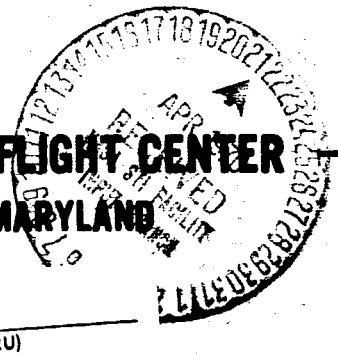
X-643-69-425
PREPRINT

ON THE BENDING OF A RAY THROUGH A SPHERICALLY SYMMETRIC ATMOSPHERE

HARVEY G. SAFREN

NOVEMBER 1969

GODDARD SPACE FLIGHT CENTER
GREENBELT, MARYLAND



X-643-69-425

ON THE BENDING OF A RAY THROUGH A SPHERICALLY SYMMETRIC ATMOSPHERE

Harvey G. Safren
Laboratory for Theoretical Studies

November 1969

Goddard Space Flight Center
Greenbelt, Maryland

ON THE BENDING OF A RAY THROUGH A SPHERICALLY SYMMETRIC ATMOSPHERE

Harvey G. Safren
Laboratory for Theoretical Studies

ABSTRACT

An equation is derived which explicitly gives the path of a ray through a spherically symmetric atmosphere. The refraction of the ray is assumed to be governed by the laws of geometric optics. The equation involves the arrival angle of the ray at the ground; this angle is obtained by solving an auxiliary equation. The latter equation is not explicitly solvable, but approximate solution formulas are obtained and a numerical solution method is described. Some of the approximate formulas are shown to be accurate to better than one percent down to elevation angles of ten degrees or less. The refractivity profile includes both the troposphere and ionosphere, and is arbitrary within a very large class of functions.

CONTENTS

ILLUSTRATIONS

TABLES

LIST OF SYMBOLS

D	a region of the $\alpha_t - R_s$ plane; $D = [0, \pi/2) \times (R_e, \infty)$
$\mathring{D}$	interior of the region D
E	a region of $\alpha_t - R_s - \delta_g$ space
$\mathring{E}$	interior of E
$\mathcal{E}, \mathcal{E}^+$	classes of ϵ-functions
f	frequency of radar or radio wave
M	electron density in the ionosphere
N	refractivity; $N = (n - 1) \times 10^6$
n	index of refraction
n_g, n_s	indexes of refraction at ground level and at the satellite, respec.
p	a point in the $\alpha_t - R_s$ plane
$\vec{R}$	radius vector from the center of the earth
R	magnitude of $\vec{R}$
R_e	radius of the earth
R_s	distance of the satellite from the center of the earth
$\vec{s}$	unit vector tangent to the ray path at a given point
α_a	arrival angle of the ray path at the ground station (apparent elevation angle of the satellite)
α_t	true elevation angle of the satellite
β_s	total bending angle of ray (through entire atmosphere)
δ_g	angle between ray path and slant path at the ground station (elevation angle error); $\delta_g = \alpha_a - \alpha_t$.

δ_s angle between ray path and slant path at the satellite

ϵ symbol denoting class membership

ϵ a quantity related to the index of refraction; $\epsilon = n_g/n - 1$.

$\tilde{\epsilon}$ the transformed ϵ -function; $\tilde{\epsilon}(\rho) = \epsilon(R)$.

ϵ_m the supremum of $|\epsilon(R)|$ on the interval $[R_e, \infty)$

ψ angle at the center of the earth between $\vec{R}$ and a line drawn to the ground station

ψ_s value of ψ at the satellite

ρ a variable related to R; $\rho = R_e/R$.

ρ_s R_e/R_s

θ angle between $\vec{s}$ and $\vec{R}$ at a point on the ray path

ξ a variable related to δ_g; $\xi = \cos^2 \alpha_t / \cos^2 (\alpha_t + \delta_g)$.

ON THE BENDING OF A RAY THROUGH A SPHERICALLY SYMMETRIC ATMOSPHERE

INTRODUCTION

To determine the orbit of a satellite with great precision it is necessary to correct ground-based radar observations for the effects of refraction by the atmosphere. In theory it is possible to compute these corrections very accurately, but in practice it is not. The atmosphere is so complex, and its behavior so variable, that one cannot hope to have a very complete description of it at the time the radar observations are made. The best one can ordinarily hope to do is to assume a model of the atmosphere which is greatly simplified, but still realistic enough so that the corrections computed from it can be regarded as reasonably accurate. In this report the main simplifying assumption made is that the atmosphere is spherically symmetric. It is also assumed that the tracking frequencies used are well above 100 megahertz, so that the effects of the earth's magnetic field are not too important, and that the variation of refractive index with height is gradual enough so that the laws of geometric optics can be assumed to hold. With these assumptions the problem of computing the arrival angle becomes manageable. (The computation of range and range-rate errors will be treated in a later report.) It is not necessary to assume any specific law for the variation of the refractive index with altitude.

The mathematical formulation of the problem follows the method outlined in Ref. 1 (pp. 120-122), but the derivation given here is more complete. The basic idea is to make use of the differential equation for a ray path. This equation does not give the ray path explicitly, because it uses the path length along the ray as the independent variable, but it is easily integrated for a spherically symmetric medium. The result is the formula known as Bouguer's Law, or the spherical form of Snell's Law: $nR \sin \theta = \text{constant}$, where n is the index of refraction at the distance R from the center of the earth, and θ is the angle between the ray path and the radius vector $\vec{R}$. This relation also does not explicitly give the ray path in cartesian or polar coordinates, but it is easy to deduce from it, by geometrical reasoning, a first-order differential equation for the ray path in terms of R and the central angle ψ (see Fig. 1). This equation can be immediately integrated, but the integrated equation still contains the unknown constant in Bouguer's law. To find the constant we impose the condition that the ray path must join the satellite to the ground station. This gives an equation for δ_g, the elevation angle error (i.e., the angle between the ray path and the slant path at the ground station). Unfortunately δ_g appears in a definite integral and it doesn't seem possible to explicitly solve for it. But it is possible to deduce a considerable amount of information about δ_g as a function of the altitude and elevation angle of the satellite, and to obtain good approximations for it.

Because it seems to be impossible to get an explicit solution for the elevation angle error δ_g, a computer program (which will be documented in a companion report) was designed to obtain accurate numerical solutions. This program can also evaluate any approximate formula for δ_g, and may be used to test the accuracy of such formulas. The approximate formula must be coded in a standard form and added to the main deck. The refractivity profile (i.e., the variation of the index of refraction with altitude) is arbitrary; it also is coded in a standard form and added to the main deck. The program was used to test the accuracy of the various approximations obtained for δ_g, and some of them were found to be quite accurate.

The work presented in this report was inspired by the writer's desire to have a mathematically rigorous, systematic and detailed treatment of the problem of the refraction of a ray through a spherically symmetric atmosphere. It is felt that the treatment of the problem given below satisfies those requirements reasonably well. Many of the results appear elsewhere in the literature, and some are well-known. Some of the results, however, are believed to be new: e.g., the simple upper bound for the elevation angle error given by the inequality (30), the very general approximate formula for the same quantity given by Equation (61), and the inequality (23), which gives a necessary and sufficient condition for the existence of a ray path joining a satellite to a ground station, in terms of the satellite's position and the refractivity profile. Even though some of the results obtained are well-known, the writer feels that the very completeness of the mathematical treatment given in this report is valuable in itself, because it brings to light questions (such as the existence of ray paths and their general behavior) which are usually not considered.

BACKGROUND MATERIAL

The material in this section is not necessary for the mathematical discussion which follows. It is inserted here to provide some background information and to motivate the various assumptions made in the mathematical treatment.

Tropospheric Index of Refraction

In most regions of the electromagnetic spectrum the troposphere is essentially non-dispersive; i.e., the index of refraction does not change appreciably with frequency. The index of refraction n of a medium is defined by $n = c/v$, where c is the velocity of light in vacuum and v is the velocity of light in the medium. Because the value of n is nearly unity throughout the atmosphere (including the ionosphere), it is customary to use a related quantity called the "refractivity," defined by $N = (n - 1) \times 10^6$. The following empirical formula

for the radio refractivity of the troposphere is in widespread use by radio scientists (see Ref. 2, p. 7):

$$N = \frac{77.6}{T}\left(P + \frac{4810\,e}{T}\right), \tag{1}$$

where T is the temperature in degrees Kelvin, P is the total atmospheric pressure in millibars and e is the partial pressure of water vapor in millibars. This formula is considered to be accurate to 0.5% for radio frequencies up to 30,000 megahertz, for normally encountered ranges of pressure, temperature and humidity. The fact that the formula does not involve the frequency shows how nearly frequency-independent the refractivity is in this region of the spectrum.

The ratio e/P is about 0.01 at ground level for 60% relative humidity, and decreases rapidly with altitude (Ref. 2, p. 11). It follows that the second term in Equation (1) is at least an order of magnitude smaller than the first term. Since the pressure P (which is the dominant term) normally decreases exponentially with height, and since the variation of T with height is not great, it would seem that an exponential decrease of N with height might be a good approximation to the actual refractivity profile. In Ref. 2 it is shown that an exponential radio refractivity profile is in fact a quite accurate model for the troposphere.

For optical frequencies the index of refraction of air is also very weakly frequency-dependent. The refractive bending of violet light differs from that of red light by only about one percent. The dependence of N on pressure and temperature, for visible light, is given in Ref. 3 (p. 62) by the formula

$$N = 82\frac{P}{T}, \tag{2}$$

where N, P and T are as defined above (the units used in Ref. 3 have been changed to the units used here). This formula is almost the same as formula (1), with the second term deleted. It thus appears that the tropospheric refractivity profile for visible light should also be closely approximated by an exponential model.

Ionospheric Index of Refraction

The refractive behavior of the ionosphere is very different from that of the troposphere. The density of air above 40 kilometers is too small to cause any appreciable refractive effects; the refractive behavior of the ionosphere is due primarily to the rarified electron gas arising from ionization by solar ultraviolet and X-radiation. (For a recent review of what is known about the formation and composition of the ionosphere, see Ref. 4.) A discussion of the propagation of electromagnetic waves in such a medium is given in Ref. 5. The following formula is derived there (p. 329) for the index of refraction n of an electron gas in the presence of a static magnetic field (the earth's field):

$$n^2 = 1 - \frac{\frac{Me^2}{m\epsilon_0}}{\omega^2 \pm \frac{e\mu_0 H_0 \omega}{m}}, \tag{3}$$

where M is the number of electrons per unit volume, e is the electronic charge, m is the electronic mass, ϵ_0 is the permittivity of free space, μ_0 is the permeability of free space, H_0 is the intensity of the component of the magnetic field in the direction of propagation of the wave and ω is the circular frequency of the wave. All quantities are in the rationalized m.k.s. system. This formula shows that the ionosphere has some remarkable refractive properties. Perhaps the most striking thing about the formula is the presence of the ± sign; this means that there are two modes of propagation with two distinct velocities. In fact it can be shown that a linearly polarized wave entering the medium is resolved into right- and left-circularly polarized components; thus a rarified electron atmosphere with a static magnetic field imposed on it acts like an anisotropic, doubly-refracting crystal. Further, it is apparent from Equation (3) that for a certain frequency the index of refraction of one of the waves becomes infinite, and that for another frequency the index of refraction of the other wave becomes zero. This behavior accounts for many of the remarkable properties of the Kennelly-Heaviside layers at radio frequencies.

The effect of the earth's magnetic field becomes less important for higher frequencies, as can be seen from Equation (3). In fact, if we use the values $e = 1.602 \times 10^{-19}$ coulomb, $e/m = 1.759 \times 10^{11}$ coulombs/kilogram, $\mu_0 = 4\pi \times 10^{-7}$ henry/meter and $\epsilon_0 = 8.854 \times 10^{-12}$ farad/meter, and if we take the maximum value of H_0 to be about 0.4 gauss, then we see that the ratio of the two terms involving ω in Equation (3) is

$$\frac{e\mu_0 H_0 \omega}{m\omega^2} = \frac{1.407}{f} \times 10^4 ,$$

where $f = \omega/2\pi$ is the frequency in cycles per second. Hence for $f > 100$ megahertz, which is about the lowest frequency likely to be used for satellite tracking, the above ratio is less than 1.4×10^{-4}. We may therefore neglect the effect of the earth's magnetic field and represent the ionospheric index of refraction by the simpler formula

$$n^2 = 1 - \frac{Me^2}{m\epsilon_0 \omega^2} . \tag{4}$$

This simplified formula still has some interesting properties. For the critical frequency $\omega = \omega_c = (Me^2/m\epsilon_0)^{1/2}$ the index of refraction becomes zero. Physically this means that a wave entering the medium is totally reflected. More generally, if M is increasing monotonically along the direction of propagation, a wave of frequency ω will be totally reflected when it reaches the region where $M = m\epsilon_0 \omega^2/e^2$. This phenomenon is used to measure the electron density profile of the ionosphere by beaming radio pulses vertically upward and measuring the time between transmission and reception of the pulses. As the transmitter frequency is increased, higher regions of the ionosphere are probed. Only that part of the ionosphere below the F_2 maximum (which lies between 200 and 400 kilometers) can be probed by this method, however, because the electron density above that point decreases again. To measure the top part of the ionosphere it is necessary to use satellites to beam radio pulses downward.

If we use the values given above for e, m and ϵ_0, and if we assume that the largest value of M in the ionosphere is about 10^{12} electrons per cubic meter (a realistic assumption, corresponding to the F_2 maximum), we find that $f_c = \omega_0 / 2\pi = 8.98$ megahertz. This is far below our lower bound of 100 megahertz; we therefore do not need to concern ourselves with the possibility of total reflection.

There is an interesting property of the above formulas for n which we have not yet mentioned: the value of n is less than unity. This means that a wave of any frequency greater than f_c is propagated in a rarified electron gas with a velocity greater than the velocity of light in free space. This of course does not contradict relativity theory. In the derivation of Equation (3) it is implicitly assumed that a sinusoidal plane wave of frequency f, of infinite extent, is being propagated through a rarified electron gas which is also of infinite extent. But every physical wave is limited in space and time, and has a frequency spectrum of finite (i.e., non-zero) width. Such a waveform can be represented mathematically by a Fourier integral, which may be regarded intuitively as the sum of an (uncountably) infinite number of sinusoids of infinitesimal amplitude, each sinusoid being of infinite extent. Each of these hypothetical sinusoidal waves is

propagated with a velocity greater than c, the velocity of light in free space; but their sum, the actual physical waveform, is propagated with a "velocity" less than c. The reason for using quotation marks is the impossibility of precisely and unambiguously defining the velocity of the physical wave. Each of the component sinusoids is propagated with a different velocity, because the index of refraction varies with frequency (i.e., the medium is "dispersive"). Therefore their phases are constantly varying relative to each other, and the shape of their sum is constantly changing. The physical waveform is thus continuously deformed as it moves through the medium, and it is not possible to uniquely define its velocity. However, it is possible to define velocities which are useful under certain circumstances. The one most commonly used is the group velocity; it is meaningful only for waveforms whose frequency spectrum is confined to a very narrow band. If this band lies in a region where the dispersion of the medium is normal and moderate (e.g., away from absorption bands), the group velocity is approximately the velocity of energy propagation. In Ref. 5 the following expression is derived for the group velocity:

$$v_{group} = \frac{1}{\frac{dk}{d\omega}} ,$$

where k is the wave number ($k = 2\pi/\lambda$, where λ is the wavelength in the medium), and the derivative is evaluated at ω_0, the center of the narrow frequency band. If we use the relation $k = \omega/v_{phase}\ (\omega)$, where $v_{phase}\ (\omega)$ is the velocity (in the medium) of the component sinusoid of circular frequency ω, we obtain a relation between the group index of refraction n_{group} ($= c/v_{group}$) and the phase index n_{phase} ($= c/v_{phase}$):

$$n_{group} = n_{phase} \cdot \left(1 + \frac{\omega}{n_{phase}} \cdot \frac{d\, n_{phase}}{d\omega}\right) . \qquad (5)$$

The phase velocity is so called because it is the rate at which surfaces of constant phase, of a pure sinusoidal plane wave of infinite extent, are propagated through the medium. If the medium is non-dispersive (as the troposphere is, for example), the derivative is zero and the group velocity is the same as the phase velocity (which is then independent of frequency).

If we apply Equation (5) to the expression for n_{phase} given by Equation (4), we obtain the following expression for the group index of refraction of the ionosphere (neglecting the earth's magnetic field):

$$n_{group} = \left(1 - \frac{M e^2}{m \epsilon_0 \omega^2}\right)^{-1/2} . \qquad (6)$$

It is apparent from Equation (6) that $v_{group} < c$, as it should be, if the group velocity represents the rate at which energy is propagated through the medium. It is also worth pointing out that $v_{group} \cdot v_{phase} = c^2$.

The problem of the propagation of electromagnetic signals through a dispersive medium is a very complex one. A very illuminating discussion of this problem is given in Ref. 5; it is based on an investigation conducted by Sommerfeld and Brillouin in 1914 (Ref. 6) The results of this investigation show that the time t of arrival of a signal at a given point in a dispersive (homogeneous and isotropic) medium can never be less than d/c, where d is the distance of the given point from the signal source, which is assumed to begin emitting the signal at time t = 0.

It is important to note that the phase index must be used to compute the bending of a radio wave through the ionosphere, but the group index must be used to compute the travel time of the signal. To see why this is so, consider a signal which has a narrow frequency spectrum, in the sense that it has a small ratio of spectral width to carrier frequency. (A modulated radio signal will ordinarily have a narrow spectrum in this sense, and even a radar pulse consists of so many cycles that its spectrum is fairly narrow.) Each (hypothetical) component wave of frequency f suffers refraction, the degree of which depends on its index of refraction n(f), which we have called the phase index. Because of the narrowness of the spectrum, the variation (throughout the spectral width) in the amount of refraction is small. In other words, the angular dispersion of the signal in space is small. Thus the curved path followed by the physical wave is essentially the same as that followed by any one of its hypothetical components, for the calculation of which the phase index must naturally be used.

On the other hand, if the travel time is to be computed the velocities of the hypothetical component waves are not of interest. It is the velocity of the resultant physical wave that matters; and this, as we have seen, is given by the group index of refraction.

Good approximations for the phase and group refractive indices may be derived from equations (4) and (6) by using the binomial approximation $(1 - a)^{\pm 1/2} \approx 1 \mp a/2$, with $a = M e^2/m \epsilon_0 \omega^2$. If we use the values for the constants given previously, and if we again take 10^{12} electrons per cubic meter as the maximum value of M and 10^8 cycles per second as the minimum value of f, we see that the maximum value of a is about 0.00806. The binomial approximation is

thus quite accurate. If we use it, and if we insert the values of the constants, equations (4) and (6) become

$$n_{phase} = 1 - 40.3 \frac{M}{f^2} \tag{7}$$

and

$$n_{group} = 1 + 40.3 \frac{M}{f^2} , \tag{8}$$

where M is the electron density in electrons per cubic meter and f is the carrier frequency in cycles per second. In terms of the refractivity $N = (n - 1) \times 10^6$ these equations become

$$N_{phase} = - 40.3 \times 10^6 \frac{M}{f^2} \tag{9}$$

and

$$N_{group} = 40.3 \times 10^6 \frac{M}{f^2} . \tag{10}$$

From these formulas it is clear that both the phase and group refractivity profiles in the ionosphere are proportional to the electron density profile, for a given frequency, and that they decrease rapidly with increasing frequency. For radio frequencies of a few hundred megahertz the ionospheric contribution to refraction is important. For frequencies of thousands of megahertz it becomes very small, and for optical frequencies it is infinitesimal compared to the effects of the troposphere.

The Approximation of Geometric Optics

In many published papers dealing with ionospheric refraction the analysis is based on the laws of geometric optics, with no comment as to the validity of these laws in the ionosphere. It is not difficult to derive the laws of geometric optics from Maxwell's equations, for the case of dielectric media; such a derivation is given in chapter 3 of Ref. 1. The derivation is more complicated for a plasma such as the ionosphere. It turns out that the laws of geometric optics do

hold for the ionosphere, provided the spatial variation of electron density is sufficiently gradual; just as for dielectrics, these laws will hold provided the properties of the medium do not change much over a distance of one wavelength of the radiation. Fortunately this condition is usually satisfied for radio waves in the ionosphere. Very detailed analyses of the applicability of geometric optics to ionospheric propagation are given in References 7 and 8.

