NASA CONTRACTOR REPORT

NASA CR-1001

GUIDANCE, FLIGHT MECHANICS AND TRAJECTORY OPTIMIZATION

Volume II - Observation Theory and Sensors

by B. J. Miller and A. S. Abbott

Prepared by
NORTH AMERICAN AVIATION, INC.
Downey, Calif.
for George C. Marshall Space Flight Center

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • MARCH 1968

GUIDANCE, FLIGHT MECHANICS AND TRAJECTORY OPTIMIZATION

Volume II - Observation Theory and Sensors

By B. J. Miller and A. S. Abbott

Issued by Originator as Report No. SID 65-1200-2

Prepared under Contract No. NAS 8-11495 by
NORTH AMERICAN AVIATION, INC.
Downey, Calif.

for George C. Marshall Space Flight Center

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

For sale by the Clearinghouse for Federal Scientific and Technical Information
Springfield, Virginia 22151 – CFSTI price $3.00

FOREWORD

This report was prepared under contract NAS 8-11495 and is one of a series intended to illustrate analytical methods used in the fields of Guidance, Flight Mechanics, and Trajectory Optimization. Derivations, mechanizations and recommended procedures are given. Below is a complete list of the reports in the series.

Volume I	Coordinate Systems and Time Measure
Volume II	Observation Theory and Sensors
Volume III	The Two Body Problem
Volume IV	The Calculus of Variations and Modern Applications
Volume V	State Determination and/or Estimation
Volume VI	The N-Body Problem and Special Perturbation Techniques
Volume VII	The Pontryagin Maximum Principle
Volume VIII	Boost Guidance Equations
Volume IX	General Perturbations Theory
Volume X	Dynamic Programming
Volume XI	Guidance Equations for Orbital Operations
Volume XII	Relative Motion, Guidance Equations for Terminal Rendezvous
Volume XIII	Numerical Optimization Methods
Volume XIV	Entry Guidance Equations
Volume XV	Application of Optimization Techniques
Volume XVI	Mission Constraints and Trajectory Interfaces
Volume XVII	Guidance System Performance Analysis

The work was conducted under the direction of C. D. Baker, J. W. Winch, and D. P. Chandler, Aero-Astro Dynamics Laboratory, George C. Marshall Space Flight Center. The North American program was conducted under the direction of H. A. McCarty and G. E. Townsend.

TABLE OF CONTENTS

FIGURES

Nomenclature List

A	azimuth angle
B	brightness
B_n	noise bandwidth
D	planet diameter
$\underline{e}_x$, $\underline{e}_y$, e_z	unit vectors of a cartesian coordinate system
E	elevation angle
$\underline{E}_1$, $\underline{E}_2$	vectors from planet center to occultation point
f	frequency
f_d	Doppler frequency
F	noise figure
g	gravitational field strength
G	gain
$\underline{H}$	angular momentum
$\underline{i}$, $\underline{j}$, $\underline{k}$	unit vectors of a rectangular cartesian coordinate system
$\underline{\underline{I}}$	inertia matrix
K	Boltzmann's constant
$\underline{m}$	unit vector
n	refractive index
$\underline{n}_i$	unit vector (defined in text)
N	noise power, N units of refractivity
p, q, r	gimbal angular velocities
$\underline{P}$	vector from planet center to a landmark
$\underline{R}$	position vector
s	Laplace transform operator
S	flux density of a radiation source
T_B	brightness temperature
T_e	effective temperature
$\underline{V}$	inertial velocity with respect to some reference
$\underline{V}_p$	velocity of center of planet with respect to some reference
$\underline{V}_r$	radar target velocity with respect to some radar station
$\underline{V}_R$	relative velocity of a vehicle with respect to some planet velocity
α	effective time duration of a radar signal
β	effective beam area of antenna
γ	effective aperture width of an antenna
δ	deviation symbol, measurement uncertainty symbol
θ^*	angular error due to planet horizon diffraction
λ	wavelength
μ	gravitational constant; micron
ϕ	phase; transformation rotation angle

Special Notation and Conventions

When several coordinate systems are being used at one time in the analysis of a problem, such as in the inertial theory section of this monograph, it is convenient to introduce a special notation in order to prevent confusion as to which system is being used to express a particular vector. In the notation adopted in this monograph, each coordinate system is assigned a number; such as 1, 2, 3... . When a vector or a component of a vector is written, a left superscript appears before the symbol to designate the coordinate system in which it is expressed. For example, if a vector $\underline{A}$ is expressed in terms of some coordinate system named "2", it is expressed as

$$ {}^{2}\underline{A} = {}^{2}x\,{}^{2}\underline{e}_{x} + {}^{2}y\,{}^{2}\underline{e}_{y} + {}^{2}z\,{}^{2}\underline{e}_{z} $$

where ${}^{2}\underline{e}_{x}$, ${}^{2}\underline{e}_{y}$ and ${}^{2}\underline{e}_{z}$ are unit vectors of coordinate system 2.

A transformation between one coordinate system and another is represented by T_{abcd} where ab represents the number of new coordinate system and cd represents the number of the old coordinate system defined by the transformation. For example, the transformation that takes a vector expressed in components of coordinate system 11 to a vector expressed in terms of coordinate system 07 is written as follows:

$$ \begin{bmatrix} {}^{7}x \\ {}^{7}y \\ {}^{7}z \end{bmatrix} = T_{0711} \begin{bmatrix} {}^{11}x \\ {}^{11}y \\ {}^{11}z \end{bmatrix} $$

Note that the matrix of the transformation is merely the direction cosine matrix of the new coordinate system with respect to the old system.

1.0 STATEMENT OF THE PROBLEM

The purpose of this monograph is to present a comprehensive technical discussion of observation theory and sensors applicable to the navigation of boost and space vehicles.

Navigation measurements, which are the practical consequences of implementation of observation theory and sensors, involve complex physical phenomena. It is thus a necessity that the technical discussion of observation theory and sensors in this monograph requires the crossing of the boundaries of a number of technical disciplines. The major technical disciplines involved are physical optics, geometrical optics, electromagnetic theory, classical mechanics, geophysics, noise theory and servo theory. The obvious impossibility of providing a thorough discussion in each of the technical disciplines involved has forced a critical selection of material for inclusion in the monograph. This selection has required the enunciation of the following specific objectives:

(1) To develop quantitative descriptions of the physical processes associated with quantitative navigation measurements; i.e., to develop a thorough description of the physics of the measurement processes.

(2) To develop expressions for evaluation of the biases and random errors which affect the accuracies of the navigation measurement processes.

(3) To develop the relationships between navigation observables (measurements) and the vehicle state vector (position and velocity).

(4) To present the solution of the navigation problem, i.e., the determination of past, present, and future values of the vehicle state vector from a set of measurements, for the simple case in which no redundant measurement information is present.

(5) To discuss vehicle constraints imposed by sensor requirements, and criteria for selecting observables to be measured.

The general objective, of course, is the production of a self-contained, reference document for formulating, analyzing, and solving the practical problems of navigation measurements for boost and space vehicles.

2.0 STATE OF THE ART

The state of the art in observation theory and sensors is defined here to be composed of two kinds of information: (1) the collection of all known physical laws and empirically derived data associated with navigation measurements, and (2) the collection of all known techniques of analysis which may be operationally applied to the physical laws and empirically derived data to generate analytical interrelationships between physical quantities associated with navigation measurements.

The problem of selection and organization of the material within this section has been difficult. It is an obvious impossibility to present in detail the extensive collections of physical laws, physical data, and analytical techniques associated with navigation measurements within the one or two hundred pages allocated to this discussion.

The material which has been developed and selected for presentation here is the result of seeking the best compromise between completeness of coverage and the conciseness of the important derivations. The emphasis in the presentation is upon the interconnection of physical theory with the navigation process. Although no "beyond the state-of-the-art" navigation techniques are discussed, the reader is, by way of example, provided with several modes for developing the navigation observation theory from any set of physical laws through the application of analytical techniques.

In Section 2.1 which follows, a structure of the physical laws, physical data, and techniques of analysis is presented which serves the dual purpose of providing (1) a general framework into which the laws, data, and techniques associated with navigation measurements may be fitted; and (2) a starting point for the orderly sifting of the huge bulk of subject matter related to navigation measurements, so that in the succeeding sections the technical discussions may be concentrated upon the really significant areas of observation theory and sensors.

2.1 CLASSIFICATION OF NAVIGATION OBSERVATIONS AND MEASUREMENTS

Navigation observations and measurements are those parameters which are quantitatively related to the vehicle state vector, i.e., position and velocity. Hence, an operational test as to whether any observation or measurement is potentially a "navigation observation or measurement" is to determine whether the value of the measurement in question varies with a variation in the vehicle state vector. If the value of the measurement varies with state vector variation, the measurement in question is a potential navigation observation or measurement. As an example, the measurement of the angles between two stars in a vehicle in interplanetary space is not a navigation observation, since for practical purposes there are no variations of this angle with variations of spacecraft position and velocity (aberration measurements are considered beyond the state of the art).

This definition of a navigation observation is closely related to the concept of "observability" in linear theory. If an equation may be written of the form:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_j \\ y_m \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \cdots a_{1k} \cdots a_{1n} \\ a_{21} & a_{22} \cdots a_{2k} \cdots a_{2n} \\ a_{j1} & a_{j2} \cdots a_{jk} \cdots a_{jn} \\ a_{m1} & a_{m2} \cdots a_{mk} \cdots a_{mn} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_k \\ X_n \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

where $Y_1, \ldots Y_m$ are measured quantities and where $X_1, \ldots X_n$ are components of the state vector, and where $b_1, \ldots, b_m$ are some known quantities, then a component of the state vector X_k is said to be unobservable if all elements of the kth column of the $\{a\}$ matrix are zero, i.e., if $a_{1k} = a_{2k} = \ldots = 0$. More generally, if a linear transformation of the n component x state vector is made to another n state vector Z such that

$$\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} c_{11} \cdots c_{1n} \\ \vdots \\ c_{m1} \cdots c_{mn} \end{bmatrix} \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} + \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

then the "system is not completely observable" if any transformation can be found which converts the kth column to a zero column in the $\{c\}$ matrix. The simplicity of this definition of observability is apparent if the matrix equation is expanded to a set of simultaneous equations, thereby indicating that the values of the measurements $\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$ are unaffected by the kth component of the Z state vector, z_k . Since the same

measurements result for an infinity of different values for z_k, it is obvious that no information concerning the value of z_k is contained in the measurements themselves, and the statement that "z_k is unobservable" is reasonable.

Physical measurements themselves must consist of one or a combination of measurements on particles, fields, or waves. Although a detailed examination of these three categories will yield numerous possible candidate "navigation measurements", they may be more simply classified for the present as follows:

Frequently Used Navigation Measurements:

1. Angular measurements of solar and celestial bodies, including stadiametric and diameter measurements.

2. Radar range and range-rate measurements.

3. Infrared and other electromagnetic intensity measurements to yield directional and range information, including sun and horizon sensors and star trackers.

4. Inertial measurements using gyroscopes and accelerometers.

5. Radio directional measurements.

Infrequently Used and Potential Navigation Measurements:

Any physical quantity which is a function of vehicle position or velocity may be a possible source of navigation information. Such quantities are:

1. Gravitational, magnetic and electric field measurements, magnitude and/or direction.

2. Radioactivity measurements.

3. Pressure measurements in atmospheric flight.

4. Angle-of-attack measurements for atmospheric flight.

The "infrequently used and/or potential" category of navigation measurements is impractical to treat in this monograph for the following reasons:

1. Gravitational field measurements are currently measurable via gravity gradient techniques; the accuracy required is currently considered beyond the state of the art for the flight phases of interest.

2. Electric and magnetic field measurements and radioactivity measurements are time variant, and greatly affected by solar radiation. Use of these measurements for navigation is currently considered beyond the state of the art.

3. Pressure measurements in atmospheric flight are common, but are rarely used for navigation purposes in the vehicles and for the flight phases of interest.

4. Angle-of-attack measurements in atmospheric flight are common, but are usually used within the vehicle control loop, and are rarely used for navigation purposes in the vehicles and for the flight phases of interest.

The phenomena which appear in the frequently used measurement lists above are seen to involve two major sensor areas:

1. Sensors which operate upon electromagnetic radiation, which includes telescopes, sextants, radars, infrared devices, laser devices, and radio directional devices.

2. Sensors which operate upon inertial principles, namely gyroscopes, accelerometers, and inertial measuring units.

The two areas of discussion which follow, namely, Radiation Theory and Sensors, and Inertial Theory and Sensors, have been selected because these areas appear to include all of the frequently used navigation measurements which are currently within the state of the art.

2.2 RADIATION THEORY AND SENSORS

As discussed previously, many of the phenomena associated with radiation measurements are essentially electromagnetic radiation phenomena. The sensors which are used for these radiation measurements include optical devices, infra-red devices, radar devices, etc. Each of these sensors is associated with a number of complex electromagnetic phenomena. The purpose of this section is to present a review of the physical phenomena which are involved in the generation of a measured quantity by the use of these sensors.

In line with one of the major objectives of the monograph series, to present a unified treatment of the formulations and techniques of analysis, the general theory describing electromagnetic radiation phenomena of interest is presented first, followed by a more specific discussion of the sensing techniques themselves.

2.2.1 General Theory

The general theory presented in this section is concerned with a description of the phenomena associated with current state-of-the-art navigation sensors. Hence, included in the following is a discussion of

diffraction, noise, and atmospheric phenomena such as penetration, absorption, and refraction.

The general theory developed herein is for two purposes: Firstly, to provide an introduction to the basic physics of the phenomena which play a role in the navigation measurement process; secondly, to provide the basic quantitative general equations which are used in later, more detailed discussions of the sensors and their error characteristics. Since all forms of electromagnetic radiation are different only in as much as their frequencies differ, the description of the electromagnetic spectrum in the following section provides a convenient and logical starting point for the discussion.

Acknowledgment is given to Space Technology Laboratories whose study in atmospheric refraction (Reference 2) was quite useful in the preparation of this monograph, and to Merrill A. Skolnik (Reference 3) for the material on noise and radio techniques.

2.2.1.1 The Electromagnetic Spectrum

Those frequencies which are useful for observations are usually the frequencies that have line-of-sight propagation. Although there is not a sharp line that divides the propagation nature of waves of different frequencies, 30 MC is generally considered the lowest frequency for line-of-sight propagation (space wave) of any useful degree. Frequencies lower than 30 MC are by no means useless for observation measurements, although the nature of their propagation restricts their use to navigational techniques that are used on, or very near, the earth's surface. Waves in this frequency range propagate by ground and sky waves and, as a result, cannot be used outside the Kennelly-Heaviside layers. Any electromagnetic waves lower than 30 MC suffer from a large attenuation in the space wave component and become progressively more difficult to use for line-of-sight purposes as the frequency is decreased.

2.2.1.2 Atmospheric Penetration and Absorption

Atmospheric penetration and absorption phenomena are important in navigation measurements for two reasons: because all radiation measurements, made from space, of the earth are subject to re-radiation produced by the atmospheric absorption of solar energy; and because there exist "windows," or penetration bands, of the spectrum which enable the observation and communication between a point in space from the earth. The radio "window" enables the use of radio waves to track and communicate with vehicles outside the atmosphere. The optical "window" enables infrared satellite tracking, laser communications in space, laser tracking, and the normal visible phenomenon.

Acknowledgment is given to G. P. Kuiper (Reference 1) for the detailed information on the absorption spectrum in the following discussion.

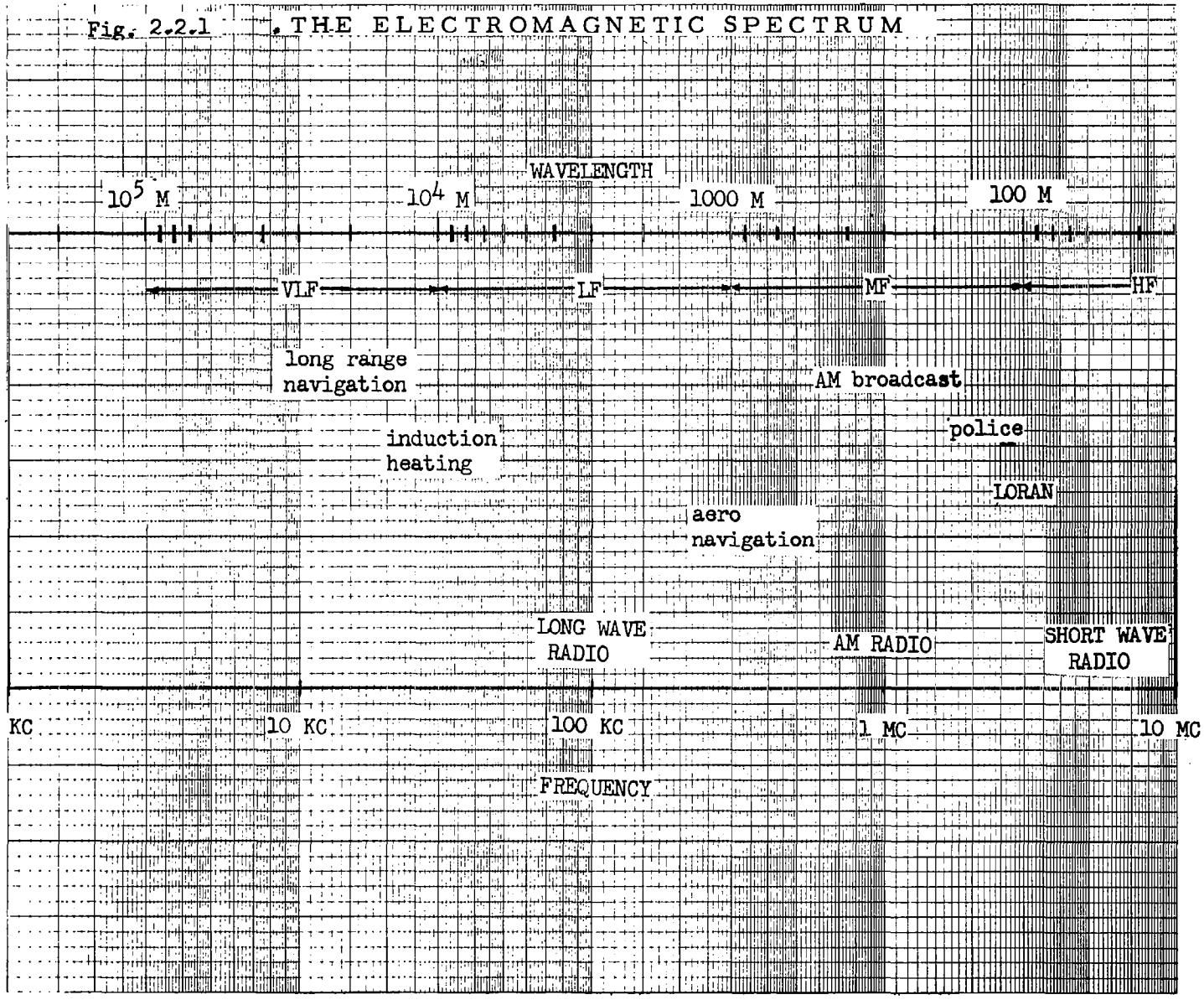


Fig. 2.2.1 THE ELECTROMAGNETIC SPECTRUM
WAVELENGTH
10^5 M
10^4 M
1000 M
100 M
VLF
LF
MF
HF
long range navigation
induction heating
AM broadcast
police
LORAN
aero navigation
LONG WAVE RADIO
AM RADIO
SHORT WAVE RADIO
1 KC
10 KC
100 KC
1 MC
10 MC
FREQUENCY

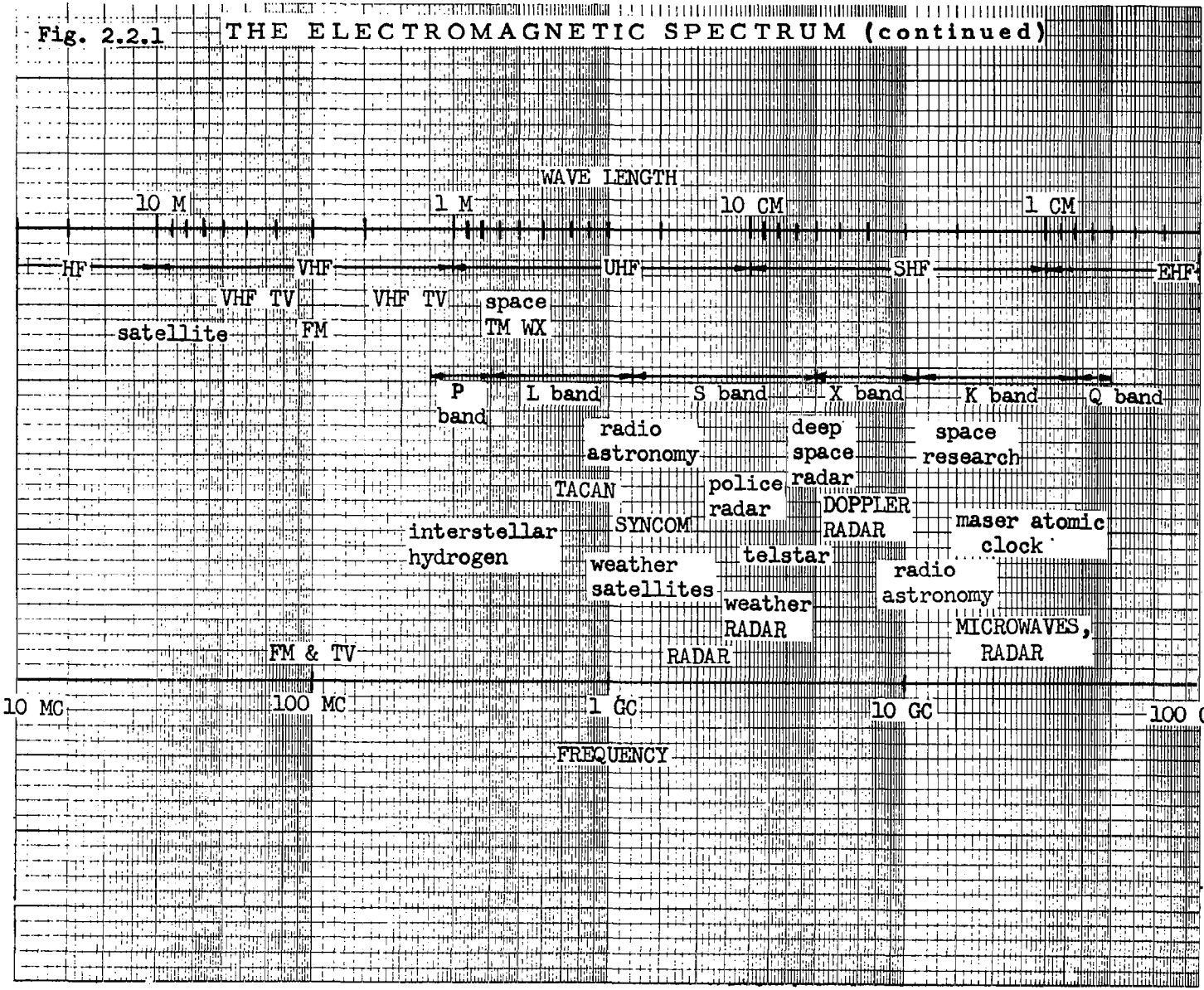


Fig. 2.2.1
THE ELECTROMAGNETIC SPECTRUM (continued)
WAVE LENGTH
10 M
1 M
10 CM
1 CM
HF
VHF
UHF
SHF
EHF
VHF TV
VHF TV
space
TM WX
satellite
FM
P band
L band
S band
X band
K band
Q band
radio astronomy
deep space radar
space research
TACAN
police radar
DOPPLER RADAR
interstellar hydrogen
SYNCOM
maser atomic clock
telstar
weather satellites
radio astronomy
weather RADAR
MICROWAVES, RADAR
FM & TV
RADAR
10 MC
100 MC
1 GC
10 GC
100 GC
FREQUENCY

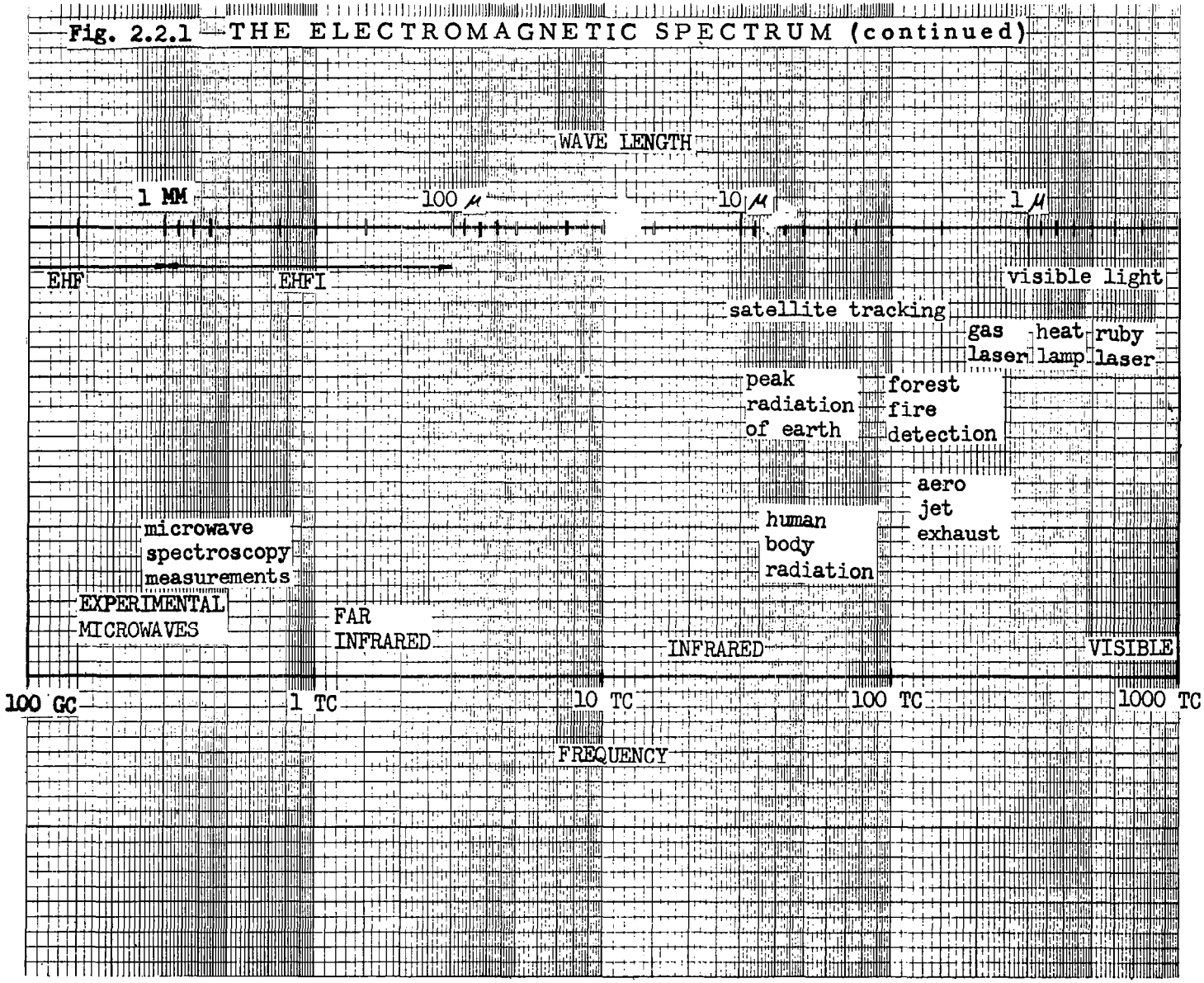


Fig. 2.2.1 THE ELECTROMAGNETIC SPECTRUM (continued)
WAVE LENGTH
1 MM
100 μ
10 μ
1 μ
EHF
EHFI
visible light
satellite tracking
gas laser
heat lamp
ruby laser
peak radiation of earth
forest fire detection
aero jet exhaust
human body radiation
microwave spectroscopy measurements
EXPERIMENTAL MICROWAVES
FAR INFRARED
INFRARED
VISIBLE
100 GC
1 TC
10 TC
100 TC
1000 TC
FREQUENCY

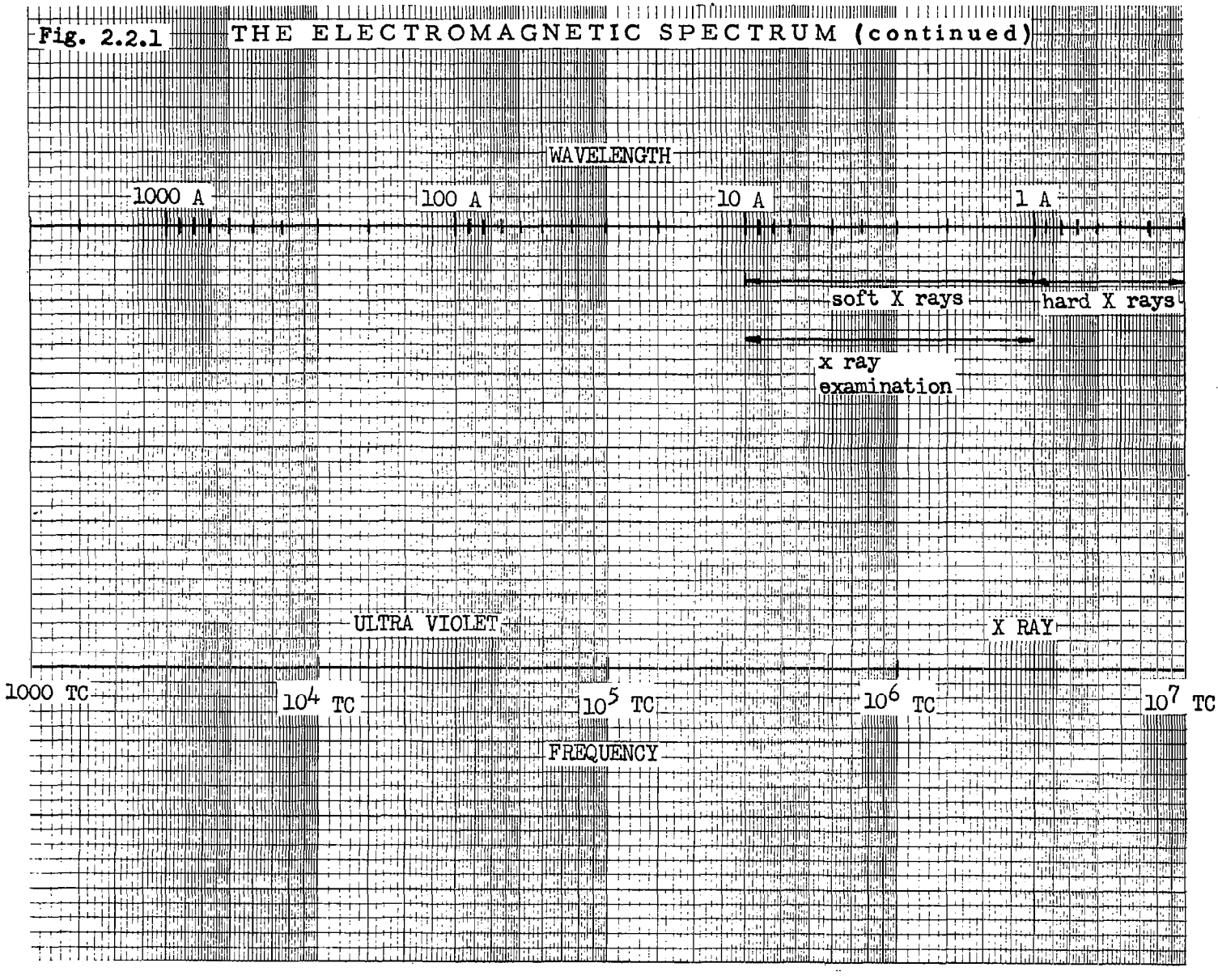


Fig. 2.2.1
THE ELECTROMAGNETIC SPECTRUM (continued)
WAVELENGTH
1000 A
100 A
10 A
1 A
soft X rays
hard X rays
x ray
examination
ULTRA VIOLET
X RAY
1000 TC
10^4 TC
10^5 TC
10^6 TC
10^7 TC
FREQUENCY

The sun emits radiation at all wavelengths, from X-rays to radio waves, though the solar spectrum observed from sea-level shows numerous gaps produced by the absorption of the radiation in the earth's atmosphere. The region from 0-3000 Å is totally obscured, as are numerous other regions centered at approximately 1.3, 1.9, 2.7, 4.2, and 6.0μ, and from 14μ to the region of millimeter waves. Between 14 and 24μ there is a region of limited transparency, which opens up with increased altitude and decreased water content. At longer wavelengths, the atmosphere is made completely opaque from 24 to about 1000μ by the pure rotational spectrum of H_2O. The long-wave end of the ultraviolet absorption, from 3000 Å to 2200Å, is produced entirely by ozone. From 2400 Å to 1300 Å, radiation is absorbed by discrete bands and by continua of O_2. From 1300 Å to about 300 Å are found similar bands of continua of both N_2 and O_2. Beginning at about 900Å the bound-free continua of atomic nitrogen and oxygen abosrb strongly to about 200 Å. Below 200 Å energy is absorbed by the K and L X-ray continua of atomic nitrogen and oxygen. Very little absorption takes place in the visible region of the spectrum apart from the red electronic bands of O_2. In the near infrared, on the short-wave-length side of the photographic limit at 1.35μ, are found relatively weak rotation vibration bands of H_2O and electronic bands of O_2 at 1.06 and 1.27μ. The regions of 100 percent absorption in the infrared are caused entirely by H_2O and CO_2. Many intervening regions are partially obscured by weaker bands of H_2O, CO_2, CH_4, N_2O, CO, and O_3. In the radio-frequency region, the atmosphere is semi-transparent between wave lengths of a few millimeters and about 10 cm and almost completely transparent at longer wave lengths up to about 10 M wavelengths. The absorption of millimeter and centimeter waves is caused mostly by a rotational line of H_2O centered at 1.35 cm and by a series of lines of O_2 at about 5 mm and 2.5 mm that result from changes in the orientation of the spin vector in the ground electronic state.

From this discussion, it can be seen that a relatively small amount of electromagnetic energy penetrates the atmosphere. Generally speaking, there are two "windows" through which electromagnetic energy can penetrate the atmosphere. One is called the radio window which extends from approximately 30 MC to about 22,000 MC. On the low frequency end it is limited by the absorption of the ionosphere. On the high frequency end the radiation is affected by water, oxygen, and nitrogen, which begin to absorb the radiation so that at higher frequencies there is complete atmospheric absorption. Figure 2.2.2 shows the penetration of the electromagnetic spectrum through the atmosphere. **This severe attenuation outside the optical and radio windows must be considered in the selection of navigation sensors which operate at the earth's surface on radiation from outer space, and also in the selection of navigation sensors which are to operate in space on radiation generated on the earth's surface or within the earth's atmosphere.**

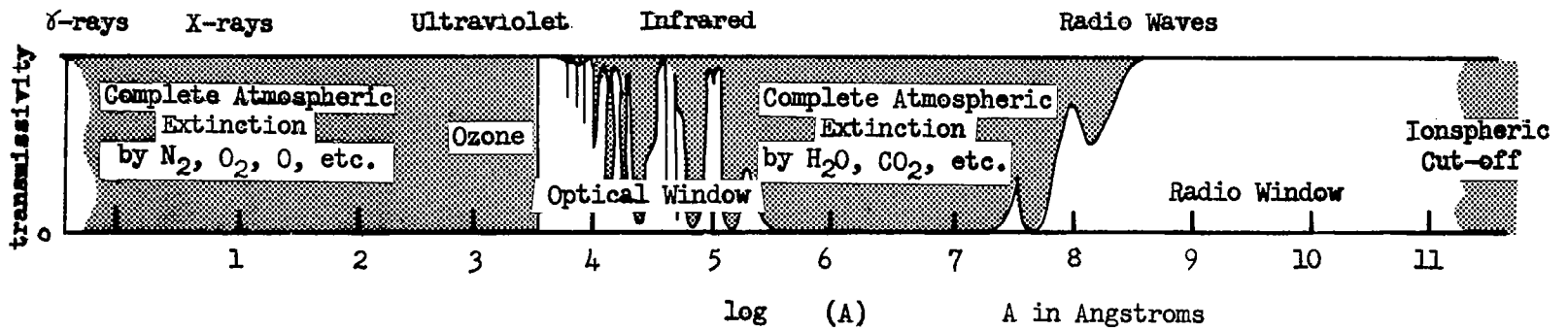


Fig. 2.2.2 Atmospheric Penetration (Reference 5)

2.2.1.3 Atmospheric Refraction

Electromagnetic waves propagating within the earth's atmosphere do not travel in straight lines, but are generally bent or refracted. One effect of refraction is to appear to extend the distance to the horizon. Another effect is the introduction of errors in the measurement of the elevation angle. Bending, or refraction, of electromagnetic waves in the atmosphere is caused by the variation with altitude of the velocity of propagation, or the index of refraction, defined as the ratio of the velocity of propagation in free space to that in the medium in question. The index of refraction in the lower atmosphere may be decomposed into a mean static profile η_0 (h) which depends only on height, and a component $\Delta\eta$ which varies randomly in time and space

$$\eta = \eta_0(h) + \Delta\eta(\ell, t) \tag{2.2.1}$$

Since η is usually very near unity, it is convenient to measure the refractive index in terms of parts per million or N units of refractivity, i.e.,

$$N = (\eta - 1)\,10^6 \tag{2.2.2}$$

The empirical formula for refractivity at microwave frequencies is

$$(\eta - 1)\,10^6 = N = \frac{77.6}{T} P + \frac{3.73 \times 10^5}{T^2} q \tag{2.2.3}$$

where T = air temperature in degrees Kelvin for the point of interest
P = total pressure in millibars for the point of interest
q = partial water vapor pressure in millibars for the point of interest

This expression is independent of frequency from 100 mc to 10,000 mc and is valid within 0.5 per cent up to 30,000 mc. At optical frequencies, water vapor has a negligible effect upon refraction and consequently the second term of equation (2.2.3) may be neglected. The equation for refractivity at optical frequencies is thus

$$(\eta - 1)\,10^6 = N = \frac{77.6}{T} P \tag{2.2.4}$$

Refractivity of the atmosphere at frequencies other than those discussed above is not considered of interest in the monograph since, as discussed in Section 2.2.1.2, the atmospheric absorption of electromagnetic energies at frequencies other than those discussed above appears to preclude the possibility of using such signals for practical state-of-the-art navigation

measurements. Since the barometric pressure P and water-vapor content q decrease rapidly with height, while the temperature T decreases slowly with height, the index of refraction normally decreases with increasing altitude. A typical value of the index of refraction near the surface of the earth is 1.0003. In a standard atmosphere, the index decreases at the rate of about 4×10^{-8} m^{-1} of altitude.

The decrease in refractive index with altitude means that the velocity of propagation increases with altitude, causing the rays, which enter the atmosphere from outer space, to bend downward. Variations of the refractive index in the horizontal plane may also exist, but they do not materially alter the bending. Refraction of electromagnetic waves in the atmosphere is analogous to bending of light rays by an optical prism. The path of the wave through the atmosphere may be plotted using ray-tracing techniques, provided the variation of the refractive index is known.

The ionosphere is the ionized portion of the atmosphere which begins at approximately 70 km and extends to an undetermined distance. The theoretical problem of propagation in an ionized medium has received extensive consideration in the literature. The interested reader is referred to Reference 2.

For carrier frequencies much greater than the collision frequency and gyro frequency of the ionosphere the following simplified expression for the equivalent refractive index of an ionized plasma may be used:

$$\eta^2 = \frac{e^2 N_e}{\epsilon_0 m \omega^2} = 1 - \frac{p^2}{\omega^2} \qquad (2.2.4)$$

where ω = carrier frequency in radians per second

$p = [N_e e^2/\epsilon_0 m]^{1/2}$ = critical frequency in radians per second.
N_e = number of electrons per unit volume.
e = charge of the electron (= 4.80×10^{-10} in cgs).
ϵ_0 = dielectric constant of free space (= 1 in cgs).
m = mass of electron (=9.11×10^{-28} gm)

Using N_e measured in electrons/cc and f in kc/sec the above expression simplifies as

$$\eta = \left[1 - \frac{81 N_e}{f}\right]^{1/2} \qquad (2.2.5)$$

Because the refractive index depends on the carrier frequency employed, it is not convenient to tabulate N as a function of altitude. Rather, attempts to establish profiles of the single scalar quantity N_e as a function of height for various locations and times are made.

Like the frequency-independent refractive index of the lower atmosphere, the electron density N_e can be decomposed into a mean static profile N_e (h) and a component ΔN_e which varies randomly in space and time.

$$N_e = N_e(h) + \Delta N_e \ (\underline{R}, t) \qquad (2.2.6)$$

To calculate ray bending in the ionosphere it is necessary to utilize the complete height profile of density and not merely a single maximum value. For purposes of such calculation, the following simplified model for the daylight electron density profile may be used.

$$N_e(h) \begin{cases} 0 & h < 100 \text{ km} \\ 1.5 \times 10^6 \left(\frac{h-100}{200}\right) \frac{\text{electrons}}{\text{cc}} & 100 < h < 300 \text{km} \end{cases} \qquad (2.2.7)$$

During periods of low solar activity, N_e should be down by a factor of approximately 4. At night it should be down by a factor of 2 or 3 below the corresponding daylight values.

2.2.1.4 Noise

The purpose of this section is to describe in a qualitative fashion those noise sources that are of a major concern to navigation observation measurements. Whenever possible, a quantitative estimate of the noise sources will be given.

Since the accuracy limit of an observation is ultimately determined by the noise in the measurement, various sources of noise will be discussed. In most general terms, noise can be considered as any disturbance that in some way interferes with the desired signal used for a measurement. It may originate within the device used for the measurement or may enter the device along with the desired signal. The former will be referred to as internal noise whereas the latter will be called external noise.

Section 2.5.1, Radiation Sensor Errors, will utilize the information presented in this section in order to determine the noise power that can be expected from these various sources of noise. This information will be used in turn to determine the signal-to-noise ratio and measurement errors induced by noise.

2.2.1.4.1 External Noise

This section introduces the various sources of external noise of interest to navigation observation measurements and some of the terminology that is used to quantitatively evaluate these noise sources. Consideration is given to the effects of noise sources on different frequency regions.

External noise in general limits the sensitivity of long-range navigation, broadcast, and shortwave frequency transmissions. At microwave frequencies, the external noise level is relatively low and the sensitivity of conventional radar receivers is determined primarily by internal noise. There are special devices that have extremely low internal noise levels such as masers and parametric amplifiers in which the external noise again becomes the limiting factor for the sensitivity.

The magnitudes of external noise sources are given in terms of brightness, B, or flux density, S. These quantities cannot be used as such for noise calculations. The following discussion will demonstrate how these quantities can be used to determine the noise power and the effective noise temperature due to these noise sources.

The term "brightness" is used to describe those noise sources that span a very large portion of the sky, as opposed to point or discrete sources whose intensities are described by the term "flux density." The amount of power that an antenna receives from discrete noise sources can be calculated by

$$N = \int A_0 \, S(f) \, d B_n$$

where A_0 = effective aperture

dB_n = differential bandwidth

N = noise power (watts)

If the flux density can be assumed to be constant over a given bandwidth B_n, then the noise power becomes

$$N = A_0 \, S \, B_n$$

If the noise is given in terms of the brightness, B, the power is more difficult to determine. Since the brightness is a function of both azimuth and elevation angles and frequency, the maximum amount of flux density that the antenna could receive must be found by the following integral

$$S = \iint B(A, E, f) \, d\Omega$$

The antenna gain pattern reduces the amount of flux density seen by the antenna to

$$S_A = \iint B(A, E, f) G(A, E, f) \, d\Omega$$

where $G(A, E, f)$ = normalized antenna gain pattern

It is convenient to introduce an equivalent beam area of the antenna in order to facilitate calculations. This equivalent beam area is defined to be β where

$$\beta = \iint G(A, E, f) \, d\Omega$$

Now, if the antenna pattern is smaller than the extended source of brightness B_o (assumed to be constant in the equivalent antenna pattern and on the bandwidth), the flux density may be written as

$$S_A = B_o \beta$$

The noise power can now be written as

$$N = A_o B_o \beta B_n$$

where

A_o= effective aperture area, m^2

B_o= brightness, $\frac{\text{watts}}{(m^2)(CPS)\ \text{steradians}}$

β= effective beam area in steradians

B_n= bandwidth (CPS)

The typical variations of the brightness over the antenna pattern and bandwidth do not significantly restrict the use of this equation. The following is a sample calculation of noise power from brightness.

Example:

At a certain frequency, an antenna has an equivalent beam area of 1 steradian and is directed at a point in the sky where the brightness is constant over the beam area and is equal to 5×10^{-25} watts/m^2(CBS)(steradian). If the antenna has an effective aperture area of 20 m^2 and the noise bandwidth of the receiver is 1 MC, what is the power received if the flux density ($B_o \beta$) is constant over the bandwidth?

$$N = A_o B_o \beta B_n$$

$= (20m^2)$ $(5 \times 10^{-25}$ watts/m^2 (CPS)(steradian)) (1 steradian)$(10^6$ CPS)

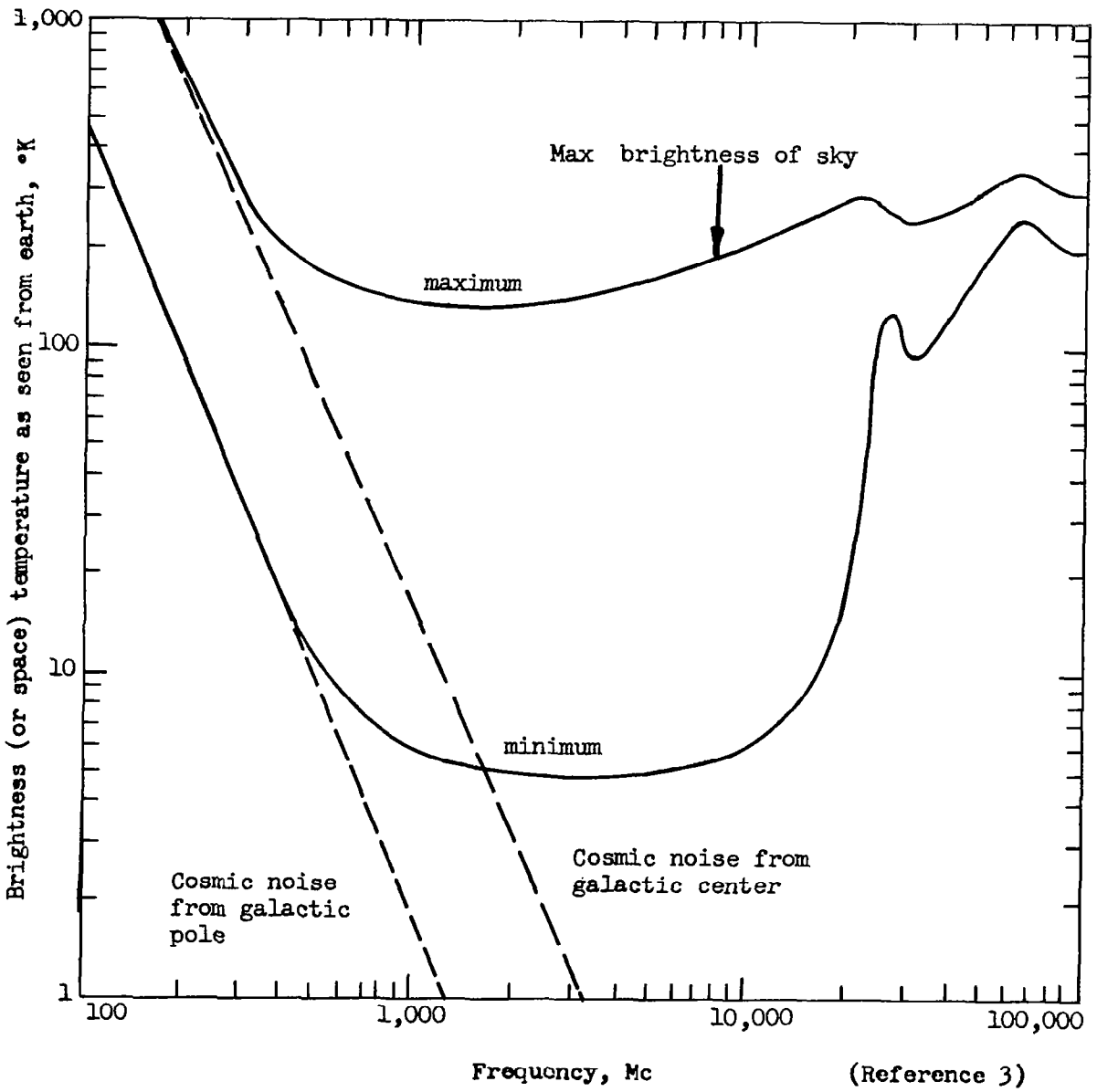
$= 10^{-17}$ Watts

2.2.1.4.1.1 Cosmic Noise

There is a continuous background of noiselike electromagnetic radiation which arrives from outer space. This extra terrestrial noise comes from our own galaxy (the Milky Way), from extra galactic sources, discrete "radio stars" and the sun. In general, cosmic noise decreases with increasing frequency. Cosmic noise is of considerable importance in the design of radars which operate in the VHF or the lower UHF bands, but it may usually be neglected at L-band frequencies or higher (frequencies greater than approximately 1 GC).

