Ionospheric Ray-Tracing Equations and their Solution

Christopher J. Coleman

Abstract

The Haselgrove ray-tracing equations are derived directly from Maxwell's equations. Some methods for their solution are discussed. In particular, we look at reduced versions of the equations that allow fast numerical solution and, in some cases, analytic solution.

1. Introduction

It is now over 50 years since Jenifer Haselgrove introduced the ray-tracing equations that have become synonymous with her name. During this time, her equations have become a major tool for investigating radiowave propagation in the ionosphere. In the early days, such studies were needed because of the importance of ionospheric propagation for long-range terrestrial communications. Such communications take place at high frequencies (HF), and use the fact that radiowaves at HF (3 to 30 MHz) are refracted back down to the Earth by the ionosphere. However, with the advent of artificial satellites, ionospheric communication has become less important. Nevertheless, such communications are still an important tool for the military, aid agencies, and remote communities. Furthermore, the introduction of over-the-horizon radar (OTHR) has significantly increased the use of ionospheric propagation. The extreme demands of over-the-horizon radar have made it necessary to understand ionospheric propagation at a more refined level, and the Haselgrove equations have played an important part in such studies.

Starting with the Haselgrove equations, we review some of the ray-tracing approaches that are available for the study of ionospheric propagation. In Section 2, we derive the Haselgrove equations directly from Maxwell's equations (together with some basic plasma physics). Derivations of the Haselgrove equations tend to use ray optics, in particular the Hamiltonian equations, as their starting point [1-3]. Section 2 attempts to start from a more fundamental position, and to cast the equations as part of a procedure for the solution of Maxwell's equations in the high-frequency limit. In [4], it was found that the Runge-Kutta-Fehlberg numerical scheme constituted a very efficient means of solving the Haselgrove equations, and so we include a brief description of this algorithm. Unfortunately, there still exist ray-tracing applications for which computer solutions to the Haselgrove equations are not fast enough. A particular case is the coordinate registration (CR) problem of over-the-horizon radar. In the coordinate-registration problem, fast ray tracing is required to convert the radar range (the time for the radio signal to travel to the target) into the actual ground range. Due to the ever-changing nature of the ionosphere, these calculations need to be done in real time. Consequently, in Sections 3 to 5 we look at some simplifications to the Haselgrove equations that can provide this increased speed. In Section 3, we look at the situation where the background magnetic field can be regarded as being weak (a good approximation for most HF frequencies above 10 MHz). In Section 4, we look at the simplification that results when we totally ignore the background magnetic field, an approximation that can be made more respectable by the use of effective wave frequencies. Finally, in Section 5, we consider some first integrals of the ray-tracing equations, and we also consider analytic solutions that can be derived from these first integrals.

2. The Haselgrove Equations

For time-harmonic fields in a vacuum, Maxwell's equations yield the field equations

$$\nabla\times\nabla\times\underline{E}-\omega^2\mu_0\varepsilon_0\underline{E}=-j\omega\mu_0\underline{J}\,, \tag{1}$$

where $\underline{E}$ is the time-harmonic electric field, $\underline{J}$ is the time-harmonic current density, and ω is the wave frequency. Within the ionospheric plasma, the motion of an electron satisfies

Christopher J. Coleman is with the School of Electrical and Electronic Engineering, University of Adelaide, Adelaide, SA 5005, Australia;
e-mail: ccoleman@eleceng.adelaide.edu.au

$$m\dot{\underline{w}} = e\underline{E} + e\underline{w}\times\underline{B}_0 - m\nu\,\underline{w}, \qquad (2)$$

where m is the electron mass, e is the electron charge, ν is the plasma collision frequency, $\underline{w}$ is the electron velocity, and $\underline{B}_0$ is the magnetic field of the Earth. The plasma current is given by $eN\underline{w}$, where N is the electron density, Thus, for a time-harmonic field, Equation (2) reduces to

$$-j\omega\varepsilon_0 X\underline{E} = U\underline{J} + j\underline{Y}\times\underline{J}, \qquad (3)$$