The main result from geometric optics that will be used in the ensuing analysis is the spherical form of Snell's law, sometimes called Bouguer's Law. It applies to a spherically symmetric medium, and states that

$$n R \sin \theta = \text{constant},$$

where n is the index of refraction at the distance R from the center of symmetry (in our case, the center of the earth), and θ is the angle between the ray path and the radius vector (from the center of symmetry) at distance R from the center. This law is the spherical analogue of the more commonly used planar form of Snell's law. It should be mentioned that the index of refraction n is the phase index, for the frequency of the radiation being used.

FORMULATION AND SOLUTION OF THE PROBLEM OF FINDING RAY PATHS

Formulation of Problem

Brief Statement of Problem – We want to solve the following problem: Given a point source of radio or radar waves in or above the earth's atmosphere, and given a station on the ground, does there exist a ray path joining the source to the ground station? If so, how can we compute the path? In particular, how can we find the apparent elevation of the source as seen from the ground? The ray path will of course deviate from the slant path, and the measured elevation angle will differ from the true elevation angle, because of atmospheric refraction of the radio waves.

Simplifying Assumptions – To make the mathematical analysis tractable we make the following assumptions:

- The atmosphere is spherically symmetric; i.e., the index of refraction depends only on height (above a spherical earth).
- The effects of the earth's magnetic field are negligible.

- The propagation of radio waves through the atmosphere can be adequately described by the laws of geometric optics (i.e., by ray-tracing methods).

The second and third assumptions will generally be valid for radio frequencies above 100 megahertz. The first assumption is only a rough approximation, since the ionospheric refractivity profile varies considerably with latitude.

Bouguer's Law as the Characterizing Property of Ray Paths – With the above assumptions we can briefly state our problem as follows: Find a curve (the ray path) joining the source to the ground station, which has the property that it satisfies the spherical form of Snell's law (Bouguer's law),

$$n\vec{R} \times \vec{s} = \overrightarrow{\text{constant}}, \tag{11}$$

at every point along the curve. (See Chapter III of Ref. 1 for a derivation of this relation.) In particular, find the arrival angle; i.e., the angle between the ray path and the horizontal at the ground station. Figure 1 illustrates the geometry. In equation (11) and Figure 1, $\vec{R}$ is the radius vector from the center of the earth to any point on the ray path, R is the magnitude of $\vec{R}$, n is the phase index of refraction (for the radio frequency being used) at distance R from the center of the earth and $\vec{s}$ is a unit vector tangent to the ray path, directed outward along the path (i.e., away from the ground station). The arrival angle is denoted by α_a, and the true elevation angle of the satellite by α_t. The angle ψ is the central angle, as shown in Figure 1. The other quantities are defined below. It is shown in References 7 and 8 that the derivative dn/dR must be continuous, and in fact slowly varying, in order for geometric optics to be applicable. We will therefore assume that n is a continuously differentiable function of R, from R_e (the radius of the earth) to infinity, and we will look for ray paths which are continuously differentiable curves.

Planarity of the Ray Path – It follows from Equation (11) that any ray path (i.e., any curve which satisfies Equation (11)), lies entirely in a plane containing the earth's center. To see this, consider any point p along the curve, and let P_p be the plane determined by the vectors $\vec{R}_p$ and $\vec{s}_p$ at the point p. Now let q be any other point along the curve, and let P_q be the plane determined by $\vec{R}_q$ and $\vec{s}_q$. It is clear from Equation (11) that the normals to P_p and P_q are parallel; since both planes must pass through the center of the earth, it follows that they must be identical. Since p and q are arbitrary points, it is clear that the entire curve must lie in one plane.

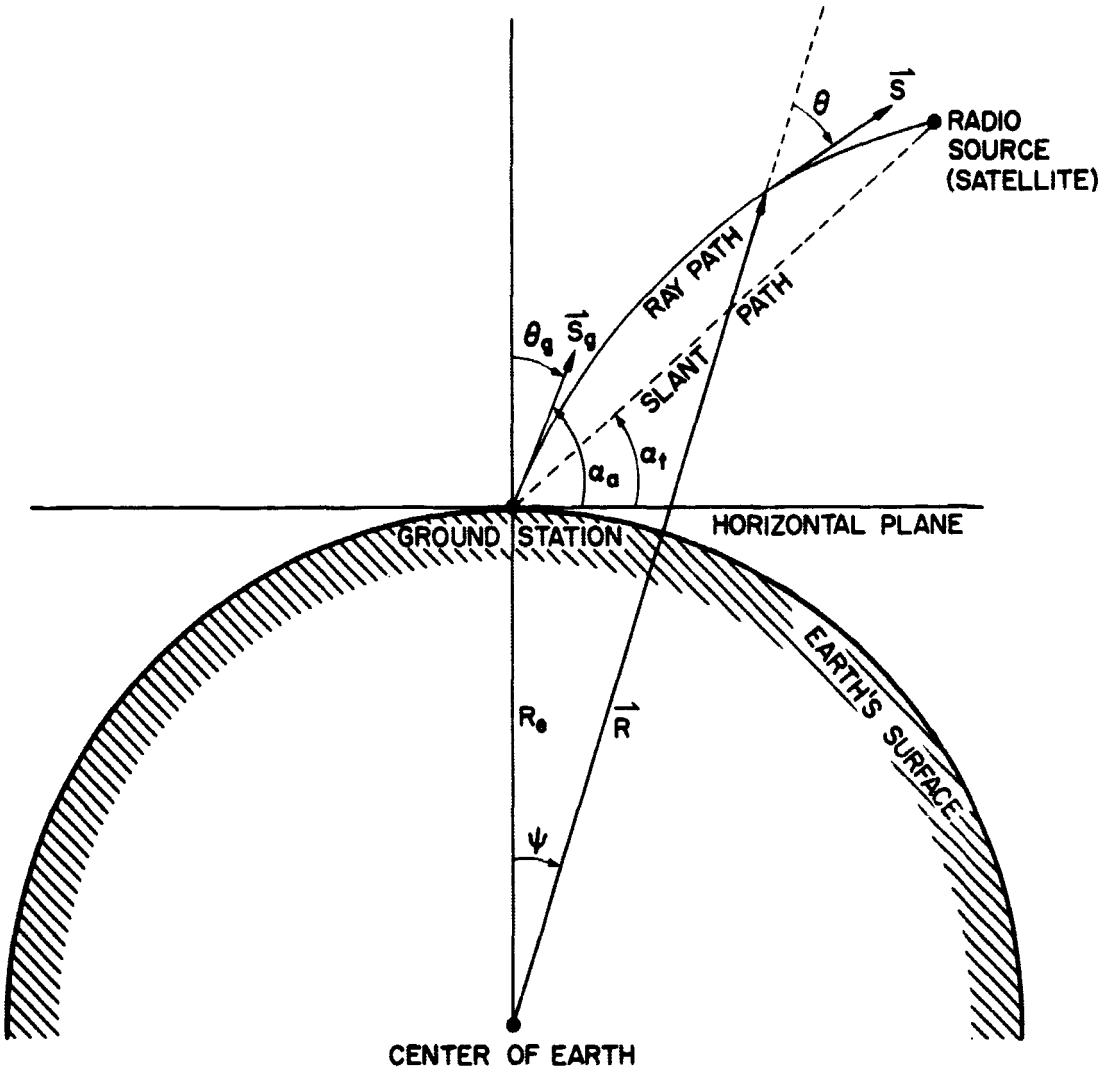


Figure 1–The Ray Path

Bouguer's Law in Scalar Form – If there exists a ray path which joins the satellite to the ground station, then of course it lies in the plane determined by those two points and the earth's center. The equation of such a ray path can therefore be written as a single scalar equation. Probably the most natural way of deriving a scalar equation from Equation (11) is to take the magnitude of each side; if we take account of algebraic signs, the result can be written in the form

$$n R \sin \theta = c, \tag{12}$$

where θ is the angle between $\vec{R}$ and $\vec{s}$, taken positive in the clockwise sense from $\vec{R}$ toward $\vec{s}$, with $-\pi < \theta \leq \pi$, and c is a constant. Equation (12) applies to any ray path, whether or not it passes through the ground station or the satellite. If we explicitly require that the ray path terminate at the ground station by evaluating the constant c at that point, Equation (12) becomes

$$n\, R \sin \theta = n_g\, R_e \sin \theta_g = n_g\, R_e \cos \alpha_a \,, \tag{13}$$

where n_g, R_e (the radius of the earth) and θ_g are the values of n, R and θ, respectively, at the ground station and $\alpha_a = \pi/2 - \theta_g$. Note that Equation (13) is merely a condition that a ray path must satisfy; it does not explicitly give the ray path in any coordinate system. In fact, Equation (13) is really a differential equation for the ray path, because the angle θ is clearly related to the slope of the curve.

The Epsilon Function – It is convenient to introduce a quantity ϵ, related to the index of refraction n, and defined by the relation

$$\epsilon = \frac{n_g}{n} - 1. \tag{14}$$

Since n differs only slightly from unity, it follows that ϵ is a small quantity; further, $\epsilon(R_e) = 0$, since $n(R_e) = n_g$. It is also worth noting that the condition $0 < n(R) < \infty$ (for all $R \geq R_e$) implies that $-1 < \epsilon(R) < \infty$ (for all $R \geq R_e$); this condition will certainly hold in the earth's atmosphere, including the ionosphere, for frequencies above 100 megahertz. In terms of ϵ, Equation (13) becomes

$$\sin \theta = (1 + \epsilon) \cdot \frac{R_e}{R} \cdot \cos \alpha_a \,. \tag{15}$$

Equation (15) is the relation which must be satisfied by a ray path emanating from the ground station at the angle α_a.

Condition for the Existence of a Complete Set of Ray Paths – It is important to note that there may not be any curve which satisfies Equation (15), for a given

value of α_a. For such a curve to exist, it is obviously necessary that the right-hand side of Equation (15) be less than or equal to unity for every value of R in some interval starting at R_e. If we want to ensure the existence of a "complete set" of ray paths, i.e., a ray path of infinite extent (emanating from the ground station) for every value of α_a in the interval $0 \leq \alpha_a \leq \pi/2$, it is necessary to assume that the condition

$$(1 + \epsilon)\frac{R_e}{R} \leq 1$$

is satisfied for every value of R in the interval $R_e \leq R < \infty$. If we solve this inequality for ϵ, it becomes

$$\epsilon(R) \leq \frac{R - R_e}{R_e},$$

for all $R \geq R_e$. It turns out that this condition is not sufficient for the existence of a complete set of ray paths, but it will be shown that the slightly stronger condition

$$-1 < \epsilon(R) \leq (1-\sigma) \cdot \frac{R - R_e}{R_e} \tag{16}$$

for all $R \geq R_e$, where $0 < \sigma < 1$ and equality holds only for $R = R_e$, is sufficient. (The fact that $\epsilon > -1$ is obvious from the definition of ϵ.)

The Class $\mathcal{E}$ of Epsilon Functions – For convenience we will denote by $\mathcal{E}$ the set of ϵ-functions which have the following properties:

- $\epsilon(R)$ is continuously differentiable for all $R \geq R_e$;
- $\epsilon(R)$ satisfies condition (16) for some value of σ in the interval (0,1).

It is worth noting that it is not clear whether the condition (16) is necessary for the existence of a complete set of ray paths, and it seems possible that it night ever be unduly restrictive. It turns out that it is not too restrictive, at least in the sense that any physically realistic profile n(R) will probably satisfy the condition (see the appendix for an example).

Derivation of an Explicit Equation for Ray Paths

We will now show that, for any ϵ-function in the class $\mathcal{E}$, equation (15) can be written as a differential equation in the variables R and ψ, for any $\alpha_a \in [0, \pi/2)$, and that this equation can be integrated to yield an explicit equation for the ray paths (with α_a as a parameter).

Some General Properties of Ray Paths – We will begin the proof by showing that, for any $\alpha_a \in [0, \pi/2)$, the ray path (assuming it exists) will have the property that R is a single-valued and strictly monotonically increasing function of ψ. It is clear that such a path will never dip below the earth's surface and that it will never "double back" on itself across any given radius vector.

The case $\alpha_a = \pi/2$ must be treated separately. It is clear that in this case the ray path will just be a vertical straight line, whether or not condition (16) is satisfied, because such a line satisfies Equation (15). No other path can be a ray path in this case, since such a path (which necessarily would be initially vertical and would later deviate from the vertical) would have to intersect some non-vertical line from the center of the earth; at the point of intersection, $\sin\theta$ would not be zero. (It is also clear from the symmetry of the situation that the ray path must be a vertical straight line.)

We now return to the general case where $\alpha_a < \pi/2$. We first show that each ray path has the property that R is a single-valued function of ψ, whether or not condition (16) is satisfied. From Equation (15) it is clear that, for any $\alpha_a \in [0, \pi/2)$, $\sin\theta > 0$ at every point along the associated ray path, assuming one exists. Since $-\pi < \theta \leq +\pi$ (by definition), it follows that $0 < \theta < \pi$ at every point along the path. It is therefore clear that the ray path cannot double back across any given radius vector; in other words, for each value of ψ there is at most one value of R.

We next show that, for each $\alpha_a \in [0, \pi/2)$, the associated ray path (if it exists) has the property that R is a strictly monotonically increasing function of ψ, if condition (16) is satisfied. If we use the inequality (16), it follows from Equation (15) that

$$\sin\theta \leq \left[1 - \sigma\left(1 - \frac{R_e}{R}\right)\right] \cdot \cos\alpha_a \leq \cos\alpha_a$$

for all $R \geq R_e$, with equality holding only for $R = R_e$. It follows that $\sin\theta < 1$ at all points of every ray path, except that $\sin\theta = 1$ at the single point $R = R_e$ for the ray path for which $\alpha_a = 0$. We recall that we are looking for ray paths

which are continuously differentiable curves, and are therefore continuous; for such a ray path it follows from the above condition on sin θ that $\theta < \pi/2$ at every point, except of course that $\theta = \pi/2$ at the ground station for the ray path for which $\alpha_a = 0$. It follows that for each such ray path R increases strictly monotonically with ψ.

Transformation of Bouguer's Law to a Differential Equation in Polar Coordinates – We are now ready to write Equation (15) as a differential equation. This will be accomplished by deriving a relation between the angle θ and the derivative $d\psi/dR$; the desired relation can be easily found with the help of Figure 2.

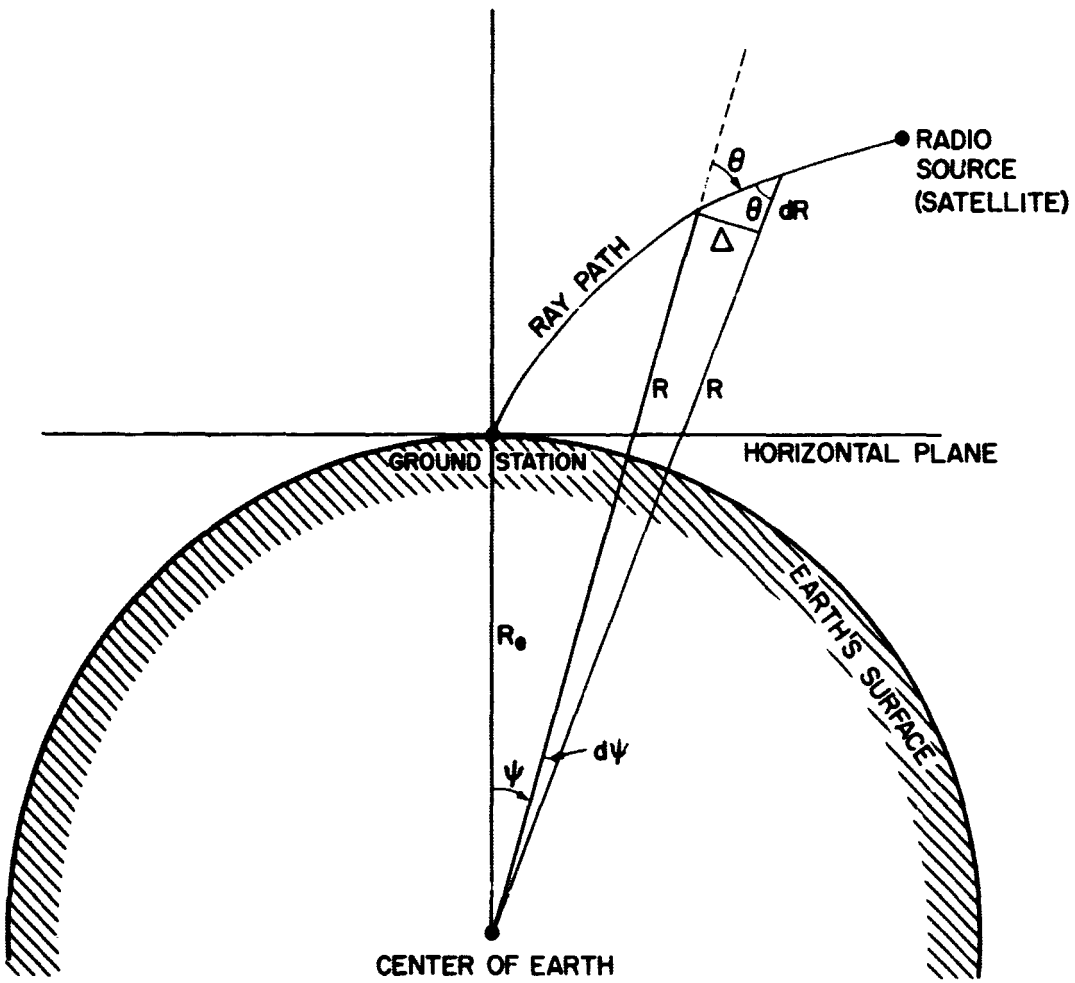


Figure 2–Diagram for the Calculation of $dR/d\psi$

From the figure we see that

$$d\psi = \frac{\Delta}{R}$$

and

$$\tan\theta = \frac{\Delta}{dR} .$$

The second equation is invalid only for the point $R = R_e$ along the ray path (if it exists) for which $\alpha_a = 0$. With this one exception, it follows that

$$\frac{dR}{d\psi} = \frac{R}{\tan\theta} .$$

The presence of $\tan\theta$ in the denominator causes no difficulty; as we have shown above, $0 < \theta < \pi/2$ along every ray path, and therefore $0 < \tan\theta < \infty$ (except that $\tan\theta$ is infinite at the one point mentioned above). It follows that $0 < dR/d\psi < \infty$, except that $dR/d\psi = 0$ at the one exceptional point. Because of the strict monotonicity of $R(\psi)$ we know that the inverse function $\psi(R)$ exists; from the above condition on the derivative of $R(\psi)$ we have $0 < d\psi/dR < \infty$, except that $d\psi/dR = \infty$ at the exceptional point. We may therefore write

$$\frac{d\psi}{dR} = \frac{1}{\frac{dR}{d\psi}} = \frac{\tan\theta}{R} . \tag{17}$$

From Equation (15) we see that

$$\tan\theta = \frac{\sin\theta}{+\sqrt{1 - \sin^2\theta}} = \frac{(1+\epsilon)\cdot\frac{R_e}{R}\cdot\cos\alpha_a}{\sqrt{1-(1+\epsilon)^2\left(\frac{R_e}{R}\right)^2\cos^2\alpha_a}} . \tag{18}$$

The plus sign is included to explicitly show that $\cos\theta > 0$, since θ is known to be in the first quadrant; this equation is therefore completely equivalent to Equation (15). From Equations (17) and (18) it follows that

$$\frac{d\psi}{dR} = \frac{(1+\epsilon)\cdot\left(\frac{R_e}{R^2}\right)\cdot\cos\alpha_a}{\sqrt{1-(1+\epsilon)^2\cdot\left(\frac{R_e}{R}\right)^2\cdot\cos^2\alpha_a}}. \tag{19}$$

This is a differential equation for the ray path, for any given angle α_a. The quantity under the square-root sign will always be greater than zero, except at the one exceptional point, where it is zero. (It should be noted that this may fail to be true if condition (16) does not hold.)

