gravitational acceleration at this location, however, is a decent, albeit still incorrect, indicator of the vertical direction in practice.

Geocentric and geodetic coordinates are relatable, both on the spheroid at zero height and off at nonzero height, though the relations for these two cases are different. For an object *on the spheroid's surface* \mathfrak{X}_s at (x_s, y_s, z_s) , which is the special case of zero height, the two latitudes are directly related through (2.44).

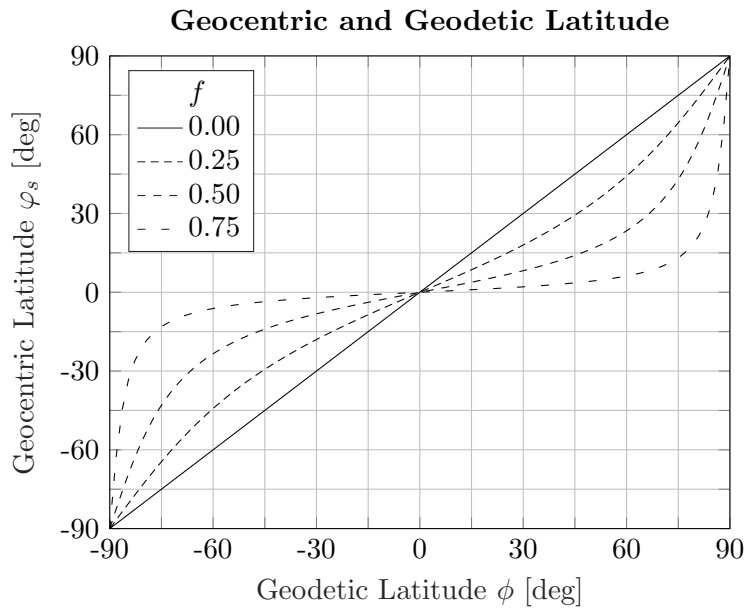
$$\tan \varphi_s = (1 - e^2) \tan \phi \quad (2.58)$$

Of course, both latitudes are centered at zero on the equator, so applying inverse tangents to either causes no issues. The geocentric radius (2.35) is then writable in terms of the geodetic latitude for an object at (x_s, y_s, z_s) as

$$r_s = a \sqrt{\frac{\cos^2 \phi + (1 - e^2)^2 \sin^2 \phi}{\cos^2 \phi + (1 - e^2) \sin^2 \phi}}, \quad (2.59)$$

which and likewise, the prime vertical radius (2.49) is writable in terms of the geocentric latitude for the same object at (x_s, y_s, z_s) .

$$R_n = a \sqrt{\frac{\cos^2 \varphi_s + (1 - e^2)^2 \sin^2 \varphi_s}{\cos^2 \varphi_s + (1 - e^2) \sin^2 \varphi_s}} \quad (2.60)$$

Fig. 2.11: Geocentric and geodetic latitude relationship as parameterized by the flattening f .

Now considering an object \mathfrak{X} at (λ, ϕ, h) , which is a height h above (x_s, y_s, z_s) , the geocentric radius r , with no subscript s , is given by the usual spherical representation as

$$r = \sqrt{(R_n + h)^2 \cos^2 \phi + ((1 - e^2)R_n + h)^2 \sin^2 \phi}, \quad (2.61)$$

and the geocentric latitude φ , again with no subscript, is also obtained by the simple spherical coordinate relation

$$\tan \varphi = \left(1 - e^2 \frac{R_n}{R_n + h}\right) \tan \phi. \quad (2.62)$$

In other words, the spherical coordinates $(r, \pi/2 - \varphi, \lambda)$ are directly recoverable given the position of a location on or above the surface of Earth (λ, ϕ, h) with respect to a reference ellipsoid as determined by GPS satellites. These relations should come really at no surprise since they are merely bypassing the ECEF (x, y, z) representation to which both spherical and geodetic coordinates are related.

As a final note about the spheroid, the quantity R_n is a radius of curvature in only one dimension. The *meridian radius (of curvature)*, denoted R_m , is given by the direct mathematical definition of radius of curvature applied to (2.29). The result yields

$$R_m = \frac{a(1 - e^2)}{(1 - e^2 \sin^2 \phi)^{3/2}}. \quad (2.63)$$

Together, the *mean curvature* H^* at a geodetic latitude ϕ is

$$H^* = \frac{1}{2} \left(\frac{1}{R_m} + \frac{1}{R_n} \right) = \frac{1}{2} \frac{\sqrt{1 - e^2 \sin^2 \phi} (-2 + e^2(1 + \sin^2 \phi))}{a(1 - e^2)}. \quad (2.64)$$

The average radius of the ellipsoid is taken simply as

$$R_{avg} = \sqrt{\frac{a^2 + b^2}{2}}. \quad (2.65)$$

This value for the average radius is typically used when choosing a reference radius for working in spherical coordinates applied to the earth, though the choice should still be explicitly stated to avoid ambiguity.

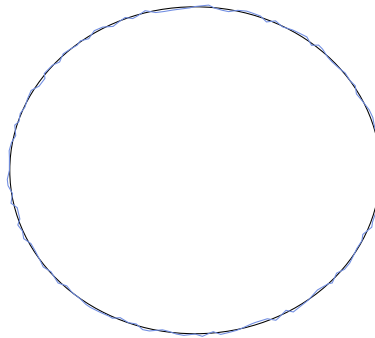


Fig. 2.12: An exaggerated two-dimensional visualization of an ellipsoid (black) fitted to the earth (blue).

3. EARTH MODEL

3.1 Reference Ellipsoid

The reference spheroid (2.28) is fully characterized by defining any two of a , b , f , or e . For Earth, the reference spheroid is typically defined by the equatorial radius a and flattening f (or inverse flattening $1/f$) and is also commonly called the *reference ellipsoid*, of which there exist several dating back so far into the early 17th century. Only a select few, however, stand out as real contenders in the modern study of geodesy.

Tab. 3.1: Some reference ellipsoids (spheroids).

Model	Equatorial a (m)	Polar b (m)	Flattening f
GRS80	6 378 137.0	6 356 752.3141(40)	1/298.257 222 101
WGS84	6 378 137.0	6 356 752.3142(45)	1/298.257 223 563
IERS03	6 378 136.6	6 356 751.8579(72)	1/298.256 42

The 1980 Geodetic Reference System (GRS80) ellipsoid is used with the International Terrestrial Reference System (ITRS), maintained by the IERS, to establish an accurate means of measuring positions near Earth’s surface. Such measuring with the ITRS means careful consideration of very fine details like the movement of Earth’s tectonic plates and resulting earthquakes, precession and nutation of the spin axis, etc. over the reference ellipsoid.

The 1984 World Geodetic System (WGS84) ellipsoid is perhaps the most commonly used reference ellipsoid in geodesy, astronomy, astronautics, and other like fields. It serves as the reference model for *all* GPS satellites, which measure ellipsoidal coordinates (λ, ϕ, h) over the entirety of Earth. With modern precision of GPS, the error in h between measured heights and actual heights is believed to be less than 2 cm [GISGeography WGS84]. Despite being so popularly used, WGS84 differs from GRS80 by only 0.1 mm on the polar radius.

The 2003 Internation Earth Rotation and Reference Systems Service (IERS03) ellipsoid is the current best estimate from the IERS of Earth’s reference ellipsoid. This model though is not used so commonly as GRS80 and WGS84 because of the dominant history that, especially WGS84, has held throughout several decades.

With these details considered, WGS84 is chosen as the reference ellipsoid because of its commonality in the field and use with GPS satellites. Despite this choice, caution should be taken regardless when using real data since these data may reference a different ellipsoid from WGS84. Despite the closeness of the models given above, the references *are* different, so mismatching reference ellipsoids between real data and modeling could lead to some, albeit slight, errors. Additional defining parameters of WGS84 are given below.

$$\omega = 0.000\,072\,921\,15 \text{ [rad/s]} \qquad GM = 3.986\,004\,418 \cdot 10^{14} \text{ [m}^3\text{/s}^2\text{]} \qquad (3.1)$$

where, according to WGS84, ω is the Earth’s rate of rotation and GM , also commonly denoted μ , is the gravitational parameter. The arbitrarily-chosen line of zero longitude passes 102 m east of the Greenwich meridian at the same (geodetic) latitude of the Royal Observatory in Greenwich, England. This line of zero longitude is defined and maintained by the IERS.

3.2 Gravitational Model

The *gravitational intensity*, or *gravity* [Geodesy], experienced by an object resting on the earth's surface is a result of both gravitational and centrifugal effects following

$$\vec{g} = \nabla(V + \Phi). \quad (3.2)$$

Here, V is the gravitational potential and Φ is the centrifugal (rotational) potential *for an object rotating with the earth*. As such, gravitational attraction on Earth's surface is actually *not* directed towards the earth's center.

To good approximation, the earth's atmosphere rotates along with the earth and can therefore be considered an extension of earth's surface itself such that the centrifugal potential remains a part of gravitation while in atmosphere. For high-orbiting satellites, this centrifugal potential drops away but the gravitational potential remains.

The gravitational potential V generally satisfies the Poisson's equation

$$\Delta V = 4\pi G\rho, \quad (3.3)$$

where G is the gravitational constant and ρ is the describing density function. Supposing the maximum attained air density is $10 \text{ [kg/m}^3\text{]}$ during atmospheric flight (a value greatly exceeding reality), this right member is still on the order of only $10^{-9} \text{ [m}^2\text{/s}^2\text{/m}^2\text{]}$. Thus, flight in both atmosphere and vacuum will be considered the same such that the gravitational potential is *harmonic*, i.e. solving Laplace's equation

$$\Delta V = 0. \quad (3.4)$$

Unlike the gravitational potential which can be difficult to obtain in practice, the centrifugal potential is actually known. On the surface of Earth rotating at a constant angular rate ω , the potential is written

$$\Phi = \frac{1}{2}\omega^2(x^2 + y^2), \quad (3.5)$$

where x and y are written with respect to the ECEF $\{\hat{i}, \hat{j}, \hat{k}\}$ basis. An air density ratio is applied to the gradient of centrifugal potential in (3.2) to include this effect continuously from flights beginning in atmosphere and ending in vacuum. This scaling is used to attenuate the contribution of Earth's rotation to the gravitational field while exiting the atmosphere such that the Earth and vehicle communicate only through gravitational forces once fully separated and in orbit.

$$\nabla\Phi \rightarrow \frac{\rho}{\rho_0}\nabla\Phi \quad (3.6)$$

The term ρ_0 here is a constant representing the actual air density at the launch site at no (zero) altitude from Earth's topological surface.

3.2.1 Newtonian Gravity (Smooth, Nonrotating Sphere)

The Newtonian model of gravity has long served as the workhorse for limiting relativistic (nonrelativistic) applications under the spherically symmetric potential (up to a sign) which indeed satisfies (3.4)

$$V = \frac{GM}{r}, \quad (3.7)$$

where r is the standard notation for radial distance from Earth's center (in place of the symbol ρ introduced in spherical coordinates). This symbol r will persist in place of ρ going forwards. The gradient of the gravitational potential in accordance with (3.2) is then

$$\nabla V = -\frac{GM}{r^2} \hat{e}_r, \quad (3.8)$$

This form is often taken as “the gravitational acceleration” without considering the component of rotational potential; Newtonian gravity is written as such.

$$\vec{g} = -\frac{GM}{r^2} \hat{e}_r \quad (3.9)$$

This model of gravity is a point mass representation of Earth, void of any details on the surface so that the potential is completely symmetric everywhere for a given radial distance r . Such a model does well when orbiting Earth from so far away that the oblateness has very little effect and mass fluctuations on the earth’s surface are unresolvable. Such a distance is much too far away for atmospheric and VLEO flights, though Newtonian gravity is still useful regardless. For low-altitude atmospheric flights, Earth’s oblateness plays little role again (as in the case of far away spacecraft) and mass fluctuations are almost zero if flying at least almost vertically. For this reason, an alternate form of (3.9) is

$$\vec{g} = -g_0 \left(\frac{R}{R+H} \right)^2 \hat{e}_r, \quad (3.10)$$

where the direction is implied “downwards”, g_0 is a reference value given at sea level a distance R from Earth’s center, and h is the elevation above sea level. *This representation, however, implies a truly spherical earth from the addition of R with h , which is not the case.* Though since the flattening of Earth is small (about 0.0033), this estimate is decent for small h (so that the earth’s topography is kept locally flat) with longitude and latitude held approximately the same as those for the launch site.

The internationally recognized value for gravitational acceleration at sea level, sourcing the practically famous quantity 9.81 [m/s²], is precisely

$$g_0 = 9.80665 \text{ [m/s}^2\text{]} \quad (3.11)$$

called *standard gravity*. While accepted by the International Service of Weights and Measures (ISWM) and true at (only) some locations on Earth because of oblateness, *this value, compatible with WGS84, is generally incorrect for the majority of actual applications in atmospheric rocket flight and in all applications of VLEO flight.* For low atmospheric flights with nearly vertical ascent, the basic estimate of gravity (3.10) uses a *locally* measured or calculated value g_0 which can be kept constant throughout. Otherwise, the oblateness of Earth requires that this term be more complicated.

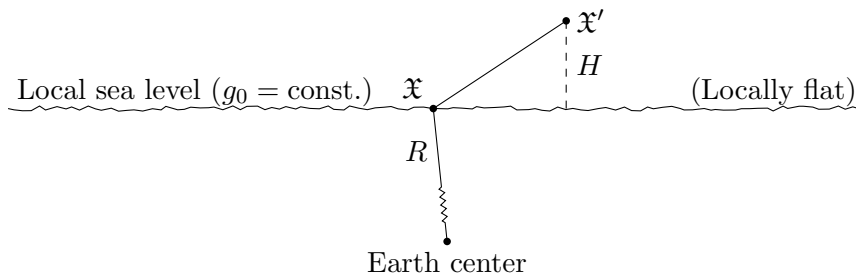


Fig. 3.1: Graphical representation of Newtonian gravity on the surface of Earth. The local elevation above sea level H , in general, is not in the same direction as the local direction to the earth’s center.

3.2.2 Normal Gravity (Smooth, Rotating Spheroid)

A *normal gravitational field* outfit with ellipsoidal parameters can be fit to model the Earth’s field, independent of longitude λ , as it would be if Earth were a smooth spheroid. The potential

of *normal gravity* is obtained by solving Laplace's equation (3.4) in spherical coordinates with symmetry about the spheroid's spin axis and equatorial plane. The solution is written

$$V = \frac{GM}{r} \left(1 - \sum_{n \text{ even}}^{\infty} \left(\frac{a}{r} \right)^n J_n P_n(\cos \theta) \right) \quad (n = 2, 4, \dots), \quad (3.12)$$

where $J_n = -C_{n0}$ is simply the negative of the zonal harmonics. With the rotational potential expressed in spherical coordinates as well, the normal gravity, denoted γ by convention (instead of g), is obtained by differentiation of the potential with respect to r .

$$\gamma = -\frac{GM}{r^2} \left(1 - \sum_{n \text{ even}}^{\infty} (1+n) \left(\frac{a}{r} \right)^n J_n P_n(\cos \theta) \right) + \omega^2 r \sin^2 \theta \quad (3.13)$$

Though a scalar quantity, the direction of γ is such that gravity on the spheroid's surface at the geodetic height $h = 0$ acts in the direction of the inwards normal to the spheroid. When extending off of the spheroid's surface for small heights h (to Earth's surface), the direction of gravity remains roughly normal to the spheroid's surface for which a series expansion in height h shows

$$\gamma = \sum_{n=0}^{\infty} \frac{h^n}{n!} \frac{\partial^n \gamma}{\partial h^n} \Big|_{h=0}. \quad (3.14)$$

In the space exterior to the spheroid, i.e. near Earth's surface, this derivative is evaluated via Bruns' equation [Geodesy] as

$$\frac{\partial \gamma}{\partial h} = -2\gamma H^* - 2\omega^2, \quad (3.15)$$

where H^* is the spheroid's mean curvature (2.64) [Geodesy]. With a series expansion dropping $O(f^2)$ terms, this derivative becomes

$$\frac{\partial \gamma}{\partial h} = -2\frac{\gamma}{a}(1 + f + m - 2f \sin^2 \phi). \quad (3.16)$$

By the work of Somigliana in 1924, the normal gravity *on the surface of the spheroid*, denoted γ_0 , is written

$$\gamma_0 = \frac{a\gamma_a \cos^2 \phi + b\gamma_b \sin^2 \phi}{\sqrt{a^2 \cos^2 \phi + b^2 \sin^2 \phi}}. \quad (3.17)$$

Because of the rapid convergence of (3.12) and compounding order of f^2 [Geodesy], however, the normal gravity *on the surface of the spheroid* (3.13) approximately (up to $n = 2$) is also

$$\gamma_0 = \gamma_a(1 + \beta \sin^2 \phi + \beta_1 \sin^2 2\phi), \quad (3.18)$$

where

$$\beta = \frac{\gamma_b - \gamma_a}{\gamma_a} \quad \beta_1 = \frac{1}{8}f^2 - \frac{5}{8}mf, \quad m = \frac{\omega^2 a^2 b}{GM}. \quad (3.19)$$

The quantity β , understood not to be confused with the reduced latitude β , is called the *gravity flattening* and m , not to be confused with any sort of mass, is approximately the ratio of centrifugal acceleration to normal gravity at the equator [Geodesy]. With zonal harmonics

$$J_2 = -\frac{1}{3}f^2 + \left(\frac{2}{3} + \frac{2}{21}m \right) f - \frac{m}{3} \quad J_4 = -\frac{4}{5}f^2 + \frac{4}{7}mf, \quad (3.20)$$

the gravity flattening may be estimated using (3.13) such that γ_b results when placing r to b and θ to 0° and likewise γ_a when r is a and θ is 90° since the derivative was taken with respect to r .

Using the WGS84 spheroid parameters, the normal gravity at the equator and poles of the WGS84 spheroid become

$$\gamma_a = -9.78058 \qquad \gamma_b = -9.83288. \quad (3.21)$$

In addition to these series-truncated approximations of equatorial and polar gravity, WGS84 provides its own official normal surface gravity (from which γ_a and γ_b are implied) following a slight rearrangement of (3.17), as [earth-info-nga.mil/GandG/.../WGS84/...]

$$\gamma_0 = -9.780\,326\,771\,4 \frac{1 + 0.001\,931\,851\,386\,39 \sin^2 \phi}{\sqrt{1 - 0.006\,694\,379\,990\,13 \sin^2 \phi}}. \quad (3.22)$$

At a small height h above the WGS84 surface at a geodetic latitude ϕ , where height and latitude are consistent with ellipsoidal coordinates, the gravity is, up to second order, accordingly [wgs84fin.pdf]

$$\gamma = \gamma_0 \left(1 - \frac{2}{a}(1 + f - 2f \sin^2 \phi)h + \frac{3}{a^2}h^2 \right), \quad (3.23)$$

where γ_0 is any of (3.17), (3.18), or (3.22). Though numerical differences do exist because of series truncation and approximations, the three forms are basically indistinguishable for all intents and purposes regarding atmospheric and VLEO flight, so whichever is most convenient may be used. These relations for normal gravity, however, fall apart when estimating gravitational accelerations too far beneath the spheroid's surface as they were derived for use in the spheroid's exterior and include no longitudinal variation.

3.2.3 Earth Gravity (Rough, Rotating Spheroid)

The most general gravitational potential is obtained by solving Laplace's equation (3.4) in spherical coordinates by assuming no simplifying symmetries unlike in normal gravity to obtain a solution in the form (B.18). Applying the appropriate boundary conditions yields the most general solution of the earth's gravitational potential.

$$V = \frac{GM}{r} \left(1 + \sum_{n=2}^{\infty} \sum_{m=0}^n \left(\frac{a}{r} \right)^n (C_{nm} \cos m\lambda + S_{nm} \sin m\lambda) P_{nm}(\cos \theta) \right) \quad (3.24)$$

The determination of these (*zonal, sectoral, and tesseral*) harmonic coefficients C_{nm} and S_{nm} is what makes up a particular gravity model for any planetary body paired with appropriate values of that body's gravitational parameter GM and semi-major axis a . Several such models are available. Some of the most recent Earth gravity models are Joint Gravity Model 3 (JGM-3), Earth Gravity Model 1996 (EGM96), and Earth Gravity Model 2008 (EGM2008), and an updated set of coefficients composing Earth Gravity Model 2020 (EGM2020) is anticipated.

Such detail is typically necessary only at low altitudes above the body's surface such that spaceflight modeling uses only the nearest body's gravity model, if available, upon close approach. Having the ability to describe minute gravitational attractions over a global topography is especially of great importance for low altitude orbital (i.e. VLEO) trajectories since measurable perturbations caused by local mass variations affect the spacecraft's flight path and attitude.

Though technically defined for a (geocentric) radius $r \geq a$ for guaranteed convergence, the gravitational potential (3.24) in actuality (with finitely-many terms) may be used on the Earth's true surface (at $r < a$) with minimal, if not negligible, error [wgs84fin.pdf]. Evaluating the potential under Earth's surface, however, steps outside of this formulation's framework and therefore begins to give wildly inaccurate results due to failure of convergence.

The gravity potential (3.24), like other models expressed in spherical harmonics, is (very) often converted to a "normalized" (also "fully normalized") form to avoid computational underflows

with the harmonic coefficients and overflows with the associated Legendre polynomials.

$$V = \frac{GM}{r} \left(1 + \sum_{n=2}^{\infty} \sum_{m=0}^n \left(\frac{a}{r} \right)^n (\bar{C}_{nm} \cos m\lambda + \bar{S}_{nm} \sin m\lambda) \bar{P}_{nm}(\cos \theta) \right) \quad (3.25)$$

This rewriting is nothing more than, albeit slightly fancy, termwise multiplication by unity which, of course, does not change the final outcome. Motivated by the orthogonality of the associated Legendre polynomials, a standard normalization in this context follows the form [wgs84fin, everywhere]

$$\bar{P}_{nm}(\cos \theta) = \sqrt{\frac{(2 - \delta_{0m})(2n+1)(n-m)!}{(n+m)!}} P_{nm}(\cos \theta) \quad (3.26)$$

$$\begin{pmatrix} \bar{C}_{nm} \\ \bar{S}_{nm} \end{pmatrix} = \sqrt{\frac{(n+m)!}{(2 - \delta_{0m})(2n+1)(n-m)!}} \begin{pmatrix} C_{nm} \\ S_{nm} \end{pmatrix}, \quad (3.27)$$

where δ_{ij} is the Kronecker delta and the associated Legendre polynomial P_{nm} contains the Condon-Shortley phase

$$P_{nm}(\cos \theta) = (-1)^m P_n^m(\cos \theta). \quad (3.28)$$

The first and largest harmonic coefficient, \bar{C}_{20} , is that corresponding to the flattening of Earth from a perfect sphere and is related to the popular so-called “ J_2 perturbation” by

$$J_2 = -\sqrt{5} \bar{C}_{20} \approx 0.001082. \quad (3.29)$$

The negation comes from convention that a general J_n perturbation is simply the negative of the n th degree zonal (zeroth order) harmonic coefficient C_{n0} . Hence, the prefactor $\sqrt{5}$ is simply the unnormalization of \bar{C}_{20} . This simple relation provides a quick check when using a set of gravity coefficients to determine if those provided are normalized or unnormalized, though the case is typically the former [Wiesel].

Carrying out numerical derivatives may or may not be more efficient than an analytic evaluation of the gradient depending on its implementation in live simulation. A numerical derivative requires only a three-dimensional matrix, or “multi-dimensional array,” to be constructed and then operated upon in its various dimensions, each representing a varying of spherical coordinate variables (say r as the arrays, θ as the rows, and λ as the columns). Thus, the gradient is obtained according to the usual relation in spherical coordinates

$$\nabla V = \frac{\partial V}{\partial r} \hat{e}_{r'} + \frac{1}{r} \frac{\partial V}{\partial \theta} \hat{e}_{\theta'} + \frac{1}{r \sin \theta} \frac{\partial V}{\partial \lambda} \hat{e}_{\lambda'}. \quad (3.30)$$

It is reiterated that the derivatives here are performed numerically (individually along the multi-dimensional array directions) about the point (r, θ, λ) . Otherwise, the gradient is obtained exactly as

$$\begin{aligned} \nabla V = & -\frac{GM}{r^2} \left(1 + \sum_{n=2}^{\infty} \sum_{m=0}^n (1+n) \left(\frac{a}{r} \right)^n (\bar{C}_{nm} \cos m\lambda + \bar{S}_{nm} \sin m\lambda) \bar{P}_{nm}(\cos \theta) \right) \hat{e}_{r'} \\ & + \frac{GM}{r^2} \sum_{n=2}^{\infty} \sum_{m=0}^n \left(\frac{a}{r} \right)^n (\bar{C}_{nm} \cos m\lambda + \bar{S}_{nm} \sin m\lambda) \frac{\partial \bar{P}_{nm}(\cos \theta)}{\partial \theta} \hat{e}_{\theta'} \\ & - \frac{GM}{r^2} \sum_{n=2}^{\infty} \sum_{m=0}^n m \left(\frac{a}{r} \right)^n (\bar{C}_{nm} \sin m\lambda - \bar{S}_{nm} \cos m\lambda) \frac{\bar{P}_{nm}(\cos \theta)}{\sin \theta} \hat{e}_{\lambda'}, \end{aligned} \quad (3.31)$$

where the derivative of the associated Legendre polynomial is [Nonlinear PDEs pg. 705, wikipedia]

$$\frac{\partial \bar{P}_{nm}(\cos \theta)}{\partial \theta} = -(n+m)(n-m+1) \bar{P}_{n(m-1)}(\cos \theta) - m \cot \theta \bar{P}_{nm}(\cos \theta). \quad (3.32)$$

In either approach, the gravity coefficients and associated Legendre polynomials must be read and computed respectively at least once per term; that is, one method may be more efficient than the other, but perhaps at the cost of accuracy. Supposing similar accuracy is desired between the two options, numerically evaluating the gradient is likely to be much more computationally expensive than doing a direct evaluation. The total gravity still includes the effects of centrifugal potential Φ such that, again in spherical coordinates,

$$\nabla\Phi = \omega^2 r \sin^2\theta \hat{e}_{r'} + \omega^2 r \sin\theta \cos\theta \hat{e}_{\theta'}. \quad (3.33)$$

Applying the atmospheric “corrective” term to try and smooth this effect out to vacuum, the total gravity experienced is

$$\vec{g} = \nabla V + (\rho/\rho_0)\nabla\Phi. \quad (3.34)$$

Earth Gravity Model 2008 Details and Corrections

The Earth Gravity Model released in 2008 (EGM2008) is consistent with TT (Terrestrial Time), compatible with the WGS84 ellipsoid (tide-free), and uses data collected by the GRACE orbiters to provide harmonic coefficients up to degree (n) 2190 and order (m) 2159. This model originally was of order and degree 2159 and expressed in ellipsoidal coordinates, but was converted to spherical coordinates which caused an extension of the maximum degree (2190) by necessity. Because of its statistical nature stemming from real data, *this gravity model describes the earth’s gravitational field very well*. The parameters describing the reference ellipsoid consistent with EGM2008 are exactly

$$GM = 3986004.415 \cdot 10^8 \text{ [m}^3/\text{s}^2] \quad a = 6378136.3 \text{ [m]}. \quad (3.35)$$

Depending on the required accuracy, however, coefficients may be cut out in extreme amounts to greatly shorten computation time. The IERS recommends that flights at about 950 [km] altitude (mid-LEO) may truncate the expansion at a degree of 90 whereas flights at an elevation of 5900 [km] (low MEO) may truncate at order 20 and GPS satellites at an altitude of about 20 200 [km] (mid/high MEO) may truncate it at a degree of 12 [Geopotential (EGM2008).pdf]. Though less accurate still, computations using a normal gravity model (obtained by either direct computation of the series-truncated formula (3.23) or by permanently setting the orders in all degrees of an EGM to zero) are generally faster than any EGM by nature of non-series or single-series solutions.

EGM2008 provides two sets of gravity coefficients $\{\bar{C}_{nm}, \bar{S}_{nm}\}$ to degree (n) 2190 and order (m) 2159, the first few of which are found in Tab. 3.7, that describe *tide-free* and *zero-tide* gravitational models concerning tidal effects, or rather the lack thereof. The tide-free gravitational model removes all direct and indirect tidal effects of the sun and moon. The zero-tide gravitational model removes all permanent direct effects of the sun and moon, but retains the indirect effect related to the earth’s elastic deformation. The *mean-tide* is the last of the three main models, which accounts for the permanent tidal effects of the sun and moon. The two sets of coefficients, tide-free and zero-tide, effected at the epoch 2008.0 are identical except for the second degree zonal harmonic coefficient \bar{C}_{20} .

Secular rates are provided for the (zonal) C_{n0} harmonics from their values at the epoch J2000, though the rate of advancement is quite slow, on the order of 10^{-12} per year, for the first three zonal harmonics [IERS Technical Note No. 36 (pg. 80)]. These rates form the gravitational coefficient at the current time t in linear fashion as

$$\bar{C}_{n0}(t) = \bar{C}_{n0}(t_0) + (t - t_0)\dot{\bar{C}}_{n0}(t_0). \quad (3.36)$$

Note that there are no given secular rates for the S_{nm} harmonics. A stronger dependence on time may be established, however, by considering the effects of the moon and sun, implying the

necessity of an ephemeris, in a 3 step procedure to account for the effects of the solid Earth tides. Step 1 begins by applying a corrective term to all of the low-degree coefficients ($n = 2, 3$), which holds for all m except \bar{C}_{20} , which requires additional corrections, as

$$\Delta C_{nm} - i\Delta S_{nm} = \frac{k_{nm}}{2n+1} \sum_{j=2}^3 \frac{\mu_j}{\mu_e} \left(\frac{R_e}{r_j} \right)^{n+1} \bar{P}_{nm}(\sin \Phi_j) e^{-im\lambda_j}. \quad (3.37)$$

Here, k_{nm} is the nominal Love number of degree n and order m given in Tab. 3.2, μ_j is the gravitational parameter (GM_j) of body j (bodies 1, 2, and 3 are the earth, moon, and sun, respectively), R_e is the earth's equatorial radius, and r_j , Φ_j , and λ_j are the ECEF distance, geocentric latitude, and longitude (east of Greenwich, England) of body j , respectively. Step 1 also includes a correction to the degree 4 coefficients induced by the degree 2 tides as

$$\Delta C_{4m} - i\Delta S_{4m} = \frac{k_{2m}^{(+)}}{5} \sum_{j=2}^3 \frac{\mu_j}{\mu_e} \left(\frac{R_e}{r_j} \right)^3 \bar{P}_{2m}(\sin \Phi_j) e^{-im\lambda_j}, \quad (m = 0, 1, 2). \quad (3.38)$$

The form describing the variation of the fourth degree coefficients is of the same form as (3.37), but states that the fourth degree coefficients are affected by the second degree coefficients by inserting $n = 4$ into the LHS and $n = 2$ into the RHS while also exchanging k_{nm} for $k_{nm}^{(+)}$. The Love numbers are given again by Tab. 3.2.

Tab. 3.2: Nominal values of solid Earth tide external potential Love numbers [IERS Technical Note 36]. Complex numbers z (k_{nm} of the anelastic earth) are separated into their real and imaginary parts such that $z = \text{Re } z + i \text{Im } z$.

Elastic Earth				Anelastic Earth		
n	m	k_{nm}	$k_{nm}^{(+)}$	$\text{Re } k_{nm}$	$\text{Im } k_{nm}$	$k_{nm}^{(+)}$
2	0	0.295 25	-0.000 87	0.301 90	0.000 00	-0.000 89
	1	0.294 70	-0.000 79	0.298 30	-0.001 44	-0.000 80
	2	0.298 01	-0.000 57	0.301 02	-0.001 30	-0.000 57
3	0	0.093				
	1	0.093				
	2	0.093				
	3	0.094				

Step 2 involves frequency-dependent corrections to the corrections ΔC_{nm} and ΔS_{nm} found in step 1. These frequency-dependent corrections are written

$$\Delta(\Delta \bar{C}_{2m} - i\Delta \bar{S}_{2m}) = (-i)^m \sum_{f(2,m)} (A_m \delta k_f H_f) e^{i\theta_f}, \quad (m = 0, 1, 2). \quad (3.39)$$

Here,

$$A_m = \begin{cases} (\sqrt{4\pi} R_e)^{-1} & (m = 0) \\ (-1)^m (\sqrt{8\pi} R_e)^{-1} & (m > 0), \end{cases} \quad (3.40)$$

$\delta k_f = k_f - k_{2m}$ is the difference between the Love number at frequency f (k_f , or $k_{2m}^{(0)}$) and the nominal value (k_{2m}) and H_f is the amplitude (in meters) of the term at frequency f from the harmonic expansion of the tide generating potential in accordance with Cartwright and Tayler (1971). The quantity A_0 is explicitly given to be $4.4228 \cdot 10^{-8}$ [1/m], implying $R_e = 6378194.623$ [m] is taken to be Earth's equatorial radius.

The real part of $A_m \delta k_f H_f$ is called *amplitude in phase (ip)* and the imaginary part of $A_m \delta k_f H_f$ is called *amplitude out of phase (op)*. The angle θ_f is given in terms of the five

fundamental arguments (Delaunay variables) of nutation theory F_j and their (integer) multipliers N_j as

$$\theta_f = m(\theta_g + \pi) - \sum_{j=1}^5 N_j F_j, \quad (3.41)$$

where θ_g is the Greenwich Mean Sidereal Time (GMST) (2.5) and the five Delaunay variables F_j , $(F_1, F_2, F_3, F_4, F_5) = (l, l', F, D, \Omega)$, are

$$l = 134.96340251^\circ + 1717915923.2178'' t + 31.8792'' t^2 + 0.051635'' t^3 - 0.00024470'' t^4 \quad (3.42)$$

$$l' = 357.52910918^\circ + 129596581.0481'' t - 0.5532'' t^2 + 0.000136'' t^3 - 0.00001149'' t^4 \quad (3.43)$$

$$F = 93.27209062^\circ + 1739527262.8478'' t - 12.7512'' t^2 - 0.001037'' t^3 + 0.00000417'' t^4 \quad (3.44)$$

$$D = 297.85019547^\circ + 1602961601.2090'' t - 6.3706'' t^2 + 0.006593'' t^3 - 0.00003169'' t^4 \quad (3.45)$$

$$\Omega = 125.04455501^\circ - 6962890.5431'' t + 7.4722'' t^2 + 0.007702'' t^3 - 0.00005939'' t^4, \quad (3.46)$$

where t is measured in Dynamical Time Julian centuries since J2000. The Delaunay variables describe the mean anomaly of the moon (F_1), the mean anomaly of the sun (F_2), the mean longitude of the moon less the mean longitude of the ascending node of the moon ($F_3 = L - F_5$), the mean elongation of the moon from the sun (F_4), and the mean longitude of the ascending node of the moon (F_5) [iers_1996_conventions.pdf].

In this correction, $m = 0$ ($\Delta\Delta\bar{C}_{20}$) accounts for long period tidal constituents of various frequencies f , $m = 1$ ($\Delta(\Delta\bar{C}_{21} - i\Delta\bar{S}_{21})$) for diurnal tidal constituents, and $m = 2$ ($\Delta(\Delta\bar{C}_{22} - i\Delta\bar{S}_{22})$) for semidiurnals. Of course, the relation for $m = 0$ may yield a nonzero $\Delta\Delta\bar{S}_{20}$, and this “correction” may be added to the permanent value $\bar{S}_{20}(t_0)$, but its resulting effect on the gravitational potential (3.25) is necessarily nothing. The appropriate Delaunay multipliers and in/out-of phase amplitudes for each of the corrections follow the following tables.

Tab. 3.3: Corrections for frequency dependence of $k_{20}^{(0)}$ of the zonal tides due to anelasticity. Units for the amplitude (*ip* and *op*) are 10^{-12} , where each amplitude is respectively given by the real and imaginary parts of $(A_0 H_f \delta k_f)$ (both A_0 and H_f are real numbers). Units for δk_f are 10^{-5} . The nominal value k_{20} for the zonal tides is taken as 0.30190. The amplitudes are to be used in determining the $\Delta\Delta\bar{C}_{20}$ correction.

