18

Full wave methods for anisotropic stratified media

18.1. Introduction

This chapter deals with problems where the approximations of ray theory cannot be used and where it is necessary to take account of the anisotropy of the medium. It is therefore mainly concerned with low and very low frequencies, so that the change of the medium within one wavelength is large. The problem is discussed here for the ionosphere, which is assumed to be a plane stratified plasma in the earth's magnetic field. A radio wave of specified polarisation is incident obliquely and it is required to find the reflection coefficient as defined in §§ 11.2–11.6. In some problems it is also required to find the transmission coefficients. It is therefore necessary to find solutions of the governing differential equations, and the form used here is (7.80), (7.81). This type of problem has been very widely studied and there are many methods of tackling it and many different forms of the differential equations. These methods are too numerous to be discussed in detail, but the main features of some of them are given in §§ 18.3–18.5. The object in this chapter is to clarify the physical principles on which the various operations are based, and for this the equations (7.80), (7.81) are a useful starting point. In a few cases solutions can be expressed in terms of known functions, and some examples are given in §§ 19.2–19.6. Usually, however, it is necessary to use a numerical integration of the differential equations, and the present chapter is largely concerned with this.

The basic equations (7.80), (7.81), and most of the alternative forms described in §§ 18.4, 18.9, 18.10, apply for a single frequency and it is assumed that the incident wave is a plane wave. Thus the solutions cannot immediately give information about the propagation of modulated signals, and particularly of pulses, nor about the angular spectrum of waves from a point source; § 11.14. Information about pulse propagation can be obtained by using additional differential equations with dependent variables $\partial/\partial\omega$ of the field variables. These are integrated at the same time as the basic equations. For an example see § 18.10. Information about the angular

dependence of the signals originating from a point source might similarly be obtained by using additional differential equations whose dependent variables are $\partial/\partial S_1$ and $\partial/\partial S_2$ of the field components, where S_1, S_2 are the x and y direction cosines of the wave normal of the incident wave in free space. But as far as the author knows this has never been done.

Throughout this chapter it is assumed that the wave normal of the incident wave is in the x–z plane so that $S_2 = 0$ and S is written for S_1. This has the advantage that for the matrix $\mathbf{T}$ (7.81) used in the basic equations (7.80), six of the sixteen elements are zero, which helps to shorten the computing time. The direction cosines of the vector $\boldsymbol{Y}$ are l_x, l_y, l_z. To study different magnetic azimuths of the plane of incidence l_x and l_y are changed. Some authors, for example Pitteway (1965), Pitteway and Jespersen (1966) have preferred to retain both S_1 and S_2 and to use a form of $\mathbf{T}$ with all sixteen elements non-zero.

The calculation of reflection and transmission coefficients is conveniently divided into three parts. They are described here for the form (7.80) of the equations in which the four dependent variables are the field components E_x, $-E_y$, $\mathscr{H}_x$, $\mathscr{H}_y$, that is the elements of the column matrix $\mathbf{e}$. But they apply also to other forms of the equations, that use other dependent variables; see §§ 18.9, 18.10. The three parts are:

(1) A solution is found that satisfies the physical conditions at a great height within the ionosphere. These conditions arise because the only influx of energy is in the incident wave below the ionosphere, and at a high enough level the solution must represent upgoing waves only. The differential equations are equivalent to a single equation of the fourth order, which has four independent solutions. At a high enough level where the W.K.B. solutions may be used, two of these are upgoing and two are downgoing waves. This is true when the angle of incidence θ is real, and was proved in § 6.4. For complex θ it is not necessarily true, and there may then be only one upgoing wave, or there may be three. These cases have not been fully explored. Complex θ is needed in problems of guided waves and in the study of the Goos–Hänchen effect, § 11.15. In this chapter it is assumed that there are two and only two upgoing waves at a great height, so that there are two solutions that satisfy the physical conditions. Methods of finding them are described in § 18.6.

The same methods can be used when the incident wave comes down from a great height. Then the starting solution is found at the bottom where it must consist of downgoing waves only, and the integration proceeds upwards.

(2) Starting with one of the solutions from (1), the step-by-step integration of the differential equations is performed, proceeding downwards until the free space below the ionosphere is reached. It is repeated with the second solution from (1). Some forms of the differential equations incorporate both solutions from (1) and then only one integration is needed; see § 18.9, 18.10. Some possible integration methods are described in § 18.2.

(3) The solution resulting from one integration gives values of the set E_x, $-E_y$, $\mathscr{H}_x$, $\mathscr{H}_y$ below the ionosphere. It is now separated into an upgoing and a downgoing wave, and each is in general elliptically polarised. The ratio of the amplitudes of these two waves would give a reflection coefficient, but it would apply only for an incident wave of a particular elliptical polarisation, of no special interest. The process is therefore repeated with the solution from the second integration, and the resulting upgoing and downgoing waves below the ionosphere in general have different polarisations from the first pair. Two linear combinations of the two solutions are now formed so that the incident wave of the combination has unit amplitude and is linearly polarised. For one combination it is in the plane of incidence, and for the other it is perpendicular to it. The associated reflected waves then give the two components R_{11}, R_{21} or R_{12}, R_{22} respectively of the reflection coefficient $\boldsymbol{R}$ (11.21). The details of these operations are given in § 18.7. If transmission coefficients are also needed, the same linear combinations are formed of the two solutions found in (1).

18.2. Integration methods

In stage (2) above, the differential equations are integrated over a range of height extending from the starting height $z = z_1$ used in stage (1) down to the finishing height, usually in the free space at the bottom of the ionosphere, used in stage (3). For doing this there are many different standard computer routines that use a step-by-step process. For the problems studied in this chapter the equations are written as r simultaneous first order differential equations. In (7.80) $r = 4$ and the dependent variables are the four elements of the column matrix $\mathbf{e}$, but it is sometimes necessary to use additional dependent variables. The equations may be written

$$\partial y_i/\partial z = F_i(y_1, y_2, \ldots, y_r, z), \quad i = 1, 2, \ldots, r. \tag{18.1}$$

The r dependent variables y_i are known at the starting height z_1. In any method the derivatives F_i must be calculated and the user must supply a computer subroutine for doing this. It is called the 'auxiliary subroutine'.

In one method, (18.1) is used first to calculate the derivatives where $z = z_1$ whence estimates of $y_1, \ldots y_r$ are made for a new height $z = z_2$. A first rough estimate would be $y_i(z_2) = y_i(z_1) + (z_2 - z_1)F_i$. This estimate is improved by recalculating the derivatives (18.1) at $z = z_2$ or at intermediate points. The final result is a corrected set of values of the y_i at the new height $z = z_2$. The interval $z_2 - z_1 = \delta z$ is called the 'step size'. One widely used method of this kind is the Runge–Kutta–Merson method. Details are given in books on computing methods, for example Hall and Watt (1976). The auxiliary subroutine is used four times for each step. Computer storage locations are needed for the r variables y_i and the r derivatives F_i. In addition storage locations for a further $4r$ variables are needed for numbers used during the calculation. For the applications of this chapter the dependent variables are complex so that each variable needs storage locations for two real numbers.

In another method the values of y_i and F_i are stored in the computer at four successive values of z with equal intervals δz, and the values after the next interval δz are computed. This is known as an 'Adams method'. It needs far more computer storage space for numbers, but the auxiliary subroutine is used only twice per step so this method is faster. For the first four steps, however, the four preceding $\dot{y}_i$ and F_i are not all available. The Runge–Kutta–Merson method can be used for the first four steps and the faster Adams process for the subsequent steps.

Usually the size of the step δz must be specified by the user. If it is too large accuracy is lost because of truncation errors, and if it is too small the computing time is wastefully long. Many computer library routines give error indications that show if the step is too large or too small. When first using a given set of differential equations it is advisable to make some trial integrations with various step sizes, so as to be sure that a suitable size has been found.

In most of the problems of this chapter the integration is downwards in the ionosphere so that the step δz is real and negative. In some applications, however, δz may be complex; see §§ 18.6, 18.13. For the basic equations (7.80) the dependent variables represent a wave motion, so that the wavelength λ in free space gives a guide to the step size. The wavelength in the ionosphere is of order $\lambda/|n|$ which may be small where the refractive index $|n|$ is large. For many purposes a step size $\delta z = -\frac{1}{50}\lambda$ has proved satisfactory in the author's experience. If, however, the electron concentration $N(z)$ varies very rapidly with height z, or if the path of integration passes close to a point of resonance where one refractive index n is very large (see § 18.13), it may be necessary to use a smaller step.

18.3. Alternative methods 1. Discrete strata

There are many ways of finding solutions of the basic equations (7.80), (7.81) or of equivalent equations as given in §§ 18.9, 18.10. Some of these methods are described briefly in this and the two following sections. The discussion of the main problem is resumed in § 18.6. The alternative methods are very numerous and no attempt has been made to include them all.

It has often proved convenient to imagine that the ionosphere is divided into discrete strata of thickness δz. By making δz small enough the stratified medium can be made to approximate as closely as desired to the actual ionosphere. In the commonest method of this kind the plasma in each stratum is assumed to be homogeneous. The method was used by Bremmer (1951) for an isotropic medium and was described in § 7.7 as one method for deriving and illustrating the W.K.B. solutions. It has been applied to the general anisotropic cold plasma by several authors, for example Johler and Harper (1962), Price (1964), Altman and Cory (1969a, b, 1970), Altman and Fijalkow (1970), Altman, Cory and Fijalkow (1970), Altman and Postan (1971), Nagano, Mambo and Hutatsuishi (1975), Nygrén (1981), Cheng and Fung (1977a, b), Fung and Cheng (1977).

It is assumed that the incident wave below the ionosphere is a plane wave with angle of incidence $\theta = \arcsin S$. Then in each homogeneous stratum there are in general four characteristic plane waves with refractive index vectors $(S, 0, q_i)$ where the q_i are the four roots of the Booker quartic equation, ch. 6, and are in general complex. The field components in each of the four waves are proportional to the elements E_x, $-E_y$, $\mathscr{H}_x$, $\mathscr{H}_y$ of the column matrix $\mathbf{s}_i$ as defined in § 7.14(1) and given by (7.118) or (7.134), and they contain a factor $\exp(-ikq_i z)$. These waves impinge on the boundaries between the strata and are partially reflected and transmitted. The reflection and transmission coefficients are found by imposing the boundary conditions for the electromagnetic fields. These conditions are that the column matrix $\mathbf{e}$, whose four elements are the horizontal components E_x, $-E_y$, $\mathscr{H}_x$, $\mathscr{H}_y$ of the total fields, is the same on the two sides of the boundary. The boundary conditions have been discussed by Johler and Walters (1960), Askne and Lisak (1976); see also §§ 11.8, 11.10.

Let $\mathbf{e}_r$ be the value of $\mathbf{e}$ at the bottom of the r^{th} stratum. Then it is also the value of $\mathbf{e}$ at the top of the $(r-1)^{\text{th}}$ stratum next below. The amplitudes f_i of the four characteristic waves are then the column $\mathbf{f} = \mathbf{S}_{r-1}^{-1}\mathbf{e}_r$ from (6.53) where $\mathbf{S}_{r-1}$ is the matrix of the four columns $\mathbf{s}_i$, given in § 7.16. In the homogeneous stratum each of these amplitudes depends on height through a factor $\exp(-ikq_i z)$ so that at the bottom of the stratum

$$\mathbf{f} = \Delta_{r-1}\mathbf{S}_{r-1}^{-1}\mathbf{e}_r \tag{18.2}$$

where Δ_{r-1} is a diagonal matrix with elements $\exp(-ikq_i\delta z)$ and the thickness δz of the stratum is negative. Then $\mathbf{e}$ at the bottom of the $(r-1)^{\text{th}}$ stratum is

$$\mathbf{e}_{r-1} = \check{\mathbf{P}}_{r-1}e_r, \quad \check{\mathbf{P}}_{r-1} = \mathbf{S}_{r-1}\Delta_{r-1}\mathbf{S}_{r-1}^{-1} \tag{18.3}$$

and $\mathbf{e}_{r-1}$ is also the $\mathbf{e}$ at the top of the $(r-2)^{\text{th}}$ stratum. In this way the fields in successive strata can be found, proceeding downwards. This form of the method was described by Nagano, Mambo and Hutatsuishi (1975). The topmost stratum is at a great height where it is known that there can be only upgoing waves, so that here only two of the four characteristic waves are present. Then by applying the matrix operations at successive boundaries, proceeding downwards, the amplitudes of the four waves in the free space below the ionosphere are found. These four waves are then combined to give the reflection coefficient matrix $\boldsymbol{R}$, (11.21).

This type of method has been widely and successfully used. A detailed account of one form of it was given by Johler and Harper (1962) who gave many results for various models of the ionosphere and included a study of the use of the Sen–Wyller formula, § 3.12, for dealing with collisions. The same type of method was used by Wait and Walters (1964) to study propagation from magnetic east to west and west to east at the magnetic equator. The waves then separate into two independent systems, see § 18.9, and they studied the case where $\boldsymbol{E}$ is in the plane of incidence so

$E_y = 0$, and $\mathscr{H}$ is perpendicular to the plane of incidence so $\mathscr{H}_x = 0$. In each stratum there are then only two characteristic waves, one upgoing and the other downgoing.