The magnitude of cosmic noise depends upon the portion of the celestial sphere in which it is observed. It is a maximum when looking toward the center of our own galaxy, and it is a minimum when observing along the pole about

which the galaxy revolves. A plot of the maximum and minimum cosmic-noise brightness temperatures, or space temperature, as a function of frequency is shown by the dotted lines in the following sketch.



(Reference 3)

The brightness temperature of an extended source of radiation measured in a particular direction is the temperature of a black-body which yields a brightness equal to that of the source under consideration. Brightness is defined as the power received per unit area of aperture per cycle of bandwidth per unit solid angle. The brightness B and the brightness temperature T_B at radio and radar frequencies are related by the Rayleigh-Jeans formula

$$B = \frac{2KT_B}{\lambda^2} \qquad (2.2.8)$$

where K = Boltzmann's constant

λ = wavelength

The brightness temperature specifies the intensity in a given direction at a given frequency. The measurable temperature is the mean brightness temperature in the field of the antenna pattern, and is called the antenna temperature.

The physicist and astronomer are generally concerned with unpolarized radiation, but most radar antennas are responsive to a single polarization. Therefore, the brightness (or space) temperature plotted in the sketch assumes a receiver with a single polarization and is one-half the brightness which would be measured by an antenna responsive to two orthogonal polarizations.

The term "space temperature" is sometimes used synonymously with the brightness temperature of cosmic noise. It is the temperature seen by an ideal antenna (one with no sidelobes, backlobes, or resistive losses) which looks at cosmic noise in the absence of the earth's atmosphere or any other sources of noise.

2.2.1.4.1.2 Atmospheric Absorption Noise

It is known that any body in equilibrium which absorbs energy, radiates the same amount of energy that it absorbs. An example of this is the lossy transmission line that absorbs a certain amount of energy and re-radiates it as noise. The same is true of the atmosphere, since it also attenuates or absorbs microwave energy. The radiation arising in the atmosphere (or any other absorbing body) must just compensate for the partial absorption of the black-body radiation.

Consider an absorbing atmosphere at an ambient temperature T_a surrounded by an imaginary black-body at the same temperature. The loss L is the factor by which energy is attenuated in passing through the atmosphere. The noise power available over a bandwidth B_n from the imaginary black-body is $K T_a B_n$. The noise power after passing through the atmosphere is $(K T_a B_n)/L$. Thus the amount of power absorbed by the atmosphere is $K T_a B_n(1-1/L)$ and is equal to the noise power radiated by the atmosphere itself. This corresponds to an effective noise temperature of

$$T_e = T_a (1 - L) \qquad (2.2.9)$$

A plot of the single-polarization brightness temperature or space temperature due to both cosmic noise and atmospheric absorption is shown by the solid curves of the previous sketch. An ambient temperature of 260°K is assumed in the computation of atmospheric absorption noise. At the higher frequencies (X band or above), atmospheric absorption is the predominant contributor to the brightness temperature, while at the lower frequencies (L band or lower), the cosmic noise predominates. There exists a broad minimum in brightness temperature extending from about 1 GC to 10 GC. It is in this region that it is advantageous to operate low-noise receivers to achieve maximum system sensitivity.

The minimum atmospheric absorption occurs when the antenna is vertical (pointed at the zenith), while the maximum occurs when the antenna is directed along the horizon. The noise is greater along the horizon than at the zenith since the antenna "sees" more atmosphere. Experience shows that antenna beams must be oriented at elevation angles greater than about 5 percent to avoid excessive space noise in the main beam.

2.2.1.4.1.3 Atmospheric Noise

Noise that arises from the lightning-stroke radiation is called atmospheric noise (not to be confused with noise produced by atmospheric absorption as previously described). A single lightning stroke radiates considerable RF noise power. There are at any one moment an average of 1,800 thunderstorms in progress in different parts of the world. From all these storms, about 100 lightning flashes take place every second. The combined effect of all the lightning strokes gives rise to a noise spectrum which is especially large at broadcast and short-wave radio frequencies. The spectrum of atmospheric noise falls off rapidly with increasing frequency and is usually of little consequence above 50 MC. Hence, atmospheric noise is seldom an important consideration in radar except for long-range navigation and radars in the VLF region.

Another source of noise predominant at the lower radar frequencies is man-made noise. Noise from automobile ignition, electric razors, power tools, and fluorescent lights are examples. Just as with atmospheric noise, man-made noise is usually of little concern to radars at UHF or higher frequencies.

2.2.1.4.1.4 Solar Noise

The sun is a strong emitter of electromagnetic radiation, the intensity of which varies with time. The minimum level of solar noise is due to black-body radiation at a temperature of about 6000°K. The flux density received on earth from a thermal source at the distance of the sun is

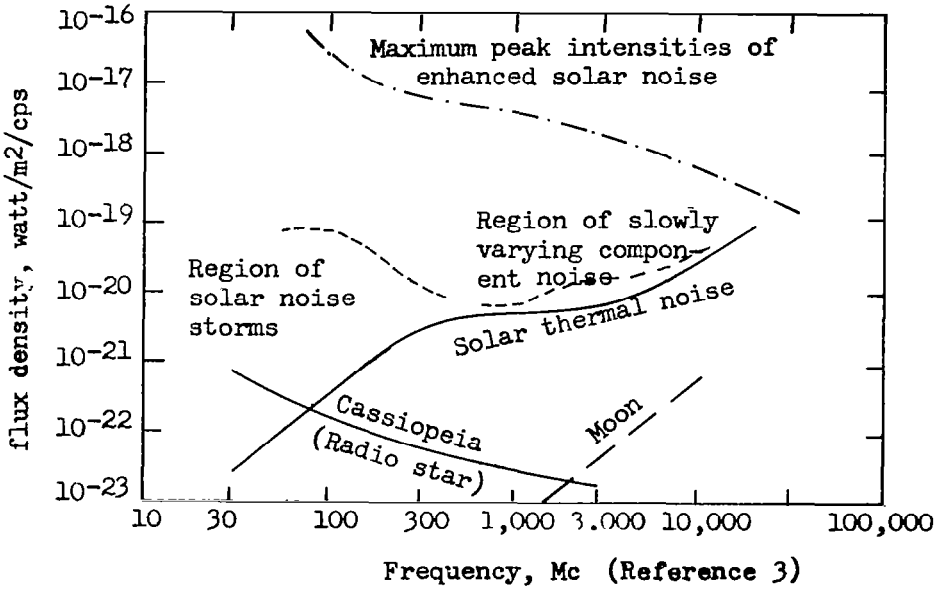
$$S = \frac{1.88 \times 10^{-27} T_d}{\lambda^2} \qquad (2.2.10)$$

where S = flux density watts/(m^2)(cps)

Td = apparent disk temperature, °K

λ = wavelength, m

The above equation indicates that solar noise power increases approximately as the square of the frequency. This is unlike most other noise mechanisms which produce less power with increasing frequency. A plot of the flux density as function of frequency for the basic thermal-noise component from the "quiet" sun is shown in the following sketch.



It does not exactly follow the relationship of Equation (2.2.8) since it takes into account the absorption in the solar atmosphere.

The solar-noise level can increase orders of magnitude over that of the "quiet" or undisturbed sun when its surface is disturbed by solar storms (sunspots and flares). The enhanced noise from the disturbed sun is complex, and its mechanism is not well understood. It might last for but a fraction of a second, or it might last for days. Enhanced solar noise is also shown in the previous sketch. At VHF the solar noise can exceed the thermal component by 40 db or more, while at the upper end of the microwave region there is but a slight increase in the noise level during the periods of enhanced activity. In general, the greater the intensity of the enhanced noise, the shorter its duration. The "noise storms" indicated in the sketch last for hours or days, during which the level shows a series of bursts on seconds' duration super imposed on a more slowly varying background. The "slowly varying component" is believed to originate in thermal radiation from localized regions of abnormally high density and temperature.

2.2.1.4.1.5 Discrete Radio Star Noise

There are a number of discrete radio noise sources in the sky, called radio stars. One of the largest is located in the Cassiopeia constellation. Its flux density is also plotted in the previous sketch. In general, radio stars are too weak at radar frequencies to be a serious source of interference.

2.2.1.4.2 Internal Noise

There are several sources of noise that originate within the sensor itself. These noise sources are usually due to the random fluctuations of the flow of electrons in a sensor, and fluctuations in the stream of photons incident on a radiation sensor.

If an electrical resistor is held at thermal equilibrium at temperature T, a random noise voltage appears across its terminals. This phenomenon is called Nyquist noise, Johnson noise, or thermal noise and is due to the random thermal motion of electrons. Since minimal thermal noise is used to determine the "noise quality" of a sensor, the quantitative discussion of it has been placed in section 2.5.1.5, Signal-to-Noise Ratio.

Shot noise is another type of internal noise found in simple diodes, grid-controlled tubes, traveling-wave tubes, klystrons, magnetrons, crystal diodes, transistors and other current carrying devices. Shot noise is attributed to the passage of electrons through an electronic device after randomly attaining enough energy to overcome some potential energy barrier. The quantitative analysis of shot noise depends on the nature of the device being considered. These numerous discussions are beyond the scope of this monograph. The interested reader is referred to References (3) and (11).

Noise is also introduced in photo-detectors due partly to the fluctuations in the rate of generation of the current carriers, and partly from fluctuations in their rate of recombination. Photo-detectors are also subjected to variations in a stream of photons that they receive.

Mechanical vibrations also induce noise in sensors because they have minute effects on the stray capacitances and inductances that are found in sensors. These microphonic effects can be eliminated by proper design of sensor.

There exists in semi-conductors at low frequencies a noise mechanism whose spectral density is inversely proportional to frequency. This is called flicker noise or 1/f noise. Several theories have been developed to explain this effect, but it is difficult to account for the lack of temperature dependence and the 1/f dependence over all frequencies. Because of the inverse relationship between flicker noise power and frequency, flicker noise will be the predominant effect in devices at low frequencies.

2.2.2 Radiation Sensing Techniques

The general theory describing radiation phenomena presented in the previous section provides a basis for discussing the subject of primary interest; namely, the radiation sensing techniques useful in navigation.

Navigation sensors which exploit radiation phenomena may be divided into two groups. One group consists of those devices used in conjunction with electromagnetic radiation energy which passes through the atmosphere. The second group consists of those devices employed to sense electromagnetic radiation which is unfiltered by the atmosphere. The electromagnetic radiation sensors which have naturally been developed to the greatest degree are those sensitive to frequencies which are not severely attenuated by the atmosphere. As discussed in the previous section, there are only two ranges of frequencies of electromagnetic radiation capable of penetrating the atmosphere. These frequency ranges are called windows, and are called (1) the optical window, and (2) the radio window. All other frequencies suffer severe attenuation

As discussed in Section 2.2.1.1 and illustrated in Figure 2.2.2, the devices which are currently considered state of the art for making navigation measurements reflect this historical interest in the optical and radio frequencies. It may be anticipated in the future that other parts of the electromagnetic spectrum will be exploited for specialized space navigation purposes in which the incident energy exposed to the sensors has not been pre-filtered by the atmosphere to remove all but the optical and radio frequencies. The discussion of this section is concerned with a description of the sensing techniques used in generating information via optical and radio devices.

2.2.2.1 Optical Techniques

The visible portion of the electromagnetic spectrum, usually defined as extending from .4 to .7 microns, and the infrared portion of the spectrum, usually defined as extending from .7 to 4 microns, dominates those wavelengths which pass through the atmosphere with little attenuation in the optical window. The common use of optical techniques to generate navigation measurements is in the sensing of the direction in space of a single line-of-sight instrument or the use of a two line-of-sight instrument to measure the included angle between two objects which radiate electromagnetic energy in space. Although infrared frequencies may be sensed by using the IR sensors, more often visible frequencies are sensed with the more common optical devices. The classification of the single line-of-sight instrument includes devices such as the telescope, star tracker, horizon scanner, and sun sensor. The two line-of-sight instrument includes the sextant, trisextant, and stadiametric devices. The navigation measurements which result from the

use of the single line-of-sight instrument may be either intensity measurements or direction measurements. The intensity measurements are not often used as navigation measurements. The direction measurements are common but require the use of some form of platform having a known orientation. The inertial platforms discussed in Section 2.3.2.1 are often used for this purpose and permit the generation of azimuth and elevation angular measurements. The two line-of-sight instruments are used to generate a measure of an included angle between two radiation sources, as mentioned above, and are usable in making sun-planet, planet-planet, planet-star, sun-star, and star-elevation measurements. In addition, stadiametric measurements which determine the curvature of a planet horizon may be used to infer planet diameter, and are essentially a variation of a sextant. Star occulation is in essence a time measurement but is associated with a single line-of-sight measurement. These single and two line-of-sight measurement techniques are discussed in detail in this section. Section 2.4 will present detailed analyses of the important line-of-sight techniques.

2.2.2.2 Radio Techniques

Since radar is used quite extensively in some types of observations, a general discussion of some of the more important techniques has been included in this section. The primary use of radar in observations is to determine the following "observables": range, range-rate, azimuth and elevation of some target. The measurement of these "observables" involves inaccuracies,as is expected in any measurement. However, two of these "observables," namely, range and range-rate are not measured directly, but by the measurement of time delays, phase measurements and frequency measurements. It is therefore necessary to relate the inaccuracies of these parameters to the "observables" so that the inaccuracies in the "observables" can be determined. The following sections formulate the various range and range-rate measurement techniques so that these relations can be established.

Section 2.5, Observation Errors, describes the errors associated with the parameters that are actually measured. The combination of the information from section 2.5 and this section provide the information of the accuracies of the observables that will be necessary to use in the state determination problem which is discussed in the following monograph.

2.2.2.2.1 Radar Range Measurement

The range of a target is determined by measuring the time required for a pulse to travel from a radar station to the target and return. Since electromagnetic energy travels at the speed of light, the range R can be written as

$$R = \frac{c\,\Delta t}{2} \tag{2.2.11}$$

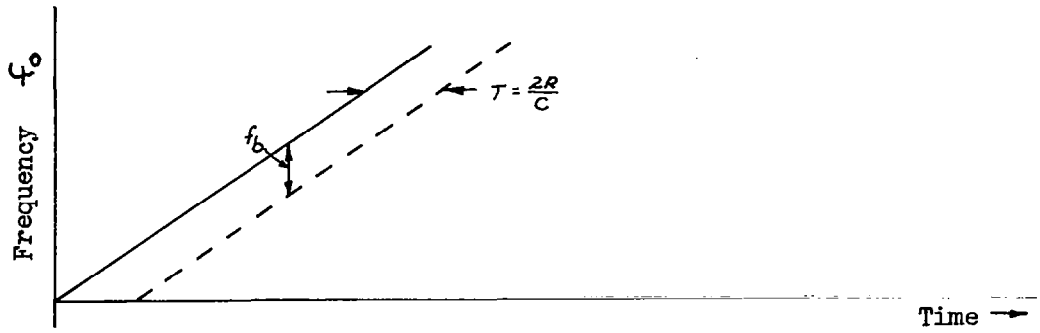
where

Δt = time required for wave to travel to target and back

c = velocity of light 3×10^8 m/sec

Range measurement commonly uses a train of narrow pulses for its waveform. The measurement of range is then proportional to the measure of the time delay between the leading edge of the transmitted and reflected pulse. The accuracy with which one can measure this time delay is discussed in section (2.5.1.2.2).

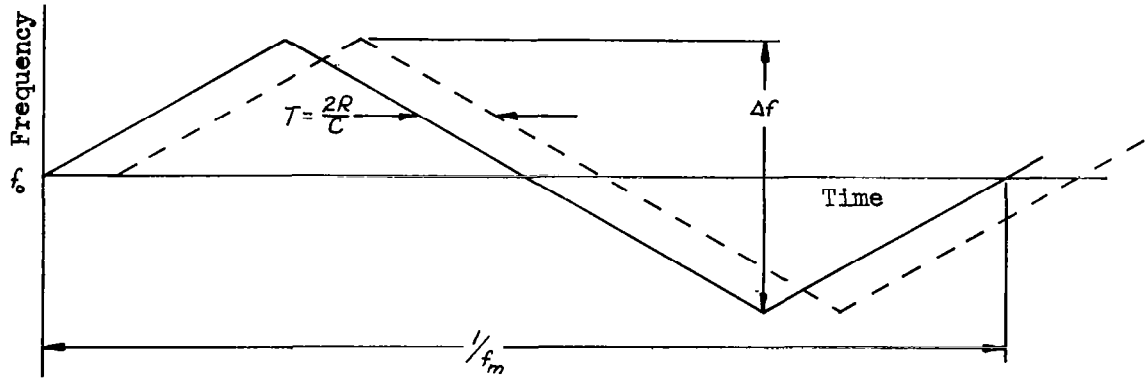
Another range measurement technique uses Frequency Modulated Continuous Wave Radar (abbreviated FM-CW Radar). In this technique, the transmitter frequency is changed as a function of time in a known manner. Assume that the transmitter frequency increases linearly with time, as shown by the solid line in the following sketch:



If there is a reflecting object at a distance R, an echo signal will return after a time $\Delta t = 2R/c$. The dashed line in the sketch represents the echo signal as it is received. If the echo signal is heterodyned with a portion of the transmitter signal in a nonlinear element such as a crystal diode, a beat frequency f_b will be produced. If there is no doppler frequency shift, the beat frequency (difference frequency) is a measure of the target's range and $f_b = f_r$, where f_r is the beat frequency due only to the target's range. If the rate of change of the carrier frequency is f_o, the beat frequency is

$$f_r = \dot{f}_o \, \Delta t = \frac{2R}{C} \dot{f}_o \qquad (2.2.12)$$

In any practical CW radar, the frequency cannot be continually changed in one direction only. Periodicity in the modulation is necessary, as in the triangular - frequency modulated waveform shown below.



The modulation need not necessarily be triangular; it can be sawtooth, sinusoidal, or some other shape. The resulting beat frequency as a function of time is shown below for triangular modulation.

The beat frequency is of constant frequency except at the turn around region. If the frequency is modulated at a rate f_m over a range Δf, the beat frequency is

$$f_r = \frac{2R}{c} \; 2 f_m \Delta f = \frac{4 R f_m \Delta f}{c} \qquad (2.2.13)$$

Thus, the measurement of the beat frequency determines the range R. It should be noted that the target was assumed to be stationary in the previous analysis. It will be shown in the following section, Radar Range-Rate Measurement, that this method is still valid for moving targets after proper treatment of the waveforms.

The use of single frequency CW radar for range measurements is not practical because the time delay that is to be measured manifests itself as a phase difference between the transmitted and reflected signals, and at radio frequencies the ambiguous distances that correspond to the true phase plus 2π multiples of phase are too numerous in the areas of interest to give useful information.

The region of unambiguous range may be extended considerably by transmitting two separate CW signals differing only slightly in frequency. It will be shown that the measurement of range using two CW frequencies results in an unambiguous range which corresponds to a half wavelength at the difference frequency. Consequently, the unambiguous range can be made considerably greater than that obtained when only a single frequency is transmitted.

The transmitted waveform is assumed to consist of two continuous sine waves of frequency f_1 and f_2. For convenience, the amplitudes of all signals are set to unity. The voltage waveforms of the two components of the transmitted signal V_{1T} and V_{2T} may be written as

$$V_{1T} = \sin(2\pi f_1 t + \phi_1)$$

$$V_{2T} = \sin(2\pi f_2 t + \phi_2) \qquad (2.2.14)$$

where ϕ_1 and ϕ_2 are arbitrary (constant) phase angles. Due to the transit time delay, the reflected signal can be expressed as

$$V_{1R} = \sin\left[2\pi f_1 \left(t - \frac{2R}{c}\right) + \phi_1\right]$$
$$V_{2R} = \sin\left[2\pi f_2 \left(t - \frac{2R}{c}\right) + \phi_2\right] \qquad (2.2.15)$$

The phase differences between each transmitted and reflected signal are

$$\phi_1' = 2\pi f_1 \frac{2R}{c}$$
$$\phi_2' = 2\pi f_2 \frac{2R}{c} \qquad (2.2.16)$$

If the difference between ϕ_1' and ϕ_2' is found, it is seen that it is an indirect measurement of range; i.e.,

$$\Delta\phi' = 2\pi f_1 \frac{2R}{c} - 2\pi f_2 \frac{2R}{c} = \frac{4\pi (f_1 - f_2) R}{c} \qquad (2.2.17)$$

or

$$R = \frac{c\,\Delta\phi'}{4\pi (f_1 - f_2)} \qquad (2.2.18)$$

Note that the unambiguous range is now a function of the <u>difference</u> in frequencies rather than the frequency itself and, hence, is much larger. The range accuracy using this method may now be determined by using the phase measurement equations in section 2.5.1.2.1.

2.2.2.2.2 <u>Radar Range-Rate Measurement</u>

Standard Doppler frequency shift equations can be used in order to derive the relation between the range-rate measurement and frequency shift. Consider a stationary radar station that is transmitting a frequency f_0. As this signal is received by a target in motion, the frequency appears to be slightly different because of the radial component of the target velocity with respect to the radar station. This new frequency, f_1, can be related

to the original frequency by the standard Doppler equation for a stationary source and a moving receiver.

$$f_1 = f_o \frac{c \pm v}{c} \tag{2.2.19}$$

where a (+) is used for approaching targets
and a (-) is used for receding targets

The target reflects the signal back to the radar station, but the frequency again appears to change because the source (which is now the target) is moving with respect to the stationary receiver (the radar station). This reflected signal appears to have a frequency f_2 and can be found by again using the standard Doppler equation for a moving source and a stationary receiver.

Hence

$$f_2 = f_1 \frac{c}{c \mp v} \tag{2.2.20}$$

where a (-) is used for approaching targets
and a (+) is used for receding targets

Now f_2 can be related to f_o as

$$f_2 = f_1\left(\frac{c}{c \mp v}\right) = f_o\left(\frac{c \pm v}{c}\right)\left(\frac{c}{c \mp v}\right) \tag{2.2.21}$$

or

$$f_2 = f_o\left(\frac{c \pm v}{c \mp v}\right) \tag{2.2.22}$$

The difference between f_o and f_2 is

$$f_d = f_2 - f_o = f_o\left(\frac{c \pm v}{c \mp v}\right) - f_o \tag{2.2.23}$$

or

$$f_d = f_o\left[\frac{c \pm v}{c \mp v} - \frac{c \mp v}{c \mp v}\right] = \frac{2 f_o(\pm v)}{c \mp v} \tag{2.2.24}$$

Since any target velocity will be very much smaller than the speed of light, the following approximation is valid.

$$c \mp V \cong c \tag{2.2.25}$$

So

$$f_d = \frac{2f_o(\pm V)}{c} \tag{2.2.26}$$

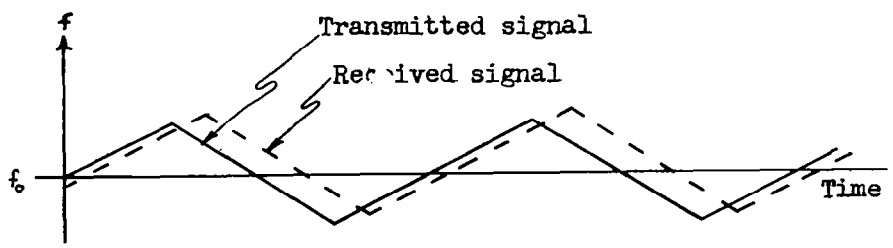
The above equation is the well known range-rate equation in terms of the Doppler frequency shift. The appropriate sign to use should be obvious from the previous discussion. Letting V_R be the radial component of velocity, the above equation becomes

$$V_R = \pm \frac{c f_d}{2f_o} \tag{2.2.27}$$

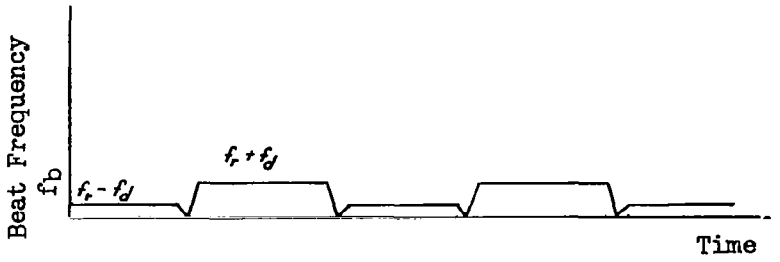
It is apparent that the accuracy that can be expected in the range-rate measurement is related to the accuracy of the frequency measurement accuracy.

In section (2.2.2.2.1) the FM-CW radar measurement assumed a stationary target. It will now be shown that with the proper treatment of the resulting waveforms, the FM-CW is also valid, not only for range, but also range-rate determination.

The Doppler frequency shift causes the frequency time plot of the echo signal to be shifted up or down depending on the sign of Δf. The following sketch shows a typical shift:



A corresponding change in the beat frequency occurs because of the Doppler frequency shift. On one portion of the frequency-modulation cycle, the beat frequency is increased by the Doppler shift, while on the other portion, it is decreased. The sketch below illustrates this point.



If, for example, the target is approaching the radar, the beat frequency f_b (up) produced during the increasing, or up, portion of the FM cycle will be the difference between the beat frequency due to the range f_r and the Doppler frequency shift f_d. Similarly, on the decreasing portion, the beat frequency f_b (down) is the sum of the two; i.e.,

$$f_b\,(UP) = f_r - f_d \tag{2.2.28}$$

$$f_b\,(DOWN) = f_r + f_d$$

The range frequency, f_r, may be extracted by measuring the average beat frequency; that is

$$f_r = \frac{1}{2}\left[f_b(UP) + f_b(DOWN)\right] \tag{2.2.29}$$

If f_b (up) and f_b (down) are measured separately by switching a frequency counter every half modulation cycle, one-half the difference between the frequencies will yield the Doppler frequency; i.e.,

$$f_d = -\frac{1}{2}\left[f_b(UP) - f_b(DOWN)\right] = -\frac{1}{2}\left[(f_r - f_d) - (f_r + f_d)\right] \tag{2.2.30}$$

2.3 INERTIAL THEORY AND SENSORS

It is the purpose of this section to present the fundamental dynamic causal relationships which enable the generation of position and velocity information from inertial sensors. The presentation is introductory to the extent that only the broad fundamental principles of inertial navigation are included. A tremendous variety of naval, aircraft, and spacecraft inertial systems have been developed for a myriad of purposes. These systems and technical approaches are well documented in the literature and will not be discussed here.

The discussion commences with a discussion of the dynamic characteristics of the various gyroscopic sensors, and is followed by a discussion of the various linear accelerometer sensors. Based upon the theory of the gyroscope and accelerometer, inertial systems capable of sensing position and velocity are then presented.

2.3.1 General Theory

2.3.1.1 Rotational Theory

The rotational theory presented in this section consists of the development of the equations describing the dynamics of a gyroscope which is subjected to external torques in general three degree of rotational freedom dynamic space. The development commences with a rigorous treatment of a single axis gyroscope, which is subsequently used as the basis for describing the dynamics of the rate gyro, the attitude gyro, and the angular integrating gyro.

2.3.1.1.1 Single Axis Gyro Equations

Figure 2.3.1 presents the terminology to be utilized in this development. The angular velocity of the gimbal is denoted $\begin{bmatrix} p' \\ q' \\ r' \end{bmatrix}$, with the upper left superscript used to denote the cartesian coordinate frame used to define the axis in which gimbal angular rate is specified.

From Figure 2.3.1

$$ {}^{3}\underline{\omega}' = \begin{bmatrix} {}^{3}p' \\ {}^{3}q' \\ {}^{3}r' \end{bmatrix} \triangleq {}^{3}p'\,{}^{3}\underline{e}_x + {}^{3}q'\,{}^{3}\underline{e}_y + {}^{3}r'\,{}^{3}\underline{e}_z \qquad (2.3.1) $$

and expressing the angular velocity of the rotor $\underline{\omega}''$ in the 4 coordinate frame:

$$ {}^{4}\underline{\omega}'' = \begin{bmatrix} {}^{4}p'' \\ {}^{4}q'' \\ {}^{4}r'' \end{bmatrix} \triangleq {}^{4}p\,{}^{4}\underline{e}_x + {}^{4}q\,{}^{4}\underline{e}_y + {}^{4}r\,{}^{4}\underline{e}_z \qquad (2.3.2) $$

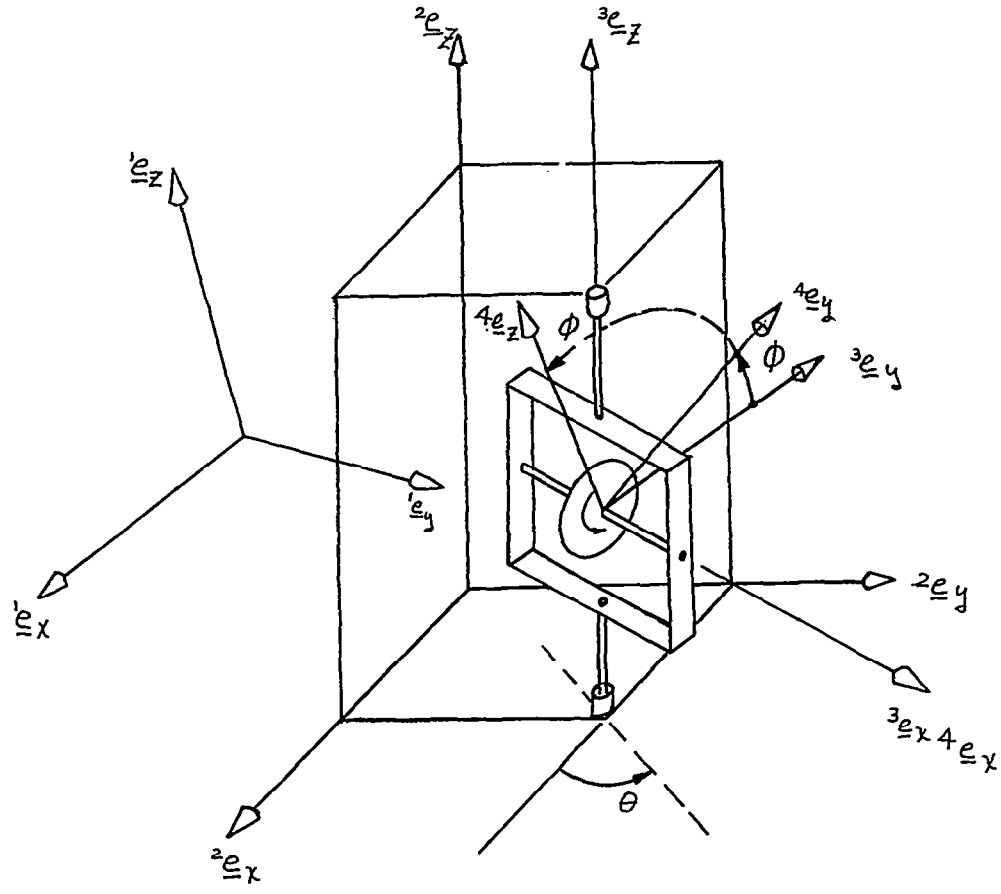


Figure 2.3.1 Single Axis Gyroscope

The total angular momentum of the gyroscope, including the gimbal and rotor, is:

$$ {}^{3}\underline{H}_{GYRO} = {}^{3}\underline{H}_{GIMBAL} + {}^{3}\underline{H}_{ROTOR} \qquad (2.3.3) $$

where

$$ {}^{3}\underline{H}_{GIMBAL} = \underline{\underline{I}}' \, {}^{3}\underline{\omega}' = \begin{bmatrix} I'_{11} & I'_{12} & I'_{13} \\ I'_{21} & I'_{22} & I'_{23} \\ I'_{31} & I'_{32} & I'_{33} \end{bmatrix} \begin{bmatrix} {}^{3}p' \\ {}^{3}q' \\ {}^{3}r' \end{bmatrix} \qquad (2.3.4) $$

$$ {}^{4}\underline{H}_{ROTOR} = \underline{\underline{I}}'' \, {}^{4}\underline{\omega}'' = \begin{bmatrix} I''_{11} & I''_{12} & I''_{13} \\ I''_{21} & I''_{22} & I''_{23} \\ I''_{31} & I''_{32} & I''_{33} \end{bmatrix} \begin{bmatrix} {}^{4}p'' \\ {}^{4}q'' \\ {}^{4}r'' \end{bmatrix} \qquad (2.3.5) $$

Note that in Equations (2.3.4) and (2.3.5) that the inertia matrices $\underline{\underline{I}}'$, $\underline{\underline{I}}''$ refer to the inertias of the gimbal and rotor in coordinate frames which are fixed to the gimbal and rotor, respectively. Now the angular velocities of equations (2.3.4) and (2.3.5) can be related by referring to the figure.

$$ \begin{bmatrix} {}^{4}p'' \\ {}^{4}q'' \\ {}^{4}r'' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & \sin\phi \\ 0 & -\sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} {}^{3}p'' \\ {}^{3}q'' \\ {}^{3}r'' \end{bmatrix} \qquad (2.3.6) $$

The notation $T0403 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & \sin\phi \\ 0 & -\sin\phi & \cos\phi \end{bmatrix}$ will be used in the place of the 3x3 matrix.

Hence

$$ {}^{4}\underline{\omega}'' = T0403 \; {}^{3}\underline{\omega}'' \qquad (2.3.7) $$

is the abbreviated form of Equation (2.3.6).

Now, since

$$ {}^{3}\underline{H}_{ROTOR} = (T0304) \, {}^{4}\underline{H}_{ROTOR} \qquad (2.3.8) $$

Then

$$ {}^{3}\underline{H}_{ROTOR} = (T0304) \, \underline{\underline{I}}'' (T0403) \; {}^{3}\underline{\omega}'' \qquad (2.3.9) $$

Furthermore, since the rotor is constrained by the gimbal,

$$ {}^{3}\underline{\omega}'' = {}^{3}\underline{\omega}' + \Omega \, {}^{3}\underline{e}_{x} $$

or

$$\begin{bmatrix} {}^3p'' \\ {}^3q'' \\ {}^3r'' \end{bmatrix} = \begin{bmatrix} {}^3p' \\ {}^3q' \\ {}^3r' \end{bmatrix} + \begin{bmatrix} \Omega \\ 0 \\ 0 \end{bmatrix} \tag{2.3.10}$$

where Ω is the angular velocity of the rotor with respect to the gimbal, substitution of (2.3.10) into (2.3.9) yields:

$${}^3\underline{H}_{ROTOR} = T0304\ \underline{\underline{I}}''\ T0403 \begin{bmatrix} {}^3p' \\ {}^3q' \\ {}^3r' \end{bmatrix} + \begin{bmatrix} I_{11}'' \Omega \\ 0 \\ 0 \end{bmatrix} \tag{2.3.11}$$

Substituting (2.3.4)and (2.3.11) into (2.3.3) yields

$${}^3\underline{H}_{GYRO} = \left[\underline{\underline{I}}' + T0304\ \underline{\underline{I}}''\ T0403\right] \begin{bmatrix} {}^3p' \\ {}^3q' \\ {}^3r' \end{bmatrix} + \begin{bmatrix} I_{11}'' \Omega \\ 0 \\ 0 \end{bmatrix} \tag{2.3.12}$$

Equation (2.3.12) is a general expression for the angular momentum of a gyro in terms of the angular velocity of the gimbal and the relative angular velocity of the rotor with respect to the gimbal. This equation is useful for analyzing perturbations induced by errors in the direction of the principal axes of the inertia matrix.

Symmetrical Rotor Simplifications:

The expression of Equation (2.3.12) may be grossly simplified by noting the I'' matrix usually takes on a very simple form, due to the symmetry of the rotor, i.e.,

$$\underline{\underline{I}}'' = \begin{bmatrix} I_{ROT} & 0 & 0 \\ 0 & I_c & 0 \\ 0 & 0 & I_c \end{bmatrix} \tag{2.3.13}$$

Hence, under these conditions, it is easily verified that:

$$(T0304)\,(I'')\,(T0403) = I'' \tag{2.3.14}$$

Equation (2.3.12) thus simplifies to:

$${}^3\underline{H}_{GYRO} = \left[\underline{\underline{I}}' + \underline{\underline{I}}''\right] \begin{bmatrix} {}^3p' \\ {}^3q' \\ {}^3r' \end{bmatrix} + \begin{bmatrix} I_{ROTOR}\,\Omega \\ 0 \\ 0 \end{bmatrix} \tag{2.3.15}$$

or

$$ {}^3\underline{H}_{GYRO} = \begin{bmatrix} I'_{11}+I_{ROTOR} & I'_{12} & I'_{13} \\ I_{12} & I'_{22}+I_c & I'_{23} \\ I_{13} & I'_{23} & I'_{33}+I_c \end{bmatrix} \begin{bmatrix} {}^3p' \\ {}^3q' \\ {}^3r' \end{bmatrix} + \begin{bmatrix} I_{ROTOR}\Omega \\ 0 \\ 0 \end{bmatrix} \quad (2.3.16) $$

Gimbal Simplifications

Usually the gimbal inertia is distributed so that the gimbal principal axes coincide with the ${}^3\underline{e}_x$, ${}^3\underline{e}_y$, and ${}^3\underline{e}_z$ axes. Under these circumstances,

$$ I'_{12} = I'_{13} = I'_{12} = 0 $$

and Equation (2.3.16) simplifies to:

$$ {}^3\underline{H}_{GYRO} = \begin{bmatrix} I_1 & 0 & 0 \\ 0 & I_2 & 0 \\ 0 & 0 & I_3 \end{bmatrix} \begin{bmatrix} {}^3p' \\ {}^3q' \\ {}^3r' \end{bmatrix} + \begin{bmatrix} I_{ROTOR}\Omega \\ 0 \\ 0 \end{bmatrix} \quad (2.3.17) $$

where

$$ I_1 = I'_{11} + I_{ROTOR} $$
$$ I_2 = I'_{22} + I_c $$
$$ I_3 = I'_{33} + I_c $$

$$ {}^3\underline{\omega}' = \begin{bmatrix} {}^3p' \\ {}^3q' \\ {}^3r' \end{bmatrix} = \text{angular velocity of gimbal} $$

Equation (2.3.17) is the standard equation describing idealized gyro angular momentum. Equation (2.3.17) can be written in shorthand form as:

$$ {}^3\underline{H}_{GYRO} = \underline{\underline{I}}\,{}^3\underline{\omega}' + {}^3\underline{H}_{SPIN} \quad (2.3.18) $$

where

$$ {}^3\underline{H}_{SPIN} = \begin{bmatrix} I_{ROTOR}\Omega \\ 0 \\ 0 \end{bmatrix} $$

The coordinates of any general vector in coordinate frames 2 and 3 are related as follows:

$$\begin{bmatrix} {}^2B_x \\ {}^2B_y \\ {}^2B_z \end{bmatrix} = (T\,0203) \begin{bmatrix} {}^3B_x \\ {}^3B_y \\ {}^3B_z \end{bmatrix} \tag{2.3.19}$$

where

$$(T\,0203) = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Hence,

$$\begin{aligned} {}^3\underline{H}_{GYRO} &= T0302\ {}^2\underline{H}_{GYRO} \\ {}^3\underline{\omega}' &= T0302\ {}^2\underline{\omega}' \end{aligned} \tag{2.3.20}$$

and substitution of (2.3.20) into (2.3.18) yields:

$${}^2\underline{H}_{GYRO} = (T\,0203)\,\underline{\underline{I}}\,(T\,0302)\,{}^2\underline{\omega}' + (T\,0203)\,{}^3\underline{H}_{SPIN} \tag{2.3.21}$$

Expansion of Equation (2.3.21) now produces

$${}^2\underline{H}_{GYRO} = \begin{bmatrix} I_1\cos^2\theta + I_2\sin^2\theta & (I_1-I_2)\sin\theta\cos\theta & 0 \\ (I_1-I_2)\sin\theta\cos\theta & I_1\sin^2\theta + I_2\cos^2\theta & 0 \\ 0 & 0 & I_3 \end{bmatrix} \begin{bmatrix} {}^2p' \\ {}^2q' \\ {}^2r' \end{bmatrix} + \begin{bmatrix} I_{ROT}\Omega\cos\theta \\ I_{ROT}\Omega\sin\theta \\ 0 \end{bmatrix}. \tag{2.3.22}$$

Now, since the angular velocity of the gimbal equals the vector sum of the angular velocity of the 2 coordinate frame plus the relative angular velocity of the gimbal with respect to the 2 coordinate frame (Figure 2.3.1)

$${}^2\underline{\omega}' = {}^2\underline{\omega} + \dot{\theta}\left({}^2\underline{e}_z\right), \tag{2.3.23}$$

where ${}^2\underline{\omega}$ = angular velocity of the 2 coordinate frame.

Substituting (2.3.23) into (2.3.22) yields:

$${}^2\underline{H}_{GYRO} = \begin{bmatrix} I_1\cos^2\theta + I_2\sin^2\theta & (I_1-I_2)\sin\theta\cos\theta & 0 \\ (I_1-I_2)\sin\theta\cos\theta & I_1\sin^2\theta + I_2\cos^2\theta & 0 \\ 0 & 0 & I_3 \end{bmatrix} \begin{bmatrix} {}^2p \\ {}^2q \\ {}^2r \end{bmatrix} + \begin{bmatrix} I_{ROT}\Omega\cos\theta \\ I_{ROT}\Omega\sin\theta \\ I_3\dot{\theta} \end{bmatrix} \tag{2.3.24}$$

where it is recalled that

$${}^2\underline{\omega} = \begin{bmatrix} {}^2p \\ {}^2q \\ {}^2r \end{bmatrix}$$

Equation (2.3.24) simplifies significantly if $I_1 = I_2$; i.e., if the sum of the inertias of the gimbal and rotor about the ${}^3\underline{e}_X$ axis is numerically equal to the sum of the gimbal and rotor inertias about the ${}^3\underline{e}_Y$ axis (recall that the rotor has been assumed to be symmetrical about the ${}^3\underline{e}_X$ axis, so that its inertia about the ${}^3e_Y$ axis is time invariant, even though the rotor is spinning with respect to the ${}^3e_Y$ axis).

Hence, assuming

$$I_1 = I_2 \tag{2.3.25}$$

Equation (2.3.24) becomes:

$${}^2\underline{H}_{GYRO} = \begin{bmatrix} I_1 & 0 & 0 \\ 0 & I_1 & 0 \\ 0 & 0 & I_3 \end{bmatrix} \begin{bmatrix} {}^2p \\ {}^2q \\ {}^2r \end{bmatrix} + \begin{bmatrix} I_{ROT}\Omega\cos\theta \\ I_{ROT}\Omega\sin\theta \\ I_3\dot{\theta} \end{bmatrix} \tag{2.3.26}$$

and the time derivative of (2.3.26) is:

$$\frac{d}{dt}{}^2\underline{H}_{GYRO} = \begin{bmatrix} I_1\dot{p} - I_{ROT}\Omega\dot{\theta}\sin\theta + qr(I_3 - I_1) + I_3\dot{\theta}q - I_{ROT}\Omega r\sin\theta \\ I_1\dot{q} + I_{ROT}\Omega\dot{\theta}\cos\theta + pr(I_1 - I_3) + I_{ROT}\Omega r\cos\theta - I_3\dot{\theta}p \\ I_3(\dot{r} + \ddot{\theta}) + I_{ROT}\Omega p\sin\theta - I_{ROT}\Omega q\cos\theta \end{bmatrix} \tag{2.3.27}$$

where $p = {}^2p$, $q = {}^2q$, $r = {}^2r$

Letting the externally applied torques be termed ${}^2\underline{T}$ then:

$${}^2\underline{T} = \begin{bmatrix} {}^2T_X \\ {}^2T_Y \\ {}^2T_Z \end{bmatrix} = \frac{d}{dt}{}^2\underline{H}_{GYRO} \tag{2.3.28}$$

Hence,

$${}^2T_X = I_1\dot{p} - I_{ROT}\Omega\dot{\theta}\sin\theta + qr(I_3 - I_1) + I_3\dot{\theta}q - I_{ROT}\Omega r\sin\theta \tag{2.3.29}$$

$${}^2T_Y = I_1\dot{q} + I_{ROT}\Omega\dot{\theta}\cos\theta + pr(I_1 - I_3) + I_{ROT}\Omega r\cos\theta - I_3\dot{\theta}p \tag{2.3.30}$$

$${}^2T_Z = I_3(\dot{r} + \ddot{\theta}) + I_{ROT}\Omega p\sin\theta - I_{ROT}\Omega q\cos\theta \tag{2.3.31}$$

Equations (2.3.29), (2.3.30), and (2.3.31) are a key set of differential equations describing the motion of a single degree of freedom gyro. This set of equations is exact, and contains no mathematical approximations. The assumptions which have been made, both explicitly and implicitly, are

presented in the table below:

Assumptions which qualify Equations (2.3.29), (2.3.30), and (2.3.31) as an exact set of differential equations are:

1. The rotor is symmetrical about the $\underline{e}_x$ axis (its spin axis). Thus, all cross product inertia terms are zero, and the 2 principal axes of the rotor which are perpendicular to the spin axis may be chosen arbitrarily oriented in a plane which contains the rotor center of mass, and which is perpendicular to the spin axis. These arbitrarily oriented principal axes must, of course, be mutually orthogonal.

2. The gimbal principal axes are along the ${}^3\underline{e}_x$, ${}^3\underline{e}_y$, and ${}^3\underline{e}_z$ directions.

3. The angular speed of the rotor with respect to the gimbal is constant.

4. The gimbal plus rotor moments of inertia, as measured about the ${}^3\underline{e}_x$ and ${}^3\underline{e}_y$ axes with the rotor frozen to the gimbal, are equal (see Equation (2.3.25)).