The Explicit Equation of a Ray Path – We have shown that, for any $\epsilon \in \mathcal{E}$, the defining equation for a ray path may be written in the form of the differential equation (19). It must now be asked whether this equation has solutions; if it does, then of course those solutions are the ray paths we are looking for, and they have all the properties described above (such as monotonicity). To find out whether Equation (19) has solutions for all α_a we integrate it formally to obtain the relation

$$\psi = \int_{R_e}^{R} \frac{[1+\epsilon(s)]\,\frac{R_e}{s^2}\cos\alpha_a}{\sqrt{1-[1+\epsilon(s)]^2\left(\frac{R_e}{s}\right)^2\cos^2\alpha_a}}\,ds, \tag{20}$$

which is the equation of a ray path in the polar coordinates ψ and R, provided the integral exists. For $\alpha_a \neq 0$ the quantity in the denominator is greater than zero for all $s \geq R_e$, and the integral therefore exists. The case $\alpha_a = 0$ requires a special proof of the existence of the integral, since the integrand becomes infinite at $R = R_e$. The proof makes use of the inequality (16) to obtain a majorant function for the integrand; the integral obtained by replacing the original integrand by this majorant function is then shown to exist. If the condition (16) is assumed to hold, then a simple calculation shows that

$$1 - (1+\epsilon)^2 \left(\frac{R_e}{s}\right)^2 \geq \sigma \left(1 - \frac{R_e}{s}\right) \left[2 - \sigma \left(1 - \frac{R_e}{s}\right)\right]$$

$$\geq \sigma \left(1 - \frac{R_e}{s}\right)$$

$$\geq 0$$

for all $s \geq R_e$, with equality holding only for $s = R_e$. It also follows from (16) that

$$(1+\epsilon)\frac{R_e}{s^2} \leq \left(1 - \sigma + \sigma \cdot \frac{R_e}{s}\right)\frac{1}{s}$$

$$< \left(1 - 0 + 1 \cdot \frac{R_e}{R_e}\right)\frac{1}{\sqrt{R_e} \cdot \sqrt{s}}$$

$$= \frac{2}{\sqrt{R_e} \cdot \sqrt{s}}$$

for all $s \geq R_e$. With these inequalities we may form the majorant function M(s):

$$\frac{(1+\epsilon)\dfrac{R_e}{s^2}}{\sqrt{1 - (1+\epsilon)^2\left(\dfrac{R_e}{s}\right)^2}} < \frac{2}{\sqrt{R_e} \cdot \sqrt{\sigma\,(s - R_e)}} = M(s),$$

for all $s > R_e$. As $s \to R_e$, both functions become infinite. It is easy to show that the integral exists with this majorant function as the integrand; the substitution $x = s - R_e$ yields

$$\int_{R_e}^{R} M(s)\,ds = \frac{4}{\sqrt{\sigma}}\sqrt{\frac{R}{R_e} - 1},$$

which is finite for all $R \geq R_e$. It follows that the integral exists even when $\alpha_a = 0$. (It is not difficult to show that if we were to allow σ to be zero the integral would diverge for $\alpha_a = 0$.) We conclude that, for any $\epsilon \in \mathcal{E}$, a complete set of ray paths exists; they are given explicitly by Equation (20).

The Existence of a Ray Path Connecting a Given Satellite Position to the Ground Station

We will now investigate more closely the character of this bundle of ray paths. First, we note that no two ray paths can intersect. To see this, it is sufficient to note that the integral in Equation (20) is a strictly monotonically increasing function of $\cos \alpha_a$; it therefore cannot take the same value (for a given value of R) for two different values of α_a. Because no two paths intersect it is clear that the ray path (if it exists) joining a given point to the ground station is unique.

The second property worth noting about the bundle of ray paths is that the paths change continuously with α_a. A rigorous statement of this property would require that the set of ray paths be made into a normed space, perhaps with the "sup norm;" but the property can be seen intuitively by noting the fact that the integral in Equation (20) is a continuous function of α_a, for any given value of R.

The third and most important property of the set of ray paths concerns the region of space which is "covered" by the paths; i.e., the set of points which lie on some ray path. To investigate this question will be our next task.

General Condition for the Existence of a Ray Path – Suppose we are given a point (α_t, R_s), where α_t is the (true) elevation angle of a satellite (taken positive upwards), and R_s is its distance from the earth's center. We can determine whether this point lies on one of our ray paths by setting $\psi = \psi_s$ (the value of ψ corresponding to α_t and R_s) and $R = R_s$ in Equation (20); if there exists a value of α_a in the interval $[0, \pi/2)$ which satisfies the resulting relation, then the point lies on the ray path corresponding to that value of α_a.

To write the relation explicitly we must express ψ_s in terms of α_t and R_s; this is easily done by applying the law of sines to the triangle formed by the earth's center, the ground station and the satellite, as shown in Figure 3. The law of sines states that

$$\frac{\cos \alpha_t}{R_s} = \frac{\sin \gamma}{R_e}$$

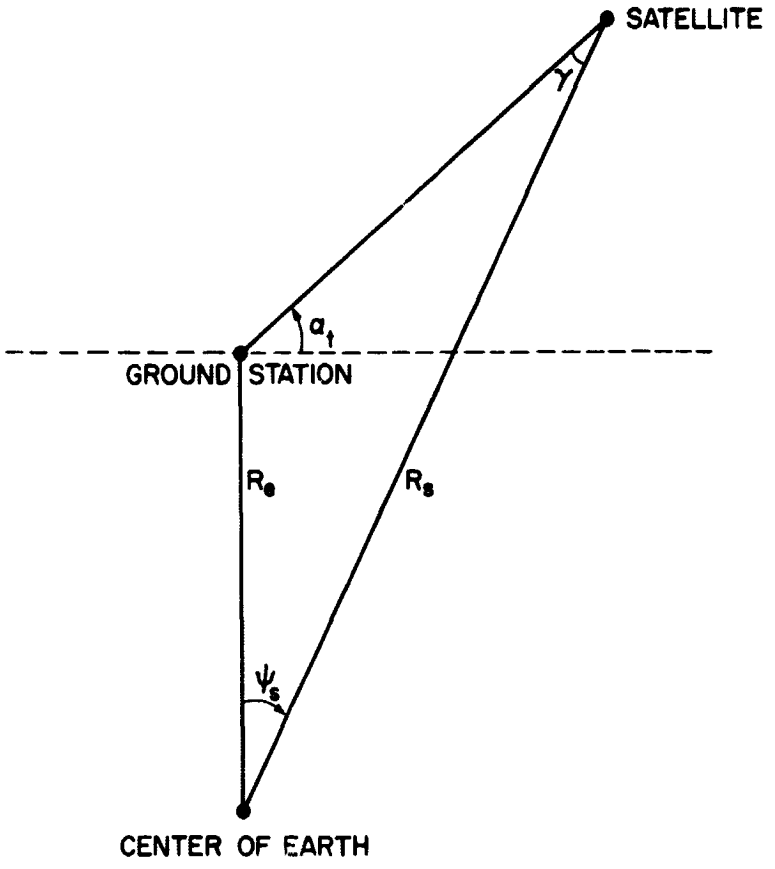


Figure 3–Diagram for the Calculation of ψ_s

or

$$\sin\gamma = \frac{R_e}{R_s}\cos\alpha_t \,.$$

This expression for $\sin\gamma$ is easily seen to be valid for $R_s > R_e$ and $\alpha_t \in (-\pi/2, \pi/2)$; i.e., for any point outside the earth and to the right of the ground station. It is clear that $\sin\gamma < 1$ everywhere in this region; since γ is clearly a continuous function of α_t and R_s in this region, and since it is obvious that $\gamma < \pi/2$ for some points, it follows that $\gamma < \pi/2$ at every point of the region. We may therefore write

$$\gamma = \sin^{-1}\left(\frac{R_e}{R_s}\cdot\cos\alpha_t\right) ,$$

where the value of the arc sin is taken in the first quadrant. It follows that

$$\psi_s = \pi - \left(\frac{\pi}{2} + \alpha_t\right) - \gamma$$

$$= \frac{\pi}{2} - \alpha_t - \sin^{-1}\left(\frac{R_e}{R_s}\cos\alpha_t\right),$$

or

$$\psi_s = \cos^{-1}\left(\frac{R_e}{R_s}\cos\alpha_t\right) - \alpha_t, \tag{21}$$

where the value of the arc cos is taken in the first quadrant; this expression for ψ_s is valid for $R_s > R_e$ and $\alpha_t \in (-\pi/2, \pi/2)$.

We can now state that the point (α_t, R_s) will lie on some ray path if and only if the relation

$$\int_{R_e}^{R_s} \frac{(1+\epsilon)\frac{R_e}{R^2}\cos\alpha_a \, dR}{\left[1 - (1+\epsilon)^2\left(\frac{R_e}{R}\right)^2\cos^2\alpha_a\right]^{1/2}} = \cos^{-1}\left(\frac{R_e}{R_s}\cos\alpha_t\right) - \alpha_t \tag{22}$$

is satisfied for some value of $\alpha_a \in [0, \pi/2)$. Since the integral decreases monotonically as α_a increases, we see that Equation (22) will have a (unique) solution if and only if

$$\int_{R_e}^{R_s} \frac{(1+\epsilon)\frac{R_e}{R^2}\,dR}{\left[1 - (1+\epsilon)^2\left(\frac{R_e}{R}\right)^2\right]^{1/2}} \geq \cos^{-1}\left(\frac{R_e}{R_s}\cos\alpha_t\right) - \alpha_t. \tag{23}$$

In other words, given any ϵ-function in the class $\mathcal{E}$ and any point $(\alpha_t, R_s) \in (-\pi/2, \pi/2) \times (R_e, \infty)$, there will exist a ray path connecting that point to the ground station if and only if condition (23) is satisfied.

We have now answered the question posed earlier concerning the region of space covered by the complete set of ray paths (which is guaranteed to exist for

any ϵ -function in $\mathcal{E}$); we have shown that, for any given ϵ-function in $\mathcal{E}$, the points which lie on some ray path are just those points which satisfy condition (23). As it stands, this condition is not very transparent; to get explicit results we must consider some special cases.

The Existence of Ray Paths in Some Special Cases – One special case is suggested by the practical requirement that the satellite be on or above the horizon; we might therefore be interested in values of α_t greater than zero. Since the expression on the right-hand side of (23), considered as a function of α_t, decreases (strictly) monotonically for $\alpha_t \in [-\pi/2, \pi/2]$, we see that (23) will hold for all $\alpha_t \in [0, \pi/2]$ if and only if

$$\int_{R_e}^{R_s} \frac{(1+\epsilon)\frac{R_e}{R^2}\,dR}{\left[1-(1+\epsilon)^2\left(\frac{R_e}{R}\right)^2\right]^{1/2}} \geq \cos^{-1}\left(\frac{R_e}{R_s}\right).$$

If we note that

$$\int_{R_e}^{R_s} \frac{(1+\epsilon)\frac{R_e}{R^2}\,dR}{\left[1-(1+\epsilon)^2\left(\frac{R_e}{R}\right)^2\right]^{1/2}} \geq \int_{R_e}^{R_s} \frac{\frac{R_e}{R^2}\,dR}{\left[1-\left(\frac{R_e}{R}\right)^2\right]^{1/2}} \qquad (24)$$

$$= \cos^{-1}\left(\frac{R_e}{R_s}\right)$$

for any ϵ -function in $\mathcal{E}$ which has the property that $\epsilon(R) \geq 0$ for all R, we see that condition (23) does in fact hold for all $\alpha_t \in [0, \pi/2]$ and for all R_s, for any such ϵ -function. In other words, if $\epsilon(R)$ is everywhere non-negative, then the entire region in and above the horizontal plane and to the right of the ground station is covered by the complete set of ray paths. Of course the condition $\epsilon(R) \geq 0$ is merely sufficient, not necessary, for this region to be covered. Also, for a given non-negative ϵ -function the total region covered will be even larger, unless $\epsilon(R) \equiv 0$.

To see the truth of this last statement, suppose that $\epsilon(R) \geq 0$ for all R, but $\epsilon(R) \not\equiv 0$; then (24) becomes a strict inequality. From the fact that the right-hand side of (23) decreases monotonically with α_t for $-\pi/2 \leq \alpha_t \leq \pi/2$, it then follows that, for each value of R_s, there exists $\beta \in [-\pi/2, 0)$ such that

$$\int_{R_e}^{R_s} \frac{(1+\epsilon)\frac{R_e}{R^2}\,dR}{\left[1-(1+\epsilon)^2\left(\frac{R_e}{R}\right)^2\right]^{1/2}} \geq \cos^{-1}\left(\frac{R_e}{R_s}\cos\alpha_t\right) - \alpha_t$$

for all $\alpha_t \geq \beta$, for the given value of R_s. Comparison of this relation with (23) shows that for any non-negative ϵ-function which is not identically zero, and for any value of R_s, there is an angle $\beta < 0$ such that every point (α_t, R_s) for which $\alpha_t > \beta$ is connected to the ground station by a ray path. If, in addition, it happens that

$$\beta_m = \sup\{\beta(R_s) : R_s > R_e\} < 0,$$

then the entire region

$$\{(\alpha_t, R_s) : \beta_m \leq \alpha_t < \pi/2 \text{ and } R_s > R_e\}$$

is covered by ray paths.

The condition $\epsilon(R) \geq 0$ will actually be satisfied by a realistic refractivity profile, provided that unusual conditions, such as temperature inversions near the ground, do not occur. To see this, one might note that the tropospheric refractivity is commonly assumed to decrease exponentially with altitude; thus $n < n_g$, and therefore $\epsilon = n_g/n - 1 > 0$, for all $R > R_e$. In the ionosphere $n < 1$ and therefore (since $n_g > 1$) we again have $\epsilon > 0$.

As another special case, suppose we ask what happens if $\epsilon(R) \leq 0$ for all R, but $\epsilon(R) \not\equiv 0$. One thing at least is easy to see in this case: no point on the horizontal plane lies on a ray path. To see this we just note that for such a point to lie on a ray path the condition

$$\int_{R_e}^{R_s} \frac{(1+\epsilon)\,\frac{R_e}{R^2}\,dR}{\left[1-(1+\epsilon)^2\left(\frac{R_e}{R}\right)^2\right]^{1/2}} \geq \cos^{-1}\left(\frac{R_e}{R_s}\right)$$

must be satisfied; but for $\epsilon(R) \leq 0$ and not identically zero this is impossible, because for such an ϵ-function

$$\int_{R_e}^{R_s} \frac{(1+\epsilon)\,\frac{R_e}{R^2}\,dR}{\left[1-(1+\epsilon)^2\left(\frac{R_e}{R}\right)^2\right]^{1/2}} < \int_{R_e}^{R_s} \frac{\frac{R_e}{R^2}\,dR}{\left[1-\left(\frac{R_e}{R}\right)^2\right]^{1/2}} = \cos^{-1}\left(\frac{R_e}{R_s}\right).$$

Since no ray path passes through or even touches the horizontal plane, it follows that all ray paths lie completely above that plane for any non-positive (and non-zero) ϵ-function.

More detailed results on the existence of ray paths could be derived, but for our purposes we do not need to go any further with this analysis.

Recapitulation

Summary of Results – Before continuing with our analysis we will recapitulate what we have found so far. We showed that any ray path joining a satellite to a ground station through a spherically symmetric atmosphere must lie entirely in the plane determined by the satellite, the ground station and the center of the earth. Since there may not be any ray path joining a given satellite position to the ground station, it seemed most logical to focus our attention on the bundle of ray paths emanating from the ground station, rather than on the satellite position; we may limit this bundle to a plane, because of symmetry, and may parametrize it by the variable α_s, the elevation angle of a ray path at the ground station. We stated a condition on the refractivity profile (condition (16)) which we claimed is sufficient to ensure the existence of a complete set of ray paths; i.e., a ray path of infinite extent for every value of α_s in the interval $[0, \pi/2]$. (This condition is actually satisfied by realistic refractivity profiles; see the appendix for an example.) On the assumption that condition (16) holds, we then actually proved that a complete set of ray paths exists; we derived an explicit equation for them

(Equation (20)), with α_a as a parameter, and we showed that they have certain properties. Among these properties are a single-valued dependence of R on ψ, with R increasing strictly monotonically with ψ (for $\alpha_a < \pi/2$). We also showed that no two ray paths intersect, which is equivalent to saying that the ray path joining a satellite position to the ground station is unique, if it exists at all. We next investigated the question of the region covered by the bundle of ray paths. We derived a condition (condition (23)) which is both necessary and sufficient for the existence of a ray path joining a given satellite position to the ground station, for a given refractivity profile. We then investigated some special cases; the most important of these, and the one we will be exclusively concerned with in the rest of this report, is the case where the ϵ-function is non-negative for all R. In this case we showed that any point above the horizontal plane is joined to the ground station by a ray path. This fact, together with the fact that realistic refractivity profiles actually have the property that $\epsilon(R) \geq 0$ for all R, justifies concentrating our attention on this type of profile.

Procedure for Calculating a Ray Path – We can now outline the procedure for finding a ray path. Suppose we are given an ϵ-function (in the class $\mathcal{E}$) and a satellite position; to determine whether a ray path exists, we test to see whether these satisfy the inequality (23). If the inequality is not satisfied, there is no ray path; if it is satisfied, there is a unique ray path joining the satellite position to the ground station. If $\epsilon(R) \geq 0$ for all R and the satellite position is above the horizon, we know that the inequality will be satisfied. To find the ray path, if it exists, we must first solve equation (22) for α_a; this value of α_a can then be put into equation (20) to obtain the equation of the ray path in the polar coordinates R and ψ. The only step in this procedure which offers any difficulty is the solution of Equation (22); unfortunately this equation cannot be explicitly solved for α_a, and we must resort to a numerical solution or to approximate formulas for α_a in terms of α_t and R_s (which describe the satellite position). The rest of this report will be mainly devoted to finding such approximate solutions and numerical solution methods; we will also try to deduce, directly from Equation (22), as much information as possible concerning the properties of the function $\alpha_a(\alpha_t, R_s)$.

INVESTIGATION OF THE ELEVATION ANGLE ERROR AS A FUNCTION OF SATELLITE POSITION, FOR A SPECIAL CLASS OF REFRACTIVITY PROFILES

The Class $\mathcal{E}^+$ of Epsilon Functions

As was mentioned above, we will limit our attention in the rest of this report to ϵ-functions which are non-negative, and to satellite positions on or above the horizon ($\alpha_t \geq 0$).

For convenience we will denote the class of non-negative, bounded ϵ-functions by $\mathcal{E}^+$; i.e.,

$$\mathcal{E}^+ = \{\epsilon \mid \epsilon \in \mathcal{E},\ \epsilon(R) \geq 0 \text{ for all } R \geq R_e \text{ and } \epsilon(R) \text{ is bounded on } [R_e, \infty)\}.$$

(Note that $\mathcal{E}^+ \subset \mathcal{E}$.) In other words, an ϵ-function is in the class $\mathcal{E}^+$ if and only if it has the following properties:

- $\epsilon(R)$ is continuously differentiable for all $R \geq R_e$;
- $\epsilon(R) \geq 0$ for all $R \geq R_e$;
- There exist a number $\sigma \in (0,1)$ and a positive number B such that

$$\epsilon(R) \leq \text{Min}\left\{(1-\sigma) \cdot \frac{R - R_e}{R_e},\ B\right\} \tag{25}$$

for all $R \geq R_e$.

We will also denote by D the region of the $\alpha_t - R_s$ plane in which we are interested:

$$D = \{(\alpha_t, R_s) \mid 0 \leq \alpha_t < \pi/2 \text{ and } R_s > R_e\}$$

$$= [0,\ \pi/2) \times (R_e, \infty).$$

A point (α_t, R_s) in the $\alpha_t - R_s$ plane will sometimes be denoted by p, for brevity.

The Equation for the Elevation Angle Error

For $\alpha_t \geq 0$ it is possible to write Equation (22) in a more transparent form, by expressing the right-hand side as an integral similar in form to the integral

on the left-hand side. We only need to note that, for $\epsilon(R) = 0$, the ray path will be a straight line and $\alpha_a = \alpha_t$; we can therefore use Equation (21) and Equation (20), with $\epsilon(s) = 0$, $\alpha_a = \alpha_t$ and $R = R_s$, to write

$$\psi_s = \int_{R_e}^{R_s} \frac{\frac{R_e}{R^2} \cos \alpha_t \, dR}{\left[1 - \left(\frac{R_e}{R}\right)^2 \cos^2 \alpha_t\right]^{1/2}} = \cos^{-1}\left(\frac{R_e}{R_s} \cos \alpha_t\right) - \alpha_t .$$

The correctness of this expression can be easily checked by direct evaluation of the integral, using the substitution $(R_e/R) \cos \alpha_t = \cos x$. Thus, for $\alpha_t \geq 0$, Equation (22) takes the form

$$\int_{R_e}^{R_s} \frac{(1+\epsilon)\frac{R_e}{R^2} \cos(\alpha_t + \delta_g) \, dR}{\left[1 - (1+\epsilon)^2 \left(\frac{R_e}{R}\right)^2 \cos^2(\alpha_t + \delta_g)\right]^{1/2}} = \int_{R_e}^{R_s} \frac{\frac{R_e}{R^2} \cos \alpha_t \, dR}{\left[1 - \left(\frac{R_e}{R}\right)^2 \cos^2 \alpha_t\right]^{1/2}}$$

$$= \cos^{-1}\left(\frac{R_e}{R_s} \cos \alpha_t\right) - \alpha_t , \qquad (26)$$

where we have written $\alpha_a = \alpha_t + \delta_g$. The quantity δ_g is the "elevation angle error;" i.e., it is the difference between the satellite's true elevation angle α_t and its apparent elevation angle α_a (the arrival angle) as viewed from the ground station. Equation (26) is the equation for the elevation angle error; the rest of this report will be concerned with solving this equation, and with finding properties of the function $\delta_g(\alpha_t, R_s)$ implicitly defined by it.

