NTIA Report 87-219

A General Theory of Radio Propagation through a Stratified Atmosphere

George Hufford

U.S. DEPARTMENT OF COMMERCE
Malcolm Baldrige, Secretary

Alfred C. Sikes, Assistant Secretary
for Communications and Information

July 1987

TABLE OF CONTENTS

A GENERAL THEORY OF RADIO PROPAGATION THROUGH A STRATIFIED ATMOSPHERE

George Hufford*

When wave propagation through a stratified atmosphere is formulated in operator theoretic terms, it becomes evident that the problem does not follow the guide of the usual examples of mathematical physics. Nevertheless, such a formulation is useful to reveal why many of the usual procedures are valid and where they may be deficient. In particular, even if the refractivity profile is merely required to belong to a general class of perhaps badly discontinuous functions, the problem always has a well-behaved solution that may be subjected to a modal analysis. On the other hand, the resulting mode series converges neither rapidly nor uniformly, and, as the asymptotic behavior of the modes shows, one is advised to use that series only with caution.

Key Words: Airy functions; contraction semi-groups; creeping wave modes; modal analysis; radio wave propagation; stratified atmospheres; wave guide modes; weak convergence

1. INTRODUCTION

The problem of electromagnetic propagation through a horizontally stratified atmosphere has been extensively studied by a great many authors. Among the first were the contributors to Kerr (Freehafer et al., 1951) in their descriptions of efforts made during World War II to explain anomalous radar returns. The following decade saw a flurry of activity marked especially by publication of the books of Wait (1962) and Fock (1965). These summarize the problems involved and the advances made in obtaining satisfactory solutions. Subsequent work has been abstracted in the extensive bibliography of Arora and Wait (1978); recent examples include the numerically oriented reports of Pappert and Goodhart (1977) and of Marcus and Stuart (1981).

In most of these previous studies one encounters such words as "eigenvalues," "modes," and "orthogonality"--words that make one think of inner products, Hilbert spaces, and self-adjoint transformations. There are, however, two difficulties with these notions. First, despite the appearance of

*The author is with the Institute for Telecommunication Sciences, National Telecommunications and Information Administration, U. S. Department of Commerce, Boulder, CO 80303 3328.

the Helmholtz equation, the problem is not a self-adjoint one. The principal reason for this is the radiation condition at infinity, which is an explicitly nonreal condition. And second, it is difficult to see how Hilbert space can enter at all since even the simple spherical wave e^{ikr}/r is not square-integrable in 3-space. The situation here can be contrasted with the theory of cavities as described, e.g., by Jones (1964, Ch. 4). In that theory the finiteness of a cavity (and the assumption of perfectly conducting walls) gives rise to a satisfactorily self-adjoint problem, the solutions to which have qualitative properties that are immediate consequences of this fact.

Despite the difficulties, there are functional analytic reasons for using the above terms and it is our purpose here to explore these reasons using modern operator theory. We believe that such an analysis will clarify many of the mysteries of mode theory and that, indeed, we shall find some new and useful results. Terminology and background for our approach may be found in such texts as those of Stone (1932), Helmberg (1969), and Dunford and Schwartz (1958, particularly Chapters 7 and 8).

Before proceeding to a particular problem, let us first recall how to introduce Hilbert space and square-integrable functions into propagation problems. We do this by the simple device of assuming that space is filled with a small, but positive, conductivity. Using the time convention $e^{-i\omega t}$ we shall assume that the wave number is complex and has the form

$$k = k_o e^{i\delta} \tag{1}$$

where k_o and δ are both real and positive. An outgoing wave will then be characterized by the property that it decreases to zero exponentially.

The problem we shall treat is that of propagation around a spherical earth and through a horizontally stratified atmosphere. We use a formulation that can be traced back to Fock (1965, Ch. 13). Although this formulation is scalar and two-dimensional and an avowed approximation, its solutions are the same as the usual approximations to the solution of the more physically realistic problem.

We suppose a rectangular coordinate system in which x is the distance along the path of propagation, y is transverse to the path, and z is the altitude above the earth's surface. Then we seek a function $\phi(x,z)$ that will represent a component of the electromagnetic field and that satisfies

$$\frac{\partial\phi}{\partial x} = \frac{i}{2k}\frac{\partial^2\phi}{\partial z^2} + ikM(z)\phi, \qquad x > 0,\ z > 0 \tag{2}$$

where M is the modified refractivity

$$M(z) = N(z) + \gamma z. \tag{3}$$

Here, N is the refractivity of the atmosphere (with an order of magnitude of perhaps 10^{-3}) and γ (equal to about $157 \cdot 10^{-9}\ m^{-1}$ in the case of Earth) is the curvature of the assumed earth.

In addition, we require that at the surface of the earth the field satisfies an impedance boundary condition

$$\frac{\partial}{\partial z}\phi(x,0) = -ikZ\phi(x,0), \qquad x > 0 \tag{4}$$

with $Re(Z) \geqq 0$, and that there be a "source function" u(z) so that

$$\phi(0,z) = u(z), \qquad z > 0. \tag{5}$$

The complex number Z is the "normalized surface impedance" and the condition that its real part is nonnegative implies the earth absorbs energy. Very often it is desirable to allow Z to become infinite so that the boundary condition becomes $\phi(x,0) = 0$. Here, however, we shall, to simplify our arguments, assume that Z is always finite.

It will also be necessary to impose conditions on the refractivity N. Now we do not want to exclude the case where N has a jump discontinuity--i.e., where there are atmospheric layers of different properties, one lying directly above the other. Therefore, particularly since it will not harm our arguments, we shall require only that N is a bounded measurable function. In addition we assume that N is real and nonnegative with, say,

$$0 \leqq N(z) \leqq N_1 \tag{6}$$

and that N vanishes identically at sufficiently great heights so that there exists a height z_a such that

$$N(z) = 0 \qquad \text{for } z > z_a \tag{7}$$

These latter two requirements--that N be nonnegative and that it vanish at sufficiently great heights--are probably not crucial for the solution to the problem. For example, it should be sufficient to say that at great heights N(z) approaches zero rapidly enough. However, the two requirements seem physically realistic and they introduce enough simplification in what follows to warrant their retention.

Of course, the introduction of a possibly discontinuous $N(z)$ means that such equations as (2) must be understood to be valid only almost everywhere. But this sort of qualification is always necessary when we are dealing with spaces of integrable functions and we shall not mention the point again.

2. RESTATEMENT OF THE PROBLEM AND A FIRST SOLUTION

The only reason $\phi = 0$ is not a solution to the problem described in (2), (4), and (5) is the presence of the source function u in (5). Furthermore, since the equations are all linear in ϕ, we expect the solution will be linear in u. Indeed, we would expect that for each $x > 0$ the solution defines a function of z which is linear in u, and we would write $\phi(x,z) = T_x u(z)$ where T_x is a linear operator that is presently unknown and whose discovery will constitute the solution to the problem.

Adopting this approach, we note that each T_x carries functions of z into functions of z. We must now specify what kind of functions we want to consider. For this we choose the Hilbert space of square-integrable functions of z defined on the interval $[0, \infty)$. The inner product is the usual one given by

$$(u,v) = \int_0^\infty u(z)v^*(z)\, dz \tag{8}$$

where the star is used to denote the complex conjugate. The norm is then defined by

$$\|u\|^2 = (u,u) = \int_0^\infty |u(z)|^2 dz. \tag{9}$$

Before continuing, let us note that the norm as thus defined has a useful physical significance. Let us suppose that the original electromagnetic problem involved vertically polarized waves and that the function ϕ was defined so that the magnetic field has the component $H_y = \phi e^{ikx}$. Now the basic premise on which are based the approximations leading to Fock's formulation given above is that ϕ is slowly varying, particularly as compared with the exponential. Thus, taking the curl of the magnetic field and discarding the derivatives of ϕ we find that the dominant component of the electric field is $E_z = -Z_0\phi e^{ikx}$, where Z_0 is the intrinsic impedance of space. The corresponding component of Poynting's vector is then $S_x = Z_0|\phi|^2 \exp[-2\mathrm{Im}(k)x]$, and it follows that for each x the square-norm of $\phi(x,z)$ is directly related to the dominant part of the total flow of power across the vertical plane at x.

We can now restate our problem in operator theoretic terms. We seek a family $\{T_x;\ x>0\}$ of linear operators on the space of square-integrable functions u such that for each u

$$\frac{d}{dx}T_x u = ikAT_x u \tag{10}$$

and

$$\lim_{x\to 0+} T_x u = u \tag{11}$$

where A is the unbounded operator defined, within its domain, by

$$Au(z) = \frac{1}{2k^2}u''(z) + M(z)u(z) \tag{12}$$

the primes denoting differentiation with respect to z. The domain of A, which we shall call D(A), is the set of functions on which A is defined. Its description is as important for the definition of A as is the formula (12). It consists of those functions u whose first derivatives are absolutely continuous functions (i.e., functions that equal the integrals of their derivatives), for which both u and Au are square-integrable, and for which

$$u'(0) + ikZu(0) = 0. \tag{13}$$

As is usually the case in such operator theoretic formulations, the boundary conditions have been incorporated into the definition of D(A). Note that the condition at infinity has been taken care of by the simple requirement that u be square-integrable. In the end, A is a "closed" linear operator whose domain is a linear manifold that is dense in the space of all square-integrable functions.