The Input Axis and Output Axis Equations

Since the angular deflection of the gyro $\theta(t)$ is considered as the "output" of the gyro, the ${}^2\underline{e}_z$ axis is referred to as the "output axis", and the differential equation describing the force balance about this axis is termed the "output axis equation".

The "output equation" can be fully developed as follows:

Assumption #5

In the general situation, the external torques applied about the ${}^2\underline{e}_z$ axis may be produced by:

a retarding "spring" force $= -k\theta$

a retarding "viscous damping" force $= -c\dot{\theta}$

a disturbance torque $= {}^2T_z'$

Hence, (Assumption #5)

$$ {}^2T_z = -k\theta - c\dot{\theta} + {}^2T_z' \qquad (2.3.32) $$

Substituting Equation (2.3.32) into Equation (2.3.31) yields:

$$I_3\ddot{\theta} + c\dot{\theta} + k\theta + I_3\dot{r} + I_{ROT}\Omega p \sin\theta - {}^2T_z' \qquad (2.3.33)$$

$$= I_{ROT}\Omega \cos\theta \, q \qquad (2.3.34)$$

Now, if $H = I_{ROT}\Omega$, and

if it is assumed that the angular deflection θ is small (in practice, usually less than one degree), then

$$\cos\theta \simeq 1$$

$$\sin\theta \simeq \theta . \qquad (2.3.35)$$

Substitution of Equations (2.3.34) and (2.3.35) into (2.3.33) yields:

$$I_3\ddot{\theta} + c\dot{\theta} + k\theta + I_3\dot{r} + Hp\theta - {}^2T_z' = Hq . \qquad (2.3.36)$$

If it is assumed that p is small, then $p\theta$ may be neglected since it is a "second order" term. If the disturbance torque ${}^2T_z'$ is also negligible, Equation (2.3.36) becomes:

$$\boxed{I_3\ddot{\theta} + c\dot{\theta} + k\theta + I_3\dot{r} = Hq} . \qquad (2.3.37)$$

The Laplace transform of Equation (2.3.37) assuming zero initial conditions is:

$$\theta(s) = \underbrace{\left(\frac{H}{I_3 s^2 + cs + k}\right) q(s)}_{\text{Gyroscopic effect term}} - \underbrace{\left(\frac{I_3 s}{I_3 s^2 + cs + k}\right) r(s)}_{\text{Inertia effect term}} \qquad (2.3.38)$$

The first term on the right side of Equation (2.3.38) is associated with the angular momentum of the gyro, H, and vanishes as H approaches zero (the situation of a non-spinning rotor). The second term on the right side of Equation (2.3.38) is associated with the total inertia of the output axis; i.e., the motion of the vehicle about the output axis induces motion of the gimbal as caused by the total effective gimbal inertia, the damping term, and the spring term.

Equation (2.3.38) may be rewritten as

$$\theta(s) = \frac{H}{I_3 s^2 + cs + k}\left[q(s) - \frac{I_3 s}{H} r(s)\right] \qquad (2.3.39)$$

Equation (2.3.39) indicates that by increasing the angular momentum of the gyro, H, and decreasing the inertia of the gimbal - rotor about the output axis I_3, the gyro output can be made insensitive to vehicle motion about the "yaw" axis, $r(t)$. It is noteworthy that while low frequency angular accelerations in yaw may be adequately removed from the gyro output signal $\theta(t)$, the coefficient of $r(s)$, $I_3 s/H$, indicates that at high frequencies, such as vibrational frequencies of a vehicle structure, the gyro output can have a significant component produced at that frequency. This characteristic may require either placement of these sensing gyros in parts of a vehicle structure having low vibratory environments, or insulation from high frequency vibration by utilizing flexible "mounts", or use of filters in the electronics to remove high frequency signals.

If the term $\frac{I_3 s\, r(s)}{H}$ is negligible, Equation (2.3.39) becomes:

$$\frac{\theta(s)}{q_g(s)} \cong \frac{H}{I_3 s^2 + cs + k} \qquad (2.3.40)$$

The expression $\frac{H}{I_3 s^2 + cs + k}$ is the "gyro transfer function" for a single axis gyro.

The input axis equation is obtained from Equation (2.3.30) in a way completely analogous to the generation of the output axis Equation (2.3.37). Assuming that the products pr and $\dot{\theta}p$ are negligibly small, and that $\cos\theta = 1$, $H = I_{ROT}\Omega$, $\dot{\phi}_y \triangleq q_g$ and $\dot{\phi}_z \triangleq r$, Equation (2.3.30) becomes:

$$T_y = I_1 \ddot{\phi}_y + H(\dot{\theta} + \dot{\phi}_z) \qquad (2.3.41)$$

2.3.1.1.2 Rate Gyro Equations

From Equation (2.3.40),

$$\frac{\theta(s)}{q_g(s)} \cong \frac{H}{I_3 s^2 + cs + k}$$

In a rate gyroscope, the value of k is made large compared with $c\omega$ or $I\omega^2$, where ω is the frequency range of interest. Hence, the angular output is proportional to angular rate about the input axis - i.e.,

$$\frac{\theta(s)}{q_g(s)} \cong \frac{H}{k} \qquad (2.3.42)$$

2.3.1.1.3 Attitude Gyro Equations

If the coefficient "C" is large such that the value of $c\omega$ is much greater than $I_3\omega^2$ or k over the frequencies ω of interest, Equation (2.3.40) becomes

$$\frac{\Theta(s)}{\Phi(s)} \cong \frac{H}{Cs} \qquad (2.3.43)$$

If ϕ_Y is the time integral of $\Phi(t)$, then ϕ_Y represents an attitude change, and

$$\phi_Y(s) = \frac{\Phi(s)}{s} \qquad (2.3.44)$$

Equation (2.3.43) can thus be re-written

$$\frac{\Theta(s)}{\phi_Y(s)} \cong \frac{H}{C} \qquad (2.3.45)$$

Equation (2.3.45) indicates that the gyro output angular changes θ will be proportional to input axis angular changes ϕ_Y. Such a gyro is called an "attitude gyro," or an "integrating rate gyro," or a "proportional gyro."

2.3.1.1.4 Angular Integrating Gyro Equations

If the viscous damping term, c, and the spring rate k are very small compared with I_3, then Equation (2.3.40) can be written:

$$\frac{\Theta(s)}{\Phi(s)} = \frac{H}{I_3 s^2} \qquad (2.3.46)$$

or

$$\frac{\Theta(s)}{\phi_Y(s)} = \frac{H}{I_3 s} \qquad (2.3.47)$$

In this instance, the gyro output angular change $\Theta(t)$ is proportional to the time integral of the output angular change.

2.3.1.2 Translational Theory

The translational theory presented in this section consists of a development of the equations describing the linear accelerometer, which is subjected to external forces and accelerations, in general six degree of freedom dynamic space.

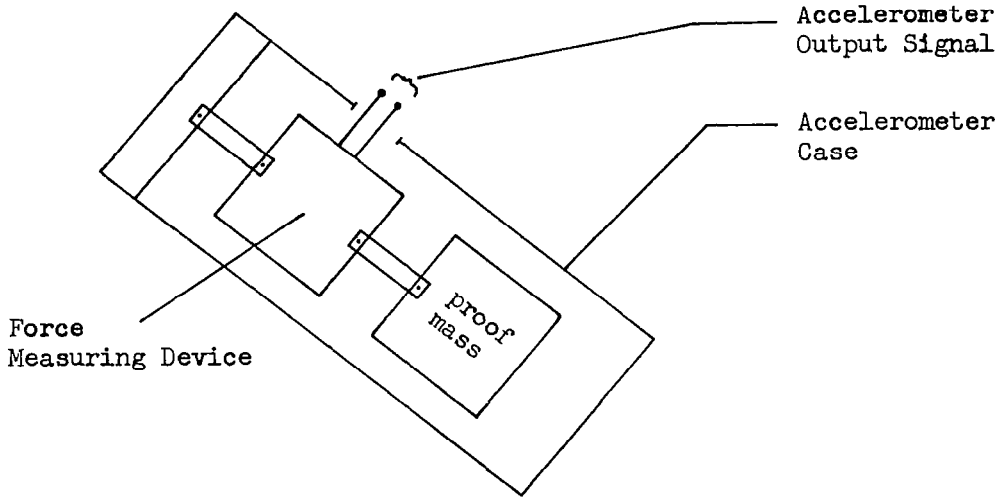
The development commences with a rigorous treatment of a single axis accelerometer, which is subsequently used as the basis for describing the dynamics of the integrating accelerometer or velocity meter, the true

accelerometer, and the double-integrating accelerometer or distance meter.

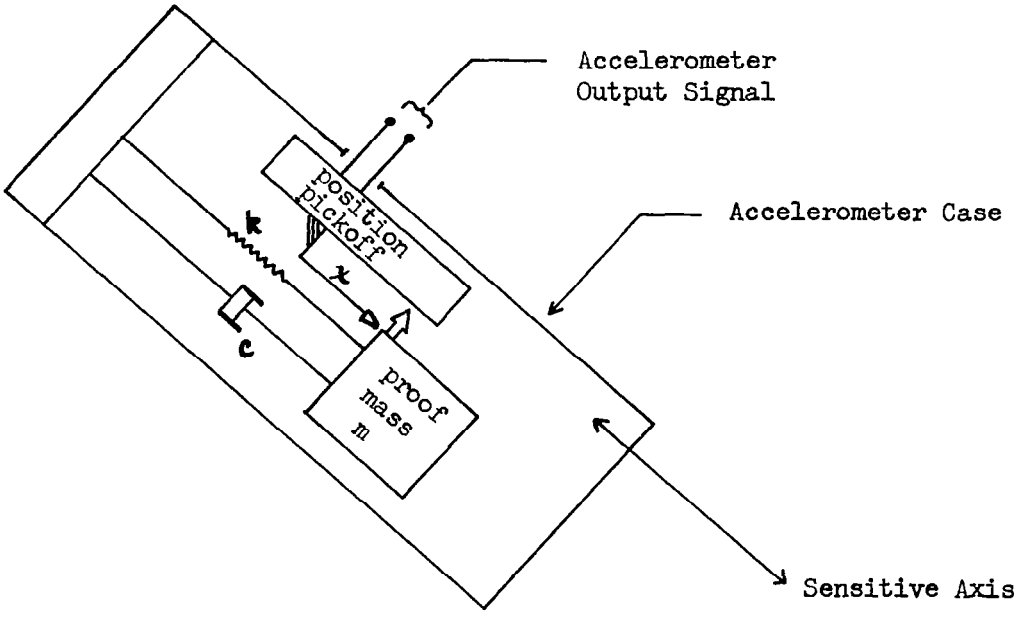
2.3.1.2.1 Single Axis Linear Accelerometer Equations

The device required for sensing translational motion is a linear accelerometer. Although many different designs of inertial linear accelerometers have evolved, all inertial accelerometers operate on the same fundamental principles.

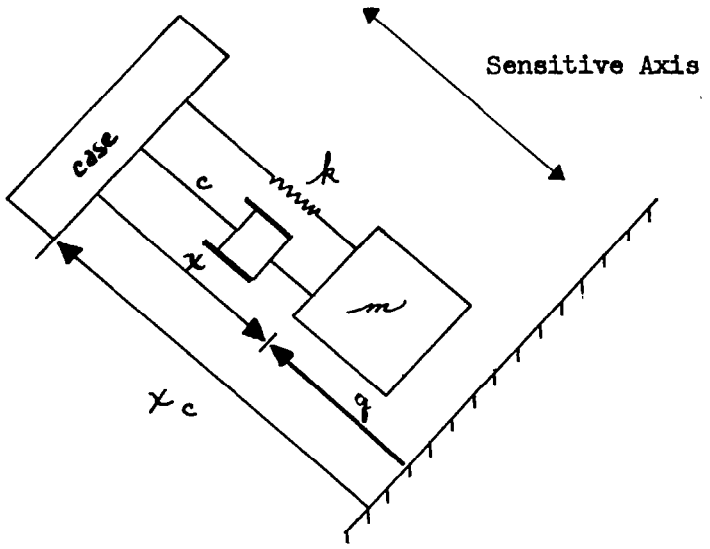
A linear accelerometer consists basically of a "proof mass" suspended by a force measuring device as shown below. The output of the force measuring device is the accelerometer output signal.



A number of different forms of linear accelerometers exist, but the simplest conceptually consists of a proof mass suspended by a linear spring in parallel with the linear viscous damper, as shown below. The accelerometer output signal is generated by an electrical pickoff which generates a signal proportional to "x", the mass deflection from its null position.



The spring and damper force the mass deflection x to deviate in known fashion (excluding errors) as a function of the forces acting upon m ; hence, the spring, damper, and position pickoff constitute the force measuring device.



Letting x_c = case position (inertial)

The equation of motion for the proof mass along the "sensitive axis" is:

$$\sum F = m\,\ddot{g} \tag{2.3.48}$$

$$k x + c\dot{x} + F_{gx} = m\,\ddot{g} \tag{2.3.49}$$

But

$$\ddot{g} = \ddot{x}_c - \ddot{x} \tag{2.3.50}$$

Where F_{gx} = gravitational force acting upon the proof mass in the sensitive axis direction, i.e., along the "x" axis.

Hence, from (2.3.49) and (2.3.50):

$$m\ddot{x}_c - m\ddot{x} = kx + c\dot{x} + F_{gx} \tag{2.3.51}$$

$$m\ddot{x} + c\dot{x} + kx = m\ddot{x}_c - F_{gx} \tag{2.3.52}$$

$$m\ddot{x} + c\dot{x} + kx = m(\ddot{x}_c - G_x) \tag{2.3.53}$$

where G_x is the x component of gravitational acceleration.

Taking the Laplace transform of Equation (2.3.53) (assuming zero initial conditions) yields:

$$(ms^2 + cs + k)\,X(s) = m\,(\ddot{x}_c(s) - G_x(s))$$

Hence:

$$\chi(s) = \left(\frac{m}{m s^2 + c s + k} \right) \ddot{X}_c(s) - G_x(s) \qquad (2.3.54)$$

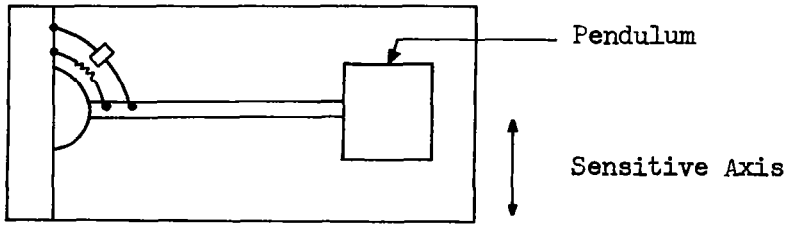
The ratio $\frac{m}{m s^2 + c s + k}$ is called the accelerometer transfer function.

Equation (2.3.54) indicates that an accelerometer output is a function of both gravitational acceleration $G_x(s)$ and the vehicle acceleration (or case acceleration) $\ddot{X}_c(s)$. Note that for zero case acceleration the accelerometer output becomes $\chi(s) = \left(\frac{-m}{m s^2 + c s + k} \right) G_x(s)$, or in the steady state (as $S \rightarrow 0$), approaches the value $-G_x/k$ for a true accelerometer, and becomes in essence a spring force balance. It is interesting to note that during "free fall" an accelerometer case has an acceleration $\ddot{X}_c = + G_x$, hence, substituting this into Equation (2.3.54) yields $\chi(s) = 0$ (assuming zero initial conditions) - i.e., during free fall an accelerometer will read zero, even though the vehicle or accelerometer case is being accelerated by the force of gravity. These brief examples serve to illustrate the fact that an accelerometer is incapable of sensing inertial acceleration in a gravitational force field. The significant consequence of this observation is that if the inertial acceleration of a vehicle is to be generated on board a vehicle, then for each of the three components of acceleration it is necessary to:

1. utilize an accelerometer to sense χ, which provides a measure of $(\ddot{X}_c - G_x)$, and
2. utilize a computer to compute the value of G_x, and generate the value of $\ddot{X}_c$ as $(\ddot{X}_c - G_x) + G_x$.

The details of this computation are presented in Section 2.3.2.3.

There are a number of mechanizations of true accelerometers and integrating accelerometers (or velocity meters). Although a linear suspension of a proof mass yields perhaps conceptually the simplest design, the suspension of the pendulous accelerometer design is more commonly used. In this design a pendulum is oriented so that the arm of the pendulum is perpendicular to the sensitive axis.



The pendulous gyro integrating accelerometer (PGIA), which is a form of velocity meter, utilizes a gyroscope having an unbalanced gimbal as the acceleration sensitive element.

2.3.1.2.2 True Accelerometer Equations

For a true accelerometer, i.e., a device which indicates acceleration directly, the spring term k dominates; hence,

$$\chi(s) \cong \frac{m}{k}\left(\ddot{\chi}_c(s) - G_K(s)\right) \qquad (2.3.55)$$

2.3.1.2.3 Integrating Accelerometer Equations

For an integrating accelerometer or velocity meter, the damping term predominates and:

$$\chi(s) \cong \frac{m}{cs}\left(\ddot{\chi}_c(s) - G_x(s)\right) \qquad (2.3.56)$$

2.3.1.2.4 Double Integrating Accelerometer Equations

For a distance meter the term $m s^2$ predominates and:

$$\chi(s) \cong \frac{1}{s^2}\left(\ddot{\chi}_c(s) - G_x(s)\right) \qquad (2.3.57)$$

2.3.2 Inertial Sensing Techniques

The basic inertial sensors used in generating navigation measurements, the gyrospope and the accelerometer, have been described in previous sections. In this section, the application of these inertial sensors to the problem of determining the position and velocity of a vehicle is presented.

The process of determining position and velocity from inertial sensors is called inertial navigation. Several significant features of inertial navigation warrant emphasis:

1. Unlike all other conventional navigation techniques, inertial navigation requires no external information to be received at the vehicle to determine position and velocity. Other techniques require the reception of either visual of radio information; or, in the case of radar, to radiate and receive back reflected information. Inertial navigation is unique in that it is selfcontained. It navigates by sensing motion of the vehicle and calculating the changes in position and velocity after an initial alignment.

2. By its very nature, the fundamental inertial measurements of the vehicle motion must be made on board the vehicle. Hence, if vehicle position and velocity information must be known at some point other than the vehicle, e.g. at an earth-fixed command and control center, then the use of inertial techniques requires communication of the inertial measurement information from the vehicle.

The presentation which follows is descriptive of only the most elementary forms of inertial navigation. The discussion has been formulated to emphasize the principles which form the basis for the techniques. In the past 25 years

a number of sophisticated inertial systems have evolved through development of aircraft, cruise missile, ballistic missile, boost vehicle, and spacecraft navigation systems. For more detailed descriptions of these systems the reader is referred to the references in this monograph and to the technical literature.

2.3.2.1 Inertial Platform Mechanization - Inertial Measuring Unit

The inertial platform mechanization of an inertial measuring unit is a sensor which provides acceleration signals that are resolved along known coordinates. It is conventionally a platform which is supported by a set of three (or four) gimbals, as shown in Figure 2.3.2. Usually the inner gimbal serves as the stable element and contains the three gyros and three accelerometers. The function of the three gyros is to maintain the orientation of the platform, usually to be inertially fixed or to maintain a locally level orientation. The functions of the three accelerometers are to generate total measured accelerations, i.e., to measure the vector sum of the inertial and gravitational acceleration. The accelerometers may be either "true accelerometers" or may be integrating accelerometers which generate the integral of inertial plus gravitational acceleration.

The orientation of the stable member is maintained by an attitude control loop around the platform such that angular errors sensed by the gyroscopes are amplified to drive the motors on the appropriate gimbals; the gimbals reorient the platform to drive the gyroscope angular error to zero. Major variations from the inertial measuring unit described above consist of four gimballed platforms, inside-out configurations (stable member as the outermost gimbal), and variations on the number and kind of accelerometers mounted on the platform itself.

The platform orientation loop will now be presented for a three gimbal platform maintained to be angularly fixed in inertial space.

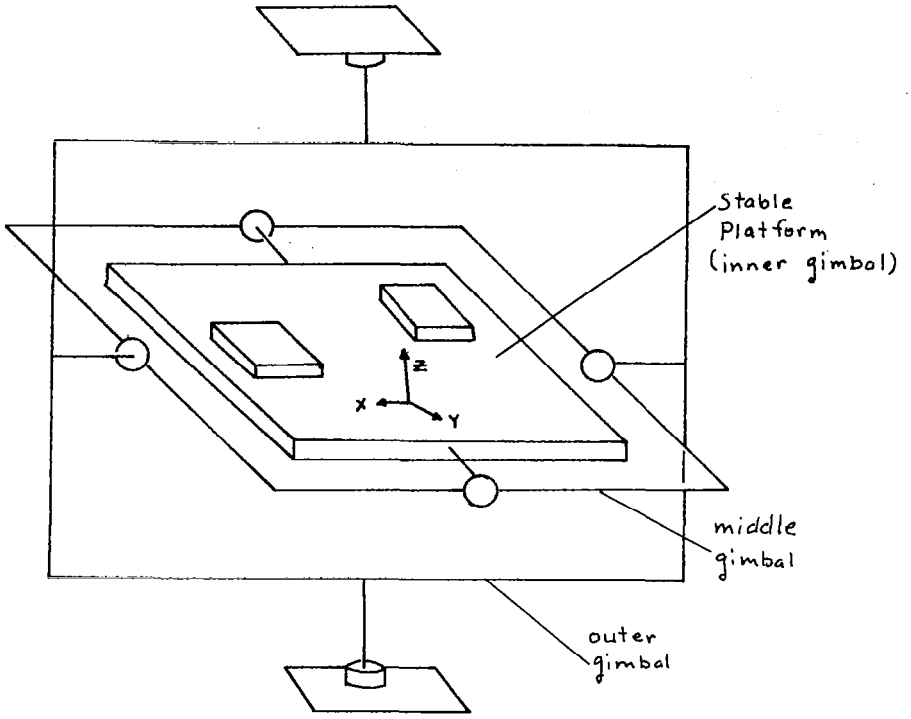


Figure 2.3.2

Three Gimbal Inertial Platform

From Equation (2.3.36), assuming $H_p\dot{\theta}$ negligible and taking the Laplace transform yields:

$$(I_3 s^2 + cs + k)\theta + I_3 \ddot{\phi}_z - {}^2T_z = H\dot{\phi}_y$$

Hence:

$$\theta = \frac{H s \phi_y + {}^2T_z - I_3 s^2 \phi_z}{I_3 s^2 + cs + k} \qquad (2.3.58)$$

Taking the Laplace transform of Equation (2.3.40) yields:

$$\phi_y = \frac{{}^2T_y - H s(\theta + \phi_z)}{I_1 s^2} \qquad (2.3.59)$$

It will now be assumed that a platform servo is used to correct any deviations of the platform by appropriately torquing the platform through torquers mounted along each gimbal. Let the platform servo transfer function be:

$$\frac{T_{sy}}{\theta} = -G_{T_y}(s) \qquad (2.3.60)$$

Hence, the total torque acting about the y axis will be

$$^{2}T_{Y} = T_{DY} + T_{sY} \qquad (2.3.61)$$

where T_{DY} = some disturbance torque

T_{sY} = servo torque

Equations (2.3.1), (2.3.2), (2.3.3), and (2.3.4) may be combined to form a control loop as shown in Figure (2.3.3).

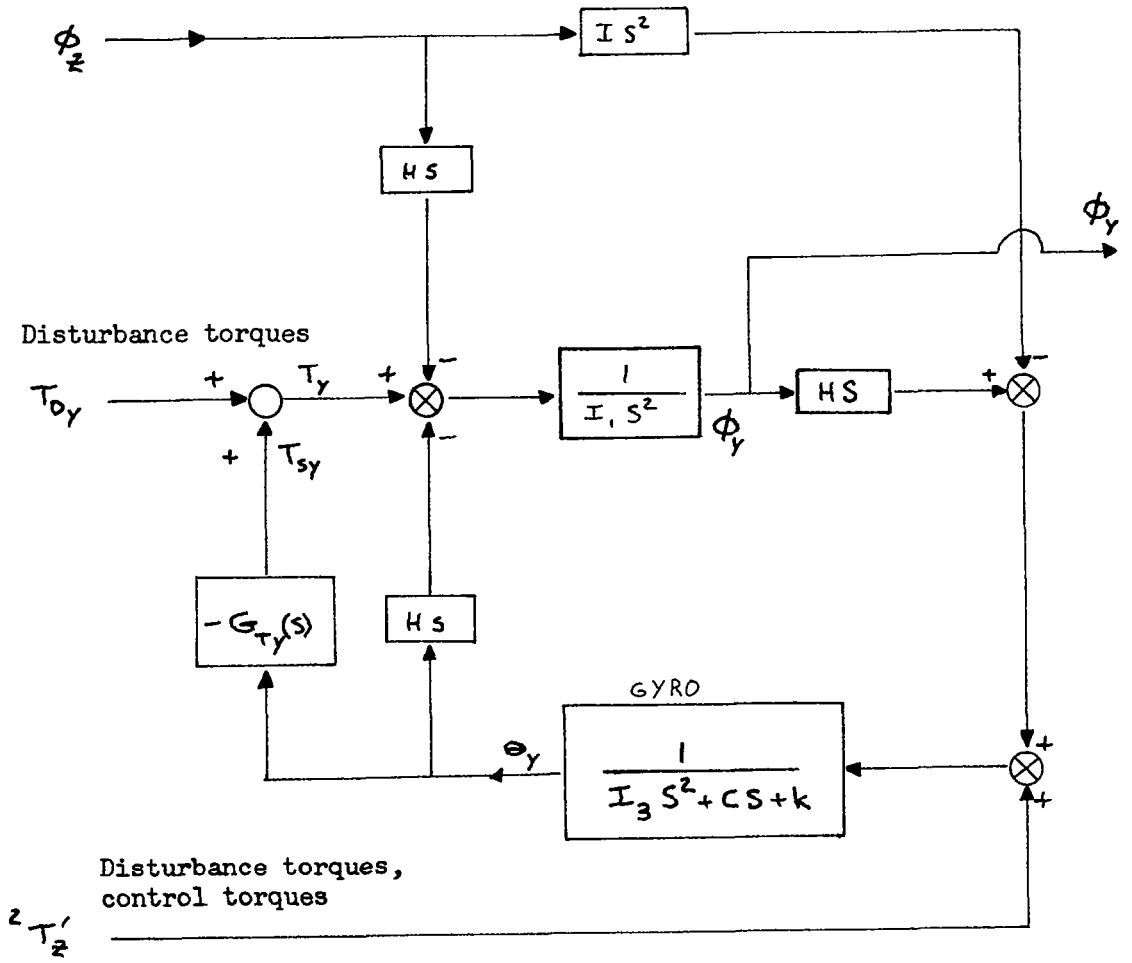


Figure 2.3.3

Single Axis Stabilization Loop for
the Y Axis of an Inertial Platform

requires simultaneous torquing of the middle and outer gimbals. This physical coupling of the middle and outer gimbals is usually done using resolvers.

The three gimballed platform just discussed is effective and practical when the middle gimbal angular deflection can be limited to approximately plus or minus 60 or 70 degrees. This angular deflection restriction may be too stringent a requirement upon vehicle motion. If the vehicle motion forces the middle gimbal of Figure 2.3.1 through 90°, the platform outer gimbal axis becomes aligned with the inner gimbal axis, and it becomes impossible for the platform (inner gimbal) to rotate about its Z axis. The consequence of this condition is that the platform cannot null attitude errors in this condition and may lose accuracy or "tumble," losing the attitude reference altogether. This situation is prevented by incorporating a fourth gimbal, as shown in Figure 2.3.4, which is automatically driven to try to maintain orthogonality between the second and third gimbals. The condition wherein the middle gimbal of a three gimballed platform becomes aligned with the outer gimbal is called gimbal lock.

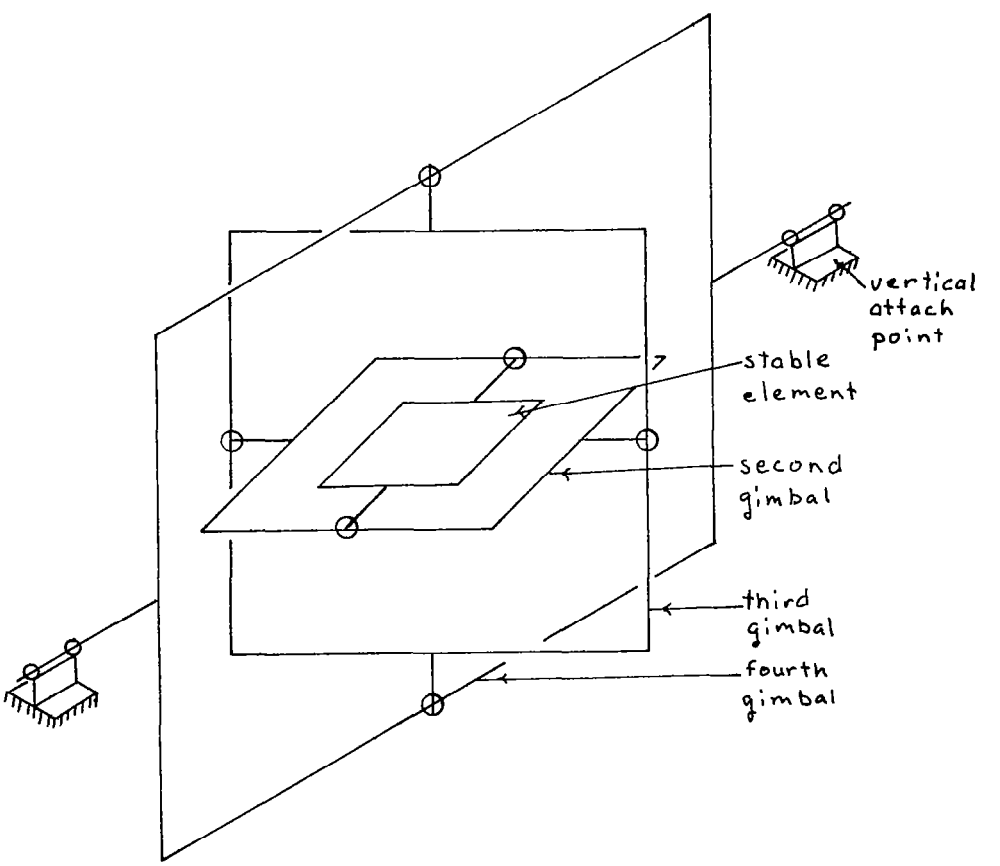


Figure 2.3.5 Four Gimbal Platform

2.3.2.2 Strap-Down System Mechanization - Inertial Measuring Unit

The strap-down system mechanization of an inertial measuring unit is similar to the inertial platform mechanization described above, in that it provides acceleration signals that are resolved along known coordinates. In addition, like the inertial platform mechanization, the strap-down system mechanization utilizes three gyroscopes and three accelerometers. Unlike the inertial platform mechanization, however, these accelerometers and gyroscopes of the strap-down mechanization are mounted directly to the vehicle frame. Although the accelerometers generate acceleration signals along vehicle axes, the integration of acceleration to determine inertial position and velocity requires that the components of acceleration be either explicitly or implicitly determined along the coordinates of a non-rotating reference frame. Many possible mechanizations may be used to perform the required integration. For simplicity and clarity of presentation, the coordinate frame for integration chosen for the following development will be a non-rotating inertial frame. This frame may be impractical or inconvenient for some particular applications; the reader is referred to the references for detailed discussion of the advantages of different coordinate frames for various applications.

Consider an inertial coordinate frame having orthogonal unit vectors $\underline{e}_X$ $\underline{e}_Y$ $\underline{e}_Z$, and a vehicle-fixed coordinate frame having orthogonal unit vectors $\underline{e}_x$, $\underline{e}_y$, $\underline{e}_z$. Assume that the accelerometer outputs are generated along the $\underline{e}_x$ $\underline{e}_y$ and $\underline{e}_z$ axes, i.e.,

$$\underline{A} = A_x \underline{e}_x + A_y \underline{e}_y + A_z \underline{e}_z \qquad (2.3.62)$$

Let the matrix of direction cosines from the vehicle fixed frame to the inertially fixed frame be T_{0102}, i.e.,

$$\underline{A} = T_{0102}\, \underline{A}$$

where

$$\underline{A} = A_x \underline{e}_x + A_y \underline{e}_y + A_z \underline{e}_z$$

and

$$T_{0102} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \qquad (2.3.63)$$

Since it is necessary to determine the inertial components of sensed vehicle acceleration, at least implicitly, in any mechanization capable of generating the position and velocity of a vehicle, it is necessary to define the equivalent of a mechanization technique for computing the matrix of direction cosines, T_{0102}.

The techniques for computing the direction cosine matrix, T0102, introduced in the above paragraphs, merit an introductory discussion here. Several methods of computing T0102 may be used. All methods require some form of angular information as a basis for computation. The customary angular information is the vehicle angular rate, as sensed by body mounted rate gyros, although variations of attitude gyros with periodic torquing, free gyros, etc., may be employed. Assume the vehicle angular rate information to be available, where the angular rate ${}^2\underline{\omega}$ is

$$ {}^2\underline{\omega} = {}^2p\,{}^2\underline{e}_x + {}^2q\,{}^2\underline{e}_y + {}^2r\,{}^2\underline{e}_z \qquad (2.3.64) $$

then two fundamental techniques for computing the transformation matrix T_{0102} may be utilized, based upon the availability of the angular rate components ${}^2p, {}^2q, {}^2r$, namely:

(1) Define a set of three Euler angles between the inertial and vehicle axes, ϕ, θ, ψ, and derive expressions of the form:

$$ \begin{aligned} \dot{\phi} &= f_\phi(p, q, r, \phi, \theta, \psi) \\ \dot{\theta} &= f_\theta(p, q, r, \phi, \theta, \psi) \\ \dot{\psi} &= f_\psi(p, q, r, \phi, \theta, \psi) \\ \phi(0) &= c_1 \\ \theta(0) &= c_2 \\ \psi(0) &= c_3 \end{aligned} \qquad (2.3.65) $$

These equations may be integrated to yield the values of $\phi(t)$, $\theta(t)$ and $\psi(t)$. The integration of this set of equations then defines the instantaneous value of the matrix T0102.

An alternate technique is to:

(2) Define the initial value of the transformation matrix, and compute each of the nine elements of the transformation matrix by direct integration.

Although technique (1) has the advantage of only three integrations versus the nine integrations of technique (2) above, technique (1) suffers the disadvantage that a discontinuity in the equations occurs if a particular Euler angle passes through a $\pm$ 90° value, while technique (2) suffers no such discontinuity. The equations for technique (2) may be developed as follows:

From equation (2.3.63),

$$\underline{e}_x = a_{11}\,{}^2\underline{e}_x + a_{12}\,{}^2\underline{e}_y + a_{13}\,{}^2\underline{e}_z \tag{2.3.66}$$

Thus, the time derivative of (2.3.66) (noting that $\dot{\underline{e}}_x = 0$ since $\underline{e}_x$ is inertially fixed) is:

$$\dot{a}_{11}\,{}^2\underline{e}_x + \dot{a}_{12}\,{}^2\underline{e}_y + \dot{a}_{13}\,{}^2\underline{e}_z = -\underline{\omega} \times (a_{11}\underline{e}_x + a_{12}\underline{e}_y + a_{13}\underline{e}_z) \tag{2.3.67}$$

Substituting (2.3.64) into (2.3.67), expanding, and equating corresponding components of the vector equation of (2.3.67) yields:

$$\begin{aligned}
\dot{a}_{11} &= {}^2q\,a_{13} - {}^2r\,a_{12} \\
\dot{a}_{12} &= {}^2r\,a_{11} - {}^2p\,a_{13} \\
\dot{a}_{13} &= {}^2p\,a_{12} - {}^2q\,a_{11}
\end{aligned} \tag{2.3.68}$$

Similarly for vectors $\underline{e}_y$ and $\underline{e}_z$, the following equations may be derived:

$$\begin{aligned}
\dot{a}_{21} &= {}^2q\,a_{23} - {}^2r\,a_{22} \\
\dot{a}_{22} &= {}^2r\,a_{21} - {}^2p\,a_{23} \\
\dot{a}_{23} &= {}^2p\,a_{22} - {}^2q\,a_{21}
\end{aligned} \tag{2.3.69}$$

$$\dot{a}_{31} = {}^2q\, a_{33} - {}^2r\, a_{32}$$

$$\dot{a}_{32} = {}^2r\, a_{31} - {}^2p\, a_{33}$$

$$\dot{a}_{33} = {}^2p\, a_{32} - {}^2q\, a_{31}$$

Equations (2.3.68) and (2.3.69) indicate that, given the initial values of $\{a_{ij}(0)\}$, all future values of $\{a_{ij}(t)\}$ may be computed provided the vehicle body angular rates ${}^2p(t), {}^2q(t), {}^2r(t)$ are sensed.

The techniques for generating $\{a_{ij}(t)\}$ using Euler angular rates are simply derived by summing the components of ${}^2p$, ${}^2q$ and ${}^2r$ along the $\dot{\phi}$, $\dot{\theta}$ and $\dot{\psi}$ axes of rotation. The derivation of this is quite lengthy and would not supplement the discussion. The interested reader is referred to References 8, 9, and 10.

With the determination of the $\{a_{ij}\}$, the values of ${}^1A_x, {}^1A_y, {}^1A_z$ may be computed using equation (2.3.63). The only remaining step in the development is to compute inertial position and velocity from ${}^1A_x, {}^1A_y, {}^1A_z$. This computation is derived in the next section.

2.3.2.3 Position and Velocity Computations and Schuler Tuning

The use of an inertially stabilized platform, as discussed in section 2.3.2.1, permits the three platform-mounted orthogonal accelerometers to directly sense the components of acceleration (inertial minus gravitational) in a non-rotating inertial reference frame. Denote this measured acceleration as:

$$\underline{{}^1A} = A_x\, {}^1\underline{e}_x + A_y\, {}^1\underline{e}_y + A_z\, {}^1\underline{e}_z \qquad (2.3.70)$$

where ${}^1\underline{e}_x, {}^1\underline{e}_y, {}^1\underline{e}_z$ are inertially fixed (non-rotating)unit vectors

The use of a strap-down system mechanization, as described in section (2.3.2.2), now permits the computation of the components of acceleration ${}^1A_x, {}^1A_y, {}^1A_z$, as indicated in equation (2.3.70).

The inertial computation of position and velocity may now be carried out as follows:
Since the response of the accelerometers is usually extremely rapid compared with the comparatively steady acceleration of the vehicle, the sensed acceleration $\underline{{}^1A}$ may be well approximated by the equation:

$$\underline{{}^1A} = \ddot{\underline{R}} - \underline{G} \qquad (2.3.71)$$

where $$\underline{G} = -\frac{g}{R}\underline{R} = G_x\, {}^1\underline{e}_x + G_y\, {}^1\underline{e}_y + G_z\, {}^1\underline{e}_z$$

and

$$g \triangleq \left|\frac{\mu}{R^2}\right|$$

$$\underline{R} = R_x \underline{e}_x + R_y \underline{e}_y + R_z \underline{e}_z$$

The vector $\underline{R}$ is the position of the vehicle in inertial coordinates.

Equation (2.3.71) is the vector equation equivalent of equation (2.3.55) derived in section 2.3.1.2, except that case acceleration $\ddot{x}_c(s)$ has been replaced with the vehicle acceleration. Note that the g in the above equation is the magnitude of the local acceleration.

The objective of this section, to indicate the inertial techniques by which the position and velocity of a vehicle may be determined, may now be developed; i.e., the position $\underline{R}(t)$ and velocity $\underline{V}(t)$ of the vehicle may be computed from:

$$\underline{V}(t) = \dot{\underline{R}}(t) = \int_{t_1=0}^{t} \ddot{\underline{R}}(t_1)\, dt_1 + \underline{V}(0)$$

$$\underline{R}(t) = \int_{t_2=0}^{t}\int_{t_1=0}^{t_2} \ddot{\underline{R}}(t_1)\, dt_1\, dt_2 + \underline{V}(0)\int_{t_1=0}^{t} dt_1 + \underline{R}(0) \qquad (2.3.72)$$

Equation (2.3.71) and (2.3.72) may be combined to form the block diagram indicated in Figure 2.3.6.

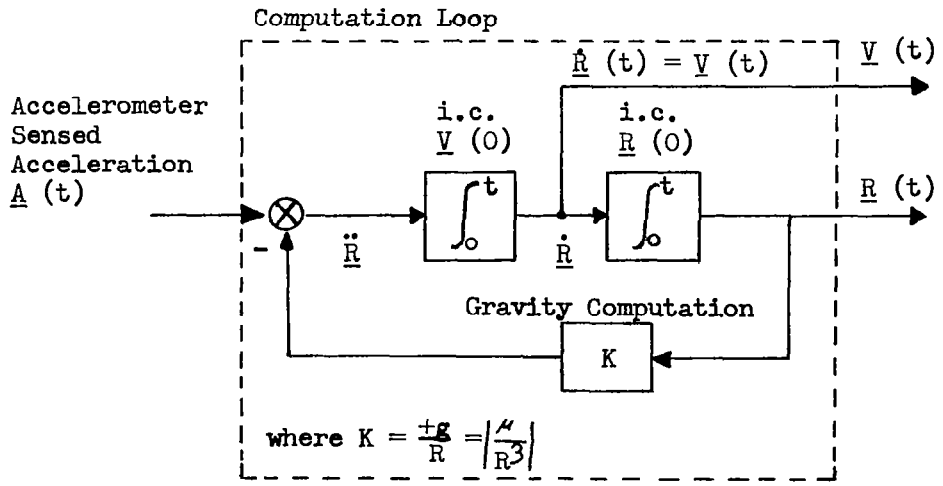


Figure 2.3.6

Mechanization of Position and Velocity Computation in an Inertial Measuring Unit

If K is slowly varying or constant in Figure 2.3.6, then the transfer function of the components of $\underline{R}(t)$ and $\underline{V}(t)$ may be written as:

$$\frac{R_x}{A_x} = \frac{R_y}{A_y} = \frac{R_z}{A_z} = \frac{1}{s^2+K}$$
$$\frac{V_x}{A_x} = \frac{V_y}{A_y} = \frac{V_z}{A_z} = \frac{s}{s^2+K} \qquad (2.3.73)$$

Equations (2.3.73) display an undamped natural frequency from the characteristic equation of the computation system:

$$\omega = (K)^{1/2} = \left(\frac{g}{R}\right)^{1/2} \qquad (2.3.74)$$

The frequency ω is called the Schuler frequency, and at the surface of the earth has a value of 84.4 minutes.

The fact that the pure inertial system has an undamped frequency characteristic produces difficulties with respect to error propagation; these difficulties are discussed in section 2.5.

2.4 OBSERVATION-STATE VECTOR RELATIONSHIP

As discussed previously, the objective of making navigation observations is to acquire sufficient information to permit accurate determination of the vehicle state vector, i.e., position and velocity. An essential relationship that must accompany the measured values of the navigation observables (where essential means essential in the sense that the vehicle state cannot be computed without it) is a set of general equations relating the vehicle state vector to the value of the navigation observable. The set of general equations relating the vehicle state vector to the value of the navigation observables is simply an identity relationship for the inertial systems previously discussed (for the inertial measuring unit and strap-down inertial systems described in paragraph 2.3.2), since through integration the output of the inertial systems discussed is the vehicle state vector. Also, in the case of radar range and range-rate under radiation sensors, again the relationship of the vehicle state vector to the measured value is an identity relationship for the radially oriented components of position and velocity sensed by the range and range-rate radars, respectively. The situation for the angular measuring sensors and for occultation measurements is, however, entirely different.

Several difficulties with the use of angular measurements as navigation observables are noted: Firstly, the measurement of angle is strictly a function of position (discounting the negligible effects of aberration); hence, velocity dependent information can only be acquired by making angular observations separated in time. Secondly, a single measurement of angle provides information only that the position of the vehicle is on a surface of revolution formed by rotating a circle about a chord line segment formed by the navigation objects. If one of the navigation objects is a star, the length of AB is very large, and vehicle motion, near some near body B (within the solar system), occurs very close to B, in which case the actual surface can be well approximated by the familiar navigation cone of position. If the navigation objects A and B are on near bodies, as, for example, on the lunar and earth surfaces for mid-cislunar flight, the conical approximation for the surface is invalid although linear approximations may still be used. The navigation problem becomes complex if a set of angular measurements must be processed in an exact fashion to generate the vehicle position and velocity. Even in the case of generating vehicle position from three exact independent simultaneously measured angles, the situation is unwieldy, since an exact computation must carry out the following steps: (1) derive the equations for each surface which is the locus of points producing the value of each of the three angular measurements, (2) compute the three lines of intersections which the three surfaces form, and (3) compute the point or points of intersection of the three intersection lines.

The navigation computations are also complex if an exact (non-linearized) method of solution is followed for the generation of vehicle position and velocity from planet occultation-time measurements. Any planet occultation-time measurement requires only that the vehicle position must lie on an approximately conical surface generated by rotating a ray from the radiation source

around the surface of the planet (which degenerates to a cylindrical surface for star radiation sources). Since occultation measurements cannot in general be made simultaneously, the exact computation of vehicle position and velocity can only be carried out by utilizing the vehicle equations of motion.

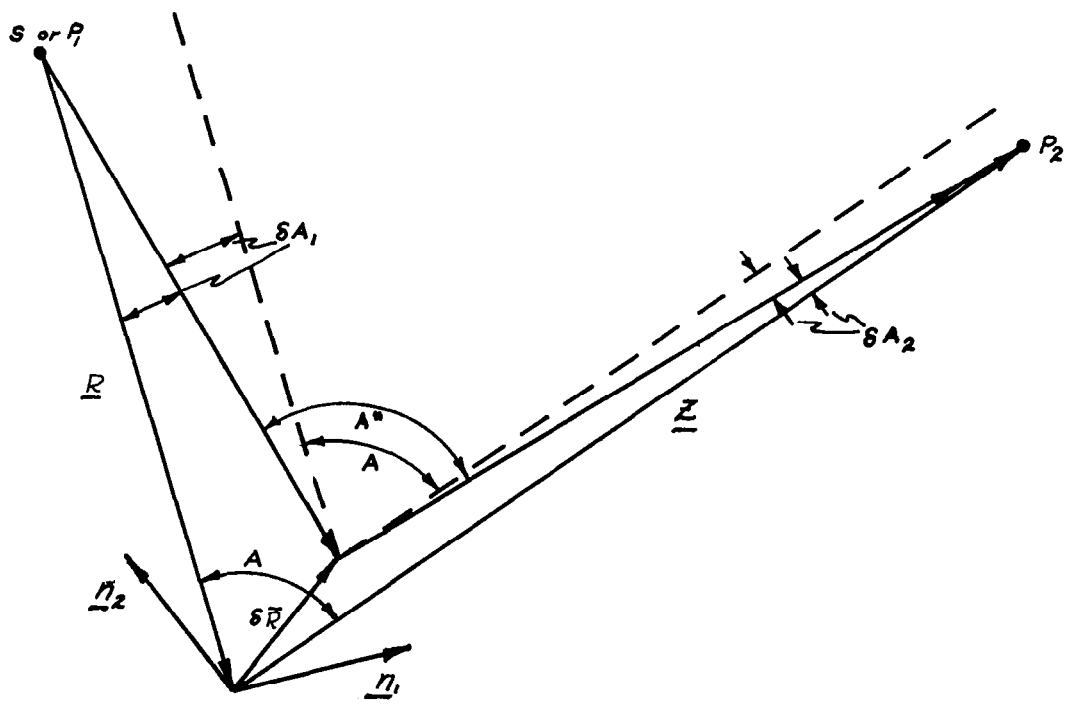
The state of the art in navigation has not advanced to the point where these more exact techniques are utilized, due **principally to the fact that** sufficiently excellent accuracy in the navigation computations can be obtained by using simplified linear relationships between the navigation observables and the vehicle state vector.