where $U = 1 - j\nu/\omega$, $X = \omega_p^2/\omega^2$, $\underline{Y} = -e\underline{B}_0/m\omega$, and $\omega_p = \sqrt{Ne^2/\varepsilon_0 m}$ is the plasma frequency. In the high-frequency limit ($\omega \to \infty$), we normally assume an electric field of the form $\underline{E} = \underline{E}_0 \exp(-j\beta\varphi)$, where $\underline{E}_0$ and φ are both slowly varying functions of position and $\beta = \omega\sqrt{\mu_0\varepsilon_0}$ (see [5] and [6] for more information concerning the high-frequency approximation). Noting that

$$\nabla\times\nabla\times\underline{E} = \nabla(\nabla\bullet\underline{E}) - \nabla^2\underline{E},$$

and substituting the assumed form of solution into Equation (1), we find the leading order in ω yields

$$-\nabla\varphi(\nabla\varphi\bullet\underline{E}_0) + (\nabla\varphi\bullet\nabla\varphi)\underline{E}_0 - \underline{E}_0 = \frac{-j\omega\mu_0}{\beta^2}\underline{J}_0 \qquad (4)$$

where $\underline{J} = \underline{J}_0\exp(-j\beta\varphi)$. From Equation (3), it follows that

$$-j\omega\varepsilon_0 X\underline{E}_0 = U\underline{J}_0 + j\underline{Y}\times\underline{J}_0. \qquad (5)$$

The dot product of Equation (4) with $\nabla\varphi$ yields

$$\nabla\varphi\bullet\underline{E}_0 = \frac{j\omega\mu_0}{\beta^2}\nabla\varphi\bullet\underline{J}_0. \qquad (6)$$

Eliminating $\nabla\varphi\bullet\underline{E}_0$ from Equation (4) using Equation (6), we obtain

$$(\nabla\varphi\bullet\nabla\varphi - 1)\underline{E}_0 = \frac{-j\omega\mu_0}{\beta^2}\left(\underline{J}_0 - \nabla\varphi\nabla\varphi\bullet\underline{J}_0\right) \qquad (7)$$

Eliminating $\underline{E}_0$ using Equation (5),

$$(\nabla\varphi\bullet\nabla\varphi - 1)(U\underline{J}_0 + j\underline{Y}\times\underline{J}_0) = -X\left(\underline{J}_0 - \nabla\varphi\nabla\varphi\bullet\underline{J}_0\right) \qquad (8)$$

From the dot product of Equation (8) with $\nabla\varphi$, we obtain

$$-j\nabla\varphi\bullet(\underline{Y}\times\underline{J}_0) = -j(\nabla\varphi\times\underline{Y})\bullet\underline{J}_0 = (U - X)\nabla\varphi\bullet\underline{J}_0 \qquad (9)$$

Using this to eliminate $\nabla\varphi\bullet\underline{J}_0$ from Equation (8),

$$(U\nabla\varphi\bullet\nabla\varphi - U + X)\underline{J}_0$$
$$= \frac{-jX}{U - X}\nabla\varphi(\nabla\varphi\times\underline{Y})\bullet\underline{J}_0 - j\underline{Y}\times\underline{J}_0(\nabla\varphi\bullet\nabla\varphi - 1) \qquad (10)$$

Without loss of generality, we choose $Y_1 = Y$ and $Y_2 = Y_3 = 0$ (i.e., the x_1 coordinate direction is in the direction of the magnetic field). In this case, Equation (10) can be rewritten as

$$\begin{bmatrix} \left(Up^2 + X - U\right) & jYPp_1p_3 & -jYPp_1p_2 \\ 0 & \left(Up^2 + X - U\right) + jYPp_2p_3 & -jYPp_2^2 - jY\left(p^2 - 1\right) \\ 0 & jYPp_3^2 + jY\left(p^2 - 1\right) & \left(Up^2 + X - U\right) - jYPp_3p_2 \end{bmatrix}\underline{J}_0 = \underline{0} \qquad (11)$$

where $\underline{p} = \nabla\varphi$, $p^2 = \underline{p}\bullet\underline{p}$, and $P = X/(U - X)$. This will only have a nontrivial solution if the determinant of the matrix is zero, that is,

$$\left(Up^2 + X - U\right)\left[\left(Up^2 + X - U\right)^2 + Y^2P^2p_3^2p_2^2\right.$$
$$\left. -Y^2\left(Pp_3^2 + p^2 - 1\right)\left(Pp_2^2 + p^2 - 1\right)\right] = 0, \qquad (12)$$

from which either $Up^2 + X - U = 0$ (in which case, $\underline{E}_0$ is parallel to $\underline{Y}$), or

$$\left(Up^2 + X - U\right)^2 - Y^2\left(p^2 - 1\right)^2$$
$$-Y^2P\left(p^2 - 1\right)\left(p^2 - p_1^2\right) = 0, \qquad (13)$$

for which $\underline{E}_0$ has no component in the direction of $\underline{Y}$. Returning to a general system of coordinates (NB: $Yp_1 = \underline{Y}\bullet\underline{p}$ and $Y^2 = \underline{Y}\bullet\underline{Y}$), we obtain from Equation (13) that

$$\left(Up^2 + X - U\right)^2 - Y^2\left(p^2 - 1\right)^2$$
$$-P\left(p^2 - 1\right)\left[Y^2p^2 - \left(\underline{p}\bullet\underline{Y}\right)^2\right] = 0. \qquad (14)$$

Equation (14) is a first-order partial differential equation for the phase field, φ, and this can be solved by standard methods for partial differential equations. For the equation

$F(\underline{x}, \underline{p}) = 0$, we can define a solution in terms of characteristic curves (our rays). These, together with the values of φ along them, will satisfy the generalized Charpit equations (see [7], for example):

$$\frac{dx_i}{\partial F/\partial p_i} = \frac{d\varphi}{\sum_{j=1}^{3} p_j \partial F/\partial p_j} = \frac{dp_i}{-\partial F/\partial x_i}. \quad (15)$$

The above equations can be used to find the ray curves in terms of the values of the phase factor, φ, along them. We will consider the collision-free case ($U = 1$). Equation (14) then implies

$$\left(p^2-1\right)^2\left(1-Y^2-PY^2\right)+2\left(p^2-1\right)X+X^2$$

$$-P\left(p^2-1\right)Y^2+P\left(p^2-1\right)\left(\underline{Y}\bullet\underline{p}\right)^2=0. \quad (16)$$

Multiplying through by $1-X$, Equation (16) can be recast as $F\left(\underline{x},\underline{p}\right)=0$, where

$$F\left(\underline{x},\underline{p}\right)=\left(p^2-1\right)^2\left(1-X-Y^2\right)$$

$$+\left(p^2-1\right)\left[2X(1-X)-XY^2\right]$$

$$+X^2(1-X)+X\left(p^2-1\right)\left(\underline{Y}\bullet\underline{p}\right)^2. \quad (17)$$

Consequently,

$$\frac{\partial F}{\partial p_i}$$

$$=4p_i\left(1-X-Y^2\right)\left(p^2-1\right)+2p_i\left[2X(1-X)-XY^2\right]$$

$$+2p_iX\left(\underline{p}\bullet\underline{Y}\right)^2+2Y_iX\left(p^2-1\right)\underline{p}\bullet\underline{Y}. \quad (18)$$

We introduce a new parameter, t, along the characteristic curves defined by

$$dt=-Xd\varphi\Big/\sum_{j=1}^{3}p_j\partial F/\partial p_j$$

and then Equation (15), together with a rearranged version of Equation (18), yields

$$\frac{dx_i}{dt}=-\frac{4p_i}{X}\left(1-X-Y^2\right)\left(p^2-1+X\right)-2p_iY^2$$

$$-2p_i\left(\underline{p}\bullet\underline{Y}\right)^2-2Y_i\left(p^2-1\right)\underline{p}\bullet\underline{Y}. \quad (19)$$

If we now introduce the new dependent variable $q=\left(p^2-1\right)\Big/X$ (as defined in [3]), we obtain

$$\frac{dx_i}{dt}=-4p_i\left(1-X-Y^2\right)\left(q^2+1\right)-2p_iY^2$$

$$-2p_i\left(\underline{p}\bullet\underline{Y}\right)^2-2Y_i\left(p^2-1\right)\underline{p}\bullet\underline{Y}, \quad (20)$$

which is the Haselgrove equation for the advancement of $\underline{p}$ [2]. The other equation implied by Equation (15) requires the coordinated derivatives of F, that is

$$\frac{\partial F}{\partial x_i}=-\left(p^2-1\right)^2\left(\frac{\partial X}{\partial x_i}+2\underline{Y}\bullet\frac{\partial \underline{Y}}{\partial x_i}\right)$$

$$+\left(p^2-1\right)\left[\frac{\partial X}{\partial x_i}\left(2-4X-Y^2\right)-2X\underline{Y}\bullet\frac{\partial \underline{Y}}{\partial x_i}\right]$$

$$+X\frac{\partial X}{\partial x_i}(2-3X)+\frac{\partial X}{\partial x_i}\left(p^2-1\right)\left(\underline{Y}\bullet\underline{p}\right)^2$$

$$+2\underline{p}\bullet\frac{\partial \underline{Y}}{\partial x_i}X\left(p^2-1\right)\left(\underline{Y}\bullet\underline{p}\right). \quad (21)$$

Introducing the same dependent and independent variables as for Equation (20), we obtain from Equation (15) that

$$\frac{dp_i}{dt}=\left\{\left[-\left(p^2-1\right)+\left(\underline{p}\bullet\underline{Y}\right)^2+2(1-2X)-Y^2\right]q\right.$$

$$\left.+2-3X\right\}\frac{\partial X}{\partial x_i}$$

$$+\sum_{j=1}^{3}\left(p^2-1\right)\left[2\left(\underline{p}\bullet\underline{Y}\right)p_j-2(q+1)Y_j\right]\frac{\partial Y_j}{\partial x_i}$$

(22)

Equations (20) and (22) together constitute the original

Haselgrove equations [2]. Furthermore, recasting Equation (16) in terms of q [2], we obtain

$$q^2\left(1-X-Y^2\right)+q\left[\left(\underline{Y}\bullet\underline{p}\right)^2+2(1-X)-Y^2\right]+1-X=0 \tag{23}$$

from which q can be derived (note that there are two possible values of q, corresponding to two of the possible solutions to Equation (11)). The corresponding current, and hence the electric field through Equation (5), can be obtained from the homogeneous Equations (11) using the calculated values of $\underline{p}$ on the ray. However, Equations (11) do not determine the magnitude of the field. This must be calculated by tracking the power out from the source. This can be done by tracing out ray tubes, and noting that within these tubes, the power, $Ap\left|\underline{E}_0\right|^2/(2\eta_0)$, must remain constant ($A$ is the cross-sectional area of the tube). The magnitude of the electric field can then be calculated from the power.

There remains the issue of solving the above ray-tracing equations. Starting with some initial value, $\underline{w}(0)$, for $\underline{w}$, the system of ordinary differential equations, $d\underline{w}/dt=\underline{W}(\underline{w})$, can be solved using Runge-Kutta techniques. This was suggested in the Haselgrove and Haselgrove paper [2]. If we start a ray at the Earth's surface, the initial values of $\underline{p}$ are given by the direction cosines at this starting point. However, the extreme variations that can exist in the ionosphere often make it necessary to vary the increment in t at each stage of the solution process, in order to maintain accuracy. Haselgrove and Haselgrove [2] suggested the introduction of a scaling to produce this effect, but the Runge-Kutta-Fehlberg (RKF) method (see [8] for example) can also achieve this.

Consider the solution at time t and prospective increment Δt. Form the vectors $\underline{k}_1$, $\underline{k}_2$, $\underline{k}_3$, $\underline{k}_4$, $\underline{k}_5$, and $\underline{k}_6$ through the process

$$\underline{k}_1=\Delta t\underline{W}(\underline{w}), \tag{24}$$

$$\underline{k}_2=\Delta t\underline{W}\left(\underline{w}+\underline{k}_1/4\right), \tag{25}$$

$$\underline{k}_3=\Delta t\underline{W}\left(\underline{w}+3\underline{k}_1/32+9\underline{k}_2/32\right), \tag{26}$$

$$\underline{k}_4=\Delta t\underline{W}\left(\underline{w}+1932\underline{k}_1/2197-7200\underline{k}_2/2197+7296\underline{k}_3/2197\right), \tag{27}$$

$$\underline{k}_5=\Delta t\underline{W}\left(\underline{w}+439\underline{k}_1/216-8\underline{k}_2+3680\underline{k}_3/513-845\underline{k}_4/4104\right), \tag{28}$$

$$\underline{k}_6=\Delta t\underline{W}\left(\underline{w}-8\underline{k}_1/27+2\underline{k}_2-3544\underline{k}_3/2565+1859\underline{k}_4/4104-11\underline{k}_5/40\right)\cdot \tag{29}$$