An equation such as (10) is sometimes called an "evolution" equation since one can imagine watching the solution unfold as the distance x increases. It is very much like a system of first order ordinary differential equations, and one refers to the function u of (11) as the <u>initial value</u>. One even expects to be able to write $T_x = e^{ikxA}$, if only one can ascribe a meaning to the exponential. In this same vein there is a second interpretation one can give the operator T_{x+s} when x and s are both positive: one imagines, as the solution unfolds, stopping at the distance s to mark the solution and then continuing on for a distance x using now the new initial value. Consequently, one expects $T_{x+s} = T_x T_s$, and one speaks of the family $\{T_x;\ x>0\}$ as a <u>one-parameter semi-group</u> (the inverse required to form a group is absent) of transformations. If such is the case the operator ikA is called the <u>infinitesimal generator</u> of the semi-group.

Assuming the T_x exist, there are some further preliminary remarks we can make. In particular, consider the magnitudes of the T_x--i.e., the norms. Because of (11) we would suppose that $\|T_x u\| \to \|u\|$ as $x \to 0$. Furthermore,

$$\begin{aligned} \frac{d}{dx}\,|T_x|^2 &= \frac{d}{dx}(T_x u, T_x u) \\ &= (ikAT_x u, T_x u) + (T_x u, ikAT_x u) \\ &= 2\,\mathrm{Re}(ikAT_x u, T_x u), \end{aligned} \tag{14}$$

so that one expects the norm to be a smooth, continuously differentiable function whose initial value is simply $\|u\|$.

Whether or not the T_x exist, the expression Re(ikAu,u) is interesting in its own right. It is a quadratic form in u that evidently represents the additional power flow across a vertical plane that is over and above what we have called the "dominant" component given by the square-norm itself. Integrating by parts, we find that for any u in D(A)

$$\begin{aligned} (Au,u) &= \int_o^\infty \left[\frac{1}{2k^2} u''(z) + M(z)u(z) \right] u^*(z)\, dz \\ &= \frac{1}{2k^2} u'(z)u^*(z)\Big|_o^\infty - \frac{1}{2k^2}\int_o^\infty |u'|^2 dz + \int_o^\infty M|u|^2\, dz \end{aligned} \tag{15}$$

Using the boundary conditions (since u is in D(A)), the first term here can be replaced by $(iZ/2k)|u(0)|^2$ and we find

$$\begin{aligned} \mathrm{Re}(ikAu,u) &= -\frac{1}{2}\mathrm{Re}(Z)|u(0)|^2 - \frac{\sin\delta}{2k_o}\int_o^\infty |u'|^2 dz - k_o \sin\delta \int_o^\infty M|u|^2 dz \\ &\leq 0, \end{aligned} \tag{16}$$

where the inequality arises because of the restrictions on δ, Z, and N. This is an important inequality that we shall soon use to great effectiveness. It seems to say that as distance increases, power is always lost, both to the ground and to the conductive atmosphere.

For now, we can note from (14) that if the $T_x u$ exist, their norms are always nonincreasing so that $\|T_x u\| \leqq \|u\|$ for all $x > 0$. The T_x are "norm decreasing." Recalling that the norm of a bounded linear operator is given by $\|T\| = \sup \|Tu\|/\|u\|$, we have

$$\|T_x\| \leqq 1, \qquad x > 0. \tag{17}$$

In trying to solve (10) one tool that comes to mind is the Laplace transform. If we are given the T_x we can then develop a second family of operators

$$R_\lambda = \int_0^\infty e^{-ik\lambda x} T_x dx, \tag{18}$$

where λ may be complex. We would surely expect this integral to converge whenever λ is such that $\mathrm{Re}(ik\lambda) > 0$; indeed, from (17) there would follow

$$\|R_\lambda\| \leqq \int_0^\infty |e^{-ik\lambda x}| \, dx \leqq 1/\mathrm{Re}(ik\lambda) \tag{19}$$

To obtain a more direct formula for R_λ we apply ikA to both sides of (18), finding

$$\begin{aligned} ikAR_\lambda &= \int_0^\infty e^{-ik\lambda x} \frac{d}{dx} T_x \, dx \\ &= e^{-ik\lambda x} T_x \Big|_0^\infty + ik\lambda \int_0^\infty e^{-ik\lambda x} T_x dx \\ &= -1 + ik\lambda R_\lambda, \end{aligned} \tag{20}$$

whence

$$ikR_\lambda = (\lambda - A)^{-1} \tag{21}$$

We are thus led to a scheme for finding the T_x: we first solve for the inverse operator in (21) as a function of λ and then use the inverse Laplace transformation. In pursuing this scheme it will be important to know for what values of λ the inverse operator does and does not exist; indeed, in the general theory of operators this knowledge is of great help in characterizing the operator A. Values of λ for which the inverse in (21) exists and is a bounded linear operator are said to belong to the <u>resolvent set</u> and in that case ikR_λ is the resolvent of A. Other values of λ--those for which the inverse does not exist or is unbounded--comprise the spectrum of A.

According to the general theory in which A is any closed operator, the resolvent set is an open subset of the complex plane and within it the resolvent is an analytic function of λ. In particular, R_λ has a derivative

$$\frac{d}{d\lambda}R_\lambda = -ikR_\lambda^2 \tag{22}$$

which we shall need later on.

At this point it is useful to introduce the adjoint A* of the operator A. (Note that in our notation the star can mean either the adjoint of an operator or the complex conjugate of a scalar. Note also that in (21) we have used the scalar λ to indicate the operator that simply multiplies a function by λ. Thus λ^* can denote an operator that is either the adjoint of λ or the product by the complex conjugate. Fortunately, the two are the same.) Since A is unbounded, so also will be A*. Its domain is given as the set of those functions v for which (Au,v) is a bounded linear functional as u varies throughout D(A); and when v is such a function, A*v is defined by (u,A*v) = (Au,v). Using the definition of A and integrating by parts twice, we find that for any u in D(A)

$$\begin{aligned}(Au,v) &= \int_0^\infty [\frac{1}{2k^2} u'' + Mu]\, v^*\, dz \\ &= \frac{1}{2k^2}(u'v^* - uv^{*\prime})\Big|_0^\infty + \int_0^\infty \frac{1}{2k^2}uv^{*\prime\prime}dz + \int_0^\infty Muv^*dz \\ &= \frac{1}{2k^2}\, u(0)(v'(0) - ik^*Z^*v(0))^* + \int_0^\infty u\, [\frac{1}{2k^{*2}}\, v'' + Mv]^*\, dz\end{aligned} \tag{23}$$

Since the first term in this last expression is not bounded in u unless it vanishes identically, it will follow that A* is the differential operator given by

$$A^*v(z) = \frac{1}{2k^{*2}}v''(z) + M(z)v(z) \tag{24}$$

where the domain D(A*) is the class of all functions v with absolutely continuous first derivatives, for which both v and A*v are square-integrable, and for which

$$v'(0) - ik^*Z^*v(0) = 0. \tag{25}$$

Comparing this with the definition of A, we note that A* is just the same operator except that all coefficients and parametric values have been replaced by their complex conjugates. One interesting way to say this is that the problem adjoint to the original one is the same as the original except that the opposite time convention is used. A consequence is that whatever we can say about the characteristics of A will also be true of those of A* provided all quantities involved are replaced by their complex conjugates.

With these preparations in hand we are already in a position to demonstrate the existence of a solution to (10), our major technical tool being the inequality (16). Suppose that λ is any complex number satisfying $\mathrm{Re}(ik\lambda)>0$. We can then show that λ is in the resolvent set. First, $\lambda - A$ is one-to-one; for otherwise there would exist a function $v \neq 0$ such that $(\lambda - A)v = 0$. But then, because of (16),

$$\mathrm{Re}(ik\lambda)\|v\|^2 = \mathrm{Re}(ik\lambda v,v) = \mathrm{Re}(ikAv,v) \leqq 0, \tag{26}$$

which is impossible. Second, the range of $\lambda - A$ is dense in the space of square-integrable functions; for otherwise there would exist a function $v \neq 0$ such that $((\lambda-A)u,v) = 0$ for all u in D(A). But this would imply that v is in D(A*) and that $(\lambda^*-A^*)v = 0$. From our remarks concerning A* and complex conjugates, this would in turn imply that $(\lambda-A)v^* = 0$, which we have already seen is impossible. Thus $(\lambda-A)^{-1}$ exists and is defined on a dense linear manifold. Finally, we must show it is bounded. Let u be in the range of $\lambda-A$, and let v be the unique function satisfying $ik(\lambda-A)v = u$. Then again because of (16)

$$\mathrm{Re}(ik\lambda)\|v\|^2 = \mathrm{Re}(ikAv + u,v) \leq \mathrm{Re}(u,v) \leq \|u\| \; \|v\|, \tag{27}$$

so that $\|v\| \leq \|u\|/\mathrm{Re}(ik\lambda)$. It follows that the correspondence of u with v defines the bounded linear operator R_λ which satisfies (21) (and, incidentally, also (19)) and is a one-to-one correspondence of the entire space of square-integrable functions with D(A).

