7

Slowly varying medium. The W.K.B. solutions

7.1. Introduction

This chapter continues the discussion of radio waves in a stratified ionosphere and uses a coordinate system x, y, z as defined in §6.1. Thus the electric permittivity $\varepsilon(z)$ of the plasma is a function only of the height z. It is assumed that the incident wave below the ionosphere is a plane wave (6.1) with $S_2 = 0$, $S_1 = S$ so that wave normals, at all heights z, are parallel to the plane $y = \text{constant}$, and for all field components (6.48) is satisfied. It was shown in § 6.10 that if the four components E_x, E_y, $\mathscr{H}_x$, $\mathscr{H}_y$ of the total field are known at any height z, they can be expressed as the sum of the fields of the four characteristic waves, with factors f_1, f_2, f_3, f_4. In a homogeneous medium the four waves would be progressive waves, and these factors would be $\exp(-ikq_i z)$, $i = 1, 2, 3, 4$. We now enquire how they depend on z in a variable medium. This question is equivalent to asking whether there is, for a variable medium, any analogue of the progressive characteristic waves in a homogeneous medium. The answer is that there is no exact analogue. There are, however, approximate solutions, the W.K.B. solutions, which have many of the properties of progressive waves.

This problem has proved to be of the greatest importance in all those branches of physics concerned with wave propagation. It has intrigued not only physicists but many mathematicians, and the subject has been taken to levels of complication that are beyond the interests of most physicists. The studies have been devoted almost entirely to wave fields that satisfy differential equations of the second order. In the radio problem this usually means that the plasma medium must be assumed to be isotropic. An anisotropic ionosphere leads, in general, to the equivalent of a fourth order differential equation. Here the studies have been less extensive.

Most of this chapter is devoted to the isotropic ionosphere, that is to the solutions of second order differential equations. The anisotropic ionosphere is discussed in §§ 7.13 onwards.

7.2. The differential equations for an isotropic ionosphere

For an isotropic ionosphere the earth's magnetic field is neglected and the electric permittivity of the plasma is a scalar

$$\varepsilon(z) = n^2 = 1 - X/U \tag{7.1}$$

given by (3.17). Then $\boldsymbol{D} = \varepsilon_0 n^2 \boldsymbol{E}$, (2.29), and, from (6.6), the values of q are in two equal pairs

$$q = \pm(n^2 - S^2)^{\frac{1}{2}} = \pm(C^2 - X/U)^{\frac{1}{2}}. \tag{7.2}$$

When using Maxwell's equations (2.45) we may still use (6.48), but (6.49) must not be used because it is the z dependence of the fields that is to be studied. Thus (2.45) give

$$\frac{\partial E_y}{\partial z} = \mathrm{i}k\mathscr{H}_x, \quad SE_y = \mathscr{H}_z, \quad \frac{\partial \mathscr{H}_x}{\partial z} = \mathrm{i}k(n^2 E_y - S\mathscr{H}_z) \tag{7.3}$$

and

$$\frac{\partial E_x}{\partial z} = -\mathrm{i}k(\mathscr{H}_y + SE_z), \quad \frac{\partial \mathscr{H}_y}{\partial z} = -\mathrm{i}kn^2 E_x, \quad S\mathscr{H}_y = -n^2 E_z. \tag{7.4}$$

The equations have here been deliberately separated into the two sets. The first (7.3) depend only on E_y, $\mathscr{H}_x$, $\mathscr{H}_z$, and the second (7.4) only on E_x, E_z, $\mathscr{H}_y$. Thus if one of these sets is zero at any height z, it remains zero at all heights. The equations separate into two independent sets (7.3) and (7.4).

The set (7.3) is the simpler and will be discussed first. The electric field at all levels has only the component E_y, which is horizontal. The waves are sometimes said to be 'horizontally polarised' at all levels. The magnetic field has no $\mathscr{H}_y$ component so that it is in the plane of incidence. For the other set (7.4), E_y is everywhere zero. The waves are said to be 'vertically polarised' because $\boldsymbol{E}$ lies in the vertical plane of incidence $y =$ constant. This case is a little more complicated and is discussed in §7.12.

From (7.3), $\mathscr{H}_z$ may be eliminated, and (7.2) is used to give

$$\frac{\partial E_y}{\partial z} = \mathrm{i}k\mathscr{H}_x, \quad \frac{\partial \mathscr{H}_x}{\partial z} = \mathrm{i}k\{q(z)\}^2 E_y. \tag{7.5}$$

It is at once clear that there is no solution of (7.5) that represents a 'progressive' wave. For in a progressive wave all field quantities vary with z through the same factor $\exp\{\mathrm{i}\phi(z)\}$. Now the first equation (7.5) would require that $\mathrm{d}\phi/\mathrm{d}z$ is a constant while the second would require that it is proportional to $\{q(z)\}^2$, which is only possible if q is a constant. In spite of this, it is possible to derive *approximate* solutions of (7.5) which have many of the properties of progressive waves. These are the W.K.B. solutions. At every place where q is varying, the reflection process is going on, and it is this which prevents the occurrence of a true progressive wave. Except in certain places, however, the reflection process is very weak, so that the

W.K.B. solutions are often very good approximations. This argument is illustrated in more detail in later sections.

Elimination of $\mathscr{H}_x$ from (7.5) gives

$$\frac{\mathrm{d}^2E_y}{\mathrm{d}z^2}+k^2q^2E_y=0. \tag{7.6}$$

This is a very important differential equation satisfied by the electric field E_y. It is assumed that the factor $\exp\{\mathrm{i}(\omega t-kSx)\}$ is omitted so that E_y is a function of z only. We therefore use the total derivative sign $\mathrm{d}/\mathrm{d}z$ instead of the partial derivative sign $\partial/\partial z$.

Equation (7.6) is of the form which occurs in the study of nearly all kinds of wave propagation. For example if ψ is the wave function of a particle of mass m and of constant total energy E, moving in a one-dimensional potential field $V(z)$, the Schrödinger equation is

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}z^2}+\frac{2m}{\hbar^2}\{E-V(z)\}\psi=0 \tag{7.7}$$

so that q^2 in (7.6) is analogous to the classical kinetic energy $E-V$ in (7.7). Again if p is the excess pressure in a plane sound wave of small amplitude travelling in a gas of molecular weight M and temperature T, the wave equation is

$$\frac{\mathrm{d}^2p}{\mathrm{d}z^2}+\frac{\omega^2M}{\gamma RT}p=0 \tag{7.8}$$

where R is the gas constant, and γ is the ratio of specific heats; M, γ and T may be functions of z. Further examples will be found in Problems 7 at the end of this chapter. Results which will now be derived for (7.6) can be adapted to apply to these other equations.

There are other kinds of wave propagation where the differential equation is less simple than (7.6) and may be of higher order than the second, as occurs, for instance, in the study of waves in anisotropic media but in nearly all cases, the W.K.B. method of 7.6 can still be used.

7.3. The phase memory concept

Suppose that in (7.5) and (7.6) the variation of $q(z)$ with z is so slow that a range of z can be chosen where q can be treated as constant. Consider a progressive wave in this region travelling in the direction of positive z. Then its field components would be given by

$$E_y=A\exp(-\mathrm{i}kqz),\quad \mathscr{H}_x=-Aq\exp(-\mathrm{i}kqz) \tag{7.9}$$

where A is constant to a first approximation. Now the first equation (7.9) ought to satisfy (7.6) but by direct substitution it can be seen that the fit is not very good unless $\mathrm{d}q/dz$ and $\mathrm{d}^2q/\mathrm{d}z^2$ are very small.

If q were real, the part kqz of the exponents in (7.9) would be a real angle called the 'phase' of the wave. Since q is in general complex, the angle kqz is complex but it is still called the (complex) phase. Its real part is the same as the phase angle as ordinarily understood, and its imaginary part gives an additional term in the exponent. This term is real because of the factor $-\mathrm{i}$, and therefore affects only the amplitude of the wave. When the wave represented by (7.9) passes through an infinitesimal thickness δz of the medium, the change in its (complex) phase is $kq\delta z$, which therefore depends on q. When it passes through a finite thickness, from $z=0$ to $z=z_1$, say, the change of phase is $k\int_0^{z_1} q\,\mathrm{d}z$. This suggests that a better solution than (7.9) might be

$$E_y = A\exp(-\mathrm{i}k\int_0^z q\,\mathrm{d}z), \quad \mathscr{H}_x = -Aq\exp(-\mathrm{i}k\int_0^z q\,\mathrm{d}z). \tag{7.10}$$

These satisfy the first equation (7.5) exactly, but the second is approximately satisfied only when $\mathrm{d}q/\mathrm{d}z$ is small. The agreement is better, however, than for (7.9). The integral in (7.10) constitutes what is sometimes called the 'phase memory' concept. It expresses the idea that a change of phase is cumulative for a wave passing through a slowly varying medium. The solution (7.10) is adequate for the discussion of many wave propagation problems. An extension of it to three dimensions forms the basis of much of the theory of geometrical optics or ray theory, ch. 14. But for some purposes a still more accurate solution is needed. This is the W.K.B. solution, derived in the following sections.

7.4. Loss-free medium. Constancy of energy flow

To obtain a solution that is more accurate than (7.10) it is convenient to treat A as a function of z. One method of deriving an approximation to this function can be used if the medium is loss-free. We therefore now consider the special case where n^2 is real, and we study a range of z where q^2 is positive. Then, in a progressive wave, the average flow of energy in the z direction must be the same for all z. It is proportional to the z component Π_z of the time averaged Poynting vector, (2.63), that is to

$$[\mathrm{Re}(\boldsymbol{E}\wedge\boldsymbol{\mathscr{H}}^*)]_z = -\tfrac{1}{2}(E_y\mathscr{H}_x^* + E_y^*\mathscr{H}_x). \tag{7.11}$$

If q is positive, the exponents in (7.10) are purely imaginary and (7.11) is proportional to qAA^* which must be constant. Hence A is proportional to $q^{-\frac{1}{2}}$, so that the solution becomes

$$E_y = A_1 q^{-\frac{1}{2}}\exp\left(-\mathrm{i}k\int_0^z q\,\mathrm{d}z\right), \quad \mathscr{H}_x = -A_1 q^{\frac{1}{2}}\exp\left(-\mathrm{i}k\int_0^z q\,\mathrm{d}z\right) \tag{7.12}$$

where A_1 is a constant. The argument applies also to a wave travelling in the direction of negative z, so that another solution is

$$E_y = A_2 q^{-\frac{1}{2}} \exp\left(ik\int_0^z q\,\mathrm{d}z\right), \quad \mathscr{H}_x = A_2 q^{\frac{1}{2}} \exp\left(ik\int_0^z q\,\mathrm{d}z\right) \tag{7.13}$$

where A_2 is another constant.

These two pairs are the two W.K.B. solutions. They are approximate because the reflection process, here neglected, is actually occurring at all values of z where $\mathrm{d}q/\mathrm{d}z \neq 0$, so that the assumption that the energy flow is constant is not strictly true.

The above derivation applies only when q is real and positive, but it is useful because it illustrates the physical significance of W.K.B. solutions. In § 7.6 another derivation is given which applies for any value of q including complex and purely imaginary values.

The argument used here for a loss-free system may be applied to other kinds of wave. Consider, for example, the Schrödinger equation (7.7) for a particle in a potential field $V(z)$. Let the wave function $\psi(z)$ represent a progressive wave travelling in the positive z direction. Then the first equation (7.10) is replaced by:

$$\psi = \Psi \exp\left[i\int_0^z \{(E-V)2m/\hbar^2\}^{\frac{1}{2}}\mathrm{d}z \right] \tag{7.14}$$

(in accordance with the invariable practice in wave mechanics it is here assumed that the suppressed time factor is $e^{-i\omega t}$). If $V(z)$ varies sufficiently slowly, the flux of the particles must be approximately constant so that

$$\psi\frac{\mathrm{d}\psi^*}{\mathrm{d}z} - \psi^*\frac{\mathrm{d}\psi}{\mathrm{d}z} = \text{constant} \tag{7.15}$$

(see, for example, Mott, 1952, § II.5). Substitution from (7.14) gives $\Psi\Psi^* \propto (E-V)^{-\frac{1}{2}}$ which suggests that

$$\psi = \psi_0 (E-V)^{-\frac{1}{4}} \exp\left[i\int_0^z \{(E-V)2m/\hbar^2\}^{\frac{1}{2}}\mathrm{d}z \right]. \tag{7.16}$$

This is one W.K.B. solution. The other has a minus sign in the exponent and represents a particle travelling in the negative z direction.

7.5. W.K.B. solutions

The solutions that are here called W.K.B. solutions were used in wave-mechanical problems independently by Wentzel (1926), Kramers (1926) and Brillouin (1926). The name W.K.B. is derived from the initials of these three authors. These solutions had been studied earlier by Jeffreys (1924) and some authors refer to them as J.–W.K.B. or W.K.B.–J. solutions. There are, however, still earlier accounts of this type of solution. They were used by Gans (1915) in the study of radio waves in the ionosphere, and by Lord Rayleigh (1912) in the study of sound waves. Jeffreys and Jeffreys (*Methods of Mathematical Physics*, 1972) call them 'asymptotic approximations of Green's type' because Green used them in 1837 in the study of waves on the surface of water. They also draw attention to the use of this kind of approximation by Carlini in 1817. Some authors, for example Olver (1974), now refer to these solutions as L.G. approximations, after Liouville (1837) and Green (1837) (see § 7.9).

The term 'W.K.B. solutions' will be used here because it is well known and widely used. For the history of the subject and further references, see Olver (1974), Heading (1962a).

There seems to be no generally accepted *precise* definition of the term 'W.K.B. solution' that covers all cases. Many of the mathematical treatments study ordinary differential equations of the second order (Jeffreys and Jeffreys, 1972; Heading, 1962a; Olver, 1974) and it is then often possible to formulate a precise definition. But equations of higher order have not received such detailed study. For equations of the fourth order the subject is discussed in §§ 7.13–7.17.

7.6. The W.K.B. method

The W.K.B. solutions of the differential equation (7.6) are now to be derived by a method which is not restricted to real values of q. When q is complex, the convention is adopted, in this section, that its real part is positive.

In a homogeneous medium there are two solutions of (7.6) which represent progressive waves. In both of them the field E_y depends on z, only through a factor $\exp\{i\phi(z)\}$ where $\phi(z) = \mp kqz$. The function $\phi(z)$ is called the generalised phase of the wave and it is in general complex because q is complex. In a homogeneous medium q is independent of z so that $d\phi/dz = \mp kq$, and $d^2\phi/dz^2 = 0$. This suggests that we try to find the function $\phi(z)$ for a slowly varying medium, and we may expect that $d^2\phi/dz^2$ is very small, so that its square and higher powers may be neglected. This idea is the basis of the following method. The function $\phi(z)$ is an example of the eikonal function (Sommerfeld, 1954, §§ 35, 48; see also § 14.2).

Let

$$E_y = A \exp\{i\phi(z)\} \tag{7.17}$$

where A is a constant. Then

$$\frac{d^2E_y}{dz^2} = A\left\{i\frac{d^2\phi}{dz^2} - \left(\frac{d\phi}{dz}\right)^2\right\}e^{i\phi}. \tag{7.18}$$

When this is substituted in the differential equation (7.6) it shows that $\phi(z)$ must satisfy:

$$\left(\frac{d\phi}{dz}\right)^2 = k^2q^2 + i\frac{d^2\phi}{dz^2}. \tag{7.19}$$

This is a non-linear differential equation which would be very difficult to solve exactly, but an approximate solution may be found as follows. Since $d^2\phi/dz^2$ is small, it may be neglected to a first approximation. Then

$$\frac{\mathrm{d}\phi}{\mathrm{d}z} \approx \mp kq \tag{7.20}$$

and thence

$$\frac{\mathrm{d}^2\phi}{\mathrm{d}z^2} \approx \mp k\frac{\mathrm{d}q}{\mathrm{d}z}. \tag{7.21}$$

The approximate value (7.21) is now substituted in (7.19) which gives as a second approximation

$$\frac{\mathrm{d}\phi}{\mathrm{d}z} \approx \left\{k^2q^2 \mp \mathrm{i}k\frac{\mathrm{d}q}{\mathrm{d}z}\right\}^{\frac{1}{2}}. \tag{7.22}$$

The square root is expanded by the binomial theorem, and since the second term is small, only two terms of the expansion need be retained. The result is

$$\frac{\mathrm{d}\phi}{\mathrm{d}z} \approx \mp kq\left\{1 \mp \frac{\mathrm{i}}{2kq^2}\frac{\mathrm{d}q}{\mathrm{d}z}\right\}. \tag{7.23}$$

Now the second minus sign in (7.23) was obtained from the minus sign in the first approximation (7.20) and for this case the first sign in (7.23) must be minus. Similarly, if the second sign is plus, the first sign must be plus also. Hence (7.23) becomes

$$\frac{\mathrm{d}\phi}{\mathrm{d}z} \approx \mp kq + \frac{\mathrm{i}}{2q}\frac{\mathrm{d}q}{\mathrm{d}z}. \tag{7.24}$$

which may be integrated at once to give

$$\phi \approx \mp k\int^z q\,\mathrm{d}z + \mathrm{i}\ln(q^{\frac{1}{2}}). \tag{7.25}$$

The constant of integration may be made zero by a suitable choice of the lower limit of the integral, which is left unspecified here, since its value does not affect the way in which $\phi(z)$ depends on z. Substitution of (7.25) into (7.17) gives

$$E_y \approx Aq^{-\frac{1}{2}}\exp\left\{\mp \mathrm{i}k\int^z q\,\mathrm{d}z\right\}. \tag{7.26}$$

These two solutions are the W.K.B. solutions of (7.6).

If (7.26) is substituted in the first equation (7.5) it gives

$$\mathscr{H}_x = A\left[\mp q^{\frac{1}{2}} - \tfrac{1}{2}q^{-\frac{3}{2}}\frac{\mathrm{d}q}{\mathrm{d}z}\right]\exp\left\{\mp \mathrm{i}k\int^z q\,\mathrm{d}z\right\}. \tag{7.27}$$

Here the second term in the square brackets is small compared with the first, for a slowly varying medium, and can usually be neglected.

It is stressed that this derivation of the W.K.B. solutions for an isotropic stratified medium is valid for the most general case where q is a complex variable function.

The process can be continued to higher orders of approximation to obtain a still more accurate solution. This is algebraically complicated, however, and it is usually more profitable to use other methods to get a more accurate solution; see, for example, § 8.11 equation (8.30) onwards.

7.7. Discrete strata

The physical significance of W.K.B. solutions is further illustrated by the following alternative derivation, which was used by Bremmer (1951). See also Lord Rayleigh (1912), Brekhovskikh (1960, § 16), Bremmer (1949). It is supposed that the slowly varying medium is divided, by planes $z =$ constant, into thin strata in each of which n and q are constant (compare § 6.2). By making the strata sufficiently thin and numerous it is possible to approximate as closely as desired to the actual medium. Suppose that for $z < 0$ the medium is homogeneous with $q = q_0$. Let the strata be of small thickness h and let them be numbered consecutively for $z > 0$, so that the values of q are q_1, q_2, etc. Let a plane wave

$$E_y = \exp(-ikq_0 z), \quad \mathscr{H}_x = -q_0 \exp(-ikq_0 z) \tag{7.28}$$

(compare (7.9)) be incident from the direction of negative z. When it reaches the first boundary plane at $z = 0$ it gives rise to a reflected wave

$$E_y = R_0 \exp(ikq_0 z), \quad \mathscr{H}_x = R_0 q_0 \exp(ikq_0 z) \quad \text{for } z < 0 \tag{7.29}$$

and a transmitted wave

$$E_y = T_1 \exp(-ikq_1 z), \quad \mathscr{H}_x = -T_1 q_1 \exp(-ikq_1 z) \quad \text{for } 0 < z < h. \tag{7.30}$$

On crossing a boundary the components of electric and magnetic intensity parallel to it must be continuous. Hence, for the boundary at $z = 0$

$$1 + R_0 = T_1, \quad q_0(1 - R_0) = q_1 T_1 \tag{7.31}$$

whence

$$R_0 = (q_0 - q_1)/(q_0 + q_1), \quad T_1 = 2q_0/(q_0 + q_1). \tag{7.32}$$

The transmitted wave now impinges on the next boundary at $z = h$, and is there partially reflected and partially transmitted. The reflected wave from this boundary is

$$E_y = R_1 T_1 \exp(-ikq_1 h) \exp\{ikq_1(z - h)\} \quad \text{for } 0 < z < h \tag{7.33}$$

and the transmitted wave is

$$E_y = T_1 T_2 \exp(-ikq_1 h) \exp\{-ikq_2(z - h)\} \quad \text{for } h < z < 2h \tag{7.34}$$

where, by analogy with (7.32):

$$R_1 = (q_1 - q_2)/(q_1 + q_2), \quad T_2 = 2q_1/(q_1 + q_2). \tag{7.35}$$

The medium is assumed to be slowly varying so that the differences $q_0 - q_1, q_1 - q_2$, etc. for adjacent strata are very small. Thus, provided that q_0, q_1, q_2 etc. are not too small, the reflected waves have small amplitude.

The reflected waves now impinge on the boundaries and are again reflected. These second reflections add on to the waves travelling in the direction of positive z. They have extremely small amplitude, however, and as a first approximation they can be neglected. The reason is as follows. By making the thickness h of the strata and the differences $q_{i+1} - q_i$ small enough, the reflection coefficients R_0, R_1 etc. can be made indefinitely small. But this also makes the reflecting boundaries indefinitely numerous, so that the numerous first reflections have to be added. The addition must take account of the amplitude and phase of the contributions and is conveniently done by an amplitude–phase diagram of the kind used in physical optics (see, for example, Lipson and Lipson, 1969, § 6.4.5; or Jenkins and White, 1976, ch. 12). When the wave (7.33) reflected from the boundary at $z = h$ reaches the origin it is given by

$$E_y = R_1 T_1 \exp(-2ikq_1 h) \tag{7.36}$$

where the phase $2kq_1h$ contains the factor 2 because the wave has made two traverses of the stratum $0 \leqslant z \leqslant h$. Similarly when the wave reflected from $z = 2h$ reaches the origin it is given by

$$E_y = R_2 T_1 T_2 \exp\{-2ikh(q_1 + q_2)\} \tag{7.37}$$

and in general, for the boundary at $z = mh$

$$E_y = R_m T_1 T_2 \ldots T_m \exp\{-2ikh(q_1 + q_2 + \ldots + q_m)\}. \tag{7.38}$$

Thus there is a phase difference $2khq_1$ between successive contributions to this sequence, so that the amplitude–phase diagram is a polygon resembling a tightly wrapped spiral. The resultant amplitude of all the first reflections is therefore very small for two reasons: (i) the contributing vectors such as (7.38) are small because R_m is small, and (ii) the phase difference between successive contributions gives the amplitude–phase diagram a large curvature. Consequently the resultant amplitude of the twice reflected waves is smaller still.

Neglect of the twice reflected waves means that the transmitted wave in the m^{th} stratum is given by

$$E_y = A \exp\left[-ik\{h(q_1 + q_2 + \ldots + q_{m-1}) + q_m(z - h)\}\right] \tag{7.39}$$

where

$$A = T_1 T_2 \ldots T_m, \quad T_m = 2q_{m-1}/(q_{m-1} + q_m). \tag{7.40}$$

In going from the $m - 1^{\text{th}}$ stratum to the m^{th} stratum the increment of $\ln A$ is

$$\Delta(\ln A) = \ln T_m = \ln\left\{\left(1 + \frac{q_m - q_{m-1}}{2q_{m-1}}\right)^{-1}\right\} \approx -\tfrac{1}{2}\frac{\Delta q}{q}. \tag{7.41}$$

If it is assumed that the strata are so thin that this relation can be integrated, it gives

$A = Kq_m^{-\frac{1}{2}}$ and because $A = 1$ when $z < 0$, the constant K is $q_0^{\frac{1}{2}}$. The sum in the exponent of (7.39) can similarly be expressed as an integral so that (7.39) gives

$$E_y = Kq^{-\frac{1}{2}} \exp\left\{ -\mathrm{i}k \int_0^z q \, \mathrm{d}z \right\} \tag{7.42}$$

which is one W.K.B. solution.

It may happen that the wave (7.42) continues on to a region where q_m is small. Then the reflection coefficients $R_m = (q_m - q_{m+1})/(q_m + q_{m+1})$ are no longer small. Similarly the phase change $2khq_m$ between waves reflected from adjacent boundaries is very small, so that successive contributions augment each other. In this region the reflection process is strong and the W.K.B. solutions fail. A reflected wave is generated which travels in the negative z direction. After it has left this region of reflection, its field is given by the other W.K.B. solution.

For further discussion of the use of discrete strata see §18.3.

7.8. Coupling between upgoing and downgoing waves

The discussion in § 7.7 showed that a wave represented by one W.K.B. solution may give rise, as it travels, to another wave represented by the other W.K.B. solution, but that this process is negligible provided that q varies slowly enough with z, and is not small. This idea is illustrated by the following alternative derivation of the W.K.B. solutions.

In a homogeneous medium a progressive wave travelling obliquely in the direction of increasing z has field components E_y, $\mathscr{H}_x$ related by $\mathscr{H}_x = -qE_y$, from (7.9), and a progressive wave travelling obliquely in the direction of decreasing z has $\mathscr{H}_x = qE_y$. For an inhomogeneous medium these relations may be used to define the analogues of the progressive waves. Hence let

$$E_y = E_y^{(1)} + E_y^{(2)}, \quad \mathscr{H}_x = \mathscr{H}_x^{(1)} + \mathscr{H}_x^{(2)} \tag{7.43}$$

where

$$\mathscr{H}_x^{(1)} = -qE_y^{(1)}, \quad \mathscr{H}_x^{(2)} = qE_y^{(2)}. \tag{7.44}$$

These are used in the Maxwell equation (7.5) to give

$$\left.\begin{aligned} \frac{\mathrm{d}E_y^{(1)}}{\mathrm{d}z} + \frac{\mathrm{d}E_y^{(2)}}{\mathrm{d}z} &= \mathrm{i}kq\mathscr{H}_x^{(1)} + \mathrm{i}kq\mathscr{H}_x^{(2)}, \\ -q\frac{\mathrm{d}E_y^{(1)}}{\mathrm{d}z} + q\frac{\mathrm{d}E_y^{(2)}}{\mathrm{d}z} &= \mathrm{i}kq^2E_y^{(1)} + \mathrm{i}kq^2E_y^{(2)} + E_y^{(1)}\frac{\mathrm{d}q}{\mathrm{d}z} - E_y^{(2)}\frac{\mathrm{d}q}{\mathrm{d}z}. \end{aligned}\right\} \tag{7.45}$$

Divide the second of these by q and subtract it from, or add it to, the first. This gives the two equations

$$\frac{\mathrm{d}E_y^{(1)}}{\mathrm{d}z} + \mathrm{i}kqE_y^{(1)} + \frac{1}{2q}\frac{\mathrm{d}q}{\mathrm{d}z}E_y^{(1)} = \frac{1}{2q}\frac{\mathrm{d}q}{\mathrm{d}z}E_y^{(2)}, \tag{7.46}$$

$$\frac{dE_y^{(2)}}{dz} - ikqE_y^{(2)} + \frac{1}{2q}\frac{dq}{dz}E_y^{(2)} = \frac{1}{2q}\frac{dq}{dz}E_y^{(1)}. \tag{7.47}$$

These are coupled equations in the sense of the definition given later in § 16.1. They may be solved by successive approximations. The right-hand sides both contain the factor $\frac{1}{2}q^{-1}\,dq/dz$ which is small in a slowly varying medium except near where $q = 0$. As a first approximation, therefore, the right-hand sides are neglected. This is in effect the same assumption as was made in (7.20) where $d^2\phi/dz^2$ was neglected. Then (7.46), (7.47) give two independent differential equations for $E_y^{(1)}$, $E_y^{(2)}$ whose solutions are

$$\left.\begin{matrix} E_y^{(1)} \\ E_y^{(2)} \end{matrix}\right\} = q^{-\frac{1}{2}} \exp\left(\mp ik\int^z q\,dz\right). \tag{7.48}$$

These are simply the W.K.B. solutions. This method of deriving them shows how they arise by neglecting the coupling process, which in this case is a reflection process. The method is extended later, § 7.15, to deal with the anisotropic ionosphere.

The solutions (7.48) may be substituted in the right-hand sides of (7.46), (7.47) and the equations solved to give a better approximation. An example where this is useful is as follows. Suppose that a plane radio wave travels obliquely upwards in the earth's troposphere. Here the refractive index is very close to unity at all levels, but there can be small variations and there can be stratified layers caused by water vapour or by temperature changes with height. The upgoing wave, (7.48) upper sign, would travel right through such a layer, and any resulting downgoing wave would have such an extremely small amplitude that the right-hand side of (7.46) could be neglected at all levels. But the very weak reflection from a layer in the troposphere can sometimes be strong enough to be detected when it reaches the ground. It is therefore important to be able to calculate its amplitude. The solution (7.48) for $E_y^{(1)}$ is substituted in the right-hand side of (7.47). The resulting inhomogeneous first order differential equation for $E_y^{(2)}$ can then be solved at once by using an integrating factor, to give the amplitude of this downgoing wave below the reflecting layer. The formula was given by Wait (1962, § IV.6) who applied it to some special cases and gave further references.

7.9. Liouville method and Schwarzian derivative

The form (7.26) of the W.K.B. solutions is rather complicated because of the integrals $\mp\int q\,dz$ in the exponents. These represent the phases of the two waves, and it would be useful if they could be expressed in simpler form. This suggests that we should try to use a new independent variable ζ, instead of z. It is equivalent to postulating a fictitious 'comparison' medium with a new 'height' coordinate ζ, instead of z. The idea is now to be applied to the differential equation (7.6) and it is convenient to omit the subscript y. A prime $'$ is used to denote $d/d\zeta$. Then

$$E' = \frac{\mathrm{d}E}{\mathrm{d}z}z', \quad E'' = \frac{\mathrm{d}^2E}{\mathrm{d}z^2}(z')^2 + \frac{\mathrm{d}E}{\mathrm{d}z}z'' \tag{7.49}$$

and with (7.6) this gives

$$E'' - E'z''/z' + Ek^2(z')^2q^2 = 0. \tag{7.50}$$

To remove the term in E', the dependent variable is changed to F, thus

$$E = (z')^{\frac{1}{2}}F \tag{7.51}$$

so that

$$F'' + F[(kqz')^2 - (z')^{\frac{1}{2}}\{(z')^{-\frac{1}{2}}\}''] = 0. \tag{7.52}$$

So far there are no approximations. From § 7.2 we know that there are no solutions that represent true progressive waves. We therefore now examine the conditions that will make the second term in the square brackets of (7.52) negligible compared to the first. If, then, there are true progressive wave solutions, the first term must be a constant. Hence let

$$qz' = 1, \text{ so that } \zeta = \int^z q\,\mathrm{d}z. \tag{7.53}$$

The lower limit of the integral affects only the origin of ζ and may be assigned later. With this choice of ζ, the 'comparison' medium is free space and the (approximate) progressive wave solutions of (7.52) are

$$F = \exp(\mp ik\zeta). \tag{7.54}$$

From this and (7.53), (7.51)

$$E = q^{-\frac{1}{2}}\exp(\mp ik\int^z q\,\mathrm{d}z) \tag{7.55}$$

which are the two W.K.B. solutions.

Liouville (1837) obtained solutions of this form and he used a change of independent variable similar to that used here. In the same year Green (1837) obtained similar solutions. He did not change the independent variable and his method was similar to that of § 7.6.

The neglected term in (7.52) is

$$-(z')^{\frac{1}{2}}\{(z')^{-\frac{1}{2}}\}'' = \tfrac{1}{2}\{(z')^{-1}z''' - \tfrac{3}{2}(z')^{-2}z''\} = \tfrac{1}{2}\mathscr{D}(z;\zeta) \tag{7.56}$$

and $\mathscr{D}$ is called the Schwarzian derivative of z with respect to ζ. Some of its properties are described by Heading (1975c). For the approximation to be justified it is necessary that $\frac{1}{2}\mathscr{D}$ is very small compared with the first term in the square brackets in (7.52). In terms of q and z, from (7.53), this gives

$$|\tfrac{1}{2}\mathscr{D}| = \left|\frac{3}{4}\left(\frac{1}{q^2}\frac{\mathrm{d}q}{\mathrm{d}z}\right)^2 - \frac{1}{2q^3}\frac{\mathrm{d}^2q}{\mathrm{d}z^2}\right| = \left|\frac{5}{16q^6}\left\{\frac{\mathrm{d}(q^2)}{\mathrm{d}z}\right\}^2 - \frac{1}{4q^4}\frac{\mathrm{d}^2(q^2)}{\mathrm{d}z^2}\right| \ll k^2. \tag{7.57}$$

This is sometimes used as a condition of validity of the W.K.B. solutions (7.55). An alternative way of deriving it is to substitute (7.55) into the differential equation (7.6). It does not fit exactly, and the expression that remains must be small compared with either term of (7.6). When the term k^2q^2E is used, the result is (7.57).

It must be stressed that the criterion (7.57) applies only when the differential equation has the simple form (7.6). It does not apply for vertical polarisation (7.4) (see § 7.12) nor for an anisotropic plasma (see §§ 7.13–7.16).

A criterion such as (7.57) is not entirely satisfactory because although it requires that $|\frac{1}{2}\mathscr{D}|$ is small, it does not say how small, and it does not tell us how good the approximation is when $|\frac{1}{2}\mathscr{D}|$ is known. One possible exact solution of (7.6) may be written

$$E = q^{-\frac{1}{2}}\exp\left(-\mathrm{i}k\int^z q\,\mathrm{d}z\right)\{1+\mathscr{E}(z)\}. \tag{7.58}$$

The great importance of the Liouville method is that it can be extended so as to specify an upper bound for the fractional error $|\mathscr{E}|$ of the W.K.B. solution. The subject is algebraically complicated and beyond the scope of this book. A very clear account is given by Olver (1974, ch. 6) who is the main originator of the subject.

7.10. Conditions for the validity of the W.K.B. solutions

For the W.K.B. solutions (7.26) to be good approximations, the criterion is (7.57). To get an estimate of how small the left-hand side must be, it is useful to study the special case where q^2 is a linear function of z, as in (8.2). Then an exact solution of (7.6) is the Airy integral function $\mathrm{Ai}(\zeta)$, ch. 8, where, from (8.6), $\zeta = -q^2(k/a)^{\frac{2}{3}} = (k^2a)^{\frac{1}{3}}(z-z_0)$. (This ζ is not the same as that in § 7.9). If these are substituted in the left-hand side of (7.57) it gives $\frac{5}{16}k^2/|\zeta^3|$. The W.K.B. solutions in this case are the same as the asymptotic forms for $\mathrm{Ai}(\zeta)$ and it is shown in § 8.10 that if $|\zeta| \geqslant 1$, the error in using them is less than about 8 to 9%. To achieve this degree of accuracy the criterion (7.57) gives

$$\frac{1}{k^2}\left|\frac{3}{4}\left(\frac{1}{q^2}\frac{\mathrm{d}q}{\mathrm{d}z}\right)^2 - \frac{1}{2q^3}\frac{\mathrm{d}^2q}{\mathrm{d}z^2}\right| \leqslant \frac{5}{16}. \tag{7.59}$$

It must be stressed that if $q^2(z)$ is not approximately linear the error in using W.K.B. solutions may exceed 8 to 9% even when (7.59) is satisfied.

In deriving the W.K.B. solutions (7.26) it was assumed that, in the binomial expansion of (7.22)

$$\frac{1}{k}\left|\frac{1}{q^2}\frac{\mathrm{d}q}{\mathrm{d}z}\right| = \frac{1}{k}\left|\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{q}\right)\right| \ll 1. \tag{7.60}$$

This is a necessary condition, but not sufficient because (7.59) shows that there is a restriction on $\mathrm{d}^2q/\mathrm{d}z^2$ that is not included. But (7.60) is simple, and adequate for most practical applications. It can be made more quantitative by using again the special case where q^2 is a linear function of z. Then for $|\zeta| \geqslant 1$ it can be shown that (7.60) gives

$$\frac{1}{k}\left|\frac{\mathrm{d}}{\mathrm{d}z}\frac{1}{q}\right| \leqslant 0.5. \tag{7.61}$$

This result can be extended so as to apply in an anisotropic ionosphere; see end of § 16.3.

Either (7.59) or (7.60) is a quantitative definition of the term 'slowly varying'. They require that the derivatives $\mathrm{d}q/\mathrm{d}z$ and, for (7.59), $\mathrm{d}^2q/\mathrm{d}z^2$ shall be sufficiently small, and that q is not too small. Because of the factor $1/k$ or $1/k^2$ the condition is most easily satisfied at high frequencies, but no matter how great the frequency nor how small the derivatives, the condition is certain to fail near values of z where q^2 passes through a zero. Then one W.K.B. solution can generate some of the other, and this constitutes the process of 'reflection'. The approximations made in §§ 7.6, 7.8 no longer apply, and to study the reflection process it is necessary to use more accurate solutions of the differential equations. It is these that are supplied by solutions of the Stokes equation, ch. 8, or other equations, ch. 15.

The condition (7.59) fails if the derivatives $\mathrm{d}q/dz$ and $\mathrm{d}^2q/\mathrm{d}z^2$ are large, even when q is not small. An extreme example of this is the reflection at the sharp boundary between two homogeneous media, where the derivatives are infinite. It is shown in § 15.16 that reflection at a steep gradient and reflection at a zero of q can be regarded as different aspects of the same phenomenon.

7.11. Properties of the W.K.B. solutions

The two W.K.B. solutions (7.26) were derived by starting with the idea of a progressive wave in a small region of an inhomogeneous stratified medium. Equation (7.6) has no exact solution that represents a purely progressive wave, except when q^2 is independent of z. In an inhomogeneous medium the W.K.B. solutions represent waves that are the analogues of the progressive waves travelling obliquely in the positive and negative z directions. Equation (7.27) shows that, if the small second term is neglected

$$\mathscr{H}_x = \mp qE_y \tag{7.62}$$

and (7.3) gives

$$\mathscr{H}_z = SE_y. \tag{7.63}$$

$S = \sin\Theta$ can in general be complex, but it is here assumed to be real. These relations are the same as for progressive waves in a homogeneous medium, for all values of q, except that E_y and q depend on z. Thus for either wave, and for all q, the horizontal components of the time averaged Poynting vector $\mathbf{\Pi}_{\mathrm{av}}$ are

$$\Pi_x = \tfrac{1}{2}Z_0^{-1}S|E_y^2|, \quad \Pi_y = 0 \tag{7.64}$$

so Π_x has the same sign as S and is zero only for vertical incidence, $S=0$. If q is real and positive, the minus sign in (7.62) gives $\Pi_z = \frac{1}{2}Z_0^{-\frac{1}{2}}q|E_y^2|$ so that the wave is obliquely upgoing. For the plus sign in (7.62), Π_z has a minus sign and the wave is obliquely downgoing.

In the lower ionosphere, if N increases monotonically with z, and if collisions are neglected, $q^2 = n^2 - S^2 = C^2 - X$ is at first real and decreases monotonically as z increases. Then (7.26) shows that the electric field E_y of either wave increases because of the factor $q^{-\frac{1}{2}}$, and (7.62) shows that the magnetic field decreases, because of the factor $q^{\frac{1}{2}}$. As the value z where $q=0$ is approached, the electric field would become indefinitely large according to (7.26), but before that happens, a place is reached where the W.K.B. solutions can no longer be used because the condition (7.59) is violated. When this region is passed, however, for still larger z, (7.59) is again satisfied and the solutions (7.26) are again valid. Hence two W.K.B. solutions are possible above the level of reflection. Here q^2 is negative and q is purely imaginary so that one W.K.B. solution is given by

$$E_y = Aq^{-\frac{1}{2}}\exp(-k\int^z |q|\,\mathrm{d}z), \quad \mathscr{H}_x = -qE_y. \tag{7.65}$$

These fields are now in quadrature so that Π_z is zero. Thus there is no average flow of energy in the z direction but the horizontal flow Π_x, (7.64) is still present if $S \neq 0$. There is, stored in the electromagnetic field, some energy which pulsates back and forth in the z direction with twice the wave frequency. Both fields (7.65) decrease as z increases because of the exponential, and for both of them the phase is independent of z. The wave is inhomogeneous, § 2.15, if $S \neq 0$, and evanescent, § 2.14, if $S=0$.

The other W.K.B. solution where q^2 is negative is

$$E_y = Aq^{-\frac{1}{2}}\exp(k\int^z |q|\,\mathrm{d}z), \quad \mathscr{H}_x = qE_y. \tag{7.66}$$

It has similar properties to the wave in (7.65) but both fields increase indefinitely as z increases. This wave, therefore, could not be excited by waves incident from the negative z direction, but it could occur if there were sources of waves where z is very large and positive.

In lossy systems where n^2 and q^2 are complex it often happens that q never becomes exactly zero for any real value of z. In many cases of interest, however, there is still a region of the real z axis where $|q|$ is so small that the condition (7.59) is violated. Then the W.K.B. solutions fail and some reflection occurs. If (7.59) is valid for all real values of z, the reflection is extremely small and the wave incident from the negative z direction continues to travel in the direction of positive z until it is absorbed. Its fields are given with good accuracy by the W.K.B. solution, for all real z.

It is often important to know the relation between the amplitudes of the two

W.K.B. solutions where z is negative and where z is positive, with a range of z in between where (7.59) is violated and there is failure of the W.K.B. solutions. Such a relation is called a 'circuit relation'. The main purpose of the papers by Wentzel, Kramers and Brillouin cited in § 7.5 was a study of circuit relations rather than the derivation of the W.K.B. solutions. A knowledge of the circuit relations is essential in the calculation of reflection coefficients, and the subject is therefore studied in some detail in later chapters. See §§ 8.19, 15.13.

7.12. W.K.B. solutions for oblique incidence and vertical polarisation

When the electric field is parallel to the $x-z$ plane, the appropriate differential equations are (7.4), and they are not quite so simple as the set (7.3) which has been discussed in the earlier sections. The field component E_z may be eliminated from (7.4) to give

$$\frac{\mathrm{d}E_x}{\mathrm{d}z} = -\mathrm{i}k\frac{q^2}{n^2}\mathscr{H}_y, \quad \frac{\mathrm{d}\mathscr{H}_y}{\mathrm{d}z} = -\mathrm{i}kn^2E_x \tag{7.67}$$

which corresponds to (7.5) for horizontal polarisation. It is not now possible to obtain a second order differential equation of the simple form (7.6). Elimination of E_x from (7.67) gives

$$\frac{\mathrm{d}^2\mathscr{H}_y}{\mathrm{d}z^2} - \frac{1}{n^2}\frac{\mathrm{d}(n^2)}{\mathrm{d}z}\frac{\mathrm{d}\mathscr{H}_y}{\mathrm{d}z} + k^2q^2\mathscr{H}_y = 0. \tag{7.68}$$

Some important properties of this equation were given by Heading (1970a). The W.K.B. solutions may be found by the method of § 7.6 Let

$$\mathscr{H}_y = \exp\{\mathrm{i}\phi(z)\}, \tag{7.69}$$

compare (7.17), where $\phi(z)$ is the complex phase. Then the differential equation that ϕ must satisfy is

$$\mathrm{i}\frac{\mathrm{d}^2\phi}{\mathrm{d}z^2} - \left(\frac{\mathrm{d}\phi}{\mathrm{d}z}\right)^2 - \frac{\mathrm{i}}{n^2}\frac{\mathrm{d}(n^2)}{\mathrm{d}z}\frac{\mathrm{d}\phi}{\mathrm{d}z} + k^2q^2 = 0, \tag{7.70}$$

compare (7.19). Since the medium is slowly varying, the first and third terms are small and hence to a first approximation

$$\frac{\mathrm{d}\phi}{\mathrm{d}z} = \mp kq, \quad \frac{\mathrm{d}^2\phi}{\mathrm{d}z^2} = \mp k\frac{\mathrm{d}q}{\mathrm{d}z}. \tag{7.71}$$

These are inserted in the first and third terms of (7.70) which gives for the second approximation

$$\frac{\mathrm{d}\phi}{\mathrm{d}z} = \mp\left\{k^2q^2 \mp \mathrm{i}k\left(\frac{\mathrm{d}q}{\mathrm{d}z} - \frac{2q}{n}\frac{\mathrm{d}n}{\mathrm{d}z}\right)\right\}^{\frac{1}{2}} \tag{7.72}$$

where either the upper or the lower signs must be used throughout, for the same reasons as in § 7.6. The last two terms on the right of (7.72) are small. Then expanding by the binomial theorem gives

$$\frac{\mathrm{d}\phi}{\mathrm{d}z} \approx \mp kq + \frac{1}{2}\frac{\mathrm{i}}{q}\frac{\mathrm{d}q}{\mathrm{d}z} - \frac{\mathrm{i}}{n}\frac{\mathrm{d}n}{\mathrm{d}z} \tag{7.73}$$

which may be integrated, and the result inserted in (7.69). This gives

$$\mathscr{H}_y = nq^{-\frac{1}{2}}\exp\left(\mp \mathrm{i}k\int^z q\,\mathrm{d}z\right). \tag{7.74}$$

Now E_x may be found from the second equation (7.67). If derivatives of q and n are neglected

$$E_x = \pm n^{-1}q^{\frac{1}{2}}\exp\left(\mp \mathrm{i}k\int^z q\,\mathrm{d}z\right). \tag{7.75}$$

The expressions (7.74), (7.75) are the W.K.B. solutions for vertical polarisation. They contain a factor $\exp\{\mathrm{i}(\omega t - kSx)\}$ which is omitted as usual; § 7.2.

A useful criterion for the validity of (7.74) rests on the use in (7.73) of only two terms of the binomial expansion of (7.72). For this it is necessary that

$$\frac{1}{k}\left|\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{q}\right) + \frac{2}{qn}\frac{\mathrm{d}n}{\mathrm{d}z}\right| \ll 1, \tag{7.76}$$

compare (7.60). This fails near the level where $q = 0$, no matter how small are the derivatives $\mathrm{d}q/\mathrm{d}z$, $\mathrm{d}n/\mathrm{d}z$. Hence $q = 0$ gives the level of reflection. The condition (7.76) also fails near the level where $n = 0$ which is not a level of reflection if $S \neq 0$. If the electron concentration N increases monotonically with height z, the level where $n = 0$ is above the reflection level where $q = 0$. If these two levels are well separated, the reflection process is unaffected by the failure of (7.76) at the level $n = 0$, and the reflection coefficient is given by (7.152) just as for horizontal polarisation. But if the two levels where $q = 0$ and $n = 0$ are close together, the reflection coefficient may be affected, and a more detailed study of the differential equation is needed to find its true value. This is given in §§ 15.5–15.7.

7.13. Differential equations for anisotropic ionosphere

When the earth's magnetic field is allowed for, the differential equations (2.45) cannot be separated into two independent sets such as (7.3) and (7.4). They must be treated together as one set. As in the earlier sections it is assumed that the x and y dependence of all field components is through the factor $\exp(-\mathrm{i}kSx)$, so that (6.48) is used. Then Maxwell's equations (2.45) give

$$\frac{\partial E_x}{\partial z} = -\mathrm{i}k(\mathscr{H}_y + SE_z),\quad \frac{\partial E_y}{\partial z} = \mathrm{i}k\mathscr{H}_x,\quad SE_y = \mathscr{H}_z \tag{7.77}$$

and

$$\frac{\partial \mathscr{H}_x}{\partial z} = \mathrm{i}k(\varepsilon_0^{-1}D_y - S\mathscr{H}_z),\quad \frac{\partial \mathscr{H}_y}{\partial z} = -\mathrm{i}k\varepsilon_0^{-1}D_x,\quad S\mathscr{H}_y = \varepsilon_0^{-1}D_z. \tag{7.78}$$

Now

$$\boldsymbol{D} = \varepsilon_0 \boldsymbol{\varepsilon} \boldsymbol{E} = \varepsilon_0 \boldsymbol{E} + \boldsymbol{P} = \varepsilon_0(\boldsymbol{E} + \boldsymbol{M}\cdot\boldsymbol{E}). \tag{7.79}$$

For an electron plasma in which the collision frequency is not velocity dependent, the susceptibility matrix $\boldsymbol{M}$ is as given in § 3.9. It may be used in (7.78) to express $\boldsymbol{D}$ in terms of $\boldsymbol{E}$. In (7.77) and (7.78) the only field components that appear in derivatives are the four horizontal components E_x, E_y, $\mathscr{H}_x$, $\mathscr{H}_y$. We may therefore use the last equations in (7.77) and (7.78) to eliminate $\mathscr{H}_z$ and E_z. This leads to four homogeneous first order differential equations for the horizontal field components. These can very conveniently be written in matrix form. Let $\mathbf{e}$ be the column matrix with four elements E_x, $-E_y$, $\mathscr{H}_x$, $\mathscr{H}_y$. The reason for using $-E_y$ is given in § 7.14 (3), (8). Then the equations (7.77), (7.78) become

$$\mathrm{d}\mathbf{e}/\mathrm{d}z = -\mathrm{i}k\mathbf{T}\mathbf{e} \tag{7.80}$$

where $\mathbf{T}$ is the 4×4 matrix

$$\mathbf{T} = \begin{pmatrix} -\dfrac{SM_{zx}}{1+M_{zz}} & \dfrac{SM_{zy}}{1+M_{zz}} & 0 & \dfrac{C^2+M_{zz}}{1+M_{zz}} \\ 0 & 0 & 1 & 0 \\ \dfrac{M_{yz}M_{zx}}{1+M_{zz}} - M_{yx} & C^2 + M_{yy} - \dfrac{M_{yz}M_{zy}}{1+M_{zz}} & 0 & \dfrac{SM_{yz}}{1+M_{zz}} \\ 1 + M_{xx} - \dfrac{M_{xz}M_{zx}}{1+M_{zz}} & \dfrac{M_{xz}M_{zy}}{1+M_{zz}} - M_{xy} & 0 & \dfrac{-SM_{xz}}{1+M_{zz}} \end{pmatrix}. \tag{7.81}$$

The algebraic steps of the derivation of (7.80), (7.81) were given in more detail by Budden (1961a).

If the electron collision frequency is assumed to be velocity dependent, as for example in the Sen–Wyller formulae § 3.12, or if the effect of heavy ions is allowed for, the form (7.81) can still be used but the elements of $\boldsymbol{M}$ are not as given in § 3.9. In these cases they can be found from (3.55) which gives $\boldsymbol{\varepsilon} = \boldsymbol{M} + 1$. The elements of $\boldsymbol{\varepsilon}$ are written ε_{ij}, $i, j = x, y, z$. Then $\mathbf{T}$ is given (Walker and Lindsay, 1975) by:

$$\varepsilon_{zz}\mathbf{T} = \begin{pmatrix} -S\varepsilon_{zx} & S\varepsilon_{zy} & 0 & \varepsilon_{zz} - S^2 \\ 0 & 0 & \varepsilon_{zz} & 0 \\ \varepsilon_{yz}\varepsilon_{zx} - \varepsilon_{yz}\varepsilon_{zz} & \varepsilon_{yy}\varepsilon_{zz} - \varepsilon_{yz}\varepsilon_{zy} - S^2\varepsilon_{zz} & 0 & S\varepsilon_{yz} \\ \varepsilon_{xx}\varepsilon_{zz} - \varepsilon_{xz}\varepsilon_{zx} & \varepsilon_{xz}\varepsilon_{zy} - \varepsilon_{xy}\varepsilon_{zz} & 0 & -S\varepsilon_{xz} \end{pmatrix}. \tag{7.82}$$

Equations (7.80), (7.81) were first given by Clemmow and Heading (1954). An almost identical set for the application to optical waves was given by Berreman (1972).

The equations (7.80) are a set of four simultaneous linear first order differential equations. It is possible to eliminate three of the four dependent variables $\mathbf{e}$, and the result would be one linear fourth order differential equation for the remaining variable. Thus we expect (7.80) to have four linearly independent solutions. If it is studied in a range of z where the plasma is homogeneous, the elements of $\mathbf{T}$ are then all independent of z. In such a region, as shown in ch. 6, there are four solutions representing obliquely travelling progressive waves, associated with the four roots of the Booker quartic. For this case, therefore, we seek a progressive wave solution of (7.80) in which all field components depend on z only through a factor $\exp(-ikqz)$. Then it gives

$$(\mathbf{T} - q)\mathbf{e} = 0. \tag{7.83}$$

This is a set of four simultaneous algebraic equations for finding the elements of $\mathbf{e}$. It can only have a non-trivial solution if the determinant of the coefficients is zero, that is

$$\det(\mathbf{T} - q) = 0. \tag{7.84}$$

This is a fourth degree equation for finding q, and is simply the Booker quartic. It can be verified that it is the same as the various forms of the quartic given in ch. 6. It shows that the four qs are the eigen values of the matrix $\mathbf{T}$, and when they are inserted in (7.83), the resulting solutions for $\mathbf{e}$ are their associated eigen columns.

7.14. Matrix theory

We now derive further properties of the matrix form (7.80) of the differential equations. A column matrix with four elements is denoted by a bold lower case sans serif letter, for example $\mathbf{e}$. A 4×4 square matrix is denoted by a bold face capital sans serif letter, for example $\mathbf{T}$. The transpose of a matrix is indicated by a superscript T. Thus $\mathbf{e}^{\mathrm{T}}$ is a row matrix with four elements.

In this section some important results are derived and it is useful, for future reference, to set them out in a numbered sequence.

(1) *Eigen columns and eigen rows of* $\mathbf{T}$

Let the four values of q be q_i, $i = 1, 2, 3, 4$, and let the associated eigen columns of $\mathbf{T}$ be $\mathbf{s}_i$. The four $\mathbf{s}_i$ are used as the four columns of a 4×4 matrix $\mathbf{S}$. For the present it is assumed that the four q_i are distinct, so that the four $\mathbf{s}_i$ are linearly independent and the matrix $\mathbf{S}$ is non-singular. The case where two qs approach equality is dealt with in § 16.3. From (7.83)

$$\mathbf{T}\mathbf{s}_i = q_i\mathbf{s}_i, \quad \mathbf{TS} = \mathbf{SQ} \tag{7.85}$$

where $\mathbf{Q}$ is the 4×4 diagonal matrix with elements q_i. Multiplication by $\mathbf{S}^{-1}$ on both sides gives

$$\mathbf{S}^{-1}\mathbf{T} = \mathbf{Q}\mathbf{S}^{-1}, \tag{7.86}$$

which shows that the rows of $\mathbf{S}^{-1}$ are eigen rows of $\mathbf{T}$ with the eigen values q_i.

If any $\mathbf{s}_i$ is multiplied by a constant, it remains an eigen column of $\mathbf{T}$. To complete the definitions of $\mathbf{s}_i$ and $\mathbf{S}$ it is necessary to impose some condition on the field amplitudes. In other words the $\mathbf{s}_i$ must be normalised. One possible rule for this, used in most of this book, is given in (7) below, and the reason for choosing it is explained in (8). Other normalisation methods are possible and this is discussed in §§ 16.2, 16.3.

(2) *Integrating factor*

In the study of differential equations an integrating factor is often used. The technique is described in books on differential equations, for example Ince (1927, § 2.2), Birkhoff and Rota (1962, chs. I. 5, II. 5). It is applied here to the matrix equation (7.80).

We seek, as an integrating factor, a row matrix $\mathbf{g}^{\mathrm{T}}$ that multiplies (7.80) on the left so that the result is a perfect differential, thus

$$\mathbf{g}^{\mathrm{T}}\mathrm{d}\mathbf{e}/\mathrm{d}z + \mathrm{i}k\mathbf{g}^{\mathrm{T}}\mathbf{T}\mathbf{e} = \mathrm{d}W/\mathrm{d}z = 0 \tag{7.87}$$

where W is a scalar. This is satisfied if

$$\mathrm{d}\mathbf{g}^{\mathrm{T}}/\mathrm{d}z = \mathrm{i}k\mathbf{g}^{\mathrm{T}}\mathbf{T}, \tag{7.88}$$

and

$$W = \mathbf{g}^{\mathrm{T}}\mathbf{e}. \tag{7.89}$$

Then W is called the bilinear concomitant.

The transpose of (7.88) is

$$\mathrm{d}\mathbf{g}/\mathrm{d}z = \mathrm{i}k\mathbf{T}^{\mathrm{T}}\mathbf{g}. \tag{7.90}$$

(3) *Symmetry properties of* $\mathbf{T}$

The matrix $\mathbf{T}$ uses the elements of $\boldsymbol{M}$ (3.35). If the vector $\boldsymbol{Y}$ is reversed in direction, that is if l_x, l_y, l_z are all reversed in sign, $\boldsymbol{M}$ is changed to a new matrix $\bar{\boldsymbol{M}}$ which is simply $\boldsymbol{M}^{\mathrm{T}}$. When this change is made in (7.81), let the resulting matrix be denoted by $\bar{\mathbf{T}}$. Then inspection of (7.81) shows that $\bar{\mathbf{T}}$ is obtained by transposing $\mathbf{T}$ about its trailing diagonal, thus

$$\bar{\mathrm{T}}_{ij} = \mathrm{T}_{5-j,5-i} \tag{7.91}$$

(Budden and Clemmow, 1957). This property still holds when the more general form (7.82) of $\mathbf{T}$ is used. To write it in matrix notation introduce the 4×4 matrix used by Lacoume (1967) and Bennett (1976):

$$\mathbf{B} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \tag{7.92}$$

with the property

$$\mathbf{B}^{\mathrm{T}} = \mathbf{B}^{-1} = \mathbf{B}. \tag{7.93}$$

Then

$$\bar{\mathbf{T}} = \mathbf{B}\mathbf{T}^{\mathrm{T}}\mathbf{B}, \quad \mathbf{T}^{\mathrm{T}} = \mathbf{B}\bar{\mathbf{T}}\mathbf{B}. \tag{7.94}$$

This property of $\mathbf{T}$ would not have held if E_y had been given a plus sign in the definition of $\mathbf{e}$, § 7.13. This is one reason why E_y has a minus sign.

(4) *Adjoint matrix. Adjoint differential equation*

Now $\mathbf{T}^{\mathrm{T}}$ from (7.94) is substituted in (7.90) and the result is multiplied on the left by $\mathbf{B}$ to give

$$\mathrm{d}\bar{\mathbf{e}}/\mathrm{d}z = \mathrm{i}k\bar{\mathbf{T}}\bar{\mathbf{e}} \tag{7.95}$$

where

$$\bar{\mathbf{e}} = \mathbf{B}\mathbf{g}. \tag{7.96}$$

Here $\bar{\mathbf{e}}$, $\bar{\mathbf{T}}$ are called the adjoints of $\mathbf{e}$, $\mathbf{T}$, and (7.95) or its equivalent form (7.88) is the differential equation adjoint to (7.80). These adjoints, and the bilinear concomitant W, were introduced into the theory of radio propagation by Suchy and Altman (1975a) in a most important paper. For further discussion see § 14.12.

Let $\mathbf{m}$ be a square matrix with elements m_{ij} and let M_{ij} be the cofactor of m_{ij}. Then the matrix $\mathbf{M}$ whose elements are M_{ji} is commonly called the 'adjoint' of $\mathbf{m}$, and less commonly the 'adjugate'. Note that it is the transpose of the matrix of the cofactors. Here, as in Jeffreys and Jeffreys (1972, p. 117), the term 'adjugate' will be used. The adjoint form $\bar{\mathbf{T}}$ of $\mathbf{T}$ as defined by (7.94) is not the same as the adjugate of $\mathbf{T}$.

The elements of $\bar{\mathbf{e}}$ will be written $\bar{E}_x$, $-\bar{E}_y$, $\bar{\mathscr{H}}_x$, $\bar{\mathscr{H}}_y$. Then (7.95) shows that $\bar{\mathbf{e}}$ gives these field components in a fictitious 'adjoint' medium in which the direction of $\boldsymbol{Y}$, and therefore of the earth's magnetic field $\boldsymbol{B}$, is reversed, and the sign of $k = \omega/c$ is reversed. The sign of ω is reversed in the factor $\exp(\mathrm{i}\omega t)$, see (7.100) below, but not in $Z = \nu/\omega$ and not where it appears in ε_1, ε_2, ε_3. The adjoint fields $\bar{\mathbf{e}}$ are introduced only to assist the theory and need not satisfy any physical conditions. For example in a range of z where the medium is homogeneous with losses, (7.95) has a solution proportional to $\exp(+\mathrm{i}qkz)$ where q is any root of the Booker quartic. If it applies to an upgoing wave, $\mathrm{Im}(q)$ is negative, so the adjoint fields *increase* as z increases.

(5) *The bilinear concomitant*

From (7.89), (7.96):

$$W = \bar{\mathbf{e}}^{\mathrm{T}}\mathbf{B}\mathbf{e} = \bar{\mathscr{H}}_y E_x - \bar{\mathscr{H}}_x E_y - \bar{E}_y \mathscr{H}_x + \bar{E}_x \mathscr{H}_y. \tag{7.97}$$

This expression is very similar to (2.65) for $4Z_0\Pi_z$ where Π_z is the z component of the time averaged Poynting vector. The only difference is that, instead of using complex conjugates E_x^* etc., we use the adjoints $\bar{E}_x$ etc. The adjoints resemble the complex conjugates in that neither of them satisfies Maxwell's equations nor any physical conditions. But they are needed in the definitions of W and Π_z respectively.

It is shown later § 14.12, that W is the z component of a vector $\boldsymbol{W}$, the bilinear concomitant vector, which is a generalisation of the time averaged Poynting vector Π_{av}. From its definition (2.64), Π_{av} is necessarily real. But $\boldsymbol{W}$ is in general complex. This applies particularly in a medium with losses.

(6) *Eigen columns and eigen values of* $\bar{\mathbf{T}}$

Let $\mathbf{r}$ be an eigen column of $\bar{\mathbf{T}}$ with eigen value q, so that

$$q\mathbf{r} = \bar{\mathbf{T}}\mathbf{r}. \tag{7.98}$$

Transpose this, multiply on the right by $\mathbf{B}$, and use (7.94). Then

$$q\mathbf{r}^{\mathrm{T}}\mathbf{B} = \mathbf{r}^{\mathrm{T}}\bar{\mathbf{T}}^{\mathrm{T}}\mathbf{B} = \mathbf{r}^{\mathrm{T}}\mathbf{B}\mathbf{B}\bar{\mathbf{T}}^{\mathrm{T}}\mathbf{B} = \mathbf{r}^{\mathrm{T}}\mathbf{B}\mathbf{T} \tag{7.99}$$

which shows, first, that the eigen values q of $\bar{\mathbf{T}}$ are the same as those of $\mathbf{T}$. This also follows by transposition of the determinant (7.84) about its trailing diagonal. Second, (7.99) shows that $\mathbf{r}^{\mathrm{T}}\mathbf{B}$ is an eigen row of $\mathbf{T}$. Then (7.86) shows that we are free to choose $\mathbf{r}$ so that $\mathbf{r}^{\mathrm{T}}\mathbf{B}$ is a row of $\mathbf{S}^{-1}$. The eigen columns of $\bar{\mathbf{T}}$ are the rows of $\mathbf{S}^{-1}$ with the order of the elements reversed.

The field components given by the elements of $\mathbf{S}$ all depend on x, y and t through a factor $\exp\{\mathrm{i}(\omega t - kSx)\}$ which is usually omitted; § 6.2. This suggests that the adjoint fields given by the elements of $\mathbf{S}^{-1}$ should contain a factor $\exp\{-\mathrm{i}(\omega t - kSx)\}$. Then in the adjoint of a homogeneous medium, (7.95) is satisfied and the four progressive wave solutions are

$$(\mathbf{S}^{-1})_j \exp\{\mathrm{i}(-\omega t + kSx + kq_j z)\} \tag{7.100}$$

where the first factor is the j^{th} row of $\mathbf{S}^{-1}$. In all terms of the exponent the signs of k and ω have therefore been reversed.

For a progressive wave in a homogeneous medium choose the j^{th} column of $\mathbf{S}$, and for the adjoint solution $\mathbf{g}^{\mathrm{T}}$ of (7.88) choose (7.100). Then the bilinear concomitant W is $+1$.

(7) *Normalisation of* $\mathbf{S}$

In this section it is assumed that the exponential factors $\exp\{\pm\mathrm{i}(\omega t - kSx)\}$ are omitted. Specific expressions for $\mathbf{S}$ and $\mathbf{S}^{-1}$ are given by (7.118), (7.121) in §7.16, and it is there shown that for any column of $\mathbf{S}$ and the same numbered row of $\mathbf{S}^{-1}$, here taken as the j^{th} column and row,

$$(\mathbf{S}^{-1})_{j3}/\mathrm{S}_{2j} = (\mathbf{S}^{-1})_{j2}/\mathrm{S}_{3j} \quad \text{(not summed)}. \tag{7.101}$$

Hence we now choose for the normalisation of $\mathbf{S}$:

$$\mathrm{S}_{2j} = (\mathbf{S}^{-1})_{j3}, \quad \mathrm{S}_{3j} = (\mathbf{S}^{-1})_{j2}. \tag{7.102}$$

The reason for this choice is given in (8) below.

(8) *Bilinear concomitant, and Poynting vector*

Consider now the special case where the ionosphere is loss-free, so that the qs are either real or in complex conjugate pairs; see § 6.4. Then inspection of (7.81) with (3.35) or (3.55) shows that reversal of the direction of Y is just the same as changing the sign of i. The adjoint $\bar{\mathbf{T}}$ is the same as the complex conjugate $\mathbf{T}^*$.

First let q be real and let the elements of its associated column in $\mathbf{S}$ be E_x, $-E_y$, $\mathscr{H}_x$, $\mathscr{H}_y$. This is an eigen column of $\mathbf{T}$. Suppose that it applies for an upgoing wave. The elements of the corresponding row of $\mathbf{S}^{-1}$ are those of the eigen column of $\bar{\mathbf{T}}$, that is of $\mathbf{T}^*$, in reversed order, and any multiple of this may be chosen as the adjoint solution

$$\mathbf{g}^{\mathrm{T}} = \mathscr{H}_y^*, \mathscr{H}_x^*, -E_y^*, E_x^* \tag{7.103}$$

of (7.88). Then the bilinear concomitant is

$$W = E_x\mathscr{H}_y^* - E_y\mathscr{H}_x^* - E_y^*\mathscr{H}_x + E_x^*\mathscr{H}_y = 4Z_0\Pi_z, \tag{7.104}$$

where (2.64) has been used. For the upgoing wave on the real z axis this is real and positive. For a downgoing wave we may choose, for $\mathbf{g}^{\mathrm{T}}$, the set (7.103) with the signs on the right reversed, and this would make W real and negative. The normalisation in (7) above was deliberately chosen so that, for a progressive wave with a real q in a loss-free ionosphere, the adjoint set (7.103) is the complex conjugate of the horizontal field components.

Next let q and q^* be two conjugate complex solutions of the quartic and let them be associated with the first and second columns of $\mathbf{S}$. Let the elements of these two columns be

$$\mathbf{s}_1 = E_x(q),\ -E_y(q),\ \mathscr{H}_x(q),\ \mathscr{H}_y(q): \quad \mathbf{s}_2 = E_x(q^*),\ -E_y(q^*),\ \mathscr{H}_x(q^*),\ \mathscr{H}_y(q^*) \tag{7.105}$$

respectively. These are identical except for the change from q to q^*. Let the two corresponding rows of $\mathbf{S}^{-1}$ be $\mathbf{r}_1^{\mathrm{T}}$, $\mathbf{r}_2^{\mathrm{T}}$. They are not the complex conjugates of $\mathbf{s}_1$, $\mathbf{s}_2$ respectively, with the elements in reversed order, because in forming them we do not replace q by q^* or vice versa. But clearly

$$\mathbf{r}_2^{\mathrm{T}} = (\mathbf{s}_1^{\mathrm{T}}\mathbf{B})^*, \quad \mathbf{r}_1^{\mathrm{T}} = (\mathbf{s}_2^{\mathrm{T}}\mathbf{B})^*. \tag{7.106}$$

The product $\mathbf{r}_2^{\mathrm{T}}\mathbf{s}_1$ is $4Z_0\Pi_z$ for the wave associated with q, and $\mathbf{r}_1^{\mathrm{T}}\mathbf{s}_2$ is $4Z_0\Pi_z$ for that associated with q^*. But both these products are zero. This confirms the statement made in § 6.4, that for waves associated with complex conjugate qs in a loss-free medium, the z components of the time averaged Poynting vectors are zero.

The association of the bilinear concomitant product (7.104) with the Poynting vector is another consequence of the use of $-E_y$ in the definition of $\mathbf{e}$ in § 7.13, and in (7.103).

7.15. W.K.B. solutions for anisotropic ionosphere

Two different methods have been given for deriving the W.K.B. solutions when the earth's magnetic field is neglected. In the first, § 7.6, the differential equation (7.6) for

E_y was formulated and the change of variable $E_y = A\exp\{i\phi(z)\}$ (7.17) was then made, leading to a non-linear differential equation (7.19) for ϕ. To use the same method in the more general case it would be necessary to formulate the differential equation for one field variable. We have already seen, near the end of § 7.13, that this is of fourth order, and it is too complicated to be worth deriving specifically, although in principle the method could be used. It is better, however, to extend the second method, § 7.8, in which the fields were expressed as the sum of partial fields for the upgoing and downgoing waves. In this way the differential equations were expressed in coupled form (7.46), (7.47). Neglect of the coupling terms on the right-hand sides gave two separate differential equations whose solutions were the W.K.B. solutions for the upgoing and downgoing waves. This method will now be extended to the general case. The fields are to be expressed as the sum of the partial fields for the four characteristic waves, and the differential equations are expressed in the general coupled form (7.109) below. Neglect of the coupling terms gives four separate equations whose solutions are the four W.K.B. solutions.

Suppose that the ionospheric plasma varies slowly with the height z so that $\mathbf{T}$ and $\mathbf{S}$ are functions of z. At any height z a solution $\mathbf{e}(z)$ of (7.80) can be expressed as a linear combination of the four characteristic waves, that is the four columns of $\mathbf{S}$, with factors f_j as in (6.53), § 6.10. In a homogeneous ionosphere the four f_j would be proportional to $\exp(-ikq_jz)$. We now enquire how they vary with z in a slowly varying ionosphere. The following method is the exact analogue of the 'coupling' method used in § 7.8 for an isotropic ionosphere. It was first given by Clemmow and Heading (1954). In the following theory the summation convention, § 2.11, for repeated suffixes is not used.

A prime $'$ is now used to denote $k^{-1}\,\mathrm{d}/\mathrm{d}z$. Then insertion of (6.53) into (7.80) gives

$$(\mathbf{Sf})' = -\mathrm{i}\mathbf{TSf} \tag{7.107}$$

so that

$$\mathbf{f}' + \mathrm{i}\mathbf{S}^{-1}\mathbf{TSf} = \mathbf{\Gamma f}, \quad \text{where} \quad \mathbf{\Gamma} = -\mathbf{S}^{-1}\mathbf{S}' \tag{7.108}$$

whence from (7.85)

$$\mathbf{f}' + \mathrm{i}\mathbf{Qf} = \mathbf{\Gamma f} \tag{7.109}$$

where $\mathbf{Q}$ is the 4×4 diagonal matrix whose elements are the four qs. The terms on the right correspond to the terms on the right of (7.46), (7.47). They are called 'coupling terms', and $\mathbf{\Gamma}$ is called the 'coupling matrix'; see ch. 16. In a homogeneous medium they are zero because $\mathbf{S}'$ and thence $\mathbf{\Gamma}$ are zero. In a sufficiently slowly varying medium they are small and as a first approximation they are neglected. Then (7.109) is four separate first order differential equations

$$\mathsf{f}_j' + \mathrm{i}q_j\mathsf{f}_j = 0 \tag{7.110}$$

with the solutions

$$\mathsf{f}_j = \mathsf{K}_j \exp\left(-ik\int_0^z q_j \mathrm{d}z\right) \quad j = 1, 2, 3, 4 \tag{7.111}$$

where K_j is a constant. The contribution of one of these terms to the solution $\mathbf{e}$ is, from (6.53),

$$\mathbf{e} = \mathsf{K}_j \mathbf{s}_j \exp\left(-ik\int_0^z q_j \mathrm{d}z\right) \tag{7.112}$$

where $\mathbf{s}_j$ is the j^{th} column of $\mathbf{S}$. These are here taken to be the four W.K.B. solutions for the anisotropic ionosphere. The full expression is given later at (7.136).

In an exactly similar way, let the adjoint solution $\bar{\mathbf{e}}^{\mathrm{T}}$ be given by

$$\bar{\mathbf{e}}^{\mathrm{T}} = \bar{\mathsf{f}}^{\mathrm{T}} \mathbf{S}^{-1}. \tag{7.113}$$

It must satisfy the adjoint differential equation (7.88), whence it can be shown that, if coupling terms are neglected,

$$\bar{\mathsf{f}}_j' - \mathrm{i} q_j \bar{\mathsf{f}}_j = 0 \tag{7.114}$$

with the solution

$$\bar{\mathsf{f}}_j = \bar{\mathsf{K}}_j \exp\left(ik\int_0^z q_j \mathrm{d}z\right) \tag{7.115}$$

where $\bar{\mathsf{K}}_j$ is another constant.

Thus the adjoint of (7.112) is

$$\bar{\mathbf{e}} = \bar{\mathsf{K}}_j \mathbf{r}_j \exp\left(ik\int_0^z q_j \mathrm{d}z\right) \tag{7.116}$$

where $\mathbf{r}_j$ is the j^{th} row of $\mathbf{S}^{-1}$. The W.K.B. solution (7.112) and its adjoint (7.116) give a bilinear concomitant $W = \mathsf{K}_j \bar{\mathsf{K}}_j$ that is independent of z. This means that for the W.K.B. solution in the special case of a loss-free ionosphere and a progressive wave, the z component Π_z of the time averaged Poynting vector is independent of z (compare § 7.4). It is easily verified that for an isotropic ionosphere with losses, the bilinear concomitant W is independent of z for the W.K.B. solutions (7.12) and (7.13) (horizontal polarisation) and for (7.74), (7.75) (vertical polarisation).

Equation (7.110) for f_j should, strictly speaking, have a term on the right $\Gamma_{jj}\mathsf{f}_j$ (not summed) from the diagonal element of $\mathbf{\Gamma}$ in (7.109). Then (7.111) would become

$$\mathsf{f}_j = \mathsf{K}_j \exp(-ik\int^z q_j \mathrm{d}z + k\int^z \Gamma_{jj} \mathrm{d}z) \quad \text{(not summed).} \tag{7.117}$$

The exponential in (7.111) has, in § 7.3, been called the phase memory term, and the new factor $\exp(k\int^z \Gamma_{jj} \mathrm{d}z)$ is called 'additional memory' (Budden and Smith, 1976). In some important special cases the elements Γ_{jj} are zero; see § 16.2. When they are not zero they ought to be included in the W.K.B. solutions (7.111). Some consequences of including them are described in § 16.9.

Suppose that the matrix $\mathbf{S}$ used to give $\mathbf{s}_j$ in (7.112) has been normalised in a different way so that the column $\mathbf{s}_j$ is replaced by $b(z)\mathbf{s}_j$. Then it can be shown that the element Γ_{jj} now has an extra added term $-b'(z)/b(z)$ and the additional memory term $\exp(k\int^z\Gamma_{jj}\mathrm{d}z)$ therefore has a factor $1/b(z)$ that cancels the $b(z)$ multiplying $\mathbf{s}_j$. Thus, if the additional memory term is included, the W.K.B. solution (7.112) is unaffected by the choice of normalisation of $\mathbf{S}$.

7.16. The matrices $\mathbf{S}$ and $\mathbf{S}^{-1}$

Any column of $\mathbf{S}$ is a solution of the set of four homogeneous equations (7.85); see § 6.10 and end of § 7.13. It is proportional to the cofactors of any row of the matrix $\mathbf{T}-q_j\mathbf{1}$, where q_j applies to the j^{th} column of $\mathbf{S}$. Because of the zeros in the second row and third column of $\mathbf{T}$, (7.81), it is simplest to choose the third row. Then

$$\mathsf{S}_{1j}:\mathsf{S}_{2j}:\mathsf{S}_{3j}:\mathsf{S}_{4j}\propto a_3q_j+a_4:A_j:q_jA_j:a_5q_j+a_6 \tag{7.118}$$

where

$$\left.\begin{aligned}A_j&=q_j^2+a_1q_j+a_2,\\ a_1=-(\mathsf{T}_{11}+\mathsf{T}_{44}),\quad a_2&=\mathsf{T}_{11}\mathsf{T}_{44}-\mathsf{T}_{14}\mathsf{T}_{41},\quad a_3=\mathsf{T}_{12},\\ a_4=\mathsf{T}_{14}\mathsf{T}_{42}-\mathsf{T}_{12}\mathsf{T}_{44},\quad a_5&=\mathsf{T}_{42},\quad a_6=\mathsf{T}_{12}\mathsf{T}_{41}-\mathsf{T}_{11}\mathsf{T}_{42}.\end{aligned}\right\} \tag{7.119}$$

Similarly, from § 7.14(6), the j^{th} row of $\mathbf{S}^{-1}$ is proportional to the j^{th} row of the adjoint matrix $\bar{\mathbf{T}}-q_j\mathbf{1}$, with the order reversed. The adjoints $\bar{a}_k$ of the a_k are found from (7.119) by using $\bar{\mathbf{T}}$, that is $\mathbf{T}$ transposed about its trailing diagonal. Thus

$$\begin{aligned}&\bar{a}_1=a_1,\quad \bar{a}_2=a_2,\quad \bar{A}_j=A_j\\ &\bar{a}_3=\mathsf{T}_{34},\quad \bar{a}_4=\mathsf{T}_{14}\mathsf{T}_{31}-\mathsf{T}_{34}\mathsf{T}_{11},\bar{a}_5=\mathsf{T}_{31},\bar{a}_6=\mathsf{T}_{34}\mathsf{T}_{41}-\mathsf{T}_{44}\mathsf{T}_{31}.\end{aligned} \tag{7.120}$$

Hence

$$(\mathbf{S}^{-1})_{j1}:(\mathbf{S}^{-1})_{j2}:(\mathbf{S}^{-1})_{j3}:(\mathbf{S}^{-1})_{j4}\propto \bar{a}_5q_j+a_6:q_jA_j:A_j:\bar{a}_3q_j+\bar{a}_4. \tag{7.121}$$

For the electron plasma, from (7.81), the expressions for a_1-a_6 are as follows. Let

$$D=X/\{U(U^2-Y^2)-X(U^2-l_z^2Y^2)\}. \tag{7.122}$$

Then

$$\left.\begin{aligned}a_1&=2SY^2l_xl_zD,\\ a_2&=-C^2+\{U(U-X)-Y^2(l_x^2C^2+l_z^2S^2)\}D,\\ a_3&=S(l_yl_zY^2-\mathrm{i}l_xUY)D,\\ a_4&=-\{C^2l_xl_yY^2+\mathrm{i}l_zY(UC^2-X)\}D,\\ a_5&=-\{l_xl_yY^2+\mathrm{i}l_zY(U-X)\}D,\\ a_6&=S\{l_yl_zY^2-\mathrm{i}l_xY(U-X)\}D.\end{aligned}\right| \tag{7.123}$$

For the more general plasma, from (7.82) the expressions are (G and J are given by (3.51)):

$$\left.\begin{aligned}
a_1 &= -2Sl_xl_zG/\varepsilon_{zz},\\
a_2 &= [S^2\{\varepsilon_3 + (l_y^2 + l_z^2)G\} - \{\tfrac{1}{2}\varepsilon_3(\varepsilon_1+\varepsilon_2) - l_yJ\}]/\varepsilon_{zz},\\
a_3 &= -S\{l_yl_zG + \tfrac{1}{2}\mathrm{i}l_x(\varepsilon_1-\varepsilon_2)\}/\varepsilon_{zz},\\
a_4 &= [S^2\{-l_xl_yG + \tfrac{1}{2}\mathrm{i}l_z(\varepsilon_1-\varepsilon_2)\} - \{l_xl_yJ + \tfrac{1}{2}\mathrm{i}l_z\varepsilon_3(\varepsilon_1-\varepsilon_2)\}]/\varepsilon_{zz},\\
a_5 &= -\{l_yl_zJ + \tfrac{1}{2}\mathrm{i}l_z\varepsilon_3(\varepsilon_1-\varepsilon_2)\}/\varepsilon_{zz}\\
a_6 &= S\{l_yl_zJ - \tfrac{1}{2}\mathrm{i}l_z\varepsilon_3(\varepsilon_1-\varepsilon_2)\}/\varepsilon_{zz}
\end{aligned}\right\} \tag{7.124}$$

where

$$\varepsilon_{zz} = \tfrac{1}{2}(\varepsilon_1+\varepsilon_2) - l_z^2G. \tag{7.125}$$

The sets (7.118), (7.121) must be normalised by dividing them by suitable factors. Here we give the result of applying the rule in §7.14(7). In (7.118), (7.121) the condition (7.102) is already satisfied so that, for both, the required normalising factor N_j is the same. It is found from the condition that the diagonal elements of $\mathbf{S}^{-1}\mathbf{S}$ are unity. Hence

$$N_j^2 = (a_3q_j + a_4)(\bar{a}_5q_j + \bar{a}_6) + 2q_jA_j^2 + (a_5q_j + a_6)(\bar{a}_3q_j + \bar{a}_4). \tag{7.126}$$

Now let

$$F_1 = (q_1 - q_2)(q_1 - q_3)(q_1 - q_4) \tag{7.127}$$

and let F_2, F_3, F_4 be obtained from this by cyclic permutation of the suffixes. The quartic found by multiplying out the determinant (7.84) is

$$(q^2 - \mathrm{T}_{32})(q^2 + a_1q + a_2) - (a_3\bar{a}_5 + \bar{a}_3a_5)q - (a_4\bar{a}_5 + a_6\bar{a}_3) = 0 \tag{7.128}$$

from which F_j is found by differentiating with respect to q and then setting $q = q_j$. This gives

$$F_j = 2q_jA_j + (q_j^2 - \mathrm{T}_{32})(2q_j + a_1) - (a_3\bar{a}_5 + \bar{a}_3a_5) \tag{7.129}$$

whence

$$\begin{aligned}
A_jF_j &= 2q_jA_j^2 + (2q_j + a_1)\{(a_3\bar{a}_5 + \bar{a}_3a_5)q_j + a_4\bar{a}_5 + a_6\bar{a}_3\} - (a_3\bar{a}_5 + \bar{a}_3a_5)A_j\\
&= 2q_jA_j^2 + q_j^2(a_3\bar{a}_5 + \bar{a}_3a_5) + 2q_j(a_4\bar{a}_5 + a_6\bar{a}_3)\\
&\quad + a_1(a_4\bar{a}_5 + a_6\bar{a}_3) - a_2(a_3\bar{a}_5 + \bar{a}_3a_5).
\end{aligned} \tag{7.130}$$

Now from (7.119), (7.120) it can be shown that

$$a_4\bar{a}_5 + a_6\bar{a}_3 = a_3\bar{a}_6 + a_5\bar{a}_4 \tag{7.131}$$

$$a_1(a_4\bar{a}_5 + a_6\bar{a}_3) - a_2(a_3\bar{a}_5 + \bar{a}_3a_5) = a_4\bar{a}_6 + \bar{a}_4a_6 \tag{7.132}$$

whence from (7.130), (7.126)

$$A_jF_j = N_j^2. \tag{7.133}$$

Then we finally get for a column of $\mathbf{S}$ from (7.118)

$$(\mathsf{S}_{1j}, \mathsf{S}_{2j}, \mathsf{S}_{3j}, \mathsf{S}_{4j}) = (A_jF_j)^{-\frac{1}{2}}(a_3q_j + a_4, A_j, q_jA_j, a_5q_j + a_6) \tag{7.134}$$

and for a row of $\mathbf{S}^{-1}$ from (7.121)

$$\{(\mathbf{S}^{-1})_{j1}, (\mathbf{S}^{-1})_{j2}, (\mathbf{S}^{-1})_{j3}, (\mathbf{S}^{-1})_{j4}\} = (A_jF_j)^{-\frac{1}{2}}(\bar{a}_5q_j + \bar{a}_6, q_jA_j, A_j, \bar{a}_3q_j + \bar{a}_4). \tag{7.135}$$

With (7.134) it is now possible to write specific expressions for the horizontal field components of the W.K.B. solution (7.112), thus

$$(E_x, -E_y, \mathscr{H}_x, \mathscr{H}_y)_j = (A_jF_j)^{-\frac{1}{2}}(a_3q_j + a_4, A_j, q_jA_j, a_5q_j + a_6)\exp\left(-\mathrm{i}k\int^z q_j\mathrm{d}z\right) \tag{7.136}$$

where the omission of the lower limit of the integral is equivalent to the inclusion of an arbitrary constant multiplier, such as K_j in (7.112).

The factor $F_j^{-\frac{1}{2}}$ in (7.136) is infinite, from (7.127), when q_j is equal to any other q. It is in this condition that $\mathbf{S}$ is singular and the arguments of § 7.14 (1) onwards must fail. Thus we expect the W.K.B. solution (7.136) to be invalid in regions of the complex z plane near where q_j is equal to any other q. It does not necessarily fail when q_j is zero, nor when two qs are equal if neither of them is q_j.

One important case of this failure is when $X \to 0$ so that the medium is free space. Then the four qs are in two equal pairs $\pm C$. It can be shown that for small X the denominator $(A_jF_j)^{\frac{1}{2}}$ in (7.136) is proportional to X but the four elements in brackets are also proportional to X. Thus each of the four field expressions (7.136) tends to a bounded non-zero limit when $X \to 0$. The limits are functions of $U = 1 - \mathrm{i}Z$ and may vary with height z when no electrons are present. This behaviour is of interest in the study of the phenomenon of 'limiting polarisation', discussed in §§ 17.10, 17.11.

7.17. W.K.B. solutions for vertical incidence

For vertical incidence, $S = 0$, the axes may be chosen so that $l_y = 0$. They are then the same as the x, y, z axes of ch. 4. For the electron plasma (7.123) shows that

$$\begin{aligned} a_1 &= a_3 = \bar{a}_3 = a_6 = \bar{a}_6 = 0 \\ a_2 &= -1 + \{U(U-X) - l_x^2Y^2\}D \\ a_4 &= a_5 = -\bar{a}_4 = -\bar{a}_5 = -\mathrm{i}l_zY(U-X)D. \end{aligned} \tag{7.137}$$

The roots q are the refractive indices n and the quartic (7.128) is

$$q^4 + q^2(a_2 - \mathsf{T}_{32}) + a_4^2 - a_2\mathsf{T}_{32} = 0 \tag{7.138}$$

which is the same as the dispersion relation (4.65). Its solutions may be taken as

$$q_1 = -q_2 = n_{\mathrm{O}}, \quad q_3 = -q_4 = n_{\mathrm{E}} \tag{7.139}$$

where n_{O}, n_{E} are the refractive indices for the ordinary and extraordinary waves.

Then

$$A_1 = A_2 = n_O^2 + a_2, \quad A_3 = A_4 = n_E^2 + a_2, \quad A_1A_3 = a_4^2 \tag{7.140}$$

$$F_1 = -F_2 = 2n_O(n_O^2 - n_E^2), \quad F_3 = -F_4 = -2n_E(n_O^2 - n_E^2). \tag{7.141}$$

The expression (7.134) for a column of **S** may now be written down. In particular

$$(S_{11}, S_{21}, S_{31}, S_{41}) = (A_1F_1)^{-\frac{1}{2}}\{a_4, n_O^2 + a_2, n_O(n_O^2 + a_2), a_4n_O\} \tag{7.142}$$

which refers to the upgoing ordinary wave and shows that for this wave

$$\frac{E_y}{E_x} = -\frac{\mathscr{H}_x}{\mathscr{H}_y} = -\frac{n_O^2 + a_2}{a_4} = -\frac{A_1}{a_4} = \rho_O \tag{7.143}$$

from the definition (4.13), (4.14) of ρ. The same result is obtained for the downgoing ordinary wave (see end of § 4.4 for remarks on the sign of ρ). Similarly for the upgoing or downgoing extraordinary waves

$$\frac{E_y}{E_x} = -\frac{n_E^2 + a_2}{a_4} = -\frac{A_3}{a_4} = \rho_E. \tag{7.144}$$

Equations (7.140) now show that $\rho_O\rho_E = 1$, in agreement with (4.29). The field components E_x, $-E_y$, $\mathscr{H}_x$, $\mathscr{H}_y$ given by the four W.K.B. solutions (7.136) are then

Upgoing ordinary wave

$$-\{2(\rho_O^2 - 1)n_O\}^{-\frac{1}{2}}(1, -\rho_O, -\rho_On_O, n_O)\exp\left(-ik\int^z n_O dz\right). \tag{7.145}$$

Downgoing ordinary wave

$$-\{-2(\rho_O^2 - 1)n_O\}^{-\frac{1}{2}}(1, -\rho_O, \rho_On_O, -n_O)\exp\left(ik\int^z n_O dz\right). \tag{7.146}$$

Upgoing extraordinary wave

$$-\{2(\rho_O^2 - 1)n_E\}^{-\frac{1}{2}}(\rho_O, -1, -n_E, \rho_On_E)\exp\left(-ik\int^z n_E dz\right) \tag{7.147}$$

Downgoing extraordinary wave

$$-\{-2(\rho_O^2 - 1)n_E\}^{-\frac{1}{2}}(\rho_O, -1, n_E, -\rho_On_E)\exp\left(ik\int^z n_E dz\right). \tag{7.148}$$

7.18. Ray theory and 'full wave' theory

In this chapter it has been shown that the W.K.B. solutions may be used at nearly all levels in a slowly varying stratified medium. They fail near certain levels where reflection or coupling between the characteristic waves is occurring.

Every W.K.B. solution is of the form (7.136). It is the product of the exponential which expresses the 'phase memory' concept, and the other terms, which vary much more slowly than the exponential. The exponent is equal to $-i$ multiplied by the generalised or complex phase of the wave, which is calculated as though the wave travelled at each level according to the laws of geometrical optics. A W.K.B. solution is therefore often said to be a mathematical expression of 'ray' theory. A more precise meaning for the term 'ray' is given in ch. 10.

When the W.K.B. solutions fail, it is necessary to make a further study of the differential equations, and to find a solution which cannot in general be interpreted in terms of geometrical optics. Such a solution is sometimes called a 'full wave' solution. For high frequencies, greater than about 1 MHz, most of the important aspects of radio wave propagation in the ionosphere can be handled by 'ray theory'. At levels of reflection or coupling, the W.K.B. solutions fail, but an investigation of the reflection or coupling process by 'full wave' theory shows that in general the 'ray theory' can still be considered to hold, with only trivial modification. An exception to this is the phenomenon of partial penetration and reflection for frequencies near the penetration frequency of an ionised layer (§ 15.10). Here the ray theory is inadequate, and a 'full wave' treatment is necessary.

For lower frequencies, less than about 1 MHz, and especially for very low frequencies, less than about 100 kHz, the ionosphere can change appreciably within a distance of one wavelength, and cannot always be regarded as a slowly varying medium. Then the W.K.B. solutions fail to be good approximations, and a full wave solution is required for nearly every problem. Chs. 10 to 14 deal with problems that can be treated by ray theory methods. The later chapters are almost entirely concerned with 'full wave' solutions.

7.19. The reflection coefficient

Consider again the W.K.B. solutions (7.26) for horizontally polarised waves in an isotropic stratified ionosphere, as discussed in §§ 7.2–7.11. Suppose that an upgoing radio wave of unit amplitude is generated at the ground, $z = 0$. Then its W.K.B. solution

$$E_y = C^{\frac{1}{2}} q^{-\frac{1}{2}} \exp\left(-ik \int_0^z q\,\mathrm{d}z \right) \tag{7.149}$$

gives the complex amplitude at any other level where q is not near to zero. The constant $C^{\frac{1}{2}}$ is included because, in the free space at the ground, $q = C$. It is required to find the amplitude R of the resulting reflected wave when it reaches the ground. Its field at other levels is given by its W.K.B. solution

$$E_y = RC^{\frac{1}{2}} q^{-\frac{1}{2}} \exp\left(ik \int_0^z q\,\mathrm{d}z \right). \tag{7.150}$$

These W.K.B. solutions fail where $q = 0$, and the reflection occurs there. If, nevertheless, it is supposed that (7.149) and (7.150) tend to equality when $q \to 0$, then clearly

$$R = \exp\left(-2ik \int_0^{z_0} q\,\mathrm{d}z \right) \tag{7.151}$$

where z_0 is the value of z that makes $q = 0$. A more detailed study of the differential equations in the region near $z = z_0$, given in § 8.20, shows that this formula needs a small modification. The right-hand side must be multiplied by a factor i. This represents a phase advance of $\frac{1}{2}\pi$ that is not predicted by the cumulative phase change given by the integral. For many purposes it is unimportant, but will be included for completeness. Thus the formula (7.151) should read

$$R = \mathrm{i}\exp\left(-2\mathrm{i}k\int_0^{z_0} q\,\mathrm{d}z\right). \tag{7.152}$$

In a similar way (7.150) shows that the complex amplitude of the reflected wave at a height $z < z_0$ within the ionosphere is

$$E_y = \mathrm{i}C^{\frac{1}{2}}q^{-\frac{1}{2}}\exp\left(-\mathrm{i}k\int_0^{z_0} q\,\mathrm{d}z + \mathrm{i}k\int_{z_0}^{z} q\,\mathrm{d}z\right). \tag{7.153}$$

The two integrals together express the total change of the complex phase of the wave in its passage from the ground to the reflection level z_0 and then back to the level z. They may be combined into a single integral called the 'phase integral'. If the effect of collisions is included, the condition $q = 0$ does not hold for any real value of z. Often, however, q is a known analytic function of z and is zero for a complex value $z = z_0$. Then the phase integral is a contour integral in the complex z plane. This is the basis of the 'phase integral method'; see §§ 8.21, 16.7.

The same property applies for the more general anisotropic ionosphere. The definition of the reflection coefficient R is then more complicated and depends on the wave polarisations; see ch. 11. But it can be shown in a similar way that if q_1 refers to the upgoing incident wave and if q_2 refers to the reflected wave, then the reflection coefficient at the ground contains a factor

$$\mathrm{i}\exp\left\{-\mathrm{i}k\int_0^{z_0}(q_1 - q_2)\,\mathrm{d}z\right\} \tag{7.154}$$

where the reflection level z_0 is now the value of z, possibly complex, for which $q_1 = q_2$. Similarly for some other $z < \mathrm{Re}(z_0)$ the field of the reflected wave contains a factor

$$\mathrm{i}\exp\left\{-\mathrm{i}k\left(\int_0^{z_0} q_1\,\mathrm{d}z + \int_{z_0}^{z} q_2\,\mathrm{d}z\right)\right\}. \tag{7.155}$$

These results, without the first factor i, are used in the theory of ray tracing, ch. 10.

PROBLEMS 7

7.1. Formulate the differential equation governing the transverse displacement y for waves of constant frequency on a string, stretched with constant tension T, whose mass per unit length $m(x)$ is a function of position, x. Find the W.K.B. solutions and deduce how the amplitude of a 'progressive' wave depends on m.

7.2. A perfect gas is stratified so that the ambient pressure P and temperature T are slowly varying functions of z. For a very weak plane sound wave of angular frequency ω, with its wave normal parallel to the z axis, find the differential equations satisfied by the excess pressure p and the displacement ζ. Find the W.K.B. solutions for the cases when (a) P is independent of z but T varies, (b) T is independent of z but P varies. In each case find how the amplitudes of p and ζ depend on P or T, and verify that, for a progressive wave, the average energy flow is constant.

7.3. A simple pendulum has a bob of mass m. The string passes through a fixed hole and its length L is made to vary slowly in time t. The pendulum swings with small amplitude. How do the amplitude a, the period T and the energy E of the oscillations depend on L?

7.4. A radio wave travels upwards into the horizontally stratified ionosphere and its wave normal is vertical. The coordinate z is vertically upwards. The earth's magnetic field is allowed for and the vector $\boldsymbol{Y}$ is in the x–z plane. Collisions are neglected. The wave is an ordinary wave with real and positive refractive index n_{O} and polarisation $\rho_{\mathrm{O}} = E_y/E_x$. Suppose that the ionosphere is so slowly varying that no appreciable reflection is occurring. Find an expression for the z component Π_z of the time averaged Poynting vector. Show that if this is independent of z, the field components must be given by the W.K.B. solution (7.145).

In the same problem, but with collisions allowed for, show that for the characteristic polarisations ρ_{O}, ρ_{E} of the ordinary and extraordinary waves, the adjoints $\bar{\rho}_{\mathrm{O}}$, $\bar{\rho}_{\mathrm{E}}$ are $-\rho_{\mathrm{O}}$, $-\rho_{\mathrm{E}}$ respectively. Hence show that, for a W.K.B. solution the bilinear concomitant W, (7.97), is independent of height z.

7.5. A radio wave travels obliquely upwards into an ionosphere simulated by a Chapman layer with scale height $H = 10\,\text{km}$. The earth's magnetic field and collisions are neglected. The wave is reflected at height $z = z_0$. In a range $|z - z_0|$ near this the W.K.B. solutions fail to be good approximations. Use the criterion (7.61) to find an expression for this range. Find its minimum value, that is the value when z_0 is where $\mathrm{d}N/\mathrm{d}z$ is a maximum. What is this minimum value if the frequency is 1 MHz?