Tide	Doodson Number	Frequency f (deg/hr)	N_1	N_2	N_3	N_4	N_5	Re δk_f	Im δk_f	In phase Amplitude	Out of phase Amplitude
S_a	55.565	0.002 21	0	0	0	0	1	0.013 47	-0.005 41	16.6	-6.7
	55.575	0.004 41	0	0	0	0	2	0.011 24	-0.004 88	-0.1	0.1
	56.554	0.041 07	0	-1	0	0	0	0.005 47	-0.003 49	-1.2	0.8
	57.555	0.082 14	0	0	-2	2	-2	0.004 03	-0.003 15	-5.5	4.3
	57.565	0.084 34	0	0	-2	2	-1	0.003 98	-0.003 13	0.1	-0.1
M_{sm}	58.554	0.123 20	0	-1	-2	2	-2	0.003 26	-0.002 96	-0.3	0.2
	63.655	0.471 52	1	0	0	-2	0	0.001 01	-0.002 42	-0.3	0.7
	65.445	0.542 17	-1	0	0	0	-1	0.000 80	-0.002 37	0.1	-0.2
M_m	65.445	0.544 38	-1	0	0	0	0	0.000 80	-0.002 37	-1.2	3.7
	65.465	0.546 58	-1	0	0	0	1	0.000 79	-0.002 37	0.1	-0.2
	65.655	0.553 66	1	0	-2	0	-2	-0.000 77	-0.002 36	0.1	-0.2
M_{sf}	73.555	1.015 90	0	0	0	-2	0	-0.000 09	-0.002 16	0.0	0.6
	75.355	1.088 75	-2	0	0	0	0	-0.000 18	-0.002 13	0.0	0.3
	75.555	1.098 04	0	0	-2	0	-2	-0.000 19	-0.002 13	0.6	6.3
M_f	75.565	1.100 24	0	0	-2	0	-1	-0.000 19	-0.002 13	0.2	2.6
	75.575	1.102 45	0	0	-2	0	0	-0.000 19	-0.002 13	0.0	0.2
	83.655	1.569 56	1	0	-2	-2	-2	-0.000 65	-0.002 02	0.1	0.2
M_{stm}	85.455	1.642 41	-1	0	-2	0	-2	-0.000 71	-0.002 01	0.4	1.1
	85.465	1.644 62	-1	0	-2	0	-1	-0.000 71	-0.002 01	0.2	0.5
M_{sqm}	93.555	2.113 94	0	0	-2	-2	-2	-0.001 02	-0.001 93	0.1	0.2
M_{qm}	95.355	2.186 79	-2	0	-2	0	-2	-0.001 06	-0.001 92	0.1	0.1

Tab. 3.4: Corrections for frequency dependence of $k_{21}^{(0)}$. Units for the amplitude (ip and op) are 10^{-12} , where each amplitude is respectively given by the real and imaginary parts of $(A_1 H_f \delta k_f)$ (both A_1 and H_f are real numbers). Units for δk_f are 10^{-5} . The nominal value k_{21} for the diurnal tides is taken as $0.29830 - 0.00144i$. The amplitudes are to be used in determining the $\Delta\Delta\bar{C}_{21}$ correction.

Tide	Doodson Number	Frequency f (deg/hr)	N_1	N_2	N_3	N_4	N_5	Re δk_f	Im δk_f	In phase Amplitude	Out of phase Amplitude
$2Q_1$	125.755	12.854 29	2	0	2	0	2	-29	3	-0.1	0.0
σ_1	127.555	12.927 14	0	0	2	2	2	-30	3	-0.1	0.0
	135.645	13.396 45	1	0	2	0	1	-45	5	-0.1	0.0
Q_1	135.655	13.398 66	1	0	2	0	2	-46	5	-0.7	0.1
ρ_1	137.455	13.471 51	-1	0	2	2	2	-49	5	-0.1	0.0
	145.545	13.940 83	0	0	2	0	1	-82	7	-1.3	0.1
O_1	145.555	13.943 03	0	0	2	0	2	-83	7	-6.8	0.6
τ_1	147.555	14.025 17	0	0	0	2	0	-91	9	0.1	0.0
$N\tau_1$	153.655	14.414 56	1	0	2	-2	2	-168	14	0.1	0.0
	155.445	14.485 20	-1	0	2	0	1	-193	16	0.1	0.0
Lk_1	155.455	14.487 41	-1	0	2	0	2	-194	16	0.4	0.0
No_1	155.655	14.496 69	1	0	0	0	0	-197	16	1.3	-0.1
	155.665	14.498 90	1	0	0	0	1	-198	16	0.3	0.0
χ_1	157.455	14.569 55	-1	0	0	2	0	-231	18	0.3	0.0
	157.465	14.561 76	-1	0	0	2	1	-233	18	0.1	0.0
π_1	162.556	14.917 87	0	1	2	-2	2	-834	58	-1.9	0.1
	163.545	14.956 73	0	0	2	-2	1	-1117	76	0.5	0.0
P_1	163.555	14.958 93	0	0	2	-2	2	-1138	77	-43.4	2.9
	164.554	15.000 00	0	-1	2	-2	2	-1764	104	0.6	0.0
S_1	164.556	15.000 00	0	1	0	0	0	-1764	104	1.6	-0.1
	165.345	15.029 58	-2	0	2	0	1	-3048	92	0.1	0.0
	165.535	15.036 65	0	0	0	0	-2	-3630	195	0.1	0.0
	165.545	15.038 86	0	0	0	0	-1	-3845	229	-8.8	0.5
K_1	165.555	15.071 07	0	0	0	0	0	-4084	262	470.9	-30.2
	165.565	15.043 28	0	0	0	0	1	-4355	197	68.1	-4.6
	165.575	15.045 58	0	0	0	0	2	-4665	334	-1.6	0.1
	166.455	15.077 49	-1	0	0	1	0	85 693	21 013	0.1	0.0
	166.544	15.079 93	0	-1	0	0	-1	35 203	2084	-0.1	0.0
ψ_1	166.554	15.082 14	0	-1	0	0	0	22 794	358	-20.6	-0.3
	166.556	15.082 14	0	1	-2	2	-2	22 780	358	0.3	0.0
	166.564	15.084 34	0	-1	0	0	1	16 842	-52	-0.3	0.0
	167.355	15.113 92	-2	0	0	2	0	3755	-189	-0.2	0.0
	167.365	15.116 13	-2	0	0	2	1	3552	-182	-0.1	0.0
ϕ_1	167.555	15.123 21	0	0	-2	2	-2	3025	-160	-5.0	0.3
	167.565	15.125 42	0	0	-2	2	-1	2892	-154	0.2	0.0
	168.554	15.164 27	0	-1	-2	2	-2	1638	-93	-0.2	0.0
θ_1	173.655	15.512 59	1	0	0	-2	0	370	-20	-0.5	0.0
	173.665	15.514 80	1	0	0	-2	1	369	-20	-0.1	0.0
	175.445	15.583 23	-1	0	0	0	-1	325	-17	0.1	0.0
J_1	175.455	15.585 45	-1	0	0	0	0	324	-17	-2.1	0.1
	175.465	15.587 65	-1	0	0	0	1	323	-16	-0.4	0.0
So_1	183.555	16.056 97	0	0	0	-2	0	194	-8	-0.2	0.0
	185.355	16.129 89	-2	0	0	0	0	185	-7	-0.1	0.0
Oo_1	185.555	16.139 11	0	0	-2	0	-2	184	-7	-0.6	0.0
	185.565	16.141 31	0	0	-2	0	-1	184	-7	-0.4	0.0
	185.575	16.143 52	0	0	-2	0	0	184	-7	-0.1	0.0
ν_1	195.455	16.683 48	-1	0	-2	0	-2	141	-4	-0.1	0.0
	195.465	16.685 69	-1	0	-2	0	-1	141	-4	-0.1	0.0

Tab. 3.5: Corrections for frequency dependence of $k_{22}^{(0)}$. Units for the amplitude (*ip* and *op*) are 10^{-12} , where each amplitude is respectively given by the real and imaginary parts of $(A_2 H_f \delta k_f)$ (both A_2 and H_f are real numbers). The nominal value k_{22} for the sectorial tides is taken as $0.30102 - 0.00130i$. The amplitude is to be used in determining the $\Delta\Delta\bar{C}_{22}$ correction — the corrections are only to the real part.

Tide	Doodson Number	Frequency f (deg/hr)	N_1	N_2	N_3	N_4	N_5	$\text{Re } \delta k_f$	$\text{Im } \delta k_f$	(In phase) Amplitude	(Out of phase) Amplitude
N_2	245.655	28.439 73	1	0	2	0	2	0.000 06	0	−0.3	0.0
M_2	255.555	28.984 10	0	0	2	0	2	0.000 04	0	−1.2	0.0

The total variations of the gravitational coefficients \bar{C}_{2m} and \bar{S}_{2m} accounting for tidal variations are obtained by adding their corrections obtained from (3.39) using the appropriate tables of coefficients (Tab. 3.3 for $\Delta\Delta\bar{C}_{20}$, Tab. 3.4 for $\Delta\Delta\bar{C}_{21}$ and $\Delta\Delta\bar{S}_{21}$, and Tab. 3.5 for $\Delta\Delta\bar{C}_{22}$ and $\Delta\Delta\bar{S}_{22}$) to the corrections (3.37) from step 1. The corrections to \bar{C}_{4m} and \bar{S}_{4m} for $m = 0, 1, 2$ via (3.38) remain as is and are not corrected any further in step 2.

Step 3 ensures that the permanent tide is not doubly counted if using the zero-tide model. If using the zero-tide model and its coefficients, then the corrective term to \bar{C}_{20} be used in place of (3.37) is

$$\Delta C_{20}^{zt} = \Delta\bar{C}_{20} - \Delta\bar{C}_{20}^{perm}, \quad (3.47)$$

where $\Delta\bar{C}_{20}$ is what results from (3.37) and $\Delta\bar{C}_{20}^{perm}$ is the correction as to not count the time-independent part twice, which is written

$$\Delta\bar{C}_{20}^{perm} = A_0 H_0 k_{20}. \quad (3.48)$$

Assuming the same values from the free-tide model, the permanent zero tide contribution amounts to

$$\Delta\bar{C}_{20}^{perm} = (4.4228 \cdot 10^{-8})(-0.31460)(0.29525) = -4.1081465282 \cdot 10^{-9}. \quad (3.49)$$

Step 3 is otherwise completely avoided if the corrections are applied with a tide-free model since the coefficients are identically the same *except* C_{20} in EGM2008 [IERS Technical Note 36].

The effect of the ocean tides on the earth's gravitational field are again treated as variations on the gravitational coefficients $\{C_{nm}, S_{nm}\}$ provided by EGM2008 of the (periodic) form

$$\Delta\bar{C}_{nm} - i\Delta\bar{S}_{nm} = \sum_f \sum_+ (\mathcal{C}_{f,nm}^\pm \mp i\mathcal{S}_{f,nm}^\pm) e^{\pm i\theta_f}, \quad (3.50)$$

which are implicitly dependent on time because of the ephemeris involved with θ_f . The coefficients $\{\mathcal{C}_{f,nm}^\pm, \mathcal{S}_{f,nm}^\pm\}$ for the main tidal waves $\{S_a, S_{sa}, M_m, M_f, M_{tm}, M_{sgm}\}$ (long period waves), $\{Q_1, O_1, P_1, K_1\}$ (diurnal waves), $\{2N_2, N_2, M_2, T_2, S_2, K_2\}$ (semi-diurnal waves) and $\{M_4\}$ (quarter-diurnal wave) are given by FES2004 directly [IERS Technial Note 36]. The first few coefficients of the first few tidal modes are shown in Tab. 3.6. That is, use of Tab. 3.6 in (3.50) is sufficient to find the gravitational coefficient variations, but FES2004 provides more coefficients that may be used in the theory of tides to incorporate more tidal modes if needed.

Tab. 3.6: The first few FES2004 harmonic coefficients of wave modes modelling the effects of ocean tides on the gravitational coefficients. The full model is of maximum degree and order 100 for a given tidal mode and does not include the atmospheric tide. These coefficients may be used directly in (3.50).

Tide	D n	O m	Doodson Number	$\bar{C}_{f,nm}^+$	$\bar{S}_{f,nm}^+$	$\bar{C}_{f,nm}^-$	$\bar{S}_{f,nm}^-$
O_{m_1}	2	0	55.565	6.581 28	0.000 00	0.000 00	0.000 00
O_{m_2}	2	0	55.575	-0.063 30	0.000 00	0.000 00	0.000 00
S_a	2	1	56.554	0.003 57	0.000 75	0.003 41	-0.001 31
\vdots							
S_{sa}	2	1	57.555	-0.026 18	-0.005 64	-0.023 19	0.007 42
\vdots							
M_m	2	1	65.455	-0.122 92	-0.116 30	0.026 60	0.014 79
\vdots							
M_4	2	1	455.555	0.009 55	0.001 92	0.007 19	-0.001 30

Tab. 3.7: EGM2008 tide-free, fully normalized, dimensionless harmonic coefficients \bar{C}_{nm} and \bar{S}_{nm} up to degree and order 5 and the secular rates up to degree 4. All zeroth order (zonal harmonic) terms \bar{S}_{n0} are unnecessary but written as zeros to eliminate any ambiguity. The set of gravitational coefficients in the zero-tide model is exactly the same as the tide-free model (shown below) for all n and m except $(n, m) = (2, 0)$ specifically, for which $\bar{C}_{20} = -0.484\,169\,317\,366\,974 \cdot 10^{-3}$.

Degree n	Order m	Gravitational Coeff. \bar{C}_{nm} (Nominal)	Gravitational Coeff. \bar{S}_{nm} (Nominal)	Secular Coeff. $\dot{\bar{C}}_{n0}(\text{J2000})$ (1/yr)
2	0	$-0.484\,165\,143\,790\,815 \cdot 10^{-3}$	0.000 000 000 000 000	$11.6 \cdot 10^{-12}$
	1	$-0.206\,615\,509\,074\,176 \cdot 10^{-9}$	$0.138\,441\,389\,137\,979 \cdot 10^{-8}$	
	2	$0.243\,938\,357\,328\,313 \cdot 10^{-5}$	$-0.140\,027\,370\,385\,934 \cdot 10^{-5}$	
3	0	$0.957\,161\,207\,093\,473 \cdot 10^{-6}$	0.000 000 000 000 000	$4.9 \cdot 10^{-12}$
	1	$0.203\,046\,201\,047\,864 \cdot 10^{-5}$	$0.248\,200\,415\,856\,872 \cdot 10^{-6}$	
	2	$0.904\,787\,894\,809\,528 \cdot 10^{-6}$	$-0.619\,005\,475\,177\,618 \cdot 10^{-6}$	
	3	$0.721\,321\,757\,121\,568 \cdot 10^{-6}$	$0.141\,434\,926\,192\,941 \cdot 10^{-5}$	
4	0	$0.539\,965\,866\,638\,991 \cdot 10^{-6}$	0.000 000 000 000 000	$4.7 \cdot 10^{-12}$
	1	$-0.536\,157\,389\,388\,867 \cdot 10^{-6}$	$-0.473\,567\,346\,518\,086 \cdot 10^{-6}$	
	2	$0.350\,501\,623\,962\,649 \cdot 10^{-6}$	$0.662\,480\,026\,275\,829 \cdot 10^{-6}$	
	3	$0.990\,856\,766\,672\,321 \cdot 10^{-6}$	$-0.200\,956\,723\,567\,452 \cdot 10^{-6}$	
	4	$-0.188\,519\,633\,023\,033 \cdot 10^{-6}$	$0.308\,803\,882\,149\,194 \cdot 10^{-6}$	
5	0	$0.686\,702\,913\,736\,681 \cdot 10^{-7}$	0.000 000 000 000 000	
	1	$-0.629\,211\,923\,042\,529 \cdot 10^{-7}$	$-0.943\,698\,073\,395\,769 \cdot 10^{-7}$	
	2	$0.652\,078\,043\,176\,164 \cdot 10^{-6}$	$-0.323\,353\,192\,540\,522 \cdot 10^{-6}$	
	3	$-0.451\,847\,152\,328\,843 \cdot 10^{-6}$	$-0.214\,955\,408\,306\,046 \cdot 10^{-6}$	
	4	$-0.295\,328\,761\,175\,629 \cdot 10^{-6}$	$0.498\,070\,550\,102\,351 \cdot 10^{-7}$	
	5	$0.174\,811\,795\,496\,002 \cdot 10^{-6}$	$-0.669\,379\,935\,180\,165 \cdot 10^{-6}$	

3.3 Magnetic Model

The World Magnetic Model (WMM) provides a realization of the earth's resulting *primary internal* magnetic field which is largely by the internal movement of its liquid iron outer core and provided contributions due to magnetic materials in earth's surface and induced electric currents in sea water [WMM2020_Report]. Secondary internal sources are excluded from the WMM because they are typically induced by external effects, such as the primary internal magnetic field interacting with radiation (van Allen belts). These external effects propagate back towards the surface, inducing electric currents into the earth and oceans, which induce the secondary internal magnetic fields, and so on. These secondary and other effects are grouped together to form the *disturbing field* which is *not* accounted for in the WMM.

Because the sources of the primary internal fields are forever evolving, a provided WMM naturally depends on time from an epoch and eventually expires at which point the model is updated providing a new expiry date.

The magnetic field strength and the associated field directions vary in both space and time as a result of the evolving sources. As a result, the locations, and therefore the local directions, of the magnetic poles, referred to as the *magnetic north* and *magnetic south* poles, differ from the true north and south pole (which are both skewered by the spin axis) in general.

Like the earth's gravitational field, its magnetic field permeates throughout the space surrounding it and, albeit much more strongly, varies in time. Unlike the gravitational field, however, the magnetic field B is split into three contributions — the primary internal field resulting from motions of the earth's liquid core (*core field*), resulting fields from Earth's crust/upper mantle (*crustal field*), and the disturbance field.

$$\vec{B} = \vec{B}_{core} + \vec{B}_{crust} + \vec{B}_{distur}. \quad (3.51)$$

The core field is related to a (time-dependent) magnetic potential function satisfying Laplace's equation via

$$\vec{B}_{core} = -\nabla\Psi. \quad (3.52)$$

In geocentric coordinates (r, θ, λ) where r is the radial displacement from Earth's center, θ is the colatitude, and λ is the longitude, the (harmonic) potential takes the form

$$\Psi = a \sum_{n=1}^{\infty} \sum_{m=0}^n \left(\frac{a}{r}\right)^{n+1} (g_n^m \cos m\lambda + h_n^m \sin m\lambda) \check{P}_n^m(\cos \theta). \quad (3.53)$$

Here, a is the geomagnetic reference radius, the time-dependence is encapsulated in the harmonic coefficients g_n^m and h_n^m , and \check{P}_n^m is the Schmidt semi-normalized associated Legendre polynomial of order n and m . The harmonic coefficients $\{g_n^m, h_n^m\}$ have an explicit linear dependence on time as simply

$$g_n^m = g_n^m{}_0 + (t - t_0) \dot{g}_n^m{}_0 \quad (3.54)$$

$$h_n^m = h_n^m{}_0 + (t - t_0) \dot{h}_n^m{}_0. \quad (3.55)$$

The normalization (and symbolism) of the associated Legendre polynomials in geomagnetics differs from the normalization used other fields of geophysics as

$$\check{P}_n^m(\cos \theta) = \sqrt{(2 - \delta_{0m}) \frac{(n - m)!}{(n + m)!}} P_{nm}(\cos \theta), \quad (3.56)$$

where δ_{ij} is the Kronecker delta and P_{nm} contains the Condon-Shortley phase, or explicitly

$$P_{nm}(\cos \theta) = (-1)^m P_n^m(\cos \theta). \quad (3.57)$$

The contribution of magnetic fields from the earth's crust can strongly influence the field strength locally. An exact mathematical relationship is not quite yet establishable, but these contributions can be encapsulated into the harmonic coefficients describing the core field depending on the resolution of the model. That is, a more highly resolved model can include effects from the crustal field creating magnetic waves with shorter wavelengths. Magnetic wavelengths shorter than what the resolution permits the model to “see” are largely left alone as they do not belong to the core field and are not part of the disturbance field.

The disturbance field is not included in the magnetic model since its causes depend on time, change frequently, and are difficult to predict.

As the core potential is written in geocentric spherical coordinates, the action of the gradient naturally places the resulting components in a spherical basis. These components are better served to be transformed into a local ENV coordinate system to reference magnetic north against true north, though this is unnecessary if only the magnitude of the magnetic field is sought. The gradient in spherical coordinates follows is

$$\nabla\Psi = \frac{\partial\Psi}{\partial r}\hat{e}_{r'} + \frac{1}{r}\frac{\partial\Psi}{\partial\theta}\hat{e}_{\theta'} + \frac{1}{r\sin\theta}\frac{\partial\Psi}{\partial\lambda}\hat{e}_{\lambda'} \quad (3.58)$$

Like in the gravitational potential, this gradient can be computed numerically or exactly (point-wise). In the case of the latter, the gradient is explicitly written

$$\nabla\Psi = -\sum_{n=1}^{\infty}\sum_{m=0}^n(n+1)\left(\frac{a}{r}\right)^{n+2}(g_n^m\cos m\lambda + h_n^m\sin m\lambda)\check{P}_n^m(\cos\theta)\hat{e}_{r'} \quad (3.59)$$

$$+ \sum_{n=1}^{\infty}\sum_{m=0}^nm\left(\frac{a}{r}\right)^{n+2}(g_n^m\cos m\lambda + h_n^m\sin m\lambda)\frac{\check{P}_n^m(\cos\theta)}{\partial\theta}\hat{e}_{\theta'} \quad (3.60)$$

$$- \sum_{n=1}^{\infty}\sum_{m=0}^n\left(\frac{a}{r}\right)^{n+2}(g_n^m\sin m\lambda - h_n^m\cos m\lambda)\frac{\check{P}_n^m(\cos\theta)}{\sin\theta}\hat{e}_{\lambda'}, \quad (3.61)$$

where the derivative of the associated Legendre polynomial follows

$$\frac{\partial\check{P}_n^m(\cos\theta)}{\partial\theta} = (n-m+1)\check{P}_{n+1}^m(\cos\theta)\csc\theta - (n+1)\check{P}_n^m(\cos\theta)\sec\theta \quad (3.62)$$

and the harmonic coefficients are still dependent on time.

3.3.1 World Magnetic Model 2020

The 2020 World Magnetic Model (WMM2020) provides a complete set of coefficients to describe the core field and crustal field (with a resolution permitting wavelengths exceeding 28.8° arc-length, or 3200 kilometers at the earth's surface, to be included with the core field) up to degree and order 12. The geomagnetic reference radius a is defined to be

$$a = 6\,371\,200 \text{ [m]}. \quad (3.63)$$

The harmonic coefficients capturing the magnetic field due to the core and partially of the crust are given by Tab. 3.8. These 180 coefficients make up the entirety of WMM2020. The epoch t_0 is taken to be exactly 2020.0, where time is counted in fractions of a year for compatibility with the secular values.

WMM2020 is prescribed to expire on January 1, 2025 at 00:00:000.000 UTC (precisely at 2025.0), though the time zone is really of no importance. Shortly before this expiration date, the next WMM (WMM2025) will be released. For reference, WMM2020 was released on December 19, 2019 and became the active model, replacing WMM2015, nearly two weeks later.

Tab. 3.8: The complete set of magnetic coefficients and their secular rates (time derivatives) of the 2020 World Magnetic Model. The epoch is 2020.0, or January 1, 2020 00:00:00.000 UTC. Time t is counted in fractions of years to be used with the secular rates since is taken to be exactly $t_0 = 2020.0$. The harmonic degree is labelled ‘D’ and the order is labeled ‘O’ accordingly. WMM2020 is officially set to expire on 2025.0, or January 1, 2025 00:00:00.000 UTC.

D	O	Epoch Coeffs. (nT)		Secular Rates (nT/yr)	
n	m	$g_{n\ 0}^m$	$h_{n\ 0}^m$	$\dot{g}_{n\ 0}^m$	$\dot{h}_{n\ 0}^m$
1	0	-29 404.5	0.0	6.7	0.0
	1	-1450.7	4652.9	7.7	-25.1
2	0	-2500.0	0.0	-11.5	0.0
	1	2982.0	-2991.6	-7.1	-30.2
3	0	1676.8	-734.8	-2.2	-23.9
	1	1363.9	0.0	2.8	0.0
4	0	-2381.0	-82.2	-6.2	5.7
	1	1236.2	241.8	3.4	-1.0
5	0	525.7	-542.9	-12.2	1.1
	1	903.1	0.0	-1.1	0.0
6	0	809.4	282.0	-1.6	0.2
	1	86.2	-158.4	-6.0	6.9
7	0	-309.4	199.8	5.4	3.7
	1	47.9	-350.1	-5.5	-5.6
8	0	-234.4	0.0	-0.3	0.0
	1	363.1	47.7	0.6	0.1
9	0	187.8	208.4	-0.7	2.5
	1	-140.7	-121.3	0.1	-0.9
10	0	-151.2	32.2	1.2	3.0
	1	13.7	99.1	1.0	0.5
11	0	65.9	0.0	-0.6	0.0
	1	65.6	-19.1	-0.4	0.1
12	0	73.0	25.0	0.5	-1.8
	1	-121.5	52.7	1.4	-1.4
13	0	-36.2	-64.4	-1.4	0.9
	1	13.5	9.0	0.0	0.1
14	0	-64.7	68.1	0.8	1.0
	1	80.6	0.0	-0.1	0.0
15	0	-76.8	-51.4	-0.3	0.5
	1	-8.3	-16.8	-0.1	0.6
16	0	56.5	2.3	0.7	-0.7
	1	15.8	23.5	0.2	-0.2
17	0	6.4	-2.2	-0.5	-1.2
	1	-7.2	-27.2	-0.8	0.2
18	0	9.8	-1.9	1.0	0.3
	1	23.6	0.0	-0.1	0.0
19	0	9.8	8.4	0.1	-0.3
	1	-17.5	-15.3	-0.1	0.7
20	0	-0.4	12.8	0.5	-0.2
	1	-21.1	-11.8	-0.1	0.5
21	0	15.3	14.9	0.4	-0.3
	1	13.7	3.6	0.5	-0.5
22	0	-16.5	-6.9	0.0	0.4
	1	-0.3	2.8	0.4	0.1
23	0	5.0	0.0	-0.1	0.0
	1				

D	O	Epoch Coeffs. (nT)		Secular Rates (nT/yr)	
n	m	$g_{n\ 0}^m$	$h_{n\ 0}^m$	$\dot{g}_{n\ 0}^m$	$\dot{h}_{n\ 0}^m$
9	1	8.2	-23.3	-0.2	-0.3
	2	2.9	11.1	0.0	0.2
10	3	-1.4	9.8	0.4	-0.4
	4	-1.1	-5.1	-0.3	0.4
11	5	-13.3	-6.2	0.0	0.1
	6	1.1	7.8	0.3	0.0
12	7	8.9	0.4	0.0	-0.2
	8	-9.3	-1.5	0.0	0.5
13	9	-11.9	9.7	-0.4	0.2
	10	-1.9	0.0	0.0	0.0
14	1	-6.2	3.4	0.0	0.0
	2	-0.1	-0.2	0.0	0.1
15	3	1.7	3.5	0.2	-0.3
	4	-0.9	4.8	-0.1	0.1
16	5	0.6	-8.6	-0.2	-0.2
	6	-0.9	-0.1	0.0	0.1
17	7	1.9	-4.2	-0.1	0.0
	8	1.4	-3.4	-0.2	-0.1
18	9	-2.4	-0.1	-0.1	0.2
	10	-3.9	-8.8	0.0	0.0
19	0	3.0	0.0	0.0	0.0
	1	-1.4	0.0	-0.1	0.0
20	2	-2.5	2.6	0.0	0.1
	3	2.4	-0.5	0.0	0.0
21	4	-0.9	-0.4	0.0	0.2
	5	0.3	0.6	-0.1	0.0
22	6	-0.7	-0.2	0.0	0.0
	7	-0.1	-1.7	0.0	0.1
23	8	1.4	-1.6	-0.1	0.0
	9	-0.6	-3.0	-0.1	-0.1
24	10	0.2	-2.0	-0.1	0.0
	11	3.1	-2.6	-0.1	0.0
25	0	-2.0	0.0	0.0	0.0
	1	-0.1	-1.2	0.0	0.0
26	2	0.5	0.5	0.0	0.0
	3	1.3	1.3	0.0	-0.1
27	4	-1.2	-1.8	0.0	0.1
	5	0.7	0.1	0.0	0.0
28	6	0.3	0.7	0.0	0.0
	7	0.5	-0.1	0.0	0.0
29	8	-0.2	0.6	0.0	0.1
	9	-0.5	0.2	0.0	0.0
30	10	0.1	-0.9	0.0	0.0
	11	-1.1	0.0	0.0	0.0
31	12	-0.3	0.5	-0.1	-0.1

Reference to the official WMM2020 report is recommended for more information.

3.4 The Geoid

The *geoid* is a mathematical surface of Earth representing the equipotential of gravitational (3.24) and centrifugal (3.5) contributions *at mean sea level* (MSL) everywhere. The geoid is a special case of the more general *geop*, a surface of gravitational and centrifugal equipotential to which gravitational *plumb lines* are normal satisfying

$$V + \Phi = \text{const.} \quad (3.64)$$

Thus, the geoid represents a surface of total gravity (gravitational and centrifugal) equipotential determined by this constant such that its surface agrees with MSL.

The distance of an object from the geoid is called the *orthometric height*, often denoted H , which is the distance from the geoid's surface to the object *along the (curved) plumb line's path passing through that object*. Measuring this height proves to be difficult — so difficult in fact that there is still no vertical reference system using orthometric heights in the world today. The distance separating the geoid from the ellipsoid is called the *geoid undulation*, often denoted N , which is the length of the straight line normal to the geoid intersecting the ellipsoid.

Though the earth's surface can vary from the ellipsoid on the order of kilometers, the geoid varies from the ellipsoid physically by fewer than about 100 [m] [Geodesy]. As such, the majority of the geoid lies under the earth's surface along with the ellipsoid. In the presence of locally excessive mass, the geoid rises above the ellipsoid, and similarly falls below in a local region of deficient mass.

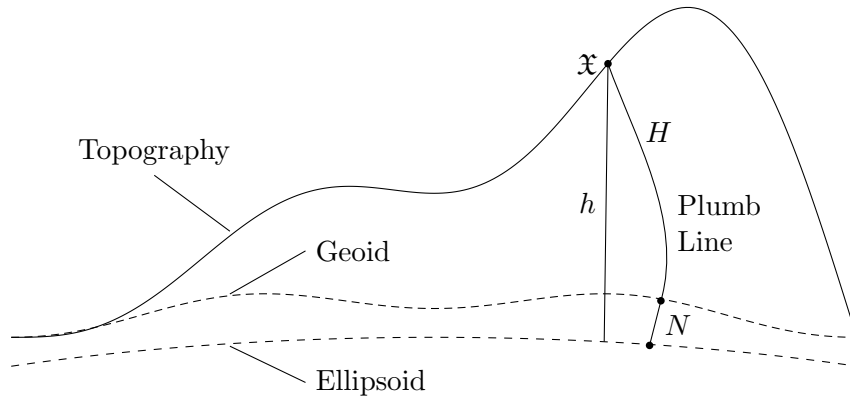


Fig. 3.2: The exaggerated difference between the ellipsoid and the geoid relative to Earth's surface. The gravitational plumb line, with greatly exaggerated curvature, emanates from the geoid's surface normally.

In reality, the geoid varies smoothly such that a good approximation for points on the surface of Earth follows

$$H = h - N. \quad (3.65)$$

That is, the plumb lines are essentially uncurved, an assumption which ignores the so-called “deflection of the vertical,” and the geoid undulation everywhere is approximately normal to both geoid and ellipsoid. As such, the elevation of a location on the surface of Earth at a longitude λ and geodetic latitude ϕ above MSL (H) is determined by that position's height above the ellipsoid h (determined by GPS) and the geoid undulation from the ellipsoid N at this same longitude and latitude pair.

3.4.1 Determining Height Above Mean Sea Level

These quantities H and N are naturally obtained by careful measurement of Earth's gravitational field as most recently monitored by the GRACE mission. The (scalar) *gravity anomaly* of

the *geoid* (MSL) is defined

$$\Delta g_0 = g_0 - \gamma_0, \quad (3.66)$$

where g_0 is the measured gravitational intensity at MSL and γ_0 is the predicted normal gravity on the ellipsoid's surface at the same ellipsoidal coordinates where the measurement was taken. A good approximation, however, is to let γ_0 take the scalar form of Newtonian gravity (3.9) such that the *disturbing potential* T , defined by the difference between actual (measured) and normal gravitational potential (thus satisfying Laplace's equation), referencing a sphere of radius a , the semi-major axis of the ellipsoid, takes on the form

$$T(r, \theta, \lambda) = \sum_{n=2}^{\infty} \left(\frac{a}{r}\right)^{n+1} T_n(\theta, \lambda). \quad (3.67)$$

The *fundamental equation of physical geodesy* as approximated on this spherical reference surface gives a simple way of relating the disturbing potential T to the gravity anomaly Δg .

$$\frac{\partial T}{\partial r} + \frac{2}{r}T = -\Delta g \quad (3.68)$$

Thus, the gravity anomaly is writable in terms of the disturbing potential

$$\Delta g(r, \theta, \lambda) = \frac{1}{r} \sum_{n=2}^{\infty} (n-1) \left(\frac{a}{r}\right)^{n+1} T_n(\theta, \lambda). \quad (3.69)$$

On the surface of the reference sphere ($r = a$), the disturbing potential T and gravity anomaly Δg then represent their values *on the geoid* (at MSL). By inserting the gravitational potentials (3.24) and (3.12) into the definition of T as evaluated on the geoid, then the gravity anomaly at MSL is expanded into spherical harmonics and becomes

$$\Delta g_0 = \frac{GM}{a^2} \sum_{n=2}^{\infty} \sum_{m=0}^n (n-1) (\delta C_{nm} \cos m\lambda + \delta S_{nm} \sin m\lambda) P_{nm}(\cos \theta). \quad (3.70)$$

Because of the linear nature of Laplace's equation that the disturbing potential satisfies, the coefficients, to good approximation, may be quantified by

$$\delta C_{n0} = C_{n0}^{meas.} - C_{n0}^{ellip.} \quad (n = 1, 2 \text{ and } C_{n0} = -J_n) \quad (3.71)$$

$$\delta C_{nm} = C_{nm}^{meas.}, \quad (n = 3, 4, \dots \text{ and } 0 < m \leq n) \quad (3.72)$$

Explicitly, the coefficients are given via the orthogonality relation

$$\begin{pmatrix} \delta C_{nm} \\ \delta S_{nm} \end{pmatrix} = \frac{2 - \delta_{0m}}{4\pi\gamma} \frac{2n+1}{n-1} \frac{(n-m)!}{(n+m)!} \iint_{\sigma} \Delta g_0 P_{nm}(\cos \theta) \begin{pmatrix} \cos m\lambda \\ \sin m\lambda \end{pmatrix} d\sigma, \quad (3.73)$$

where δ_{ij} is the Kronecker delta, γ is Newtonian gravity on the reference sphere's surface, σ is the unit sphere, and $d\sigma = \sin \theta d\theta d\lambda$ is the element of solid angle. Finally, Bruns' theorem relates the geoid undulation N to the disturbing potential T on the surface of the reference sphere by

$$N(\theta, \lambda) = \frac{T(\theta, \lambda)}{\gamma} = a \sum_{n=2}^{\infty} \sum_{m=0}^n (\delta C_{nm} \cos m\lambda + \delta S_{nm} \sin m\lambda) P_{nm}(\cos \theta). \quad (3.74)$$

Careful measurements by GPS satellites provide the elevations h above the ellipsoid. Thus, elevations H above MSL are found according to (3.65) with the geoid undulations known. Via spherical harmonics,

$$H(\theta, \lambda) = \sum_{n=0}^{\infty} \sum_{m=0}^n (\overline{HC}_{nm} \cos m\lambda + \overline{HS}_{nm} \sin m\lambda) \overline{P}_{nm}(\cos \theta), \quad (3.75)$$

where the fully-normalized coefficients $\{\overline{HC}_{nm}, \overline{HS}_{nm}\}$ are obtained from orbiter missions referencing a particular ellipsoid and geoid. Very typically, the geoid undulation N is taken with respect to the *tide-free* geoid model. The zero-tide geoid undulation relates to the tide-free geoid by the rather simple relation

$$N_{zero} = N_{free} + k(9.9 - 29.6 \cos^2 \phi) \text{ [m]}, \quad (3.76)$$

where k is the Love number internationally agreed to be defined by 0.3 [NGA_WGS84.pdf].

3.4.2 Accuracy and Resolution

The undulations of the EGM96 geoid from WGS84 are known to be within ± 1 [m] accuracy of true values, and the recent results of EGM2008 only improve on this error [Matlab “doc geoidheight”]. The EGM2008 geoid is provided coefficients up to degree and order 2190 obtained from the GRACE mission.

These values may be either calculated pointwise or tabulated into a matrix (because of the radial coordinate independence) to be referenced and interpolated upon later. One method, however, may be faster than the other. Calculating values pointwisely gives the option to decide how many terms are necessary, which could see a massive advantage in computational speed over interpolating a pre-calculated matrix (of maximum degree presumably) for a minimal loss in accuracy.