Thus all of the lower half-plane defined by $\mathrm{Re}(ik\lambda) > 0$ lies within the resolvent set and the resolvent satisfies the inequality (19). These are exactly the requirements of the basic Hille-Yosida theorem concerning semi-groups of operators (see Dunford and Schwartz, 1958, or Hille and Phillips, 1957). The consequences are that ikA is indeed the infinitesimal generator of a semi-group $\{T_x\}$ and that (10), (11), and also (17) are satisfied. One minor shortcoming here (which we shall emend shortly) is that we are assured that $T_x u$ is in D(A) only when u is already in D(A)--thus (10) is assured only for u in D(A).

From the general theory there are also available to us several ways to represent the solution T_x. These are all in terms of R_λ where λ is in the lower half-plane described above. In one of these representations we have

$$\begin{aligned} T_x u &= \lim_{n\to\infty} \left(\frac{n}{x}\right)^n R^n_{n/ikx} u \\ &= \lim_{n\to\infty} \left(1 - \frac{ikx}{n} A\right)^{-n} u, \end{aligned} \tag{28}$$

for each square-integrable u. The second formula here is merely a restatement of the first using the definition in (21). The first formula imitates one of the lesser known representations of the inverse Laplace transform, while the second formula shows how T_x may be interpreted as an exponential function. These formulas are probably not very useful for computations, and we would prefer the more standard representation of the inverse transformation. There are difficulties, but at this point we can say that when u is in D(A)

$$T_x u = \lim_{t\to\infty} \frac{k}{2\pi} \int_{L_t} e^{ik\lambda x} R_\lambda u \, d\lambda \tag{29}$$

where L_t is the straight line contour extending from $-te^{-i\delta} - i\sigma$ to $te^{-i\delta} - i\sigma$ with $\sigma > 0$. In general, the convergence here is very weak and the restriction on u and the use of the Cauchy principal value are both necessary.

In passing, we might note that these results do not depend on the strict inequalities $\gamma > 0$, $\delta > 0$. If either or both of these constants vanish, then the inequality (16) remains valid and the existence of the T_x as bounded operators on Hilbert space safisfying (17) is still assured. We shall have more to say about this later on.

When $\delta > 0$ we can sharpen (16) somewhat and then obtain considerably stronger results for the T_x. Let τ be a small positive or negative number; then as in (16) we may write

$$\mathrm{Re}(e^{i\tau}ikAu,u) = -\frac{1}{2}\,\mathrm{Re}(e^{i\tau}Z)\,|u(0)|^2 - \frac{\sin(\delta-\tau)}{2k_o}\int_0^\infty |u'|^2 dz - k_o\sin(\delta+\tau)\int_0^\infty M|u|^2 dz. \tag{30}$$

Now the second and third terms here are still negative provided $|\tau| < \delta$. The first term is also nonpositive if either $Z = 0$ or $|\tau|$ is small enough and $\mathrm{Re}(Z) > 0$. Even if Z is pure imaginary we can show that the expression above is negative when $|\tau|$ is small enough by using inequalities such as

$$\begin{aligned}|u(0)|^2 &= -\int_0^\infty \frac{d}{dz}|u(z)|^2\,dz = -2\,\mathrm{Re}(u',u)\\ &\leqq 2\|u'\|\,\|u\| \leqq (1/k_o)\|u'\|^2 + k_o\|u\|^2.\end{aligned} \tag{31}$$

Then any tendency for the first term to be positive will be cancelled out by the other two terms. In short, we can always find a $\tau_o > 0$ such that

$$\mathrm{Re}(e^{i\tau}ikAu,u) \leqq 0 \tag{32}$$

whenever u is in $D(A)$ and $|\tau| < \tau_o$.

Let S be the sector in the complex λ-plane that has its vertex at the origin and that contains all λ with $-\pi-\delta-\tau_o < \arg\lambda < -\delta+\tau_o$. Then S subtends an angle of $\pi + 2\tau_o$ and contains the lower half-plane $\mathrm{Re}(ik\lambda)>0$ within its interior. Clearly, S defines the region within which $\mathrm{Re}(e^{i\tau}ik\lambda)>0$ for all τ, $|\tau| < \tau_o$. Proceeding as we did before in developing the inequalities (26) and (27), it is straightforward to show that S lies within the resolvent set and that for any λ in S

$$\|R_\lambda\| \leqq 1/k_o d(\lambda) \tag{33}$$

where $d(\lambda)$ is the distance from the point λ to the boundary of S.

Although we have enlarged the known part of the resolvent set only slightly and although the inequality (33) is only a little stronger than (19), the consequences are striking. The T_x now form what is called an "analytical" semi-group (see Hille and Phillips, 1958, Ch. 11) in which these operators can be analytically continued into a portion of the complex x-plane. In particular, consider the difficult formula in (29). The contour there may now be deformed into a contour Γ consisting of two rays that meet near the origin and that lie above L_t making positive angles (say $\tau_o/2$) with it. The inequality

(33) will assure that integrals over the two small arcs joining the end points will vanish as t goes to infinity. In the resulting integral the term $\exp(ik\lambda x)$ is now exponentially decreasing as λ goes to infinity along either leg of Γ, and the integral is therefore absolutely convergent. We may drop the requirement for the principal value and also any restrictions on the operand u. There results the satisfyingly simple expression

$$T_x = \frac{k}{2\pi} \int_\Gamma e^{ik\lambda x} R_\lambda d\lambda \tag{34}$$

which is valid for all $x > 0$. Further consequences are that $T_x u$ is in D(A) for all $x > 0$ and all square-integrable functions u, that (10) is always satisfied, that, indeed, $T_x u$ is infinitely differentiable and belongs to $D(A^n)$, for all n, and that T_x is continuous in the "uniform operator topology." This last statement means $\|T_s - T_x\|$ tends to zero as s approaches x from either side; it is not true when $x = 0$.

3. THE RESOLVENT

With almost nothing in the way of hard analysis, we have derived several important results concerning the T_x. Further progress, however, requires a more detailed examination of the operator A. In particular, we need a more nearly explicit representation of the resolvent.

Given the square-integrable function u we want to find an expression for the function $v = R_\lambda u$. This means we want to solve the equation $ik(\lambda-A)v = u$, or equivalently to find a function v in D(A) that satisfies the ordinary differential equation

$$v''(z) + 2k^2(M(z)-\lambda)v(z) = 2iku(z). \tag{35}$$

Of course, (35) always has a unique solution to the "initial value problem" when values of v and v' are prescribed at some particular value of z (see, e.g., Coddington and Levinson, 1955). The question for us will be whether such a solution is in D(A). But any solution to (35) will have absolutely continuous first derivatives, and all that remains is to assure that the boundary conditions at the earth's surface and at infinity are satisfied.

To solve this "boundary value problem" we use the standard "variation of constants" method. The first step here is to find suitable solutions to the "homogeneous" equation where in (35) the function u is replaced by zero. In particular, consider what happens at infinity. We recall that for $z > z_a$ we have $N(z) = 0$, so that the homogeneous equation becomes

$$v''(z) + 2k^2(\gamma z-\lambda)v(z) = 0, \qquad z > z_a \tag{36}$$

A simple change of variables converts this to Airy's equation

$$\frac{d^2}{d\theta^2} w(\theta) - \theta w(\theta) = 0 \tag{37}$$

which is one of the standard second order differential equations of mathematical physics. Since the coefficients are analytic functions of θ, the equation and its solutions may be extended to the complex plane. Indeed, since the equation has no singular points in the finite plane, any solution of (37) will be an "entire" function--i.e., a function that is analytic in the entire complex θ-plane.

Because Airy's equation and its solutions are so important to us, we should like to pause here to discuss the related notation. The standard solutions to (37) are the Airy functions $Ai(\theta)$, $Bi(\theta)$. Their definitions and a list of many of their properties may be found in Abramowitz and Stegun (1964), for example. They are linearly independent and "real" in that they have real values when θ is real. The function Ai, in particular, is widely used in the study of caustics and of the general theory of asymptotics involving "turning points" of differential equations.

It would seem useful to supplement the two standard solutions with others that resemble "traveling waves" when θ is real and negative. This has been done by many authors, but always in an *ad hoc* way.

Here, we would like to suggest a more formal notation that expands upon the standard notation described above and imitates what is done to define the Hankel functions. In this notation we would write

$$\begin{aligned}
Wi^{(1)}(\theta) &= Ai(\theta) - iBi(\theta) = 2e^{-i\pi/3}Ai(e^{i2\pi/3}\theta) \\
&= e^{-i\pi/6}(-\theta/3)^{1/2}H^{(1)}_{1/3}(\tfrac{2}{3}(-\theta)^{3/2}) \\
Wi^{(2)}(\theta) &= Ai(\theta) + iBi(\theta) = 2e^{i\pi/3}Ai(e^{-i2\pi/3}\theta) \\
&= (Wi^{(1)}(\theta^*))^*
\end{aligned} \tag{38}$$

These two functions, which we might call "Airy functions of the third kind." are linearly independent solutions of (37). Except for multiplicative constants they correspond to what Fock (1965) calls w_1 and w_2 and to what Wait (1962) calls w_2 and w_1.