The following subsections present the simplified linear relationships for a single fix which are excellent approximations provided deviations of vehicle position from a "nominal" position are sufficiently small.

Acknowledgement is given to Richard H. Battin (Reference 4) whose book was of significant assistance in the preparation of this section.

2.4.1 Sun Planet Measurement

Let $\underline{R}$ be the nominal position vector of the vehicle with respect to the sun (s) and $\underline{Z}$ the vector from a nominal position of the vehicle to a planet (P_2). It is assumed that the positions of the sun and planet are known accurately. Then, a deviation in vehicle position from the nominal position produces a change in the sun-planet angle. The following sketch defines the geometry to be used in the analysis:



If A and A* are the nominal and measured angles respectively between S and P_2, it can be seen from the above sketch that

$$A^* = A + \delta A = A + \delta A_1 + \delta A_2 \qquad (2.4.1)$$

The unit vectors $\underline{\eta}_1$ and $\underline{\eta}_2$ are defined to be perpendicular to $\underline{R}$ and $\underline{Z}$, respectively, and in the plane determined by $\underline{R}$ and $\underline{Z}$. If the dot product of $\delta\underline{R}$ is now taken with $\underline{\eta}_1$ and $\underline{\eta}_2$ it is seen that

$$\delta\underline{R} \cdot \underline{\eta}_1 = |\underline{R}|\ \delta A_1$$

$$\delta\underline{R} \cdot \underline{\eta}_2 = |\underline{Z}|\ \delta A_2 \qquad (2.4.2)$$

Equations (2.4.2) are valid expressions providing δA_1, and δA_2 are very small, or alternately, $|\delta R| << |\underline{R}|$ and $|\delta\underline{R}| << |\underline{Z}|$. The very nature of the analysis automatically satisfies these conditions because $\delta\underline{R}$ is a small position deviation from the nominal position.

From Equation (2.4.2) it is seen that

$$\delta A_1 = \frac{\delta\underline{R} \cdot \underline{\eta}_1}{|\underline{R}|}$$

$$\delta A_2 = \frac{\delta\underline{R} \cdot \underline{\eta}_2}{|\underline{Z}|} \qquad (2.4.3)$$

Hence

$$\delta A = \delta A_1 + \delta A_2 = \frac{\delta\underline{R} \cdot \underline{\eta}_1}{|\underline{R}|} + \frac{\delta\underline{R} \cdot \underline{\eta}_2}{|\underline{Z}|} \qquad (2.4.4)$$

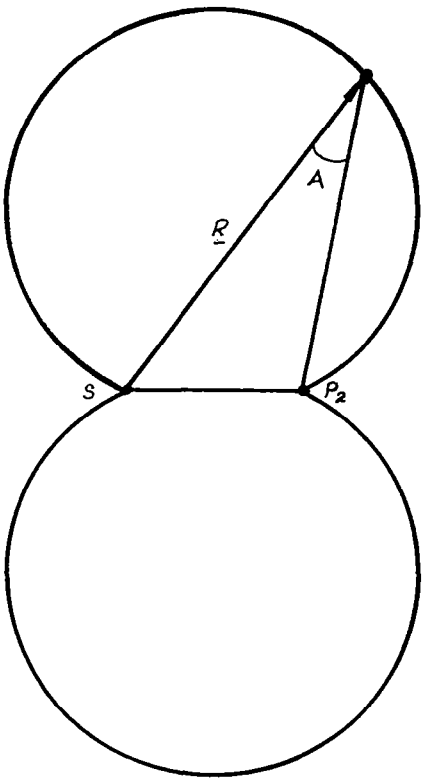
or

$$\delta A = \left(\frac{\underline{\eta}_1}{R} + \frac{\underline{\eta}_2}{Z}\right) \cdot \delta\underline{R} \qquad (2.4.5)$$

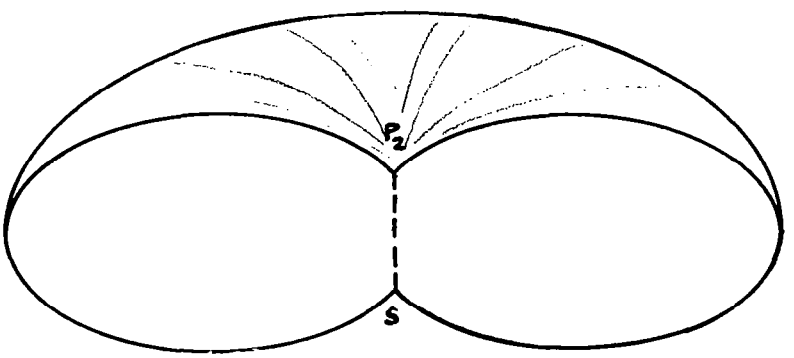
where $R = |\underline{R}|$ and $Z = |\underline{Z}|$.

Equation (2.4.5) is the final result that relates the angular measurement and position deviations for the sun-planet measurement.

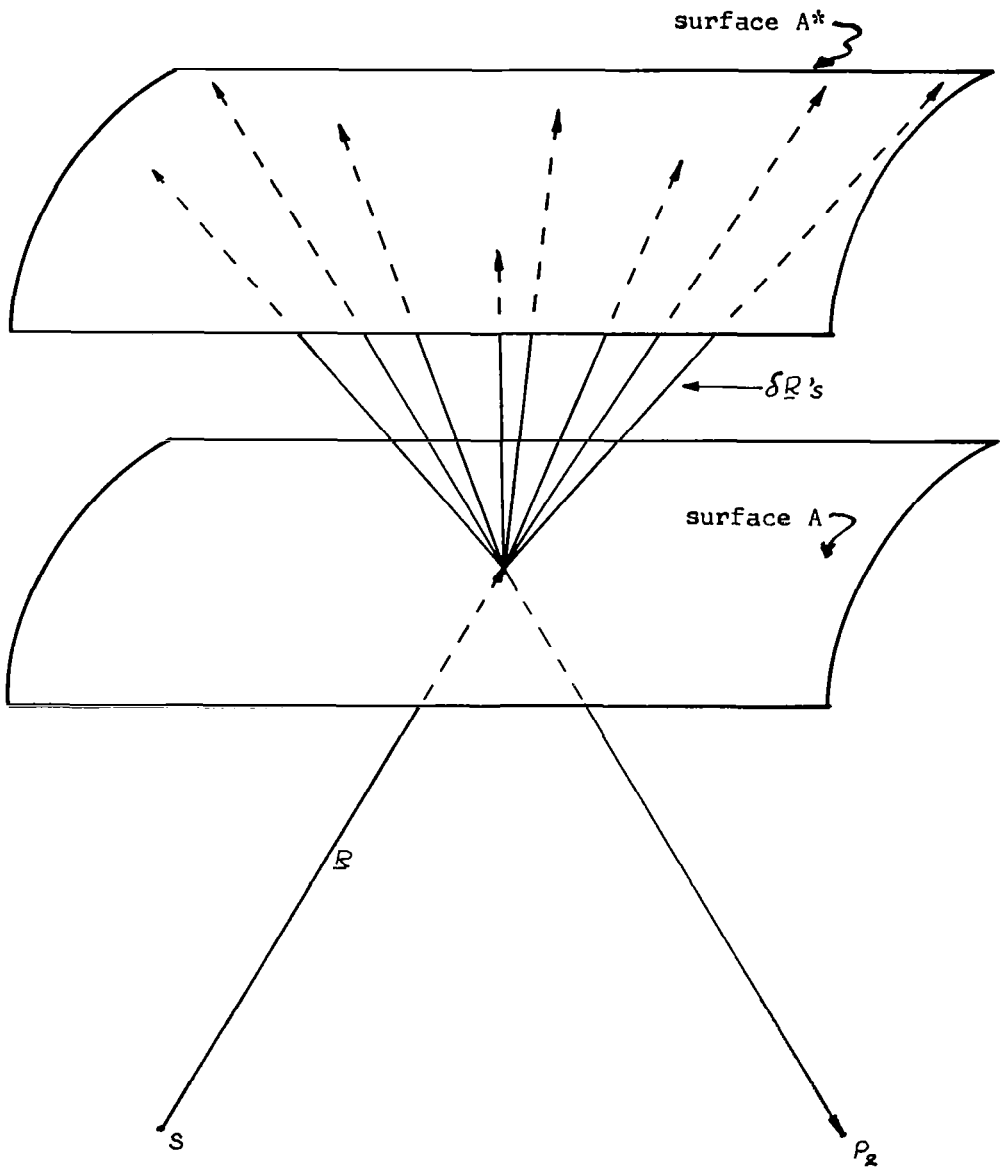
The above result can be thought of as the general expression for a measurement of the angle between two bodies of a finite distance away. Consider the locus of points defined by such a measurement. The following sketch illustrates such a locus in two dimensions.



Elementary plane geometry can be used to determine the locus as part of a circle. having a chord defined by the line S P_2 . The locus in three dimensions can be thought of as the surface of revolution of the circle about the chord. Part of this surface is shown below.



It is seen that for every angle that is measured between the two bodies, there is a corresponding surface of revolution. Hence, when a deviation in the angle is measured, it can be thought of as the measure of the separation between the two surfaces. For regions close to a nominal position $\underline{R}$, this separation thus defines a family of $\delta\underline{R}$ vectors, all of which satisfy Equation (2.4.5). The following sketch gives a pictorial representation of the family of $\delta\underline{R}$'s.



A new surface, "Surface A*" can be associated with the measured angle A*. "Surface A" is the surface determined by the nominal Angle, A. In effect, δA is an indirect measure of the separation of these surfaces. It is now apparent that the information defining a unique $\delta \underline{R}$ must come from at least three independent measurements.

2.4.2 Planet-Planet Measurement

The planet-planet measurement is analytically similar to the sun-planet measurement. All equations and sketches are exactly the same if the vector $\underline{R}$ is interpreted as the position vector from some planet (P_1) to the vehicle.

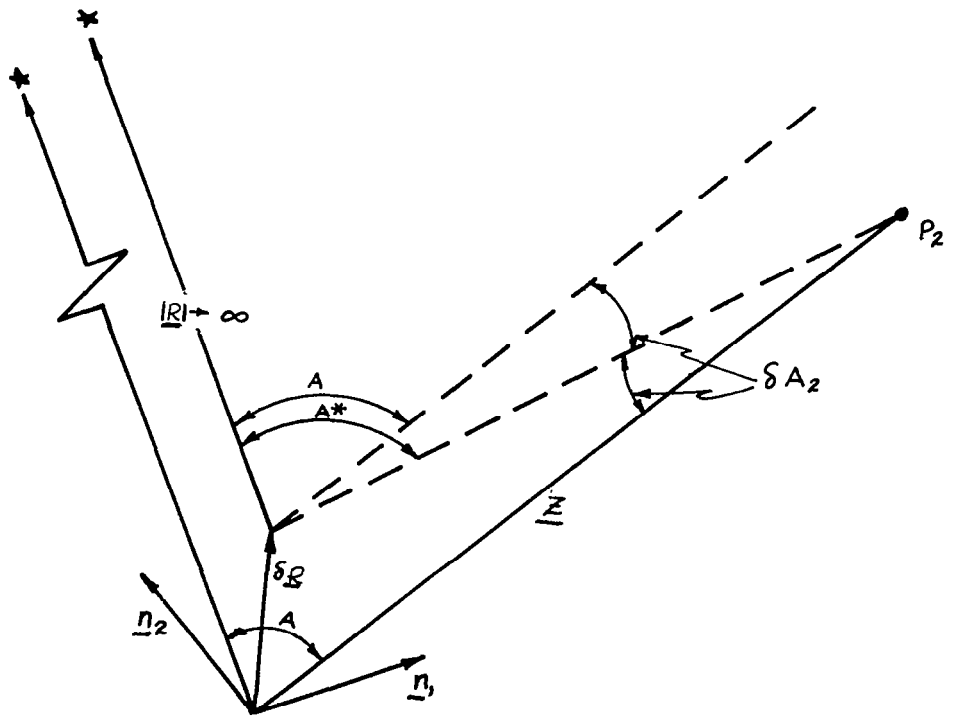
2.4.3 Planet-Star Measurement

The result of the analysis for the sun-planet measurement can be extended to the planet-star measurment by letting $|\underline{R}| \to \infty$. Under these conditions, the lines of sight to the star are parallel and the angular deviation $\delta A_1 \to 0$. Hence, the angular relations reduce to:

$$A^* = A + \delta A_2 \tag{2.4.6}$$

where $$\delta A = \delta A_2$$

The following sketch shows the geometry of the planet-star case.



Equation (2.4.5) now becomes

$$\delta A = \lim_{R \to \infty} \left(\frac{\eta_1}{R} + \frac{\eta_2}{Z} \right) \cdot \delta \underline{R} = \frac{\eta_2 \cdot \delta \underline{R}}{Z} \qquad (2.4.7)$$

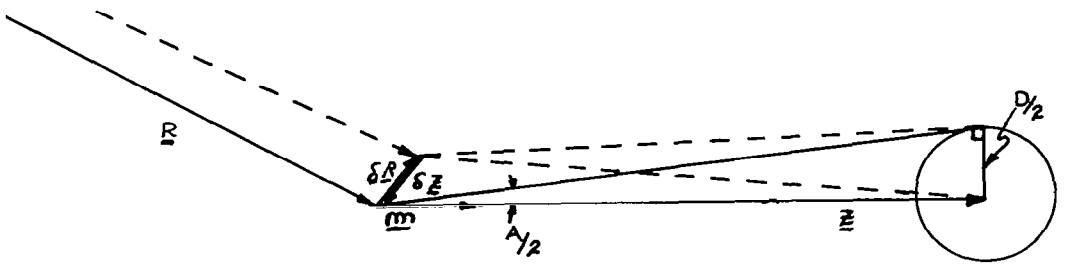
2.4.4 Sun-Star Measurement

Just as the planet-planet analysis was extended to the planet-star case, it can also be extended to the sun-star case by letting $|\underline{Z}| \to \infty$. The result is

$$\delta A = \lim_{Z \to \infty} \left(\frac{\eta_1}{R} + \frac{\eta_2}{Z} \right) \cdot \delta \underline{R} = \frac{\eta_1 \cdot \delta \underline{R}}{R} \qquad (2.4.8)$$

2.4.5 Planet Diameter Measurement

The following sketch defines the geometry to be used in the planet diameter measurement.



Let A be the diameter angle of a planet viewed from some nominal position in space. $\underline{Z}$ is position of the planet with respect to the vehicle and $\underline{R}$ is some position vector of the vehicle. For nominal conditions it is seen that

$$\sin\left(\frac{A}{2}\right) = \frac{D/2}{|\underline{Z}|} = \frac{D}{2Z} \qquad (2.4.9)$$

Thus, for small deviations from nominal, a relationship between δA and δZ is obtained as follows:

$$\cos\left(\frac{A}{2}\right) \frac{dA}{2} = -\frac{D}{2} \frac{dZ}{Z^2} \qquad (2.4.10)$$

or

$$dA = -\frac{D\, dZ}{Z^2 \cos\left(\frac{A}{2}\right)} \qquad (2.4.11)$$

The scalar quantity dZ can be thought of as the change of the vector in the radial direction to the planet center. If $\underline{M}$ is a unit vector in the $\underline{Z}$ direction, the scalar change in $|\underline{Z}|$ is

$$\delta Z = \underline{m} \cdot \delta \underline{Z} \qquad (2.4.12)$$

This equation can now be written in terms of δR by noting that the position of the planet and the origin of $\underline{R}$ are assumed to be known very accurately. Hence

$$\underline{R} + \underline{Z} = \underline{Q} \qquad \text{(a known vector)} \qquad (2.4.13)$$

Small deviations in vehicle position thus obey the equation

$$\delta \underline{R} + \delta \underline{Z} = \delta \underline{Q} = \underline{0}.$$

$\delta \underline{Q}$ is the null vector because there is no deviation in $\underline{Q}$ due to the vehicle being off the nominal position.

Thus
$$\delta \underline{R} = - \delta \underline{Z} \qquad (2.4.14)$$

and Equation (2.4.12) becomes

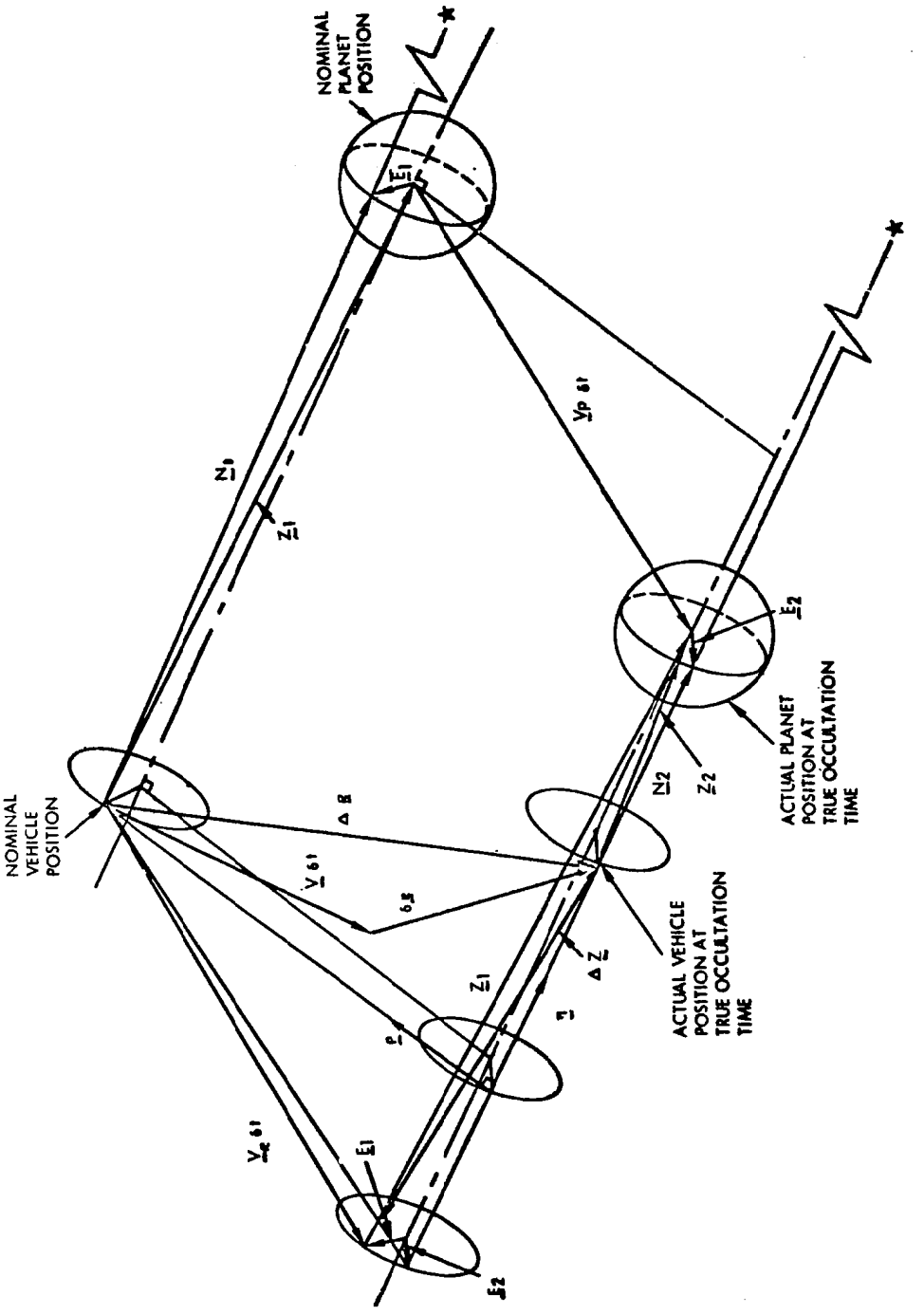
$$\delta Z = - \underline{m} \cdot \delta \underline{R} \qquad (2.4.15)$$

The diameter angle deviation can now be written in terms of the position deviation vector by combining Equations (2.4.11) and (2.4.15).

$$\delta A = \frac{D \, \underline{m} \cdot \delta \underline{R}}{Z^2 \cos\left(\frac{A}{2}\right)} \qquad (2.4.16)$$

2.4.6 Star Occultation Measurement

As a vehicle passes through a nominal trajectory, the time for which a star is aligned with the edge of a planet and the vehicle can be predicted. This event is called a star occultation. The observer in a vehicle that is slightly off the nominal trajectory will witness the alignment at a slightly different time. This information can be used in a manner that is analogous to previously mentioned measurements in order to determine the position deviation of the vehicle. The following sketch shows the nominal and actual positions of the vehicle and planet during an occultation measurement.


NOMINAL PLANET POSITION
NOMINAL VEHICLE POSITION
ACTUAL VEHICLE POSITION AT TRUE OCCULTATION TIME
ACTUAL PLANET POSITION AT TRUE OCCULTATION TIME

The vectors $\underline{E}_1$ and $\underline{E}_2$ define the position of the point of tangency of the line of sight and center of the planet at the nominal and actual positions, respectively. If the occultation time deviation is δt, both the planet and the vehicle will move a certain distance from their nominal positions. In addition to the movement in time δt, the vehicle will be displaced from the nominal position by an additional amount $\delta\underline{R}$ due to the fact that it was not on the nominal trajectory originally. If the velocity of the vehicle is $\underline{V}$, it will travel a distance $\underline{V}\,\delta t$ during the occultation time. As a result, the total displacement of the vehicle from the nominal position is

$$\Delta\underline{R} = \delta\underline{R} + \underline{V}\,\delta t \qquad (2.4.17)$$

which is the sum of the original position deviation and the movement during the occultation time.

Just as the vehicle moves during the occultation time, so does the planet. If the planet has a velocity $\underline{V}_p$, it moves a distance $\underline{V}_p\,\delta t$ during the occultation time. The vectors $\underline{N}_1$ and $\underline{N}_2$ are defined to be along the lines of sight from the point of observation to the point of tangency. Now the following vector equation may be written

$$\delta\underline{R} + \underline{V}\,\delta t + \underline{N}_2 - \underline{E}_2 - \underline{V}_p\,\delta t + \underline{E}_1 - \underline{N}_1 = \underline{0} \qquad (2.4.18)$$

The unit vector $\underline{\rho}$ is now defined to be in the plane determined by the two lines of sight and perpendicular to $\underline{N}_1$ (and $\underline{N}_2$). The dot product of $\underline{\rho}$ and equation (2.4.18) may now be taken.

$$\underline{\rho}\cdot\delta\underline{R} + \underline{\rho}\cdot\underline{V}\,\delta t + \underline{\rho}\cdot\underline{N}_2 - \underline{\rho}\cdot\underline{E}_2 - \underline{\rho}\cdot\underline{V}_p\,\delta t + \underline{\rho}\cdot\underline{E}_1 - \underline{\rho}\cdot\underline{N}_1 = \underline{0} \qquad (2.4.19)$$

But $\underline{\rho}\cdot\underline{N}_1 = \underline{\rho}\cdot\underline{N}_2 = 0$ because they are orthogonal.

Thus

$$\underline{\rho}\cdot(\underline{V}-\underline{V}_p)\,\delta t = \underline{\rho}\cdot\underline{E}_2 - \underline{\rho}\cdot\underline{E}_1 - \underline{\rho}\cdot\delta\underline{R} \qquad (2.4.20)$$

or

$$\delta t = \frac{\underline{\rho}\cdot(\underline{E}_2-\underline{E}_1) - \underline{\rho}\cdot\delta\underline{R}}{\underline{\rho}\cdot\underline{V}_R} \qquad (2.4.21)$$

where

$\underline{V}_R = \underline{V}-\underline{V}_P$ = relative velocity of vehicle with respect to the planet.

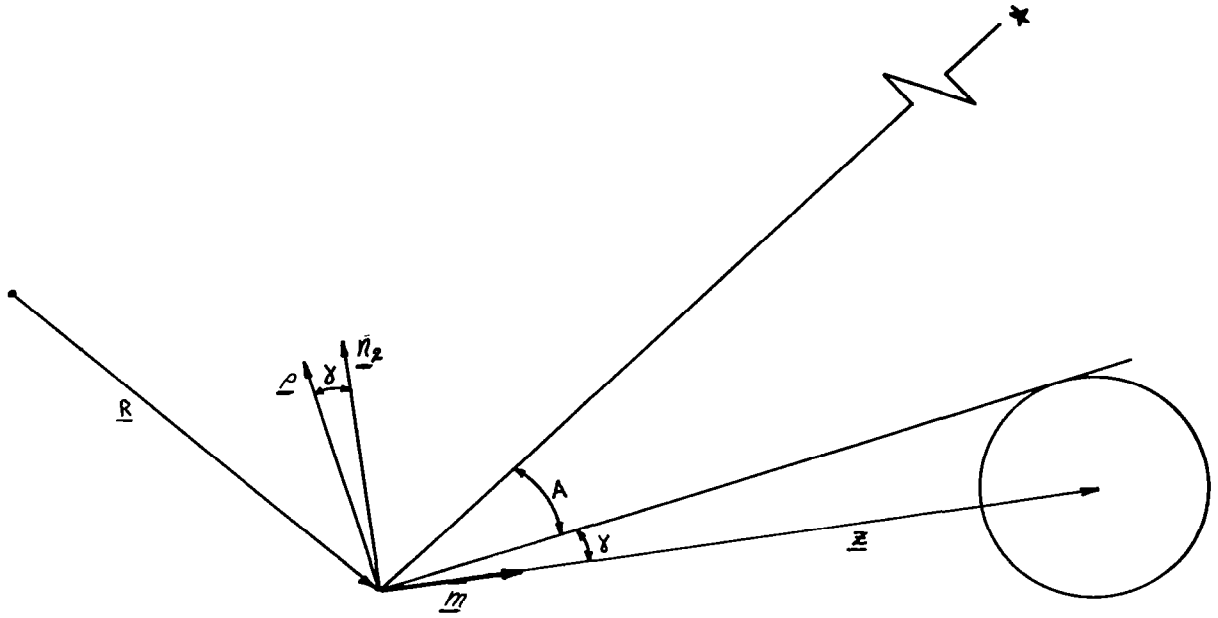
Equation (2.4.21) is a very interesting result. If the point of tangency to the planet during the actual occultation is the same as the predicted nominal position, then the $\underline{\rho}\cdot(\underline{E}_2-\underline{E}_1)$ term is zero because $\underline{E}_2 = \underline{E}_1$ and Equation (2.4.21) reduces to

$$\delta t = \frac{-\underline{\rho}\cdot\delta\underline{R}}{\underline{\rho}\cdot\underline{V}_R} \qquad (2.4.22)$$

Equation (2.4.22) is the standard occultation equation that is normally seen in the literature (Reference 4). It must be realized that the absence of the $\underline{\rho} \cdot (\underline{E}_2 - \underline{E}_1)$ term introduces an error in the expression for cases in which the points of tangency are not the same.

2.4.7 Star Elevation Measurement

Let A be the angle from a star to the edge of a planet and γ the planet diameter angle measurement as shown in the following sketch:



As before, the vector $\underline{R}$ defines the position of the vehicle and $\underline{Z}$ is a vector from the vehicle to the planet center. $\underline{m}$ is a unit vector in the $\underline{Z}$ direction and $\underline{\eta}_2$ is a unit vector perpendicular to $\underline{m}$ and in the plane of the diameter measurement. $\underline{\rho}$ is also in the diameter measurement plane but is perpendicular to the line to the planet's edge. Now consider the angle $A+\gamma$ as the angle in a planet-star measurement. Making use of Equation (2.4.7), the following relationship is obtained:

$$\frac{\underline{\eta}_2 \cdot \delta \underline{R}}{Z} = \delta(A+\gamma) = \delta A + \delta\gamma \qquad (2.4.23)$$

Again, making use of a previous derivation on the planet diameter measurement, it is seen that

$$\delta\gamma = \frac{D(\underline{m} \cdot \delta \underline{R})}{2 Z^2 \cos\gamma} \qquad (2.4.24)$$

The combination of Equations (2.4.23) and (2.4.24) yields

$$\delta A = \frac{\underline{\eta}_2 \cdot \delta \underline{R}}{Z} - \frac{D(\underline{m} \cdot \delta \underline{R})}{2 Z^2 \cos\gamma} \qquad (2.4.25)$$

or

$$\delta A = \left(\frac{\underline{n}_2}{z} - \frac{D\,\underline{m}}{2z^2 \cos\gamma} \right) \cdot \delta \underline{R} \qquad (2.4.26)$$

Remembering that $\frac{D}{2z} = \sin\gamma$, Equation (2.4.26) becomes

$$\delta A = \frac{1}{z} \left(\underline{n}_2 - \tan\gamma\, \underline{m} \right) \cdot \delta \underline{R} \qquad (2.4.27)$$

But, the vectors $\underline{\rho}$, $\underline{n}$, and $\underline{m}$ are linearly dependent so that $\underline{\rho}$ can be expressed as a linear combination of $\underline{n}_2$ and $\underline{m}$.

$$\underline{\rho} = \cos\gamma\, \underline{n}_2 - \sin\gamma\, \underline{m} \qquad (2.4.28)$$

or

$$\underline{\rho} = \cos\gamma \left(\underline{n}_2 - \tan\gamma\, \underline{m} \right) \qquad (2.4.29)$$

so

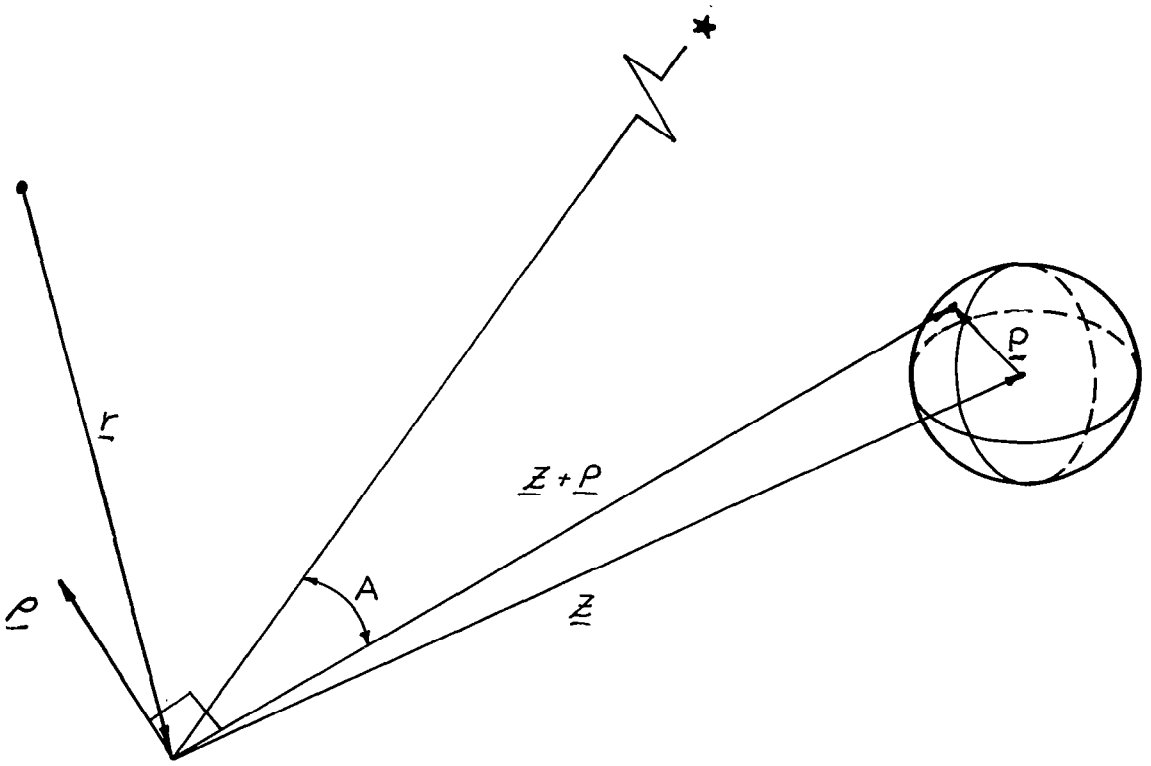
$$\frac{\underline{\rho}}{\cos\gamma} = \underline{n}_2 - \tan\gamma\, \underline{m} \qquad (2.4.30)$$

The combination of Equations (2.4.27) and (2.4.30) yields

$$\delta A = \frac{1}{z} \left(\frac{\underline{\rho}}{\cos\gamma} \right) \cdot \delta \underline{R} = \frac{\underline{\rho} \cdot \delta \underline{R}}{z \cos\gamma} \qquad (2.4.31)$$

2.4.8 Star-Landmark Measurement

The geometry of the star-landmark system is defined in the following sketch. $\underline{P}$ is the position vector of the landmark with respect to the planet centers.



If the figure is examined closely, it is apparent that this case is analogous to the star-planet measurement where $\underline{Z} + \underline{P}$ and $\underline{\rho}$ can be considered to be analogous to $\underline{Z}$ and η_2, respectively in the star-planet case.

Hence

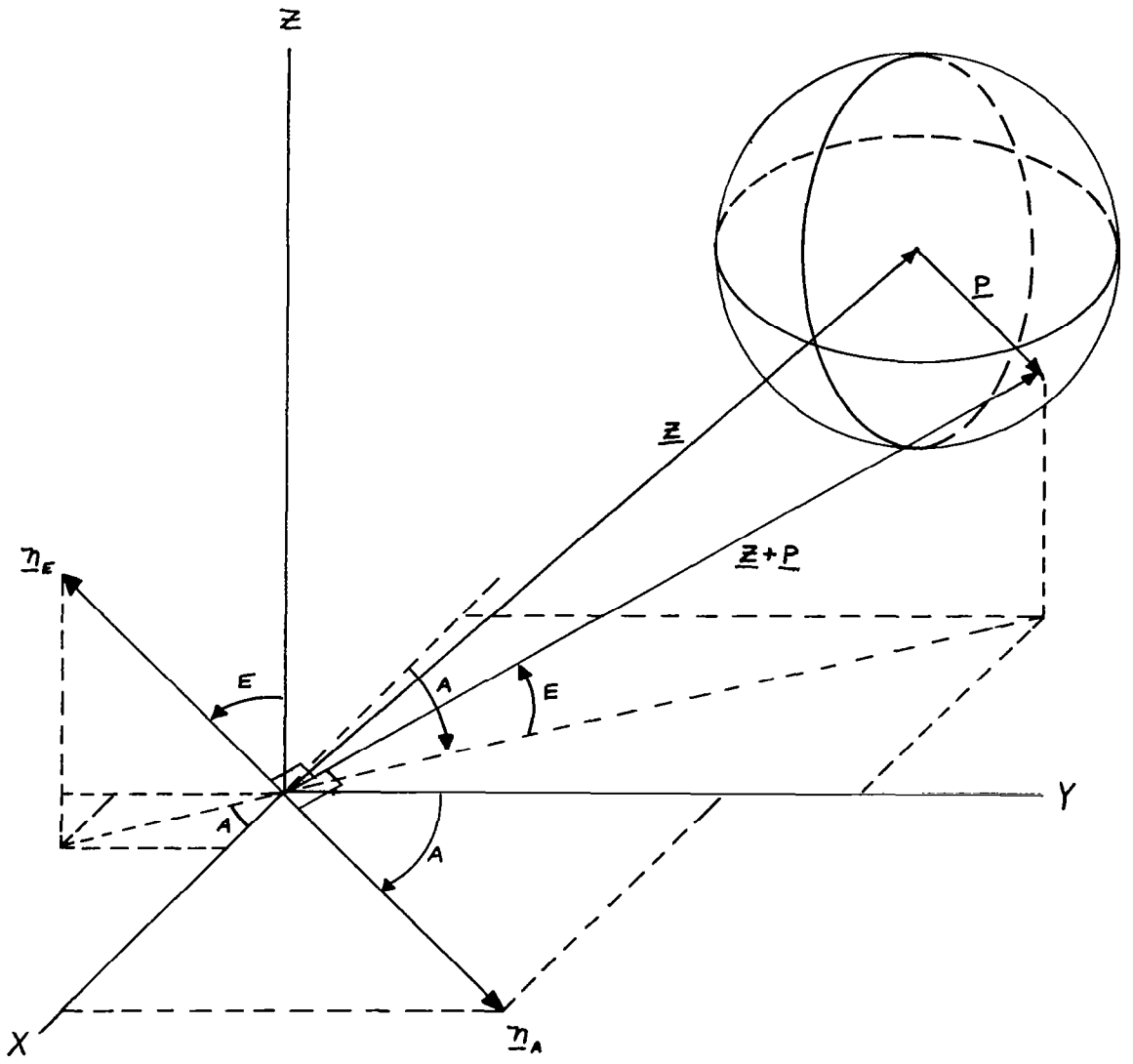
$$\delta A = \frac{\eta_2 \cdot \delta R}{Z} \qquad (2.4.32)$$

which becomes

$$\delta A = \frac{\underline{\rho} \cdot \delta \underline{R}}{|\underline{Z} + \underline{P}|} \,. \qquad (2.4.33)$$

2.4.9 Elevation and Azimuth Angle Measurement

Consider the angular measurement of a planet landmark with respect to some platform that is stabilized with respect to two stars. The angular measurement may be expressed in terms of azimuth and elevation angles "A" and "E", respectively. The following sketch shows the geometry of the analysis:



From previous work on the star-landmark measurement, the information contained in such a measurement is the component of $\delta \underline{R}$ in a direction that is perpendicular to the $\underline{Z} + \underline{P}$ vector in the plane determined by the star and the vector $\underline{Z} + \underline{P}$. The stabilized platform can be thought of as defining the direction to two stars. Hence, the information contained in the azimuth and elevation measurement is represented by the components of $\delta \underline{R}$ in the direction of $\underline{\eta}_a$ and $\underline{\eta}_e$, respectively.

$$\delta E = \frac{\underline{\eta}_e \cdot \delta \underline{R}}{|\underline{Z} + \underline{P}|} \tag{2.4.34}$$

$$\delta A = \frac{\underline{\eta}_a \cdot \delta \underline{R}}{|\underline{Z} + \underline{P}|} \tag{2.4.35}$$

The unit vectors $\underline{\eta}_a$ and $\underline{\eta}_e$ can be expressed in terms of rectangular cartesian components as follows:

$$\underline{\eta}_a = \sin A \, \underline{i} + \cos A \, \underline{j} \tag{2.4.36}$$

$$\underline{\eta}_e = \sin E \cos A \, \underline{i} - \sin E \sin A \, \underline{j} + \cos E \, \underline{k} \tag{2.4.37}$$

Rewriting equations (2.4.34) and (2.4.35) and using equations (2.4.36) and (2.4.37), it is seen that

$$\delta A = \frac{(\sin A \, \underline{i} + \cos A \, \underline{j} + 0 \, \underline{K}) \cdot \delta \underline{R}}{|\underline{Z} + \underline{P}|} \tag{2.4.38}$$

$$\delta E = \frac{(\sin E \cos A \, \underline{i} - \sin E \sin A \, \underline{j} + \cos E \, \underline{k}) \cdot \delta \underline{R}}{|\underline{Z} + \underline{P}|} \tag{2.4.39}$$

or in matrix form

$$\begin{bmatrix} \delta A \\ \delta E \end{bmatrix} = \begin{bmatrix} \frac{\sin A}{|\underline{Z} + \underline{P}|} & \frac{\cos A}{|\underline{Z} + \underline{P}|} & 0 \\ \frac{\sin E \cos A}{|\underline{Z} + \underline{P}|} & -\frac{\sin E \sin A}{|\underline{Z} + \underline{P}|} & \frac{\cos E}{|\underline{Z} + \underline{P}|} \end{bmatrix} \begin{bmatrix} \underline{i} \cdot \delta \underline{R} \\ \underline{j} \cdot \delta \underline{R} \\ \underline{k} \cdot \delta \underline{R} \end{bmatrix} \tag{2.4.40}$$

But

$$\begin{bmatrix} \underline{i} \cdot \delta \underline{R} \\ \underline{j} \cdot \delta \underline{R} \\ \underline{k} \cdot \delta \underline{R} \end{bmatrix} = \underline{R}$$

so

$$\begin{bmatrix} \delta A \\ \\ \delta E \end{bmatrix} = H \cdot \delta \underline{R} \qquad (2.4.41)$$

where

$$H = \frac{1}{|\underline{Z} + \underline{P}|} \begin{bmatrix} \sin A & \cos A & 0 \\ \sin E \cos A & -\sin E \sin A & \cos E \end{bmatrix} \qquad (2.4.42)$$

2.5 OBSERVATION ERRORS

A significant portion of the errors in the final values of the navigation estimates of position and velocity, even after smoothing of redundant data, is caused by errors in the navigation measurements themselves. Often the most difficult requirements to be met by the navigation system are its accuracy requirements. Hence, the evaluation of errors is a fundamental task in the analyses of navigation systems. It is thus appropriate to consider in detail the analytical methods and techniques which may be applied to the various kinds of errors which occur in navigation measurements. In this section the various error sources which contaminate the navigation observation measurements are introduced and analytical techniques for their evaluation are presented. The discussion consists of a treatment of radiation sensor errors and inertial sensor errors.

2.5.1 Radiation Sensor Errors

The two types of basic radiation sensors which may be categorized as current state of the art are sensors which are sensitive to electromagnetic radiation in the visible region and sensors which are sensitive to radiation in the radio frequency region. The sensors which are sensitive to the optical frequencies are for the most part line-of-sight instruments whose function is to measure the angular relationship of the line-of-sight to either a platform reference or a second line-of-sight. Since the measurements for this kind of sensor are angular measurements, the errors which dominate the total instrument inaccuracy are most frequently the mechanical read-out devices. These errors are non-electromagnetic in nature and will not be discussed here. Another type of error discussed affects optical instruments, however. These errors are associated with uncertainties in the path of the radiation through space before arriving at the instrument, and include two types of errors: atmospheric refraction and horizon-induced diffraction. The other categories of radiation sensors discussed consist of sensors sensitive to the radio frequencies, notably radar frequencies. The errors discussed describe all the major sources of inaccuracies in the optical and radar sensors.

2.5.1.1 Signal-to-Noise Ratio

The signal-to-noise ratio is fundamental to the discussion of radiation noise, and hence is the logical point of departure leading to a discussion of the random errors in radiation phenomena which are discussed in the sections immediately following.

Since noise is the chief factor limiting the sensitivity of an observation sensor, it is necessary to obtain some means of describing it quantitatively. Noise is unwanted electromagnetic energy which interferes with the ability of a sensor to detect the wanted signal. It may originate within the sensor itself, or it may enter along with the desired signal. If the sensor were to operate in a perfectly noise-free environment so that no external sources of noise accompanied the desired signal, and if the sensor itself were so perfect that it did not generate any excess noise, there would still exist an unavoidable component of noise generated by the thermal motion of the conduction electrons in the electronic components of the sensor. This is called thermal noise, or Johnson noise, and is directly proportional to the temperature of the ohmic portions of the circuits and proportional to the bandwidth of the sensor circuitry. The available thermal noise power generated by a receiver of bandwidth B_n (in cycles per second) at a temperature T (degrees Kelvin) is equal to

$$\text{available thermal-noise power} = K T B_n \qquad (2.5.1)$$

where K = Boltzmann's constant = 1.38×10^{-23} joule/deg.

If the temperature T is taken to be 290°K, which corresponds approximately to room temperature (62°F), the factor kT is 4×10^{-21} watt/cps of bandwidth. If the sensor circuitry were at some other temperature, the thermal noise power would be correspondingly different.

It should be noted that the bandwidth B_n of Equation (2.5.1) is not the 3 - db, or half-power, bandwidth commonly employed by electronic engineers. It is an integrated bandwidth and is given by

$$B_n = \frac{\int_{-\infty}^{\infty} |H(f)|^2 df}{|H(f_o)|^2} \qquad (2.5.2)$$

where H(f) = frequency-response characteristic of sensor circuitry

fo = frequency of maximum response (usually occurs at midband)

When H(f) is normalized to unity at midband (maximum response frequency), H(fo) = 1. The bandwidth B_n is called the noise bandwidth and is the bandwidth of an equivalent rectangular filter whose noise-power output is the same as the filter with characteristic H(f).

The noise power in practical sensors is usually much greater than can be accounted for by thermal noise alone. The additional internal noise components

are due to mechanisms other than the thermal agitation of the conduction electrons. A discussion of the additional noise is given in Section 2.2.1.4, "Noise," which also discusses sources of external noise.

Two useful quantities that are frequently used to describe the noise in a system are the signal-to-noise ratio and the noise figure. The signal-to-noise ratio is defined to be the ratio of the power of the desired signal-to-the-noise power.

$$\frac{S}{N} = \frac{\text{signal power}}{\text{noise power}}$$

This ratio is a measure of the relative amount of noise accompanying a signal. The lower the signal-to-noise ratio, the more difficult it becomes to detect the signal. A measure of the noise characteristics of a sensor can be obtained by comparing the noise power out of a practical sensor to that of an ideal sensor at some standard temperature. This ratio is called the noise figure, F_n. If the ideal sensor has no other than Johnson noise, its noise output is $kT B_n G_a$. Therefore,

$$F_n = \frac{N_o}{kT_o B_n G_a} = \frac{\text{noise out of practical sensor}}{\text{noise out of ideal sensor at standard temp., T}}$$

where N_o = noise output from sensor

G_a = available gain

T_o = standard temperature = 290°K

Now, the noise output of a sensor may be considered to be equal to the product of the thermal-noise power obtained from an ideal sensor and a factor called the noise figure, F. The noise figure may be interpreted as a measure of signal-to-noise ratio degradation as the signal passes through the sensor. This can be seen if the gain is thought of as the ratio of the output signal power to the input signal power (S_o/Si) and the quantity $kT_o B_n$ is taken to be the input noise N_i in an ideal sensor. Hence

$$F_n = \frac{N_o}{kT_o B_n (S_o/S_i)} = \frac{S_i}{kT_o B_n (S_o/N_o)} = \frac{(S_i/N_i)}{(S_o/N_o)} \qquad (2.5.4)$$

Using the above equation, the input signal power may be expressed as

$$S_i = kT_o B_n F_n S_o \qquad (2.5.5)$$

Assuming that the input noise is $kT_o B_n$, the minimum detectable signal power corresponding to the minimum signal-to-noise ratio at the output can be written as

$$S_{MIN} = kT_o B_n F_N (S_o/N_o)_{MIN} \qquad (2.5.6)$$

Hence, an estimate of the minimum detectable signal of a sensor can be obtained from the noise bandwidth, noise figure and the minimum signal-to-noise ratio of the output.

Another convenient parameter used to express the noise characteristic of a sensor is the effective noise temperature. Using this parameter is equivalent to using the noise figure as is shown by the following analysis.

The definition of noise figure is the ratio of the total noise out of a sensor compared to the noise out of an ideal sensor at 290°K.

Hence,

$$F = \frac{kT_o B_n G_a + \Delta N}{kT_o B_n G_a} \qquad (2.5.7)$$

where ΔN is the additional noise introduced by the sensor (not necessarily at 290°K).