The approximate solution, $\underline{w}(t)+\Delta\underline{w}$ at $t+\Delta t$, is obtained using

$$\Delta\underline{w}_4=25\underline{k}_1/216+1408\underline{k}_3/2565+2197\underline{k}_4/4104-\underline{k}_5/5 \tag{29}$$

for the fourth-order Runge-Kutta method, and by

$$\Delta\underline{w}_5=16\underline{k}_1/135+6656\underline{k}_3/12825+28561\underline{k}_4/56430-9\underline{k}_5/50+2\underline{k}_6/55 \tag{30}$$

for the fifth-order method. The global truncation error will be $O\left(\Delta t^4\right)$ for the fourth-order method and $O\left(\Delta t^5\right)$ for the fifth-order method. The magnitude of the difference between the fourth- ($\underline{w}_4$) and fifth- ($\underline{w}_5$) order estimates, $\Delta w=\left|\underline{w}_5-\underline{w}_4\right|$, gives an estimate of the global truncation error in the fourth-order method.

The Runge-Kutta-Fehlberg method proceeds by adjusting the step at each stage so that this global error remains close to a pre assigned value, ΔE. That is, at each stage we adjust Δt so that

$$\Delta t_{new}=\Delta t_{old}\left(\frac{\Delta E}{\Delta w}\right)^{\frac{1}{4}}. \tag{31}$$

This approach has been used in [4] to provide an efficient numerical implementation of the Haselgrove equations.

3. The Weak-Background-Field Limit

For the collision-free case ($U=1$), Equation (3) can be inverted to yield

$$\underline{J} = \frac{-j\omega\varepsilon_0 X}{1-Y^2}\left(\underline{E} - j\underline{Y}\times\underline{E} - \underline{Y}\bullet\underline{E}\underline{Y}\right)\cdot \quad (32)$$

Up until now, we have assumed that both the plasma frequency, ω_p, and the gyro frequency, $\omega_H = \left|e\underline{B}_0\right|/m$, are of the same order as the wave frequency, ω. We will now assume that $\omega >> \omega_p$ and, hence, we can ignore the term of second order, Y^2, in Equation (32). Substituting the trial solution $\underline{E} = \underline{E}_0 \exp(-j\beta\varphi)$ into Equation (1), the β^2 and β terms yield

$$-\nabla\varphi\nabla\varphi\bullet\underline{E}_0 + (\nabla\varphi\bullet\nabla\varphi)\underline{E}_0 - (1-X)\underline{E}_0 = 0 \quad (33)$$

and

$$-\nabla\varphi\nabla\bullet\underline{E}_0 - \nabla(\nabla\varphi\bullet\underline{E}_0) + 2\nabla\varphi\bullet\nabla\underline{E}_0 + \nabla^2\varphi\underline{E}_0$$

$$= \beta X\underline{Y}\times\underline{E}_0, \quad (34)$$

respectively. By taking the divergence of Equation (1), we obtain $\nabla\bullet\underline{E} = (j/\omega\varepsilon_0)\nabla\bullet\underline{J}$, from which and $\nabla\varphi\bullet\underline{E}_0 = 0$

$$(1-X)\nabla\bullet\underline{E}_0 = \nabla X\bullet\underline{E}_0 - \beta X\nabla\varphi\bullet(\underline{Y}\times\underline{E}_0)$$

for the two leading orders. Consequently, Equations (33) and (34) reduce to

$$(\nabla\varphi\bullet\nabla\varphi - 1 + X)\underline{E}_0 = 0 \quad (35)$$

and

$$\frac{-\nabla\varphi}{1-X}\nabla X\bullet\underline{E}_0 + \frac{\beta X}{1-X}\nabla\varphi\nabla\varphi\bullet(\underline{Y}\times\underline{E}_0)$$

$$+2\nabla\varphi\bullet\nabla\underline{E}_0 + \nabla^2\varphi\underline{E}_0 = \beta X\underline{Y}\times\underline{E}_0 \quad (36)$$

From Equation (35), we obtain the ray-tracing equations $d\underline{p}/dt = \nabla n^2/2$ and $d\underline{x}/dt = \underline{p}$, where $dt = d\varphi/p^2$ (i.e., t is the group distance) and $n^2 = 1 - X$. These equations do not contain the background magnetic field, and hence can be solved far more efficiently than the original Haselgrove equations. Once again, the Runge-Kutta-Fehlberg method can be used with great effect.