One property of these functions that will be important to us concerns their asymptotic behavior as θ becomes large. In the almost complete circle defined by $-5\pi/3 < \arg\theta < \pi/3$, it can be shown that

$$Wi^{(1)}(\theta) = -i\pi^{-1/2}\theta^{-1/4}e^{(2/3)\theta^{3/2}}(1 + O(\theta^{-3/2})) \tag{39}$$

where the remainder term is uniform so long as $\arg\theta$ remains bounded away from $\pi/3$ or $-5\pi/3$. Roughly speaking, $Wi^{(1)}$ is either exponentially large or exponentially small, depending on whether $Re(\theta^{3/2})$ is positive or negative. The radials at $\arg\theta = -5\pi/3, -\pi, -\pi/3$, and $\pi/3$ mark where $Re(\theta^{3/2}) = 0$ and separate the plane into three equal sectors within each of which the sign of this quantity is constant. Only in the sector $-\pi < \arg\theta < -\pi/3$ is the function exponentially small; in the other two it is exponentially large. At the crack where $\arg\theta = \pi/3$ (or $-5\pi/3$) one can picture two representations obtained by continuing (39) up to and beyond the two boundaries. These representations differ, particularly since their exponents have opposite signs, and the proper asymptotic expansion in this region is simply the sum of the two. Qualitatively, the appearance of $Wi^{(1)}$ along this radial is that of a "standing wave." Using complex conjugates as suggested in (38), we find immediately that $Wi^{(2)}$ is exponentially small in the sector $\pi/3 < \arg\theta < \pi$ and has the appearance of a standing wave along the radial $\arg\theta = -\pi/3$.

Returning to the problem posed in (36), if we set

$$\theta(z) = (2k^2/\gamma^2)^{1/3}(\lambda-\gamma z) \tag{40}$$

then two linearly independent solutions are

$$w_+(\lambda;z) = Wi^{(1)}(\theta(z))$$
$$w_-(\lambda;z) = Wi^{(2)}(\theta(z)) \tag{41}$$

As z tends to infinity, arg $\theta(z)$ tends to $-\pi + 2\delta/3$; thus, no matter what complex value λ may have, $\theta(z)$ eventually takes on values in the third of the complex plane where $Wi^{(1)}$ is exponentially small. Indeed, for fixed λ we have

$$\frac{2}{3}\,\theta^{3/2} = \frac{2}{3}\sqrt{2\Upsilon}\,ikz^{3/2} - ik\lambda\sqrt{2z/\Upsilon} + O(z^{-1/2}) \tag{42}$$

and since the lead term here has a negative real part, it will follow from (39) that w_+ is eventually exponentially decreasing. On the contrary, w_- increases exponentially. As the notation indicates, w_+ resembles an upgoing wave while w_- resembles a downgoing wave.

We are now in a position to describe the solution to the resolvent equation (35). We first find (which is computationally the hard part!) two functions $g_o(\lambda;z)$, $g(\lambda;z)$ satisfying the homogeneous part of (35) with prescribed initial values. For definiteness we suppose

$$\begin{aligned} &g_o(0) = 1, \qquad g'_o(0) = -ikZ \\ &g(z) = w_+(z) \qquad \text{for } z > z_a \end{aligned} \tag{43}$$

Thus g_o satisfies the boundary condition (13) and g is exponentially decreasing to zero for large z.

Since the initial values and the differential equation are analytic in the parameter λ, it will follow that for each z, these two are entire functions in λ. They will be related in that their Wronskian

$$\rho(\lambda) = g_o(z)\,g'(z) - g(z)\,g'_o(z) \tag{44}$$

is independent of z. In particular, setting z = 0 and using (43), we have

$$\rho(\lambda) = g'(0) + ikZg(0) \tag{45}$$

Of course, the two functions are linearly independent if and only if their Wronskian $\rho(\lambda)$ does not vanish.

We can now construct a "Green's function"

$$G(\lambda;z,\zeta) = \begin{cases} g(z)\ g_0(\zeta) & 0 \leq \zeta \leq z \\ g_0(z)\ g(\zeta) & 0 \leq z \leq \zeta \end{cases} \tag{46}$$

in terms of which we would want to define a linear operator

$$\begin{aligned} G_\lambda u(z) &= \int_0^\infty G(\lambda;z,\zeta)u(\zeta)\ \ d\zeta \\ &= g(z) \int_0^z g_0(\zeta)u(\zeta)\ d\zeta + g_0(z) \int_z^\infty g(\zeta)u(\zeta)\ \ d\zeta \end{aligned} \tag{47}$$

The function G is continuous in its three variables and, for fixed z and ζ, an entire function in λ. That it is symmetric in z and ζ is an expression of the "law of reciprocity"; but that it is not Hermitian symmetric shows once again that our problem is not self-adjoint. At $z = \zeta$ the first derivative of G with respect to either z or ζ has a jump discontinuity whose size is exactly the Wronskian $\rho(\lambda)$.

For fixed z, G is exponentially decreasing in ζ and therefore square-integrable. It follows that when u is square-integrable the function $v(z) = G_\lambda u(z)$ exists and is finite for each $z > 0$. Evaluating the first two derivatives of (47), it can be seen (1) that v satisfies (35) provided the right-hand side is replaced by $\rho(\lambda)u(z)$, and (2) that v satisfies the boundary condition (13). There remains the niggling question as to how v behaves for large z. Even this has an immediate answer provided u belongs to the class of square-integrable functions that vanish for all sufficiently large z; for then (for that same sufficiently large z) v is simply a multiple of g and hence is exponentially small. For this special class of functions we find that $G_\lambda u$ is in D(A) and that

$$ik(\lambda-A)G_\lambda u = \frac{\rho(\lambda)}{2ik}\ u \tag{48}$$

To further refine this statement, it is now useful to show

$$\int_0^\infty \int_0^\infty |G(\lambda;z,\zeta)|^2 d\zeta dz = 2\int_0^\infty |g(z)|^2 \int_0^z |g_0(\zeta)|^2 d\zeta dz < \infty \tag{49}$$

for then it will follow that G_λ is a bounded operator that can be applied to all square-integrable functions. Toward this end, we first note there is no difficulty with the integral in the finite portion of the (z, ζ)-plane. Indeed, if either of the two variables remains bounded, G is exponentially small in the other and convergence is assured. It is only along the diagonal $z = \zeta$ that we may find trouble. It will therefore suffice if we can show that (49) holds when both lower limits are replaced by a value η that is at our disposal. We suppose that η is, first, greater than z_a and, second, large enough so that asymptotic estimates of w_+ and w_- are fairly accurate. Because of the first criterion, it will follow that within the integral g equals w_+ and g_o is some particular linear combination of w_+ and w_-. The part of g_o that is a multiple of w_+ will not cause problems since, again, this part will be exponentially decreasing. There remains only to show

$$\int_\eta^\infty |w_+(z)|^2 \int_\eta^z |w_-(\zeta)|^2 d\zeta dz < \infty \tag{50}$$

Let $\sigma(z) = 2\,\mathrm{Re}(-(2/3)\theta(z)^{3/2})$. Then for large z we have

$$\begin{aligned}
&\sigma(z) = \frac{4}{3}\sqrt{2\gamma}\,k_o \sin\delta\, z^{3/2} + O(z^{1/2}) \\
&\sigma'(z) = 2\sqrt{2\gamma}\,k_o \sin\delta\, z^{1/2} + O(z^{-1/2}) \\
&|w_+(z)|^2 \sim c_1 z^{-1/2} e^{-\sigma}, \qquad |w_-(z)|^2 \sim c_1 z^{-1/2} e^{\sigma} \\
&\int_\eta^z |w_-(\zeta)|^2 d\zeta \sim c_1 z^{-1/2} \sigma'^{-1} e^{\sigma} \sim c_2 z^{-1} e^{\sigma}
\end{aligned} \tag{51}$$

where c_1 and c_2 are constants depending on λ. As a function of z the integrand in (50) is $O(z^{-3/2})$ and hence is integrable, although just barely.

We have thus demonstrated that the integral in (49) is finite and that, therefore, for each complex λ the operator G_λ is bounded. (Actually, we have even shown the stronger property that G_λ is "compact.") Thus (48) is valid for all square-integrable u, and consequently the resolvent exists for all λ where $\rho(\lambda) \neq 0$ and

$$R_\lambda = \frac{2ik}{\rho(\lambda)} G_\lambda \tag{52}$$

We also note that as an operator-valued function of λ, G_λ will be entire. Thus R_λ is the quotient of two entire functions and hence is "meromorphic"-- i.e., it is analytic everywhere in the complex plane except at certain isolated poles.

This rather straightforward behavior of the function R_λ can be contrasted with the situations that arise when either γ or δ is allowed to vanish. In the case of a flat earth when $\gamma = 0$, the functions in (41) become meaningless and the subsequent analysis invalid. It is, however, easy enough to solve (36) for this case; but note that $\lambda^{1/2}$ appears and that it is impossible to find a solution such as w_+ which decreases to zero for large z and which is entire in λ. Indeed, it is quickly seen that the entire ray defined by $\arg \lambda = \pi - 2\delta$ belongs to the spectrum of A, thus introducing the well-known "branch cut" into consideration.