It should be stressed that all the operations indicated in Equation (26) can be carried out, at least in theory. From our previous analysis we know that, for any ϵ -function in the class $\mathcal{E}^+$ and for any point $(\alpha_t, R_s) \in D$, the following things are true:

- $[1 - (1+\epsilon)^2 (R_e/R)^2 \cos^2 \alpha_a] \geq 0$ for all $\alpha_a \in [0, \pi/2]$ and for all $R \geq R_e$, with equality holding only for $\alpha_a = 0$ and $R = R_e$.

- The integral on the left-hand side of Equation (26) exists for all $\alpha_a \in [0, \pi/2]$.

- Equation (26) has a unique solution for α_a, which lies in the interval $[0, \pi/2)$; equivalently, it has a unique solution for δ_g, which lies in the interval $[-\alpha_t, \pi/2 - \alpha_t]$.

We therefore do not need to worry about whether the square root is real, or whether the integral converges, and we know that the function $\delta_g(\alpha_t, R_s)$ is defined at every point of the region D in the $\alpha_t - R_s$ plane. Practical difficulties in computation will arise, however, for values of α_t close to zero, because of the fact that the integrand (on the left-hand side of Equation (26)) becomes infinite as $R \to R_e$, for $\alpha_a = 0$. (For realistic refractivity profiles δ_g will be very small, so that α_a will differ only slightly from α_t.)

It will be convenient to write Equation (26) in an abbreviated notation. To this end we define the functions

$$g_{\epsilon,p} : [-\alpha_t, \pi/2 - \alpha_t] \to \mathbb{R}$$

by

$$g_{\epsilon,p}(\beta) = \int_{R_e}^{R_s} \frac{(1+\epsilon)\dfrac{R_e}{R^2}\cos(\alpha_t + \beta)\, dR}{\left[1 - (1+\epsilon)^2 \left(\dfrac{R_e}{R}\right)^2 \cos^2(\alpha_t + \beta)\right]^{1/2}}$$

for each $\epsilon \in \mathcal{E}^+$ and each $p \in D$; the g-functions corresponding to $\epsilon(R) \equiv 0$ will be denoted by $g_{0,p}$. (The symbol $\mathbb{R}$ denotes the real number system.) With this notation we can write Equation (26) as

$$g_{\epsilon,p}(\delta_g) = g_{0,p}(0). \tag{27}$$

We note that:

1. $g_{\epsilon,p}(\beta)$ decreases (strictly) monotonically to the value zero at $\beta = \pi/2 - \alpha_t$; since $\alpha_t < \pi/2$, it follows that $g_{0,p}(0) > 0$.

2. If $\epsilon_1(R) \geq \epsilon_2(R)$ for all R, with equality not always holding, then $g_{\epsilon_1,p}(\beta) > g_{\epsilon_2,p}(\beta)$ for all $\delta_g \in [-\alpha_t, \pi/2 - \alpha_t)$.

3. $g_{\epsilon,p}(0) \geq g_{0,p}(0)$, with equality holding only for $\epsilon(R) \equiv 0$.

All three of these properties hold for any ϵ-function in $\mathcal{E}^+$ and for any point p in D. The third property can either be derived from the second, if we note that $\epsilon \geq 0$ for all R, or it can be deduced from the first property and the fact (which we already know) that Equation (27) has a unique solution.

Some Properties of the Solution

Positivity – We are now ready to derive some properties of the function $\delta_g(\alpha_t, R_s)$. We already know that this function is uniquely defined on the region D. We will now show that it is everywhere positive: for every ϵ-function in the class $\mathcal{E}^+$, except the function $\epsilon(R) \equiv 0$,

$$0 < \delta_g < \frac{\pi}{2} - \alpha_t \tag{28}$$

for all $p \in D$. (Clearly, for $\alpha_t = \pi/2$, $\delta_g = 0$.) This follows from properties 1 and 3 of the g-functions; it is obvious from Figure 4. Since we now know that $\delta_g > 0$, we can limit the domain of definition of the g-functions to the interval $[0, \pi/2 - \alpha_t]$.

An Upper Bound – The quantity $(\pi/2 - \alpha_t)$ in (28) is an upper bound for the δ_g-function; it is a very crude bound, however, and we will now derive a sharper one. If we "solve" Equation (26) for the δ_g in the numerator and delete the δ_g inside the square root, it follows from the positivity of δ_g that

$$0 < \delta_g < \cos^{-1}\left(\frac{\displaystyle\int_{R_e}^{R_s} \frac{\frac{R_e}{R^2}\,dR}{\left[1 - \left(\frac{R_e}{R}\right)^2 \cos^2 \alpha_t\right]^{1/2}}}{\displaystyle\int_{R_e}^{R_s} \frac{(1+\epsilon)\frac{R_e}{R^2}\,dR}{\left[1 - (1+\epsilon)^2 \left(\frac{R_e}{R}\right)^2 \cos^2 \alpha_t\right]^{1/2}}} \cdot \cos \alpha_t \right) - \alpha_t$$

$$= B_1 (\alpha_t, R_s) \tag{29}$$

for any $\epsilon \in \mathcal{E}^+$ which is not identically zero and for all $p \in D$. (Note that the ratio of the two integrals is less than unity, since $\epsilon(R) \geq 0$, and therefore $B_1 > 0$.) Unfortunately the bound B_1 is not too useful, because it requires the computation of

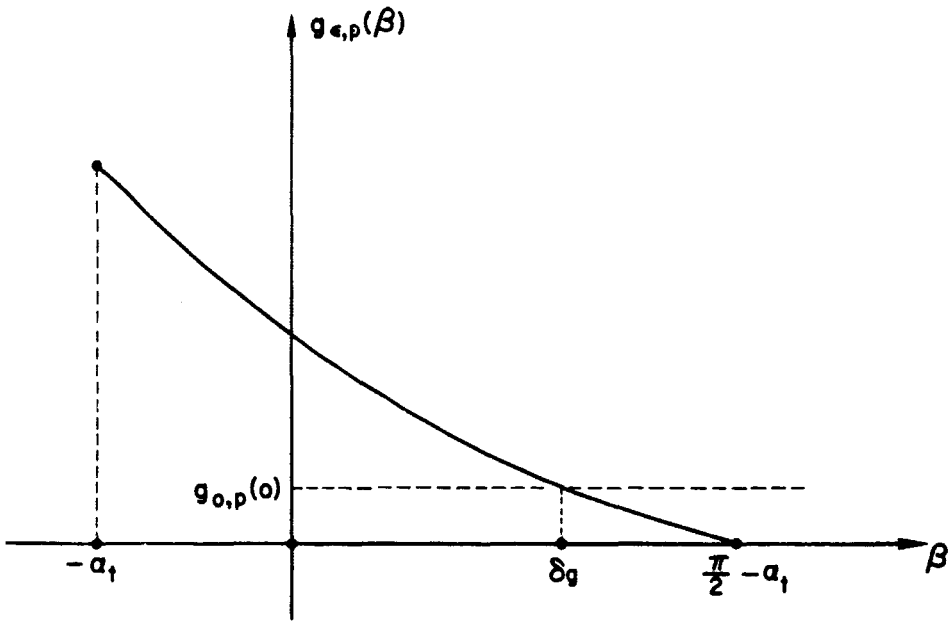


Figure 4–Graph of Equation (27)

an integral and because it is not very sharp, although it is much sharper than the bound ($\pi/2 - \alpha_t$). A computer program (which will be described in a subsequent report) was used to evaluate the bound B_1 for the fairly realistic refractivity profile described in the appendix; this profile will be referred to as the "Sperry profile," because it was taken from a report of the Sperry Gyroscope Company. The results indicate that B_1 is quite sharp for large values of α_t, but becomes progressively worse as α_t decreases; for small values of α_t it is very crude. For the Sperry profile, and for a signal frequency of 140 megahertz, B_1 exceeds the true value of δ_g by only a few percent for $\alpha_t > 75°$; for $\alpha_t = 40°$ the error is already about 100%, and for $\alpha_t = 10°$ B_1 is about ten times larger than the true value of δ_g. (The calculations were done for a satellite height of 1,000 kilometers.) An attempt was made to make B_1 more useful by replacing $\epsilon(R)$ by a simple majorant function consisting of straight-line segments (of the form given by Equation (25)) and explicitly evaluating the integral. The attempt failed on two counts: the resulting formula was undesirably complex (although the majorant function consisted only of two line segments, an initial sloped line followed by a horizontal line), and the sharpness of the bound was severely degraded, the new upper bound being almost 200% in error even for $\alpha_t = 87°$. For the Sperry profile, which is a fairly realistic one, the true value of δ_g varies from 3.9 milliradians at $\alpha_t = 5°$ to 0.026 milliradians at 87°, for a satellite height of 1,000 kilometers and a signal frequency of 140 megahertz. As in all the other calculations made with the computer program, it was found that δ_g decreases monotonically with α_t.

Another Upper Bound – A different upper bound for δ_g can be easily deduced from Equation (26); it is rather crude, but it has the advantage of great simplicity. To find it we just note that, formally, the left-hand integrand in Equation (26) differs from the right-hand integrand in only one way: the appearance of the quantity $Q = (1 + \epsilon) \cos(\alpha_t + \delta_g)$ in place of $\cos \alpha_t$. Since the left-hand integrand increases with increasing Q, for any value of R, it is clear that Equation (26) cannot possibly hold if the minimum value of Q is greater than $\cos \alpha_t$ or the maximum value of Q is less than $\cos \alpha_t$. In other words we must have

$$(1 + \min \epsilon_s) \cdot \cos(\alpha_t + \delta_g) < \cos \alpha_t$$

and

$$(1 + \max \epsilon_s) \cdot \cos(\alpha_t + \delta_g) > \cos \alpha_t$$

for any ϵ-function which is not identically zero on the interval $[R_e, R_s]$, where min ϵ_s and max ϵ_s are the infimum and supremum, respectively, of $\epsilon(R)$ on the interval $[R_e, R_s]$. Since min $\epsilon_s = 0$, the first inequality just implies that $\delta_g > 0$, which we already know. If we solve the second inequality for δ_g, taking into account the fact that $\alpha_t + \delta_g < \pi/2$ (see Equation (28)), we obtain the upper bound

$$\delta_g < \cos^{-1}\left(\frac{\cos \alpha_t}{1 + \max \epsilon_s}\right) - \alpha_t ,$$

where the arc cosine is taken in the first quadrant. This bound holds for any point $(\alpha_t, R_s) \in D$, and for any $\epsilon \in \mathcal{E}^+$ which is not identically zero on the interval $[R_e, R_s]$. We can get rid of the dependence on R_s, at the cost of weakening the bound slightly, by replacing max ϵ_s by the possibly larger number ϵ_m, the supremum of $\epsilon(R)$ on the interval $[R_e, \infty)$. We then have

$$0 < \delta_g < \cos^{-1}\left(\frac{\cos \alpha_t}{1 + \epsilon_m}\right) - \alpha_t \qquad (30)$$

$$= B_2(\alpha_t, R_s),$$

for any ϵ-function in $\mathcal{E}^+$ which is not identically zero on $[R_e, \infty)$, and for any point $p \in D$.

The bound B_2 was tested numerically for the Sperry profile. The results indicate that the ratio of B_2 to the true value of δ_g is remarkably constant over a large range of elevation angles; for the profile used, and for a signal frequency of 140 megahertz, this ratio remained between 1.79 and 1.83 for α_t between 20° and 87°, for a satellite height of 1,000 kilometers. For $\alpha_t < 20°$ the ratio increased; at 5° it was 2.75. It would appear from these results that B_2 is useful as an easily computed estimate of δ_g. Further, the relative constancy of the ratio B_2/δ_g indicates that B_2 might be used to form approximate formulas for δ_g by multiplying B_2 by a polynomial of low degree in α_t (and possibly R_s also); such a formula might be very useful over a given range of elevation angles, for a specific refractivity profile. For the profile used in the tests, it would be sufficient to divide B_2 by 1.81 to obtain a formula accurate to about 1% over the entire

Table 1
Computed Values of Elevation Angle Error, in Milliradians

Elevation Angle (deg)	Height (km)									
	100	200	300	400	500	1,000	5,000	10,000	100,000	1,000,000
1						8.00				
2						6.33				
3						5.23				
4						4.46				
5					3.89	3.89		3.84		3.83
10						2.42				
15					1.88	1.74		1.55		1.51
20						1.33				
25					1.21	1.06		0.863		0.830
30						0.856				
35					0.845	0.706		0.548		0.524
40						0.588				
45	0.290	0.301	0.435	0.599	0.604	0.492	0.385	0.370	0.356	0.354
50						0.412				
55					0.428	0.343		0.253		0.242
60						0.283				
65					0.287	0.228		0.166		0.159
70						0.178				
75					0.165	0.131		0.0947		0.0906
80						0.0860				
85					0.0540	0.0427		0.0308		0.0295
86						0.0341				
87						0.0256				
88						0.0170				
89						0.0085				

range α_t = 20° to 87°, for the satellite height of 1,000 kilometers and the given signal frequency. It would not be difficult to account for the variation of δ_g with R_s, since δ_g varies much more strongly with α_t than with R_s. To illustrate this, table 1 shows the values of δ_g (obtained with the computer program) for a range of values of elevation angle (α_t) and height ($R_s - R_e$), for the Sperry refractivity profile and a signal frequency of 140 megahertz. Some of the data in table 1 are plotted in Figure 5. (Many of the positions in the table are blank because those values of δ_g were not computed in the test runs.)

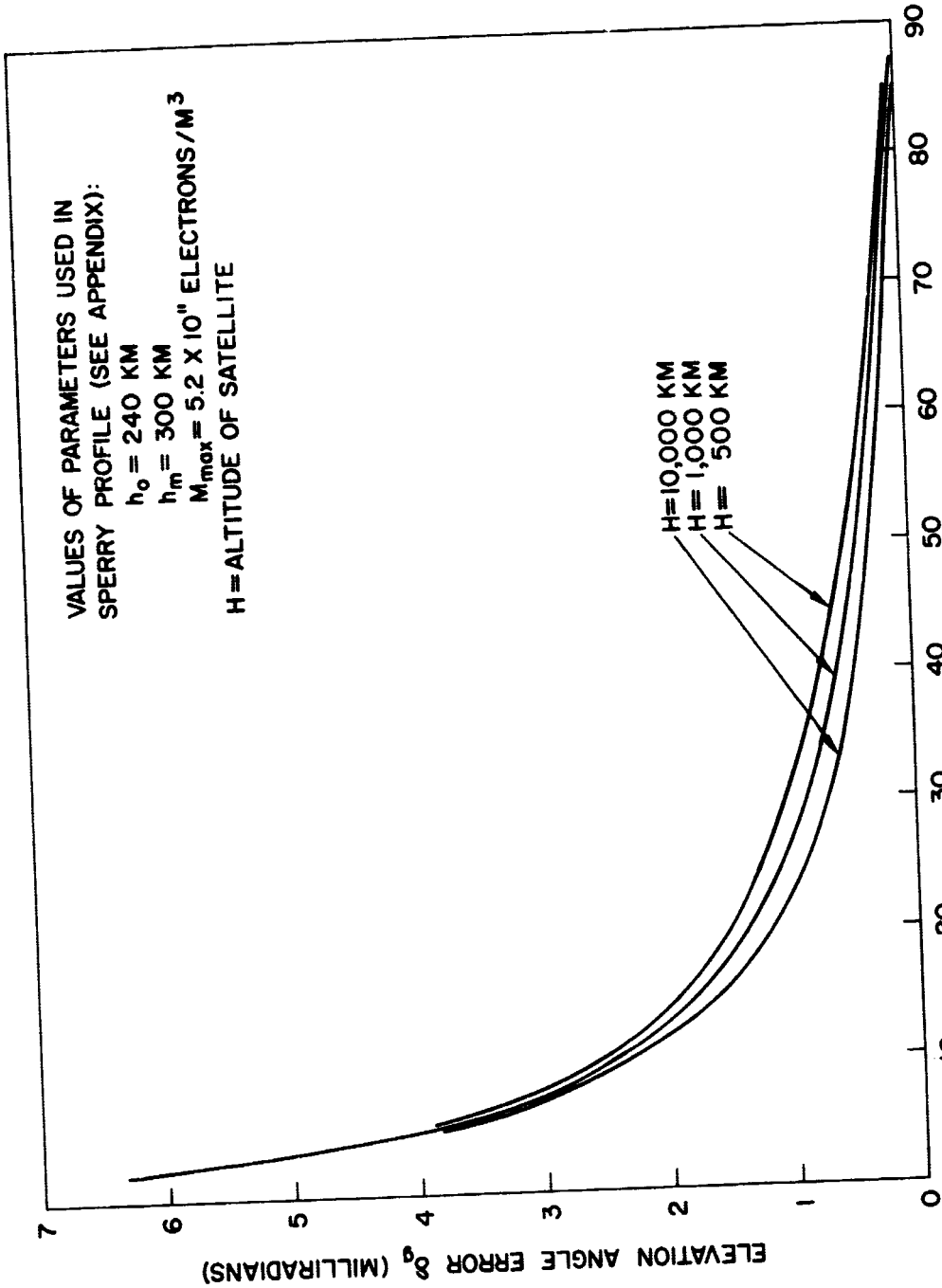


Figure 5–Elevation Angle Error as a Function of Elevation Angle, for Various Altitudes

A Lower Bound – So far we have shown that $\delta_g > 0$ and we have derived upper bounds for it. It is of some interest to also derive a lower bound, other than zero. To accomplish this we start again with Equation (26), and rewrite it in the form

$$\int_{R_e}^{R_s} \frac{\frac{R_e}{R^2}\cos\alpha_t \, dR}{\left\{\left[1 - \left(\frac{R_e}{R}\right)^2 \cos^2\alpha_t\right] + \left[\frac{\cos^2\alpha_t}{(1+\epsilon)^2\cos^2(\alpha_t + \delta_g)} - 1\right]\right\}^{1/2}}$$

$$= \int_{R_e}^{R_s} \frac{\frac{R_e}{R^2}\cos\alpha_t \, dR}{\left[1 - \left(\frac{R_e}{R}\right)^2 \cos^2\alpha_t\right]^{1/2}}.$$

Since $(\alpha_t + \delta_g) < \pi/2$ and $\alpha_t < \pi/2$ for all points $p \in D$, we see that neither of the cosines is zero; the above equation, whose left-hand integrand was derived from that of Equation (26) mainly by multiplying and dividing by the two cosines, is therefore equivalent to Equation (26). We now apply the mean-value theorem to the function $x^{-1/2}$ to obtain

$$\left\{\left[1 - \left(\frac{R_e}{R}\right)^2 \cos^2\alpha_t\right] + \left[\frac{\cos^2\alpha_t}{(1+\epsilon)^2\cos^2(\alpha_t + \delta_g)} - 1\right]\right\}^{-1/2}$$

$$= \left[1 - \left(\frac{R_e}{R}\right)^2 \cos^2\alpha_t\right]^{-1/2}$$

$$- \frac{\frac{1}{2}\left[\frac{\cos^2\alpha_t}{(1+\epsilon)^2\cos^2(\alpha_t + \delta_g)} - 1\right]}{\left\{\left[1 - \left(\frac{R_e}{R}\right)^2 \cos^2\alpha_t\right] + \theta\left[\frac{\cos^2\alpha_t}{(1+\epsilon)^2\cos^2(\alpha_t + \delta_g)} - 1\right]\right\}^{3/2}}$$

where $0 < \theta < 1$ and θ depends on R. This is valid only for $\alpha_t > 0$; if $\alpha_t = 0$, the quantity $[1 - (R_e/R)^2 \cos^2\alpha_t]$ becomes zero for $R = R_e$. If we substitute this into our rewritten equation and simplify, we are left with the equation

$$\int_{R_e}^{R_s} \frac{\frac{R_e}{R^2}\cos\alpha_t \cdot \left[\frac{\cos^2\alpha_t}{(1+\epsilon)^2\cos^2(\alpha_t+\delta_g)} - 1\right] dR}{\left\{\left[1-\left(\frac{R_e}{R}\right)^2\cos^2\alpha_t\right] + \theta\left[\frac{\cos^2\alpha_t}{(1+\epsilon)^2\cos^2(\alpha_t+\delta_g)} - 1\right]\right\}^{3/2}} \geq 0.$$