The EGM2008 geoid undulations are provided in pre-calculated matrices resolved into $2.5'$ and $1'$ (where $'$ is the arcminute, $1/60$ of a degree) equally spaced grids. Thus, a $2.5' \times 2.5'$ resolved matrix has 8640 elements corresponding to longitude and 4321 elements corresponding to latitude and contains 37333440 elements. With each floating number holding 8 bytes, the resulting matrix requires 300 MB (298667520 bytes) of memory at runtime. Likewise, a $1' \times 1'$ resolved grid contains 233301600 elements occupying 1.9 GB (1866412800 bytes) of memory at runtime. Interpolating such a grid is certainly guaranteed to be slower than pointwise computation of the height above MSL regardless of the interpolating method, but the $2.5'$ grid may contend depending on the method and implementation at the cost of accuracy.

3.4.3 Results

The surface formed by computing the height above MSL (H) everywhere over Earth’s surface approximates the *telluroid* (tellurian – earth-like), which is the best mathematical description of Earth’s surface. The achievable resolution of the telluroid due to recent advancements in the study of geodesy is on the order of centimeters to millimeters accuracy *everywhere*. An approximation of the telluroid gives approximate elevations above MSL everywhere on Earth, defining local conditions in terms of a standard atmosphere, and permits a more precise method of determining the touchdown location of a suborbital flight vehicle than by doing so using the excessively smooth ellipsoid in the case that its onboard GPS (if any) fails.

The EGM2008 geoid undulations from WGS84 shown in Fig. 3.3 were obtained by reading the encoded big endian internal binary representation file containing the $2.5'$ resolution grid pre-computed by the NGA Office of Geomatics. Fortran programs are provided to interpolate the grid properly, but these programs are not used as a spline provides, albeit it at a slight cost of speed, essentially the same functionality.

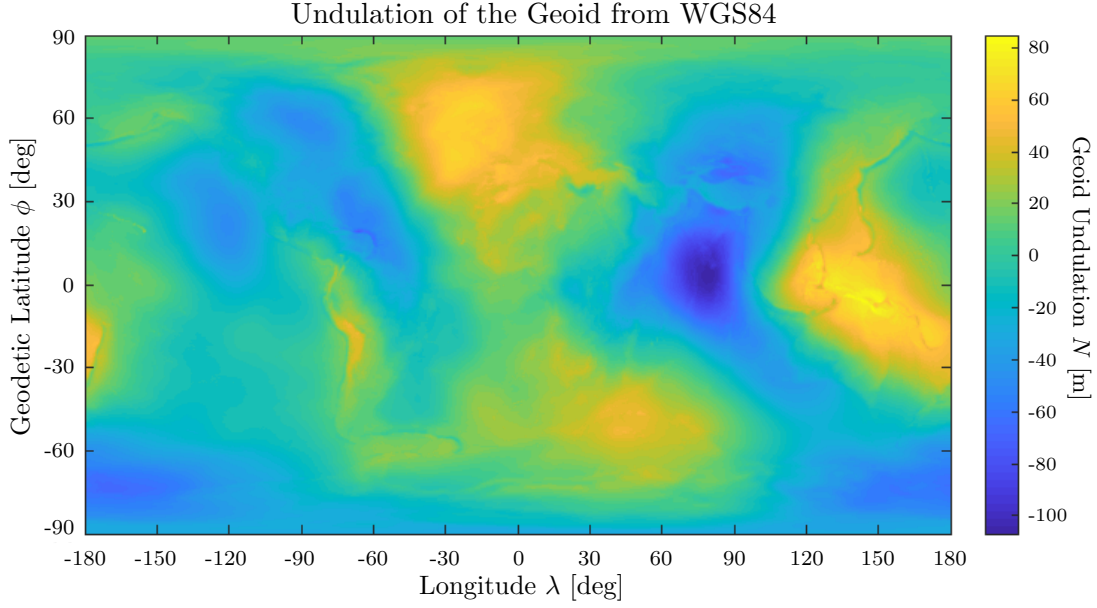


Fig. 3.3: The tide-free EGM2008 geoid undulations from WGS84

The displacements of the local topography from the geoid shown in Fig. 3.4 were computed according to (3.75) using all harmonic coefficients provided by EGM2008 to obtain a resolution of $2.5' \times 2.5'$.

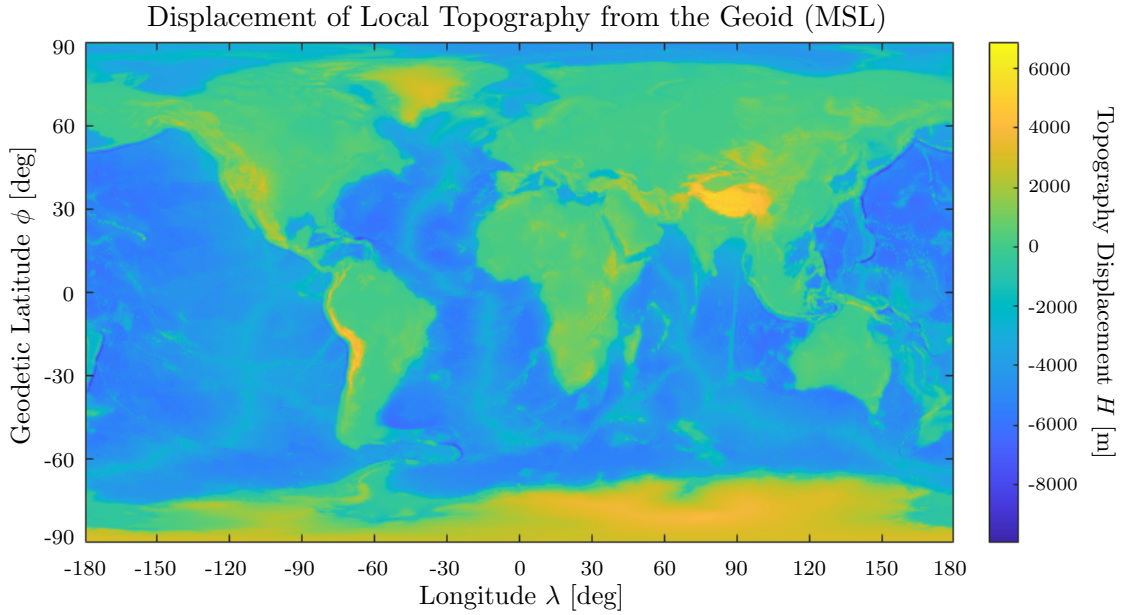


Fig. 3.4: The variation of local surface topography from the tide-free geoid

In accordance with (3.65), the combination (addition) of Figs. 3.3 and 3.4 yields the displacement of Earth's surface (without an ocean) from the WGS84 ellipsoid directly.

The majority of continental area is shown to be near zero (green/blue) since major mountain ranges and the ocean floor largely dominate the scale's extremities. As the lowest point on Earth is in the Dead Sea at an elevation of approximately 415 [m] beneath MSL, the ocean's surface (tide-free), which is the geoid over open water, can be approximated by setting values on the ocean floor (less than -415 [m]) to zero. Elevations of Earth's liquid and solid surfaces (ocean and continental) above MSL are therefore in accordance with Fig. 3.5, which approximates the

telluroid.

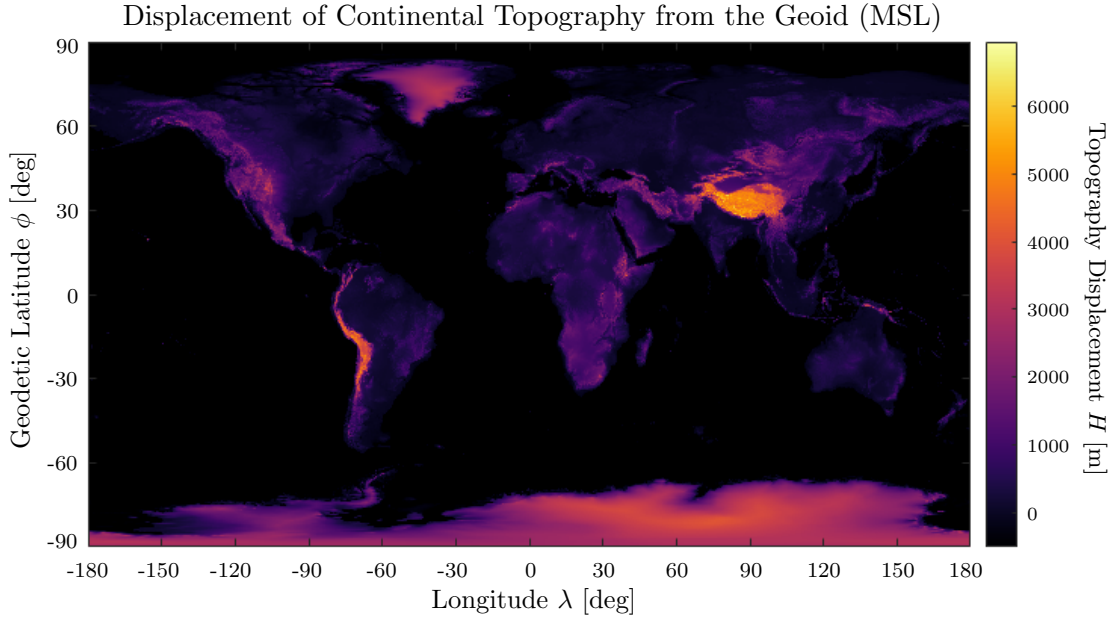


Fig. 3.5: The variation of local surface topography on the continents from the tide-free EGM2008 geoid. There is no displacement from the geoid on the open ocean by definition.

Combining (adding) Fig. 3.3 to Fig. 3.5 furnishes the elevation of Earth's liquid and solid surfaces with respect to the WGS84 ellipsoid. Subtly, the (tide-free) MSL changes globally, which is due simply to the geoid describing mass fluctuations throughout the earth.

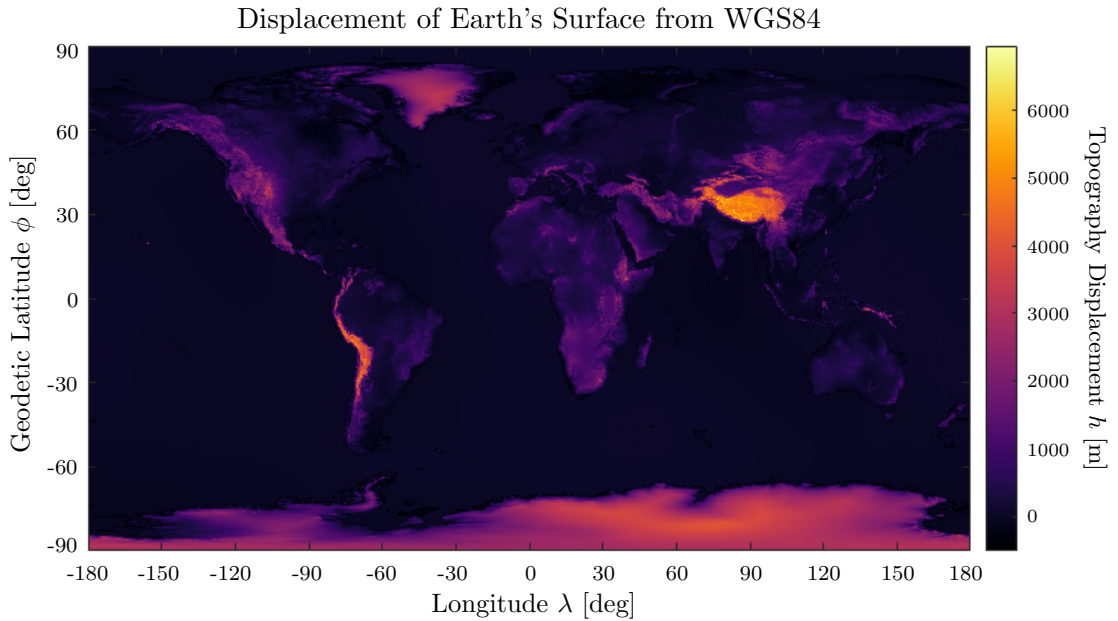


Fig. 3.6: The variation of local surface topography (ocean and land) over Earth's surface from WGS84

The variation of MSL over the oceans with respect to WGS84 is within Fig. 3.3 as this variation is what defines the geoid. For clarity, however, Fig. 3.7 overlays this variation with the continents to show only how the tide-free ocean's MSL varies over the surface of Earth.

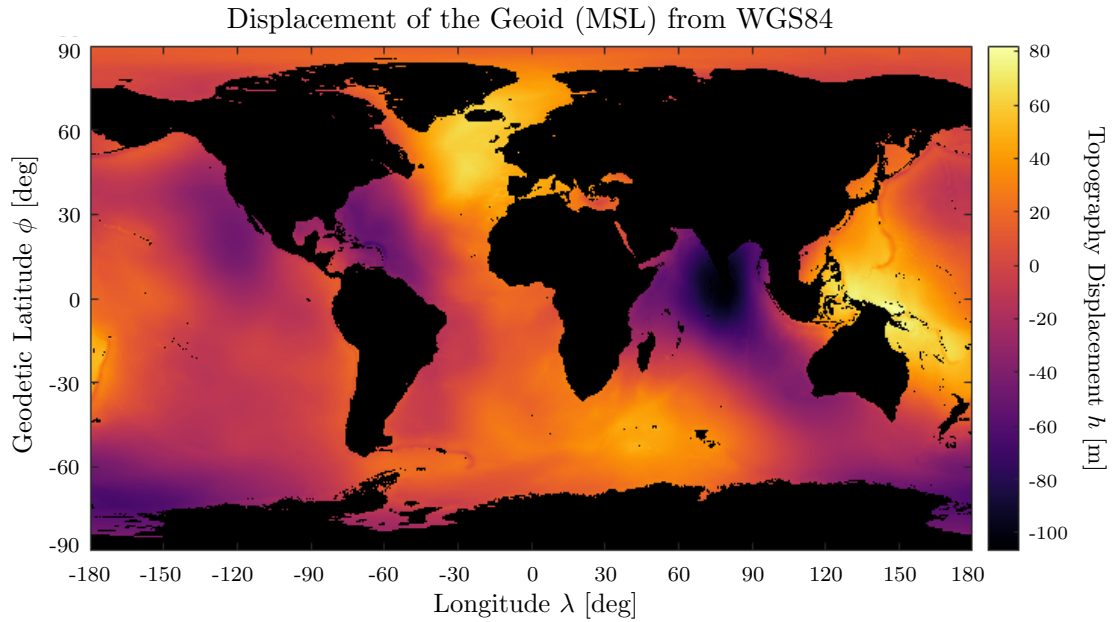


Fig. 3.7: The variation of MSL (tide-free EGM2008 geoid) over the oceans with respect to WGS84. This figure is exactly the same as Fig. 3.3 but with a different coloring and overlaid continents. Elevations exceeding the Dead Sea are ignored – some continental areas may appear fat as such.

The topography of the continental United States with respect to the EGM2008 tide-free geoid (MSL) is shown in Fig. 3.8.

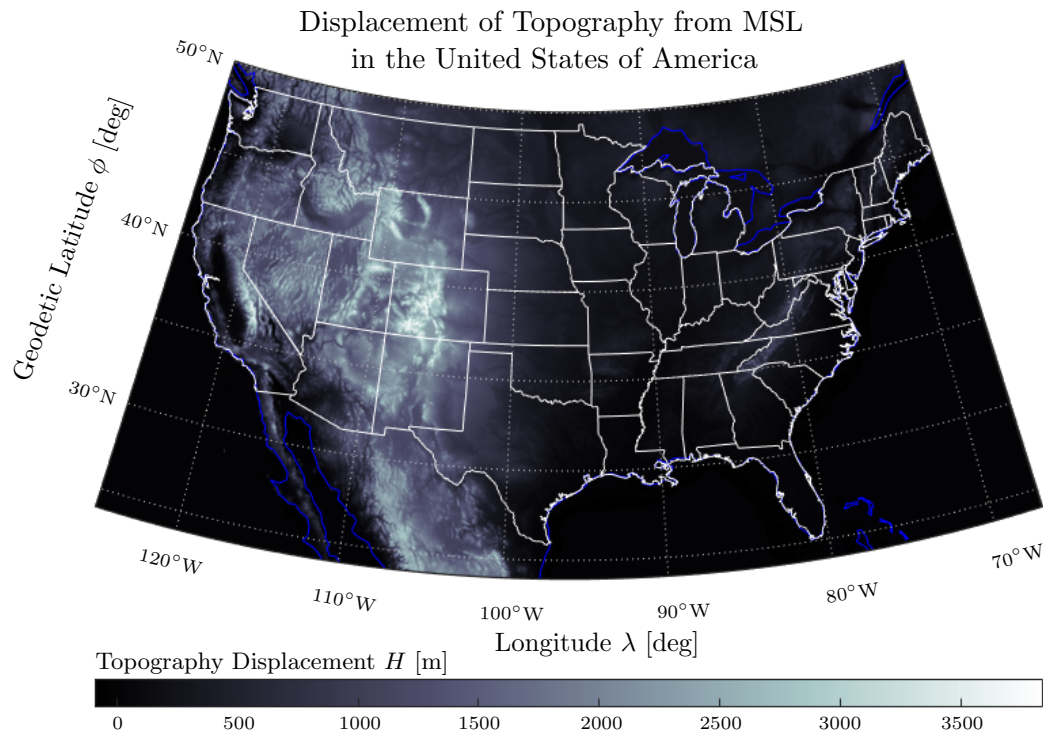


Fig. 3.8: The topography of the continental United States with respect to MSL overlaid with the 48 contiguous state borders and part of the North American coastline.

Since the geoid varies by no more than 105 meters from the WGS84 ellipsoid, a rough ap-

proximation of (3.65) yields, especially for very high altitudes, that $H \sim h$. Thus, a height h above the ellipsoid or a height H above MSL may be used for determining atmospheric properties without generating large errors, though this choice should remain consistent throughout as to, in general, avoid discontinuities in atmospheric conditions. That is, an elevation H above MSL may be used at low altitudes and switched to an elevation h above the ellipsoid if passing over a region where the geoid undulation is truly zero (or close enough). Doing so relaxes intensive computations while maintaining a continuously modelled atmosphere.

It is noted that the highest tabulated elevation above MSL (in accordance with Fig. 3.5) is 6959.9 meters whereas the summit of Mount Everest is 8848 meters (29029 feet). Likewise, the highest peak in the continental United States (in accordance with Fig. 3.8) is in California atop Mount Whitney at 4418 meters (14494 feet) above MSL. According to the figure, the highest peak in the continental United States is tabulated as 3837.3 meters (12590 feet). These errors are likely due because of grid spacing. The selected geodetic coordinates making up the grids either missed the peaks of Mount Everest and Mount Whitney (among others surely) because of its “lacking” $2.5' \times 2.5'$ resolution, or simply because of the limiting resolution that the coefficients provided from GRACE can acquire. The case seems to be the latter; inserting the coordinates of Mount Everest’s peak and checking in its surrounding area supplies no such extreme, and likewise for Mount Whitney.

3.5 Atmospheric Model

Atmospheric properties are perhaps one of the most difficult aspects of the earth to model and predict accurately at any given time because of its naturally almost chaotic and quickly evolving behavior. For this reason, a *standard atmosphere* may be established, representing not real, instantaneous atmospheric properties, but rather standardized conditions that are generally accepted by the scientific community.

3.5.1 Gas Properties

In most cases concerning the atmosphere, air is treated as an *ideal gas* (as opposed to a real gas) so that the *ideal gas law*, serving as the *equation of state* (which relates air pressure p , air temperature T , and air density ρ), is obeyed.

$$p = \rho RT \quad (3.77)$$

Here, R is the specific gas constant defined by the molecular weight \mathcal{M} of the gas with the universal gas constant \mathcal{R} , which is itself the product of fundamental quantities (Avogadro and Boltzmann constants).

$$\mathcal{R} = N_A k_B \approx 8.314\,462\,618\,153\,24 \text{ [J/K mol]} \quad (3.78)$$

The continuously-ongoing process of atmospheric motion and mixing with itself throughout the world can be leveraged to say that the composition of air close to the surface, which is a mixture of gases, is approximately uniform everywhere. However, this composition changes in time which affects, albeit slightly, the gas constant of air defined

$$R = \mathcal{R}/\mathcal{M}. \quad (3.79)$$

Thus, most recent atmospheric data is apt to give a most current value of air’s gas constant. Obtaining a single value for the gas constant of “the air” represents what is like a global average air, though it may be interpreted also in a pointwise sense. In any case, the type of considered air is *dry air* and does not take into account local variations of (particularly) humidity at any location. Regardless of details involving the “exact” composition of air, its gas constant remains

approximately constant up to the mesopause (an elevation of about 90 [km] above sea level) and is

$$R \sim 287 \text{ [J/kg K]}. \quad (3.80)$$

Specifying the composition of air to gain an accurate molecular weight is really fine-tuning the gas constant's decimals.

By physical intuition, every gas has a measure of how resistant it is to changing its temperature when supplied a heat source to do so. Two such quantities realize this resistance in the form of how much energy must be supplied to produce a unit change in the gas's temperature — the specific heat capacity at constant pressure c_p and the specific heat capacity at constant volume c_v (note the subscript v despite the typical denotation of volume as V). The term “specific” here indicates that quantities are measured per kilogram. Thus, unit masses of gas are implied and the density is consequently the inverse of the containing volume, termed the specific volume v . The quantity c_v is therefore actually a statement about the gas's specific heat capacity at constant density. Obviously, there is no need for a quantity describing specific heat capacity at constant temperature. It is a fact that these two quantities, c_p and c_v , for any gas are mainly dependent upon temperature (weakly dependent upon pressure) and both increase along with temperature.

A ratio of the two (specific) heat capacity ratios, denoted γ (not to be confused with the normal gravity γ) is defined

$$\gamma = c_p/c_v. \quad (3.81)$$

Thus, γ , called the *heat capacity ratio*, is dependent upon temperature. Mayer's relation for an ideal gas provides further insight into how these quantities relate to each other.

$$R = c_p - c_v. \quad (3.82)$$

The gas constant R is obviously a positive quantity as indicated by (3.79), but perhaps less apparent, it is *independent* of temperature. The respective conclusions are therefore $\gamma > 1$ and (3.82) is true everywhere. Since both specific heat capacities c_p and c_v increase with temperature but maintain a constant difference, the heat capacity ratio γ decreases towards unity with increasing temperature as indicated by Fig. 3.9.

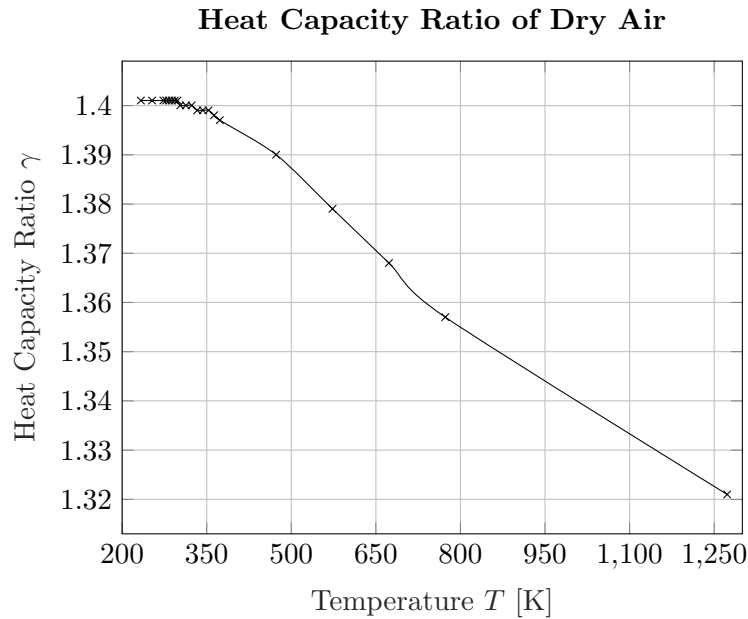


Fig. 3.9: The variation of air's heat capacity ratio on temperature (neglecting variations with pressure) [engineering toolbox].

Gases obeying the ideal gas law are said to be *thermally perfect*, and gases obeying a constant heat capacity ratio γ , at least in some specified range of temperature, are said to be *calorically perfect* (in that temperature range). Gases that are both thermally perfect and calorically perfect are *ideal*, which of course is an idealization of reality. Real gases are not perfect and consequently obey a different equation of state than the ideal gas law. Many such models exist, like van der Waals or Redlich-Kwong, which necessarily change Mayer's result (3.82) to its general form

$$-T \frac{(\partial p / \partial T)_v^2}{(\partial p / \partial v)_T} = c_p - c_v. \quad (3.83)$$

As an example, the original Redlich-Kwong equation of state yields

$$p = \frac{RT}{v-b} - \frac{a}{T^{1/2}v(v+b)} \quad (\rho = 1/v) \quad (3.84)$$

$$a = \frac{R^2 T_c^{5/2} / p_c}{9(2^{1/3} - 1)} \quad (3.85)$$

$$b = \frac{RT_c / p_c}{3/(2^{1/3} - 1)}, \quad (3.86)$$

where p_c and T_c are the pressure and temperature critical points respectively, for which the difference in specific heat capacities becomes

$$c_p - c_v = \left(\frac{R}{v-b} + \frac{1}{2} \frac{a}{T^{3/2}v(v+b)} \right)^2 / \left(\frac{R}{(v-b)^2} - \frac{a(2v+b)}{T^{3/2}v(v+b)^2} \right). \quad (3.87)$$

If a , representing the correction for the molecular potential attraction, and b , representing a correction to specific volume, are both set to zero, then the ideal gas law is recovered.

Treating air as calorically imperfect requires a quantum mechanical inspection for a full description of its behavior. The atoms within experience changes in the amount of degrees of freedom which are unlocked by increasing temperature. Merely having three spatial dimensions in which space exists and the impossibility of reaching absolute zero is enough to unlock the first mode for all molecules consisting of three degrees of freedom — translation. The other two modes (rotation and vibration) at such low temperatures are unactivated, or *frozen*.

The rotational mode is fully activated shortly after absolute zero (50 [K] or so), which unlocks an amount of degrees of freedom dependent on the kind of molecule. Monotonic molecules have no other modes than translation. Molecules arranged in a line (linear molecules, all bonds are 180° apart) unlock two additional degrees of freedom whereas nonlinear molecules (all other molecules that are not monotonic and not linear) unlock three.

The vibrational mode requires high temperatures to fully activate. A molecule consisting of N atoms unlocks an additional $3N - 5$ degrees of freedom if it is linear (to the 5 it already had) and $3N - 6$ otherwise (to the 6 it already had) [Heat Transfer Physics pg. 397]. Of course, monotonic atoms have no vibrational mode, so they still have only 3 degrees of freedom.

The heat capacity ratio γ can be written in terms of these amounts of degrees of freedom at low temperatures (such that the vibrational mode has not yet become too active) to good approximation as

$$\gamma \sim 1 + \frac{2}{\text{degrees of freedom}}. \quad (3.88)$$

As expected then by Fig. 3.9, air (a *mixture* largely composed of diatomic molecules N_2 and O_2) contains 5 degrees of freedom, so at these temperatures,

$$\gamma \sim 1.4. \quad (3.89)$$

As temperatures increase, however, the specific heats should properly be written as functions of temperature. If the vibrations are considered as a simple quantum harmonic oscillator, then the according relations follow the kinetic theory of gases [NASA GRC Calorically Imperfect Gas]

$$c_p = c_{p\text{cal.perfect}} \left(1 + (1 - 1/\gamma_{\text{perfect}}) \left(\frac{\Theta}{T} \right)^2 \frac{e^{\Theta/T}}{(e^{\Theta/T} - 1)^2} \right) \quad (3.90)$$

$$c_v = c_{v\text{cal.perfect}} \left(1 - (1 - \gamma_{\text{perfect}}) \left(\frac{\Theta}{T} \right)^2 \frac{e^{\Theta/T}}{(e^{\Theta/T} - 1)^2} \right) \quad (3.91)$$

$$\gamma = 1 + \frac{\gamma_{\text{perfect}} - 1}{1 + (\gamma_{\text{perfect}} - 1) \left(\frac{\Theta}{T} \right)^2 \frac{e^{\Theta/T}}{(e^{\Theta/T} - 1)^2}}. \quad (3.92)$$

The quantities $c_{p\text{cal.perfect}}$ and $c_{v\text{cal.perfect}}$ are the specific heat capacities of the gas if it were calorically perfect (i.e., if the heat capacity ratio were constant in temperature) corresponding with the constant value γ_{perfect} . For air, the value (3.89) may be used comfortably as the perfect ratio. The quantity Θ is the scale temperature set to 5500 [°R], or

$$\Theta = \frac{5}{9} \cdot 5500 \sim 3056 \text{ [K]}. \quad (3.93)$$

These relations provide a functional description of gas characteristics in temperature and nicely approach back to the perfect values for $T \ll \Theta$. Though, they should be used with the possibility of computational overflows in mind due to the nature of exponentiating several thousand (in the event that temperature drops too low).

Whether thermally and/or calorically imperfect or perfect, the speed of a sound wave, or *speed of sound* denoted a (not to be confused with the semi-major axis of Earth's ellipsoid a), travels through the gas according to

$$a^2 = \gamma \frac{\partial p}{\partial \rho}. \quad (3.94)$$

A thermally perfect gas (obeying the ideal gas law) therefore transmits sound at speed

$$a = \sqrt{\gamma R T} \quad (3.95)$$

whereas a thermally imperfect gas otherwise transmits sound depending on its equation of state (van der Waals, etc.). The speed of sound at sea level, where $\gamma = 1.4$, $R = 287 \text{ [J/kg K]}$, and $T = 59 \text{ [°F]}$, is therefore

$$a_{059} \sim 340 \text{ [m/s]}. \quad (3.96)$$

The kinetic theory of gases gives implications as to how (dynamic) viscosity, denoted μ , should roughly behave. Perhaps the most fundamental model which poses the hardest assumptions on the gas is that of atoms made of solid, perfectly elastic spheres of mass m and diameter σ . The corresponding viscosity is

$$\mu = \frac{1.016 \cdot 5}{16 \sigma^2} \sqrt{m k_B T / \pi}, \quad (m = \mathcal{M} / N_A). \quad (3.97)$$

Thus, the viscosity of a gas under this model *increases only with temperature as $\mu \propto T^{1/2}$* (liquids typically decrease with increasing temperature). Though simple, this viscosity power law model grows too slowly to match measured data and hence is not a very good description of viscosity.

Sutherland published his model in 1893, coined *Sutherland's Law*, which is consistent with the kinetic theory of gases. It is realized by utilizing weakly attracting intermolecular forces sourced from a potential model for solid spheres of mass m and diameter σ . The viscosity is written

$$\mu = \frac{5}{16 \sigma^2} \frac{\sqrt{m k_B T / \pi}}{1 + S/T} = \mu_0 \left(\frac{T}{T_0} \right)^{3/2} \frac{T_0 + S}{T + S}. \quad (3.98)$$

Thus, a referenced (measured) viscosity μ_0 at temperature T_0 defines the viscosity as a function of temperature and parameter S . For air,

$$\mu_0 = 0.00001716 \text{ [kg/m s]} \quad T_0 = 273.15 \text{ [K]} \quad S = 110.4 \text{ [K]}. \quad (3.99)$$

The region of validity for Sutherland's law for N_2 and O_2 falls between 290 and 1100 [K] whereas for Ar, the range is merely 290 to 375 [K]. Dry air is therefore properly limited to the range of Ar, though it may be extended to 1100 [K] since Ar makes up such a trace amount of air in comparison. Sutherland's law, however, is disputed as to whether it actually gives good results or not.

The Lennard-Jones potential provides a viscosity through the kinetic theory of gases for solid spheres written

$$\mu = \frac{5}{16 \sigma^2} \frac{\sqrt{m k_B T / \pi}}{\Omega(T)} \stackrel{?}{=} \mu_0 \left(\frac{T}{T_0} \right)^{1/2} \frac{\Omega(T_0)}{\Omega(T)}, \quad (3.100)$$

where Ω is called the *collision integral*, which is determined by the expression of the potential. Much like other models (except Sutherland, oddly), the collision integral for Lennard-Jones is not expressible in terms of elementary functions, but is well approximated by

$$\Omega(T) \sim 1.16145 T^{*-0.14874} + 0.52487 e^{-0.77320 T^*} + 2.16178 e^{-2.43787 T^*}, \quad \left(T^* = \frac{k_B T}{\epsilon} \right), \quad (3.101)$$

where σ and ϵ are both constants determined by experiment. These values for various gases commonly found in the atmosphere are tabulated below in Tab. 3.9.

Tab. 3.9: Lennard-Jones gas parameters

Gas	Molecular Diameter σ (Å)	Temperature Scale k_B/ϵ (1/K)
Dry Air	3.617	1/97.0
Helium	2.576	1/10.2
Hydrogen	2.915	1/38.0
Argon	3.432	1/122.4
Nitrogen	3.667	1/99.8
Oxygen	3.433	1/113.0
Carbon dioxide	3.996	1/190.0
Methane	3.780	1/154.0

Despite Lennard-Jones being derived with solid spheres playing the role of atoms, this representation of the viscosity realizes an average deviation from the measured gas viscosity within 0.064% for the temperature ranges $0.3 < T^* < 100$. Thus, the model does well in describing nonspherical atoms too. [Temperature_dependence_of_viscosity wikipedia]

Upon choosing a model of the dynamic viscosity μ , the *kinematic viscosity* ν is then defined

$$\nu = \mu / \rho. \quad (3.102)$$

Though still a valid gas property, the kinetic viscosity plays no role in building a standard atmosphere. Its main feature lies in aerodynamics, but is included here to have both kinds of viscosities together for reference.

3.5.2 Layers of the Atmosphere

The real atmosphere is divided into a few basic layers separated by the behavior of air temperature — the troposphere, stratosphere, mesosphere, thermosphere, and exosphere. The ionosphere and magnetosphere are also regions of the atmosphere, but they have more specialized roles than separating temperature behaviors.

The Troposphere

The troposphere is the most dense atmospheric region extending from Earth's surface to about 8–10 [km] at the poles and 14–18 [km] along the equator. It is characterized by a *reduction* of air temperature as above-ground altitude increases. Most of all weather occurs in the troposphere and also usually provides the flight ceiling for commercial planes. The upper-most region of the troposphere is called the tropopause and borders along the bottom of the stratosphere.

The Stratosphere

The stratosphere extends from the tropopause up to about 50 [km] and is characterized by an *increase* in air temperature with rising altitude. This increase of temperature is usually why commercial aircraft are often flown beneath 12 [km] (40,000 [ft]) altitude. The stratosphere is what contains the ozone layer which absorbs and scatters most ultraviolet radiation incoming from the sun. The stratopause is defined near the top of the stratosphere (50–55 [km]) which serves as the boundary between it and the mesosphere.

The Mesosphere

The mesosphere begins from the stratopause and extends to about 85 [km] characterized by, again, a *decrease* in air temperature with rising altitude. The upper part of the mesosphere (the mesopause) experiences the coldest temperatures in the atmosphere reaching so low as 170–180 [K] (–135 to –155 [°F]) at around 85 [km] and 100 [km] altitude. Meteors burn up throughout the mesosphere.

The Thermosphere

The thermosphere is the continuation from the mesopause extending up to about 600 [km] in which air temperature *increases* again with increasing altitude. Though the thermosphere's end at 600 [km] is often taken as standard, the gas within expands a considerable amount due to solar heating so the thermosphere's boundary above any fixed point on Earth's surface is highly time-dependent ranging between 500–1000 [km]. And despite “air” (really, just gas now) temperatures soaring up as much as 2000–2800 [K] (3,140–4,580 [°F]), the density is so low throughout that an object in this region receives hardly any heat transfer (from the gas) and thus does *not* experience these kinds of temperatures. Aurorae and LEO satellites exist in the thermosphere. The upper-most region is called the thermopause which serves as the boundary between the thermosphere and the exosphere.

The Exosphere

The exosphere is the final atmospheric layer (though some consider the thermosphere the final layer) in which gas molecules are so diffuse that they hardly interact with each other, follow orbital trajectories, and sometimes escape from the earth system into the “true” vacuum of space. As such, there is no solid boundary separating Earth's atmosphere from where space begins.

Special Regions

The Kármán Line, defined at 100 [km] (in the *thermosphere*), is widely recognized as a boundary for the beginning of space where the study of astronautics begins to dominate over aeronautics. In comparison, some estimates for the “end” of the exosphere extend out to 10,000 [km] and even 100,000–180,000 [km] above Earth's surface. To place these heights into perspective, the moon's orbital radius about the earth is around 384,000 [km].