On the other hand, when $\gamma > 0$ but $\delta = 0$ (so that k is real), the functions in (41) still define solutions to (36); it is only the subsequent asymptotics that are false. The first term in (42) is pure imaginary but we can still appeal to the second term to show that g is again exponentially decreasing provided, however, that $\mathrm{Re}(ik\lambda) > 0$ --i.e., that λ lies in the lower half-plane. When λ is in the upper half-plane, we can still attempt to find the resolvent in terms similar to (52) and (47). But now we cannot define g as in (42); instead we must assume $g(z) = w_-(z)$ for $z > z_a$, since now it is this function that is exponentially decreasing. This approach succeeds in the special case when $\mathrm{Re}(Z) = 0$, for then (16) becomes $\mathrm{Re}(ikAu,u) = 0$ and arguments similar to those of section 2 will show that R_λ exists and satisfies $\|R_\lambda\| \leq 1/|\mathrm{Re}(ik\lambda)|$. Thus R_λ exists on both sides of the real line but is described by two different analytic functions. The real line is a "natural boundary" separating the two branches and consequently the entire real line forms the spectrum of A.

This peculiar result is partly an artifact of our choice of function space in which to operate. Note, indeed, that when k is real and Z pure imaginary then the operator A is, according to the rules of that function space, self-adjoint. This seems to be one example of where being able to say an operator is self-adjoint is of little help.

Returning to the straightforward case where both γ and δ are positive, consider what happens when λ satisfies $\rho(\lambda) = 0$. Then because of (45), the function g satisfies the boundary condition at the earth's surface; it therefore belongs to D(A) and then $(\lambda - A)g = 0$, so that the operator $\lambda - A$ is not one to one. We say that λ is an eigenvalue (a modal value) and that g is an eigenfunction (a mode). The function

$$\phi(x,z) = g(\lambda;z)e^{ik\lambda x} \tag{53}$$

safisfies (2) and the boundary conditions, and its initial value is $g(z)$. More generally, if a proposed source function u can be expanded as a linear combination of several such eigenfunctions, then the final solution to our problem is the same linear combination of functions of the form (53).

There is a general approach to such linear combinations. Let B be a closed linear operator with adjoint B^*. Suppose $\lambda_1, \lambda_2, \ldots$ are distinct eigenvalues of B with corresponding eigenfunctions $\phi_1, \phi_2, \ldots$. Then $\lambda^*_1, \lambda^*_2, \ldots$ are in the spectrum of B^* and are very likely to be eigenvalues. Suppose they are and that $\psi_1, \psi_2, \ldots$ are corresponding eigenfunctions. Then

$$\begin{aligned}(\lambda_n - \lambda_m)(\phi_n,\psi_m) &= (\lambda_n\phi_n,\psi_m) - (\phi_n,\lambda^*_m\psi_m) \\ &= (B\phi_n,\psi_m) - (\phi_n,B^*\psi_m) = 0\end{aligned} \tag{54}$$

so that ϕ_n and ψ_m are orthogonal provided $n \neq m$. We say that $\{\phi_n\}$ and $\{\psi_m\}$ form a biorthogonal pair of sequences. Such a pair is not so powerful a tool as the usual single orthogonal sequence that plays an important role in the basic theory of Hilbert spaces. Nevertheless, we may note that if

$$u = \Sigma\, a_m\phi_m \tag{55}$$

then

$$(u,\psi_m) = a_m(\phi_m,\psi_m) \tag{56}$$

which may be solved for a_m provided $(\phi_m,\psi_m) \neq 0$, a condition that is not guaranteed.

In the case of the particular operator A, let us suppose that $\{\lambda_m\}$ is a set of distinct eigenvalues and that $g_m(z) = g(\lambda_m;z)$ are the corresponding eigenfunctions. Then indeed $\lambda_m{}^*$ are eigenvalues of A^* and the corresponding eigenfunctions are the complex conjugates $g_m{}^*$. It follows that

$$(g_n,g^*_m) = 0, \qquad n \neq m \tag{57}$$

and that if u is a linear combination of the g_m then

$$u = \Sigma\, (u,g^*_m)g_m/(g_m,g^*_m) \tag{58}$$

provided, of course, that none of the denominators vanish. Although the expression (u,g^*) seems a rather tortuous way to write $\int ugdz$, we shall retain that notation because it is a continual reminder of the relationship between the solution to our problem, the role the adjoint plays, and the particular form taken on by the adjoint.

Let us examine the conditions under which one of the denominators in (58) might vanish. Let μ be a particular eigenvalue; then the corresponding g_o will be a scalar multiple of g and, indeed, we may write

$$g_o(z) = g(z)/g(0) \tag{59}$$

(Note that $g(0)$ cannot vanish; for if it did then by (45) it would also be true that $g'(0)$ would vanish, whence g would vanish identically, contradicting (43).) It will then follow from (46) that

$$G(\mu;z,\zeta) = g(z)g(\zeta)/g(0) \tag{60}$$

On the other hand, substituting (52) in (22) we find

$$-\rho'(\lambda)G_\lambda + \rho(\lambda)\frac{d}{d\lambda}G_\lambda = 2k^2G_\lambda^2 \tag{61}$$

and, in particular,

$$\rho'(\mu)G_\mu = -2k^2G_\mu^2 \tag{62}$$

But from (60) we have

$$\begin{aligned} G_\mu^2 u &= (u,g^*)(g,g^*)g/g(0)^2 \\ &= \frac{(g,g^*)}{g(0)} G_\mu u \end{aligned} \tag{63}$$

whence

$$\rho'(\mu) = -2k^2(g,g^*)/g(0) \tag{64}$$

Thus (g,g^*) vanishes if and only if $\rho'(\mu)$ vanishes. Furthermore, if $\rho'(\mu)$

$\neq 0$ then from (52) we see that the Laurent series expansion of R_λ about μ has the form

$$R_\lambda = \frac{2ik}{\rho'(\mu)}(\lambda-\mu)^{-1}G_\mu + \ldots \tag{65}$$

so that (g,g^*) does not vanish if and only if μ is a simple pole of R_λ.

4. THE RESOLVENT FOR LARGE λ

Thus far we have said how the solution might appear if there were a sufficient set of eigenfunctions; we have not shown that even one exists. To remedy this omission we now consider what happens when the contour of the integral in (29) or (34) is deformed into a large semicircle in the upper half-plane, and we shall show that as it is deformed it must cross poles of the integrand.

Our first task in this process will be to determine the asymptotic behavior of R_λ and its components as λ becomes large. As a general note we remark that since λ appears in the differential equation (35) only in the term $N(z) + \gamma z - \lambda$, we might expect that when λ is large it completely dominates the expression and asymptotic results should be independent of the refractivity function $N(z)$. We would expect to obtain the same results as when N is, say, identically zero. As it turns out, this expectation is only partly satisfied.

First we consider the function $g_o(\lambda;z)$. Let

$$\alpha = \varepsilon k(2\lambda)^{1/2} \tag{66}$$

where $\varepsilon = \pm 1$, the sign being chosen so that $\mathrm{Re}(\alpha) \geqq 0$. Then g_o satisfies

$$\begin{aligned} e^{-\alpha z}g_o(z) &+ \frac{k^2}{\alpha}\int_0^z M(\zeta)(1 - e^{-2\alpha(z-\zeta)})e^{-\alpha\zeta}g_o(\zeta)d\zeta \\ &= \frac{1}{2}(1 + e^{-2\alpha z}) - \frac{ikZ}{2\alpha}(1 - e^{-2\alpha z}) \end{aligned} \tag{67}$$

since by direct differentiation and evaluation one can show that it would then satisfy the homogeneous part of (35) and the initial values (43). Now (67) can be treated as an integral equation for the function $e^{-\alpha z}g_o(z)$. As such the kernel is $M(\zeta)(1-e^{-2\alpha(z-\zeta)})$ and is uniformly bounded in α so long as $\mathrm{Re}(\alpha) \geqq 0$ is satisfied. Because of the coefficient α^{-1} in front of the integral, we see that when z is constrained to lie in some bounded interval the integral operator reduces to a small perturbation. It follows that

$$g_o(\lambda;z) = \tfrac{1}{2}e^{\alpha z}(1 + O(1/\alpha)) + \tfrac{1}{2}e^{-\alpha z}(1 + O(1/\alpha)) \tag{68}$$

where the remainder terms are uniform so long as z remains bounded.

Next we turn to the function $g(\lambda;z)$. Setting

$$g(z) = w_+(z) + r(z) \tag{69}$$

we note that the remainder r vanishes identically for $z > z_a$ and otherwise satisfies the inhomogeneous differential equation

$$r'' + 2k^2(M-\lambda)r = -2k^2Nw_+ \tag{70}$$

With α again as in (66) one may check that r also satisfies

$$\begin{aligned} &e^{\alpha z}r(z) + \frac{k^2}{\alpha}\int_z^{z_a} M(\zeta)(1 - e^{-2\alpha(\zeta-z)})e^{\alpha\zeta}r(\zeta)\,d\zeta \\ &\quad = -\frac{k^2}{\alpha}\int_z^{z_a} N(\zeta)(1 - e^{-2\alpha(\zeta-z)})e^{\alpha\zeta}w_+(\zeta)\,d\zeta \end{aligned} \tag{71}$$

which may be treated as an integral equation for the function $e^{\alpha z}r(z)$. As such we again find the kernel is bounded so long as $z < z_a$ (which is all that interests us anyway); as before the coefficient α^{-1} means that the integral operator will introduce only a small perturbation to the solution.