We will denote the bracketed quantity in the numerator by Q, for convenience. From the relation (30) it follows that

$$\frac{\cos^2\alpha_t}{(1+\epsilon)^2\cos^2(\alpha_t+\delta_g)} < \frac{(1+\epsilon_m)^2}{(1+\epsilon)^2}.$$

Since the right-hand side of this inequality is unity for any value of R for which ϵ takes its maximum value, we see that the left-hand side is sometimes less than unity, whence $Q < 0$. On the other hand, for any value of R for which $\epsilon = 0$, for example for $R = R_e$, the left-hand side is clearly greater than unity (since $\delta_g > 0$), whence $Q > 0$. In other words, Q is sometimes positive and sometimes negative. Now suppose we set $\theta(R) \equiv 0$; clearly, the integrand will then become larger in magnitude for those values of R for which Q is positive, and it will become smaller in magnitude for those values of R for which Q is negative. It follows that

$$\int_{R_e}^{R_s} \frac{\frac{R_e}{R^2}\cos\alpha_t \cdot \left[\frac{\cos^2\alpha_t}{(1+\epsilon)^2\cos^2(\alpha_t+\delta_g)} - 1\right] dR}{\left[1-\left(\frac{R_e}{R}\right)^2\cos^2\alpha_t\right]^{3/2}} > 0.$$

If we solve this inequality for δ_g we get our lower bound:

$$\delta_g > \cos^{-1}\left(\frac{\cos\alpha_t}{\sqrt{1+\Delta}}\right) - \alpha_t$$

$$= L(\alpha_t, R_s) > 0, \tag{31}$$

where

$$\Delta = \frac{\displaystyle\int_{R_e}^{R_s} \frac{dR}{R^2 \left[1 - \left(\frac{R_e}{R}\right)^2 \cos^2 \alpha_t\right]^{3/2}}}{\displaystyle\int_{R_e}^{R_s} \frac{dR}{R^2 (1+\epsilon)^2 \left[1 - \left(\frac{R_e}{R}\right)^2 \cos^2 \alpha_t\right]^{3/2}}} - 1.$$

It is clear that $\Delta > 0$, and therefore $L > 0$. This bound holds for any ϵ -function in $\mathcal{E}^+$ which is not identically zero on the interval $[R_e, R_s]$ and for any point $p \in D$ for which $\alpha_t > 0$. It was tested numerically with the computer program; the results indicate that it is quite accurate. For the Sperry profile, at a signal frequency of 140 megahertz and a satellite height of 1,000 kilometers, the lower bound L differs from the true value of δ_g by only about 0.7% at $\alpha_t = 5°$. For larger values of α_t the error steadily decreases. At $\alpha_t = 15°$ it is 0.2%, at 25° only 0.1%, and for $\alpha_t > 45°$ it is less than 0.06%. It thus would appear that the lower bound L, as given by Equation (31), is a rather accurate approximation for the elevation angle error δ_g.

Partial Derivatives – So far we have proved that the solution function $\delta_g(\alpha_t, R_s)$ is uniquely defined on the region

$$D = \{(\alpha_t, R_s) \mid 0 \leq \alpha_t < \pi/2 \text{ and } R_s > R_e\} = [0, \pi/2) \times (R_e, \infty)$$

of the $\alpha_t - R_s$ plane, for any ϵ -function in the class $\mathcal{E}^+$, and we have derived some upper and lower bounds for it. We will now try to get more detailed information about the behavior of the solution function. Some information can be obtained with the help of the implicit function theorem of analysis. Suppose we write Equation (26) in the form

$$F(\delta_g ; \alpha_t, R_s) = 0$$

by putting all terms on the left-hand side of the equation. The function F, which maps a region E of the 3-dimensional $\delta_g - \alpha_t - R_s$ space into the reals, is given by

$$F(\delta_g ; \alpha_t , R_s) = \int_{R_e}^{R_s} \frac{(1+\epsilon)\frac{R_e}{R^2}\cos(\alpha_t + \delta_g) \cdot dR}{\left[1 - (1+\epsilon)^2 \left(\frac{R_e}{R}\right)^2 \cos^2(\alpha_t + \delta_g)\right]^{1/2}}$$

$$- \int_{R_e}^{R_s} \frac{\frac{R_e}{R^2}\cos\alpha_t \cdot dR}{\left[1 - \left(\frac{R_e}{R}\right)^2 \cos^2\alpha_t\right]^{1/2}},$$

for $(\delta_g , \alpha_t , R_s) \in E$. The region E is given by:

$$E = \{(\delta_g, \alpha_t , R_s) \in R^3 \mid \alpha_t \in [0, \pi/2) \text{ and } R_s \in (R_e , \infty) \text{ and } \delta_g \in [0, \pi/2 - \alpha_t)\}.$$

For easy visualization this region is shown in Figure 6; it is a semi-infinite prism, with a cross-section in the form of an isosceles right triangle. The

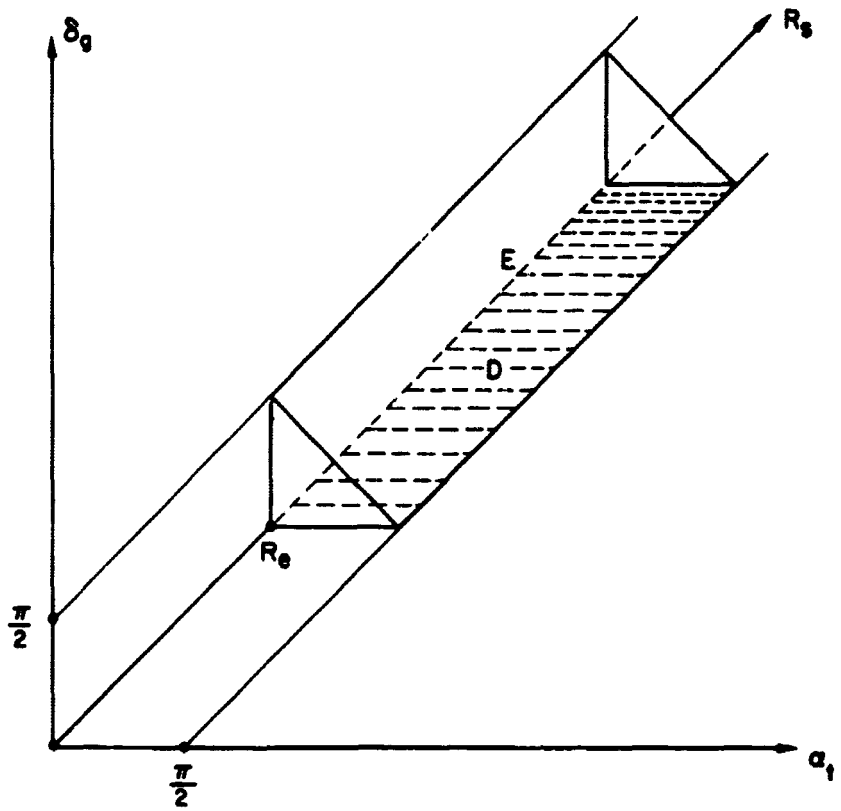


Figure 6–The Region E

region E does not include the entire surface of the prism; it includes only the interiors of the vertical and horizontal faces plus that part of the R_s -axis for which $R_s > R_e$. The base of the prism (i.e., the horizontal face) is just the region D. We will denote the interiors of E and D by $\mathring{E}$ and $\mathring{D}$, respectively. It is clear from our previous results that the graph of the solution function $\delta_g(\alpha_t, R_s)$ is a surface which lies in $\mathring{E}$ for any ϵ -function in $\mathcal{E}^+$ which is not identically zero. (If $\epsilon \equiv 0$, the surface just becomes the region D itself, since $\delta_g = 0$ for all points of D in this case.) We could have defined the function F at all points of the slab $[0, \pi/2) \times (R_e, \infty) \times (-\infty, \infty)$; but then F would be periodic in δ_g. Since we already know that the solution surface lies in the region E, there is no point in complicating the problem by defining F over a larger domain.

We are now in a position to apply the implicit function theorem; we will state this theorem as it applies to our particular case. Suppose the following hypotheses are satisfied:

1. The partial derivatives $\partial F/\partial \delta_g$, $\partial F/\partial \alpha_t$, $\partial F/\partial R_s$ exist and are continuous on the domain $\mathring{E}$.

2. There exists a point $(\delta_0, \alpha_0, R_0) \in \mathring{E}$ for which $F(\delta_0, \alpha_0, R_0) = 0$.

3. $$\left.\frac{\partial F}{\partial \delta_g}\right|_{\delta_0, \alpha_0, R_0} \neq 0.$$

The theorem then states that there exists an open neighborhood $W \subset \mathring{D}$ of the point $(\alpha_0, R_0) \in \mathring{D}$ and a unique function $h: W \to \mathbb{R}$, such that:

1. $h(\alpha_0, R_0) = \delta_0$.

2. $F(h(\alpha_t, R_s), \alpha_t, R_s) = 0$ for all points $(\alpha_t, R_s) \in W$.

3. $\dfrac{\partial h}{\partial \alpha_t}$, $\dfrac{\partial h}{\partial R_s}$ exist and are continuous on W.

4. $$\left.\frac{\partial h}{\partial \alpha_t}\right|_{(\alpha_t, R_s)} = \frac{-\left.\dfrac{\partial F}{\partial \alpha_t}\right|_{(h(\alpha_t, R_s), \alpha_t, R_s)}}{\left.\dfrac{\partial F}{\partial \delta_g}\right|_{(h(\alpha_t, R_s), \alpha_t, R_s)}}$$

and

$$\left.\frac{\partial h}{\partial R_s}\right|_{(\alpha_t, R_s)} = \frac{-\left.\frac{\partial F}{\partial R_s}\right|_{(h(\alpha_t, R_s), \alpha_t, R_s)}}{\left.\frac{\partial F}{\partial \delta_g}\right|_{(h(\alpha_t, R_s), \alpha_t, R_s)}}$$

for all points $(\alpha_t, R_s) \in W$.

In other words the theorem states that, if we know a point $(\delta_0, \alpha_0, R_0)$ for which the equation $F(\delta_g, \alpha_t, R_s) = 0$ is satisfied, then this equation can be uniquely solved for δ_g in terms of α_t and R_s for all points (α_t, R_s) sufficiently close to (α_0, R_0). It is important to note that the theorem guarantees the existence of a continuously differentiable and unique solution function only locally; even if hypotheses 2 and 3 are satisfied for every point in $\mathring{D}$, it does not follow that there exists a global solution unique on $\mathring{D}$, although there exist locally unique solutions about each point of $\mathring{D}$. However, we know from our previous work that, for each point $(\alpha_t, R_s) \in \mathring{D}$, there exists a unique δ_g such that $F(\delta_g, \alpha_t, R_s) = 0$. With this additional information we can conclude that the solution function h, or $\delta_g(\alpha_t, R_s)$, is defined uniquely everywhere on $\mathring{D}$ and has continuous partial derivatives on $\mathring{D}$. It remains only to show that hypotheses 1 and 3 are satisfied. It is easily seen that the three partial derivatives exist and are continuous on $\mathring{E}$; they are given by the following formulas:

$$\frac{\partial F}{\partial \delta_g} = -\sin(\alpha_t + \delta_g) \cdot \int_{R_e}^{R_s} \frac{(1+\epsilon)\frac{R_e}{R^2}\,dR}{\left[1 - (1+\epsilon)^2\left(\frac{R_e}{R}\right)^2 \cos^2(\alpha_t + \delta_g)\right]^{3/2}}$$

$$\frac{\partial F}{\partial \alpha_t} = -\sin(\alpha_t + \delta_g) \cdot \int_{R_e}^{R_s} \frac{(1+\epsilon)\frac{R_e}{R^2}\,dR}{\left[1 - (1+\epsilon)^2\left(\frac{R_e}{R}\right)^2 \cos^2(\alpha_t + \delta_g)\right]^{3/2}}$$

$$+ \sin\alpha_t \cdot \int_{R_e}^{R_s} \frac{\frac{R_e}{R^2}\,dR}{\left[1 - \left(\frac{R_e}{R}\right)^2 \cos^2\alpha_t\right]^{3/2}}$$

$$\frac{\partial F}{\partial R_s} = \frac{[1+\epsilon(R_s)]\frac{R_e}{R_s^2}\cos(\alpha_t+\delta_g)}{\left[1-[1+\epsilon(R_s)]^2\left(\frac{R_e}{R_s}\right)^2\cos^2(\alpha_t+\delta_g)\right]^{1/2}} - \frac{\frac{R_e}{R_s^2}\cos\alpha_t}{\left[1-\left(\frac{R_e}{R_s}\right)^2\cos^2\alpha_t\right]^{1/2}}$$

It is clear from the formula for $\partial F/\partial\delta_g$ that this derivative is non-zero (in fact, it is negative) at all points of $\mathring{E}$. We are therefore assured that the solution function $\delta_g(\alpha_t, R_s)$, which we already know exists and is unique, has continuous partial derivatives on $\mathring{D}$; by conclusion 4 of the theorem, these partials are given by the following formulas:

$$\frac{\partial\delta_g}{\partial\alpha_t} = -1 + \frac{\sin\alpha_t}{\sin(\alpha_t+\delta_g)} \cdot \frac{\displaystyle\int_{R_e}^{R_s} \frac{dR}{R^2\left[1-\left(\frac{R_e}{R}\right)^2\cos^2\alpha_t\right]^{3/2}}}{\displaystyle\int_{R_e}^{R_s} \frac{(1+\epsilon)\,dR}{R^2\left[1-(1+\epsilon)^2\left(\frac{R_e}{R}\right)^2\cos^2(\alpha_t+\delta_g)\right]^{3/2}}} \tag{32}$$

$$\frac{\partial\delta_g}{\partial R_s} = \frac{\dfrac{[1+\epsilon(R_s)]\frac{R_e}{R_s^2}\cos(\alpha_t+\delta_g)}{\left[1-[1+\epsilon(R_s)]^2\left(\frac{R_e}{R_s}\right)^2\cos^2(\alpha_t+\delta_g)\right]^{1/2}} - \dfrac{\frac{R_e}{R_s^2}\cos\alpha_t}{\left[1-\left(\frac{R_e}{R_s}\right)^2\cos^2\alpha_t\right]^{1/2}}}{\sin(\alpha_t+\delta_g)\cdot\displaystyle\int_{R_e}^{R_s}\frac{(1+\epsilon)\frac{R_e}{R^2}\,dR}{\left[1-(1+\epsilon)^2\left(\frac{R_e}{R}\right)^2\cos^2(\alpha_t+\delta_g)\right]^{3/2}}} \tag{33}$$

In these formulas, the value of δ_g must be taken as $\delta_g(\alpha_t, R_s)$, the actual solution value of δ_g at the point (α_t, R_s), in order to evaluate the partials at the point (α_t, R_s). Although we know that the solution for δ_g exists and is unique, we do not know what it is. In this sense the above formulas are incomplete; to render them complete we would have to know the solution function. In spite of this, we can still get some information from these formulas, if we use what we know about the solution function.

It should be pointed out that much more information is obtainable from the implicit function theorem than we have stated above. For example, it is easy to see that the function F is differentiable any number of times; i.e., the partial derivatives of all orders exist and are continuous. The theorem (in its more complete form) then allows us to conclude that the solution function $\delta_g(\alpha_t, R_s)$ has continuous partial derivatives of all orders. We will only use the first-order partials, however, because enough information is obtainable from them to get a general idea of the behavior of the solution function and because the higher-order partials are given by excessively complicated formulas.

We will examine the partial $\partial\delta_g/\partial R_s$ first, since it turns out to be the easier of the two to handle. If we look closely at the two terms in the numerator, we notice that the first term differs from the second in only one way: the quantity $Q = [1 + \epsilon(R_s)] \cdot \cos(\alpha_t + \delta_g)$ appears in place of $\cos\alpha_t$; furthermore, the first term increases monotonically with Q. It follows that the algebraic sign of $\partial\delta_g/\partial R_s$ is just the sign of $Q - \cos\alpha_t$; i.e.,

$$\text{sign}(\partial\delta_g/\partial R_s) = \text{sign}\{[1 + \epsilon(R_s)] \cdot \cos[\alpha_t + \delta_g(\alpha_t, R_s)] - \cos\alpha_t\}$$

$$= \text{sign}\{k_{\alpha_t}(R_s)\}.$$

From the upper bound for δ_g given by the inequality (30) it follows that

$$(1 + \epsilon_m) \cdot \cos(\alpha_t + \delta_g) - \cos\alpha_t > 0;$$

from the definition of ϵ_m as the supremum of $\epsilon(R)$ on the interval $[R_e, \infty)$ and from the continuity of ϵ and δ_g as functions of R_s it follows that, for any given value of α_t, there exists at least one point in (R_e, ∞), and an open interval containing that point, such that the function $k(R_s)$ is positive in that interval. In other words,

$$\frac{\partial\delta_g}{\partial R_s} > 0 \text{ for } R_s \in \Omega^+, \tag{34}$$

where Ω^+ is some open set contained in the interval (R_e, ∞). (Note that the set Ω^+ depends on the angle α_t.) We will now show that $\Omega^+ \neq (R_e, \infty)$. To see this, we just recall that $\epsilon(R_e) = 0$; since $\delta_g > 0$, it follows that $k_{\alpha_t}(R_e) < 0$ for all α_t. It follows from the continuity of the k-function that $k_{\alpha_t}(R_s)$ is negative in some open neighborhood of the point R_e, and possibly on other open intervals as well. In other words,

$$\frac{\partial\delta_g}{\partial R_s} < 0 \text{ for } R_s \in \Omega^-, \tag{35}$$

where Ω^- is some open set containing a neighborhood of the point R_e and contained in (R_e, ∞); note that Ω^- depends on α_t. From all this we see that, for each $\alpha_t \in (0, \pi/2)$, the partial derivative $\partial\delta_g/\partial R_s$ is sometimes positive and sometimes negative, and therefore also sometimes zero. This behavior is evident in the data shown in table 1; the table also indicates another property, which is clear from Equation (33):

$$\left|\frac{\partial \delta_g}{\partial R_s}\right| \to 0 \text{ as } R_s \to \infty, \tag{36}$$

for any $\alpha_t \in (0, \pi/2)$.

We now turn to the partial $\partial\delta_g/\partial\alpha_t$. From values of δ_g obtained with the computer program it is seen that, for the Sperry profile, δ_g decreases monotonically with α_t for any given value of R_s (e.g., see Figure 5). Since there is nothing very special about the Sperry profile, it seems reasonable to suppose that this monotonicity might hold more generally. We will investigate this question.