The ionosphere is a special region of the atmosphere reaching from about 48 [km] to about 965 [km] above Earth's surface [nasa.gov atmosphere-layers2]. *This region overlaps the stratopause,*

mesosphere, and *thermosphere*, and consequently varies in height along with the thermosphere. Further, the ionosphere is divided into three subregions — D, E, and F — based on the absorbed wavelength of incoming solar radiation. In all subregions however, atoms are ionized by solar radiation and then provide the interesting property of reflecting radio waves, thus providing the possibility of global radio communication [scied.ucar.edu, nasa.gov atmosphere-layers2].

The magnetosphere is another special region characterized by ions and free electrons being dominated by Earth's magnetic field, a process that is highly dependent upon Earth's internal activity and solar activity. Unlike the atmospheric regions, the actual profile of the magnetosphere does not resemble a sphere at all, but rather the flow field resulting from the placement of a sphere in supersonic flow. The two main channels of ion and electron travel are termed the Van Allen belts which are located about 3,000 [km] and 16,000 [km] above the earth's surface.

As mentioned earlier, the earth's atmosphere near the surface is regarded as evenly mixed and distributed due to never-ending turbulent mixing. The region in which this homogeneity is true is called the *homosphere*. Accordingly, the *heterosphere* is the region above the homosphere with strong and ordered molecular diffusion. The homosphere and heterosphere are separated by a boundary called the *turbopause*, which is located close to the mesopause about 110 [km] above Earth's surface.

3.5.3 Sourcing a Standard Atmosphere

The atmosphere at lower altitudes (not dominated by solar radiation, gravitational/magnetic attraction, etc.) has another governing relation in addition to the equation of state (3.77). Free air in equilibrium obeys the *hydrostatic equation*

$$\frac{\partial p}{\partial h} = -\rho g, \quad (3.103)$$

for which the pressure at a given altitude h above sea level (very closely related to, but not quite, the height above the ellipsoid h) immediately follows.

$$p = \int_h^\infty \rho g dh \quad (3.104)$$

That is, the air pressure (in equilibrium) at a particular point above the earth's surface at an altitude h is determined by the weight of air contained in a column of unit area directly above it [maths.ucd.ie ch03-slides-2.pdf]. Here, the direction of the gravity model g is assumed downwards (as is typical in atmospheric studies), so its value is *positive* to yield a positive pressure. This relation (3.104), albeit insightful, is not too useful when constructing an atmospheric model.

An insertion of the ideal gas law (3.77) into the hydrostatic equation (3.103), eliminating density, proves to be useful. Through study of the earth's lower atmospheric layers, the air temperature is typically modeled as having a piecewise-linear relationship with *geopotential height*, denoted h_{gp} , defined

$$h_{gp} = \frac{1}{g_0} \int_0^h g dh \quad (3.105)$$

such that in a particular atmospheric layer beneath the turbopause, the temperature at a height h takes the form

$$T = T_0 + a(h_{gp} - h_{gp0}), \quad (3.106)$$

where T_0 and h_{gp0} are the continuity-ensuring air temperature and corresponding geopotential height from the preceding layer at geometric height h_0 , unless in the troposphere in which case T_0 is the sea level temperature at zero height. The quantity a is called the *lapse rate* of temperature (dT/dh_{gp}) and may be positive, negative, or zero, but *is constant for a given atmospheric layer within the homosphere* (i.e. in the troposphere, stratosphere, and mesosphere). Thus, the pressure

at a height h is related to the pressure at the height h_0 , both heights in one atmospheric layer, via

$$p = p_0 \exp \left(- \int_{h_0}^h \frac{g}{RT} dh \right). \quad (3.107)$$

Inserting a gravity model (3.10), (3.23), or (3.34) into this relation then yields an expression for the air pressure (in the homosphere) at a height above sea level h with correct implementation of the piecewise-nature of temperature. Thus, the atmosphere model may easily also vary with latitude and/or longitude.

The real atmosphere, however, is more difficult to model than inserting a gravity model into a standard atmosphere. Intuitively, the real atmosphere also varies with time since local conditions are also dependent upon the earth's tilt relative to the ecliptic, spin, position around the sun, solar activity among other effects.

3.5.4 Jet Standard Atmosphere

The jet standard atmosphere is a very basic, yet insightful, atmosphere model that extends from mean sea level to the tropopause (about 11 [km]). Since the altitude is small from mean sea level (where h_0 is zero), a decent longitude- and latitude-independent gravity model is (3.10) using a corresponding average radius R_e (2.65) consistent with the WGS84 ellipsoid. Under this gravity model, the geopotential height (3.105) takes the form

$$h_{gp} = \frac{R_e h}{R_e + h}. \quad (3.108)$$

The temperature in the troposphere then follows (3.106) with h_0 (and h_{gp0}) taken to be zero at sea level itself,

$$T = T_0 + ah_{gp}, \quad (3.109)$$

for which the pressure according to (3.107) under this gravity model is

$$p = p_0 \left(1 + \frac{ah_{gp}}{T_0} \right)^{-g_0/aR}. \quad (3.110)$$

The density then follows by the equation of state (3.77). In the case that a is zero such that the troposphere is modeled as an isothermal layer, then the temperature (3.109) becomes a constant T_0 of course and the pressure would become

$$p \stackrel{a \rightarrow 0}{=} p_0 \exp \left(- \frac{g_0 h_{gp}}{RT_0} \right), \quad (3.111)$$

where the density would follow appropriately.

It is noted that for such low altitudes, the geopotential altitude h_{gp} may be thought of as essentially the same as geometric altitude h to within 0.2% difference at the tropopause. In fact, the gravity model might be simplified to the constant surface gravity g_0 throughout the troposphere in which case, the geopotential altitude actually becomes the geometric altitude. Doing so incurs a maximum difference of only 0.22% from the pressure model (3.110) and is similarly written

$$p = p_0 \left(1 + \frac{ah}{T_0} \right)^{-g_0/aR}. \quad (3.112)$$

In the case of an isothermal model of the troposphere (if the lapse rate a is zero), then

$$p \stackrel{a \rightarrow 0}{=} p_0 \exp \left(- \frac{g_0 h}{RT_0} \right). \quad (3.113)$$

In all cases, the pressure is seen to follow an exponential trend in addition to the temperature being linear in geopotential altitude. This trend will continue up through the atmosphere until the turbopause. The Jet Standard is quantified by Tab. 3.10.

Tab. 3.10: Jet Standard Atmosphere defining parameters.

Geop. Altitude h_{gp} (km)	Geom. Altitude h (km)	Base Pressure p_0 (Pa)	Base Temp. T_0 (K)	Lapse Rate a (K/km)
0.000 – 11.000	0.000 – 11.019	101325.000	288.15	–6.5

A Maclaurin expansion to fourth order of (3.110) using these parameters with WGS84 having an average radius of Earth (2.65), which agrees to within 0.70% difference at an altitude of 11 [km], gives a more clear sense of how the pressure behaves in the troposphere.

$$p \approx 101325 - 12.0154 h + 0.000575298 h^2 - 1.40015 \cdot 10^{-8} h^3 + 1.76465 \cdot 10^{-13} h^4 \text{ [Pa]} \quad (3.114)$$

Choosing to use a value of R_e dependent upon the latitude and a corresponding ellipsoidal gravity model, however, would make this atmosphere model one with a basic dependence also upon latitude.

3.5.5 1976 U.S. Standard Atmosphere

The 1976 U.S. Standard is like the Jet Standard in that the same gravity model (3.10) is used, but extended beyond the troposphere. To do so, the jet temperature profile (3.109) and exponential pressure (3.110) require modifying so that the atmosphere layer's initial height h_0 is nonzero. The temperature profile then follows (3.106) for which the pressure, according to (3.107) under the gravity model (3.10), becomes

$$p = p_0 \left(\frac{(h_0 + R_e)(T_0(h + R_e) - a(hh_{gp0} - hR_e + h_{gp0}R_e))}{(h + R_e)(T_0(h_0 + R_e) - a(h_0h_{gp0} - h_0R_e + h_{gp0}R_e))} \right)^{-g_0/aR} \quad (3.115)$$

where p_0 and T_0 are the pressure and temperature at height h_0 (determined by the bordering layer) and both h_0 and h are in the same atmospheric layer. The U.S. Standard is defined by repeating this procedure up until about 86 [km] near the mesopause. In the case of starting from mean sea level, (3.115) reduces to the Jet Standard pressure (3.110), which also happens to be valid for small (a few kilometers) altitudes beneath mean sea level. In the case of an isothermal layer, then

$$p \stackrel{a \rightarrow 0}{=} p_0 \exp \left(-\frac{g_0 R_e^2}{RT_0} \frac{(h - h_0)}{(R_e + h)(R_e + h_0)} \right). \quad (3.116)$$

The 1976 U.S. Standard Atmosphere below the mesopause is obtained by repeated cycling of (3.106), (3.115), and (3.77), between atmospheric layers. The according values to describe the atmosphere according to the 1976 U.S. Standard are found in Tab. 3.11.

Tab. 3.11: 1976 U.S. Standard Atmosphere defining parameters. From these values, the full atmosphere model may be determined up to a geometric altitude of 86 kilometers.

Geop. Altitude h_{gp} (km)	Geom. Altitude h (km)	Base Pressure p_0 (Pa)	Base Temp. T_0 (K)	Lapse Rate a (K/km)
0.000 – 11.000	0.000 – 11.019	101 325.000	288.15	−6.5
11.000 – 20.000	11.019 – 20.063	22 619.582	216.65	0.0
20.000 – 32.000	20.063 – 32.162	5470.481	216.65	1.0
32.000 – 47.000	32.162 – 47.350	867.316	228.65	2.8
47.000 – 51.000	47.350 – 51.412	110.990	270.65	0.0
51.000 – 71.000	51.412 – 71.802	66.986	270.65	−2.8
71.000 – 84.852	71.802 – 86.000	3.930	214.65	−2.0

This standard also includes an extended atmosphere model to 1000 [km], but the procedure of determining atmospheric properties is different because of the variable nature that the molecular composition of air undergoes at such high altitudes. The (kinetic) temperatures, i.e. the temperature of individual molecules (not necessarily “the temperature” of the surrounding region itself) is given piecewise for which the pressure and density, which both decrease exponentially with altitude, are fit to data with an exponential power law.

Between the mesopause (86 and 91 [km]), the kinetic temperature is found to be isothermal such that, in accordance with Tab 3.11.

$$T = 186.9 \text{ [K]}. \quad (3.117)$$

The following region until 110 [km] is modeled with an elliptical distribution of temperature

$$T = T_c + A \sqrt{1 + \frac{(h - h_0)^2}{H^2}}. \quad (3.118)$$

Here, the coefficients are approximately extend continuously from the mesopause conditions such that the temperature T_c is 263.1905 [K] and the “temperature” A is −76.3232 [K], the height H is −19.9429 [km], and the initial altitude h_0 is of course 91 [km].

The region extending from 110 [km] to 120 [km] geometric altitude (entering into the thermosphere) again follows a linear form, but this time with respect to the geometric altitude, as

$$T = T_0 + a(h - h_0), \quad (3.119)$$

where the initial temperature T_0 is 240 [K], the lapse rate a is 12 [K/m], and the initial altitude h_0 is 110 [km].

The final region extending all the way to 1000 [km] models the temperature as an exponential, reaching a maximum temperature at “infinity” to represent mean solar conditions. The temperature takes the form

$$T = T_\infty - (T_\infty - T_0) \exp(-\lambda \xi), \quad (3.120)$$

where

$$\lambda = \frac{L_0}{T_\infty - T_0} \quad \xi = \frac{R_e + h_0}{R_e + h}(h - h_0), \quad (3.121)$$

where L_0 is the initial temperature gradient at the geometric altitude h_0 . The Standard indicates that T_∞ is 1000 [K], T_0 is 360 [K], λ is 0.01875 [1/km], and the initial altitude h_0 is 120 [km]. [braeunig.us, all of this]

The pressure is then fit to data in the form of an exponential power law of the form

$$p = \exp (A_p h^4 + B_p h^3 + C_p h^2 + D_p h + E_p), \quad (3.122)$$

whose coefficients are given below in Tab. 3.12. Similarly, the air density follows the identical form

$$\rho = \exp (A_\rho h^4 + B_\rho h^3 + C_\rho h^2 + D_\rho h + E_\rho), \quad (3.123)$$

whose coefficients are given below in Tab. 3.13.

Tab. 3.12: Exponential power law pressure coefficients in accordance with the 1976 U.S. Standard Atmosphere.

Geom. Alt. h (km)	A_p	B_p	C_p	D_p	E_p
86 – 91	0.000 000	$2.159\,582 \cdot 10^{-6}$	$-4.836\,957 \cdot 10^{-4}$	$-1.425\,192 \cdot 10^{-1}$	$1.347\,530 \cdot 10^1$
91 – 100	0.000 000	$3.304\,895 \cdot 10^{-5}$	$-9.062\,730 \cdot 10^{-3}$	$6.516\,698 \cdot 10^{-1}$	$-1.103\,037 \cdot 10^1$
100 – 110	0.000 000	$6.693\,926 \cdot 10^{-5}$	$-1.945\,388 \cdot 10^{-2}$	1.719 080	$-4.775\,030 \cdot 10^1$
110 – 120	0.000 000	$-6.539\,316 \cdot 10^{-5}$	$2.485\,568 \cdot 10^{-2}$	$-3.223\,620$	$1.359\,355 \cdot 10^2$
120 – 150	$2.283\,506 \cdot 10^{-7}$	$-1.343\,221 \cdot 10^{-4}$	$2.999\,016 \cdot 10^{-2}$	$-3.055\,446$	$1.135\,764 \cdot 10^2$
150 – 200	$1.209\,434 \cdot 10^{-8}$	$-9.692\,458 \cdot 10^{-6}$	$3.002\,041 \cdot 10^{-3}$	$-4.523\,015 \cdot 10^{-1}$	$1.919\,151 \cdot 10^1$
200 – 300	$8.113\,942 \cdot 10^{-10}$	$-9.822\,568 \cdot 10^{-7}$	$4.687\,616 \cdot 10^{-4}$	$-1.231\,710 \cdot 10^{-1}$	3.067 409
300 – 500	$9.814\,674 \cdot 10^{-11}$	$-1.654\,439 \cdot 10^{-7}$	$1.148\,115 \cdot 10^{-4}$	$-5.431\,334 \cdot 10^{-2}$	$-2.011\,365$
500 – 750	$-7.835\,161 \cdot 10^{-11}$	$1.964\,589 \cdot 10^{-7}$	$-1.657\,213 \cdot 10^{-4}$	$4.305\,869 \cdot 10^{-2}$	$-1.477\,132 \cdot 10^1$
750 – 1000	$2.813\,255 \cdot 10^{-11}$	$-1.120\,689 \cdot 10^{-7}$	$1.695\,568 \cdot 10^{-4}$	$-1.188\,941 \cdot 10^{-1}$	$1.456\,718 \cdot 10^1$

Tab. 3.13: Exponential power law density coefficients in accordance with the 1976 U.S. Standard Atmosphere.

Geom. Alt. h (km)	A_ρ	B_ρ	C_ρ	D_ρ	E_ρ
86 – 91	0.000 000	$-3.322\,622 \cdot 10^{-6}$	$9.111\,460 \cdot 10^{-4}$	$-2.609\,971 \cdot 10^{-1}$	5.944 694
91 – 100	0.000 000	$2.873\,405 \cdot 10^{-5}$	$-8.492\,037 \cdot 10^{-3}$	$6.541\,179 \cdot 10^{-1}$	$-2.362\,010 \cdot 10^1$
100 – 110	$-1.240\,774 \cdot 10^{-5}$	$5.162\,063 \cdot 10^{-3}$	$-8.048\,342 \cdot 10^{-1}$	$5.555\,996 \cdot 10^1$	$-1.443\,338 \cdot 10^3$
110 – 120	0.000 000	$-8.854\,164 \cdot 10^{-5}$	$3.373\,254 \cdot 10^{-2}$	$-4.390\,837$	$1.765\,294 \cdot 10^2$
120 – 150	$3.661\,771 \cdot 10^{-7}$	$-2.154\,344 \cdot 10^{-4}$	$4.809\,214 \cdot 10^{-2}$	$-4.884\,744$	$1.723\,597 \cdot 10^2$
150 – 200	$1.906\,032 \cdot 10^{-8}$	$-1.527\,799 \cdot 10^{-5}$	$4.724\,294 \cdot 10^{-3}$	$-6.992\,340 \cdot 10^{-1}$	$2.050\,921 \cdot 10^1$
200 – 300	$1.199\,282 \cdot 10^{-9}$	$-1.451\,051 \cdot 10^{-6}$	$6.910\,474 \cdot 10^{-4}$	$-1.736\,220 \cdot 10^{-1}$	$-5.321\,644$
300 – 500	$1.140\,564 \cdot 10^{-10}$	$-2.130\,756 \cdot 10^{-7}$	$1.570\,762 \cdot 10^{-4}$	$-7.029\,296 \cdot 10^{-2}$	$-1.289\,844 \cdot 10^1$
500 – 750	$8.105\,631 \cdot 10^{-12}$	$-2.358\,417 \cdot 10^{-9}$	$-2.635\,110 \cdot 10^{-6}$	$-1.562\,608 \cdot 10^{-2}$	$-2.002\,246 \cdot 10^1$
750 – 1000	$-3.701\,195 \cdot 10^{-12}$	$-8.608\,611 \cdot 10^{-9}$	$5.118\,829 \cdot 10^{-5}$	$-6.600\,998 \cdot 10^{-2}$	$-6.137\,674$

On any given day, the temperature at a particular elevation h may be corrected from the predicted standard by applying a corrective term ΔT to the entire temperature spectrum. The other properties (pressure and density), however, are left as is according to the predicted standard temperature.

3.5.6 Reference Atmospheric Models

Some atmospheric models include variation of air properties with respect to latitude and, typically, also time. This scale of time, however, is large (on the order of months and not with hour/minute resolution) since the atmosphere is still of the “standard” variety. A series of three tables of standard air properties from sea level to 2000 [km] with variation in altitude, latitude, and time (monthly) is found at braeunig.us/space/atmmmodel.htm.

3.5.7 *NRLMSISE-00 Atmospheric Model*

The Naval Research Lab (NRL) Mass-Spectrometer - Incoherent Scatter (MSISE) 2001 model provides an estimate of the air composition and its neutral temperature and density from the Earth's surface up into the thermosphere. This model extensively considers details and real data from past space shuttle missions so that the atmosphere varies in full three-dimensional space with time-of-year and time-of-day resolution while considering past and nearly up-to-date values for solar activity.

4. STRUCTURES

Aerospace vehicles must very often be made of materials balancing weight, durability, heat resistance, and cost among other things. In the context of temporary, unmanned vehicles in the atmosphere or in VLEO, all four of these areas are paramount to successfully launching and returning the vehicle. A typical metal found in many such vehicles is 6061 aluminum alloy having a material density of

$$\rho_{A6061} = 2700 \text{ [kg/m}^3\text{]}, \quad (4.1)$$

though in general, the materials used on a vehicle may vary throughout. As such, the masses and first and second moments of inertia (center of mass and inertia tensor, respectively) of individual vehicle components, and thus of the whole vehicle, may be estimated via exercises in geometry. This geometry, of course, is heavily dependent upon the design, so the components and the results are dependent upon fine-tuned design parameters such as shape, length, width, thickness, and material density among many others in general.

The fundamental forms of the quantities of interest here follow

$$M = \int_{\Omega} \rho(\underline{\vec{x}}) d^3 \underline{\vec{x}} \quad (4.2)$$

$$\underline{\vec{x}} = \frac{1}{M} \int_{\Omega} \rho(\underline{\vec{x}}) \underline{\vec{x}} d^3 \underline{\vec{x}} \quad (4.3)$$

$$[I] = \int_{\Omega} \rho(\underline{\vec{x}}) (|\underline{\vec{x}}|^2 \mathbb{I}_3 - \underline{\vec{x}} \otimes \underline{\vec{x}}) d^3 \underline{\vec{x}}. \quad (4.4)$$

where Ω is the three-dimensional analogue to the striped region outlined in Fig. 4.1. The distinction between $\underline{\vec{x}}$ and \vec{x} here is important. Using $\underline{\vec{x}}^T = (\underline{x} \ \underline{y} \ \underline{z})$ in (4.3) gives the center of mass $\underline{\vec{x}}$ with respect to the basis positioned on the (outer) tip of the rocket's nose cone, whereas using $\vec{x} = (x \ y \ z)$ in (4.4) gives the inertia matrix $[I]$ with respect to the basis positioned at the center of mass. As physical intuition may (correctly) predict, the contribution to terms in the first and second moments of inertia indicating variation from the longitudinal axis of a radially symmetric body is nothing. Thus, the center of mass lies along the longitudinal axis and the inertia matrix is diagonalized. Despite calculations still taking place in three dimensions, the radial component only becomes important regarding the geometric shape, which becomes a function of the longitudinal coordinate.

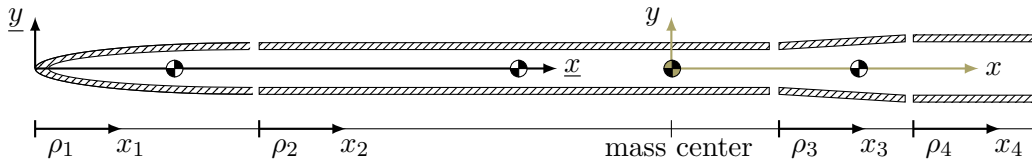


Fig. 4.1: Overviewing diagram of the cross-section of an empty, multistaged rocket body with its basic components separated from one another for individual analysis. The structure is radially symmetric from the longitudinal axis and hollow, containing material only around the outer shell. Measurements are made from the nose cone's (outer) tip. Comparatively small components like fins and nozzles are removed from consideration.

4.1 Determination of Characteristics for Fundamental Geometries

The mass of the (empty) vehicle follows (4.2) for which the material density $\rho(\vec{x})$ is likely piecewise constant in \underline{x} , i.e. it is likely that $\rho(\vec{x}) = \rho(\underline{x})$. Assuming such an equality holds and the vehicle is indeed a body of revolution consisting of n components/segments, for which the (longitudinal) length of section (or segment) $s \in \{1, 2, \dots, n\}$ is L_s with density ρ_s and the distances of the outer and inner surfaces from the longitudinal axis are described by $f(x_s)$ and $g(x_s)$ respectively for segments $x_s \in [0, L_s]$, then the mass integral is equivalently stated

$$M = \sum_{s=1}^n \underbrace{\rho_s \int_0^{L_s} A(x_s) dx_s}_{\text{segment mass } M_s}, \quad (4.5)$$

whereby the triple integration of (4.2) has been reduced to a single integral with the introduction of the *local* area A of the region at x_s .

$$A(x_s) = \pi(f(x_s)^2 - g(x_s)^2), \quad (0 \leq g < f) \quad (4.6)$$

Thus, the total mass is simply the sum of the masses of individual components composing the whole system.

With the mass of segment s known, its center of mass with respect to the position x_s indicated by Fig. (4.1) follows the form of (4.3) as

$$\bar{x}_s = \frac{1}{M_s} \rho_s \int_0^{L_s} A(x_s) x_s dx_s, \quad (4.7)$$

where vector dependence has been dropped due to radial symmetry. The center of mass for the entire system then follows the standard form

$$\bar{x} = \sum_s \frac{M_s}{M} \bar{x}_s. \quad (4.8)$$

Of course, the distances of the center of mass from the y_s and z_s axes are zero by radial symmetry.

The inertia matrix of each segment *with respect to the segment's center of mass* is then

$$[I]_s = \frac{\rho_s}{2} \int_0^{L_s} A(x_s) \text{diag} \begin{pmatrix} B(x_s) \\ C(x_s) \\ C(x_s) \end{pmatrix} dx_s - M_s(|\vec{r}_s|^2 \mathbb{I}_3 - \vec{r}_s \otimes \vec{r}_s), \quad (4.9)$$

for which the auxillary functions and quantity \vec{r}_s are defined

$$B(x_s) = f(x_s)^2 + g(x_s)^2 \quad C(x_s) = \frac{1}{2}B(x_s) + 2x_s^2 \quad \vec{r}_s = (\bar{x}_s \quad 0 \quad 0)^T. \quad (4.10)$$

Applying another translation to express each segment's inertia matrix with respect to one point then results in a single inertia matrix describing the entire system. The most convenient point at which to express this matrix is at the system's center of mass, $\vec{\bar{x}}$, for which the inertia matrix $[I]$ of the entire system due to fixed structure is

$$[I] = \sum_s [I]_s - M_s(|\vec{r}|^2 \mathbb{I}_3 - \vec{r} \otimes \vec{r}), \quad (4.11)$$

where \vec{r} is the vector pointing from the segment's center of mass $\vec{\bar{x}}_s$ to the system's overall center of mass $\vec{\bar{x}}$ (expressed in that segment's coordinate system positioned at its own center of mass). Of course, since the system is radially symmetric, this relation reduces to a simple application of the parallel axis theorem. In fact, only the I_{yy} and I_{zz} (second and third diagonals, respectively) components require adjusting; the I_{xx} (first diagonal) component is already obtained and requires no further modification as far as permanent structures are concerned.

4.1.1 Nose Cone Configurations

Solid Cone

The solid cone is characterized by $f, g : [0, L] \rightarrow \mathbb{R}$ defined by

$$f(x; \theta_c) = \tan \theta_c x \quad g(x) = 0, \quad (4.12)$$

where $\tan \theta_c = R/L$.

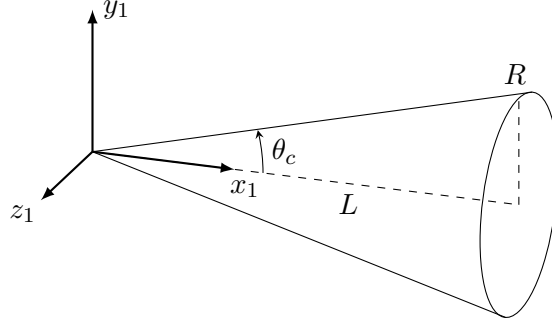


Fig. 4.2: Solid cone of constant density ρ and half (cone) angle θ_c satisfying $\tan \theta_c = R/L$. The coordinate frame is labelled appropriately with the index $s = 1$ to reinforce that the section is in agreement with Fig. 4.1 and will remain so in subsequent figures.

For a constant density ρ , the properties (mass, center of mass, and inertia matrix relative to the center of mass) considering a nondimensionalization $R = \alpha L$ such that $R/L = \alpha$ follow

$$V = \frac{M}{\rho} = \frac{\pi R^2 L}{3}, \quad \bar{x} = \frac{3}{4}L, \quad [I] = \frac{3MR^2}{20} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 + 1/(4\alpha^2) & 0 \\ 0 & 0 & 1 + 1/(4\alpha^2) \end{pmatrix}. \quad (4.13)$$

Conical Shell with Finite Thickness

A conical shell with *frontal* thickness $0 < l < L$ may be obtained upon providing a generalization of the solid cone by considering $f, g : [0, L] \rightarrow \mathbb{R}$ defined by

$$f(x; \theta_c) = \tan \theta_c x \quad g(x; l) = H(x - l)f(x - l), \quad (4.14)$$

where $\tan \theta_c = R/L = \alpha$ as before, but now g is nonzero with H acting as the Heaviside step function.

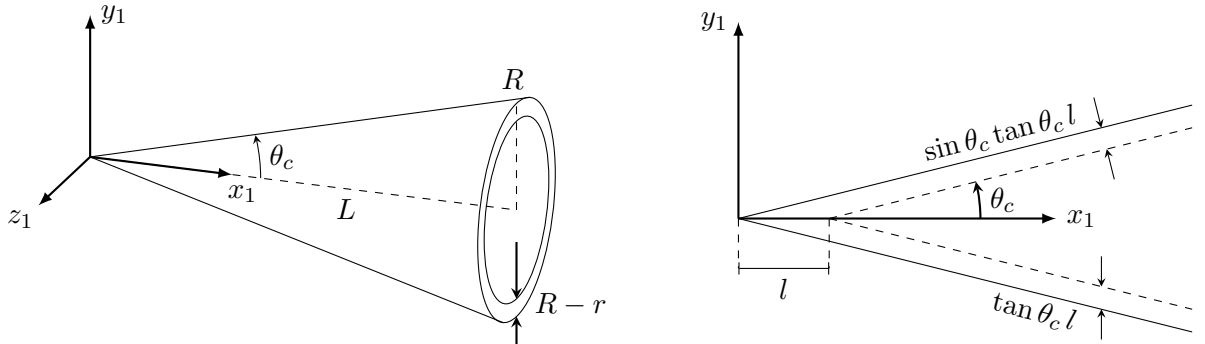


Fig. 4.3: Conical shell of constant density ρ , frontal thickness l , and half (cone) angle satisfying $\tan \theta_c = R/L$.

Defining $\ell = l/L$ as a nondimensional parameter of the shell, the properties follow

$$V = \frac{M}{\rho} = \frac{\pi R^2 L}{3} (\ell^3 - 3\ell^2 + 3\ell), \quad \bar{x} = \frac{L}{4} \left(\frac{\ell^3 - 6\ell + 8}{\ell^2 - 3\ell + 3} \right) \quad (4.15)$$

$$[I] = \frac{MR^2}{80} \text{diag} \begin{pmatrix} 24 \frac{\ell^4 - 5\ell^3 + 10\ell^2 - 10\ell + 5}{\ell^2 - 3\ell + 3} \\ \frac{1}{\alpha^2} \left(3(1 + 4\alpha^2)(1 - \ell)^2 + 2 \frac{5 - (1 - 6\alpha^2)(2 - \ell)}{\ell^2 - 3\ell + 3} - \frac{5}{(\ell^2 - 3\ell + 3)^2} \right) \\ \frac{1}{\alpha^2} \left(3(1 + 4\alpha^2)(1 - \ell)^2 + 2 \frac{5 - (1 - 6\alpha^2)(2 - \ell)}{\ell^2 - 3\ell + 3} - \frac{5}{(\ell^2 - 3\ell + 3)^2} \right) \end{pmatrix}, \quad (4.16)$$

Parabolic Shell with Finite Thickness

The parabolic cone with *frontal* thickness $0 < l < L$ may be obtained by generalizing the conical shell by considering $f, g : [0, L] \rightarrow \mathbb{R}$ defined by

$$f(x; k) = \frac{2 - k(x/L)}{2 - k} (x/L) R \quad g(x; l) = H(x - l) f(x - l), \quad (4.17)$$

where $0 \leq k \leq 1$ is the parabolic cone constant that determines the cone's overall curvature.

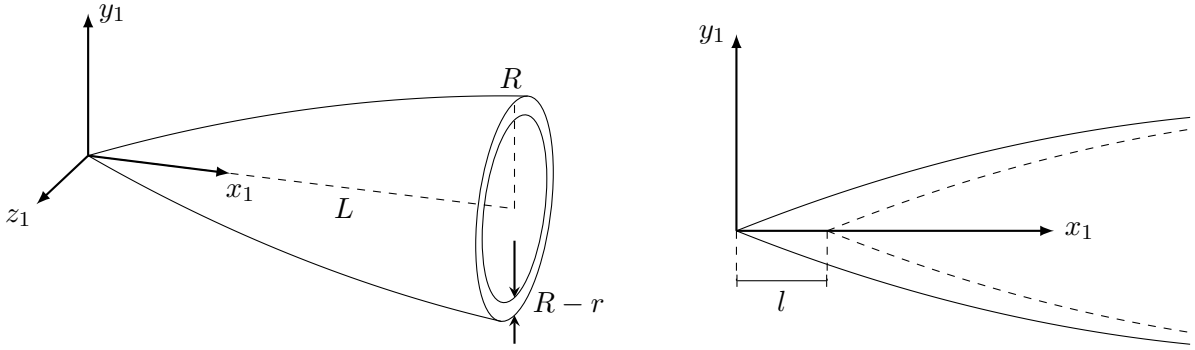


Fig. 4.4: Parabolic cone of constant density ρ , frontal thickness l , and cone constant $k = 3/4$, corresponding to a three-quarter parabolic nose cone.

With $k = 0$, all curvature vanishes such that the resulting geometry degenerates into the conical shell, and with $k = 1$, the shape is a true paraboloid such that $\partial_x f(L) = 0$. That is, the cone smoothly matches its curvature for the body on which it will be mounted in the case when $k = 1$. Otherwise, the boundary joining these two components forms an edge.

Introducing nondimensional parameters α and ℓ such that $R = \alpha L$ and $l = \ell L$, the properties of the parabolic nose cone are

$$\frac{M}{\rho} = \frac{\pi R^2 L}{3} \left(\ell \frac{20(\ell^2 - 3\ell + 3) + 15(\ell^3 - 4\ell^2 + 6\ell - 4)k + 3(\ell^4 - 5\ell^3 + 10\ell^2 - 10\ell + 5)k^2}{5(2 - k)^2} \right) \quad (4.18)$$

$$\bar{x} = \frac{L}{2} \left(\frac{10(\ell^3 - 6\ell + 8) + 6(\ell^4 - 10\ell^2 + 20\ell - 15)k + (\ell^5 - 15\ell^3 + 40\ell^2 - 45\ell + 24)k^2}{20(\ell^2 - 3\ell + 3) + 15(\ell^3 - 4\ell^2 + 6\ell - 4)k + 3(\ell^4 - 5\ell^3 + 10\ell^2 - 10\ell + 5)k^2} \right) \quad (4.19)$$

$$I_{xx} = \frac{MR^2}{42} \left\{ 1008(1 + (\ell - 1)^5) - 1680(1 - (\ell - 1)^6)k + 1080(1 + (\ell - 1)^7)k^2 \right. \\ \left. - 315(1 - (\ell - 1)^8)k^3 + 35(1 + (\ell - 1)^9)k^4 \right\} / \left\{ (k - 2)^2 (20(\ell^3 - 3\ell^2 + \ell) \right. \\ \left. + 15(\ell^4 - 4\ell^3 + 6\ell^2 - 4\ell)k + 3(\ell^4 - 5\ell^3 + 10\ell^2 - 10\ell + 5)k^2) \right\} \quad (4.20)$$

$$I_{yy} = I_{zz} \quad (4.21)$$

$$= \frac{ML^2}{84} \left\{ 336(2(15 - 10\ell + \ell^4) + 3(5 - 10\ell + 10\ell^2 - 5\ell^3 + \ell^4)\alpha^2) \right. \\ - 336((66 - 65\ell + 20\ell^2 + 2\ell^4 - \ell^5) + 5(6 - 15\ell + 20\ell^2 - 15\ell^3 + 6\ell^4 - \ell^5)\alpha^2)k \\ + 24((749 - 952\ell + 490\ell^2 - 70\ell^3 + 7\ell^4 - 14\ell^5 + 2\ell^6) \\ + 45(7 - 21\ell + 35\ell^2 - 35\ell^3 + 21\ell^4 - 7\ell^5 + \ell^6)\alpha^2)k^2 \\ + 3(-4(532 - 819\ell + 560\ell^2 - 140\ell^3 - 7\ell^5 + 4\ell^6) \\ + 105(-8 + 28\ell - 56\ell^2 + 70\ell^3 - 56\ell^4 + 28\ell^5 - 8\ell^6 + \ell^7)\alpha^2)k^3 \\ + (12(70 - 126\ell + 105\ell^2 - 35\ell^3 + \ell^6) \\ + 35(9 - 36\ell + 84\ell^2 - 126\ell^3 + 126\ell^4 - 84\ell^5 + 36\ell^6 - 9\ell^7 + \ell^8)\alpha^2)k^4 \Big\} / \\ \left\{ (k - 2)^2 (20(\ell^2 - 3\ell + 3) + 15(\ell^3 - 4\ell^2 + 6\ell - 4)k \right. \\ \left. + 3(\ell^4 - 5\ell^3 + 10\ell^2 - 10\ell + 5)k^2) \right\} - M\bar{x}^2 \quad (4.22)$$

Notice that taking a double limit as $k \rightarrow 0$ (conical shell) and $\ell \rightarrow 1$ (solid) correctly yields the desired results of the solid cone. In fact, the limiting results are not particular to the solid cone; they hold for the conical shell too as intended.

Introducing the parabolic nose cone is beneficial in the fact that its results are analytic (specifically, rational functions) and can make use of its freedom with k to closely approximate the structural properties of various others like the tangent and secant ogives, and some of the Haack nose cones, specifically the von Kármán ogive, whose properties are difficult to obtain/manage exactly.

Haack Series with Finite Thickness

The Haack series is summarized here to provide graphical comparison to the parabolic cones. The Haack ogives with *frontal* thickness $0 < l < L$ are obtained by considering $f, g : [0, L] \rightarrow \mathbb{R}$ defined by

$$f(x; c) = R \sqrt{\frac{1}{\pi} \left(\psi(x) - \frac{\sin 2\psi(x)}{2} + c \sin^3 \psi(x) \right)} \quad g(x; l) = H(x - l)f(x - l), \quad (4.23)$$

where $\psi(x) = \cos^{-1}(1 - 2(x/L))$. The von Kármán ogive, the nose cone that theoretically sources the least amount of aerodynamic drag, is obtained by setting the cone constant $c = 0$.