We need, then, to estimate the integral that forms the inhomogeneous part of (71). When $-5\pi/3 < \arg k^{2/3}\lambda < \pi/3$ we may use (39) to determine how w_+ behaves for large λ. From (40) we find

$$\begin{aligned} \tfrac{2}{3}\theta^{3/2} &= \beta - \varepsilon\alpha z + O(1/\alpha) \\ \beta &= (k/3\gamma)(2\lambda)^{3/2} \end{aligned} \tag{72}$$

where α and ε are as in (66). There follows

$$w_+(\lambda;z) = c_o(-\varepsilon\alpha)^{-1/2}e^{\beta-\varepsilon\alpha z}(1 + O(1/\alpha)) \qquad (73)$$

where c_o is the constant $(2k^2\gamma/\pi^3)^{1/6}$ and the remainder term is uniform for bounded z. The inhomogeneous part of (71) then becomes

$$c_o k^2\varepsilon(-\varepsilon\alpha)^{-3/2}e^{\beta}(1 + O(1/\alpha))\int_z^{z_a} N(\zeta)(1 - e^{-2\alpha(\zeta-z)})e^{(1-\varepsilon)\alpha\zeta}d\zeta \qquad (74)$$

Concerning ε, we note that it is positive when $|\arg k^2\lambda| \leqq \pi$ and negative in the adjacent intervals. It therefore seems useful to define two sectors of the λ-plane

$$\begin{aligned} S_1&: -\pi - 2\delta \leqq \arg\lambda < \frac{\pi}{3} - \frac{2\delta}{3} \\ S_2&: -\frac{5\pi}{3} - \frac{2\delta}{3} < \arg\lambda \leqq -\pi - 2\delta \end{aligned} \qquad (75)$$

Then ε is 1 in S_1 and -1 in S_2. To complete coverage of the plane we would also define a third sector S_o which includes a neighborhood of the ray $\arg\lambda = \pi/3-2\delta/3$.

In the sector S_1, where $\varepsilon = 1$, the last exponential under the integral in (74) disappears and the integral is a bounded function of ζ. Thus the inhomogeneous part of (71), and hence also the function $e^{\alpha z}r$, is of order $e^{\beta}\alpha^{-3/2}$. It follows from (69) that in S_1

$$g = -ic_o\alpha^{-1/2}e^{\beta-\alpha z}(1 + O(1/\alpha)) \qquad (76)$$

This is just a small perturbation of the function w_+.

In S_2, however, that same exponential does not disappear. Indeed, it becomes a large and important factor, thus making the analysis more complicated. We can separate the expression (74) into two parts obtaining

$$\begin{aligned} &-c_o\alpha^{-1/2}e^{\beta+2\alpha z_a}Q(\alpha)(1 + O(1/\alpha)) \\ &\quad + c_o k^2\alpha^{-3/2}e^{\beta+2\alpha z}(1 + O(1/\alpha)) \int_0^{z_a} N(\zeta)e^{-2\alpha\max(z-\zeta,0)}d\zeta \end{aligned} \qquad (77)$$

where

$$Q(\alpha) = \frac{k^2}{\alpha} \int_0^{z_a} N(\zeta) e^{-2\alpha(z_a - \zeta)} d\zeta \qquad (78)$$

The second part of (77) is quickly disposed of. The integral it contains is bounded and hence the complete term is of order $\alpha^{-3/2} e^{\beta + 2\alpha z}$. From (71) the corresponding part of $re^{\alpha z}$ will be of the same order of magnitude and hence the corresponding part of r will be $1/\alpha$ times the order of magnitude of w_+. When combined with w_+ as in (69) this becomes just another of the second order terms.

We are left with only the first part of (77). Substituting this in (71) and (69) we find that for λ in S_2 and $z < z_a$

$$g = c_o \alpha^{-1/2} e^{\beta + \alpha z} [1 + O(1/\alpha) - e^{2\alpha(z_a - z)} Q(\alpha)(1 + O(1/\alpha))] \qquad (79)$$

The last element of this expression is the new one; it involves a large exponential factor and is not just a small perturbation of the original function w_+. One might argue, however, that this exponential is multiplied by what is evidently the small coefficient $Q(\alpha)$; so perhaps the term is not so large after all. Certainly, from the definition (78) one notes the integral there is bounded so that Q tends to zero at least as rapidly as $1/\alpha$. But more than this, the integral will probably also tend to zero, although just how rapidly will depend on differentiability properties of $N(z)$. Nevertheless, if one supposes that z_a is particularly chosen to be the smallest elevation with the property that $N(z)$ vanishes for $z > z_a$, then Q will never go to zero so fast as an exponential in α. Thus the last term in (79) will be large whenever $\mathrm{Re}(\alpha)$ is large.

Of course (79) shows how the field represented by g can be separated into an upgoing and a downgoing wave. What is perhaps surprising is that in some circumstances the downgoing wave dominates. Note, however, that when z approaches z_a or when $\mathrm{Re}(\alpha)$ approaches zero, then the exponential approaches a moderate size, the smallness of Q takes over, and this downgoing wave is absorbed into the other second-order terms.

When λ is in S_o the proper asymptotic expression for g is simply the sum of the expressions (76) and (79). In (79) the quantity β must be replaced by

$-\beta$ to account for the fact that we have changed arg λ by 2π.

We can now write down asymptotic expressions for the Wronskian $\rho(\lambda)$. Differentiating (76) and (79) and evaluating (45) we find

$$\rho(\lambda) = \begin{cases} ic_o\alpha^{1/2}e^{\beta}(1 + O(1/\alpha)) & \text{in } S_1 \\ c_o\alpha^{1/2}e^{\beta}[1 + O(1/\alpha) + e^{2\alpha z_a}Q(\alpha)(1 + O(1/\alpha))] & \text{in } S_2 \\ ic_o\alpha^{1/2}[e^{\beta}(1 + O(1/\alpha)) - ie^{-\beta}(1 + O(1/\alpha)) & \text{in } S_o \\ \quad -ie^{-\beta-2\alpha z_a}Q(\alpha)(1 + O(1/\alpha))] & \end{cases} \tag{80}$$

It is interesting to note that in these expressions the surface impedance Z has been absorbed in the second-order terms.

As a passing remark, we also note that dominating the expressions in (80) is the factor $e^{\pm\beta}$. Since $\beta = O(\lambda^{3/2})$ we see that ρ is an entire function of order 3/2. Because the order is not an integer it follows from Picard's "little theorem" (see, e.g., Copson, 1935, Ch. 7) that ρ attains all complex values infinitely often. In particular, it follows that ρ has an infinite number of zeros, thus answering at least one of our questions.

More to the point, perhaps, a closer examination of (80) yields some details concerning the zeros of ρ. There will be two sets of zeros along the two boundaries of S_2--i.e., one set in S_o and another near where arg $k^2\lambda = -\pi$ and $\text{Re}(\alpha) = 0$; there are no (large) zeros in S_1. In S_o it becomes a matter of solving

$$e^{2\beta} = i + O(1/\alpha) + ie^{2\alpha z_a}Q(\alpha)(1 + O(1/\alpha)) \tag{81}$$

when arg $k^{2/3}\lambda \approx \pi/3$ and $\text{Re}(\beta) \approx 0$. As functions of λ the term β varies more rapidly than does α, and the left-hand side of (81) varies much more rapidly than does the right-hand side. Choosing some approximate value for λ, we may evaluate the right-hand side (which will usually be large), then move λ towards S_1 so that the real part of β becomes large enough to make the magnitude of the left-hand side match that of the right, and then adjust the imaginary part so that the arguments (phases) match. Since this latter will happen infinitely often as β steps through values separated by about $i\pi$, we deduce an infinite sequence $\lambda_m^{(1)}$ of eigenvalues with the property

$$\lim_{m\to\infty} \lambda_m^{(1)}/m^{2/3} = \tfrac{1}{2}(3\pi\gamma/k)^{2/3}e^{i\pi/3} \tag{82}$$

At the other boundary of S_2 we need to solve an equation of the form

$$e^{2\alpha z_a}\, Q(\alpha) = -1 + O(1/\alpha) \tag{83}$$

Throughout most of S_2, $\mathrm{Re}(\alpha)$ is large and the exponential in (83) dominates making the left-hand side large. When, however, α is on the boundary of S_2 where $\mathrm{Re}(\alpha) = 0$ the smallness of Q makes the left-hand side small. Since the exponential varies much more rapidly than does Q, we may adjust λ so that first the real part of α becomes just large enough to offset the smallness of Q and then the imaginary part to match phases. Again, this latter will happen when αz_a steps through values separated by about $i\pi$. We find a second infinite sequence $\lambda_m^{(2)}$ of eigenvalues, this one such that

$$\lim_{m\to\infty} \lambda_m^{(2)}/m^2 = -\pi^2/2k^2z_a^2 \tag{84}$$

The first of the above series leads to the "creeping wave" (Ekersley) modes while the second leads to the "waveguide" (Gamow) modes. Note how the first series becomes more dense as m increases while successive elements of the second series become further and further apart. In the literature one reads also of "trapped" modes and "whispering gallery" modes. These are special modes whose associated eigenvalues are small or at most moderately large; they can be only finite in number. While they are often the most important terms in a modal expansion, they do not appear in our present considerations since we have restricted ourselves to those with large eigenvalues.