It is evident from Equation (32) that a sufficient (...not necessary) condition for the derivative to be negative is that the ratio of the two integrals be less than unity:

$$\int_{R_e}^{R_s} \frac{(1+\epsilon)\,dR}{R^2\left[1-(1+\epsilon)^2\left(\frac{R_e}{R}\right)^2\cos^2(\alpha_t+\delta_g)\right]^{3/2}} > \int_{R_e}^{R_s} \frac{dR}{R^2\left[1-\left(\frac{R_c}{R}\right)^2\cos^2\alpha_t\right]^{3/2}}$$

$$= \frac{1}{R_e \sin\alpha_t}\left[1-\frac{\frac{R_e}{R_s}\sin\alpha_t}{\sqrt{1-\left(\frac{R_e}{R_s}\right)^2\cos^2\alpha_t}}\right] \tag{37}$$

$$\equiv \Phi(\alpha_t).$$

(The integral was evaluated by using the substitution $(R_e/R)\cos\alpha_t = \cos\theta$.) We can get rid of the unknown δ_g and at the same time construct conditions which imply (37) by replacing δ_g in (37) by any of its upper bounds. Our crudest and

simplest upper bound for δ_g is the one given by (28): $\delta_g < \pi/2 - \alpha_t$. The corresponding condition is

$$\int_{R_e}^{R_s} \frac{(1+\epsilon)\,dR}{R^2} > \int_{R_e}^{R_s} \frac{dR}{R^2\left[1-\left(\frac{R_e}{R}\right)^2\cos^2\alpha_t\right]^{3/2}} = \Phi(\alpha_t). \qquad (38)$$

It is evident that Φ is a monotonically decreasing function of α_t. Further, $\Phi(\alpha_t)$ becomes infinite as $\alpha_t \to 0$, and

$$\Phi\left(\frac{\pi}{2}\right) = \int_{R_e}^{R_s} \frac{dR}{R^2} < \int_{R_e}^{R_s} \frac{(1+\epsilon)\,dR}{R^2}.$$

We conclude that, for any ϵ-function in $\mathcal{E}^+$, there exists a value of α_t for each value of R_s, say $\alpha^-(R_s)$, such that $\partial\delta_g/\partial\alpha_t < 0$ in the interval $(\alpha^-(R_s), \pi/2)$ (and probably in a larger interval), where $\alpha^-(R_s)$ is the value of α_t for which (38) becomes an equality.

The value of α^- given by Equation (38) would probably be close to 90°; we can get a smaller value of α^- by using a sharper upper bound for δ_g. If we use the bound B_2 given by Equation (30), we get the condition

$$\int_{R_e}^{R_s} \frac{(1+\epsilon)\,dR}{R^2\left[1-\frac{(1+\epsilon)^2}{(1+\epsilon_m)^2}\left(\frac{R_e}{R}\right)^2\cos^2\alpha_t\right]^{3/2}} > \Phi(\alpha_t). \qquad (39)$$

We reach the same conclusion as before concerning the existence of α^-, because the left side of (39), evaluated at $\alpha_t = \pi/2$, is clearly greater than $\Phi(\pi/2)$. The value of α^- computed from (39) will in general be smaller than that computed from (38).

More detailed results can be obtained, but they are probably not worth the considerable effort required to find them. For example, after a very lengthy

calculation (which will not be given here), the writer found that $\alpha^{-}(R_s) \to 0$ as $R_s \to \infty$; however, all attempts to prove that $\alpha^{-}(R_s) = 0$ (i.e., that the derivative $\partial \delta_g / \partial \alpha_t < 0$ for all α_t and all R_s) failed.

We will now go on to a different topic - the transformation of Equation (26) to a form which is more convenient for numerical calculation.

A Transformed Version of the Elevation Angle Error Equation

Earlier we wrote Equation (26) in the abbreviated form (27), by using the functions $g_{\epsilon,p}(\beta)$. It would be useful to know something about the derivatives of these functions. We find by direct calculation that

$$g'_{\epsilon,p}(\beta) = -R_e \sin(\alpha_t + \beta) \cdot \int_{R_e}^{R_s} \frac{(1+\epsilon)\, dR}{R^2 \left[1 - (1+\epsilon)^2 \left(\frac{R_e}{R}\right)^2 \cos^2(\alpha_t + \beta)\right]^{3/2}}$$

and

$$g''_{\epsilon,p}(\beta) = -R_e \cos(\alpha_t + \beta) \cdot \int_{R_e}^{R_s} \frac{(1+\epsilon)\, dR}{R^2 \left[1 - (1+\epsilon)^2 \left(\frac{R_e}{R}\right)^2 \cos^2(\alpha_t + \beta)\right]^{3/2}}$$

$$+ 3R_e \cos(\alpha_t + \beta) \cdot \sin^2(\alpha_t + \beta) \cdot \int_{R_e}^{R_s} \frac{(1+\epsilon)^3 \left(\frac{R_e}{R}\right)^2 dR}{R^2 \left[1 - (1+\epsilon)^2 \left(\frac{R_e}{R}\right)^2 \cos^2(\alpha_t + \beta)\right]^{5/2}}.$$

The first derivative is clearly negative, but the behavior of the second derivative is not very clear. Higher derivatives will be even more complicated and difficult to analyze. It seems that the functions $g_{\epsilon,p}(\beta)$ show undesirably complex behavior. Fortunately, it is possible to avoid this difficulty by making an appropriate transformation of the dependent variable in Equation (26). To this end we define the functions

$$\varphi_{\alpha_t} : [0,\ \pi/2 - \alpha_t) \to [1,\ \infty)$$

by

$$\varphi_{\alpha_t} : \beta \vdash \frac{\cos^2 \alpha_t}{\cos^2 (\alpha_t + \beta)} = \zeta ,$$

for any value of α_t in the interval $[0, \pi/2)$. In terms of the variable ζ, we can define the functions

$$f_{\epsilon, p} : [1, \infty) \to \mathbb{R}$$

by

$$f_{\epsilon, p}(\zeta) = \int_{R_e}^{R_s} \frac{(1 + \epsilon) \frac{R_e}{R^2} dR}{\left[\zeta - (1 + \epsilon)^2 \left(\frac{R_e}{R} \right)^2 \cos^2 \alpha_t \right]^{1/2}} .$$

It is convenient also to change the variable of integration to ρ, where

$$\rho = \frac{R_e}{R} . \tag{40}$$

If we define the function $\tilde{\epsilon}(\rho)$ by

$$\tilde{\epsilon}(\rho) = \epsilon (R) = \epsilon (R_e / \rho) , \tag{41}$$

we can write the f-functions in the form

$$f_{\epsilon,p}(\zeta) = \int_{\rho_s}^{1} \frac{d\rho}{\left[\frac{\zeta}{(1+\tilde{\epsilon})^2} - \rho^2 \cos^2 \alpha_t\right]^{1/2}}, \tag{42}$$

where $\rho_s = R_e / R_s$. Clearly $\rho_s \in (0, 1)$.

We can now write Equation (26) in the form

$$\int_{\rho_s}^{1} \frac{d\rho}{\left[\frac{\xi}{(1+\tilde{\epsilon})^2} - \rho^2 \cos^2 \alpha_t\right]^{1/2}} = \int_{\rho_s}^{1} \frac{d\rho}{[1 - \rho^2 \cos^2 \alpha_t]^{1/2}} = \frac{\cos^{-1}(\rho_s \cos \alpha_t) - \alpha_t}{\cos \alpha_t}, \tag{43}$$

or

$$f_{\epsilon,p}(\xi) = f_{0,p}(1), \tag{44}$$

where $f_{0,p}$ is the f-function corresponding to $\epsilon(R) \equiv 0$ and

$$\xi = \frac{\cos^2 \alpha_t}{\cos^2 (\alpha_t + \delta_g)}. \tag{45}$$

We also write the inverse relation for reference:

$$\delta_g = \cos^{-1}\left(\frac{\cos \alpha_t}{\xi^{1/2}}\right) - \alpha_t . \tag{46}$$

To avoid confusion, it should be noted that Equation (45) defines ξ, not as a function of δ_g, but as a function of α_t and R_s (since δ_g depends on α_t and R_s). However, from the fact that Φ_{α_t} is a monotonically increasing function, it is clear that any upper (or lower) bound for δ_g generates an upper (or lower) bound for ξ, and vice versa. For example, the bounds for ξ, corresponding to the bounds for δ_g given by (30), are

$$1 < \xi < (1 + \epsilon_m)^2 . \tag{47}$$

These bounds can also be read off from Equation (43) by inspection, if we recall that $\tilde{\epsilon}(\rho) \geq 0$ for all ρ. (Note that this upper bound does not depend on α_t.)

From (47) we conclude that the quantity $\xi / [1 + \tilde{\epsilon}(\rho)]^2$ is less than unity for some values of ρ and greater than unity for other values of ρ. To see this we note first that, for any value of ρ for which $\tilde{\epsilon}(\rho) = 0$ (e.g., for $\rho = 1$),

$$\frac{\xi}{[1 + \tilde{\epsilon}(\rho)]^2} = \frac{\xi}{1} > 1 .$$

Secondly, we note that for any value of ρ for which $\tilde{\epsilon}(\rho) = \epsilon_m$,

$$\frac{\xi}{[1 + \tilde{\epsilon}(\rho)]^2} = \frac{\xi}{(1 + \epsilon_m)^2} < 1 .$$

Of course it is also evident from inspection of Equation (43) that this must be so; if the above quantity were either always less than or always greater than unity Equation (43) could not be satisfied.

Equation (43) is our new form for the elevation angle error equation. We will now show that it has more convenient properties than does Equation (26). If we look at the derivatives of the f-functions, we easily find that the n-th derivative is given by the simple formula

(48)

$$f^{(n)}_{\epsilon,\rho}(\zeta) = (-1)^n \cdot \frac{1 \cdot 3 \cdot \cdots \cdot (2n-1)}{2^n} \int_{\rho_s}^{1} \frac{d\rho}{(1 + \tilde{\epsilon})^{2n} \left[\frac{r}{(1 + \tilde{\epsilon})^2} - \rho^2 \cos^2 \alpha_t \right]^{n+1/2}} ,$$

for n = 1, 2, 3, ····· . This very clear formula contrasts sharply with the obscure behavior of the derivatives of the g-functions. In particular, we note that the first derivative is always negative and the second derivative is always positive. We will use this fact later, when we describe how to solve equation (43) numerically by the Newton-Raphson method.

If we denote the integral on the right-hand side of Equation (48) by $I_n(\zeta)$, where n = 0, 1, 2, 3, ···, it is easy to see that for any ϵ-function in $\mathcal{E}^+$:

- $\left[\frac{\zeta}{(1+\tilde{\epsilon})^2} - \rho^2 \cos^2 \alpha_t\right] \geq 0$ for all $\alpha_t \in [0, \pi/2)$, all $\zeta \in [1, \infty)$ and all $\rho \leq 1$, with equality holding only for $\alpha_t = 0$, $\zeta = 1$ and $\rho = 1$;
- $f_{\epsilon,p}(\zeta)$ (= $I_0(\zeta)$) exists (i.e., is finite) for all $\zeta \in [1, \infty)$, for any point $p \in \mathring{D}$;
- $I_n(\zeta)$, and therefore also $f^{(n)}_{\epsilon,p}(\zeta)$, exists (i.e., is finite) for all $\zeta \in [1, \infty)$ and for n = 1, 2, 3, ···, for any point $p \in \mathring{D}$;
- $I_n(1)$, and therefore also $f^{(n)}_{\epsilon,p}(1)$, where n = 1, 2, 3, ···, does not exist (i.e., is infinite) for $\alpha_t = 0$.

The first three properties follow directly from our previous results (see the discussion following Equation (26)). To see that the fourth property holds, we proceed as follows:

$$I_n(1) = \int_{\rho_s}^{1} \frac{d\rho}{(1+\tilde{\epsilon})^{2n}\left[\frac{1}{(1+\tilde{\epsilon})^2} - \rho^2 \cos^2 \alpha_t\right]^{n+1/2}}$$

$$> \frac{1}{(1+\epsilon_m)^{2n}} \int_{\rho_s}^{1} \frac{\rho\, d\rho}{[1-\rho^2 \cos^2 \alpha_t]^{1+1/2}}$$

$$= \frac{1}{(1+\epsilon_m)^{2n} \cos^2 \alpha_t}\left[\frac{1}{\sin \alpha_t} - \frac{1}{\sqrt{1-\rho_s^2 \cos^2 \alpha_t}}\right].$$

Clearly, the last expression becomes infinite as $\alpha_t \to 0$. The fact that the derivatives become infinite (at $\zeta = 1$) as $\alpha_t \to 0$ causes difficulty in the numerical solution of Equation (43), for small values of α_t.

It is clear that Equation (43) is equivalent to Equation (26), for all ϵ-functions in $\mathcal{E}^+$ and for all points $p \in D$; thus the solution of one of these equations can be obtained directly from the solution of the other by using Equation (45) or Equation (46). From this point on we will use either Equation (43) or Equation (26), whichever happens to be convenient at the time.

A Method for Solving the Transformed Equation Numerically

In terms of the f-functions (defined by Equation (42)), the transformed elevation angle error equation is given by Equation (44); we will write that equation in the simplified form

$$f(\xi) = f_0 \tag{49}$$

for convenience. A plot of this equation will have the general appearance shown in Figure 7.

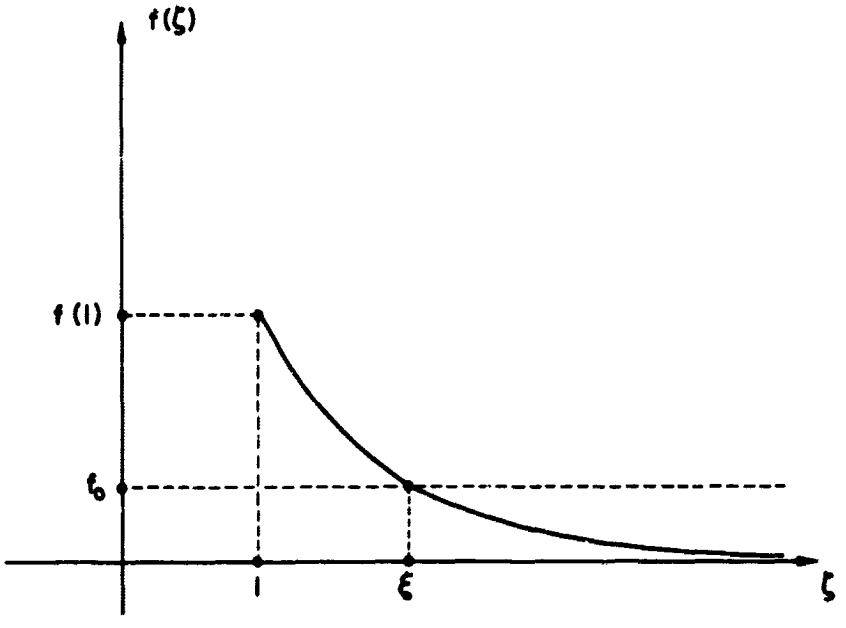


Figure 7–Graph of Equation (49)

Of the various methods available for solving such an equation, we will describe only one: the Newton-Raphson method. The reason for this choice is the fact that $f'(\zeta) < 0$ and $f''(\zeta) > 0$, for all $\zeta \in [1, \infty)$; this guarantees that the Newton-Raphson process will converge monotonically upwards to the solution value ξ, provided that the starting point is smaller than ξ. We will now describe the process in detail.

The Newton-Raphson method uses a succession of approximating tangents to the curve $y = f(\zeta)$. The first step is to pass a tangent line through the point $(\xi_0, f(\xi_0))$, where $\xi_0 < \xi$; the equation of this line is

$$y = f'(\xi_0) \cdot (\zeta - \xi_0) + f(\xi_0).$$

The value of ζ for which this line crosses the horizontal line $y = f_0$ gives us the first iterate, ξ_1; we calculate ξ_1 by setting the above expression equal to f_0 and solving for ζ. The result is

$$\xi_1 = \xi_0 + \frac{f(\xi_0) - f_0}{-f'(\xi_0)}.$$

To get the second iterate, ξ_2, we pass a tangent line through the point $(\xi_1, f(\xi_1))$ and calculate its intersection with the line $y = f_0$; we find that

$$\xi_2 = \xi_1 + \frac{f(\xi_1) - f_0}{-f'(\xi_1)}.$$

Clearly we have the general recursive relation

$$\xi_n = \xi_{n-1} + \frac{f(\xi_{n-1}) - f_0}{-f'(\xi_{n-1})}, \tag{50}$$

for $n = 1, 2, 3, \cdots$.

To start the computation we need a value for ξ_0 which is less than the true solution ξ. The simplest choice is to take $\xi_0 = 1$; this will always work in

principle, except when $\alpha_t = 0$. In practice, computational difficulties will arise even when α_t is just close to zero, because of the boundless growth (as $\alpha_t \to 0$) of $f'(1)$ and of the integrand of $f(1)$ at $\rho = 1$. Somewhat better results can be obtained by taking ξ_0 to be a lower bound for ξ, other than unity. For example the computer program (which will be described in a companion report) takes ξ_0 to be the lower bound for ξ corresponding to the lower bound for δ_g given by (31).

The iterative process is stopped when some convergence criterion is satisfied. Instead of putting a condition on the iterates ξ_n, the computer program puts a condition directly on the corresponding values of δ_g, which we will denote by δ_n, because δ_g is the quantity we are really interested in. The criterion used is that the relative error $(\delta_{n+1} - \delta_n)/\delta_{n+1}$ be less than a preset value; when this condition is first satisfied the iterative process stops.

In general, the Newton-Raphson method converges rapidly after the iterates get close to the true solution. But during the first few iterations some other process, such as the secant method, may be more efficient, and it may be advantageous to use such a method to start the computation. In our case, however, the iterates converge monotonically upwards toward the solution, and it seems simpler to start right off with the Newton-Raphson method (which is what is done in the computer program).

To complete this discussion of the numerical solution of Equation (49) we will prove by induction that the iterates ξ_n do in fact converge monotonically upward to the solution ξ, and we will show that a sufficient condition for this to happen is that $f'(\zeta) < 0$ and $f''(\zeta) > 0$ for all ζ.

Suppose then that $\xi_{n-1} < \xi$, and let ξ_n be given by Equation (50). If we could prove that

$$\xi_{n-1} < \xi_n < \xi , \tag{51}$$

then we could conclude that the sequence (ξ_n) is monotonically increasing and is bounded above by ξ. (The induction is started by noting that $\xi_0 < \xi$, so that the premise is in fact true for $n = 1$.) To prove the first inequality in (51) we use the fact that $f(\zeta)$ is monotonically decreasing. From our premise that $\xi_{n-1} < \xi$ it then follows that $f(\xi_{n-1}) > f(\xi) = f_0$. With this result we can conclude directly from Equation (50) that $\xi_n > \xi_{n-1}$, if we also use the fact that $f'(\xi_{n-1}) < 0$. To prove the second inequality in (51) we use Taylor's Theorem to write

$$f(\xi_n) = f(\xi_{n-1}) + f'(\xi_{n-1}) \cdot (\xi_n - \xi_{n-1}) + \frac{1}{2} f''(\xi^*) \cdot (\xi_n - \xi_{n-1})^2 ,$$

where $\xi_{n-1} < \xi^* < \xi_n$. If we use Equation (50) to replace the first two terms on the right by f_0 we obtain the relation

$$f(\xi_n) = f_0 + \frac{1}{2} f''(\xi^*) \cdot (\xi_n - \xi_{n-1})^2 .$$

It follows that $f(\xi_n) > f_0 = f(\xi)$, because $f''(\xi^*) > 0$. From the monotonicity of $f(\zeta)$ we conclude that $\xi_n < \xi$, which proves (51).

There is one last thing to prove - we must show that $\xi_n \to \xi$. We have just shown that the sequence (ξ_n) is monotonically increasing and is bounded above by ξ; the sequence must therefore converge to some number $\xi' \leq \xi$. To show that $\xi' = \xi$ we observe that, since (ξ_n) is convergent, $(\xi_n - \xi_{n-1}) \to 0$. From Equation (50) we then conclude that $f(\xi_n) \to f_0$. Since it is clear that the inverse function f^{-1} (which exists by virtue of the monotonicity of f) is continuous, we conclude that $\xi_n \to f^{-1}(f_0) = \xi$.