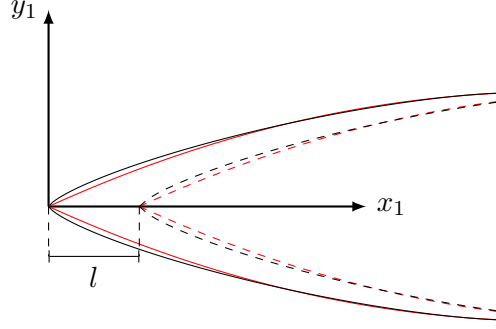


Fig. 4.5: Comparison of the parabolic cone of cone constant $k = 9/10$ (red, slightly inside) and the von Kármán ogive of the Haack series characterized by the cone constant $c = 0$ (black, slightly outside).

The mass and center of mass properties are

$$V = \frac{M}{\rho} = \frac{R^2 L}{48} \left[9c\pi - ((48 + 18c) - (64 - 12c)\ell + (64 - 144c)\ell^2 + 96c\ell^3) \sqrt{(1 - \ell)\ell} - 6(16\ell + 3c - 8) \cos^{-1} \sqrt{\ell} - 48 \sin^{-1}(1 - 2\ell) \right] \quad (4.24)$$

$$\begin{aligned} \bar{x} = \frac{L}{10} \bigg\{ & (75 + 45c)\pi - \left((330 + 90c) - (590 - 150c)\ell + (660 - 1608c)\ell^2 - (240 - 1944c)\ell^3 \right. \\ & \left. - (160 + 384c)\ell^4 - 192c\ell^5 \right) \sqrt{\frac{\ell}{1 - \ell}} - 90c(1 + 2\ell) \cos^{-1} \sqrt{\ell} + 15 \cos^{-1}(1 - 2\ell) \\ & + 240\ell(1 + \ell) \cos^{-1}(2\ell - 1) - 150 \sin^{-1}(1 - 2\ell) \bigg\} / \\ & \left\{ (24 + 9c)\pi - ((48 + 18c) + 12c\ell - 144c\ell^2 + 96c\ell^3) \sqrt{(1 - \ell)\ell} + 64\sqrt{(1 - \ell)\ell}^3 \right. \\ & \left. + (48 + 96\ell - 18c) \cos^{-1} \sqrt{\ell} - 48 \cos^{-1}(2\ell - 1) \right\}. \end{aligned} \quad (4.25)$$

The inertia matrix is spared, but it indeed may be determined from (4.9) exactly or also numerically if need be. Note that the center of mass has a finite limit as $\ell \rightarrow 1$. The limit is simply

$$\lim_{\ell \rightarrow 1} \bar{x} = L \left(\frac{11 + 3c}{16 + 6c} \right). \quad (4.26)$$

Elliptical Shell with Finite Thickness

The elliptical shell with *frontal* thickness $0 < l < L$ may be obtained by considering a new geometry $f, g : [0, L] \rightarrow \mathbb{R}$ defined by

$$f(x) = R \sqrt{1 - \frac{(x - L)^2}{L^2}} \quad g(x; l) = H(x - l) f(x - l). \quad (4.27)$$

Here, “new geometry” means one that does not generalize upon the parabolic nose cone. Upon visual inspection of Fig. 4.6, the fundamental difference between the elliptical and the parabolic cones is the tip’s sharpness (which has implications as to how shockwaves might behave when experiencing a supersonic flow field). Supposing the wall thicknesses are invariant, cones with blunter tips effectively move the center of mass more forward than cones with sharper tips, structurally.

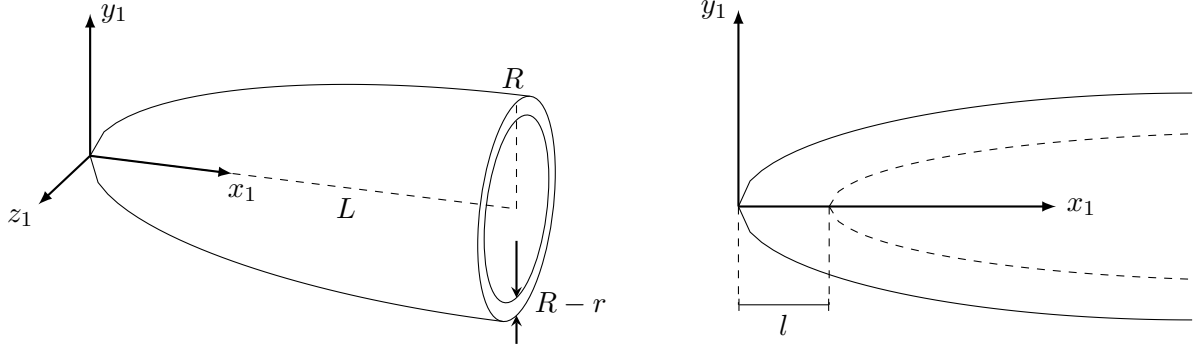


Fig. 4.6: Elliptical cone of constant density ρ and frontal thickness l .

The properties of interest follow after defining nondimensional parameters $\alpha = R/L$ and $\ell = l/L$.

$$V = \frac{M}{\rho} = \frac{\pi R^2 L}{3}(\ell^3 - 3\ell), \quad \bar{x} = \frac{L}{4} \left(\frac{\ell^3 + 4\ell^2 - 6\ell - 4}{\ell^2 - 3} \right) \quad (4.28)$$

$$[I] = \frac{ML^2}{80} \text{diag} \begin{pmatrix} 8 \frac{3\ell^4 - 10\ell^2 + 15}{\ell^2 - 3} \alpha^2 \\ \frac{(3\ell^6 - 44\ell^4 + 80\ell^3 - 60\ell^2 + 40) - 4(3\ell^6 - 19\ell^4 + 45\ell^2 - 45)\alpha^2}{(\ell^2 - 3)^2} \\ \frac{(3\ell^6 - 44\ell^4 + 80\ell^3 - 60\ell^2 + 40) - 4(3\ell^6 - 19\ell^4 + 45\ell^2 - 45)\alpha^2}{(\ell^2 - 3)^2} \end{pmatrix} \quad (4.29)$$

4.1.2 Body Configurations

Solid Cylinder

The solid cylinder is characterized by $f, g : [0, L] \rightarrow \mathbb{R}$ defined by

$$f(x) = R \quad g(x) = 0. \quad (4.30)$$

With R a fixed constant, the resulting surface described by this geometry is the simplest of all bodies of revolution to analyze.

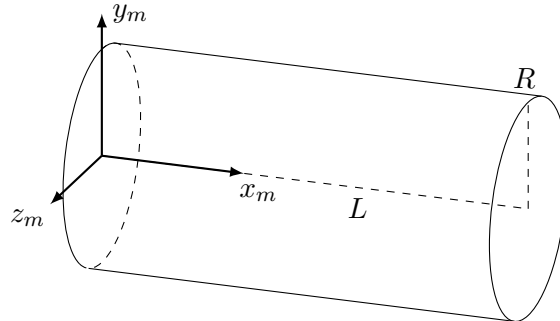


Fig. 4.7: Solid cylinder of constant density ρ .

For constant density ρ , the properties (mass, center of mass, and inertia matrix relative to

the center of mass) considering the introduction of a nondimensional parameter $\alpha = R/L$ follow

$$V = \frac{M}{\rho} = \pi R^2 L, \quad \bar{x} = \frac{L}{2}, \quad [I] = \frac{MR^2}{4} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (4.31)$$

Hollowed Cylinder

The hollow cylinder of *inner* radius $0 < r < R$ is characterized by $f, g : [0, L] \rightarrow \mathbb{R}$ defined by

$$f(x) = R, \quad g(x) = r. \quad (4.32)$$

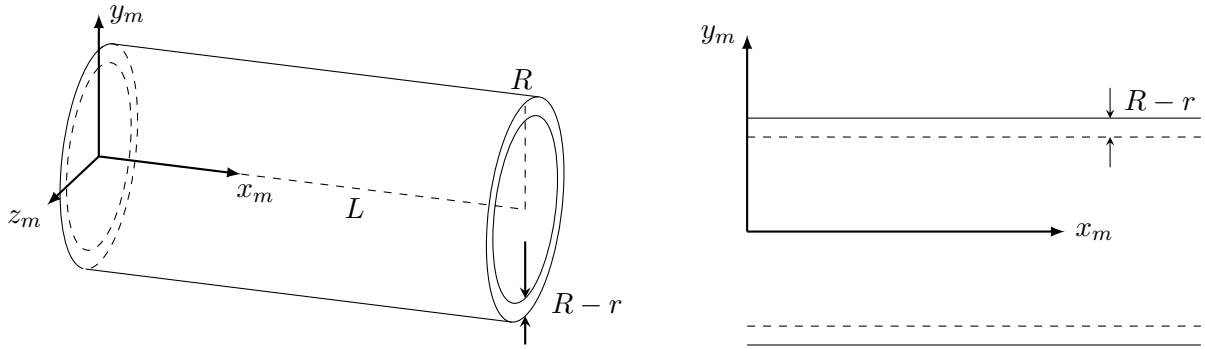


Fig. 4.8: Hollow cylinder of constant density ρ and wall thickness $R - r$.

For constant density ρ , the properties are

$$V = \frac{M}{\rho} = \pi(R^2 - r^2)L, \quad \bar{x} = \frac{L}{2}, \quad (4.33)$$

$$[I] = \frac{M}{12} \text{diag} \begin{pmatrix} 6(R^2 + r^2) \\ 3(R^2 + r^2) + L^2 \\ 3(R^2 + r^2) + L^2 \end{pmatrix}. \quad (4.34)$$

Intuitively and indeed, taking $r \rightarrow 0$ reduces back to the solid cylinder. Note, however, that taking $r = R$, where the hollow cylinder vanishes, gives nonzero properties.

Radial Frustum

The radial frustum is obtained by cutting a circular cone with two vertical planes at $x = a$ and $x = b$ with $0 < a < b$, where then the length of the frustum is $L = b - a$. The form of f, g may therefore be borrowed from the case of the conical shell (Fig. 4.3) by applying a remapping $x \mapsto x - a$, effecting functions $f, g : [0, L] \rightarrow \mathbb{R}$ of the form

$$f(x) = R_0 + \frac{R - R_0}{L}x, \quad g(x) = f(x - l), \quad (4.35)$$

where R_0 is the outer radius at $x = 0$, R is the outer radius at $x = L$, and $0 < l < R_0L/(R_0 + R)$ is the *frontal* thickness.

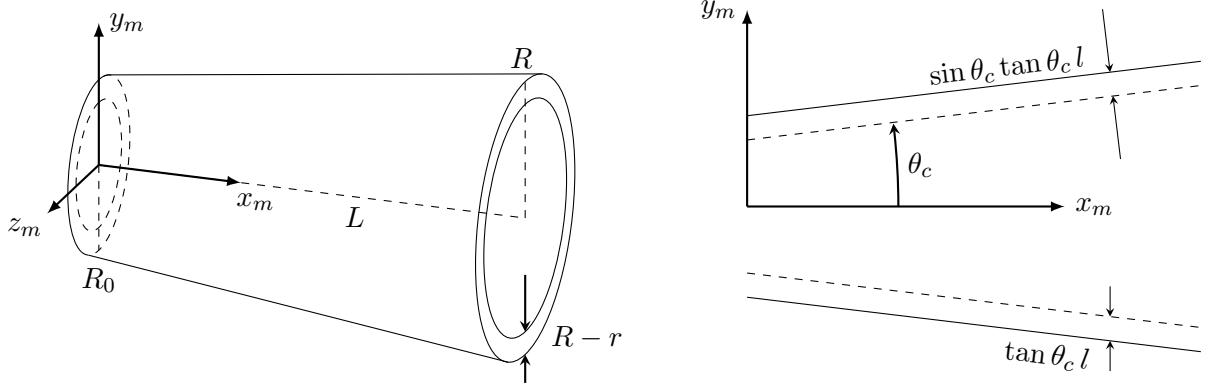


Fig. 4.9: Hollow conical frustum of constant density ρ and wall thickness l .

The properties of the frustum of constant density and nondimensional parameters $\alpha = R/L$, $\beta = R_0/L$, and $\ell = l/L$ follow.

$$V = \frac{M}{\rho} = \pi L^3 ((\alpha^2 - \beta^2)\ell - (\alpha - \beta)^2 \ell^2), \quad \bar{x} = \frac{L}{6} \left(\frac{4\alpha + 2\beta - 3(\alpha - \beta)\ell}{\alpha + \beta - (\alpha - \beta)\ell} \right) \quad (4.36)$$

$$I_{xx} = \frac{ML^2}{2} (\alpha^2 + \beta^2 - (\alpha^2 - \beta^2)\ell + (\alpha - \beta)^2 \ell^2) \quad (4.37)$$

$$I_{yy} = I_{zz} \quad (4.38)$$

$$\begin{aligned} &= \frac{ML^2}{36} \frac{1}{(\alpha + \beta - (\alpha - \beta)\ell)^2} \left\{ (2\alpha^2 + 9\alpha^4 + 8\alpha\beta + 18\alpha^3\beta + 2\beta^2 + 18\alpha^2\beta^2 + 18\alpha\beta^3 + 9\beta^4) \right. \\ &\quad - 3(\alpha^2 + 9\alpha^4 + 6\alpha^3\beta - 2\beta^2 - 6\alpha\beta^3 - 9\beta^4)\ell \\ &\quad + 3(\alpha^2 + 12\alpha^4 - 2\alpha\beta - 6\alpha^3\beta + \beta^2 - 12\alpha^2\beta^2 + 12\beta^4)\ell^2 \\ &\quad - 27(\alpha^4 - 2\alpha^3\beta + 2\alpha\beta^3 - \beta^4)\ell^3 \\ &\quad \left. + 9(\alpha^4 - 4\alpha^3\beta + 6\alpha^2\beta^2 - 4\alpha\beta^3 + \beta^4)\ell^4 \right\}. \end{aligned} \quad (4.39)$$

Note that under this construction, taking $\beta \rightarrow \alpha$ yields $R_0 \rightarrow R$ and thus $\theta_c \rightarrow 0$, which reduces the above results to those of the cylindrical sheet (of zero thickness).

If the frustum is of the form $R_0 > R$, then the above results hold as viewed from the frame opposite of the current frame with its x and y axes reversed (so that it appears to copy the form of Fig. 4.9). As such, the results hold in this case under the transformation $x \mapsto L - x$; no transformation is necessary to correct results on the y axis since the opposing frame is still radially symmetric about the x axis.

4.1.3 Cavity Tuning

The inside surface of each geometry characterized by the function $g(x; l) = H(x - l)f(x - l)$ can be further generalized by considering a remapping of the surface $g(x; l)$ as

$$g(x; l) \mapsto H(x - l)f(x - l)h(l). \quad (4.40)$$

Specifically, the previously considered nose cones were special cases of (4.40) such that $h(l) \equiv 1$. As l is simply a parameter of the geometry, $h(l)$ poses no trouble for the procedure of integration; it simply appears repeatedly throughout the results. When choosing h , the bound

$$f(x - l)h(l) < f(x) \quad (4.41)$$

should be respected to ensure physical nose cones or other geometries are realized.

5. PROPULSION

Most air and space vehicles (jets, rockets, spacecraft, etc.) utilize some sort of propulsive device to change their current state by ejecting mass out from their systems. That is, the effect of propulsion is to deliver a reactionary force imparted by a fluid particle leaving the system in which it was carried. This idea is realized by the conservation of momentum for a system of initial mass $m + dm$, where the mass of the fluid particle is dm , travelling at a velocity \vec{v} .

$$(m + dm)\vec{v} = m(\vec{v} + d\vec{v}) + dm(\vec{v} - \vec{v}_e). \quad (5.1)$$

Here, \vec{v}_e is the exit velocity of the fluid particle relative to the system (as if the system was stationary). Consequently, the force is written

$$\vec{F} = \frac{dm}{dt}\vec{v}_e, \quad (5.2)$$

where m identifies the current mass of the system. Inherently then, this rate \dot{m} is always non-positive during free flight. This negativity, however, is again negated by the fact that \vec{v}_e acts in the direction opposite to the system's velocity \vec{v} . Thus, the resulting force on the system acts in the direction of its velocity \vec{v} . This force, however, is *not* the force of thrust. A closer inspection reveals that the conservation of momentum in this context is actually realized by

$$\vec{F}_T - \iint_S p \hat{n} dS = \frac{\partial}{\partial t} \iiint_V \rho \vec{v} dV + \iint_S \rho \vec{v} (\vec{v} \cdot \hat{n}) dS, \quad (5.3)$$

where \vec{F}_T is the force of thrust, p and ρ are static air pressure and static air density, and \hat{n} is the unit outward normal to the control volume of surface area S and volume V . An assumption that the exit velocity is constant and completely perpendicular to the plane spanning across the nozzle's exit, which is decent, yields from (5.3) that

$$\vec{F}_T \approx (\dot{m}v_e + (p_e - p_n)A_e)\hat{v}. \quad (5.4)$$

Here, the subscripts n and e respectively denote positions at the vehicle's nose tip and nozzle exit. When moving at supersonic speed with a sharp nose tip, this pressure becomes the freestream pressure p_∞ since, by definition of being supersonic, perturbations cannot propagate their way upstream to influence incoming fluid properties. Otherwise, a blunt-nosed body creates a bow shock, which induces complications in finding the air pressure acting on the nose.

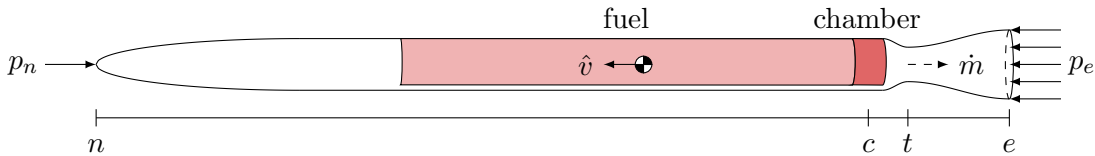


Fig. 5.1: Overviewing diagram of the systems and quantities relevant to the basics of rocket propulsion. Particularly, this diagram contains no aerodynamic considerations (shockwaves) in the visualization of p_n nor in the exhaust field. The fuel is represented by a block-cutaway to reserve space for either solid propellant or liquid fuel and oxidizer. The chamber, leading into the nozzle, is characterized by a constant total pressure, total temperature, and total density. The exhaust velocity is not shown since its reference frame (the rocket) is different from the reference frame monitoring the velocity and thrust force. The exit area A_e is simply the cross-sectional area of the nozzle at the exit.

5.1 Quasi-Unidimensional Nozzle Flow

Gas expelled from a system is typically done so by sourcing a large chamber pressure, typically denoted p_c , (which is a total pressure) to force the gas through a *nozzle* which produces an exit pressure p_e along nozzle's exit plane among many other things. Thus, gas flow within a rotationally symmetric geometry is of interest to determine the amount of thrust produced.

The *de Laval* nozzle, or simply called the *converging-diverging* nozzle, is characterized by a symmetrically converging section in which all gas flow is subsonic, a *throat*, and a diverging section in which the gas flow may be subsonic, supersonic, or a combination of both.

The throat is the critical region that determines the flow field in the diverging section of the nozzle. A true fact is that the flow here can be only sonic (Mach 1) *at most*. Thus, a flow that is subsonic at the throat is necessarily subsonic in the diverging section, so the flow is entirely subsonic. A flow, however, that is sonic at the throat and has sufficient chamber pressure p_c relative to the backpressure, denoted p_b , to push the flow into the supersonic regime will be supersonic in the diverging section and will continue to be so unless encountering a shockwave in the section. The flow regime of the nozzle is largely determined by the pressure ratio p_b/p_c of the back pressure to the chamber pressure and the area ratio A_e/A_t of the exit area to the throat area.

Because the de Laval nozzle is diverging past the throat by definition, then a safe assumption is that the area A is monotonically increasing past the throat. Consequently, the nondimensional stations x along the nozzle's longitudinal axis, where $x = 0$ indicates the throat and $x = 1$ indicates the nozzle exit, are obtainable through the areas via

$$x = \frac{A - A_t}{A_e - A_t}. \quad (5.5)$$

This nondimensional distance therefore interprets physical locations within the nozzle as a percentage downstream of the throat up until the exit.

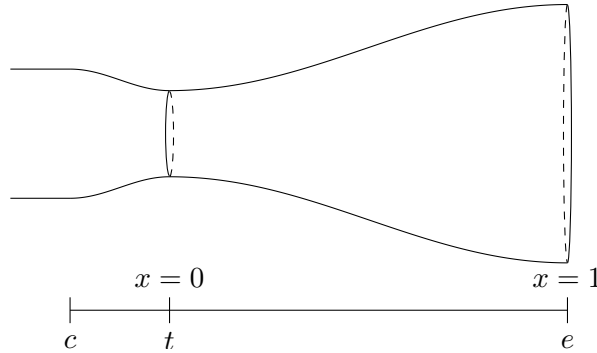


Fig. 5.2: Nondimensionalization of the quasi-unidimensional flow coordinate frame, where the throat is at $x = 0$ and the exit is at $x = 1$. The chamber is not designated a coordinate in terms of x , but the subscript c is important nonetheless in the determination of the flow.

5.1.1 Subsonic Flow

An entirely subsonic flow field through the nozzle is characterized by the theoretical throat area A^* for which the flow would be sonic under the current flow conditions. This area is found according to

$$A^* = M_e \left(\frac{2}{\gamma + 1} \left(1 + \frac{\gamma - 1}{2} M_e^2 \right) \right)^{-\frac{1}{2} \frac{\gamma + 1}{\gamma - 1}} \quad (5.6)$$

where γ is the ratio of specific heats and M_e is the Mach number at the nozzle's exit given (in the subsonic flow regime) by the pressure ratio.

$$M_e = \left(\frac{2}{\gamma - 1} \left(\left(\frac{p_b}{p_c} \right)^{\frac{1}{\gamma} - 1} - 1 \right) \right)^{1/2} \quad (5.7)$$

The (subsonic) Mach numbers M throughout the diverging section of the nozzle at nondimensional stations x are then found by solving the supersonic Mach-area relation

$$\frac{1}{M} \left(\frac{2}{\gamma + 1} \left(1 + \frac{\gamma - 1}{2} M^2 \right)^{\frac{1}{2} \frac{\gamma + 1}{\gamma - 1}} \right) = \frac{A}{A^*}. \quad (5.8)$$

The nozzle flow properties at these same stations x (corresponding with the Mach numbers M) are therefore given by the isentropic relations

$$\frac{T}{T_0} = \left(1 + \frac{\gamma - 1}{2} M^2 \right)^{-1} \quad (5.9)$$

$$\frac{p}{p_0} = \left(1 + \frac{\gamma - 1}{2} M^2 \right)^{\frac{\gamma}{1 - \gamma}} \quad (5.10)$$

$$\frac{\rho}{\rho_0} = \left(1 + \frac{\gamma - 1}{2} M^2 \right)^{\frac{1}{1 - \gamma}} \quad (5.11)$$

where the subscripted zero indicates total conditions. For subsonic flow throughout the entire nozzle, these total conditions are found in the chamber, i.e. $T_0 = T_c$, $p_0 = p_c$, and $\rho_0 = \rho_c$.

5.1.2 Supersonic Flow

Flow speeds at and in excess of Mach 1 introduce the shockwaves, or shocks, which are caused by the sudden and extreme compression of a gas. Shocks come in many different varieties, but in the context of quasi-1D nozzle flow, the *normal shock*, *oblique shock*, and *rarefaction wave* (also called an *expansion wave* or *Prandtl-Meyer fan*) are of greatest importance.

Normal Shocks

A normal shock (an adiabatic, but not isentropic, process) can result in the diverging portion of the nozzle if the chamber pressure creates too little of a difference between itself and the back pressure p_b . A normal shock *always* creates a subsonic flow on its far side (opposite of the supersonic flow) and therefore increases the exit pressure than if the shock was absent. A shock in the nozzle in the context of rocket propulsion is quite bad.

In this case, the exit pressure p_e necessarily must be the back pressure p_b and the (subsonic) Mach number at the exit is written

$$M_e^2 = \sqrt{\frac{1}{(\gamma - 1)^2} + \frac{2}{\gamma - 1} \left(\frac{2}{\gamma + 1} \right)^{\frac{\gamma + 1}{\gamma - 1}} \left(\frac{A_t/A_e}{p_b/p_c} \right)^2} - \frac{1}{\gamma - 1} \quad (5.12)$$

Finding the shock location and resulting flow properties upstream and downstream of the shock follows. The area of the shock is found from the Mach number directly upstream of the shock $M_{s,u} \geq 1$, which is obtained by solving

$$\left(1 + \frac{2\gamma}{\gamma + 1} (M_{s,u}^2 - 1) \right)^{1 - \frac{\gamma}{\gamma - 1}} \left(\frac{(\gamma + 1)M_{s,u}^2}{2 + (\gamma - 1)M_{s,u}^2} \right)^{\frac{\gamma}{\gamma - 1}} = \frac{p_b}{p_c} \left(1 + \frac{\gamma - 1}{2} M_e^2 \right)^{\frac{\gamma}{\gamma - 1}}. \quad (5.13)$$

The area of the shock is then found by rearrangement of (5.8)

$$A_s = \frac{A_t}{M_{s,u}} \left(\frac{2}{\gamma + 1} \left(1 + \frac{\gamma - 1}{2} M_{s,u}^2 \right) \right)^{\frac{1}{2} \frac{\gamma + 1}{\gamma - 1}} \quad (5.14)$$

for which the nondimensional position is obtained by (5.5). The flow is therefore categorized into two regimes as being upstream (behind) and downstream (in front of) the shock. That is, the upstream flow is located between the throat and the shock and the downstream flow is located between the shock and the exit.

Upstream of the shock, the flow behaves like usual (choked) supersonic flow without shocks for which the Mach numbers at stations x from the throat up to the shock are obtained again by (5.8). Here, the solution M_u at each station x upstream of the shock must satisfy $M_u \geq 1$. Downstream of the shock, however, the flow is in the subsonic regime so that the Mach numbers M_d at stations x between the shock and exit are, again, obtained via the Mach-area relation (5.8), but now with each satisfying $M_d < 1$. The distribution of Mach numbers throughout the nozzle at every station x are then logically formed by the relation

$$M = M_u \cup M_d, \quad (5.15)$$

for which the flow properties at these same stations follow (5.9 – 5.11).

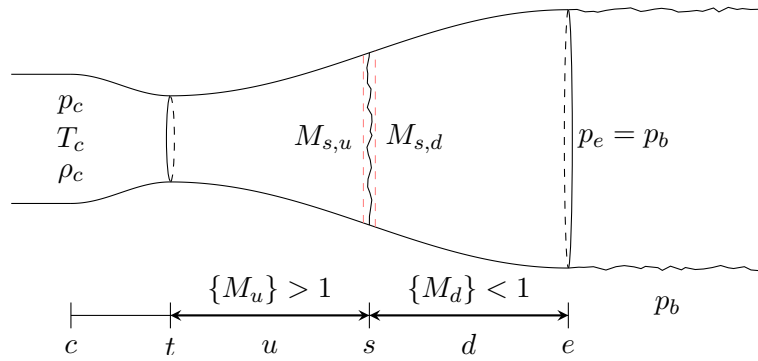


Fig. 5.3: Flow resulting from normal shock in the nozzle. The Mach numbers $M_{s,u}$ and $M_{s,d}$ are the values immediately upstream and immediately downstream of the shock, respectively.

Oblique Shocks

Oblique shocks emanating from the nozzle under a quasi-unidimensional flow assumption indicates *overexpanded flow* and means that the flow is choked throughout the entire nozzle such that there are no normal shocks. This flow regime often occurs with high-powered motors at low elevations and is characteristic of creating initially inwards-bowing exhaust plumes with repeating shock diamonds (Mach disks). It is bounded between the cases of shock at the exit and the design condition.

Because the flow throughout the nozzle is choked, the exit pressure p_e is therefore obtained via (5.10) as

$$p_e = p_c \left(1 + \frac{\gamma - 1}{2} M_e^2 \right)^{\frac{\gamma}{1 - \gamma}}, \quad (5.16)$$

where $M_e > 1$ satisfies the Area-Mach relation (5.8), in which $A/A^* = A_e/A_t$. Furthermore, the shock angle β , as measured from the exit plane outside of the shock region, is approximated as a linear function χ of the pressure ratio p_b/p_c as

$$\beta = \frac{\pi}{4}(1 + \chi), \quad (5.17)$$

where $0 < \chi < 1$ is a measure of the percentage of pressure ratios between having a normal shock at the exit and the design condition.

$$\chi = \frac{p_b/p_c - \text{pressure ratio for "design"}}{\text{pressure ratio for "normal shock at exit" - pressure ratio for "design"}}. \quad (5.18)$$

Thus, as $\chi \rightarrow 1$, the oblique shock degenerates back to a normal shock at the exit and as $\chi \rightarrow 0$, the shock is at a 45° angle from the exit plane, but has become so weak that the design condition (which means no shocks) is now met. The flow turn angle through the shock emanating from the nozzle, denoted θ , is then approximated by the θ - β - M relationship.

$$\theta = \tan^{-1} \left(2 \cot \beta \frac{M_e^2 \sin^2 \beta - 1}{M_e^2 (\gamma + \cos 2\beta) + 2} \right) \quad (5.19)$$

Effectively, the flow coming out of the nozzle's exit along the nozzle's surface immediately encounters an oblique shock, demanding a concave (inwards) turn of the flow (towards the nozzle's centerline).

The flow properties (5.9 – 5.11) are then easily obtained upon repeated solving of the Area-Mach relation (5.8), where the properties are evaluated at nondimensional stations x in the (monotonically) diverging section according to (5.5). Obtaining these flow properties may become significant in estimating heat transfer to the surrounding walls.

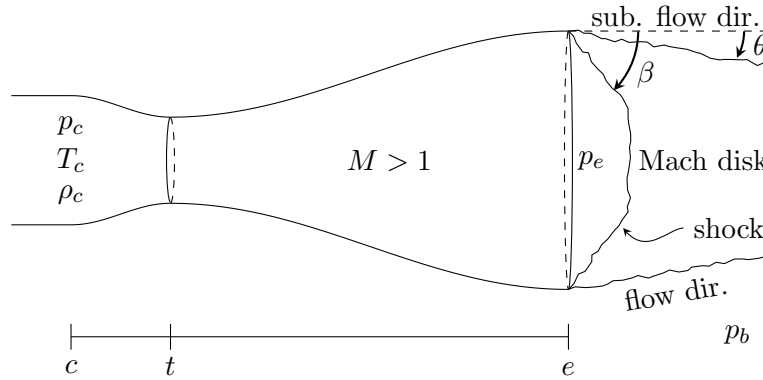


Fig. 5.4: Flow resulting from oblique shocks emanating from the nozzle's edges (resulting from pushing the normal shock out of the nozzle). Shock reflections from the Mach disk and slip line are not shown. The angles β and θ are measured from the nominal (subsonic) flow direction.

Rarefaction Waves

The creation of rarefaction waves under a quasi-unidimensional flow assumption indicates *underexpanded flow* and means that the design condition has been surpassed, and like the overexpanded case, contains choked flow throughout the nozzle. Again, like the overexpanded case, this flow regime is also common on high-powered rockets but occurs at higher elevations and is characteristic of outwards-bowing exhaust plumes.

Because the flow inside the nozzle is choked, the exit pressure is obtained in the same manner as for oblique shocks. The turning angle, however, differs fundamentally. The turning angle of the flow induced by a rarefaction wave is obtained as

$$\theta = \nu(M_{r,d}) - \nu(M_{r,u}), \quad (5.20)$$

where $M_{r,u}$ and $M_{r,d}$ (colloquially called M_1 and M_2 throughout common literature in the context of Prandtl-Meyer expansion fans) are the respective Mach numbers immediately upstream and

downstream of the rarefaction wave and the Prandtl-Meyer function ν is defined for a calorically perfect gas

$$\nu(M) = \sqrt{\frac{\gamma+1}{\gamma-1}} \tan^{-1} \sqrt{\frac{\gamma-1}{\gamma+1}} (M^2 - 1) - \tan^{-1} \sqrt{M^2 - 1}. \quad (5.21)$$

The Mach number upstream of the wave, $M_{r,u} > 1$ satisfies the Area-Mach relation (5.8) and the Mach number downstream of the wave, $M_{r,d} > M_{r,u}$, is obtained in similar fashion to the subsonic case (5.7) since there are no shocks in the nozzle; the result now is simply greater than unity (supersonic) because of the pressure ratio and the fact that rarefaction waves are isentropic. Effectively, the flow coming out of the nozzle's exit along the nozzle's surface immediately encounters a rarefaction wave, demanding a convex (outwards) turn of the flow (away from the nozzle's centerline).

The flow properties throughout the nozzle are obtained in the same way and for the same reasons as for oblique shocks.

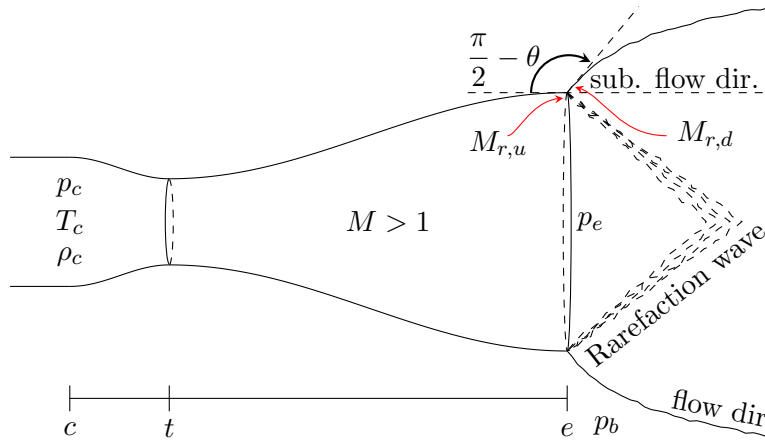


Fig. 5.5: Flow resulting from expansion fans (rarefaction waves) emanating from the nozzle's edges (resulting from pushing the oblique shock past the design condition). Shock reflections from the centerline and slip line are not shown.

5.2 Basic Grain Geometry

Solid propellants are largely characterized by their chemical composition and grain (cross-section) geometry and have the property that they burn in the direction normal (inwards) to the physically exposed surface area. An empirical approximation to the *advance rate* or *rate of regression*, typically denoted r , describing how quickly the exposed surface area is moving normally towards the casing follows Saint Robert's law

$$\tau = a p_c^n. \quad (5.22)$$

Here, a and n are experimentally determined coefficients called the *burn rate coefficient* and *burn rate exponent* (or *pressure exponent*), respectively. In general, the advance rate is a function of time depending on radial symmetry of the grain, but typically does not exceed 2 [inch/s]. Further, the *Web thickness* W is defined

$$W(t) = \int_0^t \tau(s) ds, \quad (5.23)$$

which quantifies the distance the propellant has receded normal to its burning surface along a ray emanating from the centerline to the wall. The Web thickness is defined up to the *burn time* $t_{b,ray}$, the time after which there is no propellant remaining along that ray to produce thrust. That is, the burn time along that ray, $t_{b,ray}$, is the time at which $W(t_{b,ray}) = r$, where r is the inner diameter of the casing as measured from the longitudinal axis. To good approximation for

grain geometries exhibiting fair amounts of radial symmetry (small perturbations from a concentric circle geometry), the individual burn times along each ray from the centerline approximate to each other such that the overall burn time of the motor t_b follows

$$t_b \approx \frac{1}{2\pi r} \oint_r t_{b,ray} ds \approx t_{b,ray} \quad (5.24)$$

Producing nonzero thrust and staging expended casings implies the necessity that the system's center of mass naturally varies in time. The mass of a solid propellant rocket is very typically distributed radially from the longitudinal axis, which holds to good approximation even while the propellant is burning. As such, the center of mass of a solid propellant rocket lies along its longitudinal axis but, in general, varies its longitudinal position as seen from the overall system as propellant/fuel and structure are expelled.

The center of mass and inertia matrix with respect to the center of mass follow the general form (4.3) and (4.4). Supposing a constant density of the propellant (well-mixed) and the burn characteristics are constant along the longitudinal axis, these general relations simplify enough to provide estimates for basic grain geometries.

5.2.1 BATES Grain

The simplest grain geometry to consider is the BATES grain, which is a homogeneous, radial distribution of propellant about the casing's longitudinal axis, about which, by intuition, the center of mass for the propellant remains constant.