In any one stratum, an upgoing wave in one mode, say the ordinary wave, is partially reflected at the upper boundary to give downgoing waves of both modes, and these are again reflected at the bottom boundary to give upgoing waves of both modes. Thus the total upgoing ordinary wave is the resultant of an indefinite number of reflections, and similarly for the extraordinary wave and for the two downgoing waves. The full form of the theory takes account of this. But if the reflection coefficients are small enough all reflected waves after the first reflection are small and can be ignored in a first approximation. The reasons for this were discussed in § 7.7. When this is done the calculation can be simplified. A method of this kind was described by Altman and Cory (1969a, b, 1970) who showed that the resultant ionospheric reflection coefficients calculated in this way were in most cases within 10% of the results from integration of the differential equations without approximations.

Another interesting variant of this general method must be mentioned. In the basic equations (7.80) the matrix $\mathbf{T}$ is analytic and bounded at all points in the complex z plane except at points of resonance where $\varepsilon_{zz} = 0$, see § 18.13, and at singularities of $N(z)$ and at infinity. In practice, therefore, near any real point z_k, the matrix $\mathbf{T}$ can be expanded in a power series with matrix coefficients

$$\mathbf{T} = \mathbf{T}_0 + (z - z_k)\mathbf{T}_1 + (z - z_k)^2\mathbf{T}_2 + \cdots \tag{18.4}$$

Similarly each of the four wave amplitudes can be expanded in a series of powers of $z - z_k$. Inoue and Horowitz (1966b) used strata that are not homogeneous, but with $\mathbf{T}$ given by the first two terms of (18.4). For the four solutions, the fields are matched at the boundaries as before. With this method it is found that the width δz of the strata can be about ten times the width needed with homogeneous strata, to obtain the same accuracy.

In this type of method it is necessary to solve the Booker quartic for each one of the strata, and then to sort the roots into those for upgoing and downgoing waves, in the same order for successive strata. This can use a lot of computer time and was a disadvantage with some of the older computers.

In the computer methods of integrating differential equations mentioned at the end of § 18.1, a finite step size δz is used so that they are in effect equivalent to dividing the media into discrete strata. But these strata are not homogeneous. The processes used are equivalent to the representation of $\mathbf{T}$ and of the solutions $\mathbf{e}$ by power series such as (18.4) that may go to fourth or higher powers. Moreover the computing of the elements of $\mathbf{T}$ by the auxiliary subroutine is very much simpler, and therefore less subject to human error, than calculating the coefficients (6.23) of the quartic, then solving the quartic and sorting the roots, then calculating the $\mathbf{s}_i$ (7.118), and then finally imposing the boundary conditions.

18.4. Alternative methods 2. Vacuum modes

The electromagnetic fields of a wave in a medium are in the vacuum that permeates the space between the particles. The fields act on each particle causing it to oscillate so that it emits its own field into the surrounding vacuum. The total field $\boldsymbol{E}, \mathscr{H}$ at any point is the sum of the fields contributed by the incident wave and by the oscillating particles. It is this total field that acts on each particle. This idea was used by Darwin (1924) in a study of the refracting and dispersive properties of homogeneous matter. It was applied by Hartree (1929, 1931b) to a stratified magnetoplasma for incident plane waves at both normal and oblique incidence. For a plasma and for isotropic matter the oscillating particles are Hertzian dipole oscillators. Darwin (1924) also considered atoms or molecules for which higher order multipole oscillators have to be included.

To add the fields reaching any point, the contributions may first be found from the particles in an infinitesimally thin layer of thickness δz. For a homogeneous medium, or for normal incidence on a stratified medium this layer may be chosen to be perpendicular to the wave normal, so that the particles in it all oscillate in phase with each other. The sum of their fields can then be visualised as the resultant of an amplitude–phase diagram of the kind used in physical optics (Ratcliffe, 1959, figs. 3.1, 3.2). The sum of the contributions from the various layers δz is then expressed as an integral. For oblique incidence on a stratified variable medium, both stages of the summation are expressed as integrals. There are difficulties with the convergence of the integral over one layer δz, but these can be surmounted; see Darwin (1924, pp. 139, 144), Hartree (1929, p. 101). In this way the equations for finding the fields $\boldsymbol{E}, \mathscr{H}$ are expressed as an integral equation, which can be thought of as an alternative to the basic differential equations (7.80) that are based on Maxwell's equations, and the constitutive relations. It can be shown that differentiation of the integral equation gives differential equations that agree with (7.80). The fields reaching any point are strongest from those particles that are nearest. There are many of these within a distance much less than one vacuum wavelength. The fields acting on any one particle therefore consist predominantly of the storage fields of the nearby particles. This is allowed for in the treatments of Darwin and Hartree.

When this method is used for studying the fields in a stratified medium and for finding the reflection coefficients, it shows how much of the resultant reflected wave is generated at each level. It thus shows which parts of the medium are the most important for the reflection process. It has been used in this way by Heading (1953, 1963, 1981a, b) and Westcott (1962a–d, 1964).

Poeverlein (1958a, b) has used the same idea in a slightly different way. The medium is imagined to be divided into plane layers perpendicular to the z axis and each layer is made very thin with vacuum on both sides of it. The waves propagate through the vacuum and interact with the successive thin layers. The electric fields

cause currents to flow in each layer both parallel and perpendicular to its surfaces, and these can be found when the fields are known. The total fields on the two sides of a layer are related by boundary conditions at the surfaces. They consist of two wave systems, namely a 'partial wave' system and an 'additional' wave that goes right through the layer as if it were not there. Thus in this method the effect of the particles in a layer is already summed over the horizontal plane of the layer. The definition of the partial wave and the additional wave is a matter of choice, and Poeverlein (1958a) chooses them so that the partial wave is present only on the side from which the incident wave comes. Thus the partial wave has a component travelling towards the layer and another component travelling away. It cannot be said to be 'emitted' from the layer. But this choice is useful because the amplitudes of the upgoing and downgoing component waves in the vacuum at any level are very simply related to the reflection coefficient matrix variable $\boldsymbol{R}(z)$ defined at (18.45) in § 18.10.

The following method is another way of treating the waves as though they are propagated in a vacuum. The horizontal field components at any level are the four elements E_x, $-E_y$, $\mathscr{H}_x$, $\mathscr{H}_y$ of the column matrix $\mathbf{e}$. They can be resolved into four component waves in any chosen way and in § 6.10 they were resolved, by (6.53), into the fields of the four characteristic waves of magnetoionic theory, but now we use a different way. Imagine a thin layer of vacuum parallel to the plane $z =$ constant in the ionosphere. It can be supposed to be so thin that it has no appreciable effect on the fields. Within this vacuum layer the fields are resolved, as in § 11.7, into four linearly polarised waves, two upgoing, of which one has $\boldsymbol{E}$, and the other $\mathscr{H}$, in the plane of incidence, and two downgoing with $\boldsymbol{E}$, $\mathscr{H}$ respectively in the plane of incidence, that is the x–z plane. The elements of $\mathbf{e}$ for the first of these are proportional to C, 0, 0, 1, and similarly for the other three. Hence

$$\mathbf{e} = \mathfrak{f}_1\begin{pmatrix} C \\ 0 \\ 0 \\ 1 \end{pmatrix} + \mathfrak{f}_2\begin{pmatrix} 0 \\ -1 \\ -C \\ 0 \end{pmatrix} + \mathfrak{f}_3\begin{pmatrix} -C \\ 0 \\ 0 \\ 1 \end{pmatrix} + \mathfrak{f}_4\begin{pmatrix} 0 \\ -1 \\ C \\ 0 \end{pmatrix}. \tag{18.5}$$

Here $\mathfrak{f}_1$ to $\mathfrak{f}_4$ are the complex amplitudes of the four component waves, and they are now used as the four elements of a column matrix $\mathfrak{f}$. In § 11.7 they were written a, b, c, d. Then (18.5) can be written in matrix form

$$\mathbf{e} = \mathbf{S}_\mathrm{v}\mathfrak{f}, \quad \mathbf{S}_\mathrm{v} = \begin{pmatrix} C & 0 & -C & 0 \\ 0 & -1 & 0 & -1 \\ 0 & -C & 0 & C \\ 1 & 0 & 1 & 0 \end{pmatrix}; \tag{18.6}$$

compare (11.44). This transformation is analogous to (6.53). The matrix $\mathbf{S}_\mathrm{v}$ simply replaces the matrix $\mathbf{S}$ of (6.53), (7.85), (7.108). Now (18.6) can be substituted in the basic differential equation (7.80) to give the differential equation satisfied by $\mathfrak{f}$:

$$\mathfrak{f}' = -\mathrm{i}\mathbf{S}_\mathrm{v}^{-1}\mathbf{T}\mathbf{S}_\mathrm{v}\mathfrak{f} = -\tfrac{1}{2}\mathrm{i}\mathbf{W}_\mathrm{v}\mathfrak{f} \text{ (say)}. \tag{18.7}$$

Expressions for $\mathbf{S}_v^{-1}$ and $\mathbf{W}_v$ are given later, (18.46), (18.47). Since $\mathbf{S}_v$ is independent of z, there is nothing in this equation corresponding to the coupling matrix $\mathbf{\Gamma}$ of (7.108). The amplitudes of the four 'vacuum modes' are the four elements of $\mathfrak{f}$. The coupling between them is given by the non-diagonal elements of $\mathbf{W}_v$. The equations (18.7) are now integrated, for example by the type of method described in § 18.1. This form of the equations wave suggested by Poeverlein (1967) and has been successfully used by Walsh (1967). The discussion of numerical swamping in Walsh's paper is particularly valuable; see § 18.11.

18.5. Alternative methods 3. The matrizant

If the matrices $\mathbf{e}(z)$ and $\mathbf{T}(z)$ in the basic equations (7.80) are replaced by scalars $\mathsf{e}(z)$ and $\mathsf{T}(z)$, the equation is

$$\mathsf{e}' = -ik\mathsf{T}\mathsf{e} \tag{18.8}$$

where, as usual, a prime $'$ means $k^{-1}\mathrm{d}/\mathrm{d}z$. This has the solution

$$\begin{aligned}\mathsf{e}(z) &= \mathsf{e}(z_1)\exp\left(-ik\int_{z_1}^{z}\mathsf{T}(\zeta).\mathrm{d}\zeta\right)\\ &= \mathsf{e}(z_1) - ik\mathsf{e}(z_1)\int_{z_1}^{z}\mathsf{T}(\zeta)\mathrm{d}\zeta + \frac{(ik)^2}{2!}\left(\int_{z_1}^{z}\mathsf{T}(\zeta)\mathrm{d}\zeta\right)^2 + \cdots\end{aligned} \tag{18.9}$$

This would give $\mathsf{e}(z)$ at height z when $\mathsf{e}(z_1)$ at height $z = z_1$ is known, provided that the integral of T can be found. The terms in the expansion contain powers of this single integral of T. In fact e, T are really the matrices $\mathbf{e}$, $\mathbf{T}$, but a similar result can be derived. As a first approximation assume that $\mathbf{e}$ has the constant value $\mathbf{e}(z_1)$, and substitute this on the right of (7.80). Then integration gives the second approximation

$$\mathbf{e}(z) = \mathbf{e}(z_1) - ik\int_{z_1}^{z}\mathbf{T}(\zeta)\mathrm{d}\zeta\cdot\mathbf{e}(z_1). \tag{18.10}$$

If this is now substituted on the right of (7.80), a further integration gives

$$\mathbf{e}(z) = \left[1 - ik\int_{z_1}^{z}\mathbf{T}(\zeta)\mathrm{d}\zeta - k^2\int_{z_1}^{z}\mathbf{T}(\zeta_2)\left\{\int_{z_1}^{\zeta_2}\mathbf{T}(\zeta_1)\mathrm{d}\zeta_1\right\}\mathrm{d}\zeta_2\right]\mathsf{e}(z_1). \tag{18.11}$$

Note that in the last term the integral is a double integral and there is no factor 2! in the denominator. The process can now be continued indefinitely to give further terms of the series. The fourth term has a triple integral and subsequent terms have integrals of increasing multiplicity. The result is an infinite series analogous to (18.9). It can be shown to be absolutely convergent. Its sum gives

$$\mathbf{e}(z) = \mathbf{M}(z, z_1)\mathbf{e}(z_1) \tag{18.12}$$

and the 4×4 matrix $\mathbf{M}$ is called the 'matrizant'. It was used in the theory of ionospheric reflection by Volland (1962a, c) and Bossy (1971). If it is known for a sequence of height ranges $z_1 - z_2$, $z_2 - z_3$, etc. then clearly (18.12) can be applied to

these ranges in succession to give

$$\mathbf{e}(z_r) = \mathbf{M}(z_r, z_{r-1}) \cdots \mathbf{M}(z_3, z_2)\mathbf{M}(z_2, z_1)\mathbf{e}(z_1). \tag{18.13}$$

One method of calculating the matrizants $\mathbf{M}$ is to use the series of multiple integrals whose first three terms are given by (18.11). If $|z_2 - z_1|$ is small enough it is found that only a small number of terms is needed. Further, for small $|z_2 - z_1|$, the matrix $\mathbf{T}$ can be expressed as a series of powers of $z - z_1$ and then $\mathbf{M}$ can also be expressed as a series of powers (Rawer and Suchy, 1967, p. 162). Bossy (1979) has found that if strata of thickness $z_2 - z_1 \approx 1$ km are used, and if $\mathbf{T}$ is expanded up to fifth powers of $z - z_1$, then $\mathbf{M}$ can be found with a relative precision better than 10^{-7}.