Rewriting Equation (2.5.7) as

$$F = 1 + \frac{\Delta N}{kT_o B_n G_a} \qquad (2.5.8)$$

and defining the effective temperature, Te, as that temperature at the input which would account for the noise ΔN at the output ($\Delta N = kTe\ B_n\ G_a$), the noise figure can be written as

$$F = 1 + \frac{kT_e B_n G_a}{kT_o B_n G_a} = 1 + \frac{T_e}{T_o} \qquad (2.5.9)$$

or

$$T_e = (F-1)\,T_o \qquad (2.5.10)$$

It should be emphasized that the effective noise temperature of a sensor is not necessarily equal to the physical noise temperature of the sensor input. Rather, it is a convenient parameter to use in order to represent the noise beyond that exhibited by the ideal sensor at 290°K, i.e.,

$$\Delta N = k\,T_e\,B_n\,G_a\ . \qquad (2.5.11)$$

Both the noise figure and effective temperature are measures of the additional noise introduced by a sensor due to the fact that it is not at 290°K and other sources of internal noise.

Since there are many sources of noise in a system and its environment, it sometimes becomes necessary to combine all the sources into an equivalent source which can be represented as an equivalent system noise figure or noise temperature. Usually the system can be broken down into a cascade of components each having its own noise figure, F, gain, G, and noise bandwidth, B_n. It should be noted that each component could be any transmission medium that has noise such as the atmosphere as well as any electronic equipment used to

transmit the signal. If the medium considered has losses, the gain may be less than one, in which case a loss factor, L, ($L = 1/G$) is sometimes used.

The overall noise figure for a cascade system will now be derived for two components in cascade and generalized to N components in cascade. Consider two components in cascade, each with the same noise bandwidth but with different noise figures and available gain.

Let F_1, G_1, be the noise figure and available gain, respectively, of the first component and F_2, G_2, be similar parameters for the second component. The problem is to find F_o, the over-all noise figure for the cascaded system. From the definition of noise figure, the output noise N_o of the two components in cascade is

$$N_o = F_o G_1 G_2 k T_o B_n \tag{2.5.12}$$

where $G_1 G_2$ is the total cascade gain.

This must be equal to the noise from component 1 at the output of component 2 plus the noise introduced by component 2. The total noise from component 1 at its own output is $kT_o B_n G_1 F_1$. This includes both Johnson noise and additional imperfect sensor internal noise (ΔN_1). This noise becomes $kT_o B_n G_1 G_2 F_1$ at the output of component 2. The additional noise introduced by component 2 can be found from Equation (2.5.8) as

$$\Delta N_2 = (F_2 - 1) k T_o B_n G_2 \tag{2.5.13}$$

Hence,

$$N_o = k T_o B_n F_1 G_1 G_2 + (F_2 - 1) k T_o B_n G_2 = F_o k T_o B_n G_1 G_2 \tag{2.5.14}$$

Dividing the above expression by $k T_o B_n G_1 G_2$ yields

$$F_o = F_1 + \frac{F_2 - 1}{G_1} \tag{2.5.15}$$

This expression can be generalized to the total system noise figure F_o as a function of the noise figures of N cascaded components and their respective gains as follows:

$$F_o = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \cdots + \frac{F_N - 1}{G_1 G_2 \cdots G_{N-1}} \tag{2.5.16}$$

The noise temperature for the cascaded system can be obtained by rewriting Equation (2.5.16) as

$$F_o - 1 = (F_1 - 1) + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \cdots + \frac{F_N - 1}{G_1 G_2 \cdots G_{N-1}}$$

or

$$\frac{T_{oe}}{T_o} = \frac{T_{1e}}{T_o} + \frac{T_{2e}}{G_1 T_o} + \cdots + \frac{T_{Ne}}{G_1 G_2 \cdots G_{N-1} T_o} \qquad (2.5.17)$$

or finally

$$T_{oe} = T_{1e} + \frac{T_{2e}}{G_1} + \frac{T_{3e}}{G_1 G_2} + \cdots + \frac{T_{Ne}}{G_1 G_2 \cdots G_{N-1}} \qquad (2.5.18)$$

where T_{oe} = the effective overall system temperature

T_{ne} = the effective temperature of the Nth component

2.5.1.2 Noise Errors in Radar Measurements

The ability of a radar to detect the presence of an echo signal is fundamentally limited by noise. Likewise, noise is the factor that limits the accuracy with which the radar signals may be estimated. The parameters usually of interest in radar applications are the range, or time delay, the range rate, or doppler velocity, and the angle of arrival. The amplitude of the echo signal might also be measured, but its precise value is usually not important except **insofar** as it influences the signal-to-noise ratio.

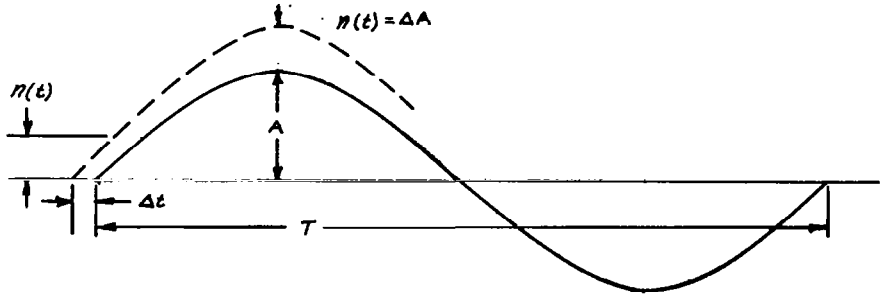
In the following analysis, the theoretical accuracies for radar measurements will be derived. To simplify the analysis, it is assumed that the signal is large compared with the noise. This is a reasonable assumption, since the signal-to-noise ratio must be relatively large if the detection is to be reliable. Furthermore, as will be shown, large signal-to-noise ratios are necessary for accurate measurements. It is also assumed that the error associated with a measurement of a particular parameter is independent of the errors in any of the other parameters. The validity of this assumption depends upon the availability of a large signal-to-noise ratio.

2.5.1.2.1 Sinusoidal Amplitude, Frequency and Phase Measurement Error

Consider a continuous sine wave

$$A \sin (2\pi f t + \phi) \qquad (2.5.19)$$

where f is the frequency and ϕ the phase. One period of the sine wave is shown in the following sketch.



The accompanying noise n(t) causes the apparent amplitude to differ from the true amplitude by an amount $\Delta A = n(t)$. The rms error in measuring amplitude is therefore

$$\delta A = \left(\overline{n^2}\right)^{1/2} \qquad (2.5.20)$$

The relative error is

$$\frac{\delta A}{A} = \frac{1}{\left(A^2/\overline{n^2}\right)^{1/2}} = \frac{1}{(2S/N)^{1/2}} \qquad (2.5.21)$$

where S/N is the signal-to-noise power ratio. The measurement of phase may be considered as the measurement of the time at which the waveform crosses the zero axis. The error in determining the time of a particular crossing is

$$\Delta t = \frac{n(t)}{\text{slope of sine wave at zero crossing}} \qquad (2.5.22)$$

The slope of the sine wave is $2\pi f A$ at the point of zero crossing. Therefore,

$$\delta t = \left[\overline{(\Delta t)^2}\right]^{1/2} = \frac{\left(\overline{n^2}\right)^{1/2}}{2\pi f A} = \frac{1}{2\pi f (2S/N)^{1/2}} \qquad (2.5.23)$$

Since $\phi = 2\pi f t$, the phase error $\delta\phi = 2\pi f \delta t$ is

$$\delta\phi = \frac{1}{(2S/N)^{1/2}} \qquad (2.5.24)$$

The period T is the time between two successive zero crossings of the same slope. The rms error in measuring the period will therefore be $\sqrt{2}$ times the single zero-crossing rms error, assuming that the zero-crossing measurements are independent.

$$(\delta T)^2 = (\delta t)^2 + (\delta t)^2 = 2(\delta t)^2$$

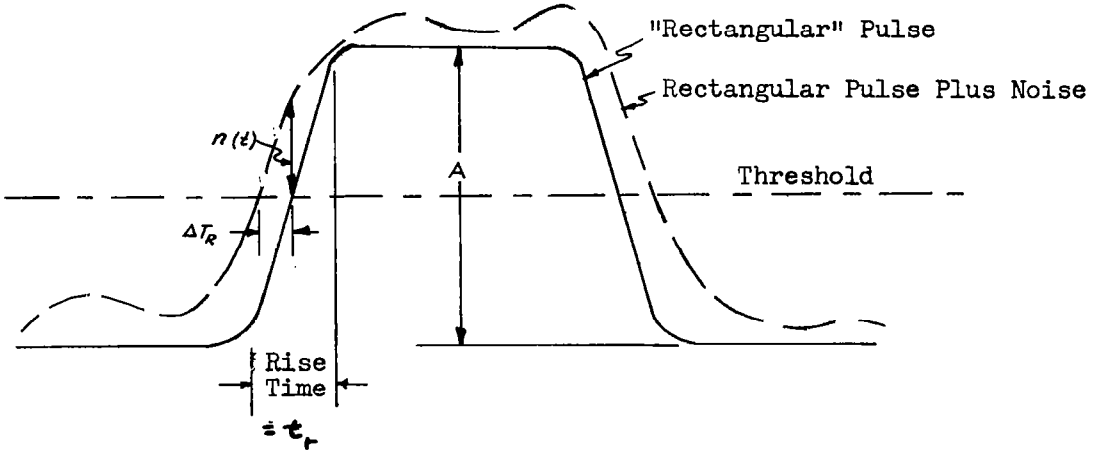
$$\delta T = \frac{\sqrt{2}}{2\pi f (2\,S/N)^{1/2}} = \frac{T}{2\pi (S/N)^{1/2}} \quad (2.5.24)$$

The error in period ΔT is related to the error in frequency Δf by $\Delta T = \frac{\Delta f}{f^2}$; therefore,

$$\delta f = \frac{f}{2\pi (S/N)^{1/2}} \quad (2.5.25)$$

2.5.1.2.2 **Range Accuracy - Leading Edge Measurement of a Pulse**

The measurement of range is the measurement of time delay $T_R = 2r/c$, where c is the velocity of light. One method of determing range with a pulsed waveform is to measure the time at which the leading edge of the pulse crosses the threshold as shown in the following sketch:



The pulse, uncorrupted by noise, is shown by the solid curve. The shape of the pulse is not perfectly rectangular; i.e., the rise and decay times are not zero, for this would require an infinite bandwidth. The effect of the noise is to perturb the shape of the pulse and to shift the time of threshold crossing as shown by the dashed curve. The maximum slope (rate of rise) of the leading edge of a rectangular pulse of amplitude A is A/t_r, where t_r is the rise time. For large signal-to-noise ratios, the slope of the pulse corrupted by noise is essentially the same as the slope of the uncorrupted pulse. From the sketch, the slope of the pulse in noise may be written as $n(t)/\Delta T_R$ where $n(t)$ is the noise voltage in the vicinity of the threshold crossing and ΔT_R is the error in the time-delayed measurement. Equating the two expressions for the slope gives

$$\Delta T_R = \frac{n(t)}{A/t_r} \qquad (2.5.26)$$

or

$$\left[\overline{(\Delta T_R)^2}\right]^{1/2} = \delta T_R = \frac{t_r}{(A^2/\overline{n^2})^{1/2}} = \frac{t_r}{(2\,S/N)^{1/2}} \qquad (2.5.27)$$

In the above expression, $A^2/\overline{n^2}$, is equal to twice the signal-to-noise ratio assuming that the signal power is $A^2/2$.

If the rise time of the pulse is limited by the bandwidth, B, then

$$t_r = \frac{1}{B}$$

Letting

$$S = \frac{E}{\tau}$$

and

$$N = N_0 B$$

where E = the signal energy,

N_0 = noise power per unit bandwidth

and τ = pulse width,

the error in the time delay can be written

$$\delta T_R = \left[\frac{\tau}{2BE/N_0}\right]^{1/2} \qquad (2.5.28)$$

Recalling that

$$T_R = \frac{2R}{c} \qquad (2.5.29)$$

and taking the derivative to find δT_R, it is seen that

$$\delta T_R = \frac{2}{c}\delta R = \left[\frac{\tau}{2BE/N_0}\right]^{1/2} \qquad (2.5.30)$$

Now,

$$\delta R = \frac{c}{2}\left[\frac{\tau}{2BE/N_0}\right]^{1/2} \qquad (2.5.31)$$

For constant pulse amplitude, the rms time-delay error given by equation (2.5.27) is proportional to the rise time and is independent of the pulse width. An improvement in the accuracy is obtained, therefore, by decreasing the rise time (increasing the bandwidth) or increasing the signal-to-noise ratio.

2.5.1.2.3 Accuracy of Frequency (or Doppler-Velocity) Measurement

A derivation which is analogous to the range error derivation in Reference (3) , can be used to find the minimum rms error in the measurement of frequency as

$$\delta f = \frac{1}{\alpha (2E/N_0)^{1/2}} \qquad (2.5.32)$$

where

$$\alpha^2 = \frac{\int_{-\infty}^{\infty} (2\pi t)^2 s^2(t)\, dt}{\int_{-\infty}^{\infty} s^2(t)\, dt} \qquad (2.5.33)$$

and s(t) the input signal as a function of time.

The parameter α is called the effective time duration of the signal. In radar, the measurement error specified by equation (2.5.32) is that of the doppler frequency shift. The value of α^2 for a perfectly rectangular pulse of width τ is $\pi^2\tau^2/3$; thus the rms frequency error is

$$\delta f = \frac{\sqrt{3}}{\pi\tau\,(2E/N_0)^{1/2}}$$

This expression shows that the frequency measurement is improved in accuracy as the pulse width is made longer. It should be noted that the expression for α used above was for a perfectly rectangular pulse which assumed an infinite bandwidth. If a bandwidth-limited case is considered, a slightly different expression for α is obtained. However, α reduces to $\pi^2\tau^2/3$ in the limit as the bandwidth goes to infinity. Since the purpose of this section was to introduce the best possible theoretical accuracy of the doppler measurement, the analysis will not be pursued any longer. The reader interested in finite bandwidths and other waveforms is referred to any radar systems book.

2.5.1.2.4 The Uncertainty Relation of Radar

The reader is no doubt familiar with the uncertainty principle of quantum mechanics which states that the position and velocity of an electron or other atomic particle cannot be simultaneously determined to any degree of accuracy desired. Precise determination of one parameter can be had only at the expense of the other. This is not the case in radar. Both position (range or time delay) and velocity (Doppler frequency) may in theory be determined simultaneously if the appropriate system parameters to be subsequently discussed are specified. The two uncertainty principles apply to different phenomena, and the radar principle based on classical concepts should not be confused with the physics principle that describes quantum-mechanical effects. In classical radar there is no theoretical limit to the minimum value of the $\delta T_R \cdot \delta f$ product since the radar system designer is free to choose system parameters that make this product as small as desired. The limits are practical ones, such as power limitations or the inability to meet tolerances. In the quantum- mechanical case, on the other hand, the observer does not have control over his system as does the radar designer since the parameters that are modified by the radar designer are fixed by the very nature of the quantum particle.

In order to derive the analytical expressions that illustrate the above principles, use is made of the accuracy relations of the time delay (range) and frequency (range-rate) measurements.

$$\delta T_R = \frac{1}{\beta (2E/N_0)^{1/2}} \qquad (2.5.34)$$

$$\delta f = \frac{1}{\alpha (2E/N_0)^{1/2}} \qquad (2.5.35)$$

where β is the effective bandwidth and α is the effective time duration of pulse.

Now expressing $\delta T_R \cdot \delta f$ as

$$\delta T_R \cdot \delta f = \frac{1}{\beta\alpha (2E/N_0)} \qquad (2.5.36)$$

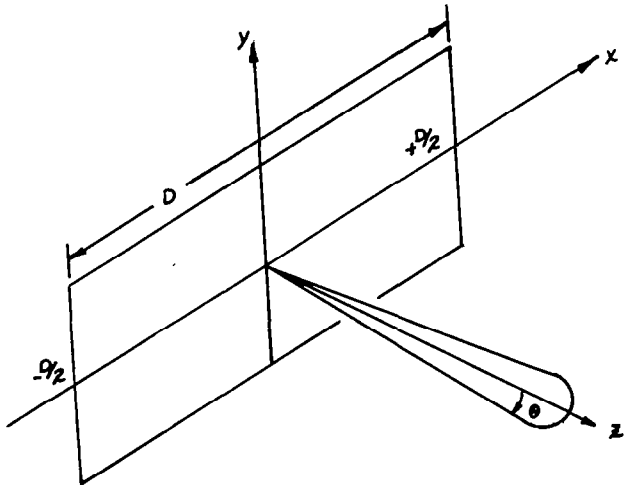
it can be seen that the time delay and the frequency may be simultaneously measured to as small a theoretical error as one desires by designing the radar to yield a sufficiently large ratio of signal energy (E) to noise, power per cycle (No), or for a fixed E/No, to select a radar waveform which results in a large value of $\beta\alpha$. Large $\beta\alpha$ products require waveforms long in duration and of wide bandwidth.

2.5.1.2.5 **Angular Accuracy**

The measurement of angular position is the measurement of the angle of arrival of the equiphase wavefront of the echo signal. The theoretical rms error of the angle measurement may be derived in a manner similar to the derivation of time (range) and frequency (range-rate) errors discussed previously. The analogy between the angular error and the time delay or frequency error comes about because the Fourier Transform describes the relationship between the radiation pattern and the aperture distribution of an antenna in a manner similar to the relationship between the time waveform and its frequency spectrum.

For simplicity, the angular error in one coordinate plane only will be considered. The analysis can be extended to include angular errors in both planes, if desired. It is assumed that the signal-to-noise ratios are large and that the noise can be described by the Gaussian probability-density function.

Consider a linear in-phase receiving antenna of length D, or a rectangular receiving aperture of width D as shown in the following sketch



The amplitude distribution across the aperture as a function of x is denoted A(x). The (voltage) gain as a function of the angle (one dimensional radiation pattern) in the xz plane is proportional to

$$G_v(\theta) = \int_{-D/2}^{D/2} A(x) \exp\left(2\pi j \frac{x}{\lambda} \sin\theta\right) dx \qquad (2.5.37)$$

When the angle θ is small, $\sin\theta \simeq \theta$ and the above equation is recognized as an inverse Fourier transform.

$$G_v(\theta) = \int_{-D/2}^{D/2} A(x) \exp\left(2\pi j x \frac{\theta}{\lambda}\right) dx \qquad (2.5.38)$$

This is analogous to the inverse Fourier transform relating the frequency spectrum $S(f)$ and the time waveform $\mathcal{A}(t)$, i.e.,

$$\mathcal{A}(t) = \int_{-\infty}^{\infty} S(f) \exp(2\pi j f)\, df \qquad (2.5.39)$$

As the antenna scans at a uniform angular rate ω_s, the received signal voltage from a fixed point source will be proportional to $G_v(\theta) = G_v(\omega_s t)$ and may be considered a time waveform. If θ/λ is associated with t in the inverse Fourier transform and x with f, the theoretical rms error can be obtained by an analogy to the time-delay accuracy expression

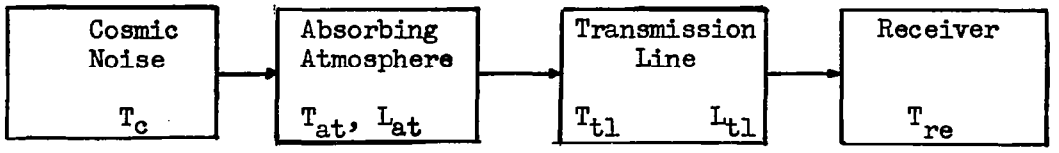
$$\delta\left(\frac{\theta}{\lambda}\right) = \frac{1}{\gamma\,(2E_0/N)^{1/2}} \qquad (2.5.40)$$

where γ is the effective aperture width defined by

$$\gamma^2 = \frac{\int_{-\infty}^{\infty} (2\pi x)^2 |A(x)|^2 dx}{\int_{-\infty}^{\infty} |A(x)|^2 dx} \qquad (2.5.41)$$

2.5.1.3 Total System Noise

The technique for computing total system noise will be explained via an example. Consider a radar system whose receiver is at T_{re} °K and sees cosmic noise and atmospheric noise. The system block diagram is shown below.



From equation (2.29), it can be seen that the effective temperature of the atmosphere is

$$T_{e_{at}} = (L_{at} - 1)\, T_{at}$$

Similarly, the effective temperature of the transmission line is

$$T_{e_{tl}} = (L_{tl} - 1)\, T_{tl}$$

Using the effective temperature equation for components in Cascade, equation (2.5.18), it is seen that

$$T_e = T_c + T_{e_{at}} + T_{e_{tl}} L_{at} + T_{re} L_{at} L_{tl}$$

or $$T_e = T_c + (L_{at} - 1)\, T_{at} + (L_{tl} - 1)\, T_{tl} L_{at} + T_{re} L_{at} L_{tl}$$

Hence, if the noise parameters of the components of a system are known, the effective noise temperature of the system can be calculated and an estimate of the expected noise power can be determined by

$$N = K\, T_e\, B_n$$

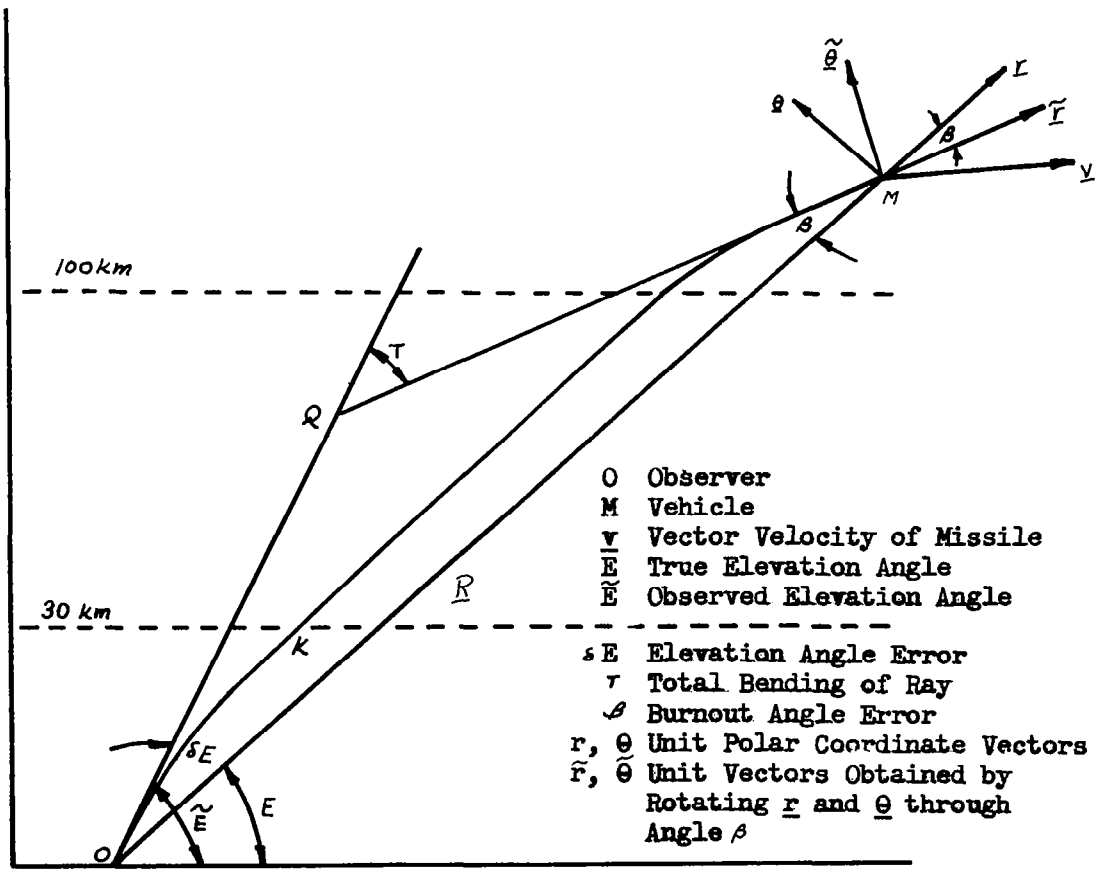
where N = Noise power in watts
K = Boltzmann's constant
T_e = Effective temperature °K
B = Noise band width (CPS)

2.5.1.4 Atmospheric Refraction Errors

Position and velocity information may be deduced by tracking boost and space vehicles from earth based tracking stations. An approximate estimate of vehicle position and velocity may be generated directly from the tracking antenna, gimbal angles, and/or elapsed time measurements for radio signals to travel from vehicle to ground, etc. These estimates of position and velocity contain a predictable or bias error component , however, which may be removed by analyzing the refractive effects of the earth's atmosphere. The analytical derivation follows:

First, a flat earth model with a horizontally stratified atmosphere will be examined and then a spherical earth with the index of refraction a function of the distance from the center of the earth will be analyzed.

Consider the electromagnetic wave traveling from some vehicle outside the atmosphere. The path that the wave follows is shown in the following sketch.



The total angle τ through which the ray will be bent, i.e., the angle through which the tangent to the ray rotates, in going from the vehicle to ground can be derived using Snell's Law

$$\eta_o \cos \tilde{E} = \eta_M \cos(\tilde{E} - \tau) \qquad (2.5.42)$$

where $\eta_M = 1 + \epsilon_M$ = index of refraction at vehicle, M

$\eta_o = 1 + \epsilon_o$ = index of refraction at ground, O

τ is sufficiently small so that one can ignore terms in τ^2. For the values of ϵ_M, ϵ_o, and E encountered in practice

$$\epsilon_M \approx 0.6 \times 10^{-6} \text{ at } 10{,}000 \text{ MC}$$

$$\approx 6.0 \times 10^{-6} \text{ at } 1000 \text{ MC}$$

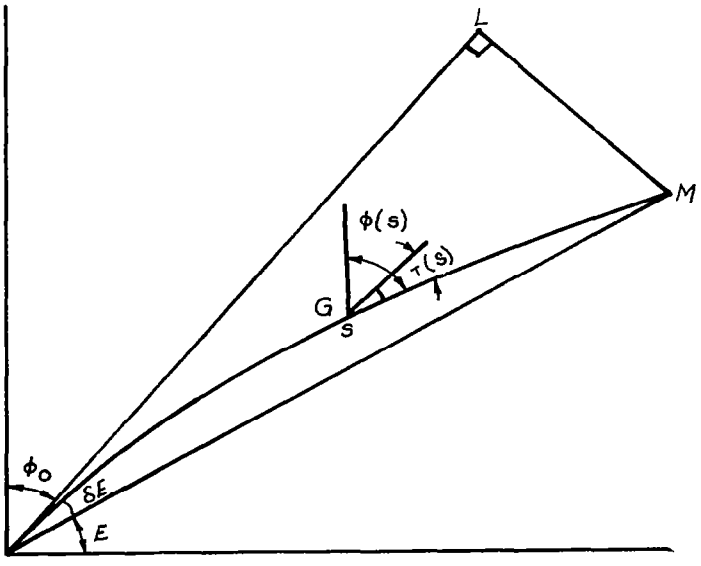
$$\epsilon_o \approx 300 \times 10^{-6}$$

$$\tilde{E} \approx 25^\circ$$

Hence, equation (2.5.42) can be approximated as

$$\tau = (\epsilon_o - \epsilon_M) \operatorname{CTN} \tilde{E} \qquad (2.5.43)$$

The error angle δE between line OM and the tangent to the ray at O may be derived with the aid of the following sketch



O Observer
M Vehicle
G Point of Ray OM
OL Tangent to Ray at O
ML Perpendicular to OL
S Distance from O to G Measured along Ray OM
$\phi(s)$ Inclination of Ray at G
T(s) Total Bending of Ray at G

Let ML be the perpendicular from the vehicle M to the extension OL of the tangent to the ray at O. Then δE is the angle whose sin is ML/OM and

$$ML = \int_{\Gamma} d\mathcal{A} \sin \tau(\mathcal{A}) \tag{2.5.44}$$

where the integration is along the ray path Γ and $\tau(\mathcal{A})$ is the total bending at point $\mathcal{A}$ relative to the ray direction at O. Since OKM is never very far away from the straight line OM, equation (2.5.44) may be approximated as

$$ML \doteq \int_0^R \sin \tau \, d\tau \doteq \int_0^R \tau \, dr = R\bar{\tau}, \tag{2.5.45}$$

where $\bar{\tau}$ is the average of τ along OM.

Now

$$\delta E \doteq \frac{ML}{OM} = \bar{\tau} \tag{2.5.46}$$

$\bar{\tau}$ can now be evaluated by averaging equation (2.5.4.3) leading to

$$\delta E = (\epsilon_0 - \bar{\epsilon}) \cot N \tag{2.5.47}$$

Note that, although a single observation at the surface is necessary to determine ϵ_0, $\bar{\epsilon}$ is a function of the entire refractive index profile. However, $\bar{\epsilon}$ can be estimated by using a standard profile. The contribution to $\bar{\epsilon}$ from the lower atmosphere (below 30 km) can be approximated by the following model

$$n_0(h) = 1 + \epsilon_0 e^{-h/H} \tag{2.5.48}$$

where $\epsilon_0 = n_0(0) - 1$

and $H = 7.6$ km.

For $h \geq 30$ km, equation (2.5.48) can be approximated as

$$\bar{\epsilon}_{LA} = \frac{7.6\,\epsilon_0}{h}$$

where h is the vehicle's altitude measured in kilometers. The contribution to $\bar{\epsilon}$ from the ionosphere can be estimated by using the electron density profile

$$\bar{\epsilon}_I = -0.2 \times 10^{-6} \text{ at } 10{,}000 \text{ mi.}$$

$$\bar{\epsilon}_I = -2.0 \times 10^{-6} \text{ at } 1000 \text{ mi.}$$

The net average value of ϵ is $\bar{\epsilon} = \bar{\epsilon}_{LA} + \bar{\epsilon}_{I}$ Thus, equation (2.5.47) can be used to determine the elevation angle error for a flat earth model.

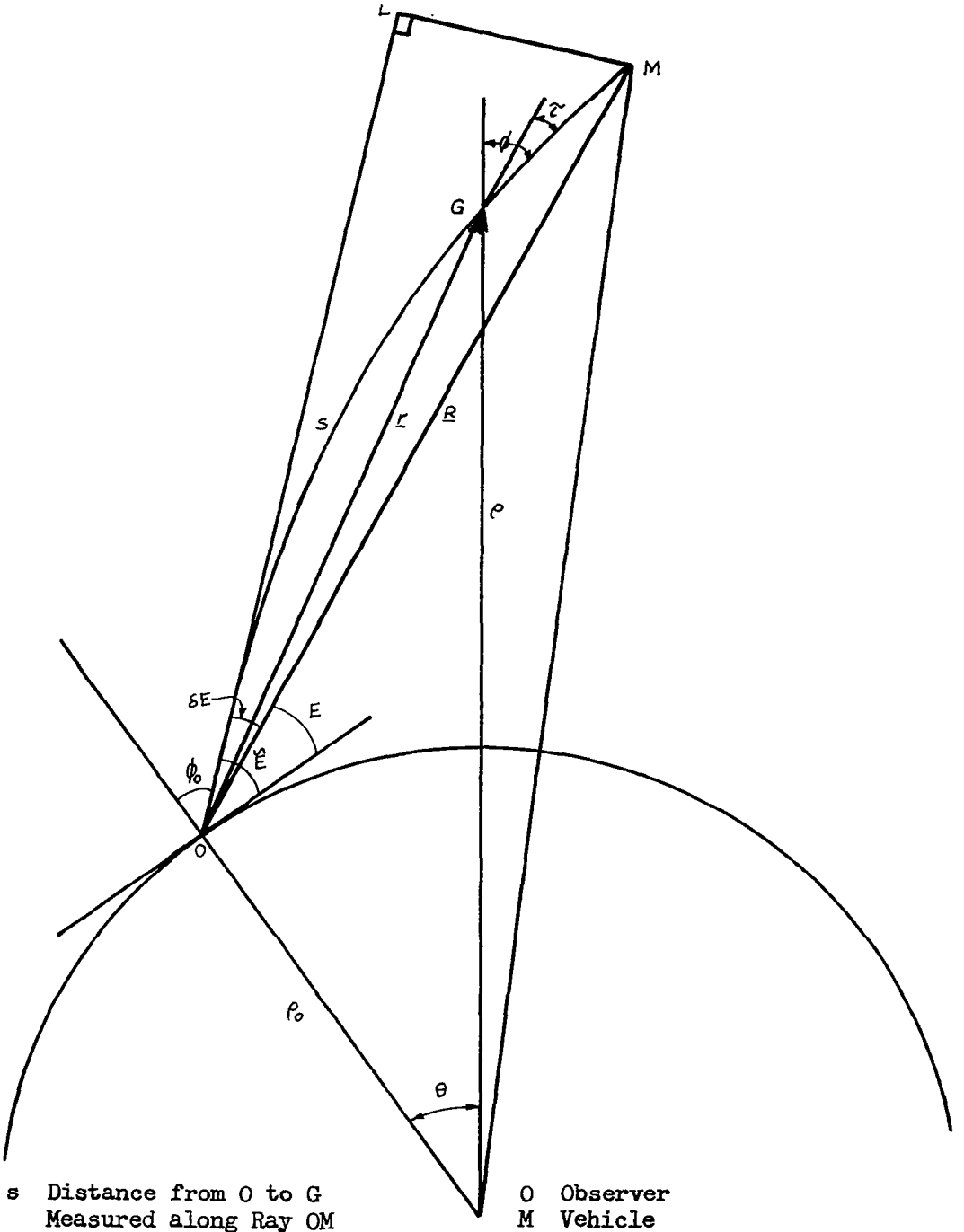
The spherical earth case is now considered. For a spherical earth with the index of refraction, a function of the distance ρ from the center of the earth Snell's Law is

$$\rho \eta \sin \phi = \text{constant along a ray} \qquad (2.5.49)$$

where η = index of refraction

ϕ = inclination of ray from the local vertical

This expression can be obtained from the generalized differential form of Snell's Law if the condition of a spherically stratified atmosphere is utilized. The reader is referred to reference 6 for the details of this derivation. The following sketch shows the geometry to be used in the following analysis.



s	Distance from O to G Measured along Ray OM	O	Observer
ϕ (s)	Inclination of Ray Local Vertical at G	M	Vehicle
(p, θ)	Polar Coordinates of G Relative to Center of Earth	G	Point of Ray on OM
r	Length of Straight Line OG	Q	Center of Earth
		QL	Tangent to Ray OM
		ML	Perpendicular to OL
			Ray Path

Insofar as the effects of the lower atmosphere are concerned, equation (2.5.49) is a good approximation as long as the elevation angle E is greater than approximately 3 degrees. This has been verified by ray tracing calculations of Bean and Cahoon. Ionospheric effects are generally negligible for radio frequencies above 1000 mc, compared to lower atmospheric effects. Correction for a spherical earth in the ionosphere is a second order perturbation.

The following derivation is a method for computing the total bending $\tau(h)$ up to the altitude h in a stratified atmosphere over a spherical earth.

Taking the differential of equation (2.5.4.9) along a ray, one finds

$$d\rho\, n \sin\phi + \rho\, dn \sin\phi + \rho n \cos\phi\, d\phi = 0 \qquad (2.5.50)$$

Using simple geometry,

$$d\theta = \frac{d\rho \tan\theta}{\rho}$$

where θ is the angle subtended at the center of the earth. Equation (2.5.50) can now be written as

$$d\phi + d\theta = -\frac{dn}{n} \tan\phi \qquad (2.5.51)$$

Thus, $\tau(s)$, the angle which the tangent at s makes with the tangent at 0 is given by

$$\tau(s) = \int_{\Gamma} d(\phi + \theta) = -\int_{\Gamma} \frac{dn}{n} \tan\phi \qquad (2.5.52)$$

where Γ is the path of the ray. The following approximation is now made

$$\frac{1}{n} \doteq 1 - \epsilon \qquad (2.5.53)$$

$$\tan\phi = \tan(\phi_0 + \Delta\phi) \doteq \tan\phi_0 + \Delta\phi \sec^2\phi_0$$

where the errors are of order ϵ^2 and $\Delta\phi^2$. To a first approximation, the change in the inclination of the ray to the local vertical is given by

$$\Delta\phi \doteq (\epsilon_0 - \epsilon)\tan\phi_0 - \frac{r\cos\phi_0}{\rho_0} \qquad (2.5.54)$$

where ρ_0 is the radius of the earth (ρ_0 = 6378 km).

Equations (2.5.53) and (2.5.54) are then substituted into equation (2.5.52) and integrated along the line OM out to r, keeping only terms which are second order in ϵ and r/ρ_0:

$$\tau = \left[\epsilon_0 - \epsilon - 1/2(\epsilon_0^2 - \epsilon^2) + 1/2(\epsilon_0 - \epsilon)^2 \cos^2 \underset{\sim}{E} - \frac{r}{\rho_0}(\bar{\epsilon} - \epsilon)\cos \underset{\sim}{E}\right] \mathrm{CTN}\ \underset{\sim}{E} \tag{2.5.55}$$

where $\underset{\sim}{E} = \pi/2 - \phi_0$ is the apparent elevation angle and $\bar{\epsilon}$ denotes the average value of ϵ along the line OM out to the value r:

$$\bar{\epsilon} = \frac{1}{r}\int_0^r \epsilon(\xi)\, d\xi \tag{2.5.56}$$

The equation for the elevation angle correction angle can be shown to be

$$\delta E = \mathrm{TAN}^{-1}\left(\frac{LM}{OM}\right) \tag{2.5.57}$$

where $$LM = \int_r d\Delta \sin \tau$$

Since the ray $\widehat{OM}$ and the line OM lie close together, δE can be written as

$$\delta E \doteq \frac{1}{OM}\int_0^R dr \sin \tau \doteq \frac{1}{OM}\int_0^R \tau\, dr = \bar{\tau} \tag{2.5.58}$$

where $\bar{\tau}$ is the average value of τ along OM.

By averaging equation (2.5.55) over OM, the spherical earth correction formula for the elevation angle is obtained.

$$\delta E = \left[\epsilon_0 - \bar{\epsilon} - \tfrac{1}{2}(\epsilon_0^2 - \overline{\epsilon^2}) + \tfrac{1}{2}\overline{(\epsilon_0 - \epsilon)^2}\cos^2 \underset{\sim}{E} - \frac{1}{\rho_0}(R\bar{\epsilon} - 2\overline{r\epsilon})\cos \underset{\sim}{E}\right] \mathrm{CTN}\ \underset{\sim}{E} \tag{2.5.59}$$

Range errors are also introduced by the refraction of the electromagnetic wave through the atmosphere. If range is measured by measuring the transit time of a pulse going from O to M, then the distance measured will really be the effective path length $\widehat{OKM}$

$$L = c\int_r \frac{d\Delta}{v_g} = c\int_0^R \frac{dr}{v_g} \tag{2.5.60}$$

where V_g is the group velocity of a pulse of a carrier frequency f_o .

It can be shown that there is a negligible difference between the arc $\overline{OM}$ and the straight line OM. Thus, integration of equation (2.5.60) may be performed along the line OM. Writing

$$\frac{c}{V_g} = \frac{c}{c - \delta V_g} \doteq 1 + \frac{\delta V_g}{c} \tag{2.5.61}$$

it is seen that

$$L = R + \delta R \tag{2.5.62}$$

where

$$\delta R = \int_0^R \frac{\delta V_g}{c}\, dr \tag{2.5.63}$$

Now, the doppler errors due to refraction will be considered. Let the vehicle at point M, moving with a velocity vector $\underline{V}$, transmit a signal $\exp(2\pi j f_o t)$. Then the signal which will be received on the ground at O will be

$$\exp\left\{2\pi j f_o \left[t - \frac{1}{c}\int_r n(s)\, ds\right]\right\} \tag{2.5.64}$$

(appropriate modifications must be made for two-way doppler transmissions). The instantaneous frequency received on the ground will be the time rate of change in the above expression, i.e.,

$$f = f_o - \frac{f_o}{c}\frac{d}{dt}\int_r n(s)\, ds \tag{2.5.65}$$

or

$$f = f_o - \frac{f_o}{c} n_m \frac{ds}{dt} - \frac{f_o}{c}\int_r \frac{\partial n}{\partial t}\, ds \tag{2.5.66}$$

Thus the observed one-way doppler shift at frequency f_o is

$$\Delta f = \frac{f_o n_m}{c}\frac{ds}{dt} + \frac{f_o}{c}\int_r \frac{\partial n}{\partial t}\, ds \tag{2.5.67}$$

The first term on the right hand side of the above equation is the usual doppler shift term. It is pointed out, however, that (1) the index of refraction n_m at the vehicle appears in this term and (2) since the direction of increase of path length (ds/dt) is tangent to the ray path, this term corresponds to a velocity component tangent to the ray and not in the direction of the straight line OM.

The primary observable quantities from which vehicle velocity V is determined are elevation angular rate E and a doppler shift Δf, together with positional coordinates r and E. Then

$$\underline{V} = \frac{ds}{dt}\,\underline{\hat{r}} + r\dot{E}\,\underline{\hat{\theta}} \qquad (2.5.68)$$

where $\underline{\hat{r}}$ and $\underline{\hat{\theta}}$ are rotated from $\underline{r}$ and $\underline{\theta}$ by the angle β,

$$(\beta = \tau - \delta E = \bar{E} - \epsilon_n \cot \bar{E})$$

and

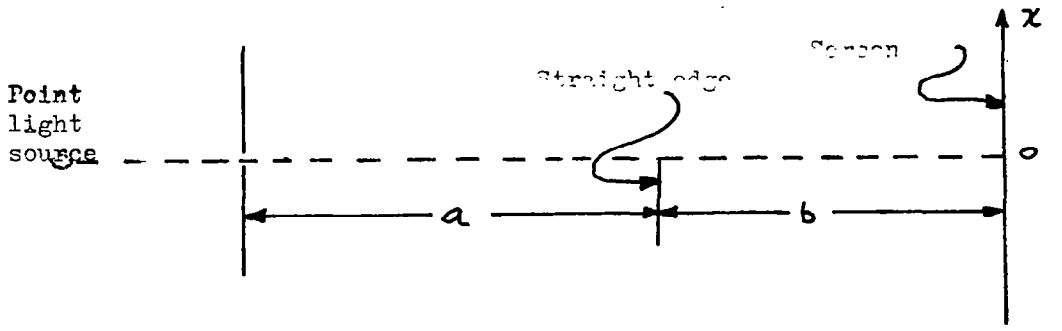
$$\frac{ds}{dt} = \frac{1}{n_m}\,\frac{c\,\Delta f}{f_o} - \int_r \frac{\partial n}{\partial t}\,ds$$

Hence, equation (2.5.68) is the doppler velocity equation with atmospheric refraction considered.

2.5.1.5 Horizon-Induced Diffraction Errors

A diffraction pattern is produced whenever electromagnetic radiation passes through a narrow slit or passes near a straight edge. Hence, diffraction occurs when electromagnetic radiation passes close to the horizon of a celestial object before arrival at a sensor, and produces an apparent bending of the radiation "rays" near a horizon "edge".

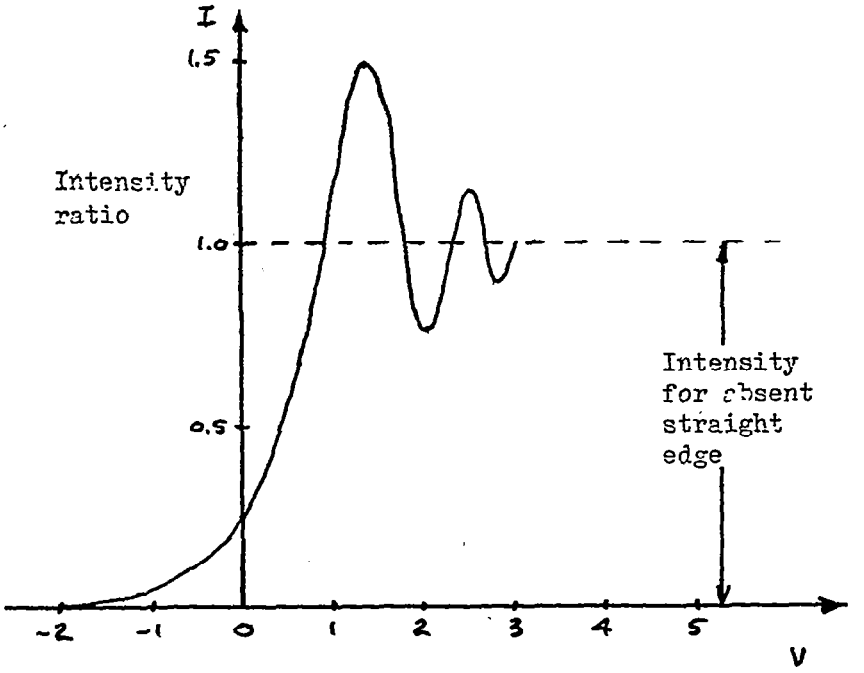
For most purposes, the horizon diffraction effect can be evaluated by applying the results of an evaluation of the Fresnel diffraction integrels describing diffraction produced by a straight edge. Assume that a straight edge is placed between a point source of light and a screen, as shown below;



The intensity of the radiation on the screen may be plotted as a function of x, the position on the screen. For simplification of the presentation, however, it is more convenient to plot intensity of the radiation as a function of V, where

$$x = V\sqrt{\frac{\lambda b(a+b)}{2a}} \qquad (2.5.69)$$

where: a,b, and x are defined in the sketch λ = wavelength of the radiation. The plot of intensity as a function of V is presented below:



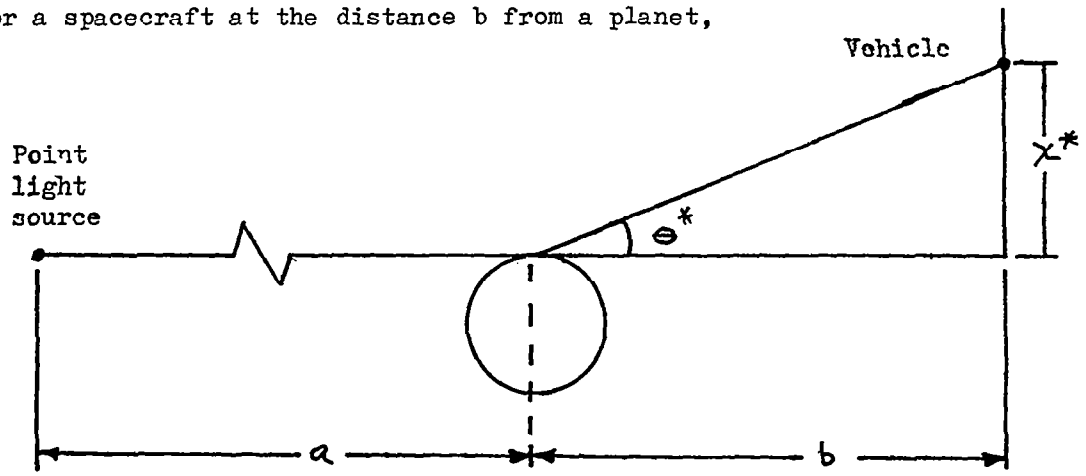
It is noted that the radiation reaches a maximum of nearly 1.5 times the intensity that would be measured in the absence of the straight edge for the value

$$V_{\substack{\text{MAX} \\ \text{INTENSITY}}} \triangleq V^* \cong 1.3 \qquad (2.5.70)$$

The corresponding value of x is:

$$x_{\substack{\text{MAX} \\ \text{INTENSITY}}} \triangleq x^* = V^* \sqrt{\frac{\lambda b(a+b)}{2a}} = 1.3 \sqrt{\frac{\lambda b(a+b)}{2a}} \qquad (2.5.71)$$

For a spacecraft at the distance b from a planet,



the angular error θ^* is thus:

$$\theta^* = \frac{x^*}{b} \approx \frac{1.3}{b}\sqrt{\frac{\lambda b\,(a+b)}{2a}} = 1.3\sqrt{\frac{\lambda(a+b)}{2ab}} \qquad (2.5.72)$$

If the distance from the vehicle to the planet (or celestial body) is much less than from the planet to the source of radiation, then $b \ll a$ and equation (2.5.72) becomes

$$\theta^* = 1.3\sqrt{\frac{\lambda}{2b}} \qquad (2.5.73)$$

Examining the angular error at visible wavelengths, as an example,

$$\lambda = 5{,}000\ \text{Å} = 5 \times 10^{-5}\ \text{cm.} = 2.69 \times 10^{-10}\ \text{n. mi.}$$

equation (2.5.73) becomes

$$\theta^* = \frac{1.51 \times 10^{-5}}{\sqrt{b}}\ \text{radians} = \frac{3.720 \times 10^{-5}}{\sqrt{b}}\ \text{degrees}$$

$$\theta^* = \frac{0.134}{\sqrt{b}}\ \text{arc sec}$$

If b = 100 n.mi., $\theta^* = 0.013$ arc sec

As the wavelength increases, equation (2.5.73) indicates that the diffraction angle θ^* increases. For example, if the wavelength were to correspond to the long wave radio wavelengths, say $\lambda = 5 \times 10^3$ meters, then assuming $a = \infty$, $b = 100$ n.mi. the diffraction angle θ^* becomes 0.372 degrees.