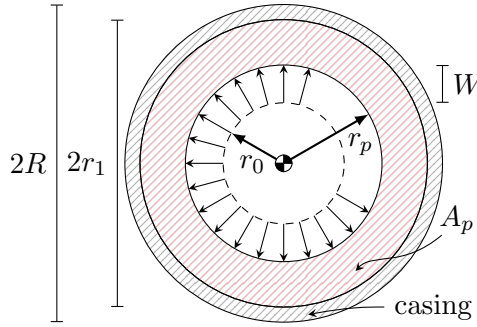


Fig. 5.6: Radially symmetric burn rate of BATES grain leaves the center of mass along the longitudinal axis. The grain borders along, or at least very near, the structural casing for which the inner diameter of the casing is $\sim r$ and the outer diameter is R . The Web thickness here at this instant in time takes on the form $W = r_p - r_0$.

Knowing the burn time t_b but lacking information regarding Saint Robert's law for a BATES grain, the web thickness W might be estimated via a constant advance rate such that

$$W(t) \sim \frac{r_1 - r_0}{t_b} t, \quad (5.25)$$

where all of r_1 , r_0 , and t_b are known constants.

Grain Variation

In accordance with the procedure outlined in 4.1.2 (Body Configurations), the characteristics of the BATES grain (mass, center of mass, and inertia matrix relative to the center of mass) follow (4.33) and (4.34) under the mapping $(R, r) \mapsto (r_1, r_0 + W(t))$, where $0 < r_0 < r_1$ and the Web thickness W is given by (5.25) for $0 \leq t < t_b$. As a result, the mass and inertia matrix of

the propellant varies with time while the rocket is producing nonzero thrust as

$$M(t) = \pi(r_1^2 - r_0^2)L \left(1 - \frac{1}{r_1/r_0 - 1} \frac{t}{t_b} - \frac{t^2}{t_b^2} \right) \rho, \quad \bar{x} = \frac{L}{2} \quad (5.26)$$

$$[I](t) = \frac{M(t)}{4} \left(r_1^2 + \left(r_0 + (r_1 - r_0) \frac{t}{t_b} \right)^2 \right) \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \frac{M(t)L^2}{12} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (5.27)$$

These properties are valid for all times $0 \leq t < t_b$, at which point there is physically no propellant remaining in the casing. Indeed, at time $t = t_b$, the propellant mass and inertia matrix vanish, though the center of mass persists. As there is physically almost no propellant remaining for times $t \geq t_b$, the properties are taken as identically zero.

Taking a time derivative of the propellant's mass reveals information about how the mass forming the solid propellant is moving, but this quantity is *not* necessarily the mass flow rate of the nozzle flow itself due to the possibility of choking at the nozzle (causing gaseous mass to remain in the casing/chamber). While the inertia matrix indeed vanishes at $t = t_b$, its time derivative reveals

$$\frac{d[I]}{dt} = \frac{\pi(r_0 - r_1)r_0^3 L \rho}{6 t_b} \begin{pmatrix} 12\Xi(t)^3 & 0 & 0 \\ 0 & ((L/r_0)^2 + 6\Xi(t)^2)\Xi(t) & 0 \\ 0 & 0 & ((L/r_0)^2 + 6\Xi(t)^2)\Xi(t) \end{pmatrix} \quad (5.28)$$

where the nondimensional quantity Ξ is defined

$$\Xi(t) = 1 - \left(1 - \frac{r_1}{r_0} \right) \frac{t}{t_b}, \quad (0 \leq t \leq t_b). \quad (5.29)$$

Thus, the distribution of mass within a BATES grain changes at a rate cubic in time t relative to the burn time t_b . Most notably, this relation only holds for $0 \leq t \leq t_b$. Otherwise for all $t > t_b$ (all propellant has been ejected), the inertia matrix *and* its derivative both vanish, thus contributing nothing more to the overall rocket system.

6. AERODYNAMICS

A geometry moving through the air at a velocity \vec{v} generates, in general, a nonzero amount of drag \vec{F}_D that *always* acts exactly opposite to the velocity direction and a lift \vec{F}_L that *always* acts perpendicular to the velocity (and therefore the drag) direction. By physical intuition, the resulting flow characteristics are independent of whether the flow is viewed from a stationary reference point or from the view of the body itself. As the interest is often focused on flow characteristics local to the geometry, viewing the flow with respect to the body is often proper.

Aerodynamics and propulsion are similar in the fact that both involve moving gases, but aerodynamics itself is more concerned about finding the lift and drag over a geometry rather than finding the thrust produced from a geometry. That is, the aerodynamics of a vehicle concern air flow about (around, on the outside of) a geometry, whereas propulsion (aerothermodynamics) concern gas flow throughout (contained by, on the inside of) a geometry. Besides these main differences, the fundamental concepts remain the same.

Most generally, the amount of lift and drag generated by a geometry in an airflow are given by

$$F_L = q_\infty S C_L, \quad F_D = q_\infty S C_D, \quad \left(q_\infty = \frac{1}{2} \rho_\infty v_\infty^2 \right) \quad (6.1)$$

where, q_∞ is called the dynamic pressure (dependent upon the freestream air density ρ_∞ and velocity v_∞), S is a fixed, yet arbitrary, reference area, and C_L and C_D are, respectively, the lift coefficient and drag coefficient with respect to the reference area S . These lift and drag coefficients are generally dependent upon everything, especially the flow parameters M (Mach number) and Re (Reynolds number), body geometry, and body orientation with respect to the direction of the freestream α (angle of attack).

6.1 Lift

In the context of rockets, lift is not to be confused with thrust, wherein thrust (which acts, to good approximation, in the direction of the velocity vector \vec{v} , i.e. opposite of drag \vec{D}) actually lifts the rocket vertically upwards, and lift (which acts perpendicular to the velocity vector \vec{v}) is generated by the rocket's fins, effectively providing a means to spin. Thus, the lift produced by a rocket in comparison to nonzero thrust is generally small. Regardless, this lift, acting on each of the fins, produces a torque about the overall center of mass in general, which induces changes in the overall attitude of the vehicle.

The general procedure of finding the lift generated by a three-dimensional fin is outlined in James S. Barrowman's dissertation, *The Practical Calculation of the Aerodynamic Characteristics of Slender Finned Vehicles* (March 1967). The idea is simply to discretize the fin into two-dimensional section to find the (two-dimensional) lift-per-length quantity and then integrate these results along the fin's span to obtain the three-dimensional lift without resorting to CFD.

6.2 Drag

The drag of the vehicle, both in the subsonic and supersonic regimes (with no discontinuity between while traversing the transonic region), may also be implemented from Barrowman's 1967 dissertation.

6.2.1 Flow over a Cone

A more accurate representation of (5.4) is obtained by returning to (5.3) and evaluating the surface integral of pressure over the nose cone as

$$\iint_S p \hat{n} dS \approx p_e A_e \hat{v} - \iint_{S_n} p \hat{n} dS_n, \quad (6.2)$$

where S_n is specifically referencing the surface of the nose cone.

6.2.2 Flow over a Circular Cone at Zero Angle of Attack

If the nose cone's (outer) surface takes the form of Fig. 4.3, the pressure distribution over the cone is completely symmetric (that is, independent of the polar angle ϕ) at zero angle of attack. The net force on the circular cone of constant cone angle θ_c (6.3) therefore takes the form

$$\iint_{S_n} p \hat{n} dS_n = \iint_{S_n} p \sin \theta_c \hat{v} dS_n \quad (6.3)$$

Thus, it remains to determine the pressure distribution on the cone's surface, p .

Steady Supersonic Flow

Within a supersonic flow regime (with an attached shock wave), airflow over a circular cone at zero angle of attack exhibits axisymmetric flow properties such that the velocity field is constant along a ray emanating from the cone's vertex and independent of the polar angle ϕ .

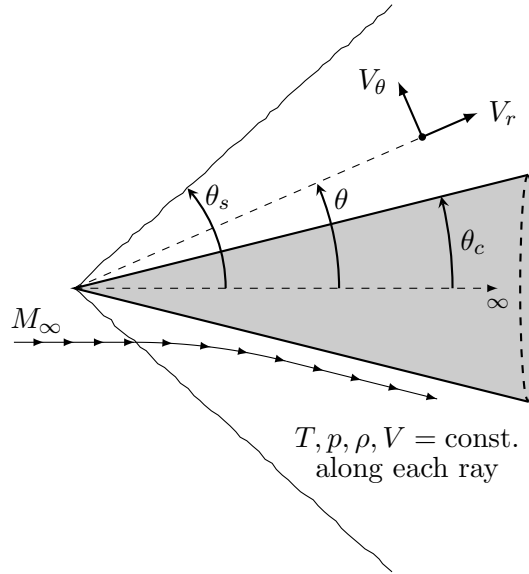


Fig. 6.1: Semi-infinite circular cone with an oblique shock attached at the nose. The incoming flow is supersonic ($M_\infty > 1$) and cone is of a constant half-angle θ_c . The flow field behind the shock is constant along the ray at an angle θ from the tip due to the axisymmetric flow field (circular cone and $\alpha = 0$).

A most general approach to determine the flow properties behind the shock would be to start with the conservation of mass (C.1) in spherical coordinates. In such an axisymmetric case, wherein derivatives of flow properties with respect to the radial and polar coordinates vanish, the conservation law yields

$$-r \frac{1}{\rho} \frac{\partial \rho}{\partial t} = 2V_r + \cot \theta V_\theta + \frac{1}{\rho} \frac{\partial \rho}{\partial \theta} V_\theta + \frac{\partial V_\theta}{\partial \theta}, \quad (6.4)$$

where $V = V_r \hat{e}_r + V_\theta \hat{e}_\theta + 0V_\phi \hat{e}_\phi$ is the velocity of the fluid behind the shockwave expressed under the spherical basis $\{\hat{e}_r, \hat{e}_\theta, \hat{e}_\phi\}$. Now, the change in entropy across the shock for all streamlines is the same since the shock is a straight line and the flow is adiabatic. If the flow is assumed further to be steady, then the change in total enthalpy is nothing throughout the entire flow field. Thus, by Crocco's theorem, the flow is irrotational, which directly implies

$$V_\theta = \frac{\partial V_r}{\partial \theta}. \quad (6.5)$$

Furthermore, the flow is now assumed steady, so all time derivatives vanish. Rewriting (6.4) yields

$$0 = 2V_r + \cot \theta \frac{\partial V_r}{\partial \theta} + \frac{1}{\rho} \frac{\partial \rho}{\partial \theta} \frac{\partial V_r}{\partial \theta} + \frac{\partial^2 V_r}{\partial \theta^2}. \quad (6.6)$$

Since the flow is irrotational, then Euler's equation applies, from which it is easily deduced

$$dp = -\rho V dV = -\frac{1}{2} \rho d(V^2) = -\frac{1}{2} \rho d(V_r^2 + V_\theta^2) \implies \frac{d\rho}{\rho} = \frac{-1}{a^2} (V_r dV_r + V_\theta dV_\theta). \quad (6.7)$$

Using the assumption again that the flow is steady and adiabatic, then a quantity of maximum velocity may be defined $V_m = \text{const.}$ Thus, since the flow has already been assumed isentropic, one may obtain

$$a^2 = \frac{\gamma - 1}{2} (V_m^2 - V_r^2 - V_\theta^2) \implies \frac{1}{\rho} \frac{\partial \rho}{\partial \theta} = \frac{2/(\gamma - 1)}{V_m^2 - V_r^2 - V_\theta^2} \left(V_r \frac{\partial V_r}{\partial \theta} + V_\theta \frac{\partial V_\theta}{\partial \theta} \right). \quad (6.8)$$

Finally, the equations determining the flow field behind the shock of a semi-infinite circular cone in an supersonic, steady, irrotational, adiabatic flow field is determined by the hailed Taylor-Maccoll equations,

$$\frac{\gamma - 1}{2} \left(1 - V_r'^2 - \left(\frac{dV_r'}{d\theta} \right)^2 \right) \left(2V_r' + \cot \theta \frac{dV_r'}{d\theta} + \frac{d^2 V_r'}{d\theta^2} \right) - \frac{dV_r'}{d\theta} \left(V_r' \frac{dV_r'}{d\theta} + \frac{dV_r'}{d\theta} \frac{d^2 V_r'}{d\theta^2} \right) = 0 \quad (6.9)$$

$$V_\theta' - \frac{dV_r'}{d\theta} = 0 \quad (6.10)$$

$$V_\theta'(\theta_c) = 0, \quad (6.11)$$

where $V_r' = V_r/V_m$ and $V_\theta' = V_\theta/V_m$ (that is, $V'^2 = V_r'^2 + V_\theta'^2$) are nondimensional velocities (defined for convenience) such that

$$M^2 = -\frac{2/(\gamma - 1)}{1 - 1/V'^2}. \quad (6.12)$$

Solving this boundary value problem provides information about the shock angle θ_s and nondimensional velocity V' , and thus the flow Mach number everywhere behind the shock wherein the flow is again isentropic (since there are no more shocks upstream). Thus, the isentropic flow relations (5.9), (5.10), and (5.11) may be used directly everywhere behind the shock with this Mach number. Furthermore, the oblique shock relations,

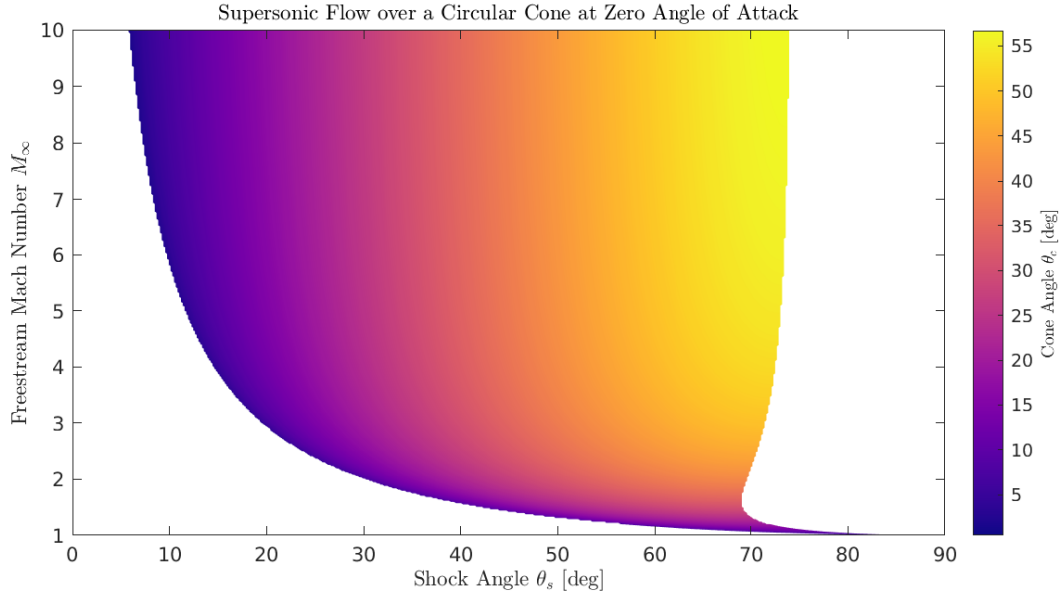
$$\frac{\rho_d}{\rho_\infty} = \frac{1 + ((\gamma - 1)/2) M_\infty^2 \sin^2 \theta_s}{\gamma M_\infty^2 \sin^2 \theta_s - (\gamma - 1)/2} \quad (6.13)$$

$$\frac{p_d}{p_\infty} = 1 + \frac{2\gamma}{\gamma + 1} (M_\infty^2 \sin^2 \theta_s - 1) \quad (6.14)$$

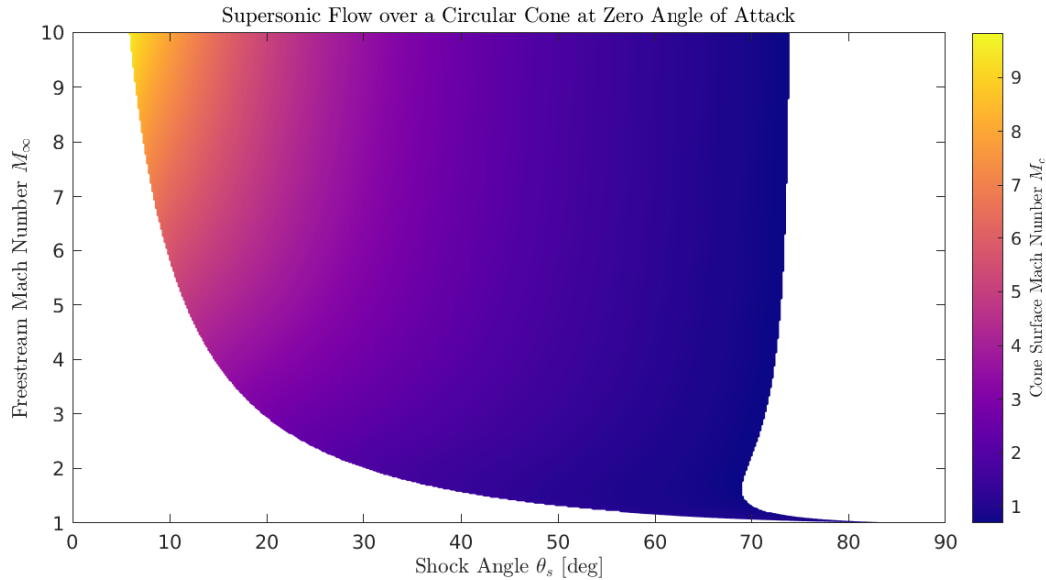
$$\frac{T_d}{T_\infty} = \frac{p_d/p_\infty}{\rho_d/\rho_\infty}, \quad (6.15)$$

hold, where the subscript d indicates conditions immediately downstream of the shock. Thus, the entire flow field behind the shock is known after finding stagnation values behind the shock by, again, using (5.9), (5.10), and (5.11), but now with the Mach number (6.12) corresponding to $\theta = \theta_s$. The easy exception to this procedure is the stagnation temperature T_0 since the flow is assumed adiabatic, $T_0 = \text{const.}$ throughout the entire flow field. That is, T_0 may be obtained from the freestream temperature and Mach number and its value remains unchanged after traversing through the shock.

The results of directly analyzing the Taylor-Maccoll equations to obtain the shock angle and Mach number on the cone's surface are respectively shown in Figs. 6.2a and 6.2b.



(a) QC plot — Cone half-angle θ_c (Res: 500×500).



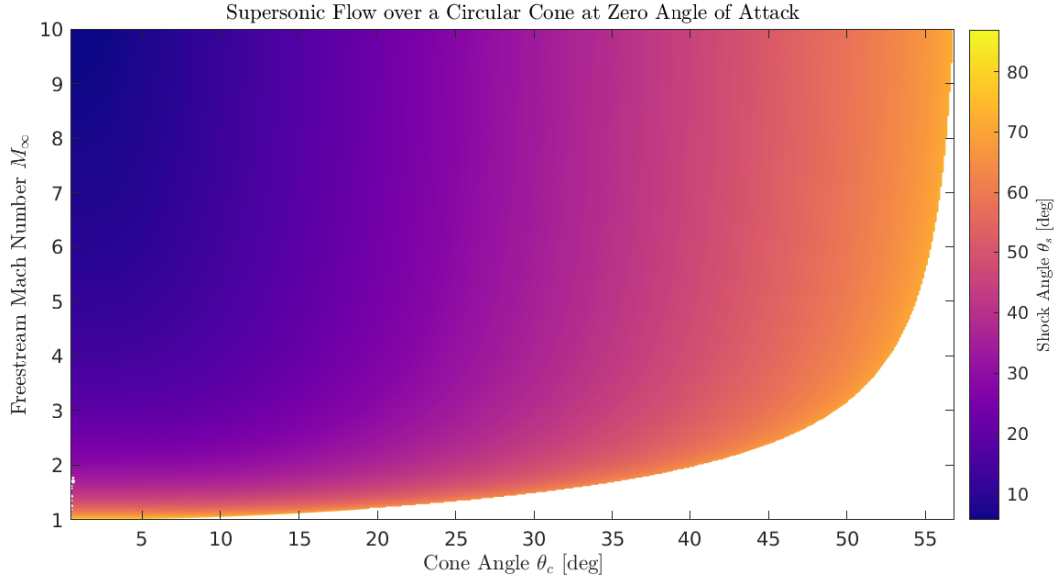
(b) QM plot — Mach number along the cone's surface M_c (Res: 500×500).

Fig. 6.2: Q plots — numerical solutions to the Taylor-Maccoll equations up to Mach 10.

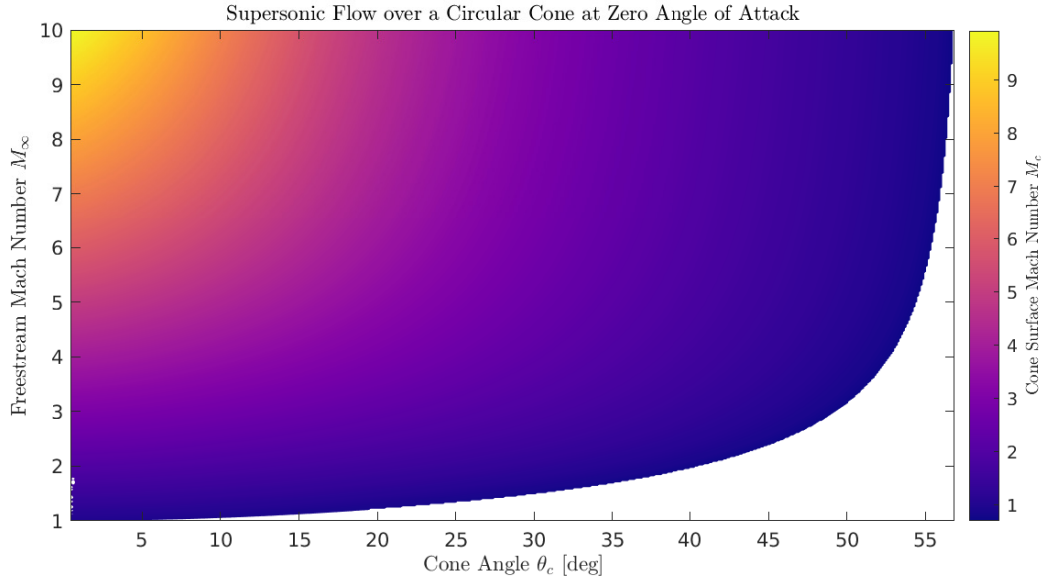
Fig. 6.2 here is said to comprise the Q-plots, or the *inverted* Taylor-Maccoll solutions. The term “inverted” refers to the fact that the solutions are functions of the freestream Mach number M_∞

and the shock angle θ_s .

In the context of rockets, it would be most convenient to swap the dependency on the shock angle for the cone's half-angle (a fixed design parameter). However, numerically manipulating the Q-solutions is an insufficient method since the shock angles are well-ordered, but the cone angles and surface Mach numbers are not. Thus, the Q-plots (specifically, the QC-plot) are used as a means to expedite obtaining the O-plots, or the *righted* (or *regular*) Taylor-Maccoll solutions, shown in Figs. 6.3a and 6.3b.



(a) US plot — Shock angle θ_s (Res: 500×500).



(b) UM plot — Mach number along the cone's surface M_c (Res: 500×500).

Fig. 6.3: U plots — numerical solutions to the Taylor-Maccoll equations up to Mach 10.

The U plots simply swap θ_s for θ_c from Figs. 6.2 in an orderly fashion.

An observation of the QC solution reveals that contours terminate normally on the right-side boundary, indicating that, along this curve, the attached shock detaches and forms a bow shock. Thus, the flow analysis of the Taylor-Maccoll equations concerns the study of *weak* shocks. Such shocks are more commonly encountered than the strong (bow) shocks for circular cones

[Anderson]. Furthermore, cones having any half-angle greater than about 56° will *always* have a strong shock in supersonic flow. The Q plots both exhibit a boundary on the left-side as well which was a numerical tool blocking shocks from surpassing the Mach angle — that is, solutions of the form $\theta_s > \mu$, where μ is the Mach angle defined

$$\mu = \sin^{-1} \frac{1}{M}, \quad (6.16)$$

were not considered.

Note that in the U plots, the right-most boundary of the Q plots persists and is remapped into a new boundary separating the weak shock solutions from the strong ones, but the left-most boundary of the Q plots has been eliminated. Also, some small regions between Mach 1 and 2 for the lowest cone angles has some missing data, likely due to computational error. These few points do not represent unphysical flow and may be safely interpolated (locally) to gain determined values.

Knowing the flow behavior as a function of freestream Mach number M_∞ and cone angle θ_c easily yields the shock angle θ_s , which gives means to evaluating (6.13), (6.14), and (6.15) if wanting to know the entire flow field, but more importantly provides a direct mapping to stagnation pressure everywhere behind the shock by combining (6.14) with (5.10), forming the relation

$$p_{0,d} = p_\infty \left(1 + \frac{2\gamma}{\gamma+1} (M_\infty^2 \sin^2 \theta_s - 1) \right) \left(1 + \frac{\gamma-1}{2} M_d^2 \right)^{\frac{\gamma}{\gamma-1}}, \quad (6.17)$$

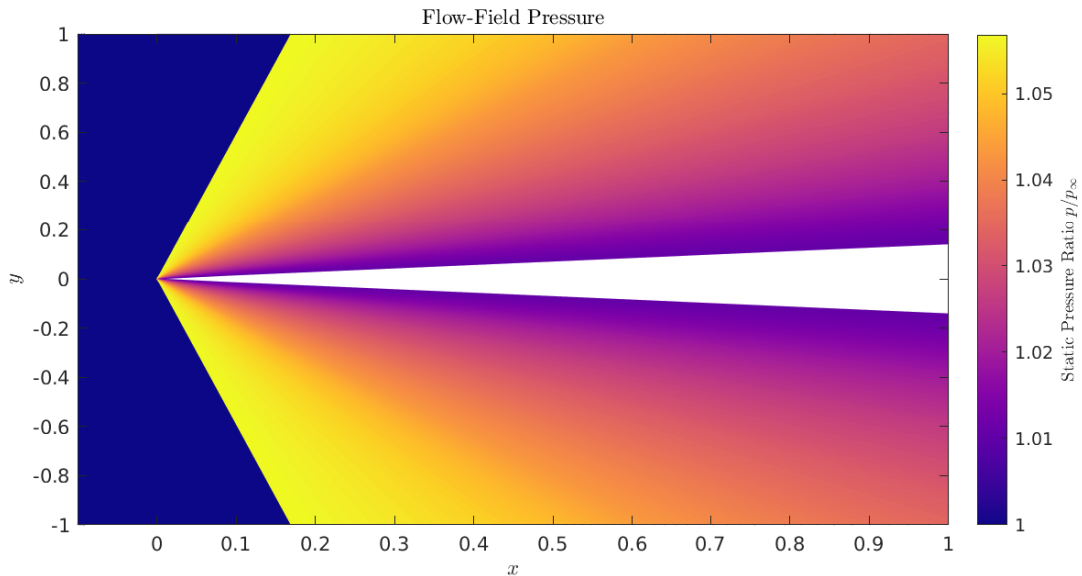
(where M_d^2 is the quantity from (6.12) using $V' = V'(\theta_s)$, i.e. the nondimensional velocity obtained from the Taylor-Maccoll solution immediately behind the shock) such that the pressure behind the shock is given again by (6.14) as

$$p_c = p_{0,d} \left(1 + \frac{\gamma-1}{2} M^2 \right)^{-\frac{\gamma}{\gamma-1}} \quad (6.18)$$

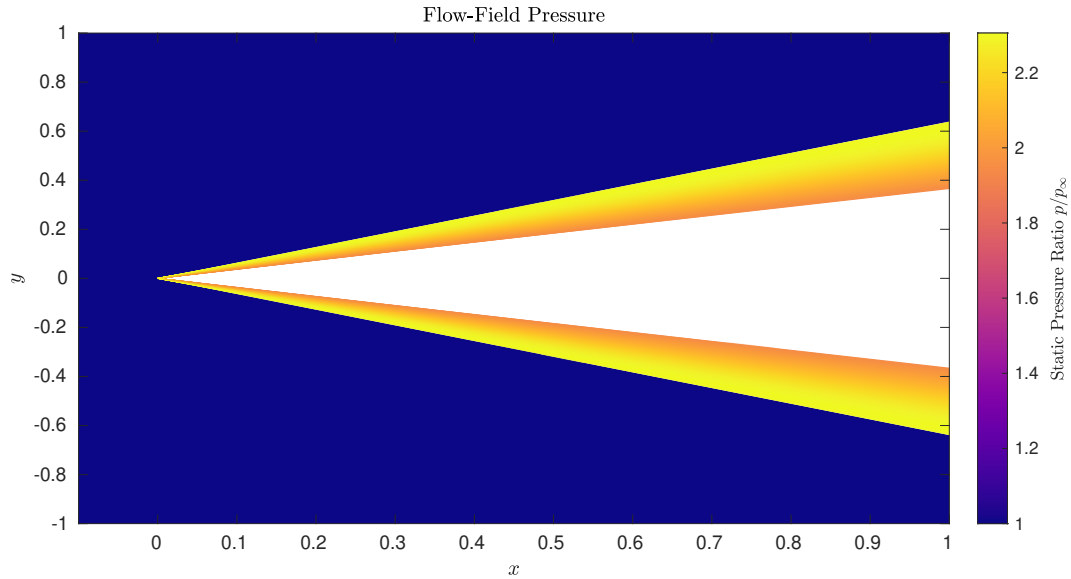
and the Mach number on the surface of the cone as shown in Fig. 6.3b. Thus, the pressure on the cone's surface, p_c is finally known and furthermore, is constant.

In this case, the pressure is constant and evenly distributed over the entirety of the cone's surface.

The Taylor-Maccoll flow field is visualized in terms of pressure for clarity.



(a) Mach 1.02 flow over an 8° circular cone to show the shockwave and flow field pattern.



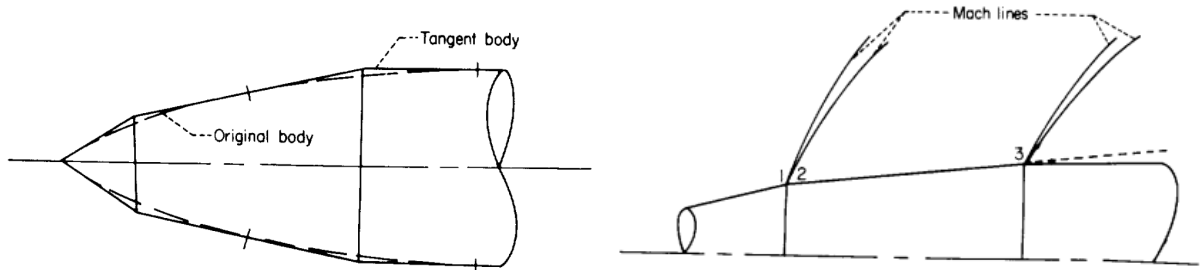
(b) Mach 2.5 flow over a 20° circular cone. The pressure on the cone's surface is nearly double that of the freestream.

Fig. 6.4: Pressure behind the shock relative to the freestream pressure for some sample cones.

The white region is the cone itself (a two-dimensional conical region, not to be confused with a two-dimensional wedge-like region), the uniform region to the left of the stark discontinuity (the shock) is the freestream, and in between is the flow behind the shock. Though shown two-dimensionally, the solution is naturally three-dimensional with rotational symmetry about the x axis. Both x and y are shown without units since the cone is semi-infinite, so the notion of distance here has no meaning. Finally, note that, generally, the shallower the cone for a constant Mach number, the weaker the shock, so the closer the flow properties behind the shock are to the freestream.

6.2.3 Second-Order Shock-Expansion Theory — Relation to Curved Nose Cones

A curved nose cone may be approximated as a circular cone followed by a series of circular frustums such that the nose cone's outline is closely followed. More precisely, a curved nose cone may be approximated by a series of circular cones tangent at various points along the curved body. In this case, the front cone and following frustums are of finite length.



(a) Comparison between the original body and the (b) The tangent body in supersonic flow with Mach waves emanating from its convex corners.

Fig. 6.5: Tangent body approximation to the original curved body using a cone and multiple conical frustums and the depiction of the points (corners) 1, 2, and 3, which lie on the tangent body's surface so that the straightforward Prandtl-Meyer relations may be used around the convex corners. Images from NASA TR1328 (Syvertson and Dennis).

Second-order shock-expansion theory from NASA Technical Report 1328 (*A Second-Order Shock-Expansion Method Applicable to Bodies of Revolution Near Zero Lift*) indicates for such conical flows that the pressure over a tangent conical frustum (for $x_2 \leq x \leq x_3$, where x_2 and x_3 are the locations of (convex) corners between frustums. In such a case, $x_0 = 0$ may be taken as the initial coordinate if the segment is the leading cone) is

$$p_{23}(x) = p_{23,c} - (p_{23,c} - p_2)e^{-\eta_{23}(x)}, \quad (6.19)$$

where $p_{23,c}$ is the pressure on the surface of frustum 2-3's equivalent semi-infinite cone satisfying the Taylor-Maccoll equations, p_2 is the pressure on the surface immediately following the expansion fans (which are isentropic processes, so the stagnation pressure from the initial cone is constant throughout) through the 1-2 corner, and

$$\eta_{23}(x) = \left(\frac{\partial p_{23}}{\partial s} \right)_2 \frac{(x - x_2) \sec \theta_{c,2}}{p_{23,c} - p_2}. \quad (6.20)$$

Here, the pressure gradient along the frustum's surface is given

$$\left(\frac{\partial p_{23}}{\partial s} \right)_2 = \frac{B_2}{r_{23}(x)} \left(\frac{\Omega_1}{\Omega_2} \sin \theta_{c,1} - \sin \theta_{c,2} \right) + \frac{B_2/B_1}{\Omega_2/\Omega_1} \left(\frac{\partial p_{01}}{\partial s} \right)_1, \quad (6.21)$$

where further, for $k = 1, 2$ (integer k in general, for frustums to come),

$$B_k = \frac{\gamma/2}{1 - 1/M_k^2} p_k, \quad \Omega_k = \frac{1}{M_k} \left(\frac{2 + (\gamma - 1)M_k^2}{\gamma + 1} \right)^{\frac{1}{2} \frac{\gamma+1}{\gamma-1}} \quad (6.22)$$

are constants and $r(x)$ is simply the radius (distance) of the frustum from the longitudinal axis x . The value $(\partial p_{01}/\partial s)_1$ is very simply zero for the initial cone (which forms the vertex) since the Taylor-Maccoll pressure and Mach number (flowing over the 0-1 cone) along the surface is constant.

The Mach number M_2 may be evaluating by inverting the Prandtl-Meyer equations, (5.20) and (5.21), since the flow deflection angle ($\theta = \theta_{c,1} - \theta_{c,2}$) and upstream Mach number ($M_{r,u} = M_1$) are known. The relation follows

$$M_2 = \nu^{-1}(\nu(M_1) + \theta). \quad (6.23)$$

This inverse relation is obtained to within 0.01% accuracy of the exact result [Carmichael, Hall] as

$$M_2 = \frac{1 + Ay + By^2 + Cy^3}{1 + Dy + Ey^2} \quad \left(y = \left(\frac{\nu}{\nu_\infty} \right)^{2/3} \text{ and } \nu_\infty = \frac{\pi}{2}(\sqrt{6} - 1) \right), \quad (6.24)$$

where the constants for perfect air in which $\gamma = 7/5$ are

$$A = 1.3604 \quad B = 0.0962 \quad C = -0.5127 \quad D = -0.6722 \quad E = -0.3278. \quad (6.25)$$

In this case, the quantity ν used within y is exactly $\nu = \nu(M_1) + \theta_{c,1} - \theta_{c,2} \geq 0$ since $\nu(M_1) \geq 0$ and $\theta_{c,1} \geq \theta_{c,2}$.

The pressure is now known over the surface of the frustum for which the pressure gradient at x_3 may be obtained by directly differentiating (6.19) with respect to the surface coordinate s , to get

$$\left(\frac{\partial p_{23}}{\partial s} \right)_3 = \frac{p_{23,c} - p_3}{p_{23,c} - p_2} \left(\frac{\partial p_{23}}{\partial s} \right)_2, \quad (6.26)$$

where of course $p_3 = p_{23}(x_3)$.