For any z_k, (18.12) shows that $\mathbf{M}(z_k, z_k)$ is the unit 4×4 matrix $\mathbf{1}$. Substitution of (18.12) in the basic equation (7.80) shows that

$$\frac{\mathrm{d}}{\mathrm{d}z}\mathbf{M}(z, z_1) = -ik\mathbf{T}\mathbf{M}(z, z_1). \tag{18.14}$$

Thus another way of finding $\mathbf{M}$ would be to start at $z = z_1$ with $\mathbf{M} = \mathbf{1}$ and integrate (18.14), for example by the methods described in § 18.1.

In its simplest form the matrizant method consists of calculating $\mathbf{M}$ for a series of strata and then using (18.13). It is therefore similar to the use of discrete strata as in § 18.3. Its advantage is that for the same amount of computation the accuracy is considerably greater. In (18.13) z_1 is at a great height and z_r is below the base of the ionosphere. For $\mathbf{e}(z_1)$ there are two different possible upgoing waves at the top and (18.13) must be applied to both. The results $\mathbf{e}(z_r)$ are then combined, as in § 18.7, to give the reflection coefficients. The form (18.13) does, however, have the disadvantage that, in many practical cases, one solution is a non-penetrating wave that is almost evanescent at great heights. With decreasing height its amplitude can increase so quickly that it gives rise to numerical swamping; see § 18.11. For this reason Bossy (1979) has used a different method for connecting the solutions at successive boundaries between strata. This will now be described in outline.

The matrix $\mathbf{e}$ of the field variables can be transformed by the relation

$$\mathbf{f} = \mathbf{S}_v^{-1}\mathbf{e} \tag{18.15}$$

from (18.6). The four elements of $\mathbf{f}$ are then the amplitudes of the four vacuum modes as described by (18.5). If this transformation is applied to (18.12) with $z = z_2$ it gives

$$\mathbf{f}(z_2) = \mathbf{P}(z_2, z_1)\mathbf{f}(z_1), \quad \mathbf{P} = \mathbf{S}_v^{-1}\mathbf{M}\mathbf{S}_v. \tag{18.16}$$

A matrix very similar to $\mathbf{P}$ was used by Lacoume (1967) (with different notation) who called it the 'propagator', but he applied it not to the vacuum modes but to the characteristic waves of magnetoionic theory. Thus he used $\mathbf{S}$ (6.53) instead of $\mathbf{S}_v$ and his $\mathbf{S}(z_2)$, $\mathbf{S}(z_1)$ were different, whereas in (18.16) the two $\mathbf{S}_v$s are the same. In (18.16) $\mathbf{P}$ relates the fields at the two boundaries $z = z_1, z_2$. Thus it plays the same part as the

matrizant when $\mathbf{f}$ is used instead of $\mathbf{e}$ to specify the wave fields. To find the fields at the boundaries of successive strata, the matrices $\mathbf{P}$ are multiplied in succession, just as the $\mathbf{M}$s are multiplied in (18.13). But by using $\mathbf{P}$ the trouble from numerical swamping is avoided.

The matrix $\mathbf{P}$ is related to the reflection and transmission coefficients of a stratum. This can be shown as follows (Volland, 1962a, b; Bossy, 1979). Let z_r, z_{r+1} be the boundaries of a stratum, with $z_r > z_{r+1}$. The elements $\mathfrak{f}_1$, $\mathfrak{f}_2$ of $\mathbf{f}$ refer to upgoing waves and $\mathfrak{f}_3$, $\mathfrak{f}_4$ refer to downgoing waves, from (18.5). Let

$$\boldsymbol{u}_r = \begin{pmatrix} \mathfrak{f}_1(z_r) \\ \mathfrak{f}_2(z_r) \end{pmatrix}, \quad \boldsymbol{d}_r = \begin{pmatrix} \mathfrak{f}_3(z_r) \\ \mathfrak{f}_4(z_r) \end{pmatrix}. \tag{18.17}$$

The 4×4 matrix $\mathbf{P}(z_{r+1}, z_r)$ from (18.16) is partitioned into the four 2×2 matrices thus

$$\boldsymbol{P}_1 = \begin{pmatrix} \mathsf{P}_{11} & \mathsf{P}_{12} \\ \mathsf{P}_{21} & \mathsf{P}_{22} \end{pmatrix}, \quad \boldsymbol{P}_2 = \begin{pmatrix} \mathsf{P}_{13} & \mathsf{P}_{14} \\ \mathsf{P}_{23} & \mathsf{P}_{24} \end{pmatrix}, \quad \boldsymbol{P}_3 = \begin{pmatrix} \mathsf{P}_{31} & \mathsf{P}_{32} \\ \mathsf{P}_{41} & \mathsf{P}_{42} \end{pmatrix}, \quad \boldsymbol{P}_4 = \begin{pmatrix} \mathsf{P}_{33} & \mathsf{P}_{34} \\ \mathsf{P}_{43} & \mathsf{P}_{44} \end{pmatrix}. \tag{18.18}$$

Then (18.16) gives

$$\boldsymbol{u}_{r+1} = \boldsymbol{P}_1 \boldsymbol{u}_r + \boldsymbol{P}_2 \boldsymbol{d}_r, \quad \boldsymbol{d}_{r+1} = \boldsymbol{P}_3 \boldsymbol{u}_r + \boldsymbol{P}_4 \boldsymbol{d}_r \tag{18.19}$$

whence

$$\begin{pmatrix} \boldsymbol{d}_{r+1} \\ \boldsymbol{u}_r \end{pmatrix} = \begin{pmatrix} \boldsymbol{P}_3 \boldsymbol{P}_1^{-1} & \boldsymbol{P}_4 - \boldsymbol{P}_3 \boldsymbol{P}_1^{-1} \boldsymbol{P}_2 \\ \boldsymbol{P}_1^{-1} & -\boldsymbol{P}_1^{-1} \boldsymbol{P}_2 \end{pmatrix} \begin{pmatrix} \boldsymbol{u}_{r+1} \\ \boldsymbol{d}_r \end{pmatrix}. \tag{18.20}$$

The elements $\boldsymbol{u}_{r+1}$, $\boldsymbol{d}_r$ on the right represent waves going in to the stratum and $\boldsymbol{d}_{r+1}, \boldsymbol{u}_r$ on the left represent waves coming out. The 4×4 matrix on the right of (18.20) is called the S-matrix of the stratum (in French: matrice de diffusion). It is here denoted by $\mathscr{S}$. Suppose that there is no downgoing wave at the top z_r of the stratum. Then $\boldsymbol{d}_r = 0$ and (18.20) shows that the reflection coefficient matrix $\boldsymbol{R}_{r+1}$ at the bottom, defined by (11.20)–(11.24), is given by $\boldsymbol{P}_3 \boldsymbol{P}_1^{-1}$. Similarly the other 2×2 component matrices of $\mathscr{S}$ are reflection and transmission coefficient matrices of the stratum, thus

$$\mathscr{S} = \begin{pmatrix} \boldsymbol{R}_{r+1} & \boldsymbol{T}_{r,r+1} \\ \boldsymbol{T}_{r+1,r} & \boldsymbol{R}_r \end{pmatrix} \tag{18.21}$$

where the first subscript of a transmission coefficient matrix $\boldsymbol{T}$ denotes the boundary where the wave goes in to the stratum, and the second shows where it comes out.

18.6. Starting solutions at a great height

Integration of a wave equation such as (7.80) usually starts at a great height in the ionosphere and proceeds downwards. The set of four field variables $\mathbf{e}$ at the start must be suitably chosen. It must represent a wave that is travelling obliquely

upwards and since there are two independent upgoing waves, two different choices are possible and two integrations are needed.

There are several different ways of calculating the starting values of **e**. One way is to use the formula (7.118) for the eigen column $\mathbf{e} \propto \mathbf{s}_i$ at the starting level. It necessitates solving the Booker quartic (6.15) for the qs at the starting level, and then selecting the two that apply for upgoing waves. This is complicated, and requires computer storage space for the necessary algebra. With modern computers it presents no problem, but it was at one time a disadvantage when computers had less available storage for programs.

Two other methods of calculating the starting **e**s are of some physical interest. The first was used by Pitteway (1965). Suppose that any arbitrary value of **e** is chosen. Then it must be expressible as the sum of multiples of the four eigen column $\mathbf{s}_i$ of **T** in (7.85), thus

$$\mathbf{e} = \sum_{i=1}^{4} a_i \mathbf{s}_i. \tag{18.22}$$

Multiply on the left by $\mathbf{T} + \Lambda\mathbf{1}$ where **1** is the unit 4×4 matrix, and Λ is a complex constant to be found as indicated below. Let this multiplication be repeated r times. The result is

$$\sum_{i=1}^{4} (q_i + \Lambda)^r a_i \mathbf{s}_i. \tag{18.23}$$

In the column (18.23), one of the eigen columns $\mathbf{s}_i$ has a coefficient very much larger than the other three. It is the one for which $|q_i + \Lambda|$ is greatest. If r is large enough this one eigen column predominates, and the others are negligible in comparison. This method is used as follows.

A value Λ is selected such that the largest $|q_i + \Lambda|$ occurs for the q_i corresponding to one of the two upgoing waves. Usually in practice one of these waves is almost evanescent so that its q_1 is almost negative imaginary. Then a positive imaginary value of Λ is suitable. The other wave is propagated with only small attenuation and its q_i is almost real and positive. Then a negative real value of Λ is suitable. It is not necessary to know the four q_is; see Pitteway (1965, appendix 1). An arbitrary column **e** is now chosen and multiplied by $\mathbf{T} + \Lambda\mathbf{1}$. It is then multiplied by some constant, which can be chosen, for example so that one of the elements of **e** is unity. This prevents the column (18.23) from becoming indefinitely large. If the new value of **e** differs from the old by more than a specified small amount, the multiplications are repeated. This iterative process is continued until two successive **e**s are the same within the specified limit. The resulting **e** is then the required starting value.

For the second method, imagine a homogeneous medium which, for all heights z, has the same composition as the actual ionosphere at the starting level. In it the four characteristic waves are propagated independently. As an illustration, suppose first

that the medium is loss-free. In practical cases for frequencies $f < f_H$ the value of q for one of the upgoing waves is then negative imaginary so that the wave amplitude decreases as z increases. This wave is an inhomogeneous wave; see § 2.15. If a value of the column matrix **e** is chosen arbitrarily it will include some of this wave; compare (18.22). Let this column be used as the starting value and let the equations (7.80) be integrated with a stepwise process for the fictitious homogeneous medium, with z decreasing, and with step size δz_h. Then the wave amplitude of the inhomogeneous wave increases, whereas the other three waves either decrease or remain constant in amplitude. Thus the inhomogeneous wave must ultimately predominate. During the integration **e** must, when necessary, be multiplied by a constant fraction to prevent its becoming too large. The absolute value of **e** is not important. It is only the ratios of its elements that are needed. The integration is continued until the ratios of corresponding elements of **e** after successive steps are all equal within specified limits. The resulting **e**, or any multiple of it, may then be used as a starting value for one of the main integrations in the actual ionosphere.

The other upgoing wave in a loss-free medium is a propagated wave in which all field quantities vary with height z through a factor $\exp(-ikqz)$ where q is real and positive. The column **e** for this wave can also be computed by a preliminary integration of (7.80) in a fictitious homogeneous medium. An imaginary step length $\delta z_h = il$ is used where l is real. Thus the amplitudes $\exp(-ikqz)$ increase at each step, whereas for the other three waves they either decrease or stay constant.

In the more general case the ionospheric medium has losses but the same method can be used. When the angle of incidence θ is real the two required starting solutions can be found by preliminary stepwise integrations with $\arg \delta z_h$ equal to π or $\frac{1}{2}\pi$ as in a loss-free medium. Occasionally it may be convenient to use other values of $\arg \delta z_h$. The derived starting **e** is an eigen column of **T**. This property is not affected by truncation errors, so the size $|\delta z_h|$ of the step can be made quite large to ensure rapid convergence. A suitable value is about one quarter of a vacuum wavelength.

This method of preliminary integration can be used with forms of the differential equations other than (7.80). It can be used, for example, with the non-linear forms (18.39), (18.49) of §§ 18.9, 18.10.