The conclusion from these examples is that for practical guidance measurements in the visible spectrum, the effects of horizon diffraction appear negligible. **It is also noted that diffraction at radio frequencies may introduce significant errors. The equations presented may be used to assess horizon diffraction effects throughout the electromagnetic spectrum.**

2.5.2 Inertial Sensing Errors

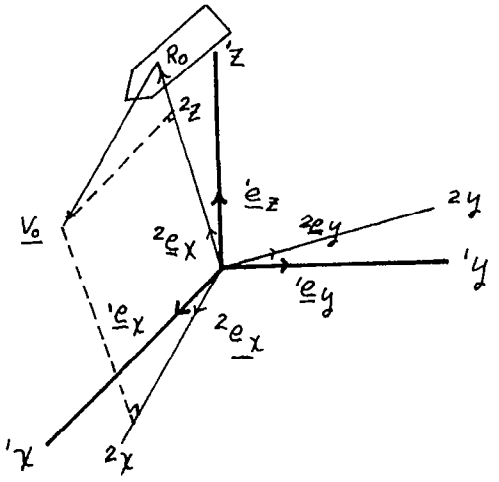
The mechanizations of position and velocity computation within an inertial navigation system are subject to several classifications of errors. This section presents an analysis of the errors which occur in these mechanizations. The sequence of subjects which follows is concerned first with a derivation of the general error equations, followed by a discussion of the solutions of the error equations and the characteristics of the individual errors.

2.5.2.1 General Error Equations

The mechanization of two forms of inertial measuring units were presented in Section 2.3.2. In the presentation of these mechanizations it was convenient to carry out the computations in a non-rotating inertial reference frame. In the derivation in this section it is more enlightening to write the equations in "local" vertical coordinates.

The fact that the navigation computations may be carried out in a number of different coordinate frames should cause no conceptual difficulty, since all of the computations amount to the same solution of the equations of motion. That this must be true is guaranteed by the uniqueness of the solution of the equations for the position vector $\underline{R}$ (t), when the initial conditions $\underline{R}$ (o) and $\underline{V}$ (o) and the accelerometer signals $\underline{A}$ (t) are given.

Consider a vehicle with a position vector $\underline{R}_o$ (t) and a velocity $\underline{V}_o$ (t). Let $\underline{e}_x, \underline{e}_y, \underline{e}_z$ define an inertial orthogonal coordinate frame located with its origin at the earth center (assumed to be inertially fixed in this derivation) and let $\underline{e}_x, \underline{e}_y, \underline{e}_z$ define an instantaneous local vertical frame with its origin located at the earth center. Let the $\underline{e}_z$ vector be oriented upward along the local vertical, and let $\underline{e}_x$ be oriented such that the plane $\underline{e}_x, \underline{e}_z$ always contains the vehicle velocity vector $\underline{V}_o$ (t). The sketch below illustrates this geometry.



Let the position and velocity vectors of the vehicle, as determined by a perfect error-free inertial navigation system, be denoted $\underline{R}_o$ (t), $\underline{V}_o$ (t), and let the acceleration as measured by perfect error-free accelerometers be denoted $\underline{A}_o$ (t). Then:

$$\underline{A}_o(t) = \ddot{\underline{R}}_o + \frac{\mu}{R_o^3}\,\underline{R}_o \qquad (2.5.74)$$

An actual inertial navigation system is now assumed in which the error corrupted computed position and velocity vectors are $\underline{R}$ (t) and $\underline{V}$ (t), and for which the error corrupted accelerometers yield signals $\underline{A}$ (t). Then the equation describing the actual navigation computations is:

$$\underline{A}(t) = \ddot{\underline{R}} + \frac{\mu}{R^3}\,\underline{R} \qquad (2.5.75)$$

Now let the position and accelerometer errors be denoted $\underline{\Delta A}$ and $\underline{\Delta R}$, i.e.,

$$\underline{R}(t) = \underline{R}_o(t) + \underline{\Delta R}(t) \qquad (2.5.76)$$

$$\underline{A}(t) = \underline{A}_o(t) + \underline{\Delta A}(t) \qquad (2.5.77)$$

Subtracting Equation (2.5.74) from (2.5.75), and substituting in Equations (2.5.76) and (2.5.77) yields:

$$\underline{\Delta A} = \underline{\Delta \ddot{R}} + \mu\left[\frac{\underline{R}}{R^3} - \frac{\underline{R}_o}{R_o^3}\right] \qquad (2.5.78)$$

Note that the definition of $\underline{\Delta \ddot{R}}$ is unambiguous and is $\frac{d^2}{dt^2}(\underline{\Delta R})$.

Equation (2.5.78) may be re-written as:

$$\underline{\Delta A} = \underline{\Delta \ddot{R}} + \frac{\mu}{R_o^3 \left(\frac{R^3}{R_o^3}\right)} \left[\underline{R} - \frac{R^3}{R_o^3} \underline{R}_o \right] \quad . \qquad (2.5.79)$$

If only very small deviations of position $\underline{R}$ (t) from $\underline{R}_o$ (t) exist, then $\frac{R^3}{R_o^3}$ is very nearly unity. This approximation cannot be made in the bracketed term $\left[\underline{R} - \frac{R^3}{R_o^3} \underline{R}_o \right]$ however, since this expression represents a small difference of two large quantities.

An approximation of the magnitude of the variation of $\frac{R^3}{R_o^3}$ from unity may be derived as follows:

$$\frac{R^3}{R_o^3} \triangleq \frac{|\underline{R}_o + \Delta \underline{R}|^3}{R_o^3}$$

or,

$$\frac{R^3}{R_o^3} = \frac{\left[(R_{ox} + \Delta R_x)^2 + (R_{oy} + \Delta R_y)^2 + (R_{oz} + \Delta R_z)^2 \right]^{3/2}}{R_o^3}$$

where

$$\underline{R}_o \triangleq R_{ox} \underline{e}_x + R_{oy} \underline{e}_y + R_{oz} \underline{e}_z$$

$$\Delta \underline{R} \triangleq \Delta R_x \underline{e}_x + \Delta R_y \underline{e}_y + \Delta R_z \underline{e}_z \qquad (2.5.80)$$

Hence,

$$\frac{R^3}{R_o^3} = \frac{\left[R_{ox}^2 + R_{oy}^2 + R_{oz}^2 + 2 \left(R_{ox} \Delta R_x + R_{oy} \Delta R_y + R_{oz} \Delta R_z \right) + \Delta R_x^2 + \Delta R_y^2 + \Delta R_z^2 \right]^{3/2}}{R_o^3}$$

$$\frac{R^3}{R_o^3} \cong \left[1 + \frac{2 \underline{R}_o \cdot \Delta \underline{R}}{R_o^2} \right]^{\frac{3}{2}} \qquad (2.5.81)$$

where it has been assumed that $|\Delta R| << R_0$. Since for $\chi << 1$, $(1+\chi)^{3/2} \simeq 1 + \frac{3}{2}\chi$, an approximation to Equation (2.5.81) is:

$$\frac{R^3}{R_0^3} = 1 + \frac{3\underline{R}_0 \cdot \Delta\underline{R}}{R_0^2} \tag{2.5.82}$$

Substituting Equation (2.5.82) into Equation (2.5.79) yields:

$$\Delta\underline{A} = \Delta\ddot{\underline{R}} + \frac{\mu}{R_0^3\left[1 + \frac{3\,\underline{R}_0 \cdot \Delta\underline{R}}{R_0^2}\right]}\left[\underline{R} - \left(1 + \frac{3\,\underline{R}_0 \cdot \Delta\underline{R}}{R_0^2}\right)\underline{R}_0\right]$$

or

$$\Delta A = \Delta\ddot{\underline{R}} + \frac{\mu}{R_0^3\left[1 + \frac{3\underline{R}_0 \cdot \Delta\underline{R}}{R_0^2}\right]}\left[\Delta\underline{R} - \frac{3(\underline{R}_0 \cdot \Delta\underline{R})}{R_0^2}\,\underline{R}_0\right] \tag{2.5.83}$$

Now, for small deviations $\Delta\underline{R}$ about $\underline{R}_0$,

$$1 + \frac{3\,\underline{R}_0 \cdot \Delta\underline{R}}{R_0^2} \simeq 1 \tag{2.5.84}$$

Also, as can be confirmed by expansion,

$$(\underline{R}_0 \cdot \Delta\underline{R})\,\underline{R}_0 = (\underline{R}_0\,\underline{R}_0^T)\,\Delta\underline{R} \tag{2.5.85}$$

where

$$\underline{R}_o \underline{R}_o^T \triangleq \begin{bmatrix} R_{ox} \\ R_{oy} \\ R_{oz} \end{bmatrix} \begin{bmatrix} R_{ox} & R_{oy} & R_{oz} \end{bmatrix} = \begin{bmatrix} R_{ox}^2 & R_{ox}R_{oy} & R_{ox}R_{oz} \\ R_{ox}R_{oy} & R_{oy}^2 & R_{oy}R_{oz} \\ R_{ox}R_{oz} & R_{oy}R_{oz} & R_{oz}^2 \end{bmatrix}$$

Thus, substitution of (2.5.84), and (2.5.85) into (2.5.83) yields

$$\underline{\Delta A} = \underline{\Delta \ddot{R}} + \frac{\mu}{R_o^3}\left[\underline{\underline{I}} - \frac{3}{R_o^2}(\underline{R}_o \underline{R}_o^T)\right] \underline{\Delta R} \qquad (2.5.86)$$

where I is the identity matrix $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

From Equation (2.5.80),

$$\underline{\Delta R} = \Delta R_x \underline{\hat{e}}_x + \Delta R_y \underline{\hat{e}}_y + \Delta R_z \underline{\hat{e}}_z, \qquad (2.5.87)$$

At this point the first and second time derivatives of (2.5.87) are formed.

$$\underline{\Delta \dot{R}} = \underline{\Delta v} + \underline{\Omega} \times \underline{\Delta R} \qquad (2.5.88)$$

$$\underline{\Delta \ddot{R}} = \underline{\Delta a} + 2\,\underline{\Omega} \times \underline{\Delta v} + \underline{\dot{\Omega}} \times \underline{\Delta R} + \underline{\Omega} \times (\underline{\Omega} \times \underline{\Delta R}) \qquad (2.5.89)$$

where:

$$\underline{\Delta v} \triangleq \Delta\dot{R}_x \underline{\hat{e}}_x + \Delta\dot{R}_y \underline{\hat{e}}_y + \Delta\dot{R}_z \underline{\hat{e}}_z$$

$$\underline{\Delta a} \triangleq \Delta\ddot{R}_x \underline{\hat{e}}_x + \Delta\ddot{R}_y \underline{\hat{e}}_y + \Delta\ddot{R}_z \underline{\hat{e}}_z \qquad (2.5.90)$$

$$\underline{\Omega} = \Omega_x \underline{\hat{e}}_x + \Omega_y \underline{\hat{e}}_y + \Omega_z \underline{\hat{e}}_z$$

and where $\underline{\Omega}$ is the angular velocity of the local vertical coordinate frame (2) with respect to the inertial frame (1).

The general error equations describing inertial system errors are given by Equations (2.5.86), (2.5.89), and (2.5.90), and are assembled here for convenience of reference:

General Error Equations	
from (2.5.86)	
$\underline{\Delta A} = \underline{\Delta\ddot{R}} + \frac{\mu}{R_o^3}\left[\underline{\underline{I}} - \frac{3}{R_o^2}(\underline{R}_o \underline{R}_o^T)\right]\underline{\Delta R}$	(2.5.91a)
from (2.5.89)	
$\underline{\Delta\ddot{R}} = \underline{\Delta a} + 2\underline{\Omega} \times \underline{\Delta v} + \dot{\underline{\Omega}} \times \underline{\Delta R} + \underline{\Omega} \times (\underline{\Omega} \times \underline{\Delta R})$	(2.5.91b)
from (2.5.90)	
$\underline{\Delta v} = \Delta\dot{R}_x \underline{\hat{e}}_x + \Delta\dot{R}_y \underline{\hat{e}}_y + \Delta R_z \underline{\hat{e}}_z$	(2.5.91c)
$\underline{\Delta a} = \Delta\ddot{R}_x \underline{\hat{e}}_x + \Delta\ddot{R}_y \underline{\hat{e}}_y + \Delta\ddot{R}_z \underline{\hat{e}}_z$	(2.5.91d)
$\underline{\Omega} = \Omega_x \underline{\hat{e}}_x + \Omega_y \underline{\hat{e}}_y + \Omega_z \underline{\hat{e}}_z$	(2.5.91e)

2.5.2.2 Special Solutions of the Error Equations

Equation (2.5.91) is too general to indicate the predominant characteristics of the errors in an inertial navigation system. Thus, consider a more restrictive case. In most circumstances, the angular velocity of the local vertical coordinate frame is dominated by its Ω_y term. Therefore, the indicated cross product multiplications with the assumption that $\underline{\Omega} \approx \Omega\, \underline{e}_y$ yields for Equation (2.5.91a)

$$\begin{aligned} \Delta\ddot{\underline{R}} = &\left[\Delta\ddot{R}_x + 2\Omega\,\Delta\dot{R}_z - \Omega^2\,\Delta R_x\right]\underline{e}_x \\ &+\left[\Delta\ddot{R}_y\right]\underline{e}_y + \left[\Delta\ddot{R}_z - 2\Omega\,\Delta\dot{R}_x - \Omega^2\Delta R_z\right]\underline{e}_z \end{aligned} \qquad (2.5.92)$$

Some additional simplifications to Equation (2.5.91) result from the fact that the local vertical reference frame has been defined as having its z axis along the radius vector. Hence, $\underline{R}_o$ has the form:

$$\underline{R}_o = 0\,\underline{e}_x + 0\,\underline{e}_y + R_o\,\underline{e}_z \qquad (2.5.93)$$

The matrix term $\frac{3}{R_o^2}\left(\underline{R}_o\,\underline{R}_o^T\right)$ can thus be written:

$$\frac{3}{R_o^2}\left(\underline{R}_o\,\underline{R}_o^T\right) = \frac{3}{R_o^2}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & R_o^2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix} \qquad (2.5.94)$$

Therefore, substitution of Equation (2.5.94) into Equation (2.5.91a) yields:

$$\underline{\Delta A} = \underline{\Delta \ddot{R}} + \frac{\mu}{R_o^3} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix} \underline{\Delta R} \qquad (2.5.95)$$

Substituting Equation (2.5.92) into Equation (2.5.95) now yields the three simultaneous equations:

$$\Delta \ddot{R}_x + 2\Omega \Delta \dot{R}_z + \left[\omega^2 - \Omega^2\right] \Delta R_x = \Delta A_x$$

$$\Delta \ddot{R}_y + \omega^2 \Delta R_y = \Delta A_y$$

$$\Delta \ddot{R}_z - 2\Omega \Delta \dot{R}_x - \left[\Omega^2 + 2\omega^2\right] \Delta R_z = \Delta A_z \qquad (2.5.96)$$

where $\omega^2 \triangleq \frac{\mu}{R_o^3}$ = local Schuler frequency

This set of equations is perhaps the most important set of equations in inertial navigation. Note that these equations are in general coupled linear differential equations having time varying coefficients. The time variation is caused by two parameters, ω and Ω. (Since, the local Schuler frequency (ω^2) is (μ/R_o^3), it is a function of the vehicle's true radius from the inverse gravitational force field.) The approximation of ω as a constant is quite accurate, of course, for earth fixed inertial reference frames (pre-launch), systems cruising at constant altitude, and systems whose trajectories are circular or near circular orbits. The approximation is usually sufficiently accurate for analyses of a great many more missions, however, as illustrated by the fact that the local Schuler frequency varies by only 4 percent for a vehicle traveling from its launch pad to a 100 n.mi. circular orbit.

It will be recalled that Equation (2.5.96) was derived under the assumption that the angular velocity of the local vertical reference frame could be represented by $\Omega^2 \underline{e}_y$, hence requiring vehicle motion to lie in a non-rotating plane containing the gravitational mass center. For constant cruise conditions (at constant altitude) and for circular orbits, Ω is a constant. For many other situations, however, Ω_y is a slowly varying quantity; therefore, the behavior of the inertial navigation errors deducible from assuming $\underline{\Omega} \doteq \Omega_y {}^2\underline{e}_y$ constant may be expected to display some of the fundamental characteristics of pure inertial navigation.

Equation (2.5.96) displays perhaps the most significant feature of inertial navigation within a gravitational force field.

For the situation in which the local vertical coordinate frame has a low value of angular velocity, the Equation (2.5.96) assumes the form (setting $\Omega = 0$):

$$\begin{aligned} \Delta\ddot{R}_x + \omega^2 \Delta R_x &= \Delta A_x \\ \Delta\ddot{R}_y + \omega^2 \Delta R_y &= \Delta A_y \\ \Delta\ddot{R}_z - 2\omega^2 \Delta R_z &= \Delta A_z \end{aligned} \qquad (2.5.97)$$

These equations demonstrate the boundedness of the horizontal components, $\Delta R_x(t)$ and $\Delta R_y(t)$, and the instability of the vertical channel $\Delta R_z(t)$, considering the homogeneous solutions (or those having zero forcing functions). This instability of the vertical channel is a fundamental deficiency of pure inertial navigation systems when the vehicle velocity is less than orbital velocity, and is the fundamental reason why the lower velocity inertial navigation systems (aircraft, submarine, cruise missiles, etc.) are augmented with additional sensors to provide damping and convergence of the vertical channel (often pressure transducers or radar altimeters). The vertical errors are seen to be dominated by the term $\exp \sqrt{2}\,\omega t$ due to the positive root of the characteristic equation for the vertical channel.

Another situation of interest and applicability is the case in which the vehicle is in a circular orbit, i.e., $\Omega = \omega$. Equation (2.5.94) then simplifies to:

$$\begin{aligned} \Delta\ddot{R}_x + 2\omega\Delta\dot{R}_z &= \Delta A_x \\ \Delta\ddot{R}_y + \omega^2 \Delta R_y &= \Delta A_y \\ \Delta\ddot{R}_z - 2\omega\Delta\dot{R}_x - 3\omega^2 \Delta R_z &= \Delta A_z \end{aligned} \qquad (2.5.98)$$

The equation $\Delta R_y(t)$ again exhibits the characteristic of oscillation at the local Schuler frequency ω. The homogeneous solution to the above equations can be easily found by taking the Laplace transform of the equations, using Cramer's rule to uncouple the equations, and by taking the inverse Laplace transform to generate the time solution as:

$$\Delta R_x = 6(\sin\omega t - \omega t)\Delta R_z(0) + \Delta R_x(0)$$

$$-\frac{2}{\omega}(1-\cos\omega t)\Delta\dot{R}_z(0) + \frac{1}{\omega}(4\sin\omega t - 3\omega t)\Delta\dot{R}_x(0) \qquad (2.5.99)$$

$$\Delta R_y = \Delta R_y(0)\cos\omega t + \frac{1}{\omega}\sin\omega t\ \Delta\dot{R}_y(0)$$

$$\Delta R_z = (4-3\cos\omega t)\Delta R_z(0) + \frac{1}{\omega}\sin\omega t\ \Delta\dot{R}_z(0)$$

$$+\frac{2}{\omega}(1-\cos\omega t)\Delta\dot{R}_x(0)$$

From these equations, it is noted that it is the horizontal channel (ΔR_x along the velocity vector) which is unstable, while the vertical channel ΔR_z is oscillatory and the other horizontal channel ΔR_y is oscillatory.

The instability of the tangential channel in circular orbital flight is thus seen to be a fundamental deficiency of pure inertial navigation, and requires the coupling of the pure inertial navigation system with auxiliary sensors to provide additional information. Optical navigation sightings, star trackers, horizon scanners, etc., have been employed to form suitable augmented inertial navigation systems having the desired characteristics of noise and error suppression.

Additional motivations for augmenting the pure inertial navigation systems result from a consideration of the way in which noise propagates through the system. As will be discussed in the monograph concerned with filter theory, the undamped inertial navigation channels have the property that, when subjected to a stationary zero mean noise, the output of the channels is nonstationary noise having an rms value which increases with the square root of time. Thus, even though the accelerometers and gyros are biased such that their noise characteristics have zero mean values, the random component of the errors still leads to unbounded channel errors.

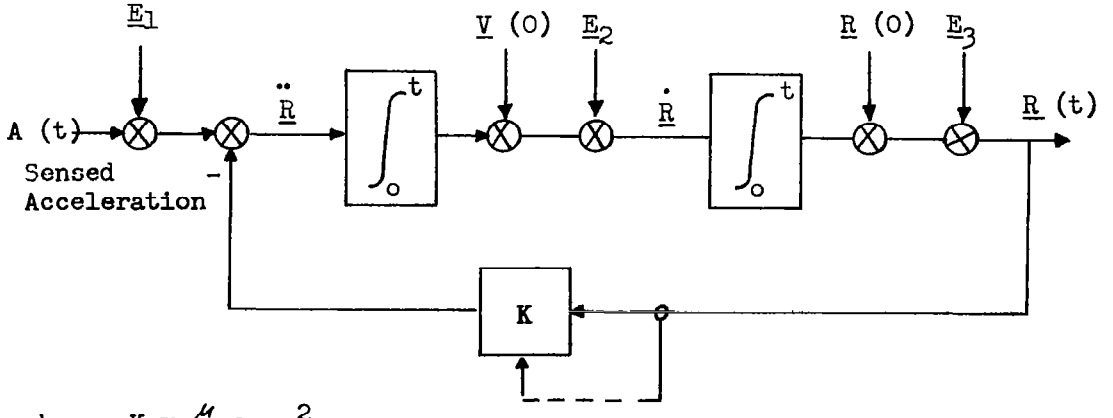
2.5.2.3 Classes of Inertial System Errors

The analysis and evaluation of errors represents a major task in inertial navigation system analysis. Since many of the most stringent requirements imposed upon inertial navigation systems are accuracy requirements, a discussion of the error characteristics in inertial navigation systems is warranted.

A detailed discussion of the errors is impossible within the space allocated here; hence, attention must be focused upon the essential and fundamental characteristics which provide some insight into the errors problem. References (8), (9), and (10) are notable for the simplicity and thoroughness of the discussions of the physical sources of errors and their effects, and for the derivations and solutions of a number of error equations describing practical inertial system mechanizations.

Some of the fundamental characteristics of pure inertial navigation systems will now be discussed. It should probably be emphasized again that in practice the pure inertial navigation system is often augmented with additional sensing devices to provide stability and reduce the effects of errors in the system. Nevertheless, the pure inertial navigation system has been selected as the means for discussing navigation system errors, since it is of considerable interest in its own right, and since the approach to an evaluation of the error effects may be applied to any particular situation.

The computation of position and velocity in a pure inertial system is represented in the sketch below:



where $K = \frac{\mu}{R^3} = \omega_s^2$

$\underline{A}$ (t) represents the "sensed" acceleration $(\ddot{\underline{R}} - \underline{G})$, and the symbols E_1, E_2, and E_3 identify three different classes of errors.

$\underline{E_1}$ represents an error in the indicated or sensed acceleration $(\ddot{\underline{R}} - \underline{G})$ and may be constant (a bias) or may be a function of time.

$\underline{E_2}$ represents a component of error in vehicle velocity and may be constant, in which case it amounts to an error in the initial velocity $\underline{V}$ (o), or it may vary as a function of time. Note that the error $\underline{E_2}$ must be summed with the integrated effects of errors $\underline{E_1}$ and $\underline{E_3}$ to derive an expression for the total velocity error.

$\underline{E_3}$ represents a component of error in vehicle position and may be constant, in which case it amounts to an error in the initial position $\underline{R}$ (o), or it may vary as a function of time. Note that the error $\underline{E_3}$ must be summed with the integrated effects of $\underline{E_1}$ and $\underline{E_2}$ to derive an expression for the total position error.

The reader is cautioned against concluding from this block diagram that the effects of all constant errors produce sinusoidal oscillations of the navigation loop, since the "feedback gain" k representing the gravity computation varies with altitude errors. It is this variation of k with altitude errors which drives the vertical channel and tangential channel unstable in the low velocity and orbital velocity situations (derived and discussed in the previous section).

The effects of each of these errors can be evaluated by applying the equations (2.5.96) derived in the previous section.

Although these general equations are readily solvable through the Laplace transform techniques, the simple case for $\Omega = 0$ is solvable almost by inspection. For this case, Equations (2.5.96) reduce to those given in Equations (2.5.97). The solutions to these equations, assuming ΔA_X, ΔA_Y, and ΔA_Z to be constants, are:

$$\Delta R_X = \left[\Delta R_X(0) - \frac{\Delta A_X}{\omega^2}\right] \cos \omega t + \frac{\Delta \dot{R}_X(0)}{\omega^2} \sin \omega t + \frac{\Delta A_X}{\omega^2}$$

$$\Delta R_Y = \left[\Delta R_Y(0) - \frac{\Delta A_Y}{\omega^2}\right] \cos \omega t + \frac{\Delta \dot{R}_X(0)}{\omega^2} \sin \omega t + \frac{\Delta A_Y}{\omega^2} \qquad (2.5.100)$$

$$\Delta R_Z = \frac{1}{2}\left[R_Z(0) + \frac{\dot{R}_Z(0)}{\omega\sqrt{2}} + \frac{\Delta A_Z}{2\omega^2}\right] e^{\omega\sqrt{2}t} + \frac{1}{2}\left[R_Z(0) + \frac{A_Z}{2\omega^2}\right] e^{-\omega\sqrt{2}t} - \frac{\Delta A_Z}{2\omega^2}$$

If the vector components of the error vectors $\underline{E_1}$ are denoted E_{1_x}, E_{1_y}, and E_{1_z} in a local vertical coordinate frame (similarly for $\underline{E_2}$ and $\underline{E_3}$), and if all are assumed small, then Equation (2.5.100) may be written:

$$\Delta R_x = \left(E_{3x} - \frac{E_{1x}}{\omega^2}\right)\cos\omega t + \frac{E_{2x}}{\omega^2}\sin\omega t + \frac{E_{1x}}{\omega^2}$$

$$\Delta R_y = \left(E_{3y} - \frac{E_{1y}}{\omega^2}\right)\cos\omega t + \frac{E_{2y}}{\omega^2}\sin\omega t + \frac{E_{1y}}{\omega^2}$$

$$\Delta R_z = \frac{1}{2}\left[E_{3z} + \frac{E_{2z}}{\omega\sqrt{2}} + \frac{E_{1z}}{2\omega^2}\right]e^{\omega\sqrt{2}t} + \frac{1}{2}\left[E_{3z} + \frac{E_{1z}}{2\omega^2}\right]e^{-\omega\sqrt{2}t} - \frac{E_{1z}}{2\omega^2} \qquad (2.5.101)$$

Several interesting features of Equations (2.5.101) merit attention. Position and acceleration errors may be made to cancel their effects in the horizontal channels provided $\Delta R_x(o) = \Delta A_x/\omega^2$ and $\Delta R_y(o) = \Delta A_y/\omega^2$. This effect suggests a technique for the initial alignment of platforms to eliminate the error propagation of leveling errors. All errors in the horizontal channels lead to oscillatory errors. Errors in the vertical channel lead to an unbounded altitude error. A similar analysis procedure can be used to evaluate errors for $\Omega \neq 0$. The reader interested in more details of vertical and horizontal channel errors is referred to References (8), (9), and (10).

The individual errors that contribute to errors $\underline{E_1}$, $\underline{E_2}$, $\underline{E_3}$, and $\underline{\Delta A}$ in the above sketch are produced by initial position and velocity errors, and errors in the inertial sensors themselves. The way in which the sensor errors contribute to these errors depends upon the mechanization of the inertial navigation system and the techniques used for initially aligning it. Such detailed considerations are discussed in References (8), (9), and (10).

Some of the more important sensor errors which must be evaluated in deriving expressions for the system errors are:

1. Accelerometer bias errors.
2. Accelerometer scale-factor errors.
3. Accelerometer non-linearity errors.
4. Misalignment error between the nominal inertial axis and the body axis of the accelerometer system.
5. Fixed gyro drift.
6. Gyro drift that is proportional to acceleration forces along input axis, acting on mass unbalance along the spin reference axis.
7. Gyro drift errors that are proportional to the acceleration forces along the spin reference axis, acting on the mass unbalance along input axis.
8. Gyro drift errors that are proportional to acceleration forces along the spin reference axis acting on mass unbalance along output axis.
9. Gyro drift errors resulting from an isoelasticity of the physical structure of the gyro.
10. Fixed gyro drift errors.
11. Misalignment error between the inertial reference axis and the input axis of the body mounted gyros.
12. Initial error sources due to the initial uncertainty in position and velocity of the vehicle.

Although deterministic error effects have been discussed in this section, a discussion of the techniques for conducting a statistical evaluation of errors is considered beyond the scope of this work. Reference (10) and the references listed from the technical literature provide an introduction to these techniques.

2.6 STATE DETERMINATION FROM EVENLY DETERMINED DATA

The process of determining the n state vector components from n pieces of independent data is called state determination from evenly determined data, as contrasted with the process of determining the n state vector components from a set of redundant data in which smoothing and averaging techniques are used to remove some of the uncertainty from the data.

If perfect measurements could be taken, and if there were no uncertainties in the physical constants, etc., and if all external forces acting upon vehicles could be accounted for, and if the complex equations of motion could be solved by even on-board vehicle computers, there would be no need to make more than six independent navigation measurements to determine a vehicle's motion, since motion would be completely predictable. The fact that measurements, miscellaneous forces, physical constants, etc., are subject to uncertainties, and the fact that computers suffer such limitations of speed, truncation errors, etc., demands that some redundancy in the information concerning a vehicle's position and velocity be available so that effects of errors may be averaged out, filtered, smoothed, etc. The filtering theory useful for this filtering and smoothing of navigation observations will be presented in a subsequent monograph. Even though measurements, physical constants, etc.,are subject to uncertainties, often an excellent value of vehicle position or of vehicle position and velocity may be determined by using evenly determined measurement data. Certainly a first step in an evaluation of the number of measurements required to attain a desired accuracy is to evaluate the accuracy of those techniques which require no filtering or smoothing of data or results, i.e., those measurements which constitute an evenly determined set of data.

In the following sections, the two basic techniques of deterministic navigation are presented, namely, determination of a vehicle's position by using data from three independent simultaneous measurements, and the determination of a vehicle's position and velocity by using data from six independent measurements.

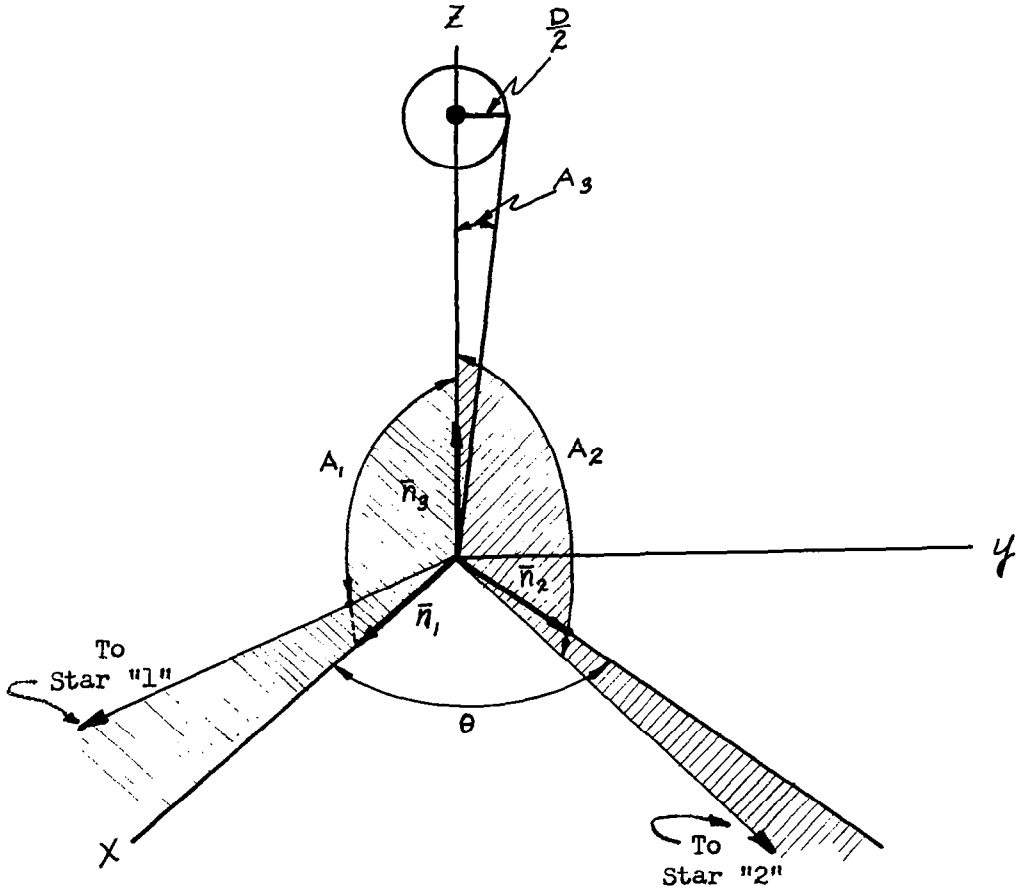
2.6.1 Simultaneous Measurements

In Section 2.4 it was shown that any single measurement can contribute a limited amount of information in determining the position deviation vector $\delta\underline{R}$. The resulting expressions showed that the measurement deviation corresponds to finding a component of $\delta\underline{R}$ in a particular direction. The direction is determined by the geometry of the measurement and is specified by the equation relating δq and $\delta\underline{R}$. In order to determine $\delta\underline{R}$ uniquely, it is obvious that at least three independent such measurements must be taken. Three measurements are considered independent if the components of $\delta\underline{R}$ which they determine are non-colinear and non-coplanar.

In the following section the techniques for computing vehicle position from practical combinations of three independent measurements are presented.

2.6.1.1 Planet-Star, Planet-Star, Planet-Diameter Measurement

The techniques discussed in Section 2.4 may be applied to the determination of position given three simultaneous independent measurements. Consider the following sketch:



In this case

$$\delta A_1 = \frac{\underline{\eta} \cdot \delta \underline{R}}{Z}$$

$$\delta A_2 = \frac{\underline{\eta}_2 \cdot \delta \underline{R}}{Z} \qquad (2.6.1)$$

$$\delta A_3 = \frac{D \underline{\eta}_3 \cdot \delta \underline{R}}{Z^2 \cos(A/2)}$$

Expressed in matrix form, these equations are

$$\begin{bmatrix} \delta A_1 \\ \delta A_2 \\ \delta A_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{Z} & 0 & 0 \\ 0 & \frac{1}{Z} & 0 \\ 0 & 0 & \frac{D}{Z^2 \cos(A/2)} \end{bmatrix} \begin{bmatrix} \underline{\eta}_1 \cdot \delta \underline{R} \\ \underline{\eta}_2 \cdot \delta \underline{R} \\ \underline{\eta}_3 \cdot \delta \underline{R} \end{bmatrix} \qquad (2.6.2)$$

Remembering that the $\underline{\eta}$ vectors are perpendicular to the planet direction, the X axis is now chosen such that it is in the $\underline{\eta}_1$ direction. Similarly, the Z axis is chosen to be positive in the $\underline{\eta}_3$ direction. If $\underline{\eta}_2$ differs from $\underline{\eta}_1$ by a rotation Θ about +Z, the unit vectors $\underline{\eta}_1$, $\underline{\eta}_2$ and $\underline{\eta}_3$ may be written in terms of the unit vectors

$$\underline{\eta}_1 = \underline{\imath}$$

$$\underline{\eta}_2 = \cos\Theta \underline{\imath} + \sin\Theta \underline{\jmath} \qquad (2.6.3)$$

$$\underline{\eta}_3 = \underline{k}$$

Using these relations, it is seen that in terms of rectangular cartesian components Equation (2.6.2) becomes

$$\begin{bmatrix} \delta A_1 \\ \delta A_2 \\ \delta A_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{Z} & 0 & 0 \\ \frac{\cos\Theta}{Z} & \frac{\sin\Theta}{Z} & 0 \\ 0 & 0 & \frac{D}{Z^2 \cos\left(\frac{A}{2}\right)} \end{bmatrix} \begin{bmatrix} \underline{\imath} \cdot \delta \underline{R} \\ \underline{\jmath} \cdot \delta \underline{R} \\ \underline{k} \cdot \delta \underline{R} \end{bmatrix} \qquad (2.6.4)$$

where

$$\begin{bmatrix} \underline{\ell} \cdot \delta \underline{B} \\ \underline{j} \cdot \delta \underline{R} \\ \underline{k} \cdot \delta \underline{R} \end{bmatrix} = \begin{bmatrix} \delta \underline{R} \\ \\ \end{bmatrix}$$

$$\begin{bmatrix} \delta A_1 \\ \delta A_2 \\ \delta A_3 \end{bmatrix} = \begin{bmatrix} \delta \underline{\rho} \\ \\ \end{bmatrix}$$

and

$$\begin{bmatrix} \frac{1}{z} & 0 & 0 \\ \frac{\cos\theta}{z} & \frac{\sin\theta}{z} & 0 \\ 0 & 0 & \frac{D}{z^2 \cos\left(\frac{A}{2}\right)} \end{bmatrix} = \underline{H}$$

In order to evaluate $|\underline{\epsilon}|$, the magnitude of the position deviation error, it is necessary to first compute $\underline{H}^{-1}$ and $(\underline{H}^{-1})^T$.

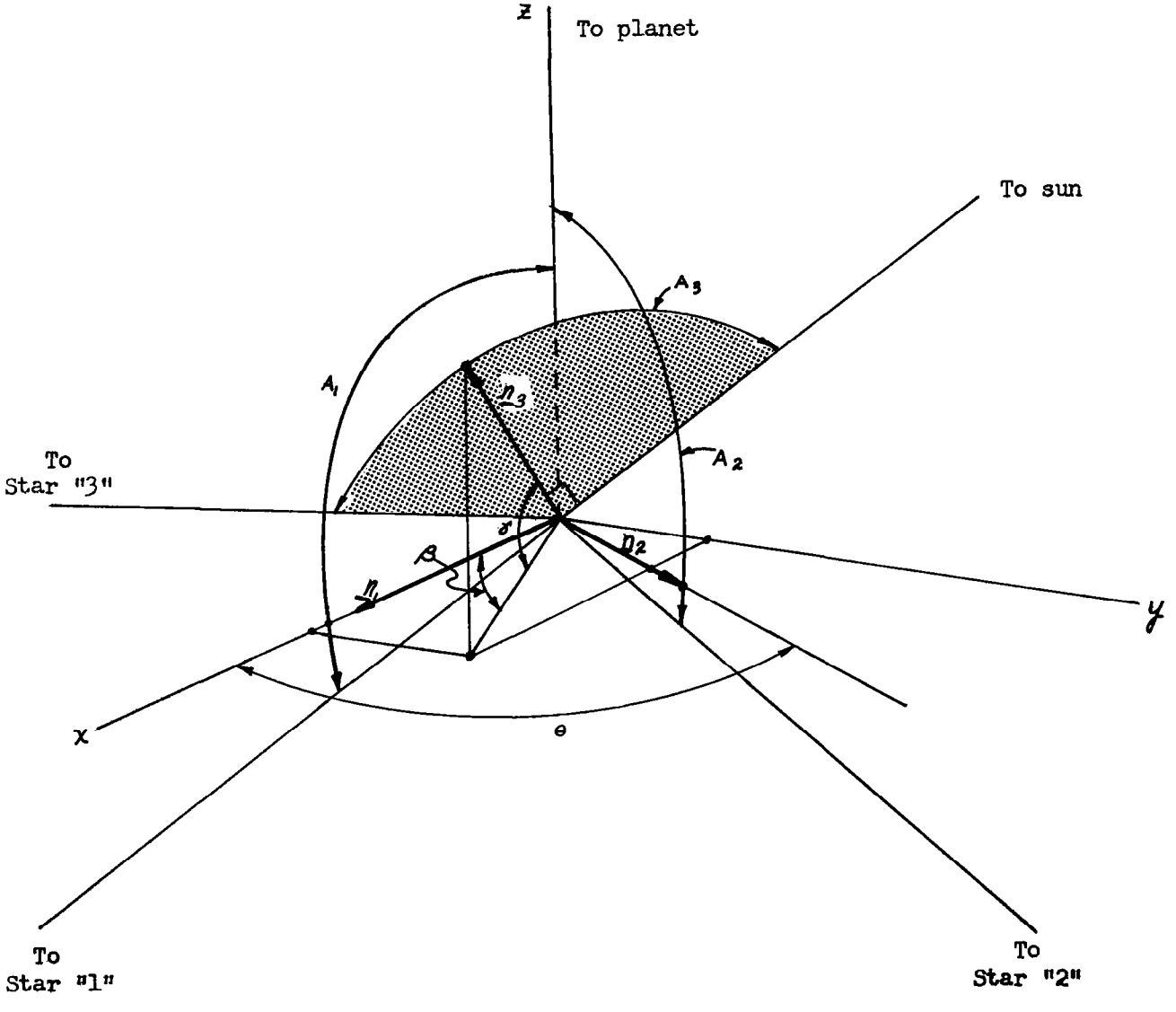
Any one of a number of methods can be used to evaluate $\underline{H}^{-1}$ and $(\underline{H}^{-1})^T$. The reader is referred to any matrix algebra book. The results are stated below.

$$\underline{H}^{-1} = \begin{bmatrix} z & 0 & 0 \\ -z\cot\theta & z\csc\theta & 0 \\ 0 & 0 & \frac{z^2\cos\frac{A}{2}}{D} \end{bmatrix} \qquad (2.6.5)$$

$$\left(\underline{H}^{-1}\right)^T = \begin{bmatrix} z & -z\cot\theta & 0 \\ 0 & z\csc\theta & 0 \\ 0 & 0 & \frac{z^2\cos\frac{A}{2}}{D} \end{bmatrix}$$

2.6.1.2 Planet-Star, Planet-Star, Sun-Star Measurement

The geometry of this measurement is similar to the planet-star, planet-star, planet diameter measurement in that the $\underline{\eta}_1$ and $\underline{\eta}_2$ vectors have the same meaning and orientation. The planet diameter measurement is replaced by a measurement of the angle between a third star and the sun. This measurement determines the component of $\delta\underline{\rho}$ in a direction which is perpendicular to the direction to the sun from the vehicle. This direction is represented by unit vector $\underline{\eta}_3$ and is shown in the following sketch.



Using the relations for planet-star and sun-star measurements that were determined earlier, the deviation equations are obtained.

$$\delta A_1 = \frac{\underline{\eta}_1 \cdot \delta \underline{R}}{Z}$$

$$\delta A_2 = \frac{\underline{\eta}_2 \cdot \delta \underline{R}}{Z} \qquad (2.6.6)$$

$$\delta A_3 = \frac{\underline{\eta}_3 \cdot \delta \underline{R}}{R}$$

where Z is the distance from the spacecraft to the planet and R is the distance from the spacecraft to the sun. In matrix form these equations may be written as:

$$\begin{bmatrix} \delta A_1 \\ \delta A_2 \\ \delta A_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{Z} & 0 & 0 \\ 0 & \frac{1}{Z} & 0 \\ 0 & 0 & \frac{1}{R} \end{bmatrix} \begin{bmatrix} \underline{\eta}_1 \cdot \delta \underline{R} \\ \underline{\eta}_2 \cdot \delta \underline{R} \\ \underline{\eta}_3 \cdot \delta \underline{R} \end{bmatrix} \qquad (2.6.7)$$

Just as in previous work, the X axis may be chosen in the direction of $\underline{\eta}_1$.