Some General Formulas Giving Approximate Solutions to the Elevation Angle Error Equation

General Form – We approach the problem of finding approximate solutions in the following way. First we obtain, by any means available, an approximation to the solution of Equation (43) in the form

$$\xi \cong 1 + \Delta , \tag{52}$$

where $0 < \Delta \ll 1$. The fact that Δ is small follows from the bound given by (47) for ξ, namely $\xi < (1 + \epsilon_m)^2$; since ϵ_m will be small (of the order of 5×10^{-3}) for a realistic refractivity profile, we see that

$$\Delta \cong \xi - 1 < 2\epsilon_m + \epsilon_m^2 \ll 1 .$$

Next we write the corresponding approximation for δ_g by using Equation (46):

$$\delta_g \cong \cos^{-1}\left(\frac{\cos \alpha_t}{\sqrt{1+\Delta}}\right) - \alpha_t \,. \tag{53}$$

If α_t is not close to zero we can simplify this formula by using a differential approximation. We regard the expression on the right as a function of Δ, say $A(\Delta)$, and consider α_t as a parameter, ignoring the fact that Δ actually depends on α_t. By using Taylor's theorem we can write

$$A(\Delta) = A(0) + A'(\eta) \cdot \Delta$$

$$= \frac{\frac{1}{2}\cos \alpha_t}{\sqrt{1 - \frac{\cos^2 \alpha_t}{1+\eta}} \cdot (1+\eta)^{3/2}} \cdot \Delta \,,$$

where $0 < \eta < \Delta$. If α_t is not too small we can neglect η without affecting the last expression too much, because η is small; in that case we may write

$$\delta_g \cong \frac{1}{2} \cdot \frac{\cos \alpha_t}{\sin \alpha_t} \cdot \Delta \,. \tag{54}$$

We can get some feeling for the range of validity of this approximation by writing

$$1 - \frac{\cos^2 \alpha_t}{1+\eta} \cong 1 - (1-\eta)\cos^2 \alpha_t = \sin^2 \alpha_t + \eta \cos^2 \alpha_t \,,$$

so that

$$A(\Delta) \cong \frac{1}{2} \cdot \frac{\cos \alpha_t}{\sin \alpha_t} \cdot \frac{\Delta}{\sqrt{1 + \eta \cot^2 \alpha_t}} \,.$$

The neglect of η will be justified if $\eta \cot^2 \alpha_t \ll 1$; in this case we can write

$$A(\Delta) \cong \frac{1}{2} \cdot \frac{\cos \alpha_t}{\sin \alpha_t} \cdot \left(1 - \frac{1}{2} \eta \cot^2 \alpha_t\right) \cdot \Delta .$$

From this expression we see that Equation (54) will be accurate to about 10% or better if we have $1/2\ \eta \cot^2 \alpha_t < 0.1$. From the comments following (52) we see that this last condition will be satisfied if $\epsilon_m \cot^2 \alpha_t < 0.1$. Since $\epsilon_m \cong 5 \times 10^{-3}$ for a realistic refractivity profile, the condition becomes $\tan \alpha_t > \sqrt{5} \cdot 10^{-1}$, or roughly $\alpha_t > 12°$. In other words, the value of δ_g given by Equation (54) will differ from that given by Equation (53) by no more than 10% or so, for $\alpha_t > 12°$.

Before we go on to obtain explicit expressions for Δ, some comments should be made about the problem of determining the accuracy of the corresponding formulas for δ_g. When the various formulas were first derived, error terms were included and attempts were made to use them to evaluate the accuracy of the formulas. This approach turned out to be intractable. The trouble is that the error terms are quite complex, involving complicated definite integrals. Attempts to simplify them by using estimates or bounds invariably result in expressions which give far too crude an estimate of the error. The original error terms could of course be evaluated numerically, but it is simpler and more direct just to compare the approximate formulas with accurate numerical solutions of the elevation angle error equation, and this is the course we will follow.

First Newton-Raphson Iterate – Perhaps the most obvious candidate for an approximate general formula is the first iterate in the Newton-Raphson numerical solution scheme which was described earlier. If we take $\xi_0 = 1$, it is seen from Equation (50) that the first iterate is given by

$$\xi_1 = 1 + \frac{f(1) - f_0}{-f'(1)} ;$$

in other words,

$$\Delta = \frac{f(1) - f_0}{-f'(1)}$$

$$= \frac{\displaystyle\int_{\rho_s}^{1} \frac{(1+\tilde{\epsilon})\, d\rho}{[1-(1+\tilde{\epsilon})^2 \rho^2 \cos^2 \alpha_t]^{1/2}} - \int_{\rho_s}^{1} \frac{d\rho}{[1-\rho^2 \cos^2 \alpha_t]^{1/2}}}{\displaystyle\frac{1}{2}\int_{\rho_s}^{1} \frac{d\rho}{(1+\tilde{\epsilon})^2 \left[\frac{1}{(1+\tilde{\epsilon})^2} - \rho^2 \cos^2 \alpha_t\right]^{3/2}}} . \tag{55}$$

It is not practicable to go further than the first iterate; higher iterates would involve integrals within integrals.

A Differential Approximation – Aside from using the first Newton-Raphson iterate (or the first iterates in other numerical schemes), the most obvious way of obtaining approximate formulas is to use differential approximations. To proceed in a systematic way, we use the notation of Equation (49):

$$f(\xi) = \int_{\rho_s}^{1} \frac{d\rho}{\left[\frac{\xi}{(1+\tilde{\epsilon})^2} - \rho^2 \cos^2 \alpha_t\right]^{1/2}} . \tag{56}$$

We also consider the integrand as a function of ξ:

$$r(\xi) = \frac{1}{\left[\frac{\xi}{(1+\tilde{\epsilon})^2} - \rho^2 \cos^2 \alpha_t\right]^{1/2}} . \tag{57}$$

The elevation angle error equation can then be written as

$$f(\xi) = \int_{\rho_s}^{1} r(\xi)\, d\rho = f_0 . \tag{58}$$

The various possibilities for differential approximations are now apparent. The most obvious are the first-order Taylor expansions of $f(\xi)$ and $r(\xi)$ about the point $\xi = 1$:

$$f(\xi) \cong f(1) + f'(1) \cdot (\xi - 1)$$

and

$$r(\xi) \cong r(1) + r'(1) \cdot (\xi - 1).$$

It is easy to see that if either of these expressions is put into Equation (58) and the resulting equation is solved for ξ, we just get back the first Newton-Raphson iterate. (It is also geometrically obvious that this is true.)

Another possibility is to write $\gamma = \xi/(1+\tilde{\epsilon})^2$, and to regard r as a function of γ, ignoring the fact that γ actually depends on ρ. Since we know (from the bounds for ξ given by (47)) that γ is close to unity, we can expand r to the first order in γ about the point $\gamma = 1$; it follows that

$$r(\xi) \cong \frac{1}{[1 - \rho^2 \cos^2 \alpha_t]^{1/2}} - \frac{1}{2} \cdot \frac{\left[\frac{\xi}{(1+\tilde{\epsilon})^2} - 1\right]}{[1 - \rho^2 \cos^2 \alpha_t]^{3/2}}.$$

If we put this expression for $r(\xi)$ into Equation (58) and solve for $\Delta(\cong \xi - 1)$, we obtain

$$\Delta = \frac{\displaystyle\int_{\rho_s}^{1} \frac{d\rho}{[1 - \rho^2 \cos^2 \alpha_t]^{3/2}}}{\displaystyle\int_{\rho_s}^{1} \frac{d\rho}{(1+\tilde{\epsilon})^2 [1 - \rho^2 \cos^2 \alpha_t]^{3/2}}} - 1. \tag{59}$$

It is interesting to note that this is precisely the lower bound given by (31). The smallness of Δ can be better displayed by writing the expression as a simple fraction:

$$\Delta = \frac{\displaystyle\int_{\rho_s}^{1} \frac{2\,\tilde{\epsilon} + \tilde{\epsilon}^2}{(1+\tilde{\epsilon})^2\,[1-\rho^2\cos^2\alpha_t]^{3/2}}\,d\rho}{\displaystyle\int_{\rho_s}^{1} \frac{1}{(1+\tilde{\epsilon})^2\,[1-\rho^2\cos^2\alpha_t]^{3/2}}\,d\rho}.$$

Another Differential Approximation – We can obtain yet another approximate formula by regarding r as a function of the two variables ξ and $\tilde{\epsilon}$, ignoring the fact that $\tilde{\epsilon}$ depends on ρ. Since ξ is close to unity and $\tilde{\epsilon}$ is small, we can expand r to the first order in ξ and $\tilde{\epsilon}$ about the point $\xi = 1$, $\tilde{\epsilon} = 0$; it follows that

$$r(\xi) \cong \frac{1}{[1-\rho^2\cos^2\alpha_t]^{1/2}} - \frac{\frac{1}{2}(\xi-1)}{[1-\rho^2\cos^2\alpha_t]^{3/2}} + \frac{\tilde{\epsilon}}{[1-\rho^2\cos^2\alpha_t]^{3/2}}.$$

This approximation should be fairly accurate (for all ρ) if α_t is not close to zero (if $\alpha_t = 0$ and $\rho = 1$, the square root vanishes). If we put this expression for $r(\xi)$ into Equation (58) and solve for $\Delta(\cong \xi - 1)$, we obtain

$$\Delta = \frac{2\displaystyle\int_{\rho_s}^{1} \frac{\tilde{\epsilon}\,d\rho}{[1-\rho^2\cos^2\alpha_t]^{3/2}}}{\displaystyle\int_{\rho_s}^{1} \frac{d\rho}{[1-\rho^2\cos^2\alpha_t]^{3/2}}} \qquad (60)$$

$$= \frac{2\sin\alpha_t}{1-\dfrac{\rho_s\sin\alpha_t}{\sqrt{1-\rho_s^2\cos^2\alpha_t}}} \cdot \int_{\rho_s}^{1} \frac{\tilde{\epsilon}\,d\rho}{[1-\rho^2\cos^2\alpha_t]^{3/2}},$$

where the integral was evaluated by using the substitution $\rho \cos \alpha_t = \cos \theta$. Note that this formula is not valid if α_t is close to zero. In fact, for $\alpha_t = 0$ formula (60) gives $\Delta = 0$. We know that this is wrong, because it implies that $\xi = 1$ for $\alpha_t = 0$, which is impossible by virtue of the inequality (47).

A Simple Formula for the Elevation Angle Error – Any one of the three expressions for Δ given by Equations (55), (59) or (60) could be used together with Equation (53) or Equation (54) to obtain an approximate formula for δ_g. It is evident, however, that the simplest of the three expressions for Δ is given by Equation (60). Further, the other two expressions for Δ can be reduced to it by using approximations based on the smallness of $\tilde{\epsilon}(\rho)$. For these reasons it would seem that this formula might be the best one to use. If we combine it with Equation (54) the factors $\sin \alpha_t$ cancel each other and we obtain the following formula for δ_g :

$$\delta_g \cong \frac{\cos \alpha_t}{\left[1 - \dfrac{\rho_s \sin \alpha_t}{\sqrt{1 - \rho_s^2 \cos^2 \alpha_t}}\right]} \cdot \int_{\rho_s}^{1} \frac{\tilde{\epsilon}\, d\rho}{[1 - \rho^2 \cos^2 \alpha_t]^{3/2}} . \qquad (61)$$

In spite of the fact that neither Equation (60) nor Equation (54) is valid for α_t close to zero, numerical comparisons with the correct solution show that the formula given by (61) is remarkably accurate down to small angles. (This is probably due in part to the fact that the $\sin \alpha_t$ factors cancelled out.) Table 2 shows the percentage error of formula (61) for various elevation angles, for the Sperry profile and for an altitude of 1,000 kilometers and a signal frequency of 140 megahertz. The % error is defined by

$$\% \text{ error} = \frac{\text{value of } \delta_g \text{ computed from (61)} - \text{correct value of } \delta_g}{\text{correct value of } \delta_g} \times 100.$$

The data in table 2 show that formula (61) gives slightly high values for δ_g, at least for this particular case. The formula is seen to be accurate to within 1% down to elevation angles of less than 10°; even for $\alpha_t = 1°$, it is only 8% too high. This accuracy is considerably better than one might have expected from the way in which the formula was derived.

Table 2
Accuracy of the Formula for δ_g

Elevation Angle (degrees)	% Error
89	0.05
87	0.05
85	0.05
80	0.05
75	0.05
70	0.05
65	0.05
60	0.05
55	0.05
50	0.05
45	0.05
40	0.05
35	0.05
30	0.05
25	0.06
20	0.09
15	0.2
10	0.4
5	2.
4	2.
3	3.
2	5.
1	8.

Numerical tests were also run on formulas for δ_g obtained by using the other two formulas for Δ. The results show slightly greater accuracy, but not enough (in the writer's opinion) to compensate for the greater complexity of those formulas.

Before closing this section, we should show that the integral in formula (61) always exists. We use the fact that $\epsilon(R) \in \mathcal{E}^+$, so that $\epsilon(R) \leq (1-\sigma)(R - R_e)/R_e$ for all $R \geq R_e$, by the condition (25). In terms of the variable $\rho = R_e/R$ this condition becomes $\tilde{\epsilon}(\rho) \leq (1-\sigma)(1/\rho - 1)$, for all $\rho < 1$. We use this condition to obtain a bound for the integral:

$$\int_{\rho_s}^{1} \frac{\tilde{\epsilon}\, d\rho}{[1-\rho^2\cos^2\alpha_t]^{3/2}} \leq (1-\sigma) \int_{\rho_s}^{1} \frac{\left(\frac{1}{\rho}-1\right) d\rho}{[1-\rho^2]^{3/2}}$$

for all $\alpha_t \in [0, \pi/2]$

$$= (1-\sigma) \int_{\rho_s}^{1} \frac{d\rho}{\rho(1-\rho)^{1/2}(1+\rho)^{3/2}}$$

$$< \frac{(1-\sigma)}{\rho_s(1+\rho_s)^{3/2}} \int_{\rho_s}^{1} \frac{d\rho}{(1-\rho)^{1/2}}$$

$$= \frac{2(1-\sigma)(1-\rho_s)^{1/2}}{\rho_s(1+\rho_s)^{3/2}}$$

$$< \infty.$$

That is, the integral is finite.

Some Comments on Methods of Finding Approximate Solutions for Particular Cases

In the preceding section we were concerned with finding general formulas which are approximate solutions of the elevation angle error equation. In this section we will comment on the problem of finding simple approximate solutions for specific refractivity profiles, with the aim of finding ways of minimizing the amount of computation needed to find δ_g. Two possible approaches to this problem will be described briefly.

Numerical Approach – The basic idea is to solve the elevation angle error equation for various values of α_t and R_s (for some given signal frequency or frequency band) and then to fit curves to the computed points. Since δ_g varies less with R_s than with α_t, it would probably be feasible to obtain a number of

simple formulas for δ_g in terms of α_t, for various values of R_s. It might also be feasible to fit a surface (i.e., a function of two variables) to the computed points. In either case, the aim would be to obtain simple, easily evaluated formulas for δ_g.

Analytic Approach – The idea here is to develop simple formulas for δ_g by using approximations for the integrals in general formulas such as formula (61). For example, one might use the trapezoid rule, or Simpson's rule, to evaluate the definite integral in formula (61). Alternatively, any one of the functions $\tilde{\epsilon}(\rho)$, $\epsilon(R)$ or $n(R)$ might be approximated by straight-line segments or by parabolic arcs (the approximation of $n(R)$ by straight-line segments would amount to assuming a spherically stratified atmosphere). In any case, the resulting expression for δ_g would have the form of a sum of easily evaluated terms. If the number of terms in such a formula is too large for easy calculation, it might be feasible to approximate the refractivity profile (or the ϵ-function) by expressions which are more complex than straight-line segments or parabolic arcs, but which are still simple enough so that the resulting integrals can be evaluated. In this way a formula might be developed which has a smaller number of terms.

A word of caution is in order here. Before replacing the refractivity profile by simpler functions, one should make sure that the corresponding approximate ϵ-function is non-negative and satisfies condition (25). It is not hard to see that the various integrals will exist for any piece-wise continuous ϵ-function which has these properties. (For example, the proof which was given for the existence of the integral in formula (61) applies to any ϵ-function which has these two properties, even if it is not continuous.) It follows that these integrals can be approximated by using such an approximate ϵ-function, even though the laws of geometric optics, and therefore our entire derivation, do not strictly hold (for radio waves in the ionosphere) if the refractivity profile is discontinuous.

Quantities Related to the Arrival Angle

There are two quantities related to the arrival angle which are widely used: the arrival angle at the satellite and the total bending angle.

The Arrival Angle at the Satellite – This angle can easily be computed by using the spherical form of Snell's law, as given by Equation (13). To carry out the calculation we refer to Figure 8, which should be compared with Figure 1. In Figure 8 the vector $\vec{s}_s$ is a unit vector tangent to the ray path at the satellite and all angles are taken positive in the directions indicated. We want to compute the angle δ_s between the slant path and the vector $\vec{s}_s$. In general it seems that δ_s may be either positive or negative, but in either case we have $\theta_s = \gamma + \delta_s$.

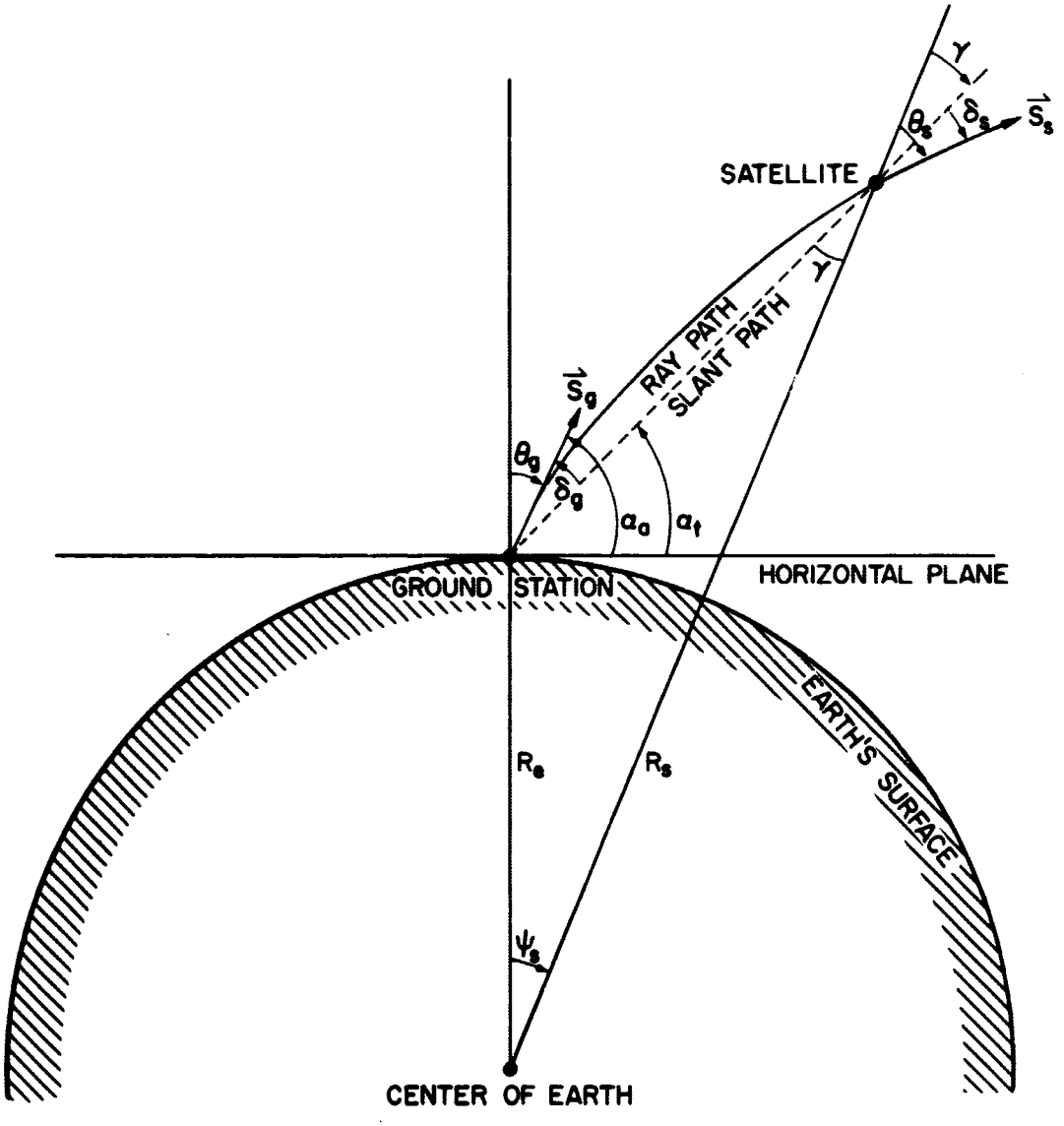


Figure 8–Diagram for the Calculation of δ_s

Recall also that, for any refractivity profile in the class $\mathcal{E}$, $0<\theta_s < \pi/2$ (see the section entitled "Some General Properties of Ray Paths"); this fact is essential for the following derivation.