18.7. Finding the reflection coefficient

The two integrations with the two independent starting solutions of the preceding section give two sets of the four field variables **e** at some level $z = z_v$ below the ionosphere. These will be written $\mathbf{e}^{(a)}$, $\mathbf{e}^{(b)}$. From them the reflection coefficient matrix $\boldsymbol{R}$, § 11.5 is to be found, with reference level at $z = z_v$; see § 11.2. The free space below the ionosphere is like a vacuum and the fields $\mathbf{e}^{(a)}$ and $\mathbf{e}^{(b)}$ are here each resolved into the four linearly polarised vacuum modes. This is done by the method used in § 18.4, equations (18.5), (18.6), and in § 11.7, equations (11.44), (11.45). The

column matrix $\mathbf{f}$ then gives the amplitudes of these modes. For the two sets of fields, its values are, from (18.15):

$$\mathbf{f}^{(a)} = \mathbf{S}_v^{-1}\mathbf{e}^{(a)}, \quad \mathbf{f}^{(b)} = \mathbf{S}_v^{-1}\mathbf{e}^{(b)} \tag{18.24}$$

where $\mathbf{S}_v$ is given by (18.6). Now the two columns (18.24) are each partitioned as in (18.17) so that

$$\mathbf{f}^{(a)} = \begin{pmatrix} \boldsymbol{u}^{(a)} \\ \boldsymbol{d}^{(a)} \end{pmatrix}, \quad \mathbf{f}^{(b)} = \begin{pmatrix} \boldsymbol{u}^{(b)} \\ \boldsymbol{d}^{(b)} \end{pmatrix} \tag{18.25}$$

where the matrices $\boldsymbol{u}, \boldsymbol{d}$ are columns with two elements; compare (11.47). From the definition of the reflection coefficient $\boldsymbol{R}$ in §11.5 and (11.16), (11.17), or (11.46), it follows that, for either superscript (a) or (b), $\boldsymbol{d} = \boldsymbol{R}\boldsymbol{u}$. Hence let

$$\boldsymbol{U} = (\boldsymbol{u}^{(a)}, \boldsymbol{u}^{(b)}), \quad \boldsymbol{D} = (\boldsymbol{d}^{(a)}, \boldsymbol{d}^{(b)}). \tag{18.26}$$

Then

$$\boldsymbol{D} = \boldsymbol{R}\boldsymbol{U}. \tag{18.27}$$

Now the solutions $\mathbf{e}^{(a)}\mathbf{e}^{(b)}$ must be independent, that is one is not a multiple of the other. Then in general $\mathbf{f}^{(a)}, \mathbf{f}^{(b)}$ are independent, though there can be exceptions; see §18.11. When they are independent $\boldsymbol{U}$ is non-singular and (18.27) gives

$$\boldsymbol{R} = \boldsymbol{D}\boldsymbol{U}^{-1}. \tag{18.28}$$

If $\mathbf{e}^{(a)}, \mathbf{e}^{(b)}$ are replaced by any two independent linear combinations of $\mathbf{e}^{(a)}, \mathbf{e}^{(b)}$, then each pair $\boldsymbol{u}^{(a)}, \boldsymbol{u}^{(b)}$ and $\boldsymbol{d}^{(a)}, \boldsymbol{d}^{(b)}$ is replaced by the same linear combination. This is achieved by multiplying $\boldsymbol{U}, \boldsymbol{D}$ on the right by an arbitrary diagonal matrix. But in (18.28) this cancels out, so that the value of $\boldsymbol{R}$ is unaffected.

The elements of $\boldsymbol{R}$ in (18.28) could be written out in full in terms of the eight field quantities $E_x^{(a)}$, $E_y^{(a)}$, $\mathscr{H}_x^{(a)}$, $\mathscr{H}_y^{(a)}$, $E_x^{(b)}$, $E_y^{(b)}$, $\mathscr{H}_x^{(b)}$, $\mathscr{H}_y^{(b)}$, but the expressions are very complicated. Equation (18.28) shows, however, that the elements of $\boldsymbol{R}$ depend only on ratios such as $\mathscr{H}_i^{(s)}/E_j^{(s)}$, $i, j = x, y$, and not on the absolute values, and the superscript s = a or b, is the same for both; the reader should prove this. Thus $\mathbf{e}^{(a)}$ or $\mathbf{e}^{(b)}$ may each be multiplied by any constant without affecting the value of $\boldsymbol{R}$.

18.8. Allowance for the earth's curvature

In most work on full wave solutions for radio waves of low or very low frequency it is assumed that the earth is flat. This approximation is nearly always satisfactory when studying propagation to short distances with angles of incidence up to about 50° to 60°. But for angles of incidence near 90°, that is grazing incidence, the errors can be serious. This applies particularly in the study of guided waves, but in some cases it may apply for a single reflection. Some method of allowing for the earth's curvature is then needed.

If the ionosphere is assumed to be isotropic and spherically stratified, the wave equation can be formulated in spherical polar coordinates r, ϑ, φ with the origin at

the earth's centre. The solution can then be expressed as the sum of terms each of which is the product of a spherical harmonic depending only on ϑ and φ and a radial part depending only on r. When the earth's magnetic field is allowed for however, so that the ionosphere is anisotropic, the problem is harder. There are some special cases where the solution can still be separated into radial and angular parts. This is possible, for example, for propagation from east to west or west to east at the magnetic equator (Wait and Walters, 1964); see § 18.3. In another case that has been studied (Wait, 1963, Krasnushkin, 1961a, b) it was assumed that the earth's magnetic field is everywhere radial.

In the more general case, with a realistic magnetic field, these simplifications are not possible. An alternative approach is to use some 'earth flattening' approximation, and perhaps the best known, for isotropic media, is the method of the modified refractive index used by Booker and Walkinshaw (1946).

Consider a spherically stratified isotropic medium and assume that the approximations of ray theory can be used. Then for a wave at any radius r, the angle ψ that its wave normal makes with the radius is given by the Bouger law $rn \sin \psi = K$, (10.18), where K is a constant. Let the ray be in the plane $\varphi = 0$, and let $r, \vartheta, 0$ be the spherical polar coordinates of a point on it. Let r_0 denote some fixed radius in free space, for example, at the surface of the earth, and there let $n = 1$, $\psi = \arcsin S$. Then

$$S = \frac{r}{r_0} n \sin \psi. \tag{18.29}$$

To derive an earth flattening formula two approaches are now possible. First let

$$n_m = nr/r_0 \quad \text{so that} \quad n_m \sin \psi = S \tag{18.30}$$

and suppose that a Cartesian coordinate system is used with

$$x = r\vartheta, \quad z = r - r_0. \tag{18.31}$$

Then the last equation (18.30) is Snell's law (6.4) provided that the modified refractive index $n_m(z)$ is used. If the height z is small compared with the earth's radius r_0

$$n_m \approx n(1 + z/r_0). \tag{18.32}$$

This is in essence the method of Booker and Walkinshaw (1946) and has proved highly successful particularly for the study of guided waves.

Alternatively, for the second approach, let

$$S_m(z) = Sr_0/r \quad \text{so that} \quad n \sin \psi = S_m(z). \tag{18.33}$$

Again the last equation is the same as Snell's law. It now uses the unmodified refractive index n, but on the right-hand side S_m is dependent on height z, thus

$$S_m \approx S(1 - z/r_0). \tag{18.34}$$

With the approximations of ray theory, either (18.30) or (18.33) can be used

without further approximations. For full wave solutions some further approximations are necessary. These are probably small in practical cases. The use of (18.32) for full wave solutions of guided wave problems has been fully discussed by Booker and Walkinshaw (1946).

When the medium is anisotropic, Bouger's law cannot be used. The refractive index n now depends on the angle ψ and has more than one value. Thus there is nothing directly analogous to the modified refractive index n_m (18.32). But S is the sine of the angle of incidence and is used for stratified anisotropic media as in earlier chapters. This suggests that (18.34) might be used as an earth flattening method for an anisotropic spherically stratified medium. The idea was proposed by Chatterjee (1953). It was used by Walker (1966) for studying guided whistler propagation in the upper ionosphere where the plasma has curved strata parallel to the lines of force of the earth's magnetic field. The suggestion is that S_m, (18.34), should be used instead of S wherever it appears in the coefficients (6.23) of the Booker quartic, and in the elements of $\mathbf{T}$ (7.81). The proof for the quartic in the special case $l_z = 0$ was given by Walker (1966). For the more general case the subject has not been fully studied. The method is difficult to test because there is no known exact solution with which the results of a test can be compared.

The assumption of a horizontally stratified medium implies also that the components of the vector $\boldsymbol{Y}$ are independent of x and y. The reflection of radio waves of very low frequency in the ionosphere occurs in a small range of height z in which the change of $\boldsymbol{Y}$ is very small, so that the approximation that $\boldsymbol{Y}$ is constant can safely be made. But for some problems this cannot be done. For most propagation problems in the earth–ionosphere wave guide, the direction of $\boldsymbol{Y}$ changes appreciably along the path. For the study of whistler propagation large ranges of height z are used and $\boldsymbol{Y}$ cannot be taken as constant. If the earth's magnetic field $\boldsymbol{B}$ is that of a magnetic dipole at the centre of the earth, then above any one place $\boldsymbol{B}$ and $\boldsymbol{Y}$ make a constant angle with the radius and they are proportional to $1/r^3$ where r is distance from the centre. Now $\boldsymbol{B}$ and $\boldsymbol{Y}$ must satisfy Maxwell's equations. If there are no currents in the upper atmosphere then curl $\boldsymbol{B} = 0$, curl $\boldsymbol{Y} = 0$. Thus if Y_x and Y_y vary with height z, it follows that Y_z must vary with x and y respectively. The properties of the medium are then no longer independent of x and y, so that it is no longer horizontally stratified. To avoid this difficulty it is sometimes implied that the components of $\boldsymbol{B}$ depend on z but are independent of x and y. This could only happen if there is a current density

$$\boldsymbol{j} = \mu_0^{-1} \operatorname{curl} \boldsymbol{B} = \mu_0^{-1}(-\partial B_y/\partial z, \partial B_x/\partial z, 0) \tag{18.35}$$

in the ionosphere.

Very few solutions of full wave radio reflection problems are known for the case where the spatial variation of $\boldsymbol{Y}$ is allowed for. This subject needs further study.

18.9. Admittance matrix as dependent variable

Various transformations of the basic differential equations (7.80) (7.81) have been used in earlier chapters, particularly the coupled equations in the forms (7.109), (16.22), (17.18) and the Försterling form (16.90), (16.91). These were introduced to elucidate some of the physical principles, but they are not recommended for computing.

Usually the electron concentration $N(z)$ and the collision frequency $\nu(z)$ are analytic functions of z, real when z is real. Then in the basic form (7.80) the elements of $\mathbf{T}$ are all bounded analytic functions of the height z at all points in the complex z plane except singularities of $N(z)$, $\nu(z)$ and at resonant points. The singularities of N and ν are nearly always remote from the real z axis or at infinity and give no trouble. The effect of resonant points is discussed in § 18.13. Thus, in particular, the coupling points are ordinary points of the differential equations (7.80). The matrix $\mathbf{T}$ and the solutions $\mathbf{e}$ are continuous and differentiable at and near coupling points. The path of integration can therefore go near or through any coupling point without invalidating the solution. In fact the part of the solution for the component waves that are coupled may vary more slowly near a coupling point than elsewhere because it behaves like an Airy integral function $\mathrm{Ai}(\zeta)$ near $\zeta = 0$; see fig. 8.5.

The transformations that give the coupled forms introduce singularities that are *not present in the basic form* (7.80). *They use solutions of the* Booker *quartic*, and any solution q has a branch point at a single coupling point associated with it; see § 16.3. Several of the elements of the coupling matrices $\mathbf{\Gamma}$ (7.108), or $\mathbf{\Lambda}$ (16.23), or $\mathbf{\Psi}$ (17.19) are infinite at the coupling points because these are singularities. Thus, for computing, the coupled equations have two disadvantages: (a) the path of integration must be carefully chosen so that it does not go near coupling points, and (b) the equations are much more complicated so that the computing time is much longer. If the amplitudes f_i (6.53) of the characteristic waves of magnetoionic theory are needed at any level z, it is best to integrate the basic equations (7.80). The integration can then be temporarily interrupted and the transformation $\mathbf{f} = \mathbf{S}^{-1}\mathbf{e}$ used, at each level where the f_i are needed.

There are, however, other forms of the equations that use transformations of a different kind and have some advantages. For finding reflection coefficients below the ionosphere the objective is to find the matrix $\boldsymbol{R}$ as given by (11.20)–(11.23) or equivalently by (18.28), from two solutions $\mathbf{e}^{(a)}$, $\mathbf{e}^{(b)}$ below the ionosphere. Instead of using two integrations to find these, as in § 18.7, we can formulate the differential equation satisfied by $\boldsymbol{R}(z)$ as given by (18.28) at all levels z, and then use only one integration of this new differential equation. The matrix $\boldsymbol{R}$ is related to the admittance matrix $\boldsymbol{A}$ (11.38)–(11.40) by the formulae (11.48)–(11.50). It is simpler to find the differential equation satisfied by $\boldsymbol{A}$ and then to find $\boldsymbol{R}$ afterwards.

As usual a prime $'$ is used to denote $k^{-1}\mathrm{d}/\mathrm{d}z$. Equation (11.40) is now differentiated once so that

$$A\begin{pmatrix}\mathscr{H}_y\\E_y\end{pmatrix}=\begin{pmatrix}E_x\\-\mathscr{H}_x\end{pmatrix},\quad A\begin{pmatrix}\mathscr{H}_y\\E_y\end{pmatrix}'+A'\begin{pmatrix}\mathscr{H}_y\\E_y\end{pmatrix}=\begin{pmatrix}E_x\\-\mathscr{H}_x\end{pmatrix}'. \tag{18.36}$$

The basic equations (7.80), (7.81) are partitioned and rearranged thus:

$$\begin{aligned}\mathrm{i}\begin{pmatrix}E_x\\-\mathscr{H}_x\end{pmatrix}'&=\begin{pmatrix}\mathsf{T}_{11}&0\\-\mathsf{T}_{31}&0\end{pmatrix}\begin{pmatrix}E_x\\-\mathscr{H}_x\end{pmatrix}+\begin{pmatrix}\mathsf{T}_{14}&-\mathsf{T}_{12}\\-\mathsf{T}_{34}&\mathsf{T}_{32}\end{pmatrix}\begin{pmatrix}\mathscr{H}_y\\E_y\end{pmatrix},\\ \mathrm{i}\begin{pmatrix}\mathscr{H}_y\\E_y\end{pmatrix}'&=\begin{pmatrix}\mathsf{T}_{41}&0\\0&1\end{pmatrix}\begin{pmatrix}E_x\\-\mathscr{H}_x\end{pmatrix}+\begin{pmatrix}\mathsf{T}_{44}&-\mathsf{T}_{42}\\0&0\end{pmatrix}\begin{pmatrix}\mathscr{H}_y\\E_y\end{pmatrix}.\end{aligned} \tag{18.37}$$

These are combined with (18.36) to give

$$\begin{aligned}\mathrm{i}A'\begin{pmatrix}\mathscr{H}_y\\E_y\end{pmatrix}=&-A\begin{pmatrix}\mathsf{T}_{41}&0\\0&1\end{pmatrix}A\begin{pmatrix}\mathscr{H}_y\\E_y\end{pmatrix}+A\begin{pmatrix}-\mathsf{T}_{44}&\mathsf{T}_{42}\\0&0\end{pmatrix}\begin{pmatrix}\mathscr{H}_y\\E_y\end{pmatrix}\\&+\begin{pmatrix}\mathsf{T}_{11}&0\\-\mathsf{T}_{31}&0\end{pmatrix}A\begin{pmatrix}\mathscr{H}_y\\E_y\end{pmatrix}+\begin{pmatrix}\mathsf{T}_{14}&-\mathsf{T}_{12}\\-\mathsf{T}_{34}&\mathsf{T}_{32}\end{pmatrix}\begin{pmatrix}\mathscr{H}_y\\E_y\end{pmatrix}.\end{aligned} \tag{18.38}$$

The column with elements $\mathscr{H}_y$, E_y appears on the right of every term. There are two independent solutions **e** giving two independent columns $\mathscr{H}_y$, E_y and (18.38) must apply to both. For this it is necessary and sufficient that the equation got by omitting the column from (18.38) shall be satisfied. This gives the differential equation for A:

$$\mathrm{i}A'=-A\begin{pmatrix}\mathsf{T}_{41}&0\\0&1\end{pmatrix}A+A\begin{pmatrix}-\mathsf{T}_{44}&\mathsf{T}_{42}\\0&0\end{pmatrix}+\begin{pmatrix}\mathsf{T}_{11}&0\\-\mathsf{T}_{31}&0\end{pmatrix}A+\begin{pmatrix}\mathsf{T}_{14}&-\mathsf{T}_{12}\\-\mathsf{T}_{34}&\mathsf{T}_{32}\end{pmatrix}. \tag{18.39}$$

As a partial check of this equation, suppose that the medium is isotropic. Then $Y=0$, the matrix M used in (7.81) is a scalar, $M=-X/U$, and only the trailing diagonal elements of **T** are non-zero. Then (18.39) gives

$$\mathrm{i}A'=-A\begin{pmatrix}\mathsf{T}_{41}&0\\0&1\end{pmatrix}A+\begin{pmatrix}\mathsf{T}_{14}&0\\0&\mathsf{T}_{32}\end{pmatrix} \tag{18.40}$$

which shows that if A_{12} and A_{21} are zero at any level z, they must be zero at all levels. Then the equations for A_{11} and A_{22} are independent. The non-zero elements of **T** are

$$\mathsf{T}_{41}=1-X/U=n^2,\quad \mathsf{T}_{32}=C^2-X/U=q^2,\quad \mathsf{T}_{23}=1,\quad \mathsf{T}_{14}=\frac{C^2-X/U}{1-X/U}=\frac{q^2}{n^2} \tag{18.41}$$

and the equations are

$$\mathrm{i}A'_{11}=-n^2A_{11}^2+q^2/n^2,\quad \mathrm{i}A'_{22}=-A_{22}^2+q^2. \tag{18.42}$$

It can be verified that these are obtained directly from the definitions $A_{11}=E_x/\mathscr{H}_y$, $A_{22}=-\mathscr{H}_x/E_y$ (11.35), (11.36), by differentiating once and then using the forms (7.3), (7.4) of Maxwell's equations.

When equations of the type (18.39) are used for finding reflection coefficients, the calculation is divided into the three parts (1)–(3) as in § 18.1, but for any one set of conditions only one main integration (2) is needed instead of the two needed for (7.80). At a great height the starting value of $\boldsymbol{A}$, stage (1), may be found from (11.38), (11.39) by using the formula (7.118) for the ratios of the field components. Alternatively it may be found by a preliminary integration as described in § 18.6, with step size δz_h and starting with an arbitrarily chosen value of $\boldsymbol{A}$. The fields are to be the two W.K.B. solutions for upgoing waves, and hence $\boldsymbol{A}$ is to be a characteristic admittance matrix and therefore independent of z; see § 11.7 (2). Thus both sides of (18.39) are to be zero. For this preliminary integration, use of the full integration routines is wasteful of computer time. All that is necessary is to use the auxiliary subroutine to compute $\boldsymbol{A}'$, and repeat this with $\boldsymbol{A}$ replaced by $\boldsymbol{A} + k\boldsymbol{A}'\delta z_h$ until the moduli of the elements of $\boldsymbol{A}'$ are all less than a specified small number. It is not permissible to multiply $\boldsymbol{A}$ by a constant during this process, as was done for (7.80), because (18.39) is a set of non-linear differential equations. A field with any two W.K.B. solutions gives $\boldsymbol{A}' = 0$. Care must be taken that the starting value found is for two upgoing waves, by suitable choice of $\arg \delta z_h$. The value $\arg \delta z_h \approx \frac{3}{4}\pi$ is usually satisfactory.

During the main integration, stage (2), $\boldsymbol{A}$ may represent fields in which both upgoing and downgoing waves are present. Its elements contain ratios of the fields and therefore have factors of order $\exp\{-ik\int(q_i - q_j)\mathrm{d}z\}$ where $\mathrm{Re}(q_i)$ and $\mathrm{Re}(q_j)$ may have opposite signs. This shows that the elements of $\boldsymbol{A}$ can vary with z at up to twice the rate of variation of the fields $\mathbf{e}$. This in turn means that the maximum permissible step size δz for the main integration is about half that for integration of (7.80). A step size of $\lambda/100$ is usually suitable, where λ is the vacuum wavelength. Thus although only one main integration is needed, it takes just as long as the two integrations needed for (7.80).

When $\boldsymbol{A}$ is found at the bottom of the ionosphere, it is to be used, stage (3), to find the reflection coefficient matrix $\boldsymbol{R}$. This is done very simply from the matrix operation (11.49).

Equations of the type (18.39) have been successfully used for calculating ionospheric reflection coefficients for radio waves of very low frequency, VLF, for example by Barron and Budden (1959), Barron (1961), who used a different form of $\boldsymbol{A}$ from that used here. One great advantage of this type of method is that it avoids troubles from numerical swamping; § 18.11. It cannot, however, be used for finding transmission coefficients. It can also have another disadvantage, especially at higher frequencies. This is best illustrated by considering the following simple special case.

Consider an isotropic ionosphere, so that $\boldsymbol{A}$ is diagonal and (18.39) separates into the two equations (18.42). We study the second of these, which applies for linearly

polarised waves where $\boldsymbol{E}$ has only one non-zero component E_y. Suppose that there is a part of the ionosphere that is homogeneous so that q^2 is independent of z. Then the differential equation has the solution

$$A_{22} = -\mathrm{i}q\tan(kqz+\gamma) \tag{18.43}$$

where γ is a complex constant of integration. In general A_{22} represents two waves, one obliquely upgoing with $E_y = a\exp(-\mathrm{i}kqz)$ and the other downgoing with $E_y = b\exp(\mathrm{i}kqz)$. Then it can be shown that

$$b/a = \mathrm{e}^{2\mathrm{i}\gamma}. \tag{18.44}$$

If $|b/a| = 1$, γ is real and there are real heights z where (18.43) is infinite. Then the use of (18.42) must fail. This condition implies that there is a perfect standing wave, and failure occurs when the E_ys for the upgoing and downgoing component waves are in antiphase. Even when γ has a small imaginary part, there are real values of z where (18.43) and its derivative are very large, so that failure still occurs. This type of failure can occur in a variable medium and more generally in the anisotropic case (18.39).

For the application to VLF reflection, the values of Z used are large, of the order 30 for frequencies near 16 kHz, so that the amplitudes of the downgoing waves never come close to those of the upgoing waves, and there is no failure. But for higher frequencies where Z is smaller, the ionosphere can be strongly reflecting and (18.39) fails at points near the real z axis. It is then necessary to use either the basic equations (7.80), or other non-linear forms as described in the following section, or to use a path of integration displaced away from the real z axis; see end of § 18.13.

18.10. Other forms, and extensions of the differential equations

One way of avoiding the failure of (18.39) when standing waves are present is to use another independent variable instead of $\boldsymbol{A}$. One such variable is the reflection coefficient matrix $\boldsymbol{R}(z)$, as defined by (11.46)–(11.49). The differential equation satisfied by $\boldsymbol{R}(z)$ will now be derived.

In § 18.4 the transformation (18.5), (18.6) gave the amplitudes $\mathfrak{f}_1$ to $\mathfrak{f}_4$ of the four vacuum modes. These are the same as a, b, c, d of (11.46) (11.47), and (11.46) shows that

$$\begin{pmatrix}\mathfrak{f}_3\\ \mathfrak{f}_4\end{pmatrix} = \boldsymbol{R}\begin{pmatrix}\mathfrak{f}_1\\ \mathfrak{f}_2\end{pmatrix}. \tag{18.45}$$

Now the column $\mathfrak{f}$ satisfies (18.7), and (18.6) gives

$$\mathbf{S}_{\mathrm{v}}^{-1} = \tfrac{1}{2}\begin{pmatrix} 1/C & 0 & 0 & 1\\ 0 & -1 & -1/C & 0\\ -1/C & 0 & 0 & 1\\ 0 & -1 & 1/C & 0\end{pmatrix} \tag{18.46}$$

so that the 4×4 matrix on the right of (18.7) is

$$\mathbf{W}_v = \begin{pmatrix} T_{11}+T_{14}/C+CT_{41}+T_{44} & -T_{12}/C-T_{42} & -T_{11}+T_{14}/C-CT_{41}+T_{44} & -T_{12}/C-T_{42} \\ -T_{31}-T_{34}/C & C+T_{32}/C & T_{31}-T_{34}/C & -C+T_{32}/C \\ -T_{11}-T_{14}+CT_{41}+T_{44} & T_{12}/C-T_{42} & T_{11}-T_{14}/C-CT_{41}+T_{44} & T_{12}/C-T_{42} \\ T_{31}+T_{34}/C & C-T_{32}/C & -T_{31}+T_{34}/C & -C-T_{32}/C \end{pmatrix}. \tag{18.47}$$

It may be written in terms of its component 2×2 matrices thus

$$\mathbf{W}_v = \begin{pmatrix} \boldsymbol{w}_{11} & \boldsymbol{w}_{12} \\ \boldsymbol{w}_{21} & \boldsymbol{w}_{22} \end{pmatrix}. \tag{18.48}$$

Equation (18.7) is now partitioned into two equations for

$$\begin{pmatrix} \tilde{f}_1 \\ \tilde{f}_2 \end{pmatrix}' = \boldsymbol{u}' \quad \text{and} \quad \begin{pmatrix} \tilde{f}_3 \\ \tilde{f}_4 \end{pmatrix}' = \boldsymbol{d}'; \quad \text{compare (18.17).}$$

Then (18.45) is differentiated once and the values of $\boldsymbol{u}'$, $\boldsymbol{d}'$ are substituted. Finally $\boldsymbol{d}$ is set equal to $\boldsymbol{Ru}$ from (18.45). The result is an equation containing $\boldsymbol{R}$ and $\boldsymbol{R}'$ with $\boldsymbol{u}$ on the right of every term. It must be valid for either of the two independent solutions so the $\boldsymbol{u}$ may be omitted. This step is similar to the step from (18.38) to (18.39). Thus the equation satisfied by $\boldsymbol{R}$ is

$$2i\boldsymbol{R}' = \boldsymbol{w}_{21} + \boldsymbol{w}_{22}\boldsymbol{R} - \boldsymbol{R}\boldsymbol{w}_{11} - \boldsymbol{R}\boldsymbol{w}_{12}\boldsymbol{R}. \tag{18.49}$$

This equation was used by Budden (1955a, b) for studying ionospheric reflection at very low frequencies. Its use is very similar to the use of (18.39) with $\boldsymbol{A}$, as described in § 18.9. The preliminary integration used in stage (1), to find the starting value of $\boldsymbol{R}$ at the top, continues until $\boldsymbol{R} = 0$. Like $\boldsymbol{A}$, $\boldsymbol{R}$ contains ratios of the fields and may vary with z at up to twice the rate of variation of the fields $\mathbf{e}$, so that for the main integration, stage (2), a step size of about $\lambda/100$ is needed, as for $\boldsymbol{A}$. At the bottom of the ionosphere $\boldsymbol{R}$ is the required solution so there are no further operations for stage (3).

The original definition of $\boldsymbol{R}$ was formulated in ch. 11 leading to (11.46) that is (18.45), for the free space below the ionosphere. In (18.49) it is being used within the ionosphere, and is still defined by (18.45). It is the reflection coefficient for the vacuum modes (§ 18.4), but not for the upgoing and downgoing characteristic waves of magnetoionic theory. Suppose that the downward integration of (18.49) is stopped at some level before the bottom of the ionosphere is reached. This is equivalent to terminating the ionosphere by a sharp boundary plane at that level with free space below. The value of $\boldsymbol{R}$ that has been reached is the reflection coefficient at that boundary. This gives a useful physical interpretation of $\boldsymbol{R}$ within the ionosphere. It follows that no element of $\boldsymbol{R}$ can have modulus greater than unity at any real height, so that there is no failure of (18.49) even if the solution contains a perfect standing wave. Compare (18.43) and § 18.9 where it was shown that failure can occur if $\boldsymbol{A}$ is used as the dependent variable.

Suppose that the equations (18.49) are integrated for a homogeneous ionosphere, as for stage (1), and the integration is stopped when $\boldsymbol{R}'$ is zero. Then $\boldsymbol{R}$ is the reflection coefficient for a homogeneous medium with a sharp lower boundary. This method of finding it is equivalent to finding a solution of the matrix equation formed by setting the right hand side of (18.49) equal to zero, which can be done by a simple iteration method. It is not necessary to use accurate integration routines. This is an alternative to the method of § 11.10 that used solutions of the Booker quartic.

If the ionosphere is isotropic the only non-zero elements of $\mathbf{T}$ are the trailing diagonal elements given by (18.41). Then it can be shown that all matrices $\boldsymbol{w}_{ij}$ are diagonal, and (18.49) gives two separate equations:

$$2\mathrm{i}R'_{11} = Cn^2(1-R_{11})^2 - \frac{q^2}{n^2C}(1+R_{11})^2, \tag{18.50}$$

$$2\mathrm{i}R'_{22} = C(1-R_{22})^2 - \frac{q^2}{C}(1+R_{22})^2. \tag{18.51}$$

These two equations were given by Jacobsson (1965) in a study of the reflection of light from isotropic stratified films. See also Schelkunoff (1951).

One integration of (18.49) or (18.50) or (18.51) gives the reflection coefficient at one frequency only. Suppose that the reflection comes predominantly from one level in the ionosphere and it is required to find the equivalent path P' (10.40) at a given receiver. Then $\boldsymbol{R}$ alone does not contain the necessary information and we need to know also how it depends on frequency. To illustrate this consider a simple case of vertical incidence on an isotropic ionosphere with $\boldsymbol{E}$ having only one component E_x. Then the reflection coefficient is R_{11} and we now omit the subscripts. With $C = 1$, $q^2 = n^2$ (18.50) then gives

$$2\mathrm{i}R' = n^2(1-R)^2 - (1+R)^2. \tag{18.52}$$

The equivalent path P' is $2h'(f)$ where h' is the equivalent height of reflection. Below the ionosphere $-\arg R$ is the phase lag that results from the reflection. If the reflection process can be described in the language of ray theory, the phase height, § 12.2, is given by

$$h(f) = -\frac{c}{4\pi f}\arg R = \tfrac{1}{2}P(f) \tag{18.53}$$

whence from (10.40)

$$h'(f) = \partial(fh)/\partial f = -\frac{c}{4\pi}\frac{\partial}{\partial f}(\arg R) = -\frac{c}{4\pi}\mathrm{Im}\left(\frac{1}{R}\frac{\partial R}{\partial f}\right). \tag{18.54}$$

Thus if $\partial R/\partial f$ can be found, $h'(f)$ can be calculated. It is assumed that $\partial^2(fh)/\partial f^2$ and higher derivatives are small. If they are not small a radio pulse would become distorted or elongated, or there might be two reflected waves present from different

levels. Then the concept of a single equivalent height ceases to be useful. Let $\partial R/\partial f = J$. Now (18.52) is to be differentiated partially with respect to f. The prime on the left means $k^{-1}\mathrm{d}/\mathrm{d}z$ and k is proportional to f. This must be allowed for, so that the result is

$$2\mathrm{i}J' = -2J\{n^2(1-R)+1+R\} + (1-R)^2\partial(n^2)/\partial f + \{n^2(1-R)^2 - (1+R)^2\}/f. \tag{18.55}$$

The whole integration process can now be applied to (18.52), (18.55) together. The method was used by Budden and Cooper (1962) to study $h'(f)$ for a Chapman layer and for a parabolic layer, including conditions where the reflection from the discontinuity of $\mathrm{d}N/\mathrm{d}z$ at the bottom could affect the results. They gave a somewhat simplified form of (18.52) (18.55) that was useful for reducing the computing time.

If the reflected wave contains two reflections from different levels, the problem is harder. In principle it could be solved by using (18.52), (18.55) together with a third differential equation for the height dependence of $\partial^2 R/\partial f^2$. This, however, is very complicated. The problem was studied by Smith, M. S. (1973b) who found that it is equivalent and easier to compute R at three different frequencies close together. Then the separate values of R and $h'(f)$ for the two reflections can be found. If the ionosphere is anisotropic, the matrix equation (18.49) must be used instead of (18.52) but the problem of separating the reflection into two components is now easier. For a single reflection of one magnetoionic component, the polarisation has its characteristic value for that component and is independent of the polarisation of the incident wave. Then $\det(\boldsymbol{R}) = 0$ and $\boldsymbol{R}$ is singular; see problem 11.1. Smith calls this a 'simple reflection'. If the reflected wave has two components with different polarisations, the $\boldsymbol{R}$ for the composite wave is not singular, but Smith showed that by calculating $\boldsymbol{R}$ at only two adjacent frequencies, it is possible to separate it into its two components and to find their polarisations and equivalent heights of reflection.

When full wave computer integrations of the differential equations are used at medium or high frequencies, the step size $\lambda/100 - \lambda/50$ is small and the number of steps needed can be extremely large with a consequent long computing time. But for large ranges of the height z the characteristic waves are not strongly coupled and the fields are given with very good accuracy by the W.K.B. solutions (7.112). The full wave equations are only needed for those parts of the path of integration where the coupling processes are appreciable, that is where the path runs near reflection or coupling points. Thus the path is divided into 'W.K.B. ranges' and 'full wave ranges'. When the equations being integrated are the basic equations (7.80), the field variables $\mathbf{e}$ are the dependent variables. At the bottom of a full wave range the transformation (6.53) is used to give $\mathbf{f}$ and this gives the amplitudes $\mathbf{K}_j$ in (7.112). Then $\mathbf{f}$ is used in the W.K.B. range that follows. The height dependence of its elements is given by the W.K.B. solutions (7.112), so that the four integrals $\int q_j \mathrm{d}z$ must be evaluated and this

means finding and sorting the four roots q_j of the Booker quartic at a sequence of levels z, but this process is much faster than the full stepwise integration. At the bottom of the W.K.B. range the transformation (6.53) is used to give **e** and the full wave integration is resumed in the full wave range that follows.

When the equations being integrated are (18.39) or (18.49), the dependent variables, $\boldsymbol{A}$ or $\boldsymbol{R}$ respectively, depend only on ratios of the field components. But there is still a transformation that allows the W.K.B. solutions to be used in a W.K.B. range. Instead of using $\boldsymbol{R}$ defined in terms of vacuum modes, we define a matrix $\mathscr{R}$ that is the reflection coefficient for the characteristic modes. Thus suppose that the elements of **f** in (6.53) refer, in order, to the upgoing ordinary, upgoing extraordinary, downgoing ordinary and downgoing extraordinary waves. Then

$$\begin{pmatrix} \mathsf{f}_3 \\ \mathsf{f}_4 \end{pmatrix} = \mathscr{R} \begin{pmatrix} \mathsf{f}_1 \\ \mathsf{f}_2 \end{pmatrix}. \tag{18.56}$$

Note that (a) the order of the elements f_1 to f_4 is not the same as in (7.145)–(7.148) nor in (16.15), (16.16), (b) these **f**s are given by (6.53) with the transforming matrix **S**, and they are not the same as the vacuum mode **f**s of (18.45) that were given by (18.6) with transforming matrix $\mathbf{S}_v$. Some of the properties of $\mathscr{R}$ and the transformations that express it in terms of $\boldsymbol{A}$ and $\boldsymbol{R}$ were discussed by Smith (1974b).

Now let $\mathbf{f}_1$ to $\mathbf{f}_4$ be the W.K.B. solutions (7.111). Apply to (18.56) the operator $k^{-1}\mathrm{d}/\mathrm{d}z$, denoted by a prime $'$. Then it follows that

$$\mathrm{i}\mathscr{R}' = \begin{pmatrix} q_3 & 0 \\ 0 & q_4 \end{pmatrix} \mathscr{R} - \mathscr{R} \begin{pmatrix} q_1 & 0 \\ 0 & q_2 \end{pmatrix} \tag{18.57}$$

where a column $\begin{pmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \end{pmatrix}$ on the right has been omitted because (18.56) applies for two independent columns. Let

$$\mathsf{E}_j = \exp\left(-\mathrm{i}k \int_{z_1}^{z} q_j \mathrm{d}z \right). \tag{18.58}$$

These are just the factors used in the W.K.B. solutions. Then (18.56) gives

$$\mathscr{R}(z) = \begin{pmatrix} \mathsf{E}_3 & 0 \\ 0 & \mathsf{E}_4 \end{pmatrix} \mathscr{R}(z_1) \begin{pmatrix} \mathsf{E}_1^{-1} & 0 \\ 0 & \mathsf{E}_2^{-1} \end{pmatrix} \tag{18.59}$$

which can now be used to give $\mathscr{R}(z)$ at the bottom of a W.K.B. range.

The differential equation satisfied by $\mathscr{R}$ in the general case is complicated and rarely needed, but there are special cases where it is useful. For example, for vertical incidence when coupling between the ordinary and extraordinary waves is negligible, $\mathscr{R}$ is diagonal and its two elements then refer to the ordinary and extraordinary waves. This case, for the ordinary wave, was studied by Fejer and Vice (1959b) who used a variable R closely related to one diagonal element of $\mathscr{R}$. They

were thus able to study absorption near the reflection level for waves of frequencies 1.83 and 2.63 MHz.

18.11. Numerical swamping

At a great height in the ionosphere where $X > 1$, in nearly all cases of practical interest, one of the two upgoing characteristic waves is a propagated wave and the other is an inhomogeneous wave whose amplitude decreases as the height increases. This inhomogeneous solution therefore increases in amplitude as z decreases, that is as the main integration proceeds downwards, whereas the propagated wave solution stays roughly constant. Even if the medium has losses, the inhomogeneous solution still increases far faster than the propagated solution. Moreover in an integration that starts at the top with a propagated solution, some of the inhomogeneous solution is generated by coupling in the medium as the integration proceeds, and once generated it grows rapidly so that it can get very large and the propagated solution is then negligible in comparison. This is numerical swamping. When it happens, both stepwise integrations of (7.80) lead to the same solution $\mathbf{e}$ at the bottom of the ionosphere and then it is not possible to find $\boldsymbol{R}$ because $\boldsymbol{U}$ in (18.28) is singular. This swamping can occur when the thickness of the ionosphere is many wavelengths. It did not occur in early calculations for 16 kHz (Budden, 1955a, b; Barron and Budden, 1959; Barron, 1961) when only reflection coefficients were calculated and the range of the integration was less than one vacuum wavelength. It can, however, occur (a) for higher frequencies, and (b) for problems where a large thickness of the ionosphere is studied, for example Pitteway's (1965) calculations of transmission coefficients.

There are two ways in which numerical swamping may be avoided. The first is to use one of the non-linear forms of the differential equation, (18.39) with dependent variable $\boldsymbol{A}$, or (18.49) with dependent variable $\boldsymbol{R}$, as described in the two preceding sections. These variables depend only on the ratios of the components of the electromagnetic field and are not affected by swamping. The same method can be used with a method of discrete strata; see for example Nygrén (1981).

The second way applies when the differential equations used are linear and homogeneous, for example (7.80) or (18.7). It uses the following property. When $\boldsymbol{R}$ is found at the bottom of the ionosphere from (18.28), two solutions $\mathbf{e}^{(a)}$, $\mathbf{e}^{(b)}$ are used. It was shown in § 18.7 that the value of $\boldsymbol{R}$ is unaffected if these two are replaced by any independent linear combinations of them, and this replacement can be made at any stage during the integration because the equations are linear and homogeneous. A starting solution that represents the inhomogeneous wave at the top is used first, and the resulting solution will be denoted by $\mathbf{e}^{(a)}$. During the downward stepwise integration a record is kept of $\mathbf{e}^{(a)}$ for a number of heights $z_A, z_B, \ldots$. If $\mathbf{e}^{(a)}$ gets large so that computer overflow is threatened, it can be multiplied by a fraction without

affecting the results, because it is only the ratios of the elements of $\mathbf{e}$ that are finally used. When this integration is complete the second integration is started, with a solution that represents the propagated wave at the top, and this solution will be denoted by $\mathbf{e}^{(b)}$. If it gets too large this is because it has acquired a contribution from the inhomogeneous wave, that grows as z decreases. Now it is permissible to add to $\mathbf{e}^{(b)}$ any multiple of $\mathbf{e}^{(a)}$. Thus $\mathbf{e}^{(b)}$ is replaced by

$$\mathbf{e}^{(c)} = \mathbf{e}^{(b)} + \gamma \mathbf{e}^{(a)}. \tag{18.60}$$

The constant γ is to be chosen so that the elements of $\mathbf{e}^{(c)}$ are not large. The simplest choice, suggested by Walsh (1967), is to use the γ that makes one element of $\mathbf{e}^{(c)}$ zero. In this way most of the unwanted contribution from $\mathbf{e}^{(a)}$ is removed from $\mathbf{e}^{(b)}$. This process can be applied at each of the levels $z_A, z_B, \ldots$ where $\mathbf{e}^{(a)}$ has been recorded.

If the integration routine used is, for example, an Adams routine, § 18.2, that requires storage of $\mathbf{e}$ and its derivative for the four preceding steps, then it must be remembered that when $\mathbf{e}$ is changed in accordance with (18.60), or when $\mathbf{e}^{(a)}$ is multiplied by a fraction to avoid impending overflow, the stored numbers for the earlier steps are not now correct for the new $\mathbf{e}$. When the integration is resumed, therefore, it must be treated as though it is newly started.

An alternative choice for γ, used by Pitteway (1965), is to use the γ that makes

$$\mathbf{e}^{(c)\mathrm{T}} \mathbf{e}^{(a)*} = 0. \tag{18.61}$$

Then $\mathbf{e}^{(c)}$ and $\mathbf{e}^{(a)}$ are said to be 'Hermitian orthogonal'. For any column $\mathbf{e}$, the product $\mathbf{e}^{\mathrm{T}}\mathbf{e}^{*}$ is the sum of the squared moduli of the elements and cannot be zero unless all the elements are identically zero. Thus (18.61) shows that $\mathbf{e}^{(c)}$ can never be a multiple of $\mathbf{e}^{(a)}$, but must include an appreciable multiple of $\mathbf{e}^{(b)}$.

For a discussion of the avoidance of numerical swamping see Walsh (1967) who gives some instructive numerical examples.

For the solution $\mathbf{e}^{(a)}$ that consists of only the inhomogeneous wave at great heights, the downward integration gives a solution $\mathbf{e}^{(a0)}$, say, at the bottom of the ionosphere. Even if there were errors in the starting value used, this would not matter because as z decreases the true inhomogeneous wave component grows and the remaining part is then negligible in comparison. Thus the solution $\mathbf{e}^{(a0)}$ is known with precision. The complete solution $\mathbf{e}^{(a)} \rightarrow \mathbf{e}^{(a0)}$ is called by Pitteway (1965) the 'non-penetrating mode'. The other solution $\mathbf{e}^{(b)}$ is equal to $\mathbf{e}^{(b0)}$, say, at the bottom of the ionosphere, and at great heights it is the propagated wave. But we can add on to it a multiple of $\mathbf{e}^{(a)}$ for which the amplitude of the inhomogeneous wave at a great height is negligible compared with the propagated wave in $\mathbf{e}^{(b)}$, but for which the amplitudes of $\mathbf{e}^{(a0)}$ and $\mathbf{e}^{(b0)}$ are comparable. This new solution would also fulfil all the requirements for the original solution $\mathbf{e}^{(b)} \rightarrow \mathbf{e}^{(b0)}$. Thus the solution $\mathbf{e}^{(b)}$ is not known with precision. For a unique definition a further condition is necessary. Pitteway (1965) chooses $\mathbf{e}^{(b)}$ so that the ratio of the power flux in the propagated

wave at a great height to that in the upgoing part of the wave $\mathbf{e}^{(b0)}$ is a maximum. This defines a unique $\mathbf{e}^{(b)}$ which he calls the 'penetrating mode'. At a great height the penetrating mode is the whistler wave, § 13.8. These definitions are important in the study of the propagation of whistlers through the ionosphere, both upwards and downwards. For details see Pitteway (1965), Pitteway and Jespersen (1966), Altman and Suchy (1980).

18.12. Reciprocity

Reciprocity in radio propagation problems has been discussed in §§ 14.13, 14.14, for conditions where the approximations of ray theory can be used. It must now be studied for the more general case. The reciprocity theorem was originally stated for electric circuits that contain only linear circuit elements, that is resistors, capacitors and inductors including mutual inductors. It used the voltages and currents at two pairs of terminals, at points A and B, and this form of it was given in § 14.13. Now it can be proved that it still applies if each terminal pair is connected to an aerial, so that the electromagnetic fields in the space containing the aerials are part of the system. The theorem is discussed here only for non-magnetic media. For it to be true it is necessary that the electromagnetic properties of the medium are linear and that the complex dielectric constant tensor $\boldsymbol{\varepsilon}$ is symmetric. Thus it does not apply for the plasmas of the ionosphere and magnetosphere, except when the earth's magnetic field can be neglected. The proof of the theorem is effected in two stages.

For the first stage let $\boldsymbol{E}^{(1)}, \mathscr{H}^{(1)}$ be the fields of some system of waves of fixed angular frequency, and let $\boldsymbol{E}^{(2)}$, $\mathscr{H}^{(2)}$ be those of a different system at the same frequency. Then at any point where Maxwell's equations and the constitutive relations are obeyed

$$\operatorname{div}(\boldsymbol{E}^{(1)} \wedge \mathscr{H}^{(2)} - \boldsymbol{E}^{(2)} \wedge \mathscr{H}^{(1)}) = 0. \tag{18.62}$$

To show this we use the vector property $\operatorname{div}(\boldsymbol{a} \wedge \boldsymbol{b}) = \boldsymbol{b} \operatorname{curl} \boldsymbol{a} - \boldsymbol{a} \operatorname{curl} \boldsymbol{b}$, and Maxwell's equations in the form (2.44). Then since $\partial/\partial t \equiv \mathrm{i}\omega$, the left-hand side of (18.62) is

$$\mathrm{i}Z_0(\boldsymbol{E}^{(2)}\cdot\boldsymbol{D}^{(1)} - \boldsymbol{E}^{(1)}\cdot\boldsymbol{D}^{(2)}) = \mathrm{i}c^{-1}\{\boldsymbol{E}^{(2)}\cdot(\boldsymbol{\varepsilon}\boldsymbol{E}^{(1)}) - \boldsymbol{E}^{(1)}\cdot(\boldsymbol{\varepsilon}\boldsymbol{E}^{(2)}). \tag{18.63}$$

Now write $\boldsymbol{E}$ and $\boldsymbol{D}$ in the suffix notation of (2.54), with the summation convention. Then (18.63) is

$$\mathrm{i}c^{-1}E_i^{(2)}E_j^{(1)}(\varepsilon_{ij} - \varepsilon_{ji}). \tag{18.64}$$

This is zero if, and in general only if $\boldsymbol{\varepsilon}$ is symmetric.

Next (18.62) is integrated through some volume V bounded by a surface Γ. This must not contain any sources because at all points within it the form (2.44) of Maxwell's equations must hold. If $\mathrm{d}\boldsymbol{\Gamma}$ is an element of the surface in the direction of the normal, the divergence theorem of vector analysis states that $\iiint \operatorname{div} \boldsymbol{a}\, \mathrm{d}V =$

$\iint_\Gamma \boldsymbol{a}\cdot d\boldsymbol{\Gamma}$. Application of this to (18.62) gives

$$\iint_\Gamma (\boldsymbol{E}^{(1)} \wedge \mathscr{H}^{(2)} - \boldsymbol{E}^{(2)} \wedge \mathscr{H}^{(1)})\cdot d\boldsymbol{\Gamma} = 0. \tag{18.65}$$

Now Γ is chosen to consist of three parts. Two of them enclose the two terminal pairs of the aerials at A and B. The remaining one is a sphere of indefinitely large radius r in a vacuum, extremely distant from the sources. Here the fields must be outward travelling radiation fields each with amplitude proportional to $1/r$. Here also $\boldsymbol{E}^{(1)}$ is perpendicular to $\mathscr{H}^{(1)}$ and similarly $\boldsymbol{E}^{(2)}$ to $\mathscr{H}^{(2)}$ whence it can be shown that the integral (18.65) tends to zero as $r \to \infty$ (Clemmow, 1973, § 5, 4.6). Thus the sum of the remaining two contributions is zero. Finally, the fields $\boldsymbol{E}^{(1)}$, $\mathscr{H}^{(1)}$ are chosen to be those that occur when a voltage is applied at the terminal pair at A, and $\boldsymbol{E}^{(2)}$, $\mathscr{H}^{(2)}$ similarly when a voltage is applied at B. The result (18.65) is then sometimes called the 'Lorentz reciprocity theorem'.

For the second stage of the proof the two field systems must be linked to the voltages and currents at the terminal pairs at A and B. This needs care and is too long to give here. Some examples within the restrictions of ray theory were given in §§ 14.13, 14.14. The problem was discussed by Dällenbach (1942). A useful treatment of it is given by Collin and Zucker (1969, § 4.2). In this way the proof of the reciprocity theorem is completed when $\boldsymbol{\varepsilon}$ is everywhere symmetric.

If $\boldsymbol{\varepsilon}$ is not symmetric the reciprocity theorem is not in general true. An alternative theorem for magnetoplasmas can, however, be proved as follows. For the fields $\boldsymbol{E}^{(2)}$, $\mathscr{H}^{(2)}$ imagine that at every point of the medium the direction of the superimposed magnetic field is reversed but no other change is made. Then for $\boldsymbol{E}^{(2)}$, $\mathscr{H}^{(2)}$ the new dielectric constant $\bar{\boldsymbol{\varepsilon}}$ is the transpose of $\boldsymbol{\varepsilon}$; see end of § 3.10. The second term of (18.64) is now $\bar{\varepsilon}_{ji} - \varepsilon_{ij}$ and the expression is zero, so that (18.62), (18.65) are now true. We say that reciprocity holds provided that, for one of the two directions of propagation, the earth's magnetic field is imagined to be reversed.

In the full wave theory of this chapter it has been assumed that the incident, reflected and transmitted waves, where they are in free space, are plane waves, with their wave normals in the x–z plane. It is then more convenient and simpler to express the reciprocity relations in terms of the fields, or of the matrices $\boldsymbol{A}$ and $\boldsymbol{R}$. For both wave systems in (18.62) the fields are independent of y and the x dependence is only through factors $\exp(\mp ikSx)$ that cancel for the two waves. Thus the operator div is simply d/dz of the z component, and (18.62) shows that

$$E_x^{(1)}\mathscr{H}_y^{(2)} - E_y^{(1)}\mathscr{H}_x^{(2)} - E_x^{(2)}\mathscr{H}_y^{(1)} + E_y^{(2)}\mathscr{H}_x^{(1)} \tag{18.66}$$

is independent of z. Note that this is very similar to the bilinear concomitant (7.97). Here, however, both field systems satisfy Maxwell's equations, unlike the adjoint fields in (7.97). For the ionospheric reflection problem we may imagine that at a great enough height there is free space. Both waves there are upgoing and it can be shown

that (18.66) is zero. Hence it is zero also below the ionosphere. This condition can be written in terms of the admittance matrices $A^{(1)}$, $A^{(2)}$ for the two field systems, from (11.40) thus

$$(\mathscr{H}_y^{(2)}, E_y^{(2)})\left\{\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}A^{(1)} - A^{(2)\mathrm{T}}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\right\}\begin{pmatrix}\mathscr{H}_y^{(1)}\\ E_y^{(1)}\end{pmatrix} = 0. \tag{18.67}$$

This must be true for any arbitrary fields and so the expression within curly brackets is zero. Thence from (11.49) it can finally be shown that

$$R^{(1)\mathrm{T}} = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}R^{(2)}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix} \tag{18.68}$$

which gives

$$R_{11}^{(1)} = R_{11}^{(2)}, \quad R_{22}^{(1)} = R_{22}^{(2)}, \quad R_{12}^{(1)} = -R_{21}^{(2)}, \quad R_{21}^{(1)} = -R_{12}^{(2)}. \tag{18.69}$$

This result is mainly of theoretical interest. It holds only if the earth's magnetic field is imagined to be reversed for finding $R^{(2)}$.

Many relations between the As or the Rs for the two fields can be found when the earth's magnetic field is not imagined to be reversed, and a few examples will now be given.

For the fields (1) let the plane of incidence, that is the x–z plane, make an angle ψ with the magnetic meridian plane measured east of north, and for (2) let the angle be $\pi - \psi$. Then for the two systems the direction cosine l_x of Y has opposite signs and there is no other difference. The effect of this on the elements of T (7.81) is

$$\mathsf{T}_{11}^{(1)} = -\mathsf{T}_{44}^{(2)}, \quad \mathsf{T}_{12}^{(1)} = \mathsf{T}_{34}^{(2)}, \quad \mathsf{T}_{31}^{(1)} = -\mathsf{T}_{42}^{(2)} \tag{18.70}$$

and a similar set with (1) and (2) interchanged. If these changes are made in the differential equation (18.39) for A, and if the equation is then transposed, it shows that $A^{(1)}$ and $A^{(2)\mathrm{T}}$ obey the same differential equation, which is to be integrated proceeding downwards. At a very great height where there is free space, the starting value of A is diagonal and the same for both systems. Hence $A^{(1)}$ and $A^{(2)\mathrm{T}}$ are the same at the bottom of the ionosphere. Then from (11.49) it follows that

$$R^{(1)} = R^{(2)\mathrm{T}}. \tag{18.71}$$

The planes of incidence for the two fields here are not in general parallel so that this is not a reciprocity property, but results of this kind are customarily discussed under the same heading. Suppose, however, that $\psi = 0$. Then the two directions are from north to south and south to north. We can imagine that the aerials are electric dipoles at distant points A, B in the magnetic meridian plane. First let them both be vertical dipoles. Then the received signals for A → B and B → A are proportional to $R_{11}^{(1)}$, $R_{11}^{(2)}$ respectively and (18.71) shows that they are equal. Thus there is reciprocity for these aerials. Similarly there is reciprocity if both aerials are horizontal dipoles because $R_{22}^{(1)} = R_{22}^{(2)}$. Suppose now that there is a vertical dipole at A and a horizontal dipole at B. When A transmits, the received current at B is proportional to $R_{21}^{(1)}$ but

when B transmits, the received current at A is proportional to $-R_{12}^{(2)}$. The minus sign is needed because of the sign convention used in the definition (11.16), (11.17) of $\boldsymbol{R}$.

The correct assignment of the signs for this problem is difficult. The following note may help. When a vertical dipole aerial transmits, then at a point many wavelengths away in free space, in the positive x direction, the field components E_z and $\mathscr{H}_y$ are in antiphase. When this same dipole receives a plane wave coming the opposite way from the same direction, the E_z and $\mathscr{H}_y$ in this wave are in phase. The $\mathscr{H}_y$s for the transmitted and received waves have opposite signs. The vertical dipole responds to E_z but it is $\mathscr{H}_y$ that is used in the definitions (11.16), (11.17) of the reflection coefficients. A horizontal dipole, parallel to the y axis, responds to E_y and there is no corresponding change of sign for transmission and reception.

Hence there is reciprocity for the amplitude in the two directions, but the phase shifts differ by 180°. The same result applies for a horizontal dipole aerial at A and a vertical dipole at B. This was called 'antireciprocity' in § 14.13, and some examples of it, with rays, were given in § 14.15.

These results have been proved for the reflection of a single plane wave but it can be shown (Budden, 1954) that they must also be true for a spherical wave from a small source, and when allowance is made for the earth's curvature. They are therefore applicable to radio transmission over any distance in the magnetic north–south direction.

Another relation for two wave systems was derived by Pitteway and Jespersen (1966). They used the property that (18.66) is independent of z and they applied it to two systems with azimuths ψ, $\pi - \psi$, as in the foregoing example. For system (1) the wave was a penetrating mode as defined at the end of § 18.11, giving an upgoing whistler wave at a great height. For system (2) the incident wave was a downgoing whistler wave at the great height. They showed that the transmission coefficients for these two systems are the same, and the polarisations of the incident wave in system (1) and of the downgoing wave below the ionosphere in system (2) are simply related.

For a stratified system a large number of relations between the reflection and transmission coefficients of two wave systems is possible, with incidence either from below or from above. This has been studied by Suchy and Altman (1975b), Altman and Suchy (1979a, b, 1980), Altman, Schatzberg and Suchy (1981), and in a series of papers by Heading (1971, 1973, 1975a, b, 1977c, 1980a). Heading finds that the structure of these relations is not clearly revealed if the equations are studied only for a three-dimensional space, and he has therefore generalised the equations so as to use an n-dimensional Cartesian axis system.

18.13. Resonance

Previous sections of this chapter have shown how the differential equations in the various forms, for example (7.80) or (18.39) or (18.49), can be integrated with respect to the height z. At the beginning of § 18.9 it was pointed out that nearly all points of

the complex z plane are ordinary points of the basic differential equations (7.80), where the coefficients $\mathbf{T}$ are bounded and the dependent variables $\mathbf{e}$ all behave smoothly. Thus the path of integration may be chosen in any convenient way. Nearly always part of the real z axis is used. But the property fails where elements of $\mathbf{T}$ are infinite and (7.81), (7.82) show that this occurs where $1 + M_{zz} \equiv \varepsilon_{zz} = 0$, at $z = z_\infty$ say. This point is called a point of resonance. It is also the condition that the coefficient α (6.16) or (6.17) or (6.23) of the Booker quartic equation is zero so that one solution q is infinite. For a plasma with electrons only the condition is

$$\varepsilon_{zz} \equiv 1 - X(U^2 - l_z^2 Y^2)/\{U(U^2 - Y^2)\} = 0. \tag{18.72}$$

Usually in practice the zero of ε_{zz} is simple, so the point z_∞ is a simple pole of the non-zero elements of $\mathbf{T}$. It is also a singular point of the basic differential equations (7.80). It can be shown (Budden, 1979, §§2.4) that at least one of the four solutions behaves like $(z - z_\infty)^p$ where the power p is complex. Thus z_∞ is an essential singularity for this solution. A branch cut must be inserted running from z_∞ to infinity. It must not cross the real z axis because the physical solution must be continuous there.

There can be more than one point z_∞ where (18.72) is satisfied, but usually there is only one that needs attention when the integration is planned. If collisions are neglected, $U = 1$ and (18.72) is satisfied for some real value of X. Thus there can be a point z_∞ on the real z axis. It is at this point that the infinities occur of n^2 in figs. 4.3, 4.6 and of q in figs. 6.4–6.7. In the lower part of the ionosphere X is usually an increasing function of z. If a small amount of collision damping is introduced so that $U = 1 - \mathrm{i}Z$, then for (18.72) X must have a small negative imaginary part, and therefore $\mathrm{Im}(z_\infty)$ is negative. The branch cut must then run to infinity in the negative imaginary half of the complex z plane. In rarer cases, when $\mathrm{Im}(z_\infty)$ is positive, the branch cut must be in the positive imaginary half plane.

For calculations of ionospheric reflection coefficients at very low frequency, of order 10 to 20 kHz, Z is large, of order 30, and z_∞ is not near the real z axis. Then the resonance does not give any trouble. For higher frequencies, however, z_∞ may be too near the real z axis. Near z_∞ one value, say q_1, of q is very large, and one possible W.K.B. solution for $\mathbf{e}$ behaves like $\exp(-ik\int q_1\,\mathrm{d}z)$, so that it changes very rapidly with z. The usual step size δz of about 1/50 of a vacuum wavelength is therefore too large. There are two ways of dealing with this.

For the first way the real z axis is used as the path of integration and a smaller δz is used where z is near to $\mathrm{Re}(z_\infty)$. This part of the integration then takes longer. It is usually necessary to make several trials to find a suitable δz.

For the second way the path of integration is moved so that it does not go near to z_∞. The new path must not cross the branch cut so that if $\mathrm{Im}(z_\infty)$ is negative, the path is moved into the positive imaginary half of the z plane. This method was used by Smith, M.S. (1974). There are several ways of applying it. For example the path can

be in four straight sections. Let a and b be real. Then the first section is the real z axis down to $z = \mathrm{Re}(z_\infty) + a$. The second is the straight line from $\mathrm{Re}(z_\infty) + a$ to $\mathrm{Re}(z_\infty) + ib$, and the third is from here to $\mathrm{Re}(z_\infty) - a$. The fourth is the real z axis below $\mathrm{Re}(z_\infty) - a$. An alternative way is to use a new variable of integration s say, that is real on the displaced part of the path. Then for example let

$$z = \mathrm{Re}(z_\infty) + s + ib(1 - s^2/a^2) \tag{18.73}$$

so that

$$\mathrm{d}z/\mathrm{d}s = 1 - 2ibs/a^2. \tag{18.74}$$

The real z axis is used for $z \geqslant \mathrm{Re}(z_\infty) + a$. Then (18.73) is used for $a \geqslant s \geqslant -a$, and finally the real z axis is again used for $z \leqslant \mathrm{Re}(z_\infty) - a$. The displaced path (18.73) is then a parabola in the complex z plane. For the middle stage here, the auxiliary subroutine must be modified. It must now include the calculation of z from (18.73), and when the derivative $\mathrm{d}v/\mathrm{d}z$ of any dependent variable v has been computed, it must be multiplied by (18.74) to convert it to $\mathrm{d}v/\mathrm{d}s$.

This technique of the displaced path can be used for the basic equations (7.80) or for the non-linear forms (18.39) or (18.49). In tests of the method Smith, M.S. (1974) found that the displaced path must not go too far from the real z axis because this may make the value of the reflection coefficient matrix $\boldsymbol{R}$ too small for accurate calculation.

The position of z_∞ is independent of the angle of incidence θ. Once a suitable integration path has been found, it can be used for a number of calculations with a sequence of values of θ.

A displaced path can also be used for avoiding failure of the form (18.39) of the equations, as illustrated by (18.43). Where this occurs, the variable $\boldsymbol{A}$ has simple poles. The integration path must not go too near these, but they are not branch points so the path may pass on either side. For an ionosphere with small losses these poles are near the real axis. It is then useful to use a line with fixed non-zero $\mathrm{Im}(z)$ for most of the integration path. Computing is then often quicker than with the more complicated equation (18.49) on the real axis.

PROBLEMS 18

18.1. A linearly polarised radio wave is vertically incident on the ionosphere from below. The earth's magnetic field is vertical and $Y = \frac{3}{8}$. The electron concentration increases linearly with height z so that $X = \alpha z$. Electron collisions are to be neglected. Prove that the reflected wave is always linearly polarised and show that its plane of polarisation is parallel to that of the incident wave when $\alpha = 1/r\lambda$, approximately, where r is an integer and λ is the vacuum wavelength. What happens if $Y > 1$?

18.2. In a full wave solution for finding reflection coefficients of the ionosphere at

very low frequency, two downward integrations of the basic differential equations are done. The angle of incidence $\theta = \arccos C$ where C is positive. The first integration uses a starting solution for the non-penetrating mode so that it is almost evanescent at the top. At the bottom, in the free space below the ionosphere the field components of the upgoing part of this solution are found to be $E_x, E_y, \mathscr{H}_x, \mathscr{H}_y = C{:}\rho_1{:} - C\rho_1{:}1 = \boldsymbol{e}^{(a)T}$. For the second integration the field components of the upgoing part of the solution at the bottom are $C{:}\rho_2{:} - C\rho_2{:}1 = \boldsymbol{e}^{(b)T}$. To find the solution for the penetrating mode a combination $\gamma\boldsymbol{e}^{(a)T} + \boldsymbol{e}^{(b)T}$ is now to be found such that the vertical power flux in the upgoing part of the solution below the ionosphere is as small as possible. What value of the constant γ is needed for this? Explain what happens when (a) $\rho_1 = -\rho_2 = \mathrm{i}$, (b) $\rho_1 = \rho_2$, (c) $\rho_1 = 0, \rho_2$ bounded and $\neq 0$.

18.3. A and B are two points in a space where there are magnetoplasmas. Prove that for any aerial system at A it is possible to construct at least one aerial system at B so that there is reciprocity between the terminals of the two aerials.
Note: This theorem was enunciated and proved by Goubau (1942) and is sometimes called 'Goubau's reciprocity theorem'. For a proof that uses the notation of this book see Budden (1961a, § 23.3).