We have seen that except when $\rho(\lambda)$ vanishes the resolvent R_λ is an integral operator with kernel $R(\lambda;z,\zeta) = 2ikG(\lambda;z,\zeta)/\rho(\lambda)$. We can now combine the expansions of g_o, g, and ρ to obtain the asymptotic behavior of this kernel as λ becomes large. The results will be uniform in z and ζ provided, however, these two variables are constrained to lie within some bounded interval. As such an interval becomes larger, the expressions will be valid only if λ becomes sufficiently larger.

In S_1 we combine (68), (76), and (80) to obtain

$$R(\lambda;z,\zeta) = \frac{-ik}{\alpha}\,[e^{-\alpha|z-\zeta|}(1 + O(1/\alpha)) + e^{-\alpha(z+\zeta)}\,(1 + O(1/\alpha))] \quad (85)$$

which seems simple enough. If z and ζ are bounded away from 0 and arg α is bounded away from $-\pi/2$, then the second term (looking like a wave reflected at the earth's surface) disappears since it is exponentially smaller than the first. We might also note it is no accident that the dominant term $e^{-\alpha|z-\zeta|}$ looks like an approximation to the delta function. From (18) and (11) it follows that for any square-integrable u the expression $ik\lambda R_\lambda u$ tends to u when λ tends, say, to $-i\infty$.

When λ is in S_2 the results are more complicated. We express them here using only dominant terms, each of which should be multiplied by a factor $1 + O(1/\alpha)$. We also assume $\zeta \leqq z$ and appeal to symmetry for the contrary case. Assembling the previous results we find for λ in S_2 and $z < z_a$

$$R(\lambda;z,\zeta) = \frac{ik}{\alpha}\,[e^{\alpha|z-\zeta|} + e^{\alpha(z+\zeta)}]\,\frac{1 - e^{2\alpha(z_a-z)}\,Q(\alpha)}{1 + e^{2\alpha z_a}\,Q(\alpha)} \quad (86)$$

When $z \to z_a$, (86) is still valid except that the term in the numerator involving $Q(\alpha)$ will disappear. We should also note that we do not mean to exclude the "degenerate" case when the refractivity vanishes identically. Then Q will also vanish and corresponding terms in both numerator and denominator of (86) will disappear.

We are now ready to attack the inverse Laplace transform as given in (29) or (34). These formulas involve an integral in the λ-plane whose contours extend roughly from left to right and lie below any poles that the integrand might have. Our plan is to deform those contours into a rough semicircle C_t having radius t and extending into the upper half-plane. Then the original integral will equal the integral over C_t plus the residues at any poles the deformation will have had to cross. If we can show that the integral over C_t tends to zero as t goes to infinity, we will have the desired result: that $T_x u$ equals the sum of the residues at all poles of $R_\lambda u$.

If we replace R_λ by its representation as an integral operator, then our integral becomes an iterated integral in the variables λ and ζ and involves the kernel $R(\lambda;z,\zeta)$ that we have just been studying. The results of our study, however, are valid only when z and ζ both remain bounded. Before applying them we must limit ourselves to those situations in which this is true.

Let B_0 be the collection of all square-integrable functions each of which vanishes for sufficiently large z. These are the functions with "bounded support." Clearly, B_0 is a linear manifold that is dense in the space of square-integrable functions. Then if in (29) or (34) we restrict u to lie in B_0, the resulting integral over ζ is effectively a finite integral and we may treat ζ as bounded. To bound z similarly we simply restrict ourselves to consider the integral only for a fixed z. Then since large t implies that λ is large all along C_t, our asymptotic results for $R(\lambda;z,\zeta)$ will apply.

Note that with u in B_0 the two integrals of our iterated integral are both over finite intervals. The integrals may therefore be interchanged and we may first turn our attention toward estimating

$$\int_{C_t} R(\lambda;z,\zeta)e^{ik\lambda x}d\lambda \tag{87}$$

for fixed z and ζ. To help here, we recall that C_t lies in the upper half-plane so that when $x > 0$ the exponential in (87) has a negative real part--except perhaps at the two end points.

The contours L_t of (29) and Γ of (34) both lie within the sector S_1. Following along C_t one finds a small arc at the beginning and about one-third of the semicircle at the end that lie in S_1. From (85) it follows that the resolvent kernel R is $O(1/\alpha)$ and hence, by Jordan's lemma, that these portions of (87) tend to zero with increasing t.

About two-thirds of C_t lies in S_2. In this sector, as one sees from (86), the function R can be exponentially large. The worst case (when R is largest) comes about in the degenerate situation when $Q(\alpha)$ vanishes identically; then R is of order $\alpha^{-1}\exp(\alpha(z+\zeta))$. But α is $O(\lambda^{1/2})$ and hence for large enough λ the integrand in (87) is still dominated by the negative exponential $\exp(ik\lambda x)$. This suffices to show that this portion of (87) also tends to zero as t increases.

There remain to be considered the sector S_0 and a small sector at the other boundary of S_2 where $\mathrm{Re}(\alpha) = 0$. In these two sectors there are points where $\rho(\lambda)$ vanishes and the function R has poles. For large λ the poles thus form two picket lines running radially out from the origin. Clearly one must deform the contours C_t somewhat so that they cross these lines transversely at about midway between two of the successive poles. Furthermore, when a C_t crosses either of the sectors in that fashion $\rho(\lambda)$ will remain large and $R(\lambda;z,\zeta)$ relatively small. For example, at the lower edge of S_2 where $Re(\alpha)$ is nearly zero when one crosses midway between poles, the denominator of (86) will be of order unity. Our previous estimates of the magnitude of the function R within S_2 will still be valid. Similarly, as C_t crosses S_o the minimum magnitude of $\rho(\lambda)$ will be $O(\lambda^{1/4})$ which is still large. Both lines of poles can be crossed in such a way as to keep the function R within bounds, and the negative exponential in (87) will assure that these portions also tend to zero as t increases.

Thus all portions of (87) tend to zero and clearly the convergence is uniform as long as z and ζ remain bounded. The iterated integral that defines the inverse Laplace transform also tends to zero and we have left only the residues at the poles of the function R. If we assume the poles are all simple poles and that their locations λ_m, m = 1,2,. . ., are numbered in order of increasing magnitude, then from (29), (58), (64), and (65) we see that when u is in B_o we have

$$T_x u(z) = -2k^2 \sum_{m=1}^{\infty} \frac{(u,g_m^*)}{\rho'(\lambda_m)g_m(0)} g_m(z)e^{ik\lambda_m x} \tag{88}$$

Concerning what multiplicity the poles have, we may note that when λ_m is large our asymptotic results show that the associated poles certainly are simple. If some of the smaller values of m correspond to multiple poles, the corresponding terms in (88) must be replaced by the proper residues. These will involve factors of the kind $x^j \exp(ik\lambda_m x)$.

The result in (88) provides for point-wise convergence only, although clearly we also have uniform convergence for bounded z. In the infinite interval, however, we cannot expect uniform convergence nor any kind of convergence in the mean. Indeed, consider one of the creeping wave modes $g_m(z)$. Its global behavior is determined fairly well by the function $\theta(z)$ defined in (40). As z goes from 0 to increasingly higher heights, θ begins high up in the first quadrant with arg $\theta \approx \pi/3$ and moves along a straight line to the left and downward (because of the positive conductivity we have imposed), first passing into the second quadrant and then eventually into the third. Thus when $\gamma z \gg |\lambda_m|$ the function g_m is exponentially small and when $\gamma z \ll |\lambda_m|$ it resembles a standing wave of only moderately large size. In between, however, $\theta(z)$ is high up in the second quadrant and $g_m(z)$ will be very large. In particular, it is clear that $\|g_m\|$ tends rapidly to infinity with increasing m. Nor can we expect help from the other terms in (88). The denominators, for example, are only moderately large since they pertain to properties of g_m at $z = 0$.

Of course, from the practical point of view of numerical evaluation, the result in (88) seems quite sufficient. That we require u to be in B_o--in other words, that the source be of finite extent--does not seem like much of a restriction; and computations will normally involve only a finite number of values of z. It is only when one wants to use the series in a further analysis that one must be wary.

Nevertheless, it is a curiosity worth mentioning that (88) is not a rapidly converging series. For example, consider only the creeping wave modes. These make up a subset of the terms of the series and consist of coefficients multiplying the exponentials $\exp(ik\lambda_m x)$ where the λ_m grow at a rate proportional to $m^{2/3}$. Estimating the coefficients is difficult since they seem to be subject to important second-order effects. But note that the proof of (88) definitely required $x > 0$; if one simply sets $x = 0$ one expects that the resulting series of coefficients alone will diverge. Indeed, under some circumstances the coefficients will actually grow in magnitude at a rate of about $\exp(bm^{1/3})$ where b is positive. Convergence is therefore controlled by the exponentials whose magnitudes are dominated by a factor $\exp(-am^{2/3})$ where a is positive. This factor tends to zero faster than any power of 1/m, but it also fails the ratio test: the ratio between successive factors tends to unity. This means that the series does not converge as rapidly as a geometric series. If x is small and more than a few creeping wave modes are needed to obtain the desired accuracy, then one can expect that a great many of these modes will be required.

Fortunately, a modal expansion is normally used for computations only when x is large and z is small so that the points of interest are all well beyond the horizon. Fortunately also, the accuracy one requires is not great; one creeping wave mode often suffices. Indeed, when ducting occurs it is often true that no creeping wave modes--and even no waveguide modes--are needed, for the miscellaneous smaller eigenvalues have very small imaginary parts and contribute the only major terms.