Indices are simply cycled by two when the pressure over a frustum is found. The next frustum ($x_4 \leq x \leq x_5$, where x_4 is on the other side of the expansion fan starting at x_3 (despite the fact

that these points are actually the same, $x_3 = x_4$) and x_5 is at the end of that next frustum in Fig. 6.5b), the equations for the pressure on the surface become

$$p_{45}(x) = p_{45,c} - (p_{45,c} - p_4)e^{-\eta_{45}(x)}, \quad (6.27)$$

where

$$\eta_{45}(x) = \left(\frac{\partial p_{45}}{\partial s} \right)_4 \frac{(x - x_4) \sec \theta_{c,4}}{p_{45,c} - p_4} \quad (6.28)$$

$$\left(\frac{\partial p_{45}}{\partial s} \right)_4 = \frac{B_4}{r_{45}(x)} \left(\frac{\Omega_3}{\Omega_4} \sin \theta_{c,3} - \sin \theta_{c,4} \right) + \frac{B_4/B_3}{\Omega_4/\Omega_3} \left(\frac{\partial p_{23}}{\partial s} \right)_3, \quad (6.29)$$

and the pressure gradient to be used for the next frustum is

$$\left(\frac{\partial p_{45}}{\partial s} \right)_5 = \frac{p_{45,c} - p_5}{p_{45,c} - p_4} \left(\frac{\partial p_{45}}{\partial s} \right)_4. \quad (6.30)$$

The quantities B_k and Ω_k for $k = 4, 5$ are evaluated accordingly from (6.22). The quantities $\theta_{c,i}$ are suggestively the equivalent semi-infinite cone angles, or simply just the angle that the frustum makes with the longitudinal axis. As such, we have trivially that $\theta_{c,2} = \theta_{c,3}$, and as the corners are assumed convex, then $\theta_{c,1} \neq \theta_{c,2}$.

This process continues repeating until the pressure is known across the surfaces of all frustums, at which point, the pressure on the curved nose cone is approximated by the pressures acting on each cone and frustum of the tangent cone. The final pressure on the tangent cone is

$$p(x) = \bigcup_{i=0,2,4,\dots}^N p_{i(i+1)}, \quad (0 \leq x \leq L) \quad (6.31)$$

where N is the number of tangent sections used in the approximation and $L = x_{2N-1}$ (where x_{2N-1} is the position of the very last corner on the tangent body – where it meets the body) is the overall nose cone length.

6.2.4 Evaluating the Pressure Drag at Zero Angle of Attack

The integral (6.3) representing the magnitude of drag on the nose cone due to pressure acting normally on the surface is finally evaluated knowing p .

$$\iint_{S_n} p \sin \theta_c d\hat{S}_n = \sum_{i=0,2,\dots}^N \iint_{S_{n,i}} p d\hat{S}_{n,i} \quad (6.32)$$

$$= \sum_{i=0,2,\dots}^N \int_0^{2\pi} \int_{r_i}^{r_{i+1}} p(x) \frac{\nabla f}{|\nabla f|} dr d\phi \quad (6.33)$$

$$= \sum_{i=0,2,\dots}^N 2\pi \int_{r_i}^{r_{i+1}} p(r \tan \theta_{c,i}) \sin \theta_c \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}^T dr \quad (6.34)$$

$$= \sum_{i=0,2,\dots}^N 2\pi \hat{v} \int_{r_i}^{r_{i+1}} p(r \tan \theta_{c,i}) \sin \theta_c dr. \quad (6.35)$$

Thus, the force enacted by the pressure on the tangent cone surfaces may readily be carried out by numerical integration. This procedure provides a better estimate of the amount of drag experienced on the rocket's nose, an important quantity when aiming to launch cheaply into the mesosphere or thermosphere, than simply estimating that the pressure acting on the nose is the freestream pressure. In many, if not the vast majority, of cases involving supersonic flow, the nose cone pressure is nowhere near the freestream pressure. Thus, such a treatment with shock-expansion theory seems to be of high importance.

7. PHASES OF FLIGHT

7.1 Prelaunch

A rocket, whether intended for purely atmospheric flight or insertion into orbit, launches from a nearly upright position, held in place only by the *launch tower*, which, depending on the size of the rocket, may or may not be situated on a *launch pad*. For smaller rockets, the launch tower might resemble a (straight) rod (on the order of several meters long) called the *guide rail* along which the rocket slides upon motor ignition.

Before liftoff occurs, however, the rocket is statically situated. As such, the rocket is held in place by the launch pad (or ground) *and* the tower in general. Suppose the rocket is of mass m and the gravity model is g (taken to be (3.10), (3.23), or (3.34)). Further suppose that the rocket rests on the guide rail which is offset from the local direction of gravity by an angle δ and offset from the ENV basis component \hat{E} by the positive (counterclockwise) angle ψ about the \hat{V} element. Then the static forces imposed onto the rail guide are

$$F_P = mg \cos \delta, \quad F_N = mg \sin \delta, \quad (7.1)$$

where F_P is the force statically supporting the rocket directed along the rail direction and F_N is the force statically supporting the rocket normal to the rail direction directed through the rocket's body. These quantities naturally vanish shortly after ignition since the rocket is powerful enough to get itself off of the rail and into free-flight.

The frame created by this 3-rotation from the ENV frame is the *firing frame*, or *local frame*, defined by basis elements $\{\hat{e}_d, \hat{e}_c, \hat{e}_u\}$ which are related to the ENV elements $\{\hat{E}, \hat{N}, \hat{V}\}$ via (A.28). As such, the coordinate variables are related to one another by

$$\begin{pmatrix} x_d \\ x_c \\ x_u \end{pmatrix} = \underbrace{\begin{pmatrix} \cos \psi & \sin \psi & 0 \\ -\sin \psi & \cos \psi & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\mathbf{T}^{\text{loc/env}}} \begin{pmatrix} x_\lambda \\ x_\phi \\ x_h \end{pmatrix}, \quad (7.2)$$

where x_d is the *downrange* position, x_c is the *crossrange* position, and x_u is the usual vertical height (up) as defined by the ENV frame. The *launch*, or *rail*, frame has its coordinate components related to that of the local frame by the elemental rotation of having the launch rail differ from true vertical by the angle δ .

$$\begin{pmatrix} \tilde{x}_d \\ \tilde{x}_c \\ \tilde{x}_u \end{pmatrix} = \underbrace{\begin{pmatrix} \cos \delta & 0 & -\sin \delta \\ 0 & 1 & 0 \\ \sin \delta & 0 & \cos \delta \end{pmatrix}}_{\mathbf{T}^{\text{rail/loc}}} \begin{pmatrix} x_d \\ x_c \\ x_v \end{pmatrix}, \quad (7.3)$$

7.1.1 Initial Conditions

A short discussion of the methods used to determine the resulting trajectory of the flight vehicle is proper. The equations of motion dictating the linear motion of the vehicle's mass center are determined via Newton's laws which necessarily yield three (3) coupled, nonlinear, nonhomogeneous, nonautonomous, *second-order* scalar ordinary differential equations. The rigid-body orientation (rotational motion) of the flight vehicle is deduced through Euler's second

law, which necessarily yields three (3) coupled, nonlinear, nonhomogeneous, autonomous, *first-order* scalar ordinary differential equations. To connect the orientation of the flight vehicle's body frame with the ENV frame (in which the equations of motion are written) simultaneously (without explicitly knowing the orientation), the quaternion kinematic equation is employed which necessarily yields four (4) coupled, nonlinear, homogeneous, autonomous, *first-order* scalar ordinary differential equations.

Put into phase-space form, there are thirteen (13) first-order differential for which there are 13 initial conditions required. These variables requiring initial conditions describe the initial position of the vehicle's mass center (x_0 , y_0 , and z_0), the initial velocity of the vehicle's mass center (\dot{x}_0 , \dot{y}_0 , and \dot{z}_0), the initial angular velocity of the vehicle (ω_{10} , ω_{20} , and ω_{30}), and the quaternions parameterizing the orientation (q_{10} , q_{20} , q_{30} , and q_{40}).

Explicitly, x_0 , y_0 , and z_0 indicate an initial position taken with respect to the ENV frame and \dot{x}_0 , \dot{y}_0 , and \dot{z}_0 describing the initial velocity are also taken with respect to the ENV frame. The quantities ω_{10} , ω_{20} , and ω_{30} describe the initial angular velocity of the rigid body about each body-frame axis (1, 2, and 3 defined with the origin at the mass center) with respect to the rail frame. The quaternion components q_{10} , q_{20} , q_{30} , and q_{40} compose the quaternion q , where (q_1, q_2, q_3) is the vector component and q_4 is the scalar component. The quaternion has the constraint that it must, under all circumstances, be unit normalized.

With a vehicle launching from rest from a launch tower at time $t = 0$, these initial conditions may most easily be expressed as

$$x_0 = 0 \quad \dot{x}_0 = 0 \quad \omega_{10} = 0 \quad q_{10} = 0 \quad (7.4)$$

$$y_0 = 0 \quad \dot{y}_0 = 0 \quad \omega_{20} = 0 \quad q_{20} = 0 \quad (7.5)$$

$$z_0 = 0 \quad \dot{z}_0 = 0 \quad \omega_{30} = 0 \quad q_{30} = 0 \quad (7.6)$$

$$q_{40} = 1. \quad (7.7)$$

A more detailed analysis of the initial position reveals that, given the straight-line distance δ_r between the guide rail's centerline and the rocket's centerline,

$$\begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} = \underbrace{\mathbf{T}^{\text{env/loc}} \mathbf{T}^{\text{loc/rail}}}_{\mathbf{T}^{\text{env/rail}}} \begin{pmatrix} \delta_r \\ 0 \\ L_{cmb} \end{pmatrix} = \begin{pmatrix} (L_{cmb} \sin \delta + \delta_r \cos \delta) \cos \psi \\ (L_{cmb} \sin \delta + \delta_r \cos \delta) \sin \psi \\ L_{cmb} \cos \delta - \delta_r \sin \delta \end{pmatrix}, \quad (7.8)$$

where L_{cmb} is the distance of the vehicle's mass center from the nozzle (b , for bottom). The effect of stating $x_0 = y_0 = z_0 = 0$ is that the rocket is skewered by the launch rail and the rocket's mass center (as opposed to the bottom of the rocket) is placed at its base. Physically, such a simplification amounts to the launch rail being shifted over slightly from its actual location and approximately lengthened by L_{cmb} .

In the sense of numerical stability, it may be necessary to perturb the initial velocity slightly. Such a perturbation is physically insignificant if on very small orders of magnitude and successfully avoids division by zero. Alternatively, any directionality associated with motion along the rail is constant and aligned with the rail itself, which completely averts the problem of division by zero in this case.

7.1.2 Defining the Body Frame

The statement that $q_{40} = 1$ initially realizes that the body axis is perfectly aligned with the rail axis at time $t = 0$. Note that the golden axes outlined in Fig. 4.1 are **not** the body axes. Although the two frames share the same origin (the vehicle's overall center of mass), some axes of the golden frame in Fig. 4.1 must be switched around such that the body-frame's 1, 2, and 3 axes align *exactly* with the rail's 1, 2, and 3 axes at time $t = 0$. Precisely, the axes must undergo

the change of basis

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix}}_{\mathbf{T}^{\text{body/struc}}} \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad (7.9)$$

where (x, y, z) reference the structure coordinates defined in Fig. 4.1 and *not* ECF coordinates here.

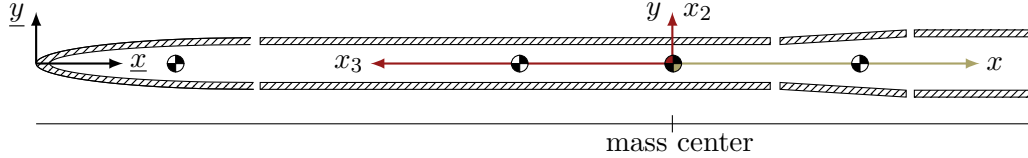


Fig. 7.1: Differentiating the structure-frame (struc) in gold and the body-frame (body) in red. The x_1 direction in the body-frame is directed outwards such that the right-hand rule is satisfied. When on the launch rail, the x_1 direction is pointing downrange as in the local frame except angled downwards by δ .

With this change of basis, the moment of inertia matrix must also be transformed into the body-frame (which is still a principal axis frame) according to

$$[I]^{\text{body}} = \mathbf{T}^{\text{body/struc}} [I]^{\text{struc}} \mathbf{T}^{\text{struc/body}}, \quad (7.10)$$

where $[I]^{\text{struc}}$ is the mass moment of inertia matrix computed as outlined in Ch. 4; its original notation followed (4.11) where it held the symbol $[I]$, but this quantity is *not* the same as the moment of inertia matrix expressed in the body-frame $[I]^{\text{body}}$. The derivative likewise follows suit since this transformation is constant in time.

$$\frac{d[I]^{\text{body}}}{dt} = \mathbf{T}^{\text{body/struc}} \frac{d[I]^{\text{struc}}}{dt} \mathbf{T}^{\text{struc/body}}. \quad (7.11)$$

7.2 Rail-Guided Flight

Supposing it produces a sufficient amount of thrust to overcome its own weight shortly after ignition, the rocket begins traversing up and along the launch/guide rail. There are *two* distinct regions of interest during this brief period of time before the rocket begins true unguided flight. These regions are distinguished by two corresponding periods of time within the overall time between launch and rail clearance.

The first region is defined from ignition until the nose just begins to break through the rail guide's end (i.e. while the entire rocket is fully contained along the rail). The other region follows immediately after the first and is defined until the active nozzle clears the rail (i.e. while the rocket is leaving the rail and, therefore, transitioning from rail-guided flight to free-flight).

7.2.1 Fully Rail-Guided Flight

The static forces imposed on the rail (7.1) are generalizable to include the dynamic effect of thrust.

$$F_P = \max(0, mg \cos \delta - F_T), \quad F_N = mg \sin \delta. \quad (7.12)$$

Here, δ , the angle between the direction of the launch rail and the local vertical direction \hat{V} , is assumed to be small. Thus, the rocket doesn't move if $F_T \leq mg \cos \delta$ by physical intuition since

$F_P \geq 0$ is still supporting some amount of the rocket's weight. Otherwise, the rail supports no weight along the rail direction and the rocket traverses up along the rail. The direction of motion is essentially constant since the vehicle is securely attached to sliding guides along the rail. Thus, any effects of wind are ignored during the phase. The effects of friction during this period of time are also assumed negligible to the amount of thrust and subsequently ignored.

Dynamic Forces

Despite the low speeds involved in this phase, aerodynamic forces begin to develop with the introduction of nonzero velocity due to nonzero thrust. The linear acceleration of the vehicle along the direction of the rail is

$$\vec{F}_{1,1}^{\text{rail}} = \begin{pmatrix} 0 \\ 0 \\ F_T - F_D - mg \cos \delta \end{pmatrix} \quad (7.13)$$

which is expressed with respect to the ENV frame as

$$\vec{F}_{1,1}^{\text{env}} = \mathbf{T}^{\text{env/rail}} \vec{F}_{1,1}^{\text{rail}}. \quad (7.14)$$

The vehicle mass m is inherently is a function of time since propellant is being expelled from the system in the form of exhaust. In the most general manner, the mass is written

$$m = m_0 + \int_0^t \dot{m}(s) ds, \quad (7.15)$$

where m_0 is the initial mass of the vehicle as measured before launch, $\dot{m} \leq 0$ is the mass flow rate (a known function of time), and $0 \leq t \leq t_f$ is the current flight time. During the first stage which, of course, is assumed to be able to clear the rail sufficiently before burning out at time t_b , we can restrict $\dot{m} < 0$ and $0 \leq t \leq t_b$ to reflect the fact that the vehicle is firing, where t_b is the burnout time. If precise knowledge of \dot{m} is unknown while firing, then a basic estimate may be made of linear form

$$\dot{m} \sim \frac{m_f - m_0}{t_b}, \quad (7.16)$$

where the quantity $m_f - m_0$ represents the entire mass of the propellant used during the first stage.

There are no torques acting on the center of mass since the vehicle is constrained to move along the rail. The angular rates and orientation, therefore, remain constant during this phase.

7.2.2 Partially Rail-Guided Flight

An approximate expression for the forces imposed on the guide rails while the rocket is transitioning to free-flight can be written as a generalization upon (7.12) by linearly introducing a scaling as

$$F_P = \max(0, mg \cos \delta - F_T), \quad F_N = \eta mg \sin \delta. \quad (7.17)$$

Here, η is a measure of the length of the rocket still in contact with the rail given by

$$\eta = \min \left(\max \left(0, \frac{L_{\text{tower}} - (r_{\text{cmp}} - L_{\text{cmb}})}{L} \right), 1 \right), \quad (7.18)$$

where L is the length of the entire flight vehicle, L_{cmb} is as previously defined, L_{tower} is the length of the launch tower from base to tip, and r_{cmp} is the position of the rocket's center of mass as measured from a position parallel along the rail. To account for this subtle shift in coordinates

(due to the launch rail being deflected by the angle δ from the vertical direction \hat{V}), $r_{cmp} \geq 0$ is written

$$r_{cmp}^2 = (x_\lambda + \delta_r \sin \text{El})^2 + x_\phi^2 + (x_h - \delta_r \cos \text{El})^2, \quad (7.19)$$

where δ_r is the straight-line distance separating the centerline of the launch rail from the centerline of the rocket and El is the rocket's *elevation* (usually paired with an *azimuth* Az) given

$$\tan \text{Az} = \frac{x_\lambda}{x_\phi} \qquad \sin \text{El} = \frac{x_h}{r_{env}}. \quad (7.20)$$

The quantity $r_{env} \geq 0$ here is simply the distance of the rocket's mass center from the ENV origin.

$$r_{env}^2 = x_\lambda^2 + x_\phi^2 + x_h^2 \quad (7.21)$$

Note that this combination of Az, El, and r_{env} provides a means to measure position much like spherical coordinates, where Az is measured from true north (whose direction is given by \hat{N}) and El is measured from the east–north plane where $\text{El} > 0$ corresponds to a positive height above the ground.

Weathervaning

Though lasting only momentarily, the vehicle's flight speed is relatively slow upon transitioning into and entering free-flight. During this time, the vehicle is susceptible to *weathervaning*, or *weathercocking*. That is, the wind speed is comparable to rocket's velocity during this time which is able to cause a slight alteration in its nominal launching direction. A blowing wind with wind speed w induces an angle of attack α_{wv} on the vehicle given by

$$\tan \alpha_{wv} = \frac{w}{v}, \quad (7.22)$$

where v is the speed of the vehicle (as expressed in the ENV frame, which is consistent with the expression of w). Here, the 2-argument arctangent function is unnecessary; the usual arctangent function may be safely employed.

As a result of the induced angle of attack, an aerodynamic force is generated about the vehicle's center of pressure *in the direction of the wind* which, in general, causes a nonzero torque about the vehicle's center of mass. Consequently, the vehicle turns the angle α_{wv} *into* the direction of the wind and necessarily fails to reach its maximum altitude if launched in a perfectly vertical configuration [NASA GRC].

For much slower flight vehicles, this effect may persist much further into the flight, well beyond the point of exiting the launching platform.

Dynamic Forces

All aerodynamic forces act through the center of pressure, which, in general, differs in location from the center of mass. The location of the vehicle's center of pressure is given, in general, by

$$\underline{x}_{cp} \int_0^L p(\underline{x}) d\underline{x} = \int_0^L p(\underline{x}) \underline{x} d\underline{x}, \quad (7.23)$$

which is trivially solvable for \underline{x}_{cp} .

With part of the vehicle still sliding along the launch rail, rolling motion is prohibited. Small amounts of yawing and pitching, however, are possible.

7.3 Free-Flight

CITATIONS

lots...

Lost the original list....Let's put the links in here for now then. (Original links in order are gone, so starting over, the links to sources should be built up again as closely as possible to originally)

Time References

1

Coordinate System References

Include sources for determining the ellipsoid and geoid here

Used to confirm geodetic coordinate transformation from geocentric spherical

https://en.wikipedia.org/wiki/Reference_ellipsoid

Useful for its graphic in depicting geocentric and geodetic latitudes

https://en.wikipedia.org/wiki/Geodetic_datum#From_ECEF_to_geodetic

Provided confirmation of ECEF to Ellip & back

https://gssc.esa.int/navipedia/index.php/Ellipsoidal_and_Cartesian_Coordinates_Conversion

Not used but provides ICRF rotation matrix

https://gssc.esa.int/navipedia/index.php/ICRF_to_CEP

(*) Provided many details about parameterizing ellipsoids, radii, etc.

<https://en.wikipedia.org/wiki/Ellipsoid>

Confirmed equation of spheroid (not necessary to cite)

<https://en.wikipedia.org/wiki/Spheroid>

(*) Gave coordinate transformation matrix from ECEF to ENV

https://en.wikipedia.org/wiki/Geographic_coordinate_conversion#From_ECEF_to_ENU

(*) Confirmation of ECEF to ENV transformation (basis and coordinates)

https://gssc.esa.int/navipedia/index.php/Transformations_between_ECEF_and_ENU_coordinates

Backup information for the coordinate transformations

USEFUL -----> <http://ads.harvard.edu/books/1989fcm..book/Chapter2.pdf>

(*) Spherical coordinate gradient, Laplacian, basis transformation, ...

math vs physics representation

https://en.wikipedia.org/wiki/Spherical_coordinate_system

(*) Basis transformation is the same as coordinate transformation & gives 2 examples

https://en.wikipedia.org/wiki/Rotation_of_axes

(*) Very useful in telling all the different types of latitudes

<https://en.wikipedia.org/wiki/Latitude>

? Not sure

<https://en.wikipedia.org/wiki/Longitude>

(***) Extremely useful in clearing up a lot of unknown or misunderstood things

file:///C:/Users/Matt/Downloads/NGA.STND.0036_1.0.0_WGS84.pdf

Provided many ellipsoid models

https://en.wikipedia.org/wiki/Earth_ellipsoid

Clarifies between ITRS and ITRF and gives links to other resources

(not really necessary to cite)

https://en.wikipedia.org/wiki/International_Terrestrial_Reference_System_and_Frame

(***) Describes EGM and WMM but also provides links to their coefficients

<https://earth-info.nga.mil/GandG/update/index.php?dir=wgs84&action=wgs84>

(****) EGM2008 WGS84 Version --- Provides geoid undulations and topography displacement coefficients

https://earth-info.nga.mil/GandG/wgs84/gravitymod/egm2008/egm08_wgs84.html

Ellipsoid is not the same as MSL (MSL is the geoid)

<https://gisgeography.com/wgs84-world-geodetic-system/>

Resulting gravity field if Earth was a smooth ellipsoid.

Provided clarity that gamma is on the surface

https://en.wikipedia.org/wiki/Theoretical_gravity

?

https://en.wikipedia.org/wiki/Geodetic_Reference_System_1980

Gives how GRS80 is different from WGS84

https://en.wikipedia.org/wiki/World_Geodetic_System

Gives info about link between EGM08 and WGS84

https://earth-info.nga.mil/GandG/update/index.php?dir=wgs84&action=wgs84#tab_wgs84-srv

Provided more background info on the geoid (Labeled as... Useful --->)

[https://www.ngs.noaa.gov/GEOID/PRESENTATIONS/2007_02_24_CCPS/Roman_A_PLSC2007notes.pdf#:~:text="](https://www.ngs.noaa.gov/GEOID/PRESENTATIONS/2007_02_24_CCPS/Roman_A_PLSC2007notes.pdf#:~:text=)

Says what the azimuth is

(not used. relates only to one of the last equations in ellip. coords.)

<https://en.wikipedia.org/wiki/Azimuth#:~:text=Normal-section%20azimuth%20is%20the%20angle%20>

Verify geodesy textbook (i.e. ensure no typos in book)

https://en.wikipedia.org/wiki/Physical_geodesy#:~:text=According%20to%20the%20famous%20Bruns

?

https://ai-solutions.com/_freeflyeruniversityguide/j2_perturbation.htm

(****) WGS84 def., normal gravity (gives expansion to h^2)

LINKS WGS84 TO GEOID AND EGM (!!)

<https://earth-info.nga.mil/GandG/publications/tr8350.2/wgs84fin.pdf>

(*) Gives official WGS84 normal gravity formula

<https://earth-info.nga.mil/GandG/publications/tr8350.2/tr8350.2-a/Chapter%204.pdf#:~:text=4.>

? - I think just to remind myself that the aerospace toolbox has this

<https://www.mathworks.com/help/aerotbx/ug/gravitysphericalharmonic.html>

(****) EGM2008 documentation (how the coefficients go into the formula)

https://earth-info.nga.mil/GandG/wgs84/gravitymod/egm2008/README_FIRST.pdf

Gravity References

https://earth-info.nga.mil/GandG/wgs84/gravitymod/egm2008/first_release.html

https://en.wikipedia.org/wiki/Theory_of_tides

Atmosphere References

Get gamma values for air from 300K to 1000K

https://www.engineeringtoolbox.com/specific-heat-ratio-d_602.html

(*) Get expressions for imperfect gas w/ simple harmonic oscillator vibration

<https://www.grc.nasa.gov/WWW/BGH/realspec.html>

<https://www.grc.nasa.gov/WWW/BGH/isentrop.html>

Verify $cp - cv$ is constant for ideal gases and $cp > cv$

https://en.wikipedia.org/wiki/Julius_von_Mayer

gamma DoF and also $cp - cv = -T dpdT^2|V / dpdV|T$

https://en.wikipedia.org/wiki/Heat_capacity_ratio

Verify results from above (not necessary to cite this)

https://en.wikipedia.org/wiki/Relations_between_heat_capacities

Real gas dynamics - told me that you can have thermally

perfect but calorically imperfect gas

(i.e. you can obey the ideal gas law but have temperature-dependent gamma)

https://www.aoe.vt.edu/content/dam/aoe_vt_edu/programs/graduate/forms/lectnotes3-09A11101812

Clearly explained thermally perfect vs calorically perfect

<https://www.aithercfcd.com/2017/05/20/thermally-perfect.html>

Lists and gives links to equations of state for imperfect gases

https://en.wikipedia.org/wiki/Real_gas

Investigate the Redlich-Kwong EOS for imperfect gas

(looked easy and reasonably accurate)

https://en.wikipedia.org/wiki/Redlich%E2%80%93Kwong_equation_of_state

Explain activation modes for atoms when exposed to increasing temperature

[https://en.wikipedia.org/wiki/Degrees_of_freedom_\(physics_and_chemistry\)](https://en.wikipedia.org/wiki/Degrees_of_freedom_(physics_and_chemistry))

(*) Heat Transfer Physics textbook (Massoud Kaviany)

Showed me a diagram for around where these modes activate

(for monotomic molecule)

<https://opentextbc.ca/universityphysicsv2openstax/chapter/heat-capacity-and-equipartition-of>

Introduced me to the kinetic theory of gases predicting viscosity

(not necessary to cite)

https://en.wikipedia.org/wiki/Kinetic_theory_of_gases#Viscosity_and_kinetic_momentum

Introduced me to basic shear viscosity (not quite the same as dynamic viscosity, I think ?)

Also introduced me to the notation and helped with symbols

(not necessary to cite)

https://en.wikipedia.org/wiki/Viscosity_models_for_mixtures#Dilute_gas_limit_and_scaled_vari

(*) Hard-spheres, Sutherland's, Lennard-Jones viscosity models

https://en.wikipedia.org/wiki/Temperature_dependence_of_viscosity

Gave me an understanding of air's diameter and some other neat things (not necessary to cite)

<https://personal.ems.psu.edu/~bannon/moledyn.html>

Reinforces that air is strongly dependent only on T if pressure isn't extreme

https://en.wikipedia.org/wiki/Viscosity#Effect_of_temperature_on_the_viscosity_of_a_gas

Sutherland's law for CFD (what I originally used for a year)

https://www.cfd-online.com/Wiki/Sutherland%27s_law

https://en.wikipedia.org/wiki/Gas_constant

https://en.wikipedia.org/wiki/Reference_atmospheric_model

https://en.wikipedia.org/wiki/Jet_standard_atmosphere

https://en.wikipedia.org/wiki/International_Standard_Atmosphere

<https://www.space.com/17683-earth-atmosphere.html>

<https://en.wikipedia.org/wiki/Atmosphere>

https://en.wikipedia.org/wiki/Atmosphere_of_Earth

https://www.engineeringtoolbox.com/individual-universal-gas-constant-d_588.html

https://www.engineeringtoolbox.com/air-composition-d_212.html

https://en.wikipedia.org/wiki/Speed_of_sound

https://en.wikipedia.org/wiki/Heat_capacity_ratio

https://en.wikipedia.org/wiki/Real_gas

<https://scied.ucar.edu/atmosphere-layers#:~:text=Layers%20of%20the%20atmosphere%3A%20troposp>
https://www.nasa.gov/mission_pages/sunearth/science/atmosphere-layers2.html
<https://niwa.co.nz/education-and-training/schools/students/layers#:~:text=The%20atmosphere%2>
https://www.nasa.gov/mission_pages/sunearth/multimedia/magnetosphere.html
<https://maths.ucd.ie/met/msc/fezzik/Phys-Met/Ch03-Slides-2.pdf>

https://en.wikipedia.org/wiki/Upper-atmospheric_models
https://en.wikipedia.org/wiki/Jet_standard_atmosphere
https://en.wikipedia.org/wiki/U.S._Standard_Atmosphere
https://en.wikipedia.org/wiki/International_Standard_Atmosphere
https://en.wikipedia.org/wiki/Earth_radius

https://math.wikia.org/wiki/Ellipsoidal_quadratic_mean_radius
<https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/swe.20064>
<https://en.wikipedia.org/wiki/NRLMSISE-00>
<https://www.mathworks.com/matlabcentral/fileexchange/56253-nrlmsise-00-atmosphere-model>
<https://www.nrl.navy.mil/ssd/branches/7630/modeling-upper-atmosphere>

<http://www.braeunig.us/space/atmmodel.htm>
<http://www.braeunig.us/space/pdf/USSA1976.pdf>
http://www.braeunig.us/space/pdf/Atmosphere_0-90.pdf
http://www.braeunig.us/space/pdf/Atmosphere_90-110.pdf
http://jimhawley.ca/downloads/Ballistics/Formulae_and_code_US_Standard_Atmosphere_1976.pdf

Missing: Layers of the atmosphere (the -spheres), sourcing a standard atmosphere (hydrostatic equation, its pressure solution, T is linear in hgp,

Aerodynamic References

Weathercocking

<https://www.grc.nasa.gov/www/k-12/rocket/rktcock.html>

APPENDIX

A. MATHEMATICAL BACKGROUND OF COORDINATE TRANSFORMATIONS

A.1 Vector Spaces

The term “coordinate system” is taken to reference a right-handed set of orthonormal vectors that form a basis, generally written $\{\hat{e}_1, \hat{e}_2, \hat{e}_3\}$, to represent objects that exist in a vector space. The term “frame of reference” is distinguished from a coordinate system in that the frame of reference properly defines the location of the coordinate system. Forming a basis of a coordinate system and defining its location of course requires a preexisting frame of reference (implying also a known coordinate system) so that basis and position components may reference something known. This existing coordinate system is taken to be the usual Cartesian coordinate system in three-dimensional Euclidean (flat) space with coordinate variables $\mathbf{x} = (x^1, x^2, x^3)$ and the corresponding basis

$$\{\hat{e}_1, \hat{e}_2, \hat{e}_3\} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}. \quad (\text{A.1})$$

These particular coordinate variables and basis elements are typically written (x, y, z) and $\{\hat{i}, \hat{j}, \hat{k}\}$ respectively and are understood to reference only the Cartesian coordinate system unless otherwise specified. Its frame of reference is taken to be the same as the coordinate system itself so that the origin of the coordinate system serves also as the origin of the reference frame.

To begin considering how changes of variables and coordinate systems relate to what is known, a general result from the study of manifolds is that the Euclidean basis $\{\hat{i}, \hat{j}, \hat{k}\}$ can be thought to identify exactly as

$$\{\hat{i}, \hat{j}, \hat{k}\} = \left\{ \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right\}, \quad (\text{A.2})$$

which comes about by thinking of directional derivatives in an abstract sense (Carroll, Fortney). This idea paired with the following result becomes particularly powerful when wanting to transform the components of an object between coordinate systems exhibiting different bases. Suppose $f(\mathbf{x}) : \mathbb{R}^\ell \rightarrow \mathbb{R}^m$ and $g(\mathbf{y}) : \mathbb{R}^m \rightarrow \mathbb{R}^n$ are maps and consider

$$\frac{\partial}{\partial x^i} (g \circ f)^j = \frac{\partial g^j}{\partial y^1} \frac{\partial f^1}{\partial x^i} + \cdots + \frac{\partial g^j}{\partial y^m} \frac{\partial f^m}{\partial x^i} = \sum_{k=1}^m \frac{\partial f^k}{\partial x^i} \frac{\partial g^j}{\partial y^k} \quad (\text{A.3})$$

This composition relates two different spaces, \mathbb{R}^ℓ and \mathbb{R}^m , to each other through the relation between \mathbf{x} and \mathbf{y} and can equivalently be thought of as

$$\frac{\partial}{\partial x^i} = \sum_{k=1}^m \frac{\partial y^k}{\partial x^i} \frac{\partial}{\partial y^k}, \quad (\text{A.4})$$

which is a statement of how the basis elements between the two spaces transform. Particularly of interest is in the case $\ell = m$, when f maps to its own space, so that \mathbf{x} and \mathbf{y} are of the same dimension. In such a case, these bases are related by a square matrix

$$\frac{\partial}{\partial \mathbf{x}} = \frac{\partial \mathbf{y}}{\partial \mathbf{x}} \frac{\partial}{\partial \mathbf{y}}. \quad (\text{A.5})$$

Thus, if given expressions for how coordinate variables relate to those coordinate variables in a new coordinate system, the new basis can be determined in terms of the known basis.

A.2 Coordinate Transformations

General transformations from this Cartesian coordinate system to another may now be established. Suppose U and V are vector spaces in \mathbb{R}^3 and X is an object in $U \cap V$. Let U be the space spanned by a general orthonormal basis $\{\hat{e}_1, \hat{e}_2, \hat{e}_3\}$ with generalized coordinate variables $\mathbf{q} = (q^1, q^2, q^3)$ and V denote the space spanned by $\{\hat{i}, \hat{j}, \hat{k}\}$ with coordinate variables $\mathbf{x} = (x^1, x^2, x^3)$ for consistency. Also, let the origins of both U and V coincide and finally suppose there exists a bijective (smooth) mapping $f : U \rightarrow V$ such that

$$x^i = f(\mathbf{q})^i, \quad (\text{A.6})$$

where $i = 1, 2, 3$ and \mathbf{x} and \mathbf{q} each describe the same object X . Then using the previous result, the basis $\{\hat{e}_1, \hat{e}_2, \hat{e}_3\}$ may be expressed in terms of the known basis elements $\{\hat{i}, \hat{j}, \hat{k}\}$ as

$$e_i = \frac{\partial x^1}{\partial q^i} \hat{i} + \frac{\partial x^2}{\partial q^i} \hat{j} + \frac{\partial x^3}{\partial q^i} \hat{k} \quad (\text{A.7})$$

with normalization following to ensure $|\hat{e}_i| = 1$. Because the object itself exists independently from the bases expressing it, we may freely write

$$x^1 \hat{i} + x^2 \hat{j} + x^3 \hat{k} = q^1 \hat{e}_1 + q^2 \hat{e}_2 + q^3 \hat{e}_3. \quad (\text{A.8})$$

Now, since the basis $\{\hat{i}, \hat{j}, \hat{k}\}$ is orthonormal, then

$$x^1 = (q^1 \hat{e}_1 + q^2 \hat{e}_2 + q^3 \hat{e}_3) \cdot \hat{i} \quad (\text{A.9})$$

$$x^2 = (q^1 \hat{e}_1 + q^2 \hat{e}_2 + q^3 \hat{e}_3) \cdot \hat{j} \quad (\text{A.10})$$

$$x^3 = (q^1 \hat{e}_1 + q^2 \hat{e}_2 + q^3 \hat{e}_3) \cdot \hat{k} \quad (\text{A.11})$$

That is to say that the *components* of the object X expressed in each coordinate system are related by a linear transformation

$$\begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} = \begin{pmatrix} \hat{e}_1 \cdot \hat{i} & \hat{e}_2 \cdot \hat{i} & \hat{e}_3 \cdot \hat{i} \\ \hat{e}_1 \cdot \hat{j} & \hat{e}_2 \cdot \hat{j} & \hat{e}_3 \cdot \hat{j} \\ \hat{e}_1 \cdot \hat{k} & \hat{e}_2 \cdot \hat{k} & \hat{e}_3 \cdot \hat{k} \end{pmatrix} \begin{pmatrix} q^1 \\ q^2 \\ q^3 \end{pmatrix} \quad (\text{A.12})$$

$$= \underbrace{\begin{pmatrix} \frac{\partial q^1 x^1}{\partial q^1} & \frac{\partial q^2 x^1}{\partial q^2} & \frac{\partial q^3 x^1}{\partial q^3} \\ \frac{\partial q^1 x^2}{\partial q^1} & \frac{\partial q^2 x^2}{\partial q^2} & \frac{\partial q^3 x^2}{\partial q^3} \\ \frac{\partial q^1 x^3}{\partial q^1} & \frac{\partial q^2 x^3}{\partial q^2} & \frac{\partial q^3 x^3}{\partial q^3} \end{pmatrix}}_{\mathbf{T}} \begin{pmatrix} q^1 \\ q^2 \\ q^3 \end{pmatrix}. \quad (\text{A.13})$$

$$(\text{A.14})$$

This matrix is called a transformation matrix because it relates the expressions, or components, of a single object X between two coordinate systems which differ from one another in general by an arbitrary amount of rotations about any axes. Furthermore, since the basis $\{\hat{e}_1, \hat{e}_2, \hat{e}_3\}$ is also orthonormal, then the argument may be repeated and we may easily deduce that this matrix is invertible and

$$\mathbf{T}^{-1} = \mathbf{T}^T. \quad (\text{A.15})$$

Matrices of this flavor are called *orthogonal* and belong to the group of three-dimensional, orthogonal matrices, denoted $O(3)$. Members of this group transform the components of an object from one basis to another while preserving the notion of length and angles and hence the notion of an inner product. This preservation, however, is not enough to ensure that the transformation

consists of pure rotations. To remedy this problem, an orthogonal matrix \mathbf{A} , which is a member of $O(3)$, can be immediately classified into two categories by observing

$$\det \mathbf{I} = \det(\mathbf{A}\mathbf{A}^{-1}) = \det \mathbf{A} \det \mathbf{A}^{-1} = \det \mathbf{A} \det \mathbf{A}^T = \det \mathbf{A} \det \mathbf{A} = 1. \quad (\text{A.16})$$

In other words, in addition to the square determinant always being unity, there are two natural behaviors that an orthogonal matrix \mathbf{A} can exhibit, which are either

$$\det \mathbf{A} = -1 \quad \text{or} \quad \det \mathbf{A} = +1. \quad (\text{A.17})$$

Matrices in $O(3)$ with a negative determinant are called *improper* rotations, and, accordingly, those with a positive determinant are called *proper* rotations. Such suggestive names are vindicated by observing, using the same previous idea, for two members \mathbf{A} and \mathbf{B} of $O(3)$,

$$(\mathbf{AB})(\mathbf{AB})^T = \mathbf{ABB}^T\mathbf{A}^T = \mathbf{AA}^T = \mathbf{I}. \quad (\text{A.18})$$

Hence the group $O(3)$ is *closed under multiplication* such that if \mathbf{A} and \mathbf{B} are members of $O(3)$, then so too is \mathbf{AB} . Checking its determinant shows

$$\det(\mathbf{AB}) = \det \mathbf{A} \det \mathbf{B}, \quad (\text{A.19})$$

so this new matrix \mathbf{AB} could be either improper or proper depending on the properness of both \mathbf{A} and \mathbf{B} . The only case, however, in which everything is consistent is that when both \mathbf{A} and \mathbf{B} are proper. Additionally, rotations should be a continuous operation intuitively beginning from the identity matrix, which has positive unit determinant. Thus, actual rotations (as opposed to reflections and of the like) have positive determinant and hence the naming scheme.

The sign of the determinant is really a statement about the consistency of each coordinate system's orientation, or ordering of coordinate axes (otherwise known as handedness). If an orthogonal matrix relates coordinate systems having the same orientation, i.e. both coordinate systems are either left- or right-handed (like \mathbf{T} does by definition) then

$$\det \mathbf{T} = 1 \quad (\text{A.20})$$

and \mathbf{T} is a proper rotation belonging to the *special orthogonal* group, denoted $SO(3)$, which is often called the *rotation group* and is a subset of $O(3)$. Consequently, the multiplication of any two members of $SO(3)$ yields another member of $SO(3)$. This group of three-dimensional, special orthogonal matrices serves as the group from which *all* coordinate system transformations regarding rotations come.

Transformations between reference frames generally involve a coordinate rotation and a coordinate translation. Such a transformation can be expressed as a simple linear transformation like in the case of rotations, but as seen from the perspective of the reference frame. If \mathbf{x} is still the position of an object X and \mathbf{y} is let to be the position of the reference frame with both \mathbf{x} and \mathbf{y} expressed in the basis $\{\hat{i}, \hat{j}, \hat{k}\}$, then the position of the object X with respect to the rotated reference frame of basis $\{\hat{e}_1, \hat{e}_2, \hat{e}_3\}$ and components \mathbf{q} describing the object X is

$$\mathbf{q} = \mathbf{T}^T(\mathbf{x} - \mathbf{y}). \quad (\text{A.21})$$

The effect of taking the difference $\mathbf{x} - \mathbf{y}$ is to produce an exact copy of the original coordinate system, basis included, but with its origin co-located with the reference frame. Then the bases between this translated coordinate system and the reference frame follow as before. The rotation matrix is transposed because the components of X are being expressed in the frame's basis.

These transformation results hold for both rotations and translations that are not constant as no assumptions, apart from existing in three spatial dimensions, have been made on any coordinate variables including time. With time-varying positions of an object X and time-varying rotations between two co-located coordinate systems, the description of the object's

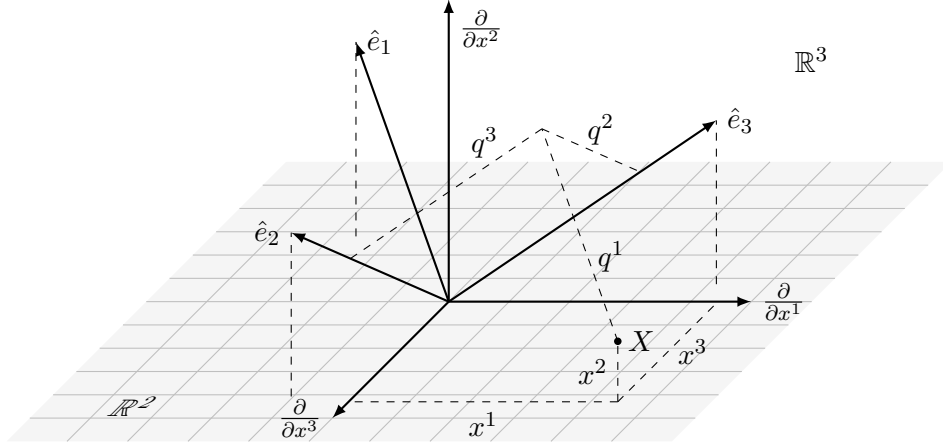


Fig. A.1: An object X is shown to exist in the space occupied by two coordinate systems, which are arbitrarily rotated from one another. Their origins overlap and each coordinate system comprises a set of three orthogonal unit (orthonormal) vectors. Normally, ∂_{x^3} will be shown to be oriented “upwards,” but is shown as such to emphasize that the visualization of the coordinate system leverages nothing in determining basis components.

location necessarily transforms differently than its rates of change. The required transformation, however, is easily obtainable. If $\{\hat{e}_1, \hat{e}_2, \hat{e}_3\}$ is the basis with components \mathbf{q} rotating with respect to the basis $\{\hat{i}, \hat{j}, \hat{k}\}$ with components \mathbf{x} , then $d\mathbf{x} = d(\mathbf{T}\mathbf{q})$ and

$$\mathbf{T}^\top \frac{d\mathbf{x}}{dt} = \frac{d\mathbf{q}}{dt} + \mathbf{T}^\top \frac{d\mathbf{T}}{dt} \mathbf{q} \iff \frac{d\mathbf{q}}{dt} = \mathbf{T}^\top \left(\frac{d\mathbf{x}}{dt} - \frac{d\mathbf{T}}{dt} \mathbf{q} \right). \quad (\text{A.22})$$

With regards to objects and rotations varying in time, the previous two results are instructive in showing that expressing the object X in the generally translating and rotating reference frame as viewed from the coordinate system with basis $\{\hat{i}, \hat{j}, \hat{k}\}$ is equivalent to the statement that the reference frame is actually neither translating nor rotating but that the coordinate system and object X are. Therefore, a reference frame with basis $\{\hat{e}_1, \hat{e}_2, \hat{e}_3\}$ can be thought of as its own coordinate system with basis $\{\hat{i}', \hat{j}', \hat{k}'\}$ so that further transformations are obtainable by induction.

A.3 Notation

A notational standard to be employed throughout is that an object X expressed in a coordinate system Y will be denoted with components X_Y . Furthermore, in another coordinate system Z , whose basis relates to that of Y via a rotation matrix \mathbf{T} , the object is expressed

$$X_Z = \mathbf{T}_Z^Y X_Y. \quad (\text{A.23})$$

The notation used on the rotation matrix is suggestive that the coordinate system to be transformed into is indicated on the bottom index, which is consistent with the notation used with vector components. This notation has the immediate implication that

$$(\mathbf{T}_Z^Y)^\top = \mathbf{T}_Y^Z, \quad (\text{A.24})$$

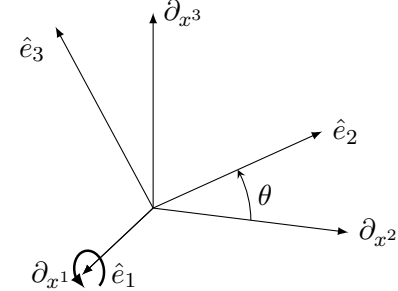
where \mathbf{T}_Y^Z is the rotation matrix describing the basis of Z relative to the basis of Y , which further has the indication that the rows are on the top index and columns on the bottom index. The top and bottom indices may also be thought of being “cancelled” on a paired, yet opposite index as demonstrated by

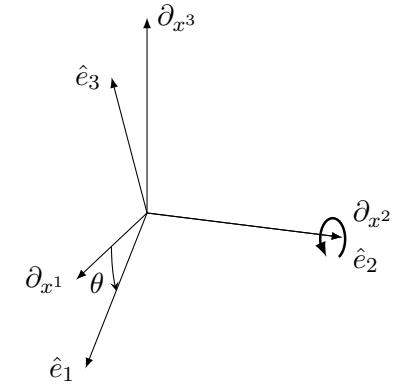
$$\mathbf{T}_Z^Y \mathbf{T}_Y^Z = \mathbf{I}, \quad (\text{A.25})$$

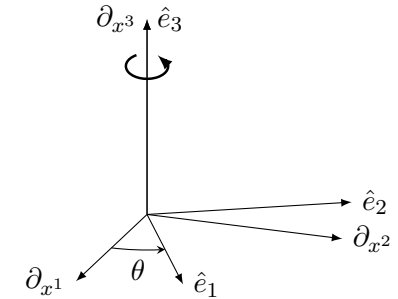
as the left member contains four total indices whereas the right member contains none.

A.4 Elemental Rotations

The core of all pure coordinate rotations consists of three elemental rotations about the usual three Cartesian axes (x^1, x^2, x^3), through positive (counterclockwise) angles $\{\theta_1, \theta_2, \theta_3\}$, though these angles need not be interpreted as respectively. Any arbitrary rotation may be achieved by combining these elemental rotations in nontrivial fasion, i.e. in such a way that permits full degree-of-freedom rotation. Noncommutativity of matrix multiplication demands, however, that in general, different sequences produce different rotations. Thus, the angles θ_1 , θ_2 , and θ_3 are interpreted in the sequential order of the chosen rotation sequence as opposed to being applied about the x^1 , x^2 , and x^3 respectively.

$$\begin{pmatrix} \hat{e}_1 \\ \hat{e}_2 \\ \hat{e}_3 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & \sin \theta \\ 0 & -\sin \theta & \cos \theta \end{pmatrix}}_{\mathbf{R}_1} \begin{pmatrix} \partial_{x^1} \\ \partial_{x^2} \\ \partial_{x^3} \end{pmatrix} \quad (\text{A.26})$$


$$\begin{pmatrix} \hat{e}_1 \\ \hat{e}_2 \\ \hat{e}_3 \end{pmatrix} = \underbrace{\begin{pmatrix} \cos \theta & 0 & -\sin \theta \\ 0 & 1 & 0 \\ \sin \theta & 0 & \cos \theta \end{pmatrix}}_{\mathbf{R}_2} \begin{pmatrix} \partial_{x^1} \\ \partial_{x^2} \\ \partial_{x^3} \end{pmatrix} \quad (\text{A.27})$$


$$\begin{pmatrix} \hat{e}_1 \\ \hat{e}_2 \\ \hat{e}_3 \end{pmatrix} = \underbrace{\begin{pmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\mathbf{R}_3} \begin{pmatrix} \partial_{x^1} \\ \partial_{x^2} \\ \partial_{x^3} \end{pmatrix} \quad (\text{A.28})$$


These rotations are so fundamental that they retain their own notation, denoted by \mathbf{R} and subscripted by 1, 2, or 3 indicating the spinning axis. As such, the clockwise rotation is directly identified with the transpose because of the lacking superscript.

The nontrivial sequences are given by Tab. A.1. Note that the rotation sequences are read forwards, but the resulting transformation reads in the opposite order. For instance, the (symmetric) 1-2-1 rotation sequence follows the rotation

$$\mathbf{R}_{121}(\theta_1, \theta_2, \theta_3) = \mathbf{R}_1(\theta_3)\mathbf{R}_2(\theta_2)\mathbf{R}_1(\theta_1) \quad (\text{A.29})$$

while the (nonsymmetric) 3-1-2 sequence is represented

$$\mathbf{R}_{312}(\theta_1, \theta_2, \theta_3) = \mathbf{R}_2(\theta_3)\mathbf{R}_1(\theta_2)\mathbf{R}_3(\theta_1). \quad (\text{A.30})$$

Tab. A.1: All combinations of fundamental rotations that span \mathbb{R}^3 . The angles $\theta_1, \theta_2, \theta_3$ are called the Euler angles.

$\mathbf{R}_1(\theta_3)\mathbf{R}_2(\theta_2)\mathbf{R}_1(\theta_1)$	$\mathbf{R}_3(\theta_3)\mathbf{R}_2(\theta_2)\mathbf{R}_1(\theta_1)$	$\mathbf{R}_1(\theta_3)\mathbf{R}_3(\theta_2)\mathbf{R}_1(\theta_1)$	$\mathbf{R}_2(\theta_3)\mathbf{R}_3(\theta_2)\mathbf{R}_1(\theta_1)$
$\mathbf{R}_2(\theta_3)\mathbf{R}_1(\theta_2)\mathbf{R}_2(\theta_1)$	$\mathbf{R}_3(\theta_3)\mathbf{R}_1(\theta_2)\mathbf{R}_2(\theta_1)$	$\mathbf{R}_1(\theta_3)\mathbf{R}_3(\theta_2)\mathbf{R}_2(\theta_1)$	$\mathbf{R}_2(\theta_3)\mathbf{R}_3(\theta_2)\mathbf{R}_2(\theta_1)$
$\mathbf{R}_2(\theta_3)\mathbf{R}_1(\theta_2)\mathbf{R}_3(\theta_1)$	$\mathbf{R}_3(\theta_3)\mathbf{R}_1(\theta_2)\mathbf{R}_3(\theta_1)$	$\mathbf{R}_1(\theta_3)\mathbf{R}_2(\theta_2)\mathbf{R}_3(\theta_1)$	$\mathbf{R}_3(\theta_3)\mathbf{R}_2(\theta_2)\mathbf{R}_3(\theta_1)$

Here, each angle $\{\theta_1, \theta_2, \theta_3\}$ is oriented positively with respect to each corresponding spinning axis and every resulting matrix is a member of $\text{SO}(3)$.

B. MATHEMATICAL BACKGROUND OF SPHERICAL HARMONICS

B.1 Function Spaces

Coordinate systems are useful on vector spaces wherein the components of, in general, an n -dimensional object may be described differently depending on the position and orientation of an observer. However, the notion of describing any such object as its dimension approaches infinity begins to break down, so a new type of space to describe such an object is required. This kind of infinite-dimensional object is called a *function*, and the space in which it exists and is described by an infinite-dimensional basis is appropriately called a *function space*.

Particularly (but quite informally), solutions u to the partial differential equation

$$Lu = f, \quad (x \in \Omega) \quad (\text{B.1})$$

where L is a linear differential operator, f is a known function, and Ω is a specified domain in \mathbb{R}^n , are combinable by linearity of the operator and expressible in terms of the basis elements contained in v satisfying

$$Lv = 0, \quad (x \in \mathbb{R}^n). \quad (\text{B.2})$$

This is all (the whole section) very informal — could use some more detail maybe ...? Go over orthogonal functions and what it means for a function to be normal, different function spaces (L^2 , \mathcal{H} , etc., their relations to one another), what it means for functions to serve as a basis (easy example is the polynomials), etc.

B.2 Spherical Harmonics — switch places with Legendre polynomials

Often, this operator takes the form of the Laplacian, denoted Δ or ∇^2 equally, which is one of four fundamental (classical) linear differential operators and thus is of extreme importance. Finding the basis is, therefore, of equally great importance because it has the ability to describe any function over a sphere existing in a Hilbert space.

Laplace's equation in spherical coordinates (r, θ, λ) , where r is the radial coordinate, θ is the polar angle, and λ is the azimuthal angle, reads

$$\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial v}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial v}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 v}{\partial \lambda^2} = 0. \quad (\text{B.3})$$

Supposing the basis element v is to take on the form $R(r)\Theta(\theta)\Lambda(\lambda)$, then the equation is separable by argument that individual contributions to the equation are actually constant. These constants, however, are important and not chosen completely at random. The basis v should be both periodic in λ and bounded at the sphere's poles where $\sin \theta = 0$. Specifically, the constants are required to take the forms

$$\frac{1}{\Lambda} \frac{\partial^2 \Lambda}{\partial \lambda^2} = -\frac{1}{R} \frac{\partial}{\partial r} \left(r^2 \frac{\partial R}{\partial r} \right) \sin^2 \theta - \frac{1}{\Theta} \sin \theta \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial \Theta}{\partial \theta} \right) = -m^2, \quad (m \in \mathbb{Z}) \quad (\text{B.4})$$

and

$$\frac{1}{R} \frac{\partial}{\partial r} \left(r^2 \frac{\partial R}{\partial r} \right) = -\frac{1}{\Theta} \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial \Theta}{\partial \theta} \right) + \frac{m^2}{\sin^2 \theta} = \ell(\ell + 1), \quad (\mathbb{Z} \ni \ell > |m|). \quad (\text{B.5})$$

Thus, the equations to be solved are

$$r^2 \frac{d^2 R}{dr^2} + 2r \frac{dR}{dr} - \ell(\ell + 1)R = 0 \quad (\text{B.6})$$

$$\frac{d^2 \Theta}{d\theta^2} + \cot \theta \frac{d\Theta}{d\theta} + \left(\ell(\ell + 1) - \frac{m^2}{\sin^2 \theta} \right) \Theta = 0 \quad (\text{B.7})$$

$$\frac{d^2 \Lambda}{d\lambda^2} + m^2 \Lambda = 0. \quad (\text{B.8})$$

The first equation, (B.6), is known as *Euler's equation* and easily seen to be solved by supposing $R(r)$ takes on the form r^μ for some number μ parameterized by ℓ . Insertion into (B.6) then gives

$$(\mu - \ell)(\mu + \ell + 1) = 0. \quad (\text{B.9})$$

Thus, the values for which μ may attain such that r^μ satisfies the Euler's equation (B.6) are immediately given.

$$\mu = \ell \quad \text{and} \quad \mu = -\ell - 1. \quad (\text{B.10})$$

Since Euler's equation is linear, the general solution of Euler's equation becomes the superposition of both solutions up to some constants determined by initial values or boundary conditions.

$$R(r) = ar^\ell + br^{-\ell-1}, \quad (\text{B.11})$$

where a and b are real constants.

The easiest equation, (B.8), is a harmonic oscillator with real solution

$$\Lambda(\lambda) = c \cos m\lambda + s \sin m\lambda, \quad (\text{B.12})$$

where c and s are, again, real constants.

The final equation (B.7) is an angular form of the *associated Legendre's equation* seen upon substitution $x = \cos \theta$ (and repeated use of the chain rule), which results in

$$(1 - x^2) \frac{d^2 \Theta}{dx^2} - 2x \frac{d\Theta}{dx} + \left(\ell(\ell + 1) - \frac{m^2}{1 - x^2} \right) \Theta = 0, \quad (-1 \leq x \leq 1). \quad (\text{B.13})$$

The special case of $m = 0$ gives *Legendre's equation* for which the solutions are called *Legendre polynomials*, denoted P_ℓ , defined by Rodrigues' formula

$$P_\ell(x) = \frac{1}{(2\ell)!!} \frac{d^\ell}{dx^\ell} (x^2 - 1)^\ell. \quad (\text{B.14})$$

The operation $!!$ is the double factorial (for which $(2\ell)!! = 2^\ell \ell!$), which is like the factorial, but counts by twos instead of ones, i.e.

$$n!! = n(n - 2)(n - 4) \dots 2 \text{ or } 1. \quad (\text{B.15})$$

For integers $0 < m \leq \ell$, solutions of (B.13) are called the *associated Legendre polynomials*, denoted P_ℓ^m , and related to the Legendre polynomials by

$$P_\ell^m(x) = (-1)^m (1 - x^2)^{m/2} \frac{d^m P_\ell}{dx^m} \quad (-1 \leq x \leq 1) \quad (\text{B.16})$$

The alternating term $(-1)^m$ is called the *Condon-Shortley phase* and is not strictly required to be present in the solution, but proves useful and is often used in the definition of the associated Legendre polynomial in quantum mechanics, and, specifically, MATLAB. Otherwise, the phase is removed by defining

$$P_{\ell m} = (-1)^m P_\ell^m. \quad (\text{B.17})$$

The linearity of Laplace's equation permits superposition of all found solutions R , Θ , and Λ . The complex solution is written as such while absorbing arbitrary constants.

$$v(r, \theta, \lambda) = \sum_{\ell=0}^{\infty} \sum_{m=0}^{\ell} ((a_{\ell} r^{\ell} + b_{\ell} r^{-1-\ell}) \underbrace{P_{\ell m}(\cos \theta)(c_m \cos m\lambda + s_m \sin m\lambda)}_{\text{eigenfunctions}}). \quad (\text{B.18})$$

The basis elements (eigenfunctions) induced by Laplace's equation in spherical coordinates are called (*Laplace*) *spherical harmonics* and are typically written

$$Y_{\ell}^{m,c}(\theta, \lambda) = \sqrt{\frac{2\ell+1}{4\pi} \frac{(\ell-m)!}{(\ell+m)!}} P_{\ell}^m(\cos \theta) \cos m\lambda \quad (\text{B.19})$$

$$Y_{\ell}^{m,s}(\theta, \lambda) = \sqrt{\frac{2\ell+1}{4\pi} \frac{(\ell-m)!}{(\ell+m)!}} P_{\ell}^m(\cos \theta) \sin m\lambda. \quad (\text{B.20})$$

The scaling factors applied to the eigenfunctions are chosen so that what results forms an orthonormal basis. The concept of orthonormality on a function space remains similar to that on a finite-dimensional vector space in that the determining operation is the inner product. The inner product defines both orthogonality and normality which, for functions, is realized through integration. Thus, the inner product of two (Laplace) spherical harmonics follows (by necessity)

$$\int_0^{2\pi} \int_{-1}^1 Y_{\ell}^m(\theta, \lambda) Y_{\ell'}^{m'}(\theta, \lambda) d(\cos \theta) d\lambda = \delta_{\ell\ell'} \delta_{mm'}. \quad (\text{B.21})$$

The result of having a basis act on the surface of a sphere is having the ability to create *any* surface off of the sphere's surface.

Spherical harmonics up to degree and order six are shown below in Fig. B.1 *using the Condon-Shortley phase*. Without the phase, the basis elements of even order m remain unchanged, but the elements with odd order flip around.

B.3 Associated Legendre Polynomials — Switch Places with Sphr Harmonics

The solutions to the associated Legendre equation (B.13) are only truly polynomials if the value m is even, but the name remains regardless. As such, the (fixed) value ℓ is called *the degree* and the (fixed) value m is called *the order* of the associated Legendre polynomial P_{ℓ}^m . The function P_{ℓ}^m is therefore called *the associated Legendre polynomial of degree ℓ and order m* . If m vanishes, then the associated Legendre polynomial degenerates into the Legendre polynomial P_{ℓ} .

$$P_{\ell}^0(x) = P_{\ell}(x). \quad (\text{B.22})$$

The associated Legendre polynomials of equal degree and, separately, equal order, are orthogonal on the closed interval $[-1, 1]$, satisfying

$$\int_{-1}^1 P_{\ell}^m(x) P_n^m(x) dx = \frac{(\ell+m)!}{(\ell-m)!} \frac{2\delta_{\ell n}}{2\ell+1} \quad (\text{B.23})$$

$$\int_{-1}^1 \frac{P_n^{\ell}(x) P_n^m(x)}{1-x^2} dx = \frac{(n+m)!}{(n-m)!} \frac{\delta_{\ell m}}{m}. \quad (\text{B.24})$$

Thus, these quantities, or something similar (depending on the field) may be used to “normalize” the polynomials so that the numerical values remain within the vicinity of unity.

Historically, the notation $P_{\ell m}$ is used to *rid* consideration of the alternating sign based on the order m . Explicitly, this quantity is defined in terms of the associated Legendre polynomial as

$$P_{\ell m}(x) = (-1)^m P_{\ell}^m(x). \quad (\text{B.25})$$

Hence, this form of the polynomial *does not* alternate signs with changing order m .

By necessity of convergence, the associated Legendre polynomials for $m > 0$ are zero on the endpoints

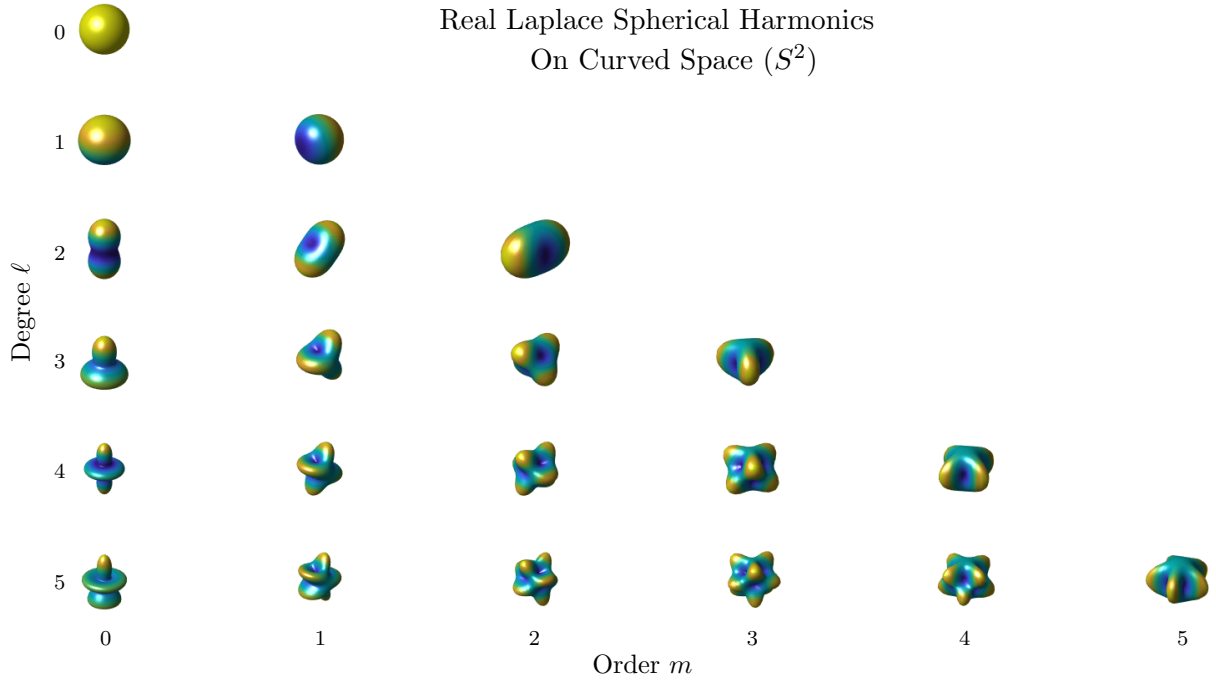
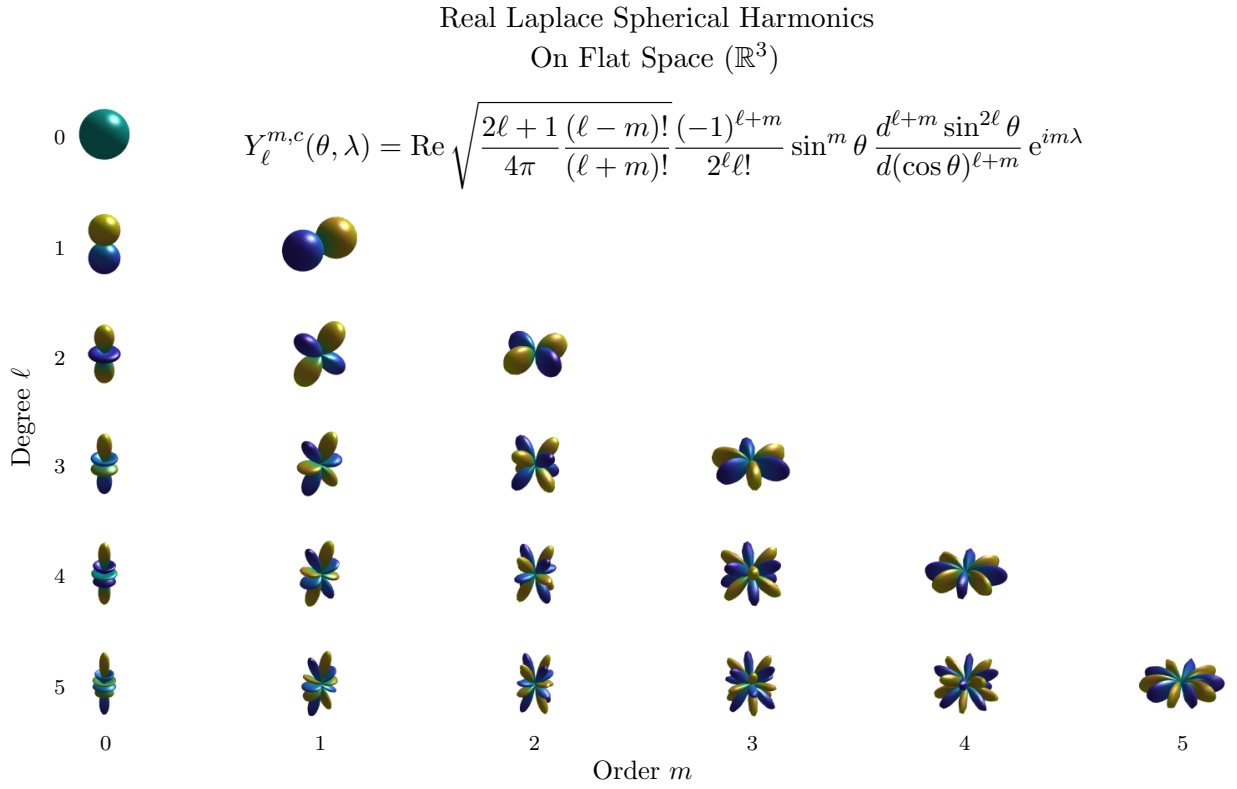


Fig. B.1: Laplace spherical harmonics up to degree and order 5. All orientations of axes are identical, but color representations between any two harmonics are *not* the same.

C. FLUID DYNAMICS (POSSIBLY REMOVE ENTIRELY)

C.1 Conservation Laws

Fluid dynamics is a rich and all-encompassing study of the behavior of liquids and gases, hereby called *fluids*, undergoing dynamic motion. The fundamental equations sourcing the study of fluid dynamics are known as the three *conservation laws*. They state that at all times and under all circumstances, mass, momentum, and energy must be conserved. The laws come in both differential and integral forms with each consisting of five scalar equations. For a fluid of pressure p , density ρ , temperature T , and enthalpy h moving at velocity \vec{v} , the differential forms read

$$0 = \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{v}) \quad (\text{mass}) \quad (\text{C.1})$$

$$\frac{D\vec{v}}{Dt} = \vec{f} - \frac{\nabla p}{\rho} \quad (\text{momentum}) \quad (\text{C.2})$$

$$\rho \frac{Dh}{Dt} = \frac{Dp}{Dt} + \nabla \cdot (k \nabla T) + \Phi. \quad (\text{energy}) \quad (\text{C.3})$$

Here, the *convective derivative* D/Dt is the differential operator involving derivatives in both time *and* space (modulated by velocity components) defined by

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \vec{v} \cdot \nabla \quad (\text{C.4})$$

while the quantity \vec{f} represents viscous forces, k is the fluid's thermal conductivity, and Φ is the viscous dissipation function. In integral form for the same fluid of internal energy e experiencing a heat flux Q ,

$$0 = \frac{\partial}{\partial t} \iiint_V \rho dV + \oint_S \rho \vec{v} \cdot \hat{n} dS \quad (\text{mass}) \quad (\text{C.5})$$

$$0 = \frac{\partial}{\partial t} \iiint_V \rho \vec{v} dV + \oint_S (\rho \vec{v} (\vec{v} \cdot \hat{n}) + p \hat{n}) dS \quad (\text{momentum}) \quad (\text{C.6})$$

$$\frac{\partial Q}{\partial t} = \frac{\partial}{\partial t} \iiint_V \rho \left(e + \frac{v^2}{2} \right) dV + \oint_S \left(\rho \left(e + \frac{v^2}{2} \right) + p \right) (\vec{v} \cdot \hat{n}) dS. \quad (\text{energy}) \quad (\text{C.7})$$

Here, the quantity \hat{n} denotes the outward normal to the surface of the fluid's control volume, which has a volume V and surface area S .

Each form (differential versus integral) states something different about the flow field that the other form struggles to convey. The differential form describes fluid properties locally (pointwise) whereas the integral forms express fluid properties over the entirety of a control surface.

C.2 Similarity Parameters

Dimensional analysis proves its strength through the Buckingham Pi theorem which states that the number of independent variables in a physical system may be reduced according to how many different fundamental physical units are being utilized. That is, given a set of p variables

containing q unique fundamental units, a new set of $p - q$ nondimensional variables may be established.

For gas flows, two of the most important nondimensional values (of which there are many) are the Mach and Reynolds numbers respectively given

$$\text{Ma} = \frac{v}{a} \qquad \text{Re} = \frac{vL}{\nu}, \qquad (\text{C.8})$$

where a is the speed of sound (3.94), L is a characteristic length (typically the smallest such length in a flow field to resolve small effects), and ν is the kinematic viscosity (3.102). The Mach number is often also commonly denoted M .

C.3 Flow Regimes

Different flow regimes may be identified by studying the conservation laws.

C.3.1 Compressibility

A fluid is considered *incompressible* if its density ρ is *constant*. Otherwise, the fluid is said to be *compressible*. In reality, an incompressible fluid is merely a mathematical convenience since the conservation of mass (C.1) is tremendously simplified to

$$\nabla \cdot \vec{v} = 0. \qquad (\text{C.9})$$

In other words, there is no such thing as a truly incompressible fluid, but some fluids indeed behave as one to very good approximation at low flow speeds.

Continuing to increase the flow speed under an incompressible assumption, however, yields results that rapidly degenerate away from results obtained in reality since every fluid is actually compressible. That is, changing the fluid's flow speed changes its flow properties — most intuitively density, but therefore also its pressure and temperature in general. These compressible subsonic effects continue until breaking the sound barrier at which point a *shockwave*, or simply just *shock*, forms. A shock is an extremely thin boundary of built-up fluid particles providing essentially a discontinuity in flow properties ρ , p , and T (as well as total conditions ρ_t and p_t).

The details of the shock are heavily dependent on the surrounding geometry. Typical types of shocks encountered are called *normal*, *oblique*, *weak* and *strong*, *attached* and *bow* (also called *detached*). Shock reflections are also possible. Reference to outside resources regarding the study of compressible flow and shockwaves is heavily recommended for more information.

C.3.2 Rotationality

An *irrotational* flow is one in which the (local) rate of angular velocity is zero everywhere throughout the flow field. That is,

$$\nabla \times \vec{v} = 0, \qquad (\text{C.10})$$

where the left member is called the *vorticity* often denoted $\vec{\Omega}$. Otherwise, the flow is called *rotational*.

As a direct result of (C.10), an irrotational flow field is defined by a scalar potential satisfying

$$\vec{v} = \nabla\psi. \qquad (\text{irrotational}) \qquad (\text{C.11})$$

The resulting velocity field is called *conservative*.

C.3.3 Viscosity

Inviscid flow, by definition, has no viscosity. In such a flow, there is no shearing and so the Reynolds number, which is the ratio of inertial forces to viscous forces, therefore goes to infinity. That is, the higher the Reynolds number, the more the gas behaves like an ideal, frictionless fluid containing no boundary layers.

Viscous flow is much different from the ideal inviscid flow. Because of viscosity alone, the fluid forms a boundary layer along the flow surface and obeys the no-slip condition, i.e. any fluid actually touching the flow surface is stationary.

C.3.4 Turbulence

Refer to outside resources.