Now

$$\underline{\eta}_1 = \underline{\imath}$$

and

$$\underline{\eta}_2 = \cos\theta\,\underline{\imath} + \sin\theta\,\underline{\jmath} \qquad (2.6.8)$$

Using the angles β and γ defined in the sketch, the unit vector $\underline{\eta}_3$ may be similarly expressed in terms of rectangular cartesian coordinates.

$$\underline{\eta}_3 = \cos\beta\cos\gamma\,\underline{\imath} + \sin\beta\cos\gamma\,\underline{\jmath} + \sin\gamma\,\underline{k} \tag{2.6.9}$$

The equations for δA_2 and δA_3 now become

$$\delta A_2 = \frac{\underline{\eta}_2 \cdot \delta\underline{R}}{z} = \frac{\cos\theta}{z}\,\underline{\imath}\cdot\delta\underline{R} + \frac{\sin\theta}{z}\,\underline{\jmath}\cdot\delta\underline{R} \tag{2.6.10}$$

$$\delta A_3 = \frac{\underline{\eta}_3 \cdot \delta\underline{R}}{R} = \frac{\cos\beta\cos\gamma}{R}\,\underline{\imath}\cdot\delta\underline{R} + \frac{\sin\beta\cos\gamma}{R}\,\underline{\jmath}\cdot\delta\underline{R} + \frac{\sin\gamma}{R}\,\underline{k}\cdot\delta\underline{R} \tag{2.6.11}$$

Now Equation (2.6.7) becomes

$$\begin{bmatrix} \delta A_1 \\ \delta A_2 \\ \delta A_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{z} & 0 & 0 \\ \frac{\cos\theta}{z} & \frac{\sin\theta}{z} & 0 \\ \frac{\cos\beta\cos\gamma}{R} & \frac{\sin\beta\cos\gamma}{R} & \frac{\sin\gamma}{R} \end{bmatrix} \begin{bmatrix} \underline{\imath}\cdot\delta\underline{R} \\ \underline{\jmath}\cdot\delta\underline{R} \\ \underline{k}\cdot\delta\underline{R} \end{bmatrix} \tag{2.6.12}$$

which is now in the form of

$$\delta\underline{q} = H\cdot\delta\underline{R} \tag{2.6.13}$$

where

$$\delta\underline{r} = \begin{bmatrix} \underline{\imath}\cdot\delta\underline{R} \\ \underline{\jmath}\cdot\delta\underline{R} \\ \underline{k}\cdot\delta\underline{R} \end{bmatrix}$$

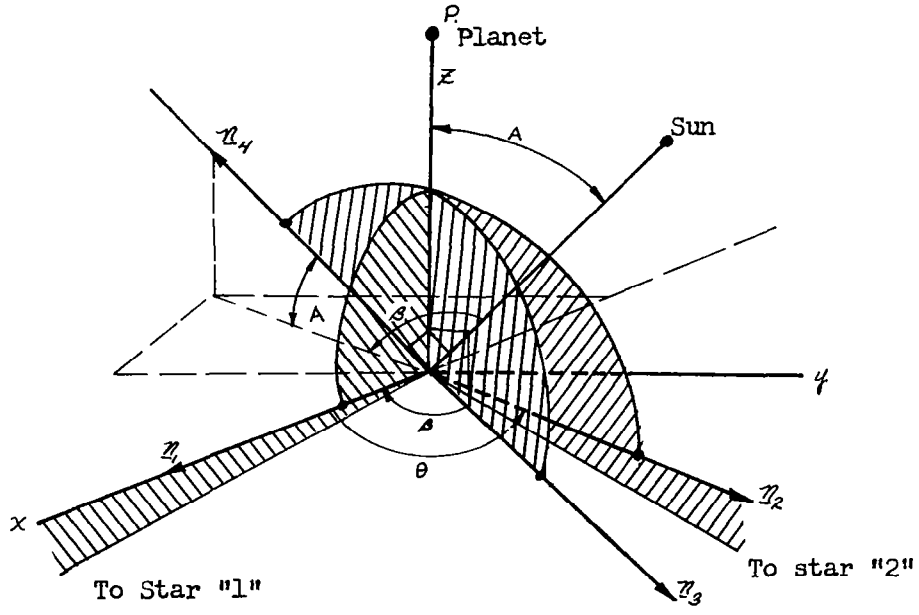
$$= H^{-1}\,\delta\underline{q}$$

and where H^{-1} is

$$H^{-1} = \begin{bmatrix} z & 0 & 0 \\ -z\cot\theta & z\csc\theta & 0 \\ z\cot\gamma\csc\theta\sin(\beta-\theta) & -z\cot\gamma\sin\beta\csc\theta & R\csc\gamma \end{bmatrix} \qquad (2.6.14)$$

2.6.1.3 Planet-Star, Planet-Star, Sun-Planet Measurement

The analysis of the planet-star, planet-star, sun-planet measurement is basically the same as the previous two except for the fact that the nature of the measurements are such that the components of the position deviation vector are originally found in terms of a linearly dependent set of base vectors, $\bar{\eta}_1$, $\bar{\eta}_2$, $\bar{\eta}_3$, $\bar{\eta}_4$. The geometry of this measurement is shown below:



After the base vectors are expressed in terms of a linearly independent set of base vectors, $\underline{i}$, $\underline{j}$, and $\underline{k}$, the analysis is very similar to the previously discussed cases. Therefore, only a brief outline of the analysis will be given.

The standard equations for the measurements are

$$\begin{aligned} \delta A_1 &= \frac{\underline{\eta}_1 \cdot \delta \underline{R}}{z} \\ \delta A_2 &= \frac{\underline{\eta}_2 \cdot \delta \underline{R}}{z} \\ \delta A_3 &= \frac{\underline{\eta}_4 \cdot \delta \underline{R}}{r} + \frac{\underline{\eta}_3 \cdot \delta \underline{R}}{z} \end{aligned} \tag{2.6.15}$$

From the sketch it is seen that

$$\begin{aligned} \underline{\eta}_1 &= \underline{i} \\ \underline{\eta}_2 &= \cos\theta\, \underline{i} + \sin\theta\, \underline{j} \\ \underline{\eta}_3 &= \cos\beta\, \underline{i} + \sin\beta\, \underline{j} \\ \underline{\eta}_4 &= -\cos\gamma^* \cos\beta\, \underline{i} - \cos\gamma^* \sin\beta\, \underline{j} + \sin\gamma^*\, \underline{k} \end{aligned} \tag{2.6.16}$$

Using these relations, Equations (2.6.15) become

$$\begin{aligned} \delta A_1 &= \frac{\underline{i} \cdot \delta \underline{R}}{z} \\ \delta A_2 &= \frac{\cos\theta}{z}\, \underline{i} \cdot \delta \underline{R} + \frac{\sin\theta}{z}\, \underline{j} \cdot \delta \underline{R} \\ \delta A_3 &= \frac{-\cos\gamma^* \cos\beta}{R}\, \underline{i} \cdot \delta \underline{R} - \frac{\cos\gamma^* \sin\beta}{r}\, \underline{i} \cdot \delta \underline{R} \\ &\quad + \frac{\sin\gamma^*}{R}\, \underline{k} \cdot \delta \underline{R} + \frac{\cos\beta\, \underline{i} \cdot \delta \underline{R}}{z} + \frac{\sin\beta\, \underline{j} \cdot \delta \underline{R}}{z} \end{aligned} \tag{2.6.17}$$

or

$$\begin{bmatrix} \delta A_1 \\ \delta A_2 \\ \delta A_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{z} & 0 & 0 \\ \frac{\cos\theta}{z} & \frac{\sin\theta}{z} & 0 \\ \frac{\cos\beta}{z} - \frac{\cos\gamma^* \cos\beta}{R} & \frac{\sin\beta}{z} - \frac{\cos\gamma^* \sin\beta}{R} & \frac{\sin A}{R} \end{bmatrix} \begin{bmatrix} \\ \delta\underline{R} \\ \\ \end{bmatrix} \qquad (2.6.18)$$

The matrix involved in equation (2.6.18) is $\underline{H}$.

$\underline{H}^{-1}$ can be determined to be

$$\left[\begin{array}{c|c|c} z & 0 & 0 \\ \hline -z\cot\theta & z\csc\theta & 0 \\ \hline z\cot\gamma^*(\cos\beta - \sin\beta\cot\theta) & \frac{-r\sin\beta}{\sin\theta\sin\gamma^*} & r\csc\gamma^* \\ -r\csc\gamma^*(\cos\beta - \sin\beta\cot\theta) & + \frac{z\cot\gamma^*\sin\beta}{\sin\theta} & \end{array}\right] \qquad (2.6.19)$$

In section 2.8.2.3 this matrix is used to evaluate the average of the error magnitude. In the analysis, the diagonal components of $[\underline{H}^{-1}]^T[\underline{H}^{-1}]$ are needed. Letting h_{ij} be the elements of $\left[\underline{H}^{-1}\right]^T\left[\underline{H}^{-1}\right]$, the diagonal terms are equal to

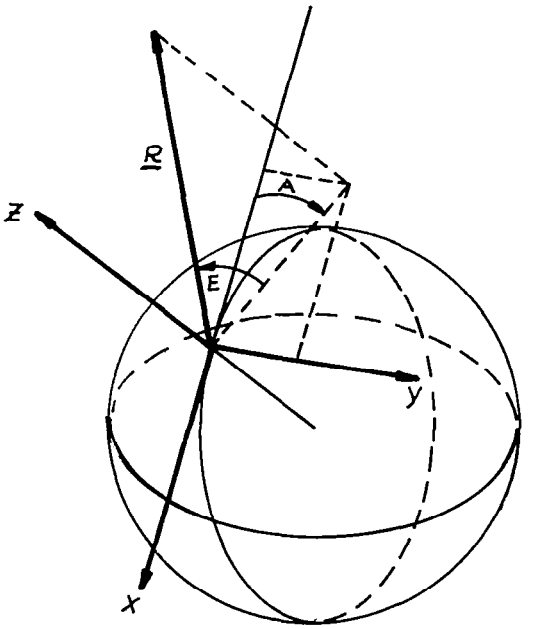
$$h_{11} = z^2 + z^2\cot^2\theta + (\cos\beta - \sin\beta\cot\theta)^2(z\cos\gamma^* - R\csc\gamma^*)^2$$

$$h_{22} = z^2\csc^2\theta + \frac{\sin^2\beta}{\sin^2\theta}\left(z\cot\gamma^* - \frac{r}{\sin\gamma^*}\right)^2 \qquad (2.6.20)$$

$$h_{33} = R^2\csc^2\gamma^*$$

2.6.1.4 Radar Range, Azimuth, Elevation Measurement

In a manner similar to the optical means of state deviation previously discussed, radio techniques may also be employed. Usually in radio measurements the observer is at some radar station and is observing some point in the atmosphere or in space. If, when tracking a vehicle, slight deviations from the nominal trajectory occur, the small deviations measured on the ground radar, such as angular deviations from nominal azimuth, can be used to calculate the state vector deviation as follows: Assume a topodetic or radar AZ - EL coordinate system erected at some site on the earth as illustrated below:



Any position vector can be represented in spherical polar form by specifying its radar range R, an azimuth angle A, and an elevation angle E. Alternately, the same vector may be expressed in rectangular cartesian coordinate as follows:

$$\underline{R} = R \begin{bmatrix} -\cos E \cos A \\ \cos E \sin A \\ \sin E \end{bmatrix} \tag{2.6.21}$$

A small position deviation in the position vector may be related to spherical polar coordinate deviations by

$$\delta \underline{R} = \frac{\partial \underline{R}}{\partial R} \delta R + \frac{\partial \underline{R}}{\partial E} \delta E + \frac{\partial \underline{R}}{\partial A} \delta A \tag{2.6.22}$$

The terms in Equation (2.6.22) may be evaluated as follows:

$$\frac{\partial \underline{R}}{\partial R} = \begin{bmatrix} -\cos E \cos A \\ \cos E \sin A \\ \sin E \end{bmatrix} \tag{2.6.23}$$

$$\frac{\partial \underline{R}}{\partial E} = R \begin{bmatrix} \sin E \cos A \\ -\sin E \sin A \\ \cos E \end{bmatrix} \tag{2.6.24}$$

$$\frac{\partial \underline{R}}{\partial A} = R \begin{bmatrix} \cos E \sin A \\ \cos E \cos A \\ 0 \end{bmatrix} \tag{2.6.25}$$

Note that

$$\left| \frac{\partial \underline{R}}{\partial R} \right| = 1 \tag{2.6.26}$$

$$\left| \frac{\partial \underline{R}}{\partial E} \right| = R \tag{2.6.27}$$

$$\left| \frac{\partial \underline{R}}{\partial A} \right| = R \cos E \tag{2.6.28}$$

$$\frac{\partial \underline{R}}{\partial R} \cdot \frac{\partial \underline{R}}{\partial E} = \frac{\partial \underline{R}}{\partial R} \cdot \frac{\partial \underline{R}}{\partial A} = 0 \tag{2.6.29}$$

In order to evaluate δR, the scalar product of $\frac{\partial \underline{R}}{\partial R}$ and Equation (2.6.22) can be used as follows:

$$\frac{\partial \underline{R}}{\partial R} \cdot \delta \underline{R} = \frac{\partial \underline{R}}{\partial R} \cdot \frac{\partial \underline{R}}{\partial R} \delta R + \cancel{\frac{\partial \underline{R}}{\partial R} \cdot \frac{\partial \underline{R}}{\partial E}}^{\,0} \delta E + \cancel{\frac{\partial \underline{R}}{\partial R} \cdot \frac{\partial \underline{R}}{\partial A}}^{\,0} \tag{2.6.29}$$

$$\delta r = \frac{\frac{\partial \underline{R}}{\partial R} \cdot \delta \underline{R}}{\left| \frac{\partial \underline{R}}{\partial R} \right|^2} = \frac{\partial \underline{R}}{\partial R} \cdot \delta \underline{R} \tag{2.6.30}$$

Expressed in terms of column and row vectors, Equation (2.6.30) becomes

$$\delta R = \begin{bmatrix} -\cos E \cos A & \cos E \sin A & \sin E \end{bmatrix} \begin{bmatrix} \delta \underline{R} \\ \end{bmatrix} \tag{2.6.31}$$

Similarly, δE and δA can be evaluated. The scalar product of $\frac{\partial \underline{R}}{\partial E}$ and Equation (2.6.22) yields

$$\frac{\partial \underline{R}}{\partial E} \cdot \delta \underline{R} = \frac{\partial \underline{R}}{\partial E} \cdot \frac{\partial \underline{R}}{\partial R} \delta R + \frac{\partial \underline{R}}{\partial E} \cdot \frac{\partial \underline{R}}{\partial E} \delta E + \frac{\partial \underline{R}}{\partial E} \cdot \frac{\partial \underline{R}}{\partial A} \delta A \qquad (2.6.32)$$

so

$$\delta E = \frac{\frac{\partial \underline{R}}{\partial E} \cdot \delta \underline{R}}{\left| \frac{\partial \underline{R}}{\partial E} \right|^2} = \frac{\frac{\partial \underline{R}}{\partial E} \cdot \delta \underline{R}}{R^2} \qquad (2.6.33)$$

Hence

$$\delta E = \frac{1}{R} \begin{bmatrix} \sin E \cos A & -\sin E \sin A & \cos E \end{bmatrix} \begin{bmatrix} \delta \underline{R} \end{bmatrix} \qquad (2.6.34)$$

The scalar product of $\frac{\partial \underline{R}}{\partial A}$ and Equation (2.6.22) yields

$$\frac{\partial \underline{R}}{\partial A} \cdot \delta \underline{R} = \frac{\partial \underline{R}}{\partial A} \cdot \frac{\partial \underline{R}}{\partial R} \delta R + \frac{\partial \underline{R}}{\partial A} \cdot \frac{\partial \underline{R}}{\partial E} \delta E + \frac{\partial \underline{R}}{\partial A} \cdot \frac{\partial \underline{R}}{\partial A} \delta A \qquad (2.6.35)$$

$$\delta A = \frac{\frac{\partial \underline{R}}{\partial A} \cdot \delta \underline{R}}{\left| \frac{\partial \underline{R}}{\partial A} \right|^2} \qquad (2.6.36)$$

so

$$\delta A = \frac{1}{R \cos E} \begin{bmatrix} \sin A & \cos A & 0 \end{bmatrix} \begin{bmatrix} \delta \underline{R} \end{bmatrix} \qquad (2.6.37)$$

Combining Equations (2.6.31), (2.6.34), and (2.6.37) the final equation relating the position and measurement deviated from nominal for the radar range, azimuth elevation measurement is obtained.

$$\begin{bmatrix} \delta r \\ \delta E \\ \delta A \end{bmatrix} = \begin{bmatrix} -\cos E \cos A & \cos E \sin A & \sin E \\ \frac{\sin E \cos A}{R} & -\frac{\sin E \sin A}{R} & \frac{\cos E}{r} \\ \frac{\sin A}{R \cos E} & \frac{\cos A}{R \cos E} & 0 \end{bmatrix} \begin{bmatrix} \\ \delta \underline{R} \\ \\ \end{bmatrix} \tag{2.6.38}$$

Equation (2.6.38) is now in the form of the general position deviation equation

$$\delta \underline{q}_0 = \underline{H} \cdot \delta \underline{R} \tag{2.6.39}$$

Hence, vehicle position may be calculated from:

$$\delta \underline{R} = \underline{H}^{-1} \begin{bmatrix} \delta R \\ \delta E \\ \delta A \end{bmatrix} \tag{2.6.40}$$

2.6.2 Sequential Measurements

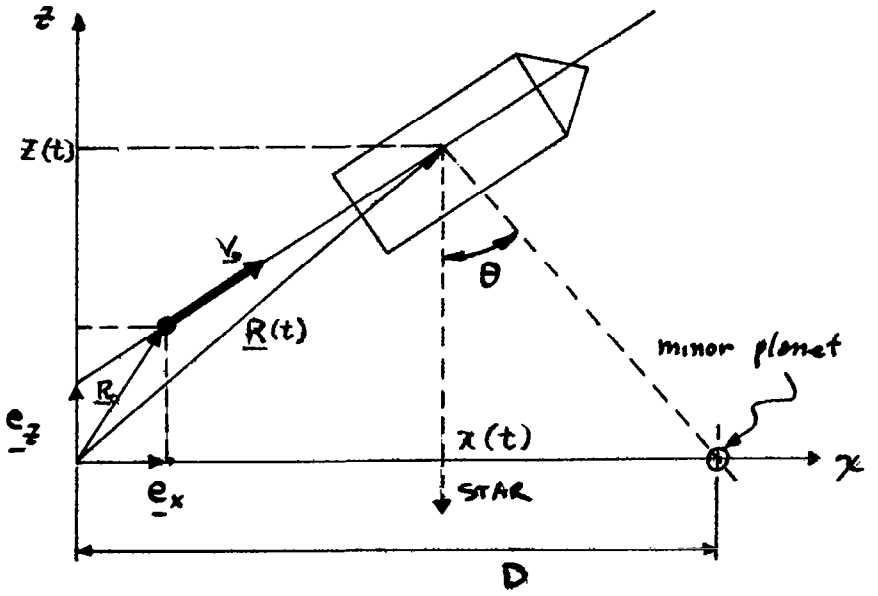
The previous section considered the determination of position from three simultaneous measurements. The determination of position from a sequence of measurements is a more difficult navigation problem. The general technique which is utilized in solution of this problem is to first assume that vehicle motion is sufficiently near a reference trajectory so that linear perturbation theory may be applied to the solution of the problem. Since terminology and trajectory mechanics complexities may delay the understanding of the basically simple concepts involved, a simple navigation example is presented in Section 2.6.2.1 to introduce the terminology and technique before the general technique is developed in Sections 2.6.2.2 and 2.6.2.3.

2.6.2.1 A Simple Example of a Navigation Problem

A simple sample problem has been prepared to illustrate the procedure and to serve as a vehicle for improving the degree of appreciation for the more precise model.

Consider the planar analysis of a spacecraft in uniform rectilinear motion. Assume that the navigation problem in this example is to compute or estimate the vehicle's position and velocity, knowing only:

1. The equations describing the motion on a nominal (or "reference") trajectory
2. The position of the minor planet from ephemeris tables
3. Successive measurements of the angle θ on the actual trajectory, and the times at which each of the measurements were made



The nominal or reference trajectory in this example is assumed to be:

$$\underline{R}_{nom}(t) = (0+3t)\,\underline{e}_x + (5+2t)\,\underline{e}_z \qquad (2.6.41)$$

Since $\underline{e}_x$ and $\underline{e}_z$ are nonrotating unit vectors, the nominal velocity is

$$V_{NOM}(t) = 3\,\underline{e}_x + 2\,\underline{e}_z \qquad (2.6.42)$$

The nominal position of the massless minor planet is assumed to be $(x,z) = (D, 0)$. Hence, from the sketch

$$\tan\theta = \frac{D-x}{z} \qquad (2.6.43)$$

Differentiating (2.6.43)

$$\sec^2\theta\,\delta\theta = -\frac{1}{z}\delta x + \frac{x-D}{z^2}\delta z$$

or

$$\delta\theta = -\frac{1}{z\sec^2\theta}\delta x + \frac{x-D}{z^2\sec^2\theta}\delta z \qquad (2.6.44)$$

or

$$\delta\theta = \begin{bmatrix} -\frac{1}{z\sec^2\theta} & \frac{x-D}{z^2\sec^2\theta} \end{bmatrix} \begin{bmatrix} \delta x \\ \delta z \end{bmatrix} \qquad (2.6.45)$$

or

$$\delta\theta = B'\,\delta\underline{R}$$

where

$$\delta\underline{R} \triangleq \begin{bmatrix} \delta x \\ \delta z \end{bmatrix} \qquad (2.6.46)$$

The basic interpretation of Equation (2.6.45) is that if a point (x,z) is chosen in the plane, then θ has a fixed value; if small deviations in the coordinate of x and z are now made, a small change in the value of θ results, as given by Equation (2.6.44) or (2.6.45).

The problem at hand is now the computation of an estimate of $\underline{R}$ (t) which satisfies the observations. This step requires consideration of the general equations of motion

$$\underline{R}(t) = (x_o + V_x t)\,\underline{e}_x + (z_o + V_z t)\,\underline{e}_z \qquad (2.6.47)$$

where x_o, z_o are the actual initial coordinates, and where V_x and V_z are actual velocity components.

Define

$$\begin{aligned}
\delta x_o &\triangleq x_o - x_{o_{NOM}} && \text{initial X position error}\\
\delta z_o &\triangleq z_o - z_{o_{NOM}} && \text{initial Z position error}\\
\delta V_x &\triangleq V_x - V_{x_{NOM}} && \text{X velocity error}\\
\delta V_z &\triangleq V_z - V_{z_{NOM}} && \text{Z velocity error}\\
\delta \underline{R}(t) &\triangleq \underline{R}(t) - \underline{R}_{NOM}(t)
\end{aligned} \qquad (2.6.48)$$

The numerical values for the nominal trajectory parameters are given by equations

$$\begin{aligned}
x_{o_{NOM}} &= 0 & V_{x_{NOM}} &= 3\\
z_{o_{NOM}} &= 5 & V_{z_{NOM}} &= 2
\end{aligned} \qquad (2.6.49)$$

Combining Equations (2.6.41), (2.6.42), (2.6.47), (2.6.48), and (2.6.49) yields

$$\delta \underline{R}(t) = \begin{bmatrix} \delta x \\ \delta z \end{bmatrix} = \begin{bmatrix} \delta x_o + t\,\delta V_x \\ \delta z_o + t\,\delta V_z \end{bmatrix} \qquad (2.6.50)$$

$$\delta \underline{V}(t) = \begin{bmatrix} \delta V_x \\ \delta V_z \end{bmatrix}$$

or

$$\begin{bmatrix} \delta \underline{R}(t) \\ \delta \underline{V}(t) \end{bmatrix} = \begin{bmatrix} \delta x \\ \delta z \\ \delta V_x \\ \delta V_z \end{bmatrix} = \begin{bmatrix} 1 & 0 & t & 0 \\ 0 & 1 & 0 & t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \delta x_0 \\ \delta z_0 \\ \delta V_{x_0} \\ \delta V_{z_0} \end{bmatrix} \quad (2.6.51)$$

Using state vector **nomenclature,** where $\underline{X}(t)$ is defined as the state vector, then

$$\underline{X}(t) = \begin{bmatrix} x(t) \\ z(t) \\ V_x(t) \\ V_z(t) \end{bmatrix} = \begin{bmatrix} \underline{R}(t) \\ \underline{V}(t) \end{bmatrix} \quad ; \quad \delta \underline{X}(t) = \begin{bmatrix} \delta x(t) \\ \delta z(t) \\ \delta V_x(t) \\ \delta V_y(t) \end{bmatrix} = \begin{bmatrix} \delta \underline{R}(t) \\ \delta \underline{V}(t) \end{bmatrix} \quad (2.6.52)$$

Equation (2.6.51) can thus be rewritten:

$$\delta \underline{X}(t) = \Phi(t, t_0)\ \delta \underline{X}(t_0) \quad (2.6.53)$$

where

$$\Phi(t,t_0) = \begin{bmatrix} 1 & 0 & t & 0 \\ 0 & 1 & 0 & t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad \delta \underline{X}(t_0) = \begin{bmatrix} \delta x_0 \\ \delta z_0 \\ \delta V_{x_0} \\ \delta V_{z_0} \end{bmatrix}$$

The matrix $\phi(t,t_0)$ is, of course, the state transition matrix. Equation (2.6.45) may be rewritten:

$$\delta\theta(t) = \left[\begin{array}{c|c|c|c} \frac{-1}{z\sec^2\theta} & \frac{x-D}{z^2\sec^2\theta} & 0 & 0 \end{array}\right] \begin{bmatrix} \delta x(t) \\ \delta z(t) \\ \delta V_x(t) \\ \delta V_z(t) \end{bmatrix} \qquad (2.6.54)$$

or

$$\delta\theta(t) = B(t)\ \delta\underline{x}(t)$$

where $$B(t) \triangleq \left[\begin{array}{c|c|c|c} \frac{-1}{z\sec^2\theta} & \frac{x-D}{z^2\sec^2\theta} & 0 & 0 \end{array}\right]$$

Hence, for measurements $\delta\theta(t_1)$, $\delta\theta(t_2)$ ... at times t_1, t_2 etc.,

$$\delta\theta(t_1) = \left[\begin{array}{c|c|c|c} \frac{-1}{z(t_1)\sec^2\theta(t_1)} & \frac{x(t_1)-D}{z^2(t_1)\sec^2\theta(t_1)} & 0 & 0 \end{array}\right] \begin{bmatrix} \delta x(t_1) \\ \delta z(t_1) \\ \delta V_x(t_1) \\ \delta V_z(t_1) \end{bmatrix} \qquad (2.6.55)$$

$$\delta\theta(t_2) = \left[\begin{array}{c|c|c|c} \dfrac{-1}{z(t_2)\sec^2\theta(t_2)} & \dfrac{x(t_2)-0}{z^2(t_2)\sec^2\theta(t_2)} & 0 & 0 \end{array}\right] \begin{bmatrix} \delta x(t_2) \\ \delta z(t_2) \\ \delta V_x(t_2) \\ \delta V_z(t_2) \end{bmatrix}$$

In these equations, the values of $X(t_1)$, $Z(t_1)$, $X(t_2)$, $Z(t_2)$, etc., may be computed along the reference trajectory, provided the deviations $\delta x(t)$ and $\delta z(t)$ from the reference trajectory are small. Hence, from Equation (2.6.41),

$$x(t_1) = 0 + 3t_1 \qquad z(t_1) = 5 + 2t_1$$

$$x(t_2) = 0 + 3t_2$$

etc.

Equations (2.6.55) thus provide a series of simultaneous equations, where each is a relationship (angle and deviation) at different times. To solve these equations requires that $\delta x(t)$, $\delta z(t)$, $\delta V_x(t)$, $\delta V_z(t)$ be expressed in terms of δx_0, δz_0, δV_{x_0}, and δV_{z_0}.

Substituting Equation (2.6.51) into Equation (2.6.55) yields

$$\delta\theta(t_1) = \left[\begin{array}{c|c|c|c} \dfrac{-1}{z(t_1)\sec^2\theta(t_1)} & \dfrac{x(t_1)-0}{z^2(t_1)\sec^2\theta(t_1)} & 0 & 0 \end{array}\right] \begin{bmatrix} 1 & 0 & t_1 & 0 \\ 0 & 1 & 0 & t_1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \delta x_0 \\ \delta z_0 \\ \delta V_{x_0} \\ \delta V_{z_0} \end{bmatrix}$$

or

$$\delta\theta(t_1) = \left[\frac{-1}{Z(t_1)\sec^2\theta(t_1)} \;\vdots\; \frac{X(t_1)-D}{Z^2(t_1)\sec^2\theta(t_1)} \;\vdots\; \frac{-t_1}{Z(t_1)\sec^2\theta(t_1)} \;\vdots\; \frac{t_1[X(t_1)-D]}{Z^2(t_1)\sec^2\theta(t_1)}\right] \begin{bmatrix}\delta X_0\\ \delta Z_0\\ \delta V_{X_0}\\ \delta V_{Z_0}\end{bmatrix} \tag{2.6.56}$$

and similarly for t_2, t_3, etc.

If four such measurements are made, Equations (2.6.56) may be written

$$\begin{bmatrix}\delta\theta(t_1)\\ \delta\theta(t_2)\\ \delta\theta(t_3)\\ \delta\theta(t_4)\end{bmatrix} = \begin{bmatrix} \frac{-1}{Z(t_1)\sec^2\theta(t_1)} & \frac{X(t_1)-D}{Z^2(t_1)\sec^2\theta(t_1)} & \frac{-t_1}{Z(t_1)\sec^2\theta(t_1)} & \frac{t_1[X(t_1)-D]}{Z^2(t_1)\sec^2\theta(t_1)} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \frac{-1}{Z(t_4)\sec^2\theta(t_4)} & \frac{X(t_4)-D}{Z^2(t_4)\sec^2\theta(t_4)} & \frac{-t_4}{Z(t_4)\sec^2\theta(t_4)} & \frac{t_4[X(t_4)-D]}{Z^2(t_4)\sec^2\theta(t_4)} \end{bmatrix} \begin{bmatrix}\delta X_0\\ \delta Z_0\\ \delta V_{X_0}\\ \delta V_{Z_0}\end{bmatrix} \tag{2.6.57}$$

Since the measurements yield the angular values of $\theta(t_1)$, $\theta(t_2)$, $\theta(t_3)$, $\theta(t_4)$, and since the nominal values $\theta_{nom}(t_1)$, $\theta_{nom}(t_2)$, $\theta_{nom}(t_3)$, $\theta_{nom}(t_4)$ are known for the reference trajectory, then the values $\delta\theta(t_1)$, $\delta\theta(t_2)$, $\delta\theta(t_3)$, $\delta\theta(t_4)$, may be computed from:

$$\delta\theta(t_i) = \theta(t_i) - \theta(t_i)_{NOM}$$

and since values of $x(t_1)$, $z(t_1)$, $x(t_2)$, $z(t_2)$, etc., along the reference trajectory may be looked up or determined by simply reading in the values of t_1 t_2. . . , every term except δx_o , δz_o , δV_{x_o} , δV_{z_o} in Equation (2.6.57) is known. Hence, the inverse of the 4 x 4 matrix may be obtained to yield the solutions for these four unknown quantities.

These equations are best written in shorthand form following the previously defined state vector approach, i.e., using the definitions of B(t) and ϕ (t, t_o), in Equations (2.6.53) and (2.6.54), Equation (2.6.56) becomes:

$$\delta\theta(t_1) = B(t_1)\ \phi(t_1, t_o)\ \delta\underline{x}(t_o) \qquad (2.6.58)$$

where

$$\delta\underline{x}(t_o) \triangleq \begin{bmatrix} \delta x_o \\ \delta z_o \\ \delta V_{x_o} \\ \delta V_{z_o} \end{bmatrix}$$

Equation (2.6.57) can thus be rewritten

$$\begin{bmatrix} \delta\theta(t_1) \\ \delta\theta(t_2) \\ \delta\theta(t_3) \\ \delta\theta(t_4) \end{bmatrix} = D\ \delta\underline{x}(t_o) \qquad (2.6.59)$$

where

$$D = \begin{bmatrix} B(t_1)\ \phi(t_1, t_o) \\ B(t_2)\ \phi(t_2, t_o) \\ B(t_3)\ \phi(t_3, t_o) \\ B(t_4)\ \phi(t_4, t_o) \end{bmatrix}$$

Hence

$$\delta \underline{X}(t_0) = \begin{bmatrix} \delta x_0 \\ \delta z_0 \\ \delta V_{x_0} \\ \delta V_{z_0} \end{bmatrix} = D^{-1} \begin{bmatrix} \delta\theta(t_1) \\ \delta\theta(t_2) \\ \delta\theta(t_3) \\ \delta\theta(t_4) \end{bmatrix} \qquad (2.6.60)$$

Thus, the numerical values of δx_0 , δz_0 , δV_{x_0} , δV_{z_0} are determined for any four measurements $\delta\theta(t_1)$, $\delta\theta(t_2)$, $\delta\theta(t_3)$, $\delta\theta(t_4)$ which are made at different times. The actual position of the spacecraft is then computed using these values of δx_0, δz_0, δV_{x_0}, δV_{z_0} from Equation (2.6.48)

$$\underline{R}(t) = \underline{R}_{nom} + \delta\underline{R}(t)$$

or

$$\underline{R}(t) = \begin{bmatrix} x(t) \\ z(t) \end{bmatrix} = \begin{bmatrix} x_{nom}(t) \\ z_{nom}(t) \end{bmatrix} + \begin{bmatrix} \delta x(t) \\ \delta z(t) \end{bmatrix}$$

Hence, using Equation (2.6.50)

$$x(t) = x_{nom}(t) + \delta x_0 + t\,\delta V_x$$

$$z(t) = z_{nom}(t) + \delta z_0 + t\,\delta V_z \qquad (2.6.61)$$

Since Equation (2.6.61) gives the position at any time t, this simple navigation problem is solved.

2.6.2.2 Perturbation Theory and the State Transition Matrix

The usual meaning of "state transition matrix" ϕ is a 6 x 6 time-varying matrix which relates the position and velocity deviations of a vehicle from a reference trajectory at time t_1 to the position and velocity deviations of the vehicle at time t_2. Expressing the position deviations and velocity deviations from a reference trajectory at time t as $\delta \underline{R}(t)$, $\delta \underline{V}(t)$, respectively, the general expression relating deviations at time t_2 to those of time t_1 is:

$$\begin{bmatrix} \delta \underline{R}(t_2) \\ \delta \underline{V}(t_2) \end{bmatrix} = \phi(t_2, t_1) \begin{bmatrix} \delta \underline{R}(t_1) \\ \delta \underline{V}(t_1) \end{bmatrix} \qquad (2.6.62)$$

Equation (2.6.62) is the general solution of a time-varying linear differential equation,

$$\frac{d}{dt}\begin{bmatrix} \delta \underline{R}(t) \\ \delta \underline{V}(t) \end{bmatrix} = \begin{bmatrix} 0 & I \\ \frac{\partial G}{\partial \underline{R}} & 0 \end{bmatrix} \begin{bmatrix} \delta \underline{R}(t) \\ \delta \underline{V}(t) \end{bmatrix} \qquad (2.6.63)$$

The purpose of this section is to outline the derivation of Equation (2.6.63), and to indicate the assumptions which must be satisfied for Equation (2.6.63) and, hence, (2.6.62) to be valid. **Equation(2.6.62) will be proven in a subsequent monograph.**

Consider the general motion of a vehicle under the influence of an externally applied force (excluding gravity) $\underline{F}$ in a central gravitational force field. If it is assumed that the applied external force $\underline{F}(t)$, the mass of the vehicle m (t), and the initial position and velocity are known for all time t, then the position and velocity of the vehicle in space is known for all time t. Call this known trajectory the reference trajectory and denote its position and velocity $\underline{R}_o(t)$, $\underline{V}_o(t)$, and its initial position and velocity by $\underline{R}_o(o)$, $\underline{V}_o(o)$.

Then, from Newton's second law

$$\ddot{\underline{R}}_o(t) = \frac{\underline{F}(t)}{m(t)} - \frac{\mu}{R_o^3(t)} \underline{R}_o(t) \qquad (2.6.64)$$

where μ = GM, M = mass of attracting body,
G = universal gravitational constant

Consider now the motion of a vehicle which is acted upon by the same force $\underline{F}(t)$, and has the same mass variation m (t), but differs from $\underline{R}_o(t)$, $\underline{V}_o(t)$ in that the initial conditions are not those of the reference trajectory, although they are considered sufficiently close that linear approximations will be assumed valid. The differential equation for the vehicle on the new trajectory, in which the position and velocity are represented as $\underline{R}(t)$ and $\underline{V}(t)$, is:

$$\ddot{\underline{R}}(t) = \frac{\underline{F}(t)}{m(t)} - \frac{\mu}{R^3(t)}\,\underline{R}(t) \tag{2.6.65}$$

Subtracting (2.6.64) from (2.6.65) yields:

$$\ddot{\underline{R}}(t) - \ddot{\underline{R}}_o(t) = \mu\left[\frac{\underline{R}_o(t)}{R_o^3} - \frac{\underline{R}(t)}{R^3}\right] \tag{2.6.66}$$

or, defining:

$$\begin{aligned} \Delta\underline{R} &\triangleq \underline{R} - \underline{R}_o \\ \Delta\ddot{\underline{R}} &\triangleq \frac{d^2}{dt^2}(\Delta\underline{R}) = \ddot{\underline{R}} - \ddot{\underline{R}}_o \end{aligned} \tag{2.6.67}$$

equation (2.6.66) may be written:

$$\underline{0} = \Delta\ddot{\underline{R}} + \mu\left[\frac{\underline{R}}{R^3} - \frac{\underline{R}_o}{R_o^3}\right] \tag{2.6.68}$$

Equation (2.6.68) is seen to be identical with Equation (2.5.78) of Section 2.5.2.1 if $\underline{\Delta A}$ is set equal to zero. If the same approximations are carried out as indicated in Section 2.5.2.1, an equation similar to Equation (2.5.86), Section 2.5.2.1 results, except that $\underline{\Delta A} = \underline{0}$; hence,

$$\underline{0} = \Delta\ddot{\underline{R}} + \frac{\mu}{R_o^3}\left[\underline{\underline{I}} - \frac{3}{R_o^2}\left(\underline{R}_o\underline{R}_o^T\right)\right]\Delta\underline{R} \tag{2.6.69}$$

The term $\frac{\mu}{R_o^3}\left[\underline{\underline{I}} - \frac{3}{R_o^2}\left(\underline{R}_o \underline{R}_o^T\right)\right]$ can be shown by a term by term expansion to be equal to $\frac{\partial \underline{G}}{\partial \underline{R}}$, where $\underline{G} = -\frac{\mu}{R^3}\underline{R}$, and where

$$\frac{\partial \underline{G}}{\partial \underline{R}} \triangleq \begin{bmatrix} \frac{\partial G_x}{\partial R_x} & \frac{\partial G_x}{\partial R_y} & \frac{\partial G_x}{\partial R_z} \\ \frac{\partial G_y}{\partial R_x} & \frac{\partial G_y}{\partial R_y} & \frac{\partial G_y}{\partial R_z} \\ \frac{\partial G_z}{\partial R_x} & \frac{\partial G_z}{\partial R_y} & \frac{\partial G_z}{\partial R_z} \end{bmatrix}$$

Hence, Equation (2.6.69) may be written:

$$\Delta \ddot{\underline{R}} = -\frac{\partial \underline{G}}{\partial \underline{R}} \Delta \underline{R} \qquad (2.6.70)$$

Letting $\Delta \dot{\underline{R}} \triangleq \frac{d}{dt}(\Delta \underline{R}) \triangleq \Delta \underline{V}$, and $\Delta \dot{\underline{V}} = \frac{d}{dt}(\Delta \underline{V})$, (2.6.71)

Equations (2.6.70) and (2.6.71) can be written:

$$\Delta \dot{\underline{R}} = \Delta \underline{V}$$

$$\Delta \dot{\underline{V}} = -\frac{\partial \underline{G}}{\partial \underline{R}} \Delta \underline{R}$$

or

$$\frac{d}{dt}\begin{bmatrix} \Delta \underline{R}(t) \\ \Delta \underline{V}(t) \end{bmatrix} = \begin{bmatrix} 0 & I \\ -\frac{\partial \underline{G}}{\partial \underline{R}} & 0 \end{bmatrix} \begin{bmatrix} \Delta \underline{R}(t) \\ \Delta \underline{V}(t) \end{bmatrix} \qquad (2.6.72)$$

Q.E.D.

The legitimate use of the state transition matrix is thus dependent upon motion being close to a reference trajectory, and must be such that the same acceleration $\underline{F}(t)/m(t)$ (excluding gravitational acceleration) must act upon the vehicle on its perturbed trajectory as on its reference trajectory.

2.6.2.3 General Solution to the Navigational Problem with Sequential Measurement Data

Let a set of six measurements, taken at varying times, be given as:

$$\delta q_i(t_j) = H_i(t_j)\ \delta \underline{R}(t_j) \tag{2.6.73}$$

where $H_i(t_j) = \left[H_{i1}(t_j)\quad H_{i2}(t_j)\quad H_{i3}(t_j)\right]$

Since motion near a reference trajectory is assumed, it is valid to write:

$$\begin{bmatrix}\delta \underline{R}(t_j)\\ \delta \underline{V}(t_j)\end{bmatrix} = \phi(t_j, t_o) \begin{bmatrix}\delta \underline{R}(t_o)\\ \delta \underline{V}(t_o)\end{bmatrix} \tag{2.6.74}$$

Let a matrix B be defined as:

$$B_i(t_j) = \left[H_{i1}(t_j)\quad H_{i2}(t_j)\quad H_{i3}(t_j)\quad 0\ 0\ 0\right] \tag{2.6.75}$$

Then Equation (2.6.73) may be written:

$$\delta q_i(t_j) = B_i(t_j) \begin{bmatrix}\delta \underline{R}(t_j)\\ \delta \underline{V}(t_j)\end{bmatrix} \tag{2.6.76}$$

Substituting (2.6.74) into (2.6.76) yields:

$$\delta q_i(t_j) = B_i(t_j)\ \phi(t_j, t_o) \begin{bmatrix}\delta \underline{R}_o\\ \delta \underline{V}_o\end{bmatrix} \tag{2.6.77}$$

Define the matrix $D_i(t_j) = \left[D_{i1}(t_j)\quad D_{i2}(t_j) \cdots\cdots D_{i6}(t_j)\right]$ as:

$$D_i(t_j) = B_i(t_j)\ \phi(t_j, t_o) \tag{2.6.78}$$

Then Equation (2.6.77) may be written for any six i's and j's, $i = 1, 2, 3, 4, 5, 6$ at times 1,2,3,4,5,6

$$\begin{bmatrix} \delta q_1(t_1) \\ \delta q_2(t_2) \\ \delta q_3(t_3) \\ \delta q_4(t_4) \\ \delta q_5(t_5) \\ \delta q_6(t_6) \end{bmatrix} = \begin{bmatrix} D_{11}(t_1) & \cdots & D_{16}(t_1) \\ D_{21}(t_2) & & \vdots \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ D_{61}(t_6) & \cdots & D_{66}(t_6) \end{bmatrix} \begin{bmatrix} \delta R_x(t_o) \\ \delta R_y(t_o) \\ \delta R_z(t_o) \\ \delta V_x(t_o) \\ \delta V_y(t_o) \\ \delta V_z(t_o) \end{bmatrix} \tag{2.6.79}$$

If all measurements $\delta q_i(t_i)$ are independent and the matrix $\begin{bmatrix} D_{11} & & D_{66} \end{bmatrix}$ can be inverted, the solution for the position and velocity deviations at time t_o can be found, i.e.,

$$\begin{bmatrix} \delta \underline{R}(t_o) \\ \delta \underline{V}(t_o) \end{bmatrix} = \begin{bmatrix} D_{11} & \cdots & D_{16} \\ & \ddots & \\ D_{61} & & D_{66} \end{bmatrix}^{-1} \begin{bmatrix} \delta q_1(t_1) \\ \vdots \\ \delta q_6(t_6) \end{bmatrix} \tag{2.6.80}$$

Knowing the initial deviations at time t_o, $\begin{bmatrix} \delta \underline{R}(t_o) \\ \delta \underline{V}(t_o) \end{bmatrix}$ is equivalent to knowing the position and velocity at any time t, as indicated from Equation (2.6.74),

$$\begin{bmatrix} \delta \underline{R}(t) \\ \delta \underline{V}(t) \end{bmatrix} = \phi(t, t_o) \begin{bmatrix} \delta \underline{R}(t_o) \\ \delta \underline{V}(t_o) \end{bmatrix}$$

since it is assumed that $\phi(t, t_o)$ is known or computable.

2.7 SENSOR REQUIREMENTS IMPOSED UPON A VEHICLE

Although it is undesirable to constrain vehicle performance and design because of navigation sensor limitations, the imposition of constraints upon a vehicle is often mandatory to guarantee successful deployment of the navigation sensors.

Unfortunately, the technical literature dealing with this subject in general terms is sparse, possibly because the constraints and limitations themselves are such individual functions of the vehicle and sensor of interest.

Even though a comprehensive listing of all of the sensor imposed vehicle constraints is impractical here, it is possible to indicate a general framework for classifying, analyzing, and evaluating the constraints in any given situation.

Three sources of navigation sensor limitations which may impose constraints upon a vehicle are presented in the following sections. Examples are included to illustrate some of the more usual sensor imposed vehicle constraints.

2.7.1 Sensor Input and Output Static and Dynamic Limitations

Frequently a navigation sensor may become damaged or destroyed if its input or output quantity varies beyond the sensor design limits. In addition, even in situations in which no damage is done to the instruments, when such instruments are over-driven by too vigorous a variation of its input variables other detrimental system effects may be associated with sensor saturation, such as the system stability margin may be decreased to the point where the system can become unstable, or the system can lose accuracy. This situation exists for all three basic types of sensors which have been discussed in previous sections of the monograph. For example, in inertial navigation systems, excessive torques imposed upon any inertial platform produced by torquing the platform gimbals to follow the vehicle motion can cause a loss of attitude reference accuracy, or can cause a complete tumbling of the platform. Additional problems with inertial platforms are associated with the three-gimbal configuration. The three gimbal configuration requires that vehicle attitude must be constrained such that the middle gimbal variations stay within approximately $\pm$ 70° of attitude change. This limitation must be imposed to preclude gimbal lock as discussed in section 2.3.2.1.

Other constraints which must be imposed upon vehicles caused by inertial sensors are those associated with rate gyroscopes and integrating rate gyroscopes in strap-down applications, and the different forms of accelerometers. All of these instruments may be saturated by excessively violent vehicle motion, and in many cases this saturation produces uncorrectable inertial navigation errors.

Other sensor induced vehicle constraints may be associated with radar sensors and star trackers. The necessity of pointing these sensors, which have a narrow beam width, requires vehicle attitude to be controlled within relatively narrow limits to preclude loss of signal.

2.7.2 Environmental Requirements

These limitations must be imposed upon a vehicle to guarantee compatibility of the sensor with its physical environment. The physical environment referred to here includes the temperature, pressure, vibration characteristics, supply voltages, input and output impedance matching requirements, etc., associated with the desired location of the sensor within the vehicle. Details describing the specific factors discussed here are strictly dependent upon the hardware characteristics of both the vehicle and the sensor; and, hence, are usually not considered in detail until the design of both are reasonably well defined in a development program.

2.7.3 Operational Limitations

These limitations reflect requirements of other hardware and functional elements which operate within the navigation sensor loop. These other functional elements may include ancillary equipments as well as functions associated with man-in-the-loop. A notable example of vehicle constraints being imposed through operational requirements is associated with the manual operation of space sextants and telescopes. For example, consider an astronaut in a spacecraft in an earth orbital environment or cis-lunar environment. Accurate measurements from the manually controlled sextant require that maximum vehicle angular rate be stringently controlled to low values; in the case of the Apollo vehicle, the vehicle rate must be constrained to 1.2 arc-minutes per second. Other examples include constraints due to landmark tracking using telescopes, etc.

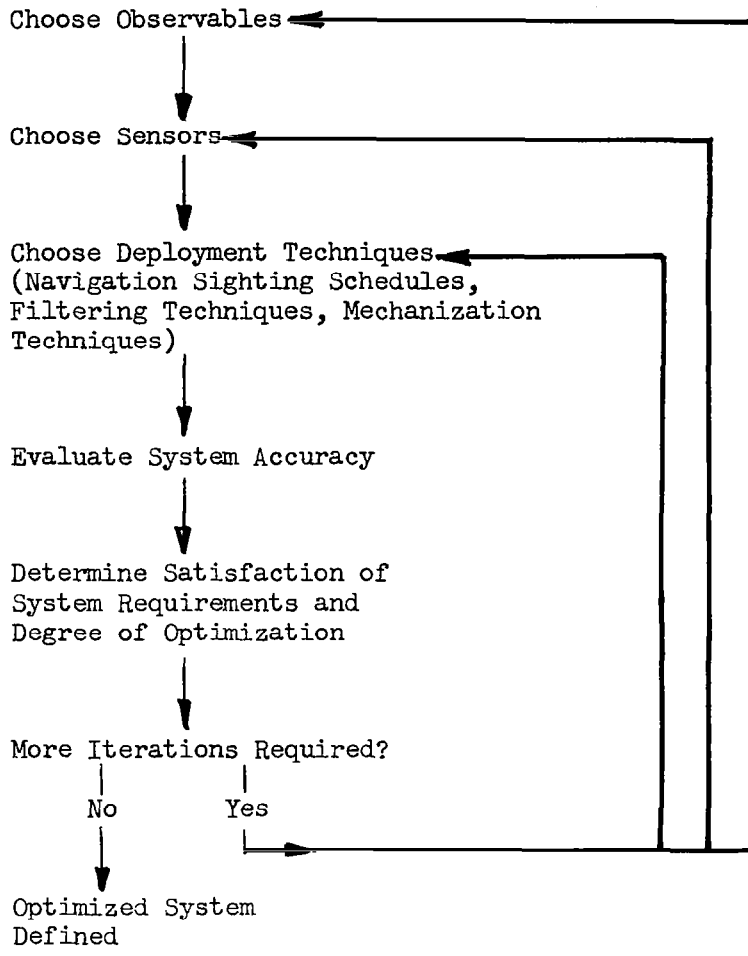
2.8 CRITERIA FOR SELECTING OBSERVABLES TO BE MEASURED

Discussed in this section will be those factors which influence the selection of navigation observables to be measured. The first part of the discussion presented in 2.8.1 concerns itself with the more realistic techniques which must actually be employed in a realistic situation to select the best navigation observables. Section 2.8.2 presents a discussion of more limited applicability of the special techniques which may be used to optimize the selection of optimum geometrical relationships between celestial objects to minimize state vector estimation errors for a evenly determined data set.