From equation (13) it follows that

$$n_s R_s \sin \theta_s = n_g R_e \cos (\alpha_t + \delta_g),$$

where n_s is the value of the index of refraction n at the satellite. If we solve this equation for δ_s $(= \theta_s - \gamma)$ we obtain

$$\delta_s = \sin^{-1}\left[\frac{n_g R_e}{n_s R_s}\cos(\alpha_t + \delta_g)\right] - \gamma,$$

where the arc sine, which is the angle θ_s, is taken in the first quadrant. We can find the angle γ by using the law of sines; we see that

$$\frac{\sin\gamma}{R_e} = \frac{\cos\alpha_t}{R_s},$$

or

$$\gamma = \sin^{-1}\left(\frac{R_e}{R_s}\cdot\cos\alpha_t\right),$$

where the arc sine is again taken in the first quadrant, since it is clear that $0 < \gamma < \pi/2$. We thus obtain the following formula for δ_s:

$$\delta_s = \sin^{-1}\left[\frac{n_g R_e}{n_s R_s}\cdot\cos(\alpha_t + \delta_g)\right] - \sin^{-1}\left(\frac{R_e}{R_s}\cos\alpha_t\right), \qquad (62)$$

where both arc sines are taken in the first quadrant. Since δ_g depends on the satellite position and on the refractivity profile, we see that formula (62) gives δ_s in terms of those same quantities. Note that $\delta_s \to 0$ as $R_s \to \infty$.

The Total Bending Angle – The total bending angle can be defined as the angle through which the vector $\vec{s}_g$ must be rotated to coincide with the vector $\vec{s}_s$, with the angle taken positive in the clockwise direction. If we denote the total bending angle by β_s, it is easy to see from Figure 8 that

$$\beta_s = \psi_s + \theta_s - \theta_g$$

$$= [\pi - \gamma - (\pi/2 + \alpha_t)] + (\gamma + \delta_s) - (\pi/2 - \alpha_t - \delta_g),$$

or

$$\beta_s = \delta_g + \delta_s . \tag{63}$$

Note that $\beta_s \to \delta_g$ as $R_s \to \infty$.

A Formula for the Total Bending Angle in Terms of Satellite Altitude and the Arrival Angle – Equations (62) and (63) give δ_s and β_s in terms of R_s (= R_e plus the satellite altitude), the arrival angle α_a (= $\alpha_t + \delta_g$) and the true elevation angle α_t. It is of some interest that we can derive a formula for β_s which depends only on R_s and α_a. If α_a could be accurately measured such a formula would be very useful, since $\delta_g \approx \beta_s$ for very distant satellites. Unfortunately angular quantities cannot be measured very precisely by radar equipment, so the formula would be useful mainly for visual observations. In fact the formula we will derive will be specialized (in the next section) to the classical astronomical formula for tropospheric bending. In contrast to the case for β_s, it does not seem possible to obtain formulas for δ_g or δ_s which depend only on R_s and α_a.

To derive the formula we first define the partial bending angle β at a given point along the ray path to be the angle through which the vector $\vec{s}_g$ must be rotated to coincide with the vector $\vec{s}$ at the given point (see Equation (11) for the definition of $\vec{s}$), with the angle taken positive in the clockwise direction. If the given point is at the satellite position, then of course β is just β_s. From Figure 1 we see that

$$\beta = (\psi + \theta) - \theta_g .$$

From our analysis of the ray paths we know that R is a strictly monotonically increasing function of ψ, for a given ray path. It follows that the angles β, ψ and θ are (single-valued) functions of R, for a given ray path. We know also that the ray path is differentiable; this can be seen from Equation (20), if we recall that any ϵ-function in the class $\mathcal{E}$ is differentiable and therefore continuous. It follows that β, ψ and θ are differentiable functions of R. We may therefore differentiate the above equation with respect to R:

$$\frac{d\beta}{dR} = \frac{d\psi}{dR} + \frac{d\theta}{dR} .$$

We can obtain an expression for $d\theta/dR$ by differentiating Equation (13):

$$\frac{d\theta}{dR} = \frac{-n_g R_e \cos\alpha_a \cdot \left(R\frac{dn}{dR} + n\right)}{n^2 R^2 \left[1 - \left(\frac{n_g}{n}\right)^2 \left(\frac{R_e}{R}\right)^2 \cdot \cos^2\alpha_a\right]^{1/2}}.$$

If we insert this into the above equation for $d\beta/dR$ together with the expression for $d\psi/dR$ given by Equation (19) some terms cancel out and we obtain the simple formula

$$\frac{d\beta}{dR} = \frac{-n_g R_e \cos\alpha_a \cdot \frac{dn}{dR}}{n^2 R \left[1 - \left(\frac{n_g}{n}\right)^2 \left(\frac{R_e}{R}\right)^2 \cdot \cos^2\alpha_a\right]^{1/2}}.$$

(In using Equation (19) recall that $1 + \epsilon = n_g/n$.) This formula gives $d\beta/dR$ explicitly as a function of R, since we assume that the function n(R) is known. If we integrate it from R_e to R_s we obtain our formula for β_s :

$$\beta_s = -n_g R_e \cos\alpha_a \cdot \int_{R_e}^{R_s} \frac{\frac{dn}{dR}dR}{n^2 R \left[1 - \left(\frac{n_g}{n}\right)^2 \left(\frac{R_e}{R}\right)^2 \cos^2\alpha_a\right]^{1/2}} \qquad (64)$$

This formula depends explicitly on R_s and α_a, but not on α_t. It is valid for any profile n(R) whose corresponding ϵ-function is in the class $\mathcal{E}$, for all $R_s > R_e$ and for $0 \leq \alpha_a \leq \pi/2$. For distant satellites the formula can be slightly simplified by noting that $dn/dR = 0$ above the sensible atmosphere. If we denote the value of R at the altitude where the atmosphere may be considered to end by

R_T, then we may write $dn/dR = 0$ for $R > R_T$. If the satellite is extremely distant, $\delta_g \approx \beta_s$ and we may then write

$$\delta_g \approx \beta_s = -n_g R_e \cos \alpha_a \cdot \int_{R_e}^{R_T} \frac{\frac{dn}{dR}\, dR}{n^2 R \left[1 - \left(\frac{n_g}{n}\right)^2 \left(\frac{R_e}{R}\right)^2 \cos^2 \alpha_a\right]^{1/2}} \quad (65)$$

This formula depends explicitly only on the arrival angle α_a. It is valid for all α_a and for all sufficiently large values of R_s. If α_a could be accurately measured this formula would be a very useful and accurate approximation for δ_g, for large satellite distances. (We could replace α_a by α_t and use the resulting expression as an approximation for δ_g, but it probably would not be very accurate except for large elevation angles.) Since α_a cannot be measured accurately for radar waves, however, formula (65) is useful mainly for visual observations. In this case it can be written in a modified form.

The Classical Formula for the Tropospheric Bending Angle – For visible light the ionospheric bending is completely negligible, because of the inverse dependence of the refractivity on frequency (see Equation (9)), and we need only consider the effect of the troposphere. We may assume that the tropospheric refractivity decreases monotonically with altitude, if we disregard local effects such as temperature inversions. In other words, we may assume that n(R) is a strictly monotonically decreasing function of R throughout the troposphere (see Equation (2) and the subsequent discussion). (It follows that R is a (single-valued) function of n, and that R(n) also decreases strictly monotonically.) We may therefore use n as the variable of integration in Equation (65); the equation then takes the form

$$\delta_g \approx \beta_s = n_g R_e \cos \alpha_a \cdot \int_1^{n_g} \frac{dn}{n^2 R \left[1 - \left(\frac{n_g}{n}\right)^2 \left(\frac{R_e}{R}\right)^2 \cos^2 \alpha_a\right]^{1/2}}. \quad (66)$$

We have used the facts that $n(R_e) = n_g$ and $n(R_T) \approx 1$, where R_T is taken to be the top of the troposphere. Equation (66) is the classical astronomical formula for the atmospheric bending of visible light. It is valid for $0 \leq \alpha_a \leq \pi/2$. For astronomical objects R_s is of course extremely large, so that (66) is a very

accurate formula for the elevation angle error δ_g. A simple derivation of this formula is given in Reference 3 (pages 62-65). Note that an accurate measurement of α_a is all that is needed to compute δ_g, assuming that the refractivity profile is known. In practice, of course, the integral is approximated to obtain simple formulas valid for objects not too close to the horizon; for objects near the horizon, empirical tables are used (see Reference 3 for a discussion of these approximations).

REFERENCES

1. Born, Max, and Wolf, Emil: Principles of Optics. Pergamon Press, 1959

2. Bean, B. R. and Dutton, E. J.: Radio Meteorology. National Bureau of Standards Monograph 92, March 1, 1966

3. Smart, W. M.: Text-Book on Spherical Astronomy. Cambridge University Press, 1956

4. Donahue, T. M.: Ionospheric Composition and Reactions. Science, Vol. 159, No. 3814, Feb. 2, 1968, pp. 489-498

5. Stratton, J. A.: Electromagnetic Theory. McGraw-Hill Book Co., Inc., 1941

6. Sommerfeld, A. and Brillouin, L: Über die Fortflanzug des Lichtes in dispergierenden Medien. Ann. d. Phys., Vol. 44, No. 177, 1914, p. 203

7. Ginzburg, V. L.: Propagation of Electromagnetic Waves in Plasma. Gordon and Breach, Science Publishers, Inc., New York, 1961 (Translated from the Russian)

8. Budden, K. G.: Radio Waves in the Ionosphere. Cambridge University Press, 1961

9. Bourdeau, R. E.: Ionospheric Research from Space Vehicles. GSFC X-Document X-615-62-234, December 31, 1962

UNCITED REFERENCES

1. Al'pert, Ya. L.: Radio Wave Propagation and the Ionosphere. Consultants Bureau, New York, 1963 (Translation from the Russian)

2. Kaula, W. M.: Celestial Geodesy. NASA TN D-1155, March, 1962

3. Schmid, Paul E.: Atmospheric Tracking Errors at S- and C-Band Frequencies. GSFC X-Document X-507-65-398, October 5, 1965

APPENDIX

THE SPERRY REFRACTIVITY PROFILE

The refractivity profile described in this appendix is referred to in this report as the "Sperry profile" because it was used in a report written by the Sperry Gyroscope Co. The profile assumes an exponentially decreasing tropospheric refractivity and a nearly parabolic shape for the F_2 layer of the ionosphere. The D and E layers of the ionosphere are neglected. The profile and the ϵ -function corresponding to it are merely piecewise continuous, but the ϵ-function satisfies condition (16) and can therefore be used in the various integrals with assurance that they will exist.

Recall that the refractivity N is defined by

$$N = (n - 1) \times 10^6 ,$$

where n is the index of refraction. The tropospheric profile is assumed to be exponentially decreasing:

$$N = N_0 e^{-h/H}, \tag{A1}$$

where N_0 (the refractivity at the earth's surface) = 313, h = height (in kilometers) above the earth's surface and H (scale height) = 7 kilometers.

The ionospheric (phase) refractivity is assumed to be given by the formula

$$N = -4.03 \times 10^{-5} \frac{M}{f^2} , \tag{A2}$$

where f is the signal frequency in megahertz and M is the number of electrons per cubic meter. Except for the units used for f, Equation (A2) is identical with

Equation (9). Equation (A2) is considered to be valid for $f \geq f_{critical} = 8.97 \times 10^{-6} M_{max}^{\frac{1}{2}}$ megahertz, where M_{max} is the maximum electron density in the F_2 layer, in electrons per cubic meter. If we take 10^{12} electrons/meter^3 as the largest likely value for M_{max} we obtain 8.97 megahertz for $f_{critical}$, which is well below any radar or even VHF frequencies.

The electron density M is assumed to be given by

$$M = M_{max} [1 - (1 - \gamma)^2] \quad \text{for } 0 \leq \gamma \leq 1 . \tag{A3}$$

and by

$$M = M_{max} \operatorname{sech} [\pi/4 \cdot (\gamma - 1)] \quad \text{for} \quad \gamma \geq 1 , \tag{A4}$$

where

$$\gamma = \frac{h - h_0}{h_m - h_0} = \frac{(R - R_e) - h_0}{h_m - h_0} ,$$

h_0 = height of the base of the F_2 layer and h_m = height of the maximum electron density in the F_2 layer. Note that

$$\operatorname{sech} x = \frac{1}{\cosh x} = \frac{2}{e^x + e^{-x}} = \frac{2 e^{-x}}{1 + e^{-2x}} ;$$

the next to the last form is probably the most convenient for numerical calculation, while the last form clearly shows the decreasing exponential behavior of sech x for large x.

The Sperry profile can now be precisely given as follows:

- For $0 \leq h \leq 40$ kilometers, (A1) holds;
- For 40 km $\leq h \leq h_0$, N = 0;
- For $h_0 \leq h \leq h_m$, (A3) holds;
- For $h_m \leq h \leq 2{,}000$ km, (A4) holds;
- For $h > 2{,}000$ km, N = 0.

Note that the ionospheric (i.e., F_2-layer) N-profile is parabolic below h_m; above h_m it is nearly parabolic, but trails off to an exponential shape at great heights. Note also that the tropospheric profile is fixed, while the ionospheric profile depends (for a given frequency) on three parameters: h_0, h_m and M_{max}.

If, instead of N, we work with ϵ, which we recall is defined by the relation

$$\epsilon = \frac{n_s}{n} - 1 = \frac{1 + 10^{-6} N_0}{1 + 10^{-6} N} - 1,$$

it is clear that the ϵ-function corresponding to the Sperry N-profile is defined as follows:

- For $R_e \leq R \leq (R_e + 40)$ kilometers,

$$\epsilon(R) = \frac{1 + 10^{-6} N_0}{1 + 10^{-6} N_0 e^{-(R - R_e)/H}} - 1;$$

- For $(R_e + 40)$ km $\leq R \leq R_e + h_0$,

$$\epsilon(R) = 10^{-6} N_0;$$

- For $R_e + h_0 \leq R \leq R_e + h_m$,

$$\epsilon(R) = \frac{1 + 10^{-6} N_0}{1 - 4.03 \times 10^{-11} \frac{M_{max}}{f^2} [1 - (1 - \gamma)^2]} - 1;$$

- For $R_e + h_m \leq R \leq (R_e + 2{,}000)$ km,

$$\epsilon(R) = \frac{1 + 10^{-6} N_0}{1 - 4.03 \times 10^{-11} \frac{M_{max}}{f^2} \operatorname{sech} [\pi/4 \cdot (\gamma - 1)]} - 1;$$

- For $R > (R_e + 2{,}000)$ km,

$$\epsilon(R) = 10^{-6} N_0 .$$

The corresponding function $\tilde{\epsilon}(\rho)$ is easily obtained by using the transformation $\rho = R_e/R$. We note that the function $\epsilon(R)$ is piecewise continuously differentiable, but that it is discontinuous at $R = R_e + 40$ and $R = R_e + 2{,}000$. In other words, it is discontinuous at the top of the troposphere and at the top of the ionosphere. These discontinuities cause no difficulty, however, because $\epsilon(R)$ is clearly always positive (except that $\epsilon(R_e) = 0$), and we will show that it satisfies condition (25). It is not hard to see that there must therefore exist functions arbitrarily close to $\epsilon(R)$, but continuously differentiable everywhere, so that they are in the class $\mathcal{E}^+$. Our entire analysis is valid for any of these functions, and the Sperry ϵ-function may be regarded as a simple approximation to any of them. All of the definite integrals in the various equations and formulas are therefore guaranteed to exist if the Sperry profile is used.

It remains only to show that $\epsilon(R)$ satisfies condition (25). To show that $\epsilon(R)$ is bounded is easy; we just note that the maximum value of $\epsilon(R)$ occurs at $R = R_e + h_m$, at which point $\gamma = 1$, and therefore

$$\epsilon(R) \leq \epsilon(R_e + h_m) = \frac{1 + 10^{-6} N_0}{1 - 4.03 \times 10^{-11} \frac{M_{max}}{f^2}} - 1$$

for all R. Further, if we assume that 100 megahertz is the smallest frequency likely to be used (corresponding to VHF frequencies), and if we assume 10^{12} electrons/M^3 as the largest value of M_{max}, we get the bound

$$B = \left[\epsilon(R_e + h_m)\right]_{max} = -1 + \frac{1 + 0.000313}{1 - 0.00403} \cong 0.00436,$$

which holds for any realistic profile and for any signal frequency in the VHF band or higher.

To show that $\epsilon(R) \leq (1 - \sigma)(R - R_e)/R_e$ for some $\sigma \in (0,1)$ is more troublesome, but still not difficult. We must first look at the tropospheric profile. For $R_e \leq R \leq (R_e + 40)$ km we have

$$\epsilon(R) = \frac{1 + a}{1 + aE} - 1 .$$

$$\frac{d\epsilon}{dR} = \frac{a(1 + a)E}{H(1 + aE)^2}$$

and

$$\frac{d^2\epsilon}{dR^2} = - \frac{a(1 + a)E\,(1 - aE)}{H^2\,(1 + aE)^3} ,$$

where we have written $E = \exp[-(R - R_e)/H]$ and $a = 10^{-6}\ N_0$ for brevity. Since $E \leq 1$ for $R \geq Re$, we have $1 - aE > 1 - a = 1 - 10^{-6} \times 313 > 0$; thus the second derivative is negative, and of course the first derivative is positive. It follows that the maximum slope of $\epsilon(R)$ occurs at $R = R_e$, so that we may write

$$\epsilon(R) \leq \left.\frac{d\epsilon}{dR}\right|_{R=R_e} \cdot (R - R_e) = \frac{aR_e}{H(1+a)} \cdot \frac{(R - R_e)}{R_e}$$

$$\cong (1 - 0.715) \cdot \frac{(R - R_e)}{R_e}$$

$$= L(R) ,$$

where we have used the value $R_e = 6{,}378$ km. This shows that the required condition is satisfied by the tropospheric profile, with the value $\sigma = 0.715$.

Next we look at the region between the top of the troposphere and the bottom of the ionosphere. We find that

$$L(R_e + 40) \cong 0.00179 > \epsilon(R_e + 40) = 0.000313;$$

thus the ϵ-function is bounded above by L(R) throughout this region, and the same is obviously also true of the region above the top of the ionosphere.

Finally we will show that $\epsilon(R) < L(R)$ throughout the part of the ionosphere below the F_2 maximum; the same will then be true of the rest of the ionosphere, because $\epsilon(R)$ decreases above the F_2 maximum. For $R_e + h_0 \leq R \leq R_e + h_m$ we find that

$$\epsilon(R) = \frac{1 + a}{1 - b[1 - (1 - \gamma)^2]} - 1 .$$

$$\frac{d\epsilon}{d\gamma} = \frac{2b(1 + a)(1 - \gamma)}{\{1 - b[1 - (1 - \gamma)^2]\}^2}$$

and

$$\frac{d^2\epsilon}{d\gamma^2} = \frac{2b(1 + a)[-1 + b + 3b(1 - \gamma)^2]}{\{1 - b[1 - (1 - \gamma)^2]\}^3} ,$$

where we have written $a = 10^{-6} N_0$ and $b = 4.03 \times 10^{-11} M_{max}/f^2$ for brevity. If we again take 10^{12} electrons/M^3 and 100 megahertz as limiting values for M_{max} and f, respectively, we find that the largest possible value of b is given by $b_{max} = 4.03 \times 10^{-11} \times 10^{12}/10^4 = 0.00403$. It follows that $[-1 + b + 3b(1 - \gamma)^2] < -1 + 4b_{max} < 0$, so that the second derivative is negative. Of more interest for our present purpose is the fact that the first derivative is positive. (Recall that $0 \leq \gamma \leq 1$ in the portion of the ionosphere below the F_2 maximum). It follows that it would be sufficient (but not necessary) to show that $\epsilon(R_e + h_m) < L(R_e + h_0)$. This condition will be satisfied if $(1 + a)/(1 - b_{max}) - 1 < 0.285\, h_0/R_e$ (recall that $\gamma = 1$ for $R = R_e + h_m$); i.e., if $h_0 > 98$ kilometers. Since the F_2 layer is usually considered to begin at the considerably higher altitude of about 140 kilometers (see for example Reference 9), we may safely conclude that the linear function L(R) is in fact a majorant for the Sperry profile, for any realistic values of the parameters h_0, h_m and M_{max}.

In the computer program, which will be described in a companion report, there is a subroutine which numerically tests the ϵ-function to determine whether it satisfies condition (25). The Sperry profile was tested by this routine and was found to satisfy the condition for arbitrarily chosen, but realistic, values of the parameters h_0, h_m and M_{max}. Several sets of values for these parameters were used in the computer runs, but representative values, used for most of the runs, are: $h_0 = 240$ km, $h_m = 300$ km and $M_{max} = 5.2 \times 10^{11}$ electrons per cubic meter.