Equation (36) can now be recast as a differential equation for $\underline{E}_0$ along a ray:

$$2\frac{d\underline{E}_0}{dt} + \frac{d\underline{x}}{dt}\nabla\ln n^2\bullet\underline{E}_0 + \nabla^2\varphi\underline{E}_0$$

$$= \beta X\left[\underline{Y}\times\underline{E}_0 - \frac{1}{1-X}\frac{d\underline{x}}{dt}\frac{d\underline{x}}{dt}\bullet(\underline{Y}\times\underline{E}_0)\right]\cdot \quad (37)$$

Taking the dot product of Equation (37) with $\underline{E}_0$, we find that $d\left|\underline{E}_0\right|^2/dt + \nabla^2\varphi\left|\underline{E}_0\right|^2 = 0$. Combining this with Equation (37), we obtain an equation for the polarization vector, $\underline{P} = \underline{E}_0/\left|\underline{E}_0\right|$, of the form

$$2\frac{d\underline{P}}{dt} + \frac{d\underline{x}}{dt}\nabla\ln n^2\bullet\underline{P}$$

$$= \beta X\left[\underline{Y}\times\underline{P} - \frac{1}{1-X}\frac{d\underline{x}}{dt}\frac{d\underline{x}}{dt}\bullet(\underline{Y}\times\underline{P})\right]. \quad (38)$$

Essentially, the background magnetic field manifests itself as a rotation of the polarization vector about the ray direction.

4. The Negligible-Background-Field Limit

When the magnetic field of the Earth can be ignored ($\underline{B}_0 = 0$), the Haselgrove equations can be reduced to

$$\frac{d}{ds}\left(n\frac{d\underline{x}}{ds}\right) = \nabla n, \quad (39)$$

where s is the distance along the ray path. The equations are themselves the Euler-Lagrange (EL) equations for the variational principle

$$\delta\int_T^R n(\underline{x})\,ds = 0, \quad (40)$$

i.e., Fermat's principle. If we now consider the case for a ray path that remains in a plane, we can use a polar coordinate system (r,θ), and the variational Euler-Lagrange principle becomes

$$\delta\int_T^R n(r,\theta)\left[\left(\frac{dr}{d\theta}\right)^2 + 1\right]^{1/2} d\theta = 0\cdot \quad (41)$$

The above ray-tracing equations can be used when any deviations from the great-circle path (certainly the case for a uniform ionosphere) are negligible. Coordinate r is the distance from the center of the Earth, and θ is an angular coordinate along the great-circle path. The Euler-Lagrange equations for the above variational principle can be reduced [9] to

$$\frac{dQ}{d\theta} = \frac{1}{2}\frac{r}{\sqrt{n^2 - Q^2}}\frac{\partial n^2}{\partial r} + \sqrt{n^2 - Q^2} \qquad (42)$$

and

$$\frac{dr}{d\theta} = \frac{rQ}{\sqrt{n^2 - Q^2}}, \qquad (43)$$

where $Q = n(dr/d\theta) \Big/ \sqrt{(dr/d\theta)^2 + r^2}$. We now only need to solve two ordinary differential equations to find a ray, a lot more efficient than the four equations for the planar Haselgrove equations. The downside is that we have ignored lateral deviations and the background magnetic field. Some measure of the background magnetic field can be incorporated by ray tracing at effective wave frequencies $\omega \pm \omega_H/2$, where ω_H is the gyro frequency at the midpoint of the ray path (the different effective frequencies correspond to the two different roots of Equation (23)). (The reader should refer to [10] for more accurate expressions of effective frequency.) Once again, the Runge-Kutta-Fehlberg technique can be used to efficiently solve the ray-tracing equations.