The result in (88) can be expressed in functional analytic form. Let $S_x^{(n)}u$ represent the sum of the first n terms of the series. Then, using our comments about uniform convergence for bounded z, we have that for any u and v in B_0

$$\lim_{n\to\infty} (S_x^{(n)}u,v) = (T_x u,v) \tag{89}$$

We say that $S_x^{(n)}u$ converges to $T_x u$ "in the weak topology generated by B_0" and that $S_x^{(n)}$ tends to T_x "in the weak operator topology." Each of the $S_x^{(n)}$ is, of course, a bounded linear operator defined for all square-integrable functions. The trouble--and the reason the introduction of a weak topology is necessary--is that the norms $\|S_x^{(n)}\|$ tend to infinity with n. We should also note that the weak topology used here is even weaker than that usually described in textbooks; in the latter, one allows the functions u and v of (89) to be any square-integrable functions.

There is a fairly simple way to think of a weak topology such as that defined above. Let u be any square-integrable function; then the inner product (u,v) is like a "coordinate" of u along the "v-axis." (The likeness is stronger if v has unit norm, but this is not important here.) Thus (89) says that each (permissible) coordinate of $S_x^{(n)}u$ tends to the corresponding coordinate of $T_x u$. If the basic space were of finite dimension this would in turn imply "strong" convergence, but in the space of square-integrable functions it is definitely a weaker statement.

Finally, we can at last reply to the question raised at the beginning of this section as to whether there are sufficiently many eigenfunctions. By "sufficiently many" we would mean that to any square-integrable function u there exist finite linear combinations of the eigenfunctions that approximate u as closely as desired--in other words, that any suggested source can be approximated by a sum of modes. The answer to this question is a qualified yes and arises from the fact that $S_x^{(n)}u$ is itself a finite linear combination of the eigenfunctions and from (11) which says that $T_x u$ approximates u. Using our present array of tools, an actual approximation would proceed in three steps: first, we would choose a u_0 in the set B_0 that adequately approximates u (simply truncate u at a sufficiently large value of z); then we would choose $x > 0$ small enough so that $T_x u_0$ adequately approximates u_0. Finally, we would choose n large enough so that, by (89), $S_x^{(n)}u_0$ adequately approximates $T_x u_0$.

This latter will also be the desired approximation for u.

The rub here comes from the last step. According to (89) we can approximate any finite number of "coordinates," but we cannot go further to approximate, for example, over the infinite z-axis. While we can say that the weak closure of the linear manifold spanned by the set $\{g_m\}$ is the entire space of square-integrable functions, we cannot say this of the "strong" closure. Although textbooks do not mention the subject, we might say that the biorthogonal pair of sequences $\{g_m\}$ and $\{g_m{}^*\}$ is "weakly complete."

5. CONCLUSION

We have seen how the Hilbert space of square-integrable functions on $(0,\infty)$ often provides a natural setting for the problem of propagation through a stratified atmosphere. The reason for this is that the resulting square-norm is closely related to power flow. We found that an almost immediate consequence is that the solution can be represented as an analytic, norm-decreasing semi-group of transformations on that Hilbert space.

When one comes to a mode analysis, however, the results lose much of their directness. Sometimes the Hilbert space seems even to get in the way, as when we seem forced to accept a nonstandard weak topology. This is rather disappointing since it was the customary use of mode theory and (bi-)orthogonality that first led us to introduce Hilbert space.

As another instance of how we seem hindered rather than helped, consider the case when the wave number k is real. Then the function $R(\lambda;z,\zeta)$ exists and is analytic in the lower-half λ-plane, and it is not impossible to use analytic continuation to extend this function (but not the operator R_λ) into the upper-half plane. Most of the results of Section 4 will still be valid; even the conclusion (88) will probably still be true, although the proof will require fussier details. In the process, however, we shall have lost our Hilbert space. The "modes" $g_m(z)$ are now not square-integrable; instead they increase exponentially to infinity. Of course, this has further consequences. It becomes difficult to claim that the modes satisfy the radiation condition at infinity and impossible to claim that the radiation condition characterizes them.

Finally we would suggest that one reason the hybrid methods (see, e.g., Felsen and Ishihara, 1979) seem so attractive is that the mode series converges so poorly and that it is particularly the creeping wave modes that give the greatest trouble. We would also suggest that cleaner and more complete results might be obtained with a change in function space. For example, the particular form of the weak topology introduced above with its emphasis on functions of bounded support reminds one of the theory of distributions and generalized functions.

REFERENCES

Abramowitz, M., and I. A. Stegun (1964), Handbook of Mathematical Functions, 1046 pp., National Bureau of Standards, Washington, D. C.

Arora, R. K., and J. R. Wait (1978), Refraction theories of radio wave propagation through the troposphere--a review, Doc. 78S-002, American Geophysical Union. See also Radio Sci. 13 No. 3, pp. 599-600.

Coddington, E. A., and N. Levinson (1955), Theory of Ordinary Differential Equations, 429 pp., (McGraw-Hill, New York).

Copson, E. T. (1935), An Introduction to the Theory of Functions of a Complex Variable, 448 pp., (Oxford University Press, London).

Dunford, N., and J. T. Schwartz (1958), Linear Operators (3 vols.), 2592 pp., (Interscience Publishers, New York).

Felsen, L. B., and T. Ishihara (1979), Hybrid ray-mode formulation of ducted propagation, J. Acous. Soc. Am. 65, pp. 595-609.

Fock, V. A. (1965), Electromagnetic Diffraction and Propagation Problems, 414 pp., (Pergamon Press, Oxford).

Freehafer, J. E., W. T. Fishback, W. H. Furry, and D. E. Kerr (1951), Theory of propagation in a horizontally stratified atmosphere, Propagation of Short Radio Waves, D. E. Kerr (Ed.), pp. 27-180, Radiation Laboratory Series 13, (McGraw-Hill, New York).

Helmberg, G. (1969), Introduction to Spectral Theory in Hilbert Space, 346 pp., North-Holland, Amsterdam.

Hille, E., and R. S. Phillips (1957), Functional Analysis and Semi-Groups, Colloq. Pub. 31, 808 pp., Amer. Math. Soc., Providence, RI.

Jones, D. S. (1964), The Theory of Electromagnetism, 807 pp., (Pergamon Press, Oxford).

Marcus, S., and W. D. Stuart (1981), A model to calculate EM fields in tropospheric duct environments at frequencies through SHF, Electronic Systems Division Tech. Rpt. ESD-TR-81-102. (NTIS Order No. AD-A107710.) See also Radio Sci. 17 No. 5, pp. 895-901, 1982.

Pappert, R. A., and C. L. Goodhart (1977), Case studies of beyond-the-horizon propagation in tropospheric ducting environments, Radio Sci. 12 No. 1, pp. 75-87.

Stone, M. H. (1932), Linear Transformations in Hilbert Space, Colloq. Pub. 15, 622 pp., Amer. Math. Soc., Providence, RI.

Wait, J. R. (1962), Electromagnetic Waves in Stratified Media (edition 1), 372 pp., (Pergamon Press, Oxford).

FORM **NTIA-29** (4-80)

U.S. DEPARTMENT OF COMMERCE
NAT'L. TELECOMMUNICATIONS AND INFORMATION ADMINISTRATION

BIBLIOGRAPHIC DATA SHEET

	1. PUBLICATION NO.	2. Gov't Accession No.	3. Recipient's Accession No.
	NTIA Report 87-219		

4. TITLE AND SUBTITLE
A General Theory of Radio Propagation through a Stratified Atmosphere

5. Publication Date
July 1987

6. Performing Organization Code
NTIA/ITS.S3

7. AUTHOR(S)
George Hufford

9. Project/Task/Work Unit No.
7 910 7108

8. PERFORMING ORGANIZATION NAME AND ADDRESS
National Telecomm & Information Administration
Institute for Telecommunication Sciences
325 Broadway
Boulder, CO 80303 3328

10. Contract/Grant No.

11. Sponsoring Organization Name and Address
NTIA
Herbert C. Hoover Building
14th & Constitution Washington, D.C. 20230

12. Type of Report and Period Covered

13.

14. SUPPLEMENTARY NOTES

15. ABSTRACT *(A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here.)*

When wave propagation through a stratified atmosphere is formulated in operator theoretic terms, it becomes evident that the problem does not follow the guide of the usual examples of mathematical physics. Nevertheless, such a formulation is useful to reveal why many of the usual procedures are valid and where they may be deficient. In particular, even if the refractivity profile is merely required to belong to a general class of perhaps badly discontinuous functions, the problem always has a well-behaved solution that may be subjected to a model analysis. On the other hand, the resulting mode series converges neither rapidly nor uniformly, and, as the asymptotic behavior of the modes shows, one is advised to use that series only with caution.

16. Key Words *(Alphabetical order, separated by semicolons)*

Airy functions; contraction semi-groups; creeping wave modes; model analysis; radio wave propagation; stratified atmospheres; wave guide modes; weak convergence

17. AVAILABILITY STATEMENT	18. Security Class. *(This report)*	20. Number of pages
☒ UNLIMITED.	unclassified	38
☐ FOR OFFICIAL DISTRIBUTION.	19. Security Class. *(This page)* unclassified	21. Price: