6

Stratified media. The Booker quartic

6.1. Introduction

The earlier chapters have discussed the propagation of a plane progressive radio wave in a homogeneous plasma. It is now necessary to study propagation in a medium that varies in space, and this is the main subject of the rest of this book. The most important case is a medium that is plane stratified, and the later chapters are largely concerned with the earth's ionosphere that is assumed to be horizontally stratified. Because of the earth's curvature the stratification is not exactly plane, but for many purposes this curvature can be neglected. In cases of curved stratification it is often possible to neglect the curvature and treat the medium as locally plane stratified. Some cases where the earth's curvature is allowed for are discussed in §§ 10.4, 18.8.

The theory of this chapter is given in terms of the earth's ionosphere. A Cartesian coordinate system x, y, z is used with x, y horizontal and z vertically upwards. These coordinates are not in general the same as the x, y, z of chs. 2–4. The composition of the plasma, that is the electron and ion concentrations N_e, N_i are assumed to be functions of z only. For the study of radio propagation the frequency is usually great enough for the effect of the ions to be ignored. When radio waves are reflected in the ionosphere the important effects occur in a range of height that is small enough for the spatial variation of the earth's magnetic induction $\boldsymbol{B}$ to be ignored. The vector $\boldsymbol{Y}$ (3.22), antiparallel to $\boldsymbol{B}$, is therefore assumed to be constant in magnitude and direction and to have direction cosines l_x, l_y, l_z.

The origin $z = 0$ is usually taken to be at the ground. Below the ionosphere there is a region of free space which extends up to a height of about $z = 70$ km or more.

In the theory of the reflection of radio waves from the ionosphere it is usually assumed that the incident wave is a plane wave in the free space, and travelling obliquely upwards. In practice the incident wave is nearly always a spherical wave coming from a transmitter of small dimensions, but at a range of 70 km or more the curvature of the wave front is small so that the wave can be treated as plane to a first approximation. The validity of this is examined in § 11.14. Alternatively the incident wave can be expressed as an integral, (6.11) below, whose integrand represents a plane wave. It is thus represented as an 'angular spectrum of plane waves' (Booker and Clemmow, 1950; Clemmow, 1966; see also §§ 5.3, 10.2, 10.3, 11.14). The propagation of one plane wave component of the incident wave is first studied, and the properties for the spherical wave are then found by an integration. For details see §§ 11.14, 11.15. Here we shall give the theory for an incident plane wave.

6.2. The variable q

Let a plane wave be incident on the ionosphere from below, and in the free space let its wave normal have direction cosines S_1, S_2, C so that it makes an angle θ with the vertical, where $\cos\theta = C$, $\sin^2\theta = S_1^2 + S_2^2$, and each field component contains the factor

$$\exp\{-ik(S_1 x + S_2 y + Cz)\}. \tag{6.1}$$

Now imagine that the ionosphere is replaced by a number of thin discrete strata, in each of which the medium is homogeneous. By making these strata thin enough and numerous enough we may approximate as closely as we please to the actual ionosphere. The incident wave is partially reflected and partially transmitted at the lower boundary of the first stratum. The transmitted wave is partially reflected and transmitted at the second boundary, and so on. In any one stratum there are upgoing and downgoing waves which are the resultants of all the partially reflected and transmitted waves entering the stratum. Because the medium is doubly refracting, there are actually two obliquely upgoing and two obliquely downgoing waves. At each boundary the field components must satisfy some boundary conditions so that they must all depend on x and y in the same way on the two sides of each boundary. Hence the waves depend on x and y only through the factor

$$\exp\{-ik(S_1 x + S_2 y)\} \tag{6.2}$$

in all the strata.

In the r^{th} stratum let the wave normal of one of the four resultant waves make an angle θ_r with the z axis, and let the refractive index for this wave be n_r. Then each field component contains the factor

$$\exp[-ik\{S_1 x + S_2 y + n_r(\cos\theta_r)z\}]. \tag{6.3}$$

Further

$$n_r \sin\theta_r = (S_1^2 + S_2^2)^{\frac{1}{2}} = \sin\theta \tag{6.4}$$

which may be regarded as the result of applying Snell's law of refraction in succession at each of the boundaries in the region between the r^{th} stratum and the free space. Now n_r in general depends on the direction of the wave normal, that is on

S_1, S_2 and θ_r. The unknown angle θ_r could be found from (6.4) if n_r were known, but it depends on the angle Θ between the wave normal and $\boldsymbol{Y}$ and this in turn depends on θ_r. Hence (6.4) cannot immediately be used to give θ_r. To overcome this difficulty introduce the quantity

$$q = n_r \cos\theta_r \tag{6.5}$$

so that

$$n_r^2 = S_1^2 + S_2^2 + q^2. \tag{6.6}$$

The expression (6.3) now becomes

$$\exp\{-\mathrm{i}k(S_1x + S_2y + qz)\} \tag{6.7}$$

and it appears as a factor in all field quantities of the wave being considered. There are in general four such waves, and it is shown in the following section that the four values of q are the roots of a quartic equation, the Booker quartic.

The total field in the stratum is the sum of four terms like (6.7) with the common factor (6.2) but with different values of q and with amplitudes that do not depend on x and y. In any one stratum these amplitudes do not depend on z, but they change in going from one stratum to another because of the partial reflections and transmissions at the boundaries. In the limit of infinitesimally thin strata, that is in a continuous stratified medium, the four amplitudes depend on z but not on x and y.

For the special case where the medium is isotropic, n_r is independent of θ_r and is known at all heights z, from § 4.2. Then (6.6) can be used at once to give q. The four values of q then consist of two equal pairs given by

$$q = \pm(n_r^2 - S_1^2 - S_2^2)^{\frac{1}{2}}. \tag{6.8}$$

Objections are sometimes raised to the assertion that the ionosphere is equivalent to the system of discrete homogeneous strata, perhaps because, however thin and numerous we make the strata, there are discontinuities at their boundaries that are not present in the actual ionosphere. This difficulty can be overcome by using the differential equations (2.45), in which the variables x, y, z are chosen to be the same as in §6.1. The electric displacement is $\boldsymbol{D} = \varepsilon_0\boldsymbol{\varepsilon}\boldsymbol{E}$ and the elements of $\boldsymbol{\varepsilon}$ are functions of z only. To find solutions of this set of six partial differential equations the method of 'separation of the variables' may be used. We seek solutions in which each of the six dependent variables E_x, E_y, E_z, $\mathscr{H}_x$, $\mathscr{H}_y$, $\mathscr{H}_z$ is a product of the form

$$E_x = \mathscr{X}(x)\mathscr{Y}(y)\mathscr{Z}(z) \tag{6.9}$$

with similar expressions for the other five variables. These are substituted into (2.45). It is then clear that the factors $\mathscr{X}(x)$ and $\mathscr{Y}(y)$ can be cancelled from the equations, provided that the functions $\mathscr{X}(x)$ and $\mathscr{Y}(y)$ are the same for all six variables, and that

$$\frac{1}{\mathscr{X}}\frac{\mathrm{d}\mathscr{X}}{\mathrm{d}x} = \text{constant}; \quad \frac{1}{\mathscr{Y}}\frac{\mathrm{d}\mathscr{Y}}{\mathrm{d}y} = \text{constant}. \tag{6.10}$$

In the differential equations that remain, the variables and the coefficients are functions only of z. These equations are given and discussed in § 7.13, where it is shown how they lead to the Booker quartic.

The two constants in (6.10) are called 'separation constants', and they are the same as $-ikS_1$, $-ikS_2$ in (6.2), (6.3), (6.7). For each plane $z=$ constant, the fields may be expressed as a double Fourier integral thus

$$E_x(x, y, z) = \iint \breve{E}_x(S_1, S_2, z) \exp\{-ik(S_1 x + S_2 y)\} \mathrm{d}S_1 \mathrm{d}S_2 \tag{6.11}$$

with similar expressions for the other five field components. The process of separating the variables, described above only in outline, shows that the differential equations for any set of Fourier amplitudes $\breve{E}_x(S_1, S_2, z)$ etc., for given S_1, S_2, are independent of those for any other pair S_1, S_2. The use of the incident plane wave (6.1) is equivalent to studying just one of the Fourier amplitudes. This technique is so widely used that the Fourier interpretation is rarely mentioned. The field components are written simply as $E_x(z)$ etc. instead of $\breve{E}_x(S_1, S_2, z)$ etc. The functional dependence on S_1 and S_2 is omitted, and the spatial dependence factor (6.2) is also omitted in the same way as the time factor $\exp(i\omega t)$ is omitted (§ 2.5). These conventions are used in this book, except where otherwise stated.

6.3. The Booker quartic. Derivation

Consider again one of the four plane waves with the properties (6.5), (6.6), in the r^{th} stratum of § 6.2. The directions cosines of its wave normal are

$$S_1/n_r,\ S_2/n_r,\ q/n_r \tag{6.12}$$

so that the angle Θ between the wave normal and the vector $\boldsymbol{Y}$ (direction cosines l_x, l_y, l_z) is given by

$$n_r \cos\Theta = (l_x S_1 + l_y S_2 + l_z q). \tag{6.13}$$

Now (6.6) and the square of (6.13) are substituted into the form (4.62) or (4.65) of the dispersion relation. This gives an equation for finding q, and inspection shows that is a quartic equation.

We use, first, the form (4.65) for an electron plasma. It gives

$$\begin{aligned}
&(q^2 + S_1^2 + S_2^2)^2\{U^2(U-X) - UY^2\} \\
&\quad + (q^2 - C^2)(l_x S_1 + l_y S_2 + l_z q)^2 X Y^2 \\
&\quad - (q^2 + S_1^2 + S_2^2)\{2U(U-X)^2 - 2Y^2(U-X) - XY^2\} \\
&\quad + (U-X)^3 - Y^2(U-X) = 0
\end{aligned} \tag{6.14}$$

where the property $1 - S_1^2 - S_2^2 = C^2$ has been used. This may be written

$$F(q) \equiv \alpha q^4 + \beta q^3 + \gamma q^2 + \delta q + \varepsilon = 0 \tag{6.15}$$

where

$$\left.\begin{aligned}
\alpha &= U(U^2 - Y^2) + X(Y^2 l_z^2 - U^2),\\
\beta &= 2(l_x S_1 + l_y S_2) l_z X Y^2,\\
\gamma &= -2U(U - X)(C^2 U - X) + 2Y^2(C^2 U - X)\\
&\quad + XY^2(1 - l_z^2 C^2 + l_x^2 S_1^2 + l_y^2 S_2^2),\\
\delta &= -2C^2(l_x S_1 + l_y S_2) l_z X Y^2,\\
\varepsilon &= (U - X)(C^2 U - X)^2 - C^2 Y^2 (C^2 U - X) - (l_x S_1 + l_y S_2)^2 C^2 X Y^2.
\end{aligned}\right\} \tag{6.16}$$

The equations (6.15), (6.16) were given by Booker (1936, 1939, 1949) in three papers which are about the most important papers in the whole of the theory of radio wave propagation. The notation used here is almost the same as Booker's original notation. The coefficients α, β must not be confused with the angles α, β used in ch. 5.

A similar process can be used with (4.62) for the more general plasma when ions are allowed for, or when the electron collision frequency is velocity dependent, for example with the Sen–Wyller formulae, § 3.12. The notation is as in § 3.10 including (3.51) for G and J. We use $S^2 = S_1^2 + S_2^2$. Then (6.16) is replaced by

$$\left.\begin{aligned}
\alpha &= \tfrac{1}{2}(\varepsilon_1 + \varepsilon_2) - l_z^2 G,\\
\beta &= -2(l_x S_1 + l_y S_2) l_z G,\\
\gamma &= -\tfrac{1}{2}\varepsilon_3(\varepsilon_1 + \varepsilon_2) - \varepsilon_1\varepsilon_2 + S^2(\varepsilon_1 + \varepsilon_2 - l_z^2 G) - l_z^2 J - (l_x S_1 + l_y S_2)^2 G,\\
\delta &= -2(l_x S_1 + l_y S_2) l_z (J + S^2 G),\\
\varepsilon &= \{\varepsilon_1\varepsilon_2 - \tfrac{1}{2}(\varepsilon_1 + \varepsilon_2) S^2\}(\varepsilon_3 - S^2) - (l_x S_1 + l_y S_2)^2 (J + S^2 G).
\end{aligned}\right\} \tag{6.17}$$

When the values (3.50) for the electron plasma are used in (6.17), and the coefficients are all multiplied by $U(U^2 - Y^2)$, it can be shown that they are then the same as (6.16).

There are several other methods of deriving the Booker quartic. An important method is given in § 7.13. Another method of some interest is worth mentioning briefly. It has the slight advantage that it makes direct use of Maxwell's equations and does not use the dispersion relation. It is required to find q for one of the four waves, in the general stratum of § 6.2, whose fields have the spatial dependence (6.7), so that for all field variables

$$\frac{\partial}{\partial x} \equiv -ikS_1, \quad \frac{\partial}{\partial y} \equiv -ikS_2, \quad \frac{\partial}{\partial z} \equiv -ikq. \tag{6.18}$$

Then Maxwell's equations (2.44) may be written in matrix notation

$$\mathbf{\Gamma} \boldsymbol{E} = \boldsymbol{\mathscr{H}}, \quad \mathbf{\Gamma}\boldsymbol{\mathscr{H}} = -(\mathbf{1} + \boldsymbol{M})\boldsymbol{E}, \tag{6.19}$$

compare (4.70), where the matrix $\mathbf{\Gamma}$ is

$$\begin{pmatrix} 0 & -q & S_2 \\ q & 0 & -S_1 \\ -S_2 & S_1 & 0 \end{pmatrix}. \tag{6.20}$$

In (6.19) **1** is the unit 3×3 matrix and $\mathbf{1} + \boldsymbol{M}$ is the permittivity expressed in the coordinate system of § 6.1. For an electron plasma the susceptibility matrix $\boldsymbol{M}$ is given by (3.35). Elimination of $\mathscr{H}$ gives

$$(\boldsymbol{\Gamma}^2 + \boldsymbol{M} + \mathbf{1})\boldsymbol{E} = 0. \tag{6.21}$$

If this set of three equations is to have a self-consistent solution $\boldsymbol{E}$, it is necessary that the determinant of the matrix shall be zero. This gives

$$\begin{vmatrix} 1 - q^2 - S_2^2 + M_{xx} & S_1 S_2 + M_{xy} & S_1 q + M_{xz} \\ S_1 S_2 + M_{yx} & 1 - q^2 - S_1^2 + M_{yy} & S_2 q + M_{yz} \\ S_1 q + M_{zx} & S_2 q + M_{zy} & C^2 + M_{zz} \end{vmatrix} = 0. \tag{6.22}$$

When this is multiplied out and (3.35) is used, it gives the Booker quartic equation (6.15), (6.16).

6.4. Some properties of the Booker quartic

The Booker quartic is a generalised form of the dispersion relation for a cold magnetoplasma. It was obtained in § 6.3 from the dispersion relation by rotating the Cartesian axes in refractive index space, so that the axis of $n_z = q$ is perpendicular to a given fixed plane. This plane is usually chosen to be parallel to the strata in a medium such as the ionosphere. The numbers S_1, S_2, q are the Cartesian components of the vector $\boldsymbol{n}$. The quartic is therefore the equation of the system of two refractive index surfaces in these new axes.

This choice of axes still leaves us free to rotate the axis system about the n_z axis. The components $n_x = S_1$ and $n_y = S_2$ were determined by the incident plane wave (6.1) in free space. It is nearly always convenient to choose the x and y axes so that the wave normal of this wave is in the x–z plane and therefore $S_2 = 0$. This is done for nearly all the problems discussed in this book. We also write S for S_1, so that $C^2 = 1 - S^2$. The coefficients (6.16) for an electron plasma then become

$$\left.\begin{aligned} \alpha &= U(U^2 - Y^2) + X(Y^2 l_z^2 - U^2), \\ \beta &= 2 l_x l_z S X Y^2, \\ \gamma &= -2U(U - X)(C^2 U - X) + 2Y^2(C^2 U - X) + XY^2(1 - l_z^2 C^2 + l_x^2 S^2), \\ \delta &= -2 l_x l_z S C^2 X Y^2, \\ \varepsilon &= (U - X)(C^2 U - X)^2 - C^2 Y^2 (C^2 U - X) - l_x^2 S^2 C^2 X Y^2. \end{aligned}\right\} \tag{6.23}$$

Similar results for the more general plasma can be derived from (6.17). Some authors, for example Pitteway (1965) have preferred to retain S_1, S_2 and to choose the axes so that $l_y = 0$. This means that the vector $\boldsymbol{Y}$ is in the x–z plane. The resulting formulae can easily be derived from (6.16) or (6.17) but are not needed in this book.

To illustrate the properties of the quartic it is useful first to ignore collisions as was done when studying the refractive indices in §§ 4.11, 4.12, 5.2. Then $U = 1$ and the coefficients (6.23) are

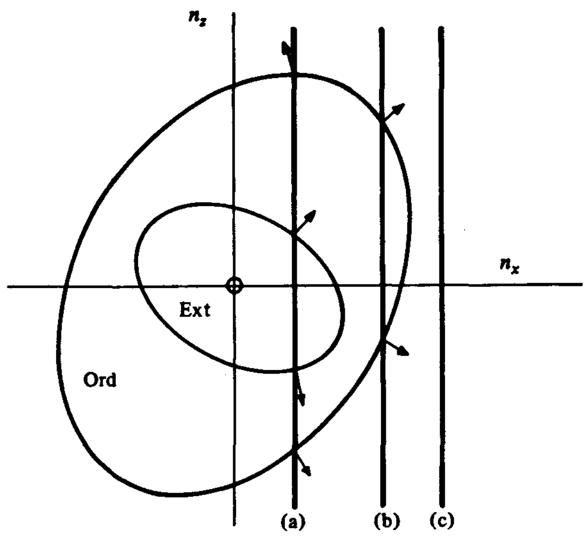
$$\left.\begin{aligned}
\alpha &= 1 - X - Y^2 + XY^2 l_z^2,\\
\beta &= 2 l_x l_z S X Y^2,\\
\gamma &= -2(1-X)(C^2-X) + 2Y^2(C^2-X) + XY^2(1 - l_z^2 C^2 + l_x^2 S^2),\\
\delta &= -2 l_x l_z S C^2 X Y^2,\\
\varepsilon &= (1-X)(C^2-X)^2 - C^2 Y^2 (C^2 - X) - l_x^2 S^2 C^2 X Y^2.
\end{aligned}\right\} \tag{6.24}$$

Some typical curves of q versus X for this case are given in §§6.7–6.9.

In some problems S, C, X can take complex values but here we consider only the simple situation where they are all real. Then the coefficients (6.24) are all real. The four roots of the quartic (6.15) are then either (a) all real, or (b) two are real and two form a conjugate complex pair, or (c) the roots are in two conjugate complex pairs.

This property can be related to the geometry of the refractive index surfaces. Imagine that in a refractive index space with axes n_x, n_y, n_z parallel to the axes x, y, z of § 6.1, the two refractive index surfaces are drawn. Fig. 6.1 shows a cross section of these surfaces by the plane $n_y = 0$. Now let a line be drawn parallel to the n_z axis with $n_x = S$, $n_y = 0$. The values of n_z where it cuts the refractive index surfaces are the four roots q of the quartic. Three possible positions for this line are shown in fig. 6.1, for the three cases (a), (b), (c). For case (a) the line may cut both refractive index surfaces

Fig. 6.1. Cross section by the plane $n_y = 0$ of the two refractive index surfaces. Each of the three lines (a), (b), (c) is for the fixed values $n_x = S$, $n_y = 0$. The points where one of these lines cuts the surfaces give the values of $q = n_z$. The radius from the origin has the direction of the wave normal and its length is the refractive index n. The arrows show the projections on to the plane of the diagram of the normals to the surfaces, that is the ray directions, at the points of intersection. This example is typical of a loss-free electron plasma with $Y < 1$, $X < 1 - Y$.



at two real points each, as shown in the figure, or it may cut one refractive index surface at four real points and not cut the other at all. This last possibility can occur, for example, when one refractive index surface has points of inflection as in figs. 5.10 or 5.11. In case (b) the line cuts only one of the refractive index surfaces, at two real points. In case (c) the line does not cut either refractive index surface at real points.

In case (a) imagine that the outward normals to the refractive index surface are drawn at the four real points of intersection. These normals give the directions of the four rays, which are also the directions of energy flux (§ 5.3). They are not in the plane of the diagram except in special cases. Their projections are shown in fig. 6.1. It is clear from the form of the refractive index surfaces, figs. 5.4–5.13, that two of these normals must make acute angles with the n_z axis and the other two must make obtuse angles with it. Thus two of the four waves are obliquely upgoing and the other two are obliquely downgoing. When two of the qs form a complex conjugate pair, it can be shown, see § 7.14(8), that for both the associated waves the component Π_z of the energy flux is zero. The wave for which $\mathrm{Im}(q)$ is negative is attenuated as z increases. If now we imagine that the collision frequency is given a very small value, energy must be absorbed from the wave at a rate H, say, per unit volume per unit time, and therefore Π_z must now have a small positive value so that

$$\mathrm{d}\Pi_z/\mathrm{d}z = 2k\,\mathrm{Im}(q)\Pi_z = -H. \tag{6.25}$$

This wave is therefore an upgoing wave. The other wave has positive $\mathrm{Im}(q)$ and is attenuated as z decreases. By a similar argument it must be a downgoing wave. Thus in all cases two of the four qs correspond to upgoing waves and two to downgoing waves. This result is still true when the collision frequency is large, provided that S is real. The proof follows.

If the reference line used in fig. 6.1 just touches one of the refractive index surfaces, the two points of intersection have moved together so that the quartic has two equal roots. This is associated with reflection of the wave and is described in detail in § 10.6.

When Z is not zero, so that the form (6.23) of the coefficients is used, no root of the quartic can be real, when S and X are real. For if q is real, (6.5), (6.6) show that n_r and Θ_r are both real. But it was shown in § 4.13 that when the wave normal is in a real direction the refractive index cannot be purely real unless $Z = 0$.

When Z is not zero and S, X are real, two of the roots of the quartic must have negative imaginary parts and the other two must have positive imaginary parts. This can be proved as follows. Consider some set of fixed values of X, U, Y, l_x, l_y, l_z for which the coefficient α is not zero. Starting with $S = 0$, let S be slowly increased (so that C^2 decreases). No root of the quartic can be real for any real value of S, so the imaginary parts of the roots cannot change sign. Now when $S = 0$, so that $q = n_z = n$, the four values of q are the four values of n given by the dispersion relation, and n^2 always has a negative imaginary part (§ 4.13). Hence two values of n have negative imaginary parts, for the upgoing waves, and two have positive imaginary parts, for the downgoing waves. This must also apply for other real values of S and therefore always applies to the four roots of the quartic.

A wave for which $\mathrm{Im}(q)$ is negative is attenuated as z increases, and it is an (obliquely) upgoing wave. Similarly when $\mathrm{Im}(q)$ is positive, the wave is downgoing. It will be shown later that for an upgoing wave, $\mathrm{Re}(q)$ may be either positive or negative. A negative value of $\mathrm{Re}(q)$ means that the wave fronts are travelling obliquely downwards. This apparent contradiction arises because the directions of the wave normal and the ray are in general different (§ 5.3).

When Z is zero some of the roots q may be real. To decide whether a given real q refers to an upgoing or downgoing wave, it is sometimes convenient to give Z a small non-zero value and then to examine the sign of $\mathrm{Im}(q)$. If it is negative, the wave is an upgoing wave.

In some physical problems a complex value of S is used. This occurs, for example, in the study of guided waves by the layers of a stratified system. It is then no longer necessarily true that two waves are upgoing and two are downgoing. It is possible to find a complex S such that three waves are upgoing and one downgoing, or three downgoing and one upgoing. But the subject of guided waves is beyond the scope of this book.

One root of the quartic is infinite when the coefficient $\alpha = 0$. For the collisionless electron plasma, (6.24), this gives

$$X = X_\infty \equiv (1 - Y^2)/(1 - Y^2 l_z^2) \tag{6.26}$$

which is independent of S, C, that is of the angle of incidence Θ. The remaining solutions are now given by a cubic equation for q, with real coefficients when the plasma is loss-free. Then at least one of these solutions must be real.

One root of the quartic is zero when the coefficient $\varepsilon = 0$. From (6.23) or (6.24) this gives a cubic equation for X. Its solutions for the collisionless case (6.24) have been studied by Booker (1939), Budden (1961a). When $S^2 \leqslant 1$, $l_x^2 \leqslant 1$ the solutions X are all real. By studying the function $\varepsilon(X)$ it can be shown that one solution is in the range $X \geqslant 1$, one is in the range $C^2 \leqslant X \leqslant 1$, and one in the range $X < C^2$. This last solution is positive if $Y^2 < 1$.

It must be remembered that the levels where $q = 0$ are not necessarily levels of reflection; see § 10.6.

When $X = 1$, the coefficients (6.24) all contain Y^2 as a factor. This may be cancelled from the quartic, which then becomes

$$\begin{aligned}(1 - l_z^2)q^4 - 2Sl_xl_zq^3 + (S^2 - C^2 + C^2l_z^2 - S^2l_x^2)q^2 + 2SC^2l_xl_zq \\ - S^2C^2(1 - l_x^2) = 0.\end{aligned} \tag{6.27}$$

Hence at the level where $X = 1$, the solutions q depend on the direction of the earth's magnetic field, but not on its magnitude. Equation (6.27) factorises and its four solutions are

$$q = \pm C, \quad q = S(l_x l_z \pm il_y)/(1 - l_z^2). \tag{6.28}$$

The last two are real only when $l_y = 0$, that is for propagation from magnetic north to south or south to north. Then the quartic has a double root equal to Sl_z/l_x, since $1 - l_z^2 = l_x^2$.

It is useful to display the solutions of the quartic as curves showing how q depends on X; see §§ 6.7–6.9. Since the quartic (6.15) is valid for all values of X, we must have $\mathrm{d}F/\mathrm{d}X = 0$, which gives

$$\frac{\partial F}{\partial q}\frac{\partial q}{\partial X} + \frac{\partial \alpha}{\partial X}q^4 + \frac{\partial \beta}{\partial X}q^3 + \frac{\partial \gamma}{\partial X}q^2 + \frac{\partial \delta}{\partial X}q + \frac{\partial \varepsilon}{\partial X} = 0 \tag{6.29}$$

whence

$$\frac{\partial q}{\partial X} = -\left(\frac{\partial \alpha}{\partial X}q^4 + \frac{\partial \beta}{\partial X}q^3 + \frac{\partial \gamma}{\partial X}q^2 + \frac{\partial \delta}{\partial X}q + \frac{\partial \varepsilon}{\partial X}\right)\Big/\frac{\partial F}{\partial q}. \tag{6.30}$$

The denominator $\partial F/\partial q$ is zero where the quartic has a double root. In general the numerator is not also zero so that $\partial q/\partial X$ is infinite. Then the curves of q vs X have a vertical tangent where there is a double root. For examples see figs. 6.2, 6.4–6.7.

The numerator of (6.30) can also be zero at a double root, in four special cases. When $X = 1$, $l_y = 0$ there is a double root given by the second equation (6.28), where $q = Sl_z/l_x$, and the numerator of (6.30) is then

$$\{Y^2(S^2 - l_x^2)^2 - S^4\}/l_x^4. \tag{6.31}$$

This is zero if

$$S = \pm l_x\{Y/(Y \pm 1)\}^{\frac{1}{2}}, \tag{6.32}$$

which gives four values of S. Only two of them are real if $Y < 1$. They are the values of S for which the reference line of fig. 6.1 passes through one of the four window points. For a double root at $X = 1$ where (6.32) is satisfied, two curves of q vs X intersect, and their tangents are not vertical. There is an example in fig. 6.6(c). The condition (6.32) is also related to the phenomena of the 'Spitze', § 10.9, and of radio windows, §§ 17.6–17.9.

It follows from (6.28) that three roots of the quartic are equal at the level $X = 1$ if $l_y = 0$ and $C = Sl_z/l_x$. This means that the wave normal of the incident wave (6.1) in free space is either parallel or antiparallel to the earth's magnetic field. It can be shown that for this condition the numerator of (6.30) is not zero. For three roots to be equal it is necessary that both $\partial F/\partial q$ and $\partial^2 F/\partial q^2$ are zero. Thence it can be shown, by multiplying (6.29) by $\partial X/\partial q$ and then differentiating with respect to q, that $\partial^2 X/\partial q^2$ is zero. Hence the curve of q vs X has a vertical tangent and a point of inflection where $X = 1$. There is an example in fig. 6.6(e).

For the illustrative examples given later in this chapter it is assumed that: (a) l_z is positive. Thus the results apply for the northern hemisphere. (b) S is positive, and l_x, when non-zero, is positive. This means that the wave normal of the incident wave (6.1) makes an acute angle with the vector $\mathbf{Y}$, and the wave normal of any reflected wave makes an obtuse angle with $\mathbf{Y}$.

If the sign of either S or l_x is reversed, the only effect is to reverse the signs of the coefficients β, δ in (6.23) or (6.24). The four solutions of the resulting quartic are the same as before except that all four signs are reversed. Similarly if the sign of l_z is reversed, the only effect is to reverse the signs of β and δ, with consequent sign reversal of all four qs. Thus all cases of interest are covered by studying positive values of S, l_x, l_z.

6.5. Some special cases of the Booker quartic

When the coefficients β and δ in (6.23) or (6.24) are both zero, the quartic reduces to a quadratic equation for q^2. This happens when $X = 0$ so that the quartic is applied to a region of free space. It is then easy to show that the solutions are in two equal pairs

$$q = \pm C. \tag{6.33}$$

It happens in three other important special cases (Booker, 1939). The first is when $S = 0$ so that the waves are vertically incident on the ionosphere. Then $q = n_z = n$ and the quartic is simply the dispersion relation (4.65). Its solutions are the four values of n and their properties have been discussed in chs. 4, 5.

The second case is when $l_x = 0$, which means that the earth's magnetic field is in the y–z plane. The plane containing the wave normal is the x–z plane so that the waves travel from magnetic east to west or west to east. The solutions of the quadratic are then

$$q^2 = \frac{(C^2U - X)\{U(U-X) - Y^2\} - \frac{1}{2}XY^2(1 - C^2l_z^2) \pm XY\{\frac{1}{4}Y^2(1 - C^2l_z^2)^2 + l_z^2(C^2U - X)(U - X)\}^{\frac{1}{2}}}{U(U^2 - Y^2) + X(Y^2l_z^2 - U^2)}$$

$$= C^2 - \frac{X(U-X)}{U(U-X) - \frac{1}{2}Y^2(1 - C^2l_z^2) \pm Y\{\frac{1}{4}Y^2(1 - C^2l_z^2)^2 + l_z^2(C^2U - X)(U - X)\}^{\frac{1}{2}}}. \tag{6.34}$$

Some of their properties, for $U = 1$, are illustrated in § 6.7.

The third special case is when $l_z = 0$ so that the earth's magnetic field is horizontal. This therefore applies for propagation at the magnetic equator. The solutions of the quadratic are then

$$q^2 = \frac{(C^2U - X)\{U(U-X) - Y^2\} - \frac{1}{2}XY^2(1 + l_x^2S^2) \pm XY\{\frac{1}{4}Y^2(1 - l_x^2S^2)^2 + l_x^2S^2U(U - X)\}^{\frac{1}{2}}}{U^2(U - X) - UY^2}$$

$$= C^2 - \frac{X(U-X)}{U(U-X) - \frac{1}{2}Y^2(1 - l_x^2S^2) \pm Y\{\frac{1}{4}Y^2(1 - l_x^2S^2)^2 + l_x^2S^2U(U - X)\}^{\frac{1}{2}}}. \tag{6.35}$$

If in addition $l_x = 0$ so that propagation is east–west or west–east at the magnetic equator, the solutions (6.35) become

$$q^2 = C^2 - \frac{X}{U}, \; q^2 = C^2 - \frac{X(U-X)}{U(U-X)-Y^2}. \tag{6.36}$$

The first is independent of Y, which means that the waves are unaffected by the earth's magnetic field. It can be shown that in this case the electric field is horizontal and parallel to the y axis, so that the electrons are moved only in directions parallel to the earth's magnetic field, and are unaffected by it. The wave is an ordinary wave.

The solutions (6.35) are not discussed further here. For a detailed discussion see Booker (1939), Chatterjee (1952).

6.6. The discriminant of the Booker quartic

The condition that two roots of the quartic shall be equal is very important. It may be associated with reflection, § 10.6 or with coupling between two of the four waves, chs. 16, 17. The condition is that

$$\partial F/\partial q = 4\alpha q^3 + 3\beta q^2 + 2\gamma q + \delta = 0. \tag{6.37}$$

From this equation and the quartic (6.15), q may be eliminated to give a condition that uses only the five coefficients α to ε. The details are given in books on algebra. See, for example, Barnard and Child (1936), or Burnside and Panton (1912). The result is

$$\Delta = 0 \tag{6.38}$$

where

$$\Delta = 4\mathscr{I}^3 - \mathscr{J}^2; \quad \mathscr{I} = 12\alpha\varepsilon - 3\beta\delta + \gamma^2,$$
$$\mathscr{J} = 72\alpha\gamma\varepsilon + 9\beta\gamma\delta - 27(\alpha\delta^2 + \varepsilon\beta^2) - 2\gamma^3. \tag{6.39}$$

Here Δ is called the discriminant of the quartic. It can also be written as a determinant

$$\Delta = \frac{-27}{4}\begin{vmatrix} 4\alpha\varepsilon + 2\beta\delta - \gamma^2 & 6\beta\varepsilon - \gamma\delta & 6\gamma\delta - \beta\gamma \\ 6\beta\varepsilon - \gamma\delta & 8\gamma\varepsilon - 3\delta^2 & 16\alpha\varepsilon - \beta\delta \\ 6\alpha\delta - \beta\gamma & 16\alpha\varepsilon - \beta\delta & 8\alpha\gamma - 3\beta^2 \end{vmatrix}. \tag{6.40}$$

The four roots q of the quartic are functions of the height z because the coefficients α to ε depend on X and Z which are functions of z. For many purposes it is useful (see §§ 7.19, 8.21, 16.5, 16.7) to allow z and thence X, Z, to take complex values. Points in the complex z plane where (6.38) is satisfied are called 'coupling points'. They include 'reflection points' because reflection is a special case of coupling. These points have their greatest influence on radio propagation when they are on or near the real z axis.

The solutions of (6.38) for real X and Z were studied by Pitteway (1959), who used the form (6.40) for Δ. He showed that Δ has a factor X^4. Thus there are four coincident coupling points at the origin of the complex X plane. Here the medium is free space and the roots q are equal in two pairs (6.33). These coupling points are important in the study of limiting polarisation, § 17.10, 17.11. The remaining factor of Δ is of degree eight in X so that there are eight other coupling points of possible interest in the complex X plane. They were studied by Jones and Foley (1972) and by Smith, M.S. (1974a) who used real Z but complex X. They are discussed in § 16.5.

For vertical incidence, $S = 0$, the four roots q have the values $\pm n_{\mathrm{O}}$, $\pm n_{\mathrm{E}}$ where $n_{\mathrm{O}} n_{\mathrm{E}}$ are the refractive indices of the ordinary and extraordinary waves, respectively. One coupling point is where $n_{\mathrm{O}} = 0$ and is associated with the condition $X = U$ of (4.74), for reflection of the ordinary wave. It is labelled O in Jones and Foley's (1972) and in Smith's (1974a) diagrams, and in figs. 6.2, 6.4, 6.5. Two are where $n_{\mathrm{E}} = 0$ and they correspond to the conditions $X = U \pm Y$ in (4.74) for reflection of the extraordinary wave. They are labelled $\mathrm{E}_+, \mathrm{E}_-$ respectively in Smith's diagrams. A fourth coupling point is also a reflection point labelled R. It coincides with the resonance (4.76) when $S = 0$. The remaining four coupling points are labelled $\mathrm{C}_1, \mathrm{C}_2, \mathrm{C}_3, \mathrm{C}_4$ in Smith's diagrams. They are where the square root in (4.47) or (4.67) is zero, that is where

$$X = U \pm \tfrac{1}{2}\mathrm{i}Y \sin^2\Theta/\cos\Theta. \tag{6.41}$$

This should be compared with (4.27) where a real value of Θ was used. Here at C_i, $i = 1$ to 4, q is in general complex so that (6.41) uses a complex Θ. Each sign in (6.41) has two different C_is and Θs, one for upgoing waves and one for downgoing. This explains why there are four C_is.

When S changes, the coupling points move in the complex X plane, and those that are coincident move apart. But they can retain the labels assigned for vertical incidence, and can thus be identified. Their behaviour, when S takes successively greater real values, is fully described in Smith's paper (1974a). In later chapters the properties of these reflection and coupling points are referred to as they are needed. See § 16.5 and fig. 16.1

6.7. The Booker quartic for east–west and west–east propagation

To illustrate the properties of the solutions q for east–west or west–east propagation it is useful to plot curves of q^2 against X for the special case where collisions are neglected as was done for n^2 in §§ 4.11–4.12. Then $U = 1$ and the solutions (6.34) are

$$q^2 = \frac{(C^2 - X)(1 - X - Y^2) - \frac{1}{2}XY^2(1 - C^2 l_z^2) \pm XY\{\frac{1}{4}Y^2(1 - C^2 l_z^2)^2 + l_z^2(C^2 - X)(1 - X)\}^{\frac{1}{2}}}{1 - Y^2 - X(1 - Y^2 l_z^2)}. \tag{6.42}$$

This is to be studied for real values of l_z and of C.

If the wave (6.1) is vertically incident, $C = 1$ and $S = 0$. Then $l_z = \cos\Theta$ where Θ is the angle between the wave normal (vertical) and the vector $\boldsymbol{Y}$. The square root in (6.42) is the same as that in (4.67) with $U = 1$. It can never be zero for any real value of X (§ 5.2 item (5)). But if $C < 1$, the square root in (6.42) is not the same as in (4.67). It is now zero when

$$X^2 - X(1 + C^2) + C^2 + \tfrac{1}{4}Y^2(1 - C^2 l_z^2)/l_z^2 = 0 \tag{6.43}$$

and this has two real solutions X if

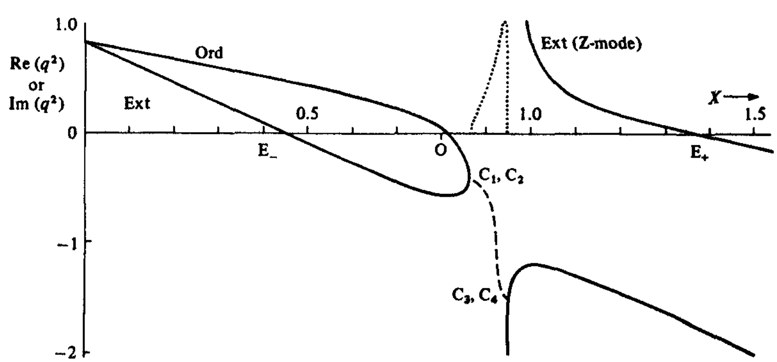
$$l_z^2 S^4 > Y^2(1 - C^2 l_z^2)^2 \tag{6.44}$$

(Booker, 1939). A case where this is satisfied is shown in fig. 6.2. When X is between the two solutions of (6.43) the square root is purely imaginary and the two values of q^2 are complex conjugates. Their real and imaginary parts are shown in fig. 6.2. If X is outside this range, both values of q^2 are real.

When the two values of q^2 are equal, that is where (6.43) is satisfied, they may be positive or negative. If they are positive they must occur for two coincident points on the same refractive index surface. For (6.6) shows that the two values of n^2 are then equal for the same real direction of the wave normal. It was shown in § 5.2 item (5) that the two refractive index surfaces cannot intersect. In the example of fig. 6.2, q^2 is negative at both values of X that satisfy (6.43).

When q^2 is positive the wave normal has a real direction (6.12) and the waves can be identified as ordinary or extraordinary, as indicated in fig. 6.1. But when q^2 is negative, q is imaginary and the wave normal direction (6.12) is complex. There can then be an ambiguity in the terms 'ordinary' and 'extraordinary' similar to the one that occurs when collisions are allowed for, as described in § 4.16. Cases could arise

Fig. 6.2. Shows how q^2 depends on X for propagation from magnetic east to west, or west to east. In this example $Y = \frac{1}{2}$, $l_z = 0.91868$, $C = 0.9$ and electron collisions are neglected. One solution is infinite where $X = 0.945$. The square root in the solution (6.42) is zero where $X = 0.865$ and 0.945. When X is between these values the two solutions are complex conjugates. The real part of q^2 is shown as a broken line and the imaginary part as a dotted line. The coupling points E$_-$, E$_+$, O, C$_1$ − C$_4$ are explained in § 6.6.



where the adjectives 'ordinary' and extraordinary' have to be interchanged. But for complex Θ there is no universally accepted convention for this.

One value of q^2 is infinite where (6.26) is satisfied. Here two roots of the Booker quartic are infinite. In the example of fig. 6.2 this X is very close to the greater of the two solutions of (6.43).

One value of q^2 is zero where the coefficient ε in (6.24), with $l_x^2 = 0$, is zero. One of the factors of ε gives

$$X = C^2. \tag{6.45}$$

This is associated with reflection of the ordinary wave. It gives the coupling point with the label O as explained in §6.6 and marked in fig. 6.2. The remaining factor of ε gives

$$X = 1 - \tfrac{1}{2}S^2 \pm (C^2Y^2 + \tfrac{1}{4}S^4)^{\frac{1}{2}}. \tag{6.46}$$

For $C = 1$, $S = 0$ these go over to the cut-off conditions $X = 1 \pm Y$ for the extraordinary wave. They correspond to the coupling points E_+, E_- of §6.6, and are associated with reflection of the extraordinary wave.

Where X is the lesser of the two solutions of (6.43) there are two coincident coupling points C_1, C_2, and for the greater solution there are again two coincident coupling points C_3, C_4. The labelling of §6.6 is not unique, however, and an alternative labelling is possible in which one of C_3, C_4 is replaced by R. (See Smith, 1974a).

For propagation from magnetic east to west or west to east, the ray is in general deviated out of the plane of incidence. This is called 'lateral deviation' and is explained in § 10.13.

For further discussion and results for east–west and west–east propagation see Booker (1939, 1949), Millington (1951, 1954), Chatterjee (1952), Budden (1961a).

6.8. The Booker quartic for north–south and south–north propagation

For propagation from magnetic north to south or south to north, the direction cosine $l_y = 0$ so that $l_x^2 + l_z^2 = 1$. This does not give any simplification of the expressions (6.24) for the coefficients α to ε, but the quartic has some properties of special interest. This is because the vector $\boldsymbol{Y}$ is in the plane of incidence. It is useful to draw a diagram like fig. 6.1 containing cross sections of the refractive index surfaces, and in it the direction of $\boldsymbol{Y}$ can be shown as a line through the origin. Fig. 6.3 gives an example. On this line are the four window points P, Q, P_2, Q_2 where $n = \pm\{Y/(Y \pm 1)\}^{\frac{1}{2}}$, though only two are real if $Y < 1$. On it also are the two points K_1, K_2 where $n = \pm 1$, $\Theta = 0$. These six points are all on the refractive index surface when $X = 1$.

In this diagram a reference line is now drawn parallel to the n_z axis, where $n_x = S$. For a given value of X the cross sections of the two refractive index surfaces are then drawn. Where they are cut by the reference line, the four values of n_z are the required values of q. To plot curves of q versus X we must imagine two sequences of these cross sections to be drawn, one for the ordinary and one for the extraordinary wave, for a sequence of values of X. The form of the q–X curves depends on the value of S chosen for the reference line. Transition cases occur when it goes through any of the six points P, Q, P_2, Q_2, K_1, K_2. Here we shall consider only positive values of S. The transition values, marked in fig. 6.3, are:

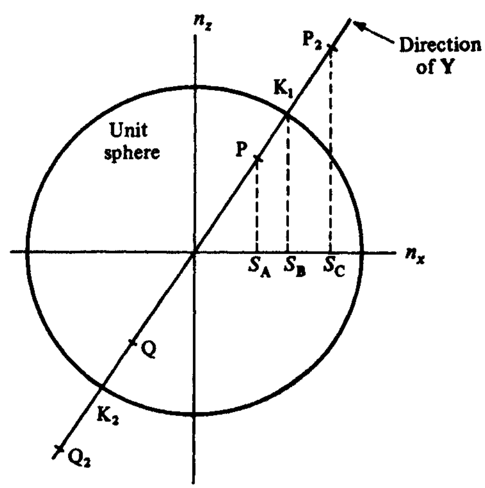
$$\begin{aligned} S_A &= \{Y/(Y+1)\}^{\frac{1}{2}} l_x \quad && \text{point P;} \\ S_B &= l_x && \text{point } K_1; \\ S_C &= \{Y/(Y-1)\}^{\frac{1}{2}} l_x && \text{point } P_2. \end{aligned} \tag{6.47}$$

S_C is real only if $Y > 1$. Its value can then be > 1.

For a wave (6.1) incident on the ionosphere from below, the angle of incidence is usually real so that S is real and < 1. A wave with S real and > 1 could not easily be launched from the ground. It could be emitted by a transmitter on a space vehicle within the ionosphere. If the transmitter is in a region where one value of n^2 is > 1, it could emit a wave with $1 < S^2 < n^2$, whose wave normal would have a real direction. Curves of q vs X for $S > 1$ are therefore of some interest but they are not considered further here. An example of a radio signal with $S > 1$ is the subprotonic whistler; see end of § 13.8 and Walker (1968a, b).

Fig. 6.4 shows curves of q vs X for a case where $1 > S > S_B$, and $Y < 1$. For $X = 0$

Fig. 6.3. Similar to fig. 6.1. Shows the plane of incidence $n_y = 0$ in refractive index space for the special case when the earth's magnetic field is in the plane of incidence. The four points P, Q, P_2, Q_2 are window points as described in § 4.11, 4.12, 5.2 item (7). The general form of the curves of q vs X depends on the value of S and changes when S passes through any of the transition values S_A, S_B, S_C.

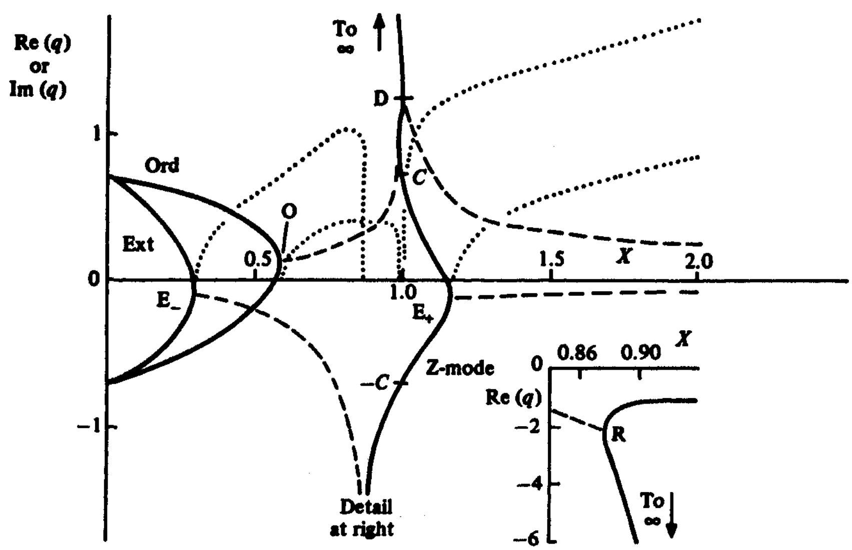


both refractive index surfaces are the unit sphere and the four values of q are in two pairs $q = \pm C$. As X increases, both refractive index surfaces shrink but that for the extraordinary wave shrinks faster (compare fig. 4.3). A value of X is reached where it just touches the reference line. Here the two qs for the extraordinary wave are equal and the tangent to the q–X curve is vertical as shown by (6.30). This is the coupling point E_- of § 6.6. For slightly larger X the two qs for the extraordinary wave are a complex conjugate pair. Similarly, for a still larger X, the reference line touches the ordinary refractive surface and the two qs for the ordinary wave are equal. This is the coupling point O of § 6.6.

The equal qs at E_- and at O have opposite signs, because the ordinary refractive index surfaces are prolate, fig. 5.4, and the extraordinary are oblate, fig. 5.7, as is shown by item (12) of § 5.2.

For $X > 1 - Y^2$ the refractive index surface for the extraordinary wave is again real, and $n > 1$. This is for the Z-mode. The surface is open and has reversed resonance cones (§ 5.2 item (11), figs. 5.8, 5.9). When X reaches the value X_∞ given by (6.26) where one root q is infinite, the reference line is parallel to the generator of the resonance cone, that is to the asymptote of the cross section of the refractive index surface in the magnetic meridian plane. The reference line then cuts the surface at

Fig. 6.4. Shows how the four values of q depend on X for propagation from magnetic north to south or south to north. In this example $Y = 0.5$, $l_x = 0.5$, $S = 0.7071$ and electron collisions are neglected. One q is infinite where $X = 0.9231$. Where two qs are a complex conjugate pair, the real part is shown as a broken line and the imaginary part as a dotted line. The points C, $-C$ are where $X = 1, q = \pm C$. The point D is the double root where $X = 1, q = Sl_z/l_x$ ($= 1.225$ in this example). The curve for X near to 1 is shown in more detail in fig. 6.6(f).



infinity and in at least one other point. This follows very simply from the geometry of the surface and its asymptotes. It was also proved in § 6.4 just after (6.26). Hence, when X is decreased, the refractive index surface for the Z-mode must cut the reference line at two points giving two real values of q for some range of X less than X_∞. This can be seen in figs. 6.4, 6.5.

Fig. 6.5 shows curves of q vs X for a case where $S < S_A$ ((6.47) and fig. 6.3), and $Y < 1$. The reference line now crosses the line segment PQ (fig. 6.3) and this is the limiting form of the refractive index surface for the ordinary wave when $X = 1$ (§ 5.2 item (7) and fig. 5.4). Hence the q–X curve for the ordinary wave extends to the value $X = 1$, and here it has a vertical tangent because there is a double root. This applies for any S in the range $-S_A < S < S_A$, but only when $l_y = 0$. For any other l_y the curve for the ordinary wave has a vertical tangent where $X < 0$.

The behaviour of the curves for X near to 1 depends on S and is shown in the six diagrams of fig. 6.6. This is for $Y < 1$ and for positive l_x. The quartic has three positive roots. There is one negative root, not shown in fig. 6.6, but visible in figs. 6.4, 6.5.

For $Y > 1$, the $q - X$ curve for the ordinary wave where $X < 1$ is similar to that for $Y < 1$. But the curves for the extraordinary wave are different in form and there is another branch of the curves for the ordinary wave corresponding to the whistler mode. The transition point S_C, (6.47) is now real. A full description would need many

Fig. 6.5. The same as fig. 6.4 except that the value of S is 0.258 819. The curve for the ordinary wave now extends to $X = 1$, and the line $X = 1$ is a tangent to it at the point O. Here there is a double root where $q = Sl_z/l_x$ ($= 0.448$ in this example). The curve for X near to 1 is shown in more detail in fig. 6.6(a).

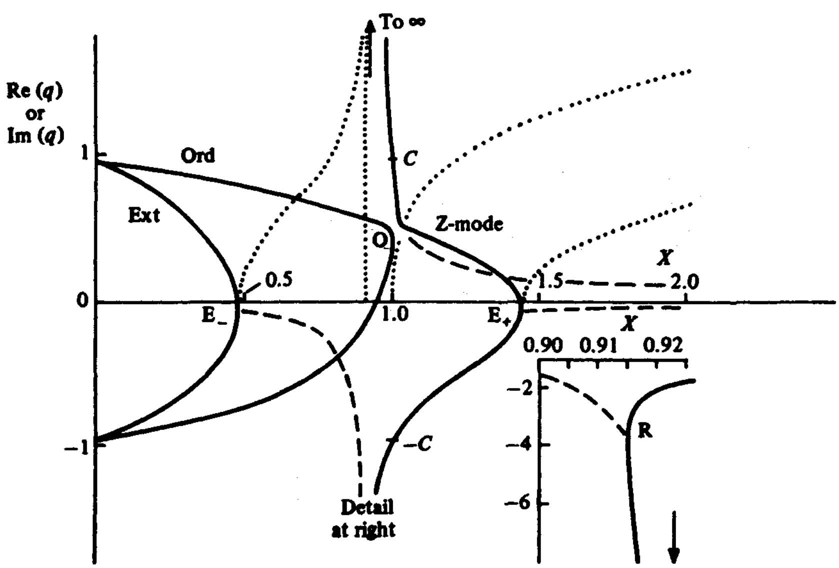


Fig. 6.6. Shows how q depends on X for various values of S and for X in a small range near $X = 1$. The other parameters are the same as in figs. 6.4, 6.5. Only three of the four qs are shown. The remaining q is real and negative in all six cases. The point C is where $q = C$, $X = 1$. The point D is where $q = Sl_z/l_x$, $X = 1$. Here there is a double root, and the tangent to the curve is vertical. In (e) the points C and D coincide and there is a triple root. The transition values (6.47) are $S_A = 0.2887$, $S_B = 0.5$. In (a) $S_A > S = 0.258\,819$ as in fig. 6.5. In (b) $S_A > S = 0.28$. In (c) $S = S_A$. The reference line goes through the window point P. In (d) $S = 0.3$; $S_A > S > S_B$. In (e) $S = S_B$. In (f) $S_B < S = 0.7071$ as in fig. 6.4. The alternative labels O, C_1, C_3 are the coupling points described in § 16.5.

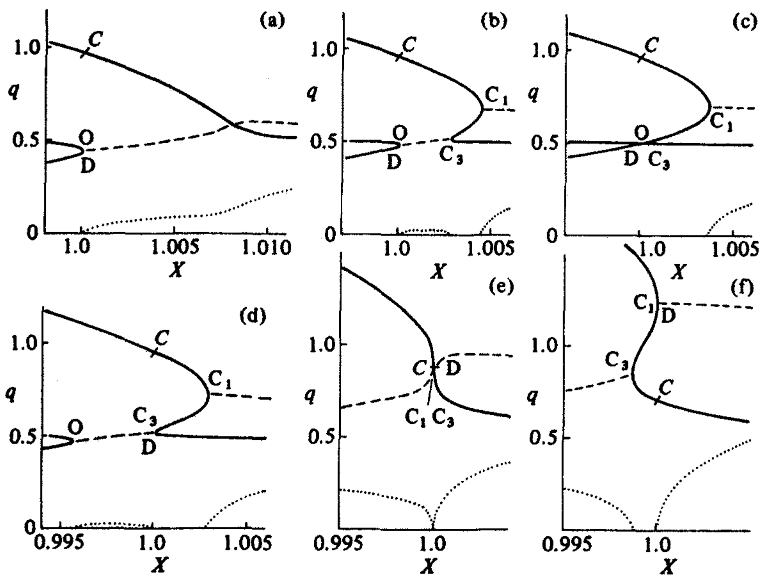
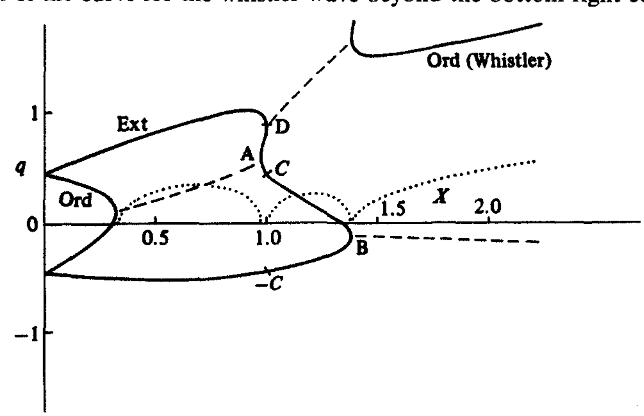


Fig. 6.7. Shows how the four values of q depend on X for propagation from magnetic north to south or south to north. In this example $Y = 2$, $l_x = 0.707$, $S = 0.9$ and electron collisions are neglected. One q is infinite where $X = 2.998$ (not shown here). Other details are as in fig. 6.4. Here $Sl_z/l_x = 0.9$. There is another branch of the curve for the whistler wave beyond the bottom right corner.



curves and would require too much space. For some examples see Booker (1939), Chatterjee (1952), Millington (1951, 1954), Budden (1961a). Here we must be content with just one example, fig. 6.7.

When $0 < X < 1$ and $Y > 1$ the refractive index for the extraordinary wave is greater than unity (see § 4.12 and figs. 4.5, 4.6). Hence the curve of q vs X for the extraordinary wave is outside that for the ordinary, and $|q|$ increases at first when X increases from zero. The curve always extends to $X = 1$ and beyond it, provided that $S < 1$. The points where $X = 1, q = \pm C$ always lie on it. If the reference line of fig. 6.3 cuts the line segment P P_2, that is $S_A < S < S_B$, (6.47), there is a point D on the curve where $X = 1, q = Sl_z/l_x$ and here there is a double root and the tangent to the curve is vertical. This case is illustrated in fig. 6.7. For a small range of X slightly less than 1 all four roots q are real and on the curve for the extraordinary wave. This is possible because here the refractive index surface has two points of inflection, fig. 5.10, and can be cut four times by the reference line. In this case, therefore, the extraordinary wave curve has three points where the tangent is vertical, representing three levels of reflection. It is used later, fig. 10.2, for illustration. If S were equal to S_B, (6.47), the points C and D would coincide and there would be a triple root where $X = 1$.

For larger X the reference line again cuts the refractive index surface for the ordinary wave where $n^2 > 1$, that is the whistler mode. One branch of the resulting $q - X$ curve is in fig. 6.7. If $S < S_A$, the ordinary wave curve for $X < 1$ extends up to $X = 1$ and the double root at D lies on it. If $S > S_C$, the branch for the whistler mode extends down to $X = 1$ and the double root at D lies on it.

6.9. Effect of electron collisions on solutions of the Booker quartic

Figs. 6.8, 6.9 are examples of how the four qs depend on X when the effect of electron collisions is allowed for. Fig. 6.8 is for the same conditions as in fig. 6.4 except that now $Z \neq 0$. The qs are now all complex when $X \neq 0$ so that curves of both Re(q) and Im(q) are shown. Only real values of X are used. When $X \neq 0$ no value of Im(q) is ever zero. Two values are negative and apply to upgoing waves; the curves are labelled 1, 2 in figs. 6.8, 6.9. The other two values are positive and apply to downgoing waves; the curves are labelled 3, 4. This property must always hold when X is real, as was proved in § 6.4.

Each of the coupling points E_-, E_+, O is a reflection point where two qs are equal, one for an upgoing and one for a downgoing wave. This cannot now occur for real X because the two qs must have non-zero Im(q) with opposite signs. The point of resonance $X = X_\infty$, now given by $\alpha = 0$ in (6.23), is no longer on the real X axis. Thus there are no values of Re(X) where one q is infinite, and in fig. 6.8 there are no values where two qs are equal, and no vertical tangents.

It is possible, however, for two qs to be equal when X and Z have real and non-zero values X_c, Z_c respectively. This is because the discriminant (6.38) can be zero for real

X, Z. The two equal qs must then be either both for upgoing or both for downgoing waves. This condition occurs when one of the coupling points C_1 to C_4 of § 6.6 lies on the real X axis. The values of X_c, Z_c were fully described by Pitteway (1959) who gave curves for various values of S, Y and l_x, l_y, l_z. Fig. 6.9 gives an example of q–X

Fig. 6.8. Shows how the four values of Re(q) and Im(q) depend on X when electron collisions are allowed for. This example is for propagation from magnetic north to south or south to north, and $Y = 0.5$, $l_x = 0.5$, $S = 0.7071$ as in fig. 6.4. But here $Z = 0.05$. The curves are identifiable by the numbers 1 – 4 which are the same for Re(q) and Im(q). If collisions were neglected, one q would be infinite where $X = 0.9231$.

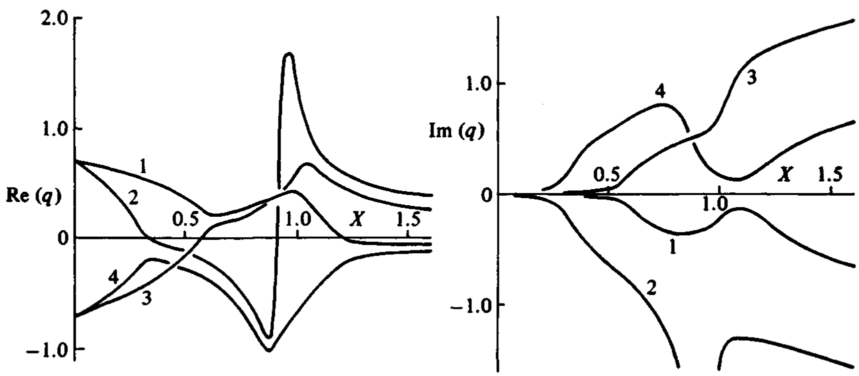
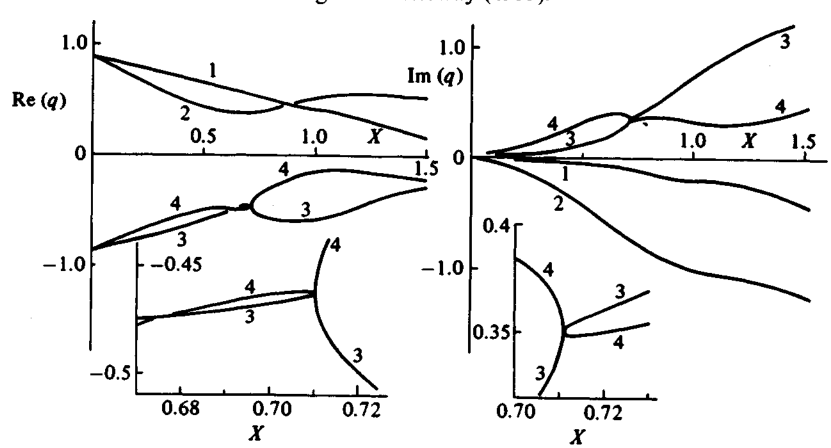


Fig. 6.9. Shows how the four values of Re(q) and Im(q) depend on X when electron collisions are allowed for. In this example $Y = 0.5$, $S = 0.5$, $l_x = l_y = 0.353\,553$, and $Z = 0.274 = Z_c$. If collisions were neglected one q would be infinite where $X = 0.9231$ but the effect of this resonance is masked by the large Z. Two values of q, for the two downgoing waves, are equal where $X = X_c = 0.711$. Here the tangents to curves 3 and 4 are vertical. The values of X_c, Z_c used here were found from fig. 3 of Pitteway (1959).



curves for one of these cases, in which $Z_c = 0.274$, $X_c = 0.711$. It can be seen that the tangents to the curves are vertical for both Re(q) and Im(q) at the coupling point, as they must be because $\partial q/\partial X \to \infty$.

6.10. The electromagnetic fields

Consider one of the thin strata described in §6.2. In it there are in general four progressive waves corresponding to the four values of q. They are called the four 'characteristic waves' for the level z of the stratum. The fields of all these waves depend on x and y only through the factor (6.2) and we now assume that $S_2 = 0$, $S_1 = S$. Hence for all field components at all heights z

$$\frac{\partial}{\partial x} \equiv -ikS, \quad \frac{\partial}{\partial y} \equiv 0. \tag{6.48}$$

For any one of these waves, we have the further property

$$\frac{\partial}{\partial z} \equiv -ikqz \tag{6.49}$$

but the qs are in general different for the four waves. Maxwell's equations (2.45) and the constitutive relation $\boldsymbol{D} = \varepsilon_0 \boldsymbol{\varepsilon} \boldsymbol{E}$ may now be applied to any one of the waves to give

$$qE_y = \mathscr{H}_x, \quad qE_x - SE_z = \mathscr{H}_y, \quad SE_y = \mathscr{H}_z \tag{6.50}$$

$$\left.\begin{aligned} q\mathscr{H}_y &= \varepsilon_{xx}E_x + \varepsilon_{xy}E_y + \varepsilon_{xz}E_z \\ q\mathscr{H}_x - S\mathscr{H}_z &= \varepsilon_{yx}E_x + \varepsilon_{yy}E_y + \varepsilon_{yz}E_z \\ S\mathscr{H}_y &= \varepsilon_{zx}E_x + \varepsilon_{zy}E_y + \varepsilon_{zz}E_z. \end{aligned}\right\} \tag{6.51}$$

These six equations determine the ratios of the six components of $\boldsymbol{E}$ and $\boldsymbol{\mathscr{H}}$. From them E_z and $\mathscr{H}_z$ may be eliminated to give the ratios of the four components

$$E_x, -E_y, \mathscr{H}_x, \mathscr{H}_y. \tag{6.52}$$

The reason for using $-E_y$ is explained in §§7.14(3), (8). Explicit expressions for these ratios are given later, §7.16, equations (7.118), (7.134). They depend on q so that there are four sets of the ratios (6.52) and they are written as the columns of a 4×4 matrix $\mathbf{S}$. The four columns are linearly independent provided that the four qs are distinct, and then $\mathbf{S}$ is non-singular; see §7.14(1). Equations (6.50), (6.51) determine only the ratios of (6.52) so that each column of $\mathbf{S}$ may be multiplied by any constant. The choice of the constants depends on the particular problem but there is one choice that has special advantages. This is explained in §7.14(7), (8).

Suppose now that, at any height z, the set of values (6.52) is given for the total field. Let it be written as a column matrix $\mathbf{e}$. Then it can always be expressed as the sum of the fields of the four characteristic waves, with amplitudes f_1, f_2, f_3, f_4; these are written as a column matrix $\mathbf{f}$. Thus

$$\mathbf{e} = \mathbf{S}\mathbf{f}; \quad \mathbf{f} = \mathbf{S}^{-1}\mathbf{e}. \tag{6.53}$$

From a knowledge of **S** and **e**, the amplitudes **f** can be found. This still applies even when we do not use the concept of discrete strata. If the fields (6.52) are given for any z in a continuous stratified medium, the amplitudes **f** can be found from (6.53). This is equivalent to considering a fictitious homogeneous medium with the properties of the actual medium at the height z, and finding what amplitudes **f** of the four characteristic waves would be needed to give the actual fields. The discussion of this topic is resumed in §§ 7.14, 16.2.

The amplitudes **f** play an important part in the theory in this book, and frequent reference is made in later chapters to the transformations (6.53).

PROBLEMS 6

Note. Problems 6.1, 6.2 below may interest those readers who like to pursue algebraic details. The results are useful when tracing curves of q or q^2 vs X. They have no particular practical application in radio propagation problems.

6.1. For a collisionless electron plasma, when $X = 1$, the Booker quartic has two solutions $q = \pm C$. Show that here the curve of q vs X has slope

$$\mathscr{S} = \partial q/\partial X = \mp\{2Y^2C(Cl_x \mp Sl_z)^2\}^{-1}.$$

For propagation from magnetic north to south or south to north ($l_y = 0$), S is given the value $l_x\{Y/(Y+1)\}^{\frac{1}{2}}$ (as in fig. 6.6(c)) so that two branches of the $q - X$ curves intersect where $X = 1$. Show that their slopes $\mathscr{S}$ are solutions of the quadratic equation

$$Yl_x^2\mathscr{S}^2 - 4l_z\{Y(Y+1)\}^{\frac{1}{2}}\mathscr{S} - 2 = 0.$$

6.2. For propagation from magnetic east to west or west to east, the Booker quartic is a quadratic equation for q^2 and, when collisions are neglected, q^2 is zero where $X = C^2$. Show that here the slope $\partial(q^2)/\partial X = -(1 - l_z^2C^2)^{-1}$. The angle of incidence is now chosen so that the quadratic equation has equal roots for two coincident real values of X. Show that $l_z^2S^4 = Y^2(1 - C^2l_z^2)^2$ and that equal roots q^2 occur where $X = \frac{1}{2}(1 + C^2)$, $q^2 = -\frac{1}{2}S^2(1 + l_z^2C^2)/l_y^2$. Here two curves of q^2 vs X intersect. Show that their slopes $\partial(q^2)/\partial X$ are $(1 - l_z^2C^2)/l_y^2$ and $-(1 - l_z^2C^2)\{2(1 + l_z^2C^2) + l_y^2\}/l_y^4$.

6.3. Show that, for propagation near the magnetic pole, the Booker quartic is a quadratic equation for q^2 and find its solutions for a collisionless electron plasma. Show that the two solutions are equal when $X = 1 - \frac{1}{2}S^2\{1 \mp (1 - Y^2)^{\frac{1}{2}}\}$ and that then $q^2 = \frac{1}{2}(1 + C^2)\{1 \pm (1 - Y^2)^{-\frac{1}{2}}\}$. Note that these are real only when $Y < 1$, and then only the upper signs give a positive value of q^2. Sketch the refractive index surface for this case and show that it applies for the Z-mode. Sketch the curves of q^2 vs X.

6.4. Show that if the ionosphere is an electron plasma and X is very small, the two values of q for obliquely upgoing waves are given by $q - C \approx bX$ where b is a solution of the quadratic equation

$$4C^2U(U^2 - Y^2)b^2 + 2\{CY^2(l_zC + l_xS)^2 - CY^2 + 2CU^2\}b + U = 0.$$

Check that this gives the correct results for the special cases (a) $S = 0$, $l_z = 0$, and (b) $S = 0$, $l_z = 1$.