If we move to a nearby ray path, the deviations (δr and δQ) in quantities r and Q can be calculated from

$$\frac{d(\delta Q)}{d\theta} = \frac{r}{\sqrt{n^2 - Q^2}}\left[\frac{\delta r}{2}\left(\frac{1}{r}\frac{\partial n^2}{\partial r} + \frac{\partial^2 n^2}{\partial r^2}\right)\right.$$

$$\left. - \left(\frac{1}{2\left(n^2 - Q^2\right)}\frac{\partial n^2}{\partial r} - \frac{1}{r}\right)\left(\frac{\delta r}{2}\frac{\partial n^2}{\partial r} - Q\delta Q\right)\right] \qquad (44)$$

and

$$\frac{d(\delta r)}{d\theta} = \frac{r}{\sqrt{n^2 - Q^2}}\left[\frac{\delta r Q}{r} + \delta Q\right.$$

$$\left. - \frac{Q}{n^2 - Q^2}\left(\frac{\delta r}{2}\frac{\partial n^2}{\partial r} - Q\delta Q\right)\right]. \qquad (45)$$

Since power flows within the confines of the rays emanating from a source, the above deviation equations are most useful in calculating the change in power density along a ray path. Lateral to the ray plane, the deviation can be estimated by assuming the rays to follow great-circle paths through the origin of the main ray. However, this approximation is only valid when the lateral variations in the ionosphere are weak.

5. Analytic Techniques

Consider the variational principle for a two-dimensional ray path in Cartesian coordinates (x, y). For the case where the refractive index depends only on the y coordinate, the variational principle becomes

$$\delta \int_T^R n(y)\left[\left(\frac{dy}{dx}\right)^2 + 1\right]^{1/2} dx = 0, \qquad (46)$$

from which a standard result of variational calculus yields a first integral of the Euler-Lagrange equations

$$n(y) \Bigg/ \left[\left(\frac{dy}{dx}\right)^2 + 1\right]^{1/2} = C, \qquad (47)$$

where C is a constant of integration. This is Snell's law for a horizontally stratified ionosphere, and is a single first-order ordinary differential equation for the ray path. In the case of radial coordinates with the refractive index depending on the radial coordinate r alone, there is a first integral of the form $n(r)r^2(d\theta/ds) = C$. This is Bouger's law, which is the generalization of Snell's law to a spherically stratified ionosphere (see [11]). In [12], it was shown that important quantities such as ground range could be derived analytically in the case of an ionospheric layer with refractive index $n = \sqrt{\alpha + \beta/r + \gamma/r^2}$ for $r_b < r < r_m + y_m$, and $n = 0$ elsewhere. Parameters α, β, and γ are given by $\alpha = 1 - (\omega_c/\omega)^2 + (r_b\omega_c/y_m\omega)^2$, $\beta = 2r_b^2 r_m\omega_c/\omega$, and $\gamma = (r_m r_b\omega_c/y_m\omega)^2$, where y_m is the thickness of the layer, r_b is the radius of the base of the layer, r_m is the radius of minimum refractive index, and ω_c is the plasma frequency at this height. Such an ionospheric layer is known as a quasi-parabolic layer. It was shown in [12] that a ray launched at the surface of the Earth with an initial elevation ϕ_0 will land at a distance

$$D = 2r_E\left[(\phi - \phi_0)\right.$$

$$\left. - \frac{r_E\cos\phi_0}{2\sqrt{\gamma}}\ln\frac{\beta^2 - 4\alpha\gamma}{4\gamma\left(\sin\phi + \frac{1}{r_b}\sqrt{\gamma} + \frac{1}{2\sqrt{\gamma}}\beta\right)^2}\right], \qquad (48)$$

where $\cos\phi = r_E/r_b\cos\phi_0$, and r_E is the radius of the Earth. Such results arguably provide the most efficient form of ray tracing.

Snell's law can be further generalized [13] to the following result. If, in two dimensions, the refractive index has the form

$$n(x,y) = R\left(\Im\{g(z)\}\right)\left|g'(z)\right|, \tag{49}$$

where $z = x + jy$, then the ray trajectories will satisfy

$$\frac{R\left(\Im\{g(z)\}\right)}{\left|g'(z)\right|}\frac{d\Re\{g(z)\}}{ds} = C. \tag{50}$$