2.8.1 General Criteria

The most realistic criteria which apply to the general problem of selecting observables to be measured is the iterative technique carried out to minimize or maximize the important parameters associated with any specific mission. For example, in selecting observables in an earth orbital mission, the basic mission requirements of, say, landing a vehicle within a small recovery area may require extremely precise position and velocity information on board the vehicle at the time of the retro-firing. In this situation, the navigation observables selected should be such that errors in position and velocity at this time are minimized. In this type of problem a number of variables affect the total accuracy obtainable, namely, such factors as variation in the navigation sighting schedule, instrumentation characteristics, operational environment (rapidity of angular motion of the vehicle), etc. A simple approach to specification of the selection criteria for this case is to specify that the navigation observables selected for implementation should provide a minimization of the position and velocity error at the retro-firing point.

The problem with this definition of a criteria is that it does not go far enough in suggesting the technique of applying the criteria for making the selection itself. A more practical definition of appropriate techniques for selecting observables is outlined in the sketch shown below.



The sketch represents an iterative procedure in which observables are chosen, the sensors are chosen, deployment techniques are chosen, followed by an evaluation of the system accuracy. If the system accuracy is unsatisfactory or non-optimized, variations of the deployment technique, a different choice of navigation observables, or a different choice of navigation sensors may be made. It is only by repeated iterations through these loops that one is assured of having reached a somewhat optimized selection of the navigation observables, navigation sensors, and the deployment technique itself. Hence, the criteria for selecting navigation observables is that set of observables which in some sense optimizes the entire navigation loop. That the optimization of the selection of observables cannot be made without considering the optimization of the entire navigation process is characteristic of the application of the methods of system engineering, in which criteria for excellence of the performance of sub-system elements becomes subordinate to the excellence of the total system performance.

2.8.2 Minimization of Position Estimation Error for Simultaneous Multiple Fix Optimal Navigation Measurements

This section will present a method of determining the desirable orientations of measurements to choose such that the magnitude of the position deviation error is minimized. The result will be applied to several standard multiple fix combinations of measurements that have been previously described.

General Approach:

Consider a measurement deviation, δq_i , of a multiple fix position deviation measurement. Due to the inherent limited accuracy of the measurement instruments, it can be expected that each measurement is in error by an amount α_i . Assuming that the correct value of the measurement deviation was δq_i, the measured quantity can be written as

$$\delta \tilde{q}_i = \delta q_i + \alpha_i \qquad (2.8.1)$$

Since there are several measurements involved in a multiple fix, Equation (2.8.1) may be written in vector form.

That is $$\delta \underline{\tilde{q}} = \delta \underline{q} + \underline{\alpha} \qquad (2.8.2)$$

The uncertainty in the measurements introduce a corresponding uncertainty in the position deviation vector. This uncertainty can be expressed by a vector $\underline{\epsilon}$. The relationship between the true position deviation vector and the computed one from the measurements can be expressed as

$$\delta \underline{\hat{R}} = \delta \underline{R} + \underline{\epsilon}$$

where $\delta \underline{\hat{R}}$ = the computed deviation

$\delta \underline{R}$ = actual deviation

$\underline{\epsilon}$ = deviation error introduced by measurement inaccuracies

In a previous section, it was shown that the actual deviations are related by

$$\delta \underline{q} = \underline{H}\, \delta \underline{R} \qquad (2.8.3)$$

or

$$\delta \underline{R} = \underline{H}^{-1} \delta \underline{q} \qquad (2.8.4)$$

With an evenly determined data set consisting of three independent measurements, an unbiased estimate for the measured quantities is

$$\delta \underline{\hat{R}} = \underline{H}^{-1} \delta \underline{\tilde{q}} \qquad (2.8.5)$$

Using Equations (2.8.2) and (2.8.3) it is seen that

$$\delta \underline{R} + \underline{\epsilon} = \underline{H}^{-1} (\delta \underline{q} + \underline{\alpha}) = \underline{H}^{-1} \delta \underline{q} + \underline{H}^{-1} \underline{\alpha} \qquad (2.8.6)$$

but

$$\delta \underline{R} = \underline{H}^{-1} \delta \underline{q} \tag{2.8.7}$$

so the error vectors are related by

$$\underline{\epsilon} = \underline{H}^{-1} \underline{\alpha} \tag{2.8.8}$$

This equation graphically displays the importance of the quality of the data in the navigation problem.

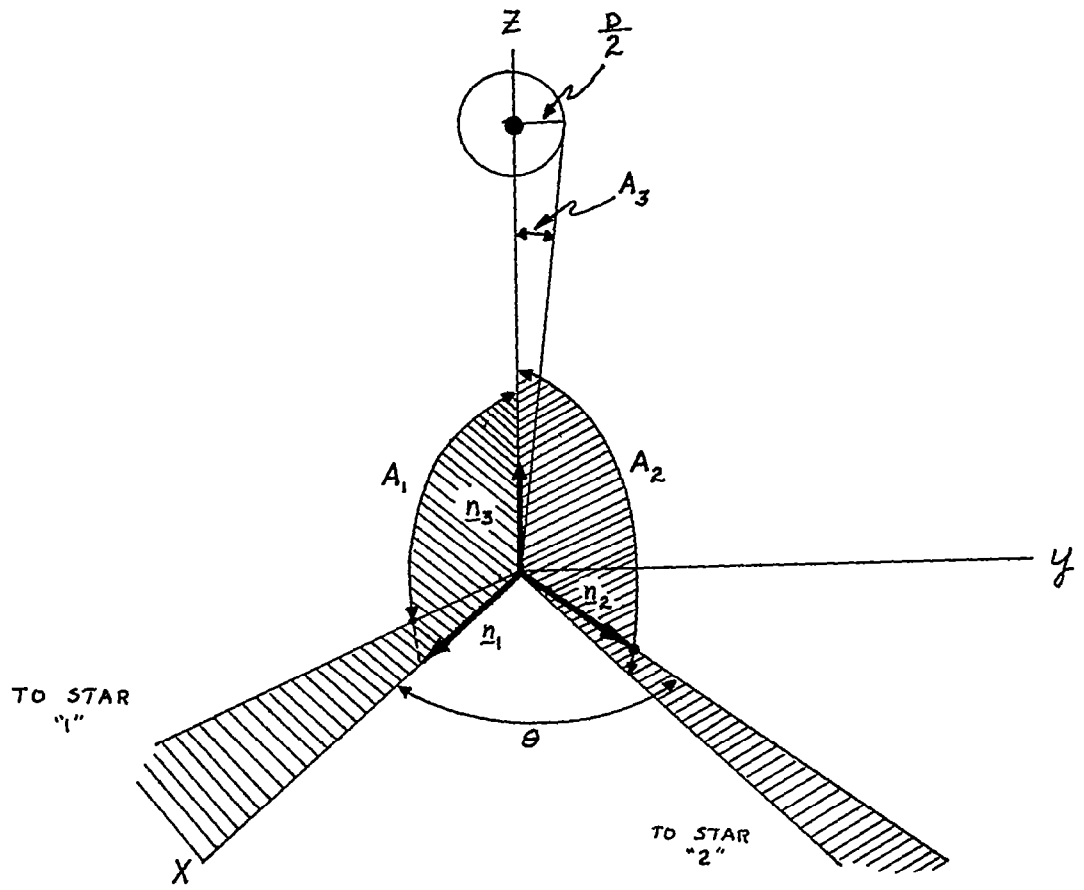
In order to determine the magnitude of $\bar{\epsilon}$, the square root of $|\bar{\epsilon}|^2$ can be found, where

$$|\bar{\epsilon}|^2 = \underline{\epsilon}^T \underline{\epsilon} = \underline{\alpha}^T (\underline{H}^{-1})^T \underline{H}^{-1} \underline{\alpha} \tag{2.8.9}$$

The "T" in the above expressions represents the transpose of a matrix, while a (-1) superscript represents the inverse.

2.8.2.1 Planet-Star, Planet Star, Planet Diameter Measurement Optimization

The method of finding the optimum positions of stars and planets for multiple fix measurements will now be applied to the planet-star measurement being used with the planet diameter measurement. The geometry is shown in the following sketch.



Using equation (2.8.9) and (2.6.5) in order to find $|\underline{\epsilon}|^2$ yields

$$|\underline{\epsilon}|^2 = \underline{\alpha}^T (\underline{H}^{-1})^T \underline{H}^{-1} \underline{\alpha} \tag{2.8.10}$$

or

$$|\underline{\epsilon}|^2 = \underline{\alpha}^T \begin{bmatrix} z^2\cos^2\theta & -z^2\cot\theta\cos\theta & 0 \\ -z^2\cos\theta\cot\theta & z^2\cos^2\theta & 0 \\ 0 & 0 & \left[\frac{z^2\cos(\frac{A}{2})}{D}\right]^2 \end{bmatrix} \underline{\alpha}$$

(2.8.11)

This matrix may be diagonalized to display the eigenvalues on the principal diagonal by employing a similarity transformation. However, since the matrix is symmetric, this transformation is equivalent to rotating the rectangular cartesian coordinate system 45° about the Z axis. The eigenvalues of the above matrix are:

$$\begin{aligned} \lambda_1 &= z^2\cos\theta\,[\cos\theta + \cot\theta] \\ \lambda_2 &= z^2\cos\theta\,[\cos\theta - \cot\theta] \\ \lambda_3 &= \left[z^2\,\frac{\cos(\frac{A}{2})}{D}\right]^2 \end{aligned} \tag{2.8.12}$$

Now equation (2.8.11) becomes

$$|\underline{\epsilon}|^2 = [\alpha_1', \alpha_2', \alpha_3'] \begin{bmatrix} z^2\cos\theta\,(\cos\theta + \cot\theta) & 0 & 0 \\ 0 & z^2\cos\theta\,(\cos\theta - \cot\theta) & 0 \\ 0 & 0 & \left[\frac{z^2\cos(\frac{A}{2})}{D}\right]^2 \end{bmatrix} \begin{bmatrix} \alpha_1' \\ \alpha_2' \\ \alpha_3' \end{bmatrix}$$

(2.8.13)

where

$$[\alpha_1', \alpha_2', \alpha_3'] = \underline{\alpha}^T$$

and

$$\begin{bmatrix} \alpha_1', \\ \alpha_2', \\ \alpha_3' \end{bmatrix} = \underline{\alpha}$$

The components used to express $\underline{\alpha}$ are the values formed by the new coordinate system that resulted when the matrix of equation (2.8.11) was diagonalized.

When equation (2.8.13) is multiplied out, the result is

$$|\underline{\epsilon}|^2 = (\alpha_1')^2 z^2 \cos\theta\,(\cos\theta + \cot\theta) + (\alpha_2')^2 z^2 \cos\theta(\cos\theta - \cot\theta) + (\alpha_3')^2 \left[z^2 \frac{\cos(\frac{A}{2})}{D}\right]^2 \quad (2.8.14)$$

It is seen that the above expression attains its smallest value when $\theta = 90^\circ$.

Now,

$$|\underline{\epsilon}|^2 = (\alpha_1')^2 + (\alpha_2')^2 + (\alpha_3')^2 \left[z^2 \frac{\cos(\frac{A}{2})}{D}\right]^2 \quad (2.8.15)$$

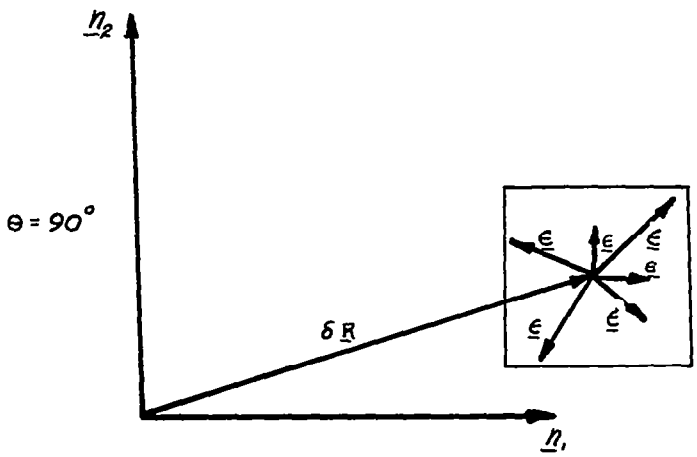
Equation (2.8.15) can be expressed in more meaningful terms by the following relation

$$\left[\frac{z^2\cos(\frac{A}{2})}{D}\right]^2 = \frac{z^4}{D^2}\left(\frac{z^2 - \frac{D}{4}}{z^2}\right) = z^2\left(\frac{z^2}{D^2} - \frac{1}{4}\right) \quad (2.8.16)$$

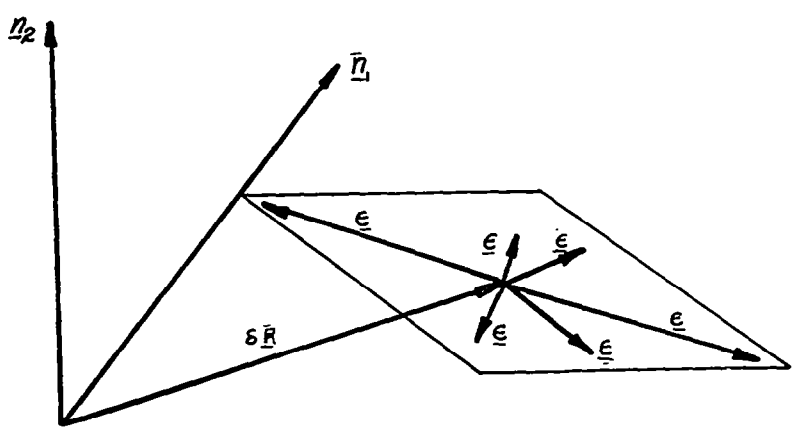
So

$$|\underline{\epsilon}|^2 = (\alpha_1')^2 + (\alpha_2')^2 + (\alpha_3')^2\, z^2 \left(\frac{z^2}{D^2} - \frac{1}{4}\right) \quad (2.8.17)$$

One would expect the same results on an intuitive basis. It seems reasonable to expect that the error is minimized if the stars are taken to be in orthogonal planes which pass through the vehicle and planet center. This behavior is due to the fact that each star-planet measurement determines the δr component in its corresponding Ω vector direction; hence, the uncertainty of δr due to each measurement is in the direction of Ω . If these planes containing the stars are taken to be orthogonal, the area of uncertainty in which $\underline{\epsilon}$ can lie in a square is shown below.



If, however, some other orientation of the star choice is selected, the area of uncertainty changes to a parallelogram; and the magnitude of the error vector, $\underline{\epsilon}$, with the same accuracy in the instruments could be much larger.



It is also apparent from equation (2.8.17) that the magnitude of the error in the position vector increases as the vehicle gets farther away from the planet. **This result is also expected.**

2.8.2.2 Planet - Star, Planet - Star, Sun - Star Measurement Optimization

The optimum positions of a planet, the sun, and stars will now be derived for the planet-star, planet-star, sun-star measurement set.

The expression for $|\dot{\underline{\epsilon}}|^2$ may now be found to be

$$|\dot{\underline{\epsilon}}|^2 = \begin{bmatrix} \alpha_1, \alpha_2, \alpha_3 \end{bmatrix} \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} \tag{2.8.18}$$

where

$$h_{11} = z^2 \cos^2\theta + z^2 \cot^2\gamma \cos^2\theta \sin^2(\beta-\theta)$$

$$h_{12} = -z^2 \cot\theta \cos\theta - z^2 \cot^2\gamma \cos^2\theta \sin\beta \sin(\beta-\theta)$$

$$h_{13} = R z \cot\gamma \cos\gamma \cos\theta \sin(\beta-\theta)$$

$$h_{21} = h_{12}$$

$$h_{22} = z^2 \cos^2\theta + z^2 \cot^2\gamma \sin^2\beta \cos^2\theta$$

$$h_{23} = -z R \cos\gamma \cot\gamma \sin\beta \cos\theta$$

$$h_{31} = h_{13}$$

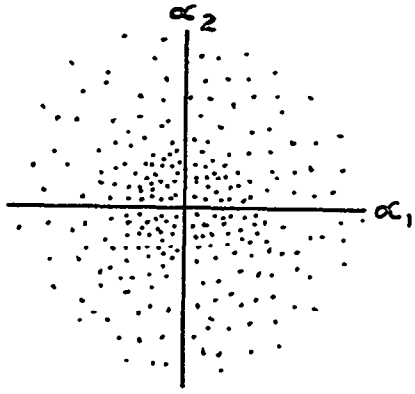
$$h_{32} = h_{23}$$

$$h_{33} = R^2 \cos^2\gamma$$

At this point in the planet-star, planet-star, planet diameter analysis, the above matrix was diagonalized. It is apparent that this approach is not very practical in this case, since the algebra involved is prohibitive. The optimum selection of the various angles can still be determined, however, by using the average value of the measurement error variables. Assuming that the errors are unbiased, it is reasonable to conclude that the average error in any single measurement is zero. Also, if the measurements are independent of each other, the average of the product of two errors will be zero, i.e.,

$$\begin{aligned} [\alpha_1 \quad \alpha_2]_{av} &= 0 \\ [\alpha_1 \quad \alpha_3]_{av} &= 0 \\ [\alpha_2 \quad \alpha_3]_{av} &= 0 \end{aligned} \tag{2.8.19}$$

This can readily be seen by examining the following sketch which illustrates the plot of two independent, unbiased errors.



If a large number of measurements are taken, there will be very nearly an equal number of points in each quadrant. When $[\alpha_1 \alpha_2]_{av}$ is evaluated, those points in quadrant I will very nearly cancel those in quadrant IV. It should also be apparent that the average value of the square of each measurement will, in general, not be zero.

Using average values for all α's, equation (2.8.18) becomes

$$|\underline{\epsilon}|^2 = [\alpha_1^2]_{av} h_{11} + [\alpha_2^2]_{av} h_{22} + [\alpha_3^2]_{av} h_{33} \qquad (2.8.20)$$

All other terms reduce to zero due to equations (2.8.19).

Substituting the values of h_{11}, h_{22}, and h_{33}, equation (2.8.20) becomes

$$\begin{aligned}\left[|\underline{\epsilon}|^2\right]_{av} = {} & z^2 \cos^2\theta \left[1 + \cot^2\gamma \sin^2(\beta-\theta)\right][\alpha_1^2]_{av} \\ & + z^2 \cos^2\theta \left[1 + \cot^2\gamma \sin^2\beta\right][\alpha_2^2]_{av} \\ & - R^2 \cos^2\gamma \, [\alpha_3^2]_{av}\end{aligned} \qquad (2.8.21)$$

This is the expression that must be used to determine the optimum orientations of the angles that are measured, i.e., A_1, A_2, and A_3. In order to find the optimum choice for β, the partial derivative of equation (2.8.21) is taken.

$$\begin{aligned}\frac{\partial}{\partial\beta}\left[|\underline{\epsilon}|^2\right]_{av} = {} & z^2 \cos^2\theta \cot^2\gamma \sin 2(\beta-\theta)[\alpha_1^2]_{av} \\ & + z^2 \cos^2\theta \cot^2\gamma \sin 2\beta [\alpha_2^2]_{av} = 0\end{aligned} \qquad (2.8.22)$$

Since two of the angle measurements are of the same type, it may be assumed that

$$\left[\alpha_1^2\right]_{av} = \left[\alpha_2^2\right]_{av}$$

Now (2.8.23)

$$\left[\alpha_1^2\right]_{av} \; z^2 \cos^2\theta \cot\gamma \left[\sin 2(\beta-\theta) + \sin 2\beta\right] = 0$$

or

$$\sin 2(\beta-\theta) = -\sin 2\beta = \sin(-2\beta) \qquad (2.8.24)$$

A solution to this equation is

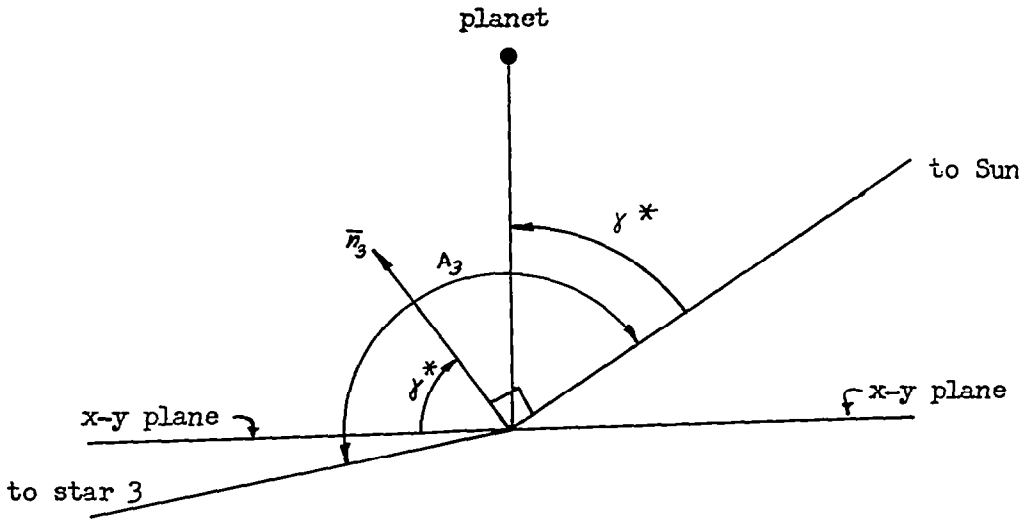
$$\beta = \frac{\theta}{2} \qquad (2.8.25)$$

This result is the first specification on the orientation of any of the $\underline{\eta}$'s. Equation (2.8.25) says that the projection of $\underline{\eta}_3$ on the x-y plane must bisect that angle formed by $\underline{\eta}_1$, and $\underline{\eta}_3$ in order to minimize the magnitude of the error, $|\underline{\epsilon}|$.

Having established β, the expression for $\left[|\underline{\epsilon}|^2\right]_{av}$ now remains a function of θ and γ. Equation (2.8.25) now reduces this expression to

$$\left[|\underline{\epsilon}|^2\right] = z^2 \cos^2\theta \; \alpha_1^2{}_{av} \; 2 + (\cot^2\gamma)(1-\cos\theta) + R^2 \cos\gamma \; \alpha_3^2{}_{av} \qquad (2.8.26)$$

Now, it must be determined how to minimize the above expression when γ is varied. Remembering that β is already specified as being $\frac{\theta}{2}$, the plane of angle measurement A_3 is now rotated about the line from the spacecraft to the sun. This rotation changes the angle γ and the location of the star that is to be chosen as "star 3." The only appearances of γ in equation (2.8.26) are as $\cot^2\gamma$ and $\cos\gamma$, both of which are minimized at $\gamma = 90^\circ$. Clearly the largest value that γ can achieve from the above rotation is that value which occurs when the plane of the sun-star measurement passes through the z axis, i.e., "star 3" is located such that it is in the plane determined by the spacecraft, planet and sun. This plane is shown below.



It is seen that the value of γ is equal to the angle between the planet and the sun, γ^*, if "star 3" is chosen as specified.

The only angle that has yet to be specified is θ. This is done by taking the partial derivative of $\left[|\underline{\epsilon}|^2\right]_{av}$ with respect to θ

$$\frac{\partial}{\partial\theta}\left[|\underline{\epsilon}|^2\right]_{av} = z^2\left[\alpha_1^2\right]_{av}\left\{\cos^2\theta\cot\gamma^*\sin\theta+\left[2+\cot^2\gamma^*(1-\cos\theta)\right]\left[-2\cos^2\theta\cot\theta\right)\right\} \tag{2.8.27}$$

Equating this expression to zero, the optimum choice for θ is found.

$$\cos\theta_o = \frac{1+\sin\gamma^*}{1-\sin\gamma^*} \tag{2.8.28}$$

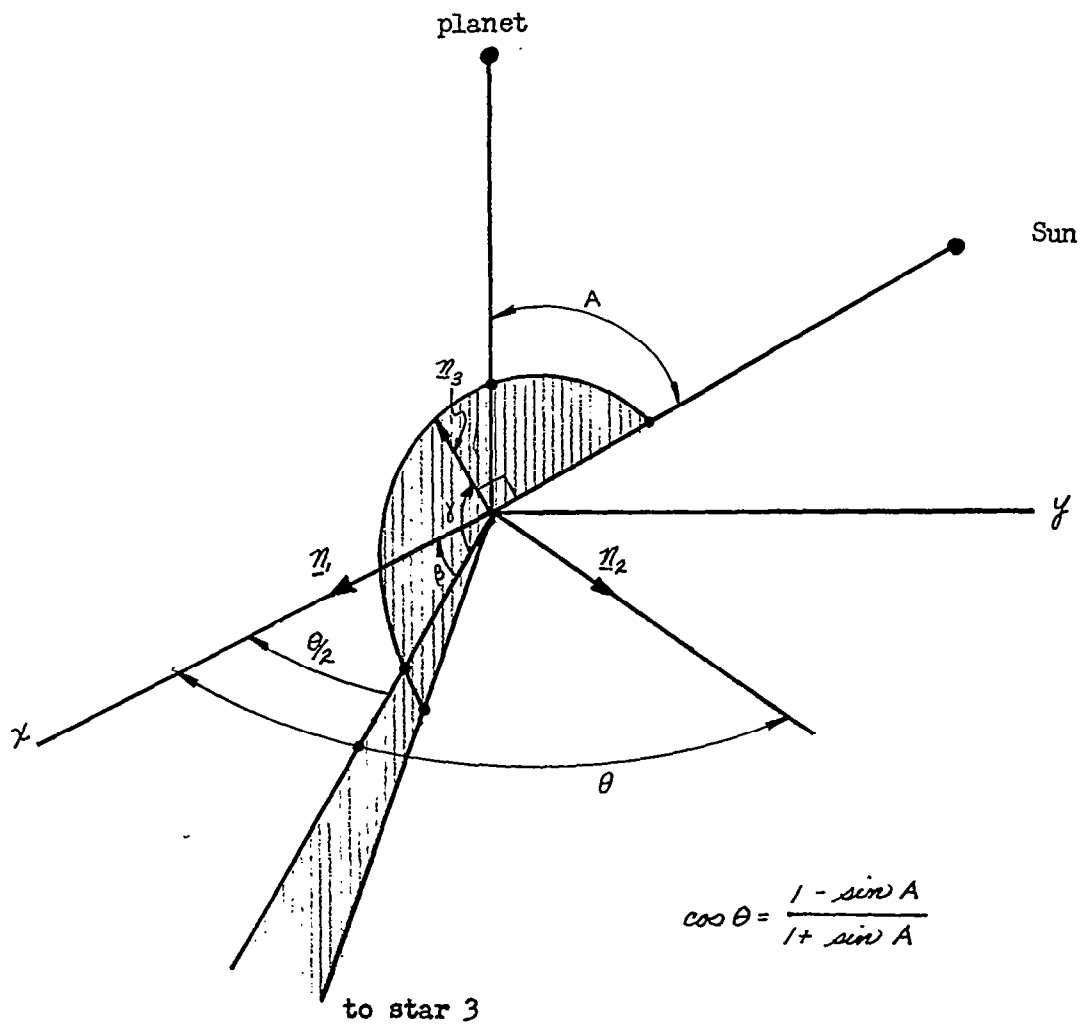
The final expression for the minimum error magnitude is now determined to be

$$\left[|\underline{\epsilon}|^2\right]_{av\,MIN} = \cos^2\gamma^*\left[\frac{z}{2}\right.(1+\sin\gamma^*)^2\left[\alpha_1^2\right]_{av}+r^2\left.\left[\alpha_3^2\right]_{av}\right] \tag{2.8.29}$$

A brief review of the results at this point would be beneficial. In order to choose the best star positions for a star-planet, star-planet, sun-star measurement, one would first select a star in the plane of the spacecraft, the planet and the sun. The angle, A_3, between the star and the sun is now determined and any deviation from the nominal vehicle position in the direction of $\bar{\eta}_3$ (the direction perpendicular to the sun) can be determined by an angular deviation, δA_3. Next, the selection of the other two stars requires that the projections of their lines of sight onto the plane perpendicular to z axis (the direction to the planet) form some angle θ where

$$\theta = \cos^{-1}\left[\frac{1+\sin\gamma^*}{1-\sin\gamma^*}\right] \tag{2.8.30}$$

Furthermore, these projections of the two stars should be symmetric about the plane of the A_3 measurement. With the selection of the last two stars, the two planet-star measurements can be made. Any deviations from the nominal values of A_1 and A_2 determine the position deviation in the η_1 and η_2 directions. If all of the above conditions are met, the minimum average error of the magnitude of the position deviation vector cannot be greater than the value of equation (2.8.29). For convenience, a sketch of the optimum star locations for a planet-star, planet-star, sun-star measurement is shown below.



$$\cos\theta = \frac{1 - \sin A}{1 + \sin A}$$

2.8.2.3 Planet-Star, Planet-Star, Sun-Planet Measurement Optimization

The optimum positions of a planet, the sun, and stars will now be derived for the planet-star, planet-star, sun-planet measurement.

The average of the error magnitude squared is

$$\left[|\underline{\epsilon}|^2\right]_{av} = \left[\alpha_1^2\right]_{av} h_{11} + \left[\alpha_2^2\right]_{av} h_{22} + \left[\alpha_3^2\right]_{av} h_{33} \qquad (2.8.31)$$

To find the optimum β, the partial derivative is again taken with respect to β,

$$\frac{\partial}{\partial \beta}\left[|\underline{\epsilon}|^2\right]_{av} = 0 \qquad (2.8.32)$$

The solution to equation (2.8.32) is $\beta = \frac{\theta}{2}$. It should be remembered that this was the optimum β for the previous case also. This is due to the fact that the A_3 measurement, although it is a slightly different type of measurement, provides information in the same plane in both cases. In this case, the components of $\delta \underline{R}$ are expressed in terms of two vectors instead of one vector.

When $\beta = \frac{\theta}{2}$ is substituted into the expression for $\left[|\underline{\epsilon}|^2\right]_{av}$, it becomes

$$\left[|\underline{\epsilon}|^2\right]_{av} = \left[2Z^2\cos^2\theta + \left(\frac{Z\cot\gamma^* - R\cos\gamma^*}{1+\cos\theta}\right)^2\right]\left[\alpha_1^2\right]_{av} + R^2\cos^2\gamma^*\left[\alpha_3^2\right]_{av} \qquad (2.8.33)$$

Similarly, the optimum θ is found as follows:

$$\frac{\partial}{\partial \theta}\,|\underline{\epsilon}|^2_{av} = 0 \qquad (2.8.34)$$

The solution to this equation is

$$\cos\theta = \frac{\sqrt{1 - 2\frac{R}{Z}\cos\gamma^* + \frac{R^2}{Z^2} - \sin\gamma^*}}{\sqrt{1 - 2\frac{R}{Z}\cos\gamma^* + \frac{R^2}{Z^2} + \sin\gamma^*}} \qquad (2.8.35)$$

Now the final minimum expression for $\left[|\underline{\epsilon}|^2\right]_{av}$ is written as

$$\left[\underline{\epsilon}_1^2\right]_{av\,MIN} = \frac{z^2}{2}\cos\gamma^*\left[\sin\gamma^* + \sqrt{1-2\frac{R}{z}\cos\gamma^* + \frac{R^2}{z^2}}\right]^2\left[\alpha_1^2\right]_{av} + R^2\cos^2\gamma^*\left[\alpha_3^2\right]_{av} \qquad (2.8.36)$$

Although only three cases have been analyzed, it is hoped that the reader can grasp the general approach one would use in order to find the minimum average position vector error magnitude for any useful combination of measurements. The analyses should also demonstrate how to select the celestial bodies to be used in a multiple fix in order to reduce the errors that are introduced by the inherent inaccuracies of the instruments that are used for the measurements.

3.0 RECOMMENDED PROCEDURES

This monograph presents two kinds of material:

(1) A body of facts and useful equations which can be applied to the problems related to navigation observations.

(2) A number of analytical techniques which are effective in quantifying the inter-relationships of fundamental physical phenomena, sensing errors, state vector determinations, etc. (These techniques can be applied to a wide class of navigation observation problems to extend the theory and generate new results).

Hence, the recommended procedures fall into two corresponding categories:

(1) The facts and equations presented herein may be utilized to: identify the physical considerations involved in radiation and inertial sensor evaluation and selection, evaluate the approximate linear relationships which hold between the vehicle state vector (position and velocity) for a class of optical sensors, evalute the observation errors which affect the outputs of radiation and inertial sensors, evaluate the vehicle state from an evenly determined data set of navigation measurements.

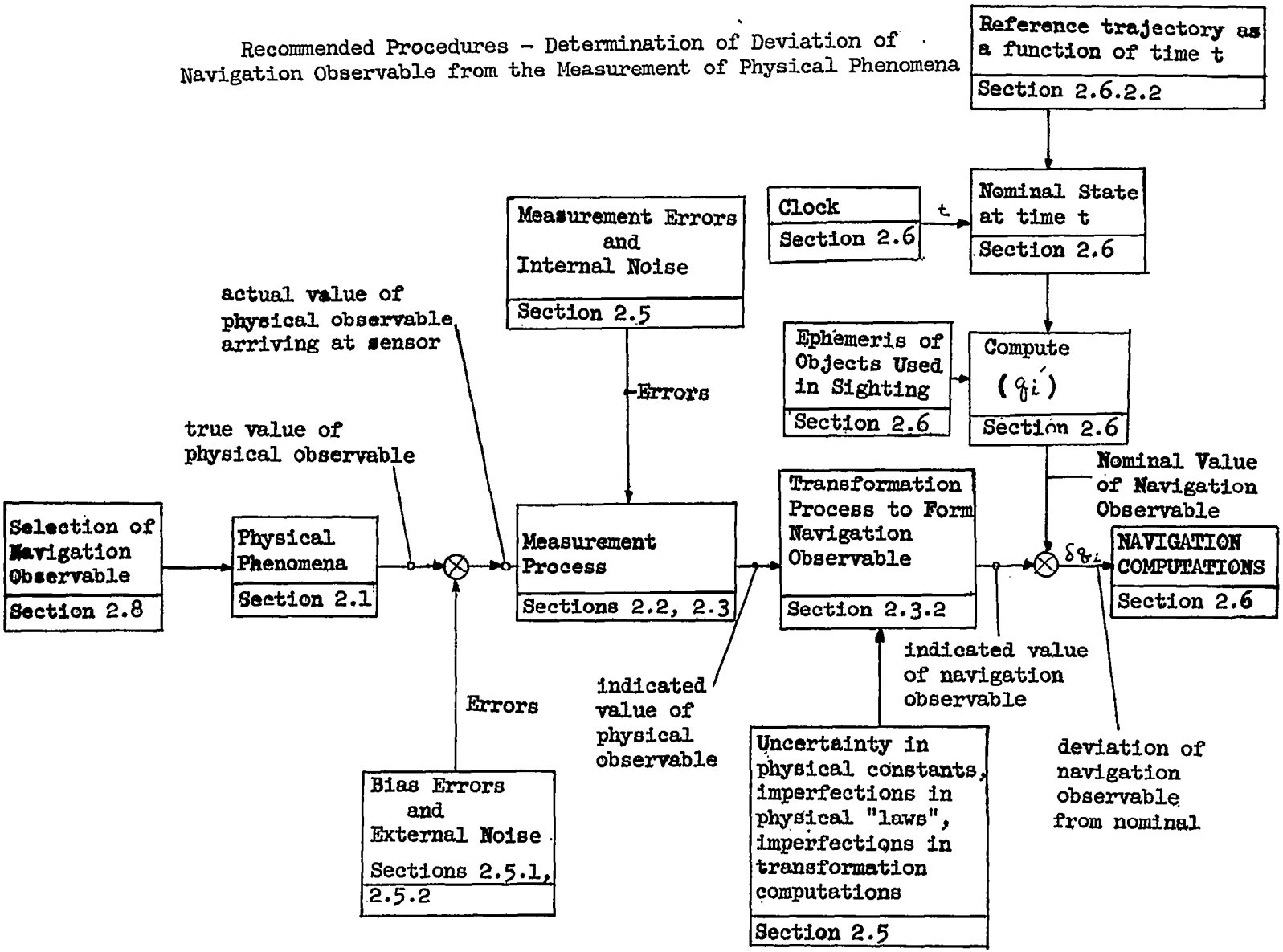
(2) The analytical techniques presented may be applied to other problems than those considered herein. Of particular importance are the techniques for developing linear approximations, techniques for noise and error analysis, techniques for generating results from difficult non-linear differential equations by suitably restricting the non-linear variations, techniques of developing observation to state vector relationships, techniques for generating solutions of the state vector from simultaneous and sequential observations, techniques for optimizing the selection of observables, and techniques for applying perturbation theory to transform non-linear to linear differential equations.

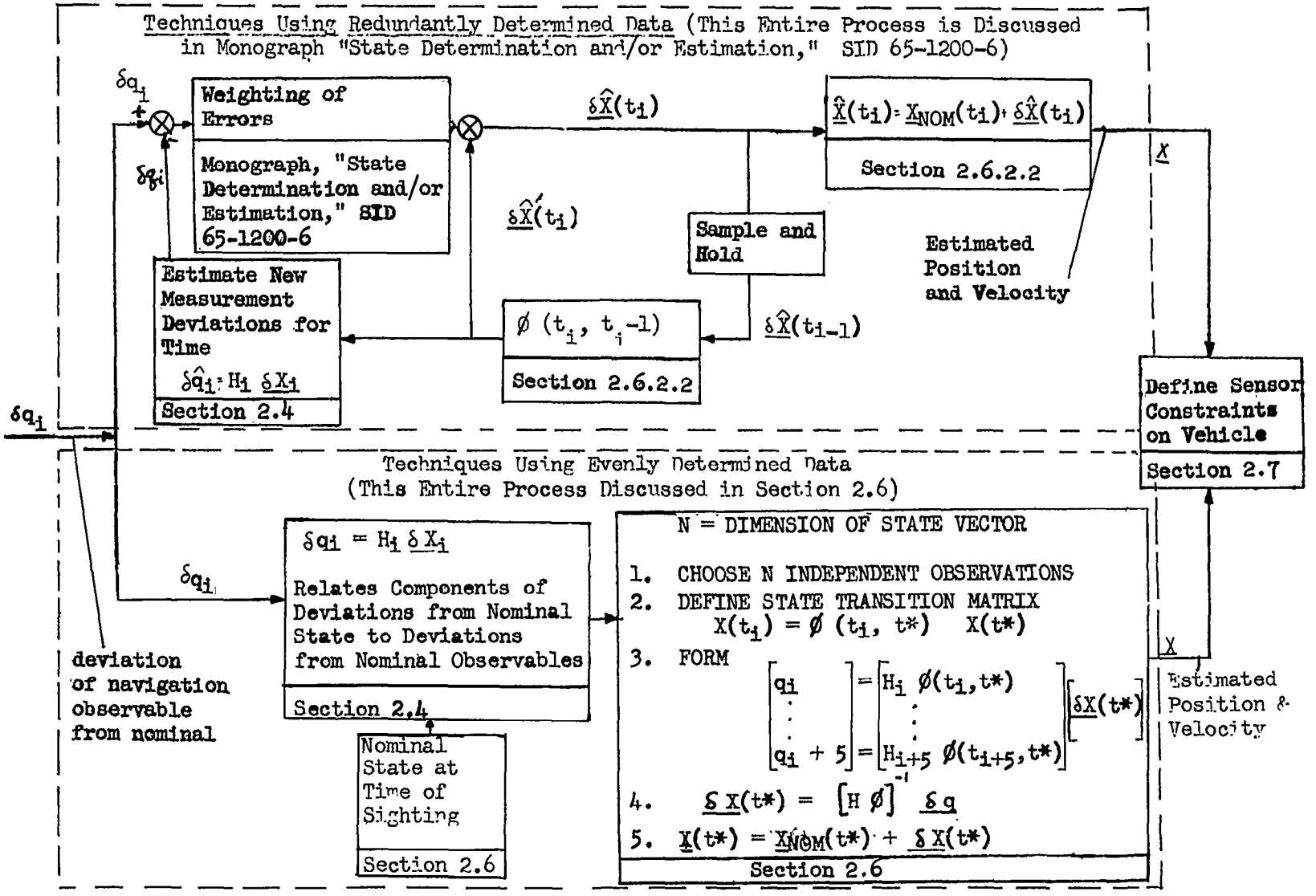
The following two block diagrams present: (1) the recommended procedures to be followed in generating the values δq_i (the deviation of the navigation observables from their nominal values) from the associated physical phenomena, and (2) the recommended procedures to be followed in generating estimates of position and velocity from the quantities δq_i .

These block diagrams are intended to give the reader a perspective of the material presented in this monograph, and to indicate the interfaces with the related monographs of this series. The sections of this monograph which are related to the various steps of the procedures are indicated in the appropriate blocks.

The navigation problem commences with the measurement of some physical phenomenon. This measurement could be any one of numerous celestial body measurements that have been mentioned, or a quantity related to the angular or translational displacement of some inertial device. Of course, as in any measurement process, there are many sources of inaccuracies. Before the

Recommended Procedures - Determination of Deviation of Navigation Observable from the Measurement of Physical Phenomena





Recommended Procedure - Navigation Computational Techniques for the Determination of an Estimate of Position and Velocity

measurement information reaches the measurement device, it is subjected to random and systematic errors that prevent the measurement device from correctly measuring the phenomenon. In addition, the measurement device itself is subjected to inaccuracies due to its limitations and internal random disturbances. As a result, the indicated value of the physical observable is in error. It has been the intent of this monograph, in part, to develop some of the theory that enables a quantitative treatment of the more important error sources that limit the performance capability of navigation sensors.

In cases such as range and range-rate measurements, inertial measurements, and angular measurements, the quantity that is measured is only an indirect measurement of the actual navigation observable that is needed. It then becomes necessary to transform the indicated measurement to the navigation observable by some physical "law." Again, there are errors introduced in this transformation process because of the limited accuracies of physical constants that must be used, and any slight imperfections of the "laws" that describe the phenomenon or its computational mechanization as used in the transformation process.

Once the measured navigation observable has been determined, it can be compared to the value of the same observable that would be observed if the vehicle were exactly on some nominal trajectory. This can be accomplished by evaluating the state vector at the time of the measurement, and calculating the nominal navigation observable using the ephemeris of the bodies used in the sighting.

The deviation of the navigation observable from nominal can now be used in one of two general ways. This monograph discusses the use of evenly determined data; i.e., the use of N independent observations to determine an N dimensional state vector. This method is discussed in detail in sections 2.6.2.1 and 2.6.2.2.

The other major method that can be used in order to determine the state involves the utilization of redundantly determined data. Although this method is extensively discussed in the following monograph ("State Determination and/or Estimation," SID 65-1200-6), it will be briefly discussed here in order that it can be related to the material presented in this monograph.

Basically, the method consists of estimating the new state and navigation observable deviations that are based on previous deviations. If the estimated and actual navigation observable deviations are identical, the process continues to update itself simply by the use of the state transition matrix. If, on the other hand, the estimated and actual navigation observables differ in value, this difference is processed in such a manner as to maximize (minimize) some measure of performance, giving attention to the accuracy that is expected from such a vector measurement. As a result, the estimate of the state deviation is defined. The new state deviation vector is then added to the nominal state vector for that particular time, and the state vector of the vehicle is thus determined.

4.0 REFERENCES

The following references were used in the preparation of some sections of this monograph. Acknowledgment is given with the appropriate reference number in the text.

1. Kuiper, G. P., ed. The Earth as a Planet/The Solar System. Vol. II, Chicago: The University of Chicago Press (1954).

2. Studies in Electromagnetic Propagation - Part I, Refractive Corrections. S.T.L. Document, GM-TM-0165-00307 (1958).

3. Skolnik, M. I. Introduction to Radar Systems. New York: McGraw-Hill Book Company, Inc. (1962).

4. Battin, R. H. Astronautical Guidance. New York: McGraw-Hill Book Company, Inc. (1964).

5. Aller, L. H. The Atmospheres of the Sun and Stars. 2nd. ed. New York: The Ronald Press Company (1963).

6. Kelso, J. M. Radio Ray Propagation in the Ionosphere. New York: McGraw-Hill Book Company, Inc. (1964).

7. Jenkins, F. A. and White, H. E. Fundamentals of Physical Optics. New York: McGraw-Hill Book Company, Inc. (1937).

8. O'Donnel, C. F., ed. Inertial Navigation Analysis and Design. New York: McGraw-Hill Book Company, Inc. (1964).

9. Parvin, R. H. Inertial Navigation. New York: D. Van Nostrand Company, Inc. (1962).

10. Leondes, C. T., ed. Guidance and Control of Aerospace Vehicles. New York: McGraw-Hill Book Company, Inc. (1963).

11. Jamieson, J. A., et al. Infrared Physics and Engineering. New York: McGraw-Hill Book Company, Inc. (1963).

The remaining references were also helpful in the preparation of this monograph, in that they provided general information that was pertinent to the discussion.

Schneider, A. M. "Vector Principles of Inertial Navigation", IRE Transactions on Aeronautical Navigational Electronics, Vol. ANE 6, No. 3 (September, 1959).

Stevens, F. Application of Optical Techniques to Interplanetary Navigation. Presented at the Third AFOSR Astronautics Symposium, Los Angeles, California, October 12-14, 1960, SAE Paper No. 230X.

Streeter, J. R. "Some Coordinate Systems for Inertial Navigation", Navigation, Vol. 6 (1958).

Vaeth, J. E. Inertial System Alignment Study. IRE 2nd National Convention on Military Electronics, Conference Proceedings (June, 1958).

Wilson, R. E. and Lewis, J. B. "Cross Coupling in Inertial Navigation Systems", Aero/Space Engineering, Vol. 17, No. 5 (May, 1958).

Broxmeyer, C. "Foucault Pendulum Effect in Schuler-tuned System", Journal Aero/Space Sciences, Vo. 27, No. 5 (May 1960).

Duncan, D. B. "Analysis of Inertial Guidance System", Jet Propulsion, Vol. 28, No. 2 (February, 1958).

Ehricke, K. A. "Error Analysis and Two-Force Field Spacecraft Orbits", Franklin Institute Monograph Series, No. 6 (December, 1958).

Mueller, R. R. Investigation of Possible Satellite Positioning Sensing Methods. Presented at ARS 14th Annual Meeting, Washington, D.C., November 16-20, 1959, Paper No. 913-59.

Palmer, R. R. and McAllister, D. E. Multiple System Operation in Long Term Inertial Navigation. Presented at ARS Guidance, Control, and Navigation Conference, Stanford University, Stanford, California, August, 1961, Paper No. 1958-61.

Wormser, E. M. and Morris, H. A. Infrared Navigation Sensors for Space Vehicles, Guidance and Control. New York: Academic Press Inc. (1962).

Index