When $g(z) = j\ln z$, we obtain Bouger's law on noting that $\ln z = \log r + j\theta$. Consider the conformal transformation $Z = g(z)$, where $Z = X + jY$. Then, Equation (50) will take the form

$$R(Y)(dX/dS) = C \tag{51}$$

in the new (X,Y) coordinates, with $dS = \left|g'(z)\right| ds$ being the distance element in these new coordinates. Equation (51) can be integrated [9] to yield

$$X + C_0 = \int \frac{C}{\sqrt{R^2(Y) - C^2}} dY. \tag{52}$$

We can effectively study the propagation through an ionosphere defined by Equation (49) by studying propagation through a horizontally stratified ionosphere. The quasi-parabolic ionosphere is obtained by introducing $R(Y) = \sqrt{\gamma + \beta\exp(Y) + \alpha\exp(2Y)}$. The integration of Equation (2) then yields

$$X + C_0 =$$

$$\frac{-C}{\sqrt{\gamma - C^2}}\left\{\ln\left[2\sqrt{\gamma - C^2}\sqrt{\gamma - C^2 + \beta\exp(Y) + \alpha\exp(2Y)}\right.\right.$$

$$\left.\left. + \beta\exp(Y) + 2\left(\gamma - C^2\right)\right] - Y\right\}. \tag{53}$$

To obtain the ray path in polar coordinates, we choose $g(z) = j\ln z$, from which we note that $X = -\theta$ and $Y = \ln r$. However, to obtain the results described in [11] we need to add the straight-line sections of the ray that connect the Earth's surface to the base of the ionosphere.

If we choose $g(z) = j\ln(z - z_0)$, instead of $g(z) = i\ln z$, we then obtain the eccentric ionospheres discussed in [13]. Another interesting example arises when $g(z) = \left[\cos(\alpha) + j\sin(\alpha)\right]z$. In this case, the two-dimensional horizontally stratified ionosphere with $\mu(y) = R(y)$ is given a tilt through angle α.

6. Conclusion

We have derived the Haselgrove ray-tracing equations directly from Maxwell's equations, and we have considered some methods for their numerical solution. Even with the fast computers of today, there are still applications for which the solution of the complete Haselgrove equations is still not fast enough. We have therefore looked at some simplifications that could deliver the required speed. In particular, we have looked at ray equations that ignore the background magnetic field and assume propagation to be in a plane. This reduces the ordinary differential equation system to two equations, rather than the six equations of the full Haselgrove formulation. For further speed increases, we have looked at generalizations of Snell's law and the analytic solutions that can be obtained from such generalizations.

7. References

1. J. Haselgrove, "Ray Theory and a New Method for Ray Tracing," *Proc. Phys. Soc.*, The Physics of the Ionosphere, 1954, pp. 355-64.
2. C. B. Haselgrove and J. Haselgrove, "Twisted Ray Paths in the Ionosphere," *Proc. Phys. Soc.*, **75**, 1960, pp. 357-63.
3. J. Haselgrove, "The Hamilton Ray Path Equations," *J. Atmos. Terr. Phys.*, **2**, 1963, pp. 397-9.
4. C. J. Coleman, "A General Purpose Ionospheric Ray Tracing Procedure," Defence Science and Technology Organisation Australia, Technical Report SRL-0131-TR, 1993.
5. D. S. Jones, *Methods in Electromagnetic Wave Propagation*, Oxford, Clarendon Press, 1994.
6. K. G. Budden, *The Propagation of Radio Waves*, New York, Cambridge University Press, 1985.
7. J. R. Smith, *Introduction to the Theory of Partial Differential Equations*, London, Van Nostrand, 1967.
8. J. F. Mathews, *Numerical Methods for Computer Science, Engineering and Mathematics*, New York, Prentice Hall, 1987.
9. C. J. Coleman, "A Ray Tracing Formulation and its Application to Some Problems in OTHR," *Radio Science*, **33**, 1998, pp. 1187-1197.
10. J. A. Bennett, J. Chen, and P. L. Dyson, "Analytic Ray Tracing for the Study of HF Magneto-Ionic Radio Propagation in the Ionosphere," *Applied Computational Electromagnetics Society Journal*, **6**, 1991, pp. 192-210.
11. J. M. Kelso, "Ray Tracing in the Ionosphere," *Radio Science*, **3**, 1968, pp. 1-12.
12. T. A. Croft and H. Hoogasian, "Exact Ray Calculations in a Quasi-Parabolic Ionosphere with no Magnetic Field," *Radio Science*, **3**, 1968, pp. 69-74.
13. C. J. Coleman, "On the Generalization of Snell's Law," *Radio Science*, **39**, 2004.
14. K. Folkestad, "Exact Ray Computations in a Tilted Ionosphere with No Magnetic Field," *Radio Science*, **3**, 1968, pp